Achievement Test Development and Validation: A Measure of Reading Comprehension Strategies for Iranian Learners of English

Mahmoud Samaie

Dept. of English, Ilam University

PO Box 516, Ilam 69315, Iran

Tel: 98-912-447-8125   E-mail: m.samaie@ilam.ac.ir

Fereshteh Khosravian (Corresponding author)

Dept. of English, Ilam University

PO Box 516, Ilam 69315, Iran

Tel: 98-916-372-5502   E-mail: fkhosravian@yahoo.com

Abstract

This study investigated the validity of an achievement test as a measure for Iranian English as a foreign language (EFL) learners' reading comprehension strategies at the pre-intermediate level at Ilam University. Different reading strategies have been introduced to aid learners in the process of comprehension, four of which, namely making connections, visualizing, inferencing (inferring), and questioning the author were selected for the purpose of the study. Adopting a reductionist approach to collecting validity evidence for the sake of practicality, the researchers made an attempt to focus merely on construct validity based on the Bachman and Palmer's (1996) framework. The evidence to examine the construct validity of the developed test was gathered through the differential-groups design, which involves selecting a sample with two mastery and non-mastery groups. The results demonstrate that the developed achievement test is a valid measure to assess the above-mentioned reading comprehension strategies for Iranian EFL learners.

1. Introduction

Language teaching has passed through a long way in search of a remedy for the language to be taught and learnt (Khany and Khosravian, 2013). Numerous methods, approaches, and theories have been recommended different intuitions to language teaching and learning. Today, there is an accord that successful language education relies much upon comprehensive teaching and learning principles (Brown, 2001; Kumaravadivelu, 2003). Language teachers globally attempt to employ different kinds of approaches to help their students to learn the fundamental skills of language including listening, speaking, reading, and writing. To assess the efficacy of educator teaching as well as pupil education of aforementioned skills, language tests can be planned and conducted (Mikhaylova, 2009). The most important index of, interpretation and use of, a useful test has almost always been considered to be that of validity, which has, of course, been conceptualized differently by interested theoreticians in the notion. According to Hughes (2003, p. 26), a test is considered to be valid if it measures accurately what it is intended to measure.

To investigate the validity of a test, both interpretation and use, different types of validity evidence need to be collected depending on one's frame of reference, such as the classic one which calls for marshalling content, criterion-related, and construct validity evidence. Viewed from within the limits of this conventional framework, construct validity has come to enjoy a superordinate status. In Bachman's (1990) words, construct validity "concerns the extent to which performance on tests is consistent with predictions that we make on the basis of a theory of abilities, or constructs" (p. 225). Due to the importance of construct validity, the present study made an attempt to develop a reading comprehension achievement test and examine its construct validity in terms of the selected strategies for Iranian EFL learners. In what follows, a review of related literature concerning reading comprehension and some of its effective strategies is presented.

2. Review of Literature

To validate the constructed test, it seems essential to describe reading comprehension along with some of its effective strategies that were selected for the purpose of this study. With regard to the definition of reading, Chastain (1988) states it is a process concerning with the activation of related knowledge and relevant language skills to perform an exchange of information from one person to another. Reading, he (ibid.) continues, necessitates the reader concentrating on the reading materials and integrating formerly acquired knowledge and skills for comprehending what someone else has written. Snow and Sweet (2003, p. 1) describe reading comprehension "as the process of simultaneously extracting and constructing meaning". Related literature indicates that reading comprehension strategies are somehow ignored by teachers in language classrooms. According to CIERA (Center for the Improvement of Early Reading Achievement), many research studies have divulged the enhancement of reading comprehension through applying different reading strategies (Palincsar and Brown, 1984; Adler, 2001; Trabasso and Bouchard, 2002). Block and Israel (2005) suggest comprehension strategies such as inferring, making connections, modeling, predicting, questioning, summarizing, and think-aloud process ought to be applied as soon as

kindergarten.

Some researchers such as Pressley (1997) and Collins and Collins (2002) identified five strategies that serve as a basis for effective reading comprehension instruction including; accessing prior knowledge, making connections to text, questioning, seeking clarification, using mental imagery, and summarizing. Such strategies obviously necessitate student involvement and teacher management. Guthrie, Schafer, and Wang (1995) suggest that comprehension strategies, when effectively implemented, empower students with the confidence to expand the frequency and level of their reading. Teele (2004) asserts that "[t]he goal of all readers should be to understand what they read" (p. 92). Research shows that good readers are actively involved with the text, and they are aware of the processes they use to understand what they read. Teachers can help improve students' comprehension through instruction of reading strategies. Alongside all of these studies, there appears to be a need to conduct a study to discover whether a useful test could be developed which would assess reading comprehension strategies. Accordingly, to implement this purpose, as mentioned before, only four strategies, namely inferring, making connections, questioning, and visualizing selected. The definition of each comes in the following.

3. Definition of Some Basic Terms

3.1 Reading Comprehension Strategies

Reading comprehension strategies refer to the conscious and flexible plan that students apply and adapt to a text when they face problems while reading. Readers use reading comprehension strategies, both cognitive and metacognitive, to better understand reading texts and in order to learn to read independently (Allen, 2003).

3.2 Cognitive Strategies

Cognitive strategies are generally deliberate, global, rather than domain-specific (Chi, 1987), activities undertaken by active learners, many times to remedy perceived cognitive failure (Garner, 1987). A cognitive strategy has four properties including: being general and domain independent, having a goal, having several components, and having finite total number of strategies in memory (Chi, 1987).

3.3 Making Connections

In order to enhance reading comprehension, readers may relate their previous learning and background knowledge to the text. This strategy is called making connections.

3.4 Visualizing

While reading, reader may possibly construct an image to be kept in his/her memory as an illustration of what he/she interprets about the text (National Reading Panel, 2000).

3.5 Inferring (Inferencing)

Pupils are required to employ their background learning as well as the knowledge of the words to conclude on their own and reach an appropriate comprehension about the text

(Serafini, 2004). Furthermore, they can perform following actions via inferring: conclude about the text, forecast the succeeding events, recognize basic topics, and employ the images and knowledge of the text to construct meaning (Harvey and Goudvis, 2000).

3.6 Questioning (the author)

In order to create meaning, improve comprehension, discover responses, resolve difficulties, discover knowledge, and find recent knowledge; readers are required to question themselves preceding, throughout, and following the reading (Harvey and Goudvis, 2000).

4. Statement of the Problem

Due to increasing demands on students in general and EFL students in particular, both in the classroom and in society, teachers are forced into the realization that their students' reading abilities would impact the success of those students in the content area classroom and that students who leave the school system without the proper reading skills and strategies are at an automatic disadvantage in society (Biancarosa and Snow, 2004). EFL learners now need some skills and strategies to read all kinds of texts for understanding. Universally, teachers of English try to employ different kinds of approaches to help students to learn the fundamental strategies and skills of language. Different reading strategies have been found to assist learners in the process of comprehension. The effectiveness of these strategies should be assessed through administering tests. Of crucial means to confirm the quality a test is to investigate its usefulness based on one of the available frameworks such as Bachman and Palmer's (1996). In case it turned out to be valid, the test developed for the study is expected to be helpful at least in terms of construct validity in Iranian educational context as an instrument to measure EFL learners' reading comprehension strategies. With regard to what has already been stated and based on the objectives of the research, the study sought answer to the following research question:

Is the developed achievement test a valid measure to assess the selected reading comprehension strategies for Iranian EFL learners?

5. Method

5.1 Participants

The participants of the study composed of 26 pre-intermediate Iranian students studying English Literature at Ilam University. The participants were both female and male and their age range was 18-25. They were at two pre-intermediate levels, upper and lower. The students at the upper level would act as a mastery group -those who are supposed to gain instruction to use reading comprehension strategies during reading comprehension- and the students in the lower group were defined as a non-mastery group -those in whose syllabus the instruction of the related strategies was not included.

5.2 Instrument

The instrument of the present study was a reading comprehension achievement test developed by the researchers. The test consisted of five texts each of which was followed by eight

comprehension questions: all of the items were in multiple-choice format. There was an attempt to select those texts with almost the same level of difficulty of the texts that the students were exposed to during the term.

5.3 Procedure and Data Analysis

After the instruction of selected strategies to the mastery group, both groups of mastery and non-mastery received a reading comprehension test at the end of the semester. In what follows, the process of test development and test evaluation is presented.

5.3.1 Test Development

Bachman and Palmer (1996) assert that test development is the entire process of creating and using a test, beginning with its initial conceptualization and design, and ends up with one or more archived test and the results of their use. They propose three stages for test development: design, operationalization, and administration.

5.3.1.1 Design Stage

In this stage, a detailed description of the components of the test design was prepared to insure that performance on the test task was compatible with language use and the test scores were maximally useful for their intended purpose. Bachman and Palmer (1996) divide this stage into six components:

1) A description of the purposes of the test; in this study the purpose of the test was to assess the effectiveness of reading comprehension strategies in promoting Iranian EFL learners' reading comprehension ability.

2) A description of the target language use domain (TLU) and task types; it was supposed that the results of this study could be generalized to the TLU domain, because all of the students probably encounter reading comprehension tasks in their daily life, e.g. in reading newspaper.

3) A description of the population of test takers for whom the test is intended; the test was prepared for Iranian EFL learners in an English class at Ilam University, they were both female and male with the 18-25 year age range.

4) A definition of the construct(s) to be measured; the construct to be measured was the students' reading comprehension strategies including visualizing, making connections, inferencing, and questioning the author (their definitions are presented in section 3).

5) A plan for evaluating the qualities of usefulness; evaluating the usefulness of the test was on the basis of Bachman and Palmer's (1996) framework, taken only construct validity into account in this study.

A list of required and available resources and the plan for their allocation and management; the required resources in this test included human resources: test developer, administrator, and rater that would be the researchers themselves; material resources: an English class at Ilam University for testing the related materials; test material: related texts, paper, and pen;

and the time resources: time to develop, administer, and score the test.

5.3.1.2 Operationalization Stage

- *Setting:* The constructed test administered in a session after the last session and in the learners' own class in order to prevent any problems related to the unfamiliarity of the setting.

- *Rubric:* Because of the test takers' level that was pre-intermediate, they were assumed to be familiar with this type of testing. The tasks would be scored objectively by the teacher.

- *Input:* The input was only the related printed texts.

- *Expected response:* The students were expected to answer the multiple-choice questions just by circling the correct responses on their answer sheets.

- *Relationship between input and response:* There was a direct relationship between the input and the expected response. The responses were formed directly from the information provided in the input i.e. the test takers read the input and answered the questions based on the information given in the text.

5.3.1.3 Test Administration Stage

Test administration stage involves giving the test to a group of individuals, collecting information, and analyzing the information for the purposes of assessing the usefulness of the test and making the inferences or decisions for which the test was intended. In this study, as mentioned above, the developed test administered in a session after the last session and in the learners' own class in order to prevent any problems related to the unfamiliarity of the setting. For the purpose of the test, it was developed into five parts; each part contained eight items, all of which were in multiple-choice format. Each multiple-choice item consisted of a stem and four options and students were expected to answer the questions by circling the correct response on their answer sheets. All of items possessed equal points; each one was worth one point. The test time limit was forty minutes, one minute for each item. Because of the nature of selected-response items, the test was scored objectively by the teacher herself.

5.3.2 Evaluating the Usefulness of the Test

In regard to the usefulness of the test, Bachman and Palmer's (1996) framework was taken into account. It emphasizes six factors: construct validity, reliability which is referred to dependability for criterion-referenced test, authenticity, interactiveness, impact, and practicality. For the purpose of the study, among the qualities of usefulness, the most important one, i.e. construct validity, was focused on.

6. Results

Differential groups approach has been applied to assess construct validity. This approach requires defining two groups of students, one group that possesses the construct (masters) and another that lacks the construct (non-masters). After determining the groups of masters and

non-masters, the test administered and the results analyzed. The students' test scores and mean scores for both master and non-master are illustrated in Table 1.

Table 1. Students' test and mean scores

	Students' scores													Mean
M	18.65	18	17.1	15.3	14.85	14	13.75	13.1	11.75	10.65	10.65	9.75	8.2	13.51
NM	16.4	15.2	14.65	13.3	12.2	11.1	10.4	10.2	10.2	10	9.3	9.3	6.65	11.45

M: Master, NM: Non-Master

The means of two groups, through paired samples t-test statistics, are also shown in Table 2.

Table 2. Paired samples t-test statistics

	Mean	N	Std. Deviation	Std. Error Mean
Master	13.51	13	3.24	.90
Non-Master	11.45	13	2.76	.76

As observed in Tables 1 and 2, the mean score for master group is considerably higher than the one for the non-master group. This difference could support the claim that the test was a valid measure of the effect of the aforementioned strategies on the reading comprehension skill of pre-intermediate EFL students. However, in order to provide a stronger support, it was required to check the statistical significance of the observed difference. In so doing, a paired samples t-test was carried out using Statistical Package for the Social Sciences (SPSS) software, version 16. Table 3 shows the SPSS output.

Table 3. SPSS output for paired samples t-test

	Paired Differences							
				95% Confidence Interval of the Difference				
	Mean	SD	SEM	Lower	Upper	t	df.	Sig. (2-tailed)
Master/ Non-master	2.06	.90	.25	1.52	2.61	8.25	12	.000*

SD: Standard Deviation, SEM: Standard Error Mean, * $p<0.05$

As observed in Table 3, the statistically significant difference between the mean scores of the participants in two groups of master and non-master suggests that students in master group performed much better in comparison with non-master group. Accordingly, it was confirmed that test was a valid device to measure the aforementioned strategies on the reading comprehension skill of pre-intermediate EFL learners.

5. Discussion and Conclusion

The present study aimed at developing and evaluating an achievement test, assessing the effect of four of the reading comprehension strategies namely, making connections, visualizing, inferencing (inferring), and questioning the author on pre-intermediate EFL learners. To answer the question arose in this regard, a plan was developed based on the usefulness model proposed by Bachman and Palmer (1996) and the techniques for assessing reliability and validity of criterion-referenced tests offered by Brown and Hudson (2002). As stated by Brown and Hudson (2002), all construct validity approaches rely on experimental studies of the degree to which a test is measuring the construct it claims to measure. Among these approaches, they have introduced intervention and differential group studies as the most practical and appropriate approaches to examine construct validity in the case of criterion-referenced tests.

Differential groups approach was applied to investigate the test construct validity. The obtained mean for master group was observed as being considerably higher than the mean for the non-master group. This observed difference was subjected to a paired samples t-test using Statistical Package for the Social Sciences (SPSS) software. The results of the paired samples t-test indicated the statistical significance of the observed difference in means and supported the claim that the test is a valid measure of the effect of the pertinent strategies on pre-intermediate EFL learners' reading comprehension skill.

Concerning the obtained results, the established and evaluated test in the present study might function well as an instrument for efficiently examining the effect of the pertinent strategies on promoting pre-intermediate EFL learners' reading comprehension skill. However, there are some limitations to be acknowledged and addressed. One limitation concerns the external validity or generalizability of the findings, since the findings were generated with a limited number of participants. Additionally, not all of the indices of usefulness were taken into account. Considering the limitations, further research is required to develop and evaluate tests of higher quality to assess the effect of the aforesaid strategies improving EFL learners' reading comprehension skill.

References

Adler, C. R. (2001). *Put reading first: The research building blocks for teaching children to read.* Jessup, MD: ED Publication.

Allen, S. (2003). An analytic comparison of three models of reading strategy instruction. *IRAL, 41,* 319-338. http://dx.doi.org/10.1515/iral.2003.015

Bachman, L. F. (1990). *Fundamental considerations in language testing.* Oxford, UK: Oxford University Press. http://dx.doi.org/10.1191/0265532203lt268xx

Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice: Designing and developing useful language tests.* Oxford: Oxford University Press. http://dx.doi.org/10.1177/136216889900300106

Biancarosa, G., & Snow, C. E. (2004). *Reading next - a vision for action and research in*

middle and high school literacy: a report from Carnegie Corporation of New York. Washington, D.C.: Alliance for Excellent Education.

Block, C., & Israel, S. (2005). *Reading first and beyond: The complete guide for teachers and literacy coaches.* Thousand Oaks, CA: Corwin Press.

Brown, H. D. (2001). *Teaching by principles* (2nd ed.). White Plains, NY: Pearson.

Brown, J. D., & Hudson, T. (2002). *Criterion-referenced language testing.* Cambridge: Cambridge University Press.

Chastain, K. (1988). *Developing second language skills: Theory and practice* (3rd ed.). Orlando, Florida: Harcout Brace Jovanovich.

Chi, M. T. H. (1987). *Knowledge structures and memory development.* Hillsdale, NJ: Erlbaum.

Collins, D., & Collins, A. (2002). *Becoming effective teachers of reading.* Tallahassee, FL: SERVE Publications.

Garner, R. (1987). *Metacognition and reading comprehension.* Norwood, NJ: Ablex.

Guthrie, J., Schafer, W., & Wang, Y. (1995). Relationships of instruction in reading: An exploration of social, cognitive and instructional connections. *Reading Research Quarterly*, *30*(1), 8-25. http://dx.doi.org/10.2307/747742

Harvey, S., & Goudvis, A. (2000). *Strategies that work teaching comprehension to enhance understanding.* York, ME: Stenhouse.

Hughes, A. (2003). *Testing for Language Teachers.* Cambridge: Cambridge University Press.

Khany, R., & Khosravian, F. (2013). *The application of Wikipedia for enhancing Iranian EFL students' reading proficiency.* Proceeding of the Global Summit on Education 2013 (GSE 2013). Kuala Lumpur, Malaysia. 135-142.

Kumaravadivelu, B. (2003). *Beyond methods: Macrostrategies for language teaching.* New Haven, CT: Yale University Press.

Mikhaylova, M. V. (2009). *Validation of the reading level achievement test of the English language center.* Unpublished Master's Thesis. Brigham Young University, Utah, USA.

National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction.* Washington, DC: Reports of the subgroups.

Palincsar, A. S., & Brown, A. L. (1984). Reading comprehension teaching of comprehension–fostering and comprehension–monitoring activities. *Cognition and instruction*, *1,* 117-175. http://dx.doi.org/10.1207/s1532690xci0102_1

Pressley, M. (1997). Skilled comprehension and its development through instruction. *School Psychology Review*, *26*(3), 448-466.

Serafini, F. (2004). *Lessons in comprehension explicit instruction in the reading workshop.* Portsmouth, NH: Heinemann.

Snow, C. E., & Sweet, A. P. (2003). *Rethinking reading comprehension.* New York: Guilford Press.

Teele, S. (2004). *Overcoming barricades to reading: A multiple intelligences approach.* Thousand Oaks, CA: Corwin Press.

Trabasso, T., & Bouchard, E. (2002). *Teaching readers how to comprehend text strategically.* New York: Guilford Press.

Copyright Disclaimer