

Assessing reading comprehension from the perspective of test-takers: A qualitative approach

Karim Sadeghi, PhD

Assistant Professor, Urmia University

Abstract

Whilst, it has been the norm to expect experts in assessment to worry about and offer solutions to issues in testing, the concerns as well as the rights of major stake-holders themselves, i.e., testees, have for the most part been neglected. Although test-takers' views may be regarded as naïve and non-specialist, no doubt obtaining an emic perspective will illuminate the path for testers and researchers and will help them to avoid pertinent biases. With such a focus, the present study takes a qualitative orientation and looks at the ideas, concerns and expectations of 20 undergraduates in EFL interviewed about their experiences in taking a variety of reading comprehension tests. The main finding to emerge from the study is that most reading tests fail the test of judgmental validation. Further findings and implications are discussed in the paper.

Key words: assessment, reading comprehension, EFL, Iranian context, qualitative approach

1.Introduction

Reading is one of the 'most complex forms of information processing' (Kolers, 1973: 29) and is probably the 'most extensively researched' language skill (Bachman, 2000: x). Developing reading comprehension in English is the most important skill focused on in Iran's state secondary and tertiary education as well as in other EFL contexts where learners have limited access to spoken communication. Assessing the material taught and giving the necessary feedback to learners as to the effectiveness of learning is an integral part of

teaching in educational settings. Attention should therefore be paid to whether the task of assessment is carried out properly.

The degree of the students' mastery of reading ability is tested through different testing techniques, the most wide-spread among which are multiple-choice and true-false questions as well as cloze procedure. While using such testing techniques for testing reading comprehension is a normal practice for the majority of test-makers (as the following review indicates), this research, with a qualitative orientation, was intended to find out whether test-takers thought that the tests they were given in various contexts could measure the degree of their comprehension of a passage effectively. It should be reminded that 'reading comprehension' and 'reading ability' are quite distinct traits: The focus of this study is on measuring text comprehension rather than on assessing general reading ability, and comprehension in a particular setting may or may not be indicative of one's overall reading ability.

2. Review of the Related Literature

As far as components of reading ability are concerned, and based on the divisibility view of language skills, reading is composed of several constructs. Constructs are not 'psychological real entities' but abstract concepts to be defined for test-design purposes (Alderson, 2000: 119). So different reading tests may have different numbers and types of constructs (Alderson, 2000). Thus, in one testing situation, the focus may be on a particular construct or a set of constructs, and, in another setting, on a different construct or a different set of constructs, and 'it is unlikely that in any given testing situation we will want to test every component in the theory' (Alderson, 2000: 137). Indeed some constructs like appreciation, enjoyment and personal response may not be measurable at all in any traditional way. Reading ability has been claimed to include such constructs as intelligence (Flahive, 1980; Farr, Carey, & Tone, 1986), reasoning ability (Johnston, 1983; Farr et al., 1986), inferring, and problem-solving (Johnston, 1983; Alderson, 2000). Specific reading comprehension may not, however, have the same number and type of constructs as general reading ability. According to Flahive (1980: 34) 'there is no universally accepted definition of language comprehension,' an assertion which applies to both reading and listening comprehension. Similarly, Johnston (1983:

1) asserts that trying to define reading comprehension is 'a bold move'. However, for the purpose of the present study, reading comprehension is considered to have at least two constructs, namely 'general' and 'detailed' comprehension. Higher level constructs such as 'evaluative' comprehension are excluded as they may well go beyond comprehension ability and include the ability to think critically, which falls outside the concern of this research. Lower-level constructs, such as understanding graphemic-phonemic relationships, are also excluded on the grounds that they do not involve real comprehension.

Despite the fact that attempts to understand the true nature of reading comprehension have not been conclusive and that there is no consensus on the nature of reading comprehension, testers and teachers have used a variety of tools to assess what they assume is the product of the process of reading, i.e. comprehension. The following is a brief survey of some of the most common ones, with a particular focus on those practiced by Iranian teachers and/or testers of EFL reading comprehension.

Throughout the history of language testing, numerous kinds of tests (including sentence comprehension, multiple-choice, matching, different forms of cloze, completion, True-False to name a few) have been used to test reading comprehension without much concern for whether they really measured what they intended to (A complete list and survey may be found in Heaton, 1988 and Farhady, Jafarpur, & Birjandi, 1995). However, having a special label 'does not guarantee that the test actually measures whatever is named' as Farhady (1983: 12) emphasises. Furthermore, even if a test is valid, performance on it may not be the same as performance in real-life situations, and therefore, as Alderson (2000) and McNamara (2000) claim, answering multiple-choice reading comprehension questions may not show how well one performs in real life.

One major problem with reading tests is the involvement of other skills such as writing in answering test questions (Aslanian, 1985; Harrison, Bailey, & Foster, 1998). Namely, the reader may have understood the text but have been unable to express his/her comprehension through another modality. The counter-argument is that making testing reading 'pure' is not authentic and is a distortion that 'biases measurement' (Alderson, 2000: 26). The problem with

questions used to test reading is that they are either too easy or too difficult. While some questions are more complex to understand than the text itself (Dehn, 1984: 97), others give clues to correct answers without the reader actually understanding the text (Aslanian, 1985). Another criticism against reading tests is that they are mainly measures of other traits in addition to reading comprehension. For example, reading tests have been criticised as being measures of intelligence (Flahive, 1980), memory and recall (Dehn, 1984), and world knowledge (Rosowsky, 2000).

Alderson (2000: 28) questions the overall authenticity of testing because 'the very act of assessing changes the sociocultural nature of the event.' Accordingly, concerns have been raised over whether informal classroom-based measurements can be more effective than formal tests carried out under constrained conditions (e.g., Lorch & van den Broek, 1997). Anxiety created by the testing situation and time-constraints under which readers operate (Walczyk, Kelly, Meche, & Braud, 1999) have also been the sources of concern for some. The main objection of cognitive psychologists to reading tests is that they try to measure the product of reading rather than the process (Johnston, 1983; Farr et al., 1986; Myers, 1991). This criticism, however, does not seem valid because it is a matter of interest to measure reading as a process or a product, and as Alderson and Urquhart (1984: xix) believe, concentration on 'reading as a product may help us for testing purposes.'

Whatever stance we take on the nature of reading comprehension, the fact is that it is assessed of necessity at least for educational purposes. This study was aimed to explore whether the measuring instruments used for testing reading comprehension are valid in the eyes of the test-takers themselves, who are the main stake-holders in the assessment process.

3. Method

3.1 Participants

The participants of this study came from undergraduates majoring in English at Urmia University. The candidates who participated in a quantitative phase of the study (231 in number) were asked to provide their contact details if they were willing to attend an interview session. 22 volunteers were randomly selected and invited for an interview. A

mutually convenient time was set up for each interview session. Out of 22 candidates invited, only 20 attended interview sessions. The age range of the interviewees was between 18 and 26 and 70 per cent of them were males.

3.2 Sources of data

The major instrument used for the purpose of data elicitation was a semi-structured interview, and as is the norm in semi-structured interviewing (Mackey and Gass, 2005), the researcher had already prepared the topics to be discussed, with a broad range of question headings. The order, type, and number of these specific questions varied from one interview context to another. Since the main purpose of the research was to investigate the validity of tests of reading comprehension as far as test-takers were concerned, the interview began by asking interviewees about what reading comprehension itself was before proceeding to the main focus of the interview. Most questions in the interview were intended to target the validity of cloze tests and other tests of reading comprehension as measures of EFL reading comprehension. For this purpose, the interview included questions on how the interviewees' reading comprehension was tested in real-life and in academic settings, whether they had the experience of sitting reading tests and what they thought such tests really measured, and whether the score they got on such tests showed their real degree of comprehension.

All interviews were conducted in a relatively silent room at the School of Literature and Humanities of Urmia University. Each interview lasted between 25 and 35 minutes. The first interview began at 8.30 a.m. and the following interviews started at 9.30, 10.30 and 11.30 a.m. The afternoon interviews started at 16.00, 17.00, and 18.00. At the beginning of each session, the purpose of the interview was explained to the participants and permission was asked to tape-record the interviews. During interview sessions, the interviewer and the interviewees sat on comfortable chairs at a 90° angle position, with the tape-recorder on a coffee-table which stood between the two. A microphone was attached to the tape-recorder to produce a higher quality recording. Candidates were asked to choose the language of interview. 11 chose to be interviewed in English and 9 in Farsi.

In order to analyse the interview data, the tape-recordings were transcribed first. During transcription, it was found that one Farsi

interview had not been recorded at all, and only about 20% of an English interview had been recorded. A few other tapes had some small unrecorded pieces as well. It was decided that Farsi transcriptions will only be necessary if they contained new ideas not included in English transcriptions. Accordingly, I listened to all Farsi interviews and transcribed all those parts which contained new information not discussed by other participants interviewed in English. Out of 11 English interviews, 10 interviews were transcribed in full. I did my best to do a kind of transcription that would involve the least degree of meaning loss. Namely, because the researcher was looking for content and ideas, it was not decided to do a phonetic-type of transcription where all meaningful and non-meaningful sounds and words would be transcribed. In the case of unintelligible words or phrases, a best guess was made based on what I remembered from the interview session or/and what the context offered.

4. Findings and discussion

To begin the analysis of transcribed data, I read all interview transcripts both with and without the interview questions and identified two general categories as follows:

- 1) Issues related to reading comprehension such as its meaning, how one shows his/her comprehension, the factors affecting reading comprehension, the differences and similarities between real-life and academic reading; and
- 2) Issues related to assessing reading comprehension such as how reading comprehension is tested in real life as opposed to academic settings, how reading comprehension should be tested and the meaning of scores on reading tests including cloze tests.

The following analysis deals with the assessment-related issues only, i.e. the second category above. All quotations from interviewees have been italicized.

1. How reading comprehension is tested in real life

When asked if their comprehension was tested in real-life reading, more than half of the interviewees gave negative answers. They had not thought of testing their comprehension in real life, and they thought that if they got the overall meaning, they did not need to worry about anything else as one student explained, *'I don't think about testing myself; I just read and go on, and if I understand I enjoy my reading'*.

For most of the candidates 'testing' was something to do with academic reading. When they were asked if their 'real-life' reading comprehension was tested and how, they began to talk about the way their reading was tested in academic settings rather than in 'real-life' reading. Very few interviewees thought they tested their comprehension in real-life settings, and those who did so, tested their understanding only occasionally; and the way they tested their comprehension was actually a kind of follow-up activity to practice their language skills. In general, no formal testing seemed to be used to check comprehension in real-life settings except for one or two cases. It seems that because the main purpose in real-life reading is pleasure and enjoyment, fewer readers tended to test their comprehension as it may intervene with the pleasure they were after. Those who did test their comprehension regarded real-life reading as more than simply a means of pleasure but as a way of gaining new information to strengthen their knowledge of language.

The following test techniques were mentioned by those who thought they tested their comprehension in real-life reading: discussing the passage with a friend or a family member who also knew English; retelling it to others; writing a summary; making an oral summary, recording it on a tape, listening to it, and matching it against the original passage; expressing relevant feelings; answering others' questions about the passage; and trying to remember the major points after reading the text, and referring to the text a few days later to see what one remembered. *'I think imagination is the key to comprehension'* commented one of the candidates, by which he meant comprehension would take place if one were able to keep the content of the passage in his memory. None of the techniques cited above was actually applied by candidates to test their comprehension; they rather talked about what they *could* do to test their real-life comprehension.

2. How reading comprehension is tested in academic settings.

The candidates were asked about whether and how their comprehension was tested during reading classes and in final-term examinations at the university. The interviewees were more or less unanimous that their comprehension of the passages they read during their courses was checked through comprehension questions which appeared at the end of each passage in their textbooks. As to final exams, some interviewees had not yet taken any final exams as they were freshmen, and they talked

about their exams at high schools or private language institutes. Those who had already experienced final exams at the university talked about more or less similar techniques used to test their reading comprehension.

In university reading courses, where the focus of the course was on developing reading 'skill', teachers were claimed to have done little to check the comprehension of the passages other than asking the questions which appeared after each text. Furthermore, such comprehension checking was done during the next session when students had read the text several times at home, as one of the interviewees made clear: '*... our professor asks the students to reading the unit at home and solve the reading questions after if for the other session.*' Different candidates mentioned different test techniques used in course books including yes/no questions, True/False (T/F) questions, multiple-choice (m/c) questions, cloze-type questions, short- and long-answer questions, synonyms/antonyms, and referential questions. The questions they talked about dealt with general ideas, detailed information, vocabulary and grammar. Accordingly, although the questions were intended to test comprehension, they tested both below and beyond comprehension.

In reading classes outside the university, some teachers asked students to summarise the text. Sometimes, teachers would ask students questions not included in the 'comprehension check' exercises. This last kind of comprehension checking seems to be better than asking students simply 'to do the exercises' after the text and checking the answers the next session. In the latter case, some may copy the answers from others or from old books without really comprehending the text, and leaving questions to be answered at home after students read the text several times may not be called 'testing comprehension' in the real sense of the word, but rather a kind of exercise to foster comprehension skills. It should be noted that 'reading' courses are intended to develop 'reading skill', and therefore, focus not only on 'comprehension skills' but also on grammar and vocabulary, which are building blocks of all language skills. Therefore, the exercises which appear after reading passages in course-books, and also in the final exams, seem to target not only comprehension of the text but also grammatical knowledge and vocabulary.

Regarding final exams, the interviewees believed that their comprehension was tested more or less in a similar way as that used during the course. The techniques used to test reading comprehension in final exams included: T/F questions, m/c questions, cloze tests, short- and

long-answer questions, and filling in the blanks. As explained above, because the tests are mainly intended to test 'reading skill' rather than 'reading comprehension', not surprisingly, many of the interviewees believed that the tests tested both comprehension and also vocabulary and grammatical knowledge. The 'comprehension' part of reading, however, does not seem to have been properly tested simply because, based on what the candidates stated, in most cases, the texts used in final exams had been exactly the same texts as those already studied during the course, and in some cases the questions had strangely enough been the same questions as those in the book. Not all final exams the students talked about, however, used seen texts. As is the norm in testing reading comprehension, a few final exams used 'unseen' texts to check real comprehension. While most of the subjects thought that having 'seen' texts in final exams was an advantage and made the test easier, others consciously argued that such tests could not measure comprehension but memorisation.

Different interviewees favoured different aspects of tests in their final examinations. One subject, for example, stated that her reading score at final exams did not show her real reading comprehension because it was based on T/F and m/c questions. She suggested '*some kind of understanding questions*' to be used instead. Another candidate criticised comprehension questions in the final exams because they were only 'related' to the text and the answers to such questions could not be found in the passage: '*I don't know where I have to find the answers: some questions are more or less related to the text and for others you need to be a good predictor.*' What this subject seemed to convey was that most of the questions were inferential rather than referential or display questions (Nation, 2009). One candidate, who was talking about how their reading comprehension was tested at high school using an 'unseen' passage, stated that students reacted unwillingly to the 'unseen' passage, saying that it was not in their book and they did not know some of the vocabulary. Some others who also talked about 'unseen' texts at the university final exams believed that the texts were far more difficult than those in the book, and that it was a way for some teachers to 'show off' and 'combat' the students: '*Our teacher Mr ... wants to annoy students at exam times; he does not teach very good but in final examination he wants to show himself off.*'

Although including inferential questions in testing reading comprehension is recommended, the right balance should be observed between different question categories. Similarly, having an 'unseen' text for testing reading comprehension is highly desirable; however, the 'unseen' text should be selected with much care so that the difficulty level would not be much higher than those already studied. Furthermore, it is better for the 'unseen' text to have familiar vocabulary items with only a few new words which could easily be guessed from context, and it is also better to include topics and issues with which students are somewhat familiar. It will not be advisable, for example, to choose a text about chemical engineering, even with low readability, for EFL subjects because the concepts discussed in the passage will require some specialist knowledge which most of the students probably do not have. On the other hand, selecting very simple and familiar texts, the questions of which can be answered without even referring back to the text itself, is not recommended.

The problem of keeping the right balance of difficulty for all test-takers is persistent, however, because, as some candidates expressed in their interviews, a text which was of optimal difficulty, and perhaps a bit difficult for others, was easy and unchallenging for them. The problem was more obvious in the case of high-school final exams, where some students did not find reading comprehension tests challenging as they had attended extra-curricular classes and were, therefore, more proficient compared to the rest of their group. Although such a problem may be solved using placement exams at the beginning of the course, such a solution is not possible in Iran's education system, because one's admission to a field in a university (English, for example) is not based only on his/her proficiency in English but on many other knowledge areas which may not be related to English at all.

The attitudes of candidates towards testing techniques were also varied. For example, while one subject argued that m/c questions were very good for testing comprehension because '*choices are usually so similar and if one does not really understand the text, he cannot choose the correct answer*', another stated that such questions could not test comprehension properly because '*when you understand a text, you don't have to look for some choices*'. One candidate criticised m/c questions used in final exams because he could not find any of the choices in the

passage. He thought that some questions tested something more than comprehension, such as one's cleverness and puzzle-solving ability. Despite this weakness, he still believed that m/c questions were a good comprehension testing technique. Similarly, while one subject stated that T/F questions were so simple and unable to test comprehension, another believed that such questions, especially in negative forms, required concentration and needed good comprehension to answer.

One of the techniques used in testing reading (comprehension) mentioned by one subject was that the teacher gave them some ideas and asked them to write which lesson they were related to in their textbook. There is great doubt in whether such a technique, and some others in which exact texts and questions are used from the textbook in final exams, can function as effective tools in testing reading (comprehension) or even in fostering reading skills. The backwash such techniques can offer teaching will undoubtedly be detrimental (Brown, 2004).

3. How reading comprehension should be tested

Most of the interviewees talked about writing a summary, retelling, explaining the text and answering open-ended questions as appropriate ways to test comprehension. Indeed, not all of them preferred a common single technique, and individuals differed in their choice. While a few preferred easier and closed techniques like m/c questions and cloze or other blanked tests, many preferred open-ended techniques such as answering questions or summarising.

One candidate favoured '*freer response*' over m/c questions, by which she meant writing answers to questions, believing that the process of writing itself would not interfere in conveying the understood message. A few others who preferred oral answers to comprehension questions, however, admitted that the spoken language itself may be a barrier for them in communicating realistically whatever they comprehended, believing that summary would not show their exact degree of comprehension. To prevent the problem of language getting into the way of comprehension, one interviewee proposed that it would be more convenient for her to tell the summary in L1, which, she thought, would be easier and would show her real reading comprehension: '*If I understand well, I should be able to tell it to my friends to my teacher in my own language.*' Some interviewees preferred to combine spoken language with written language. One candidate, for instance, stated that

she would like to *'tell and write a summary'* to show how far she had understood.

Asked about what techniques should be used if one's real degree of comprehension was to be measured, most of the interviewees suggested that a combination of the techniques mentioned above may give the true result, emphasising that the greater proportion of questions should be open-ended rather than closed. There was one subject who expressed her concern that testing one's true degree of reading comprehension may not be possible at all: *'I don't think it is possible to exactly know how much a person understood and how much she didn't'*. Her concern reflects the idea that the true nature of reading comprehension itself is not yet known. Another candidate believed that different people may understand a text differently, and therefore different techniques should be used for checking their comprehension. A similar concern was raised by another interviewee who complained that different teachers judge their answers to comprehension questions differently, and that while the answers to some questions are acceptable for some teachers, the same answers are not acceptable for others. Although subjectivity is a problem throughout the whole process of language testing, how far testers are allowed to exercise subjectivity is an issue yet to be resolved.

One interviewee, who stated that he was not an expert in testing, believed that *'discussion'* was a good technique, but he was not sure whether it would show real comprehension. A different interviewee talked about a similar technique, *'critiquing the text'* by which he meant discussing characters and the roles of people in a novel, a technique which, he thought, was better than cloze tests and T/F questions. He also criticised m/c and T/F questions on the grounds that *'some of them can be answered without any knowledge'* of the issue in question and explained how he had been able to answer 60-70% of a sample MA paper in TEFL using some *'test tricks'*. He added that although m/c was a good technique, and there was no other better technique to substitute it, it was not logical.

A few interviewees preferred the easier m/c technique, and although some of them were aware of the fact that m/c items may not be the most valid technique for this purpose, others believed that the answers to such questions would show their real reading comprehension. One subject believed that T/F questions may or may not test true reading comprehension depending on the focus of the question: *'If they are used*

to test detailed meaning, they can do this, but they can't measure general reading comprehension'. One of the interviewees favoured writing whatever she understood as her preferred testing technique. She did not like answering questions because, despite the fact that she could understand the text, she could not answer some questions, *'which makes me annoyed and causes me to forget whatever I have understood'*. The implication for testers is that they should strive to test comprehension of information stated or implied in the text and nothing beyond.

One candidate preferred questions including an idea and then asking which paragraph that idea belonged to. Another interviewee mentioned cloze as her preferred testing technique, provided that the original text is given later on and she could therefore check her comprehension against it. In a few cases where the interviewees compared their preferred techniques with cloze tests, all thought that cloze was not appropriate for testing their comprehension. For most of the subjects interviewed, cloze tests that they had previously taken were intended to test mainly grammatical structures and vocabulary. In some cases, the cloze tests had been intended to test reading comprehension. Most of the candidates felt that whatever cloze tests were intended for, they were better at testing grammar and vocabulary than reading comprehension. One interviewee criticised cloze tests on the grounds that they measured something beyond English knowledge. He believed, *'they measure one's cleverness, and if somebody is in a good mood, he can do them well'*. Asked if they needed to comprehend the whole text before they could give appropriate answers to blanks, or whether it was enough to understand only a few words and sentences around the blanks in the cloze test, most of the candidates stated that different blanks made different requirements: While for some blanks they needed to understand a larger portion of the text, and preferably a whole paragraph in order to provide a correct word, this was not the case for all items.

What such an observation implies is that only some blanks in cloze tests require comprehension of a larger part of the text, and that while overall text-comprehension is necessary to answer some blanks correctly, it is not necessary to answer all of them. If we accept that only those items which need overall comprehension to be completed correctly can measure overall reading comprehension, and that not all items need overall comprehension for restoration, then not all items are able to measure overall comprehension. Theoretically, if cloze tests can be made

in which all items require overall comprehension for their restoration, they will then be able to measure reading comprehension to a significant extent.

4. *The meaning of scores on reading tests.* The candidates were also asked to reflect on the reading exams they had so far taken at schools or at the university, and to explain whether they thought the scores they obtained showed their real comprehension of the texts they read. Except for a limited number of interviewees who seemed to be happy with their scores, more than two thirds were not satisfied with their scores and thought that their scores on reading tests did not show their exact degree of comprehension. One candidate expressed her concern in this way: *'None of my scores in any other course show my real ability.'*

One student thought that the answers to some questions seemed to be correct for her, but the teacher did not accept them. Namely, they understood the same text differently, and although both could be right, her understanding was not given credit. Another interviewee argued that some of the questions could not be answered based on the text and that they required something more than comprehending the text. A few candidates thought that their real comprehension was more than the scores they got because there were not questions from all the parts they understood, or there were too many questions from details to which they did not pay attention. They also argued that not all the questions tested comprehension: some tested their vocabulary knowledge and some others tested their grammar. Another interviewee argued that he could answer some questions even without reading the text, or even without understanding all the details in the passage, and that sometimes he read only those parts of the text where the answers to the questions could be found without reading the whole passage.

A different interviewee thought that her score could not show her true comprehension because she could answer some m/c questions by chance. Some other interviewees felt that their scores were not indicative of their true comprehension because of exam-factors like time-limit, stress, not getting enough sleep the night before the exam, not being in a good mood, and so on. They believed that they could give better answers if they sat the same exam at home. The reason why her score did not show her real comprehension for one interviewee was that the texts had already been studied during the course and had been comprehended completely,

and therefore, she felt no point in using the same text for testing comprehension, implying the need for using 'unseen' passage in reading comprehension tests.

On the other hand, one candidate who believed that her score showed her true comprehension "most often" argued that her score showed her comprehension because if she comprehended the text, she could answer the questions. Merely being able to answer a question correctly, however, may not imply that the text has been truly comprehended, as the above interviewees argued. One interviewee argued, *'cloze tests show our comprehension, vocabulary, and grammar knowledge because this is what the teachers have told us and this is the main thing that we that is said in schools'*. Evidently, the student himself had no idea what cloze tests measured and he was expressing what others have told him, which may not be true.

As evidenced by most of the interviewees' comments, rarely ever do reading tests referred to measure students' real comprehension. Although very little qualitative research exists on views of testees on reading tests, our finding here, which is in line with Alderson's (2000), can provide a justification for a major revision to reading tests and for attempting to construct testing tools that will be able to properly measure real comprehension not only as far as testers are concerned but also in the eyes of the main stake-holders themselves. With the limitations already counted of language tests on the one hand, and with the unclear nature of reading itself as well as with the notion of what comprehension is on the other (Flahive, 1984), whether such an improvement is possible is a question yet to be answered. The fact that reading comprehension test scores represent the product of the interaction between the test-taker's reading comprehension ability and a variety of test method facets among other variables (Johnston, 1983; Bachman, 1990; Bachman and Palmer, 1996) adds to the complexity of the issue.

5. Conclusion

The main purpose of this qualitative study was to investigate the problem of validity of EFL reading comprehension tests from the test-takers' point of view using a qualitative perspective. The evidence produced by the interviewees during the interview sessions suggests that there are serious problems with the way that their reading

comprehension was and is being measured and that most of the tools used for measuring reading comprehension measured something below and beyond comprehension; as such the majority of the reading tests students talked about failed to pass the test of judgemental validity.

The fact that more than half of the interviewees argued strongly that rarely ever did the scores they got on cloze tests and other tests of reading comprehension indicate their comprehension of the passage on which the tests were based, and that the scores showed something less or more than mere text-comprehension, adds to the strength of doubts about the validity of these tests as viable measures of EFL reading comprehension. Two implications emerge from such an observation: either the tests were invalid for the purpose for which they were used, or scores are inappropriate tools for indicating the degree of comprehension. Capitalising on the discussions made earlier, it seems that both implications are true. Namely, the reading tests may not be valid tests of reading comprehension partly because the true nature of reading comprehension is yet to be known, and partly because the tests seemed to have been used for testing 'general reading ability' rather than reading comprehension. Also, scores may not be indicative of the right degree of comprehension because, despite having different degrees of importance, all questions (and the answers to them) are equally weighed in m/c reading tests, and that the whole process of text-choice, question-choice, alternative-choice and so on is rather subjective; in addition, not all text-content is covered in questions.

While the current study is not claimed to be faultless, its findings are of use for language teachers, testers as well as researchers. By knowing learners' views on what test scores mean for them, classroom teachers and testers will be more careful in scoring students test papers, a goal the fulfilment of which will need striking a balance between test reliability and test validity. Professional testers will similarly be helped in better catering for customer views by adding judgmental validity to the list of validity types they account for in the new tests they make. Researchers will also illuminate the nature of entities they are studying by adding qualitative perspectives to their quantitative investigations. As an improvement over the shortcoming of this study, future researchers can seek the views of test-takers on specific tests immediately after they are taken rather than inviting candidates to talk generally about the tests they have previously taken.

Another limitation of the study which can be improved by future research is to seek the opinions of teachers and testers themselves, which when added to the data elicited from test-takers will reveal a clearer picture of the story.

References

- Alderson, J.C. (2000). *Assessing Reading*. Cambridge: Cambridge University Press.
- Alderson, J.C., & Urquhart, A.H. (1984). *Reading in a foreign language*. London: Longman.
- Aslanian, Y. (1985). Investigating the reading problems of ESL students: an alternative. *ELT Journal*, 39, 20-27.
- Bachman, L.F. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bachman, L.F. (2000). Series editor's preface. In J.C. Alderson, *Assessing reading* (pp. x-xi). Cambridge: Cambridge University Press.
- Bachman, L.F., & Palmer, A.S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Brown, H.D. (2004). *Language assessment: Principles and classroom practice*. New York: Pearson Education.
- Dehn, N. (1984). An AI perspective on reading comprehension. In J. Flood (Ed), *Understanding reading comprehension: Cognition, language, and the structure of prose* (pp. 82-100). Newark, Delaware: The International Reading Association.
- Farhady, H. (1983). On the plausibility of the unitary language proficiency factor. In J. W. Oller (Ed.), *Issues in language testing research* (pp. 11-27). Rowley, MA: Newbury House Publishers.
- Farhady, H., Jafarpour, A., & Birjandi, P. (1996). *Language skills testing: From theory to practice*. Tehran: SAMT.

- Farr, R., Carey, R., & Tone, B. (1986). Recent theory and research into the reading process: Implications for reading assessment. In J. Orasanu (Ed.), *Reading comprehension: From research to practice* (pp. 135-149). Hillsdale, New Jersey: Lawrence Erlbaum Associates Publishers.
- Flahive, D.E. (1980). Separating the g factor from reading comprehension. In J.W. Oller & K. Perkins (Eds.), *Research in language testing* (pp. 34-46). Rowley, MA: Newbury House.
- Harrison, J., Bailey, M., & Foster, C. (1998). Responsive assessment of reading: Seeking evidence on reading from students. In M. Coles & R. Jenkins (Eds.), *Assessing reading 2: Changing practice in classrooms*. London: Routledge.
- Heaton, G.B. (1988). *Writing English language tests (2nd Ed.)*. London: Longman.
- Johnston, P.H. (1983). *Reading comprehension assessment: A cognitive basis*. Newark, Delaware: International reading association.
- Kolers, P.A. (1973). Three stages of reading. In F. Smith (Ed.), *Psycholinguistics and reading* (pp. 28-49). New York: Holt, Rinehart and Winston.
- Mackey, A., & Gass, S.M. (2005). *Second Language Research: Methodology and Design*. Mahwah, NJ: Lawrence Erlbaum Associates.
- McNamara, T. (2000). *Language testing*. Oxford: Oxford University Press.
- Myers, S.S. (1991). Performance in reading comprehension: Product or process? *Educational Review*, 43, 257-272.
- Nation, I.S.P. (2009). *Teaching ESL/EFL reading and writing*. London: Routledge.
- Rosowsky, A. (2000). Reading and culture: The experience of some of our bilingual pupils. *English in Education*, 34, 45-53.
- Walczyk, J.J., Kelly, K.E., Meche, S.D., & Braud, H. (1999). Time limitations enhance reading comprehension. *Contemporary Educational Psychology*, 24, 156-165.