



The Effect of Test Task Types on Reading Comprehension Performance

Fatemeh Karimi

Department of Foreign Languages, Isfahan (Khorasgan) branch, Islamic Azad University, Isfahan, Iran.

ARTICLE INFO

Article history:

Received: 6 March 2017;

Received in revised form:
10 June 2017;

Accepted: 20 June 2017;

Keywords

Test Task Types,
Test Method Effect,
Reading Comprehension
Assessment,
Language Testing.

ABSTRACT

Test-task characteristics have been considered as an important factor affecting the test takers' performance on a test (Bachman, 1990, p. 156). Response format is one part of test-task characteristics that affects the performance of test takers. This study investigates whether the kind of response format would affect the performance of test takers. For this reason two test methods - - multiple-choice and free-response - - have been compared. The tests, together with a checklist of test taking strategies were administered to forty undergraduate English students of Rasht Azad University. The analysis showed that two tests with identical content but different formats did not yield measures of the same trait. In addition, the discrimination indices showed that two reading tests could discriminate students differently. This was more investigated by analyzing data obtained from the checklists of test taking strategies.

© 2017 Elixir All rights reserved.

Introduction

One of the most common tests that foreign language learners encounter during the course of their studies is reading comprehension tests. Bachman (1990,p.81) believes that, test takers' performance in any language test (e.g. reading comprehension test) is affected by a large number of factors that must be taken into account in the design, development and use of the test. According to Bachman (1990,p.165) these factors are: communicative language ability, personal attribute, random factor, and test method facets. As it is clear, one of factors influencing test performance is test method facets or test-task characteristics. Test method facets, according to Mousavi (1999,p.408), are the "specific characteristics of test methods which constitute the "how" of language testing. Thus test performance is affected by these "facets" or characteristics of test method".

Bachman (1990,p.118) presents a framework of test method facets consisting of five major categories including: setting, test rubric, nature of the input, nature of expected response and the relationship between input and expected response. One part of this facet is nature of the expected response which is considered in this study. In a language test, according to Bachman and Palmer (1996,p.53), "the expected response consists of the language use that we are attempting to elicit by the way instructions have been written, by the task designed, and by the kind of input provided". The characteristics of expected response include format and language of expected response.

This study focuses on the nature of the expected response by manipulating response format and examines the effect of it on the test performance. The aim of this study is to explore the effect of test format on high-ability EFL test takers performance in reading comprehension test in Iran. The formats which are used for this study are free – response and multiple – choice formats.

It was assumed that multiple-choice format is measuring the hypothesized construct of reading comprehension the

same degree as free-response format does, so test takers would perform equally well on both tests of reading comprehension and would employ the same test – taking strategies.

Review of Literature

Reading researchers have argued that different test formats seem to measure different aspects of language ability (Graesser, Hoffman, and Clark, 1980; Reder and Anderson, 1980; Kintsch and Yarbrough, 1982; Lewkowicz, 1983).

Samson (1983) used multiple-choice questions, open – ended questions, and summary tests in a reading comprehension test. The result showed there was no significant difference among the three test methods, so she concluded that the three test methods all tested the same ability or trait of the subjects. But she did find that multiple-choice questions were the easiest, and summary test the most difficult.

Shohamy (1984) investigated two aspects of test method, the test format and the language of presentation. She found significant differences between the performance of the subjects with different levels of proficiency on tests differing in terms of the test format(multiple – choice or free – response questions) and the language of presentation (target language or subjects' native language).

Hughes (1989,p.36-40) believes that there are some factors that affect performance of test takers. He mentions some of them and suggests ways of achieving consistent performance from candidates.

Weir (1990,p.42) also points out that some evidence reveals that test format can affect students' achievement. Given the lack of complete knowledge on the effect of test formats, he suggests that " the only practical approach at present is to safeguard against possible format effect by spreading the base of a test more widely through employing a variety of valid, practical and reliable formats for testing each skill".

Brown, J.D. (1996,p.188) talks about the measurement error. According to him, "measurement error is a term that describes the variance in the purpose of the test". He summarizes the types of issues that are associated in the testing literature with measurement errors and affect test performance.

Liu (1998) used multiple-choice questions, true or false questions and short answer questions in his reading comprehension test. The results showed that test methods affected the subjects' performance on reading comprehension tests, and that high-proficient students were more easily affected than low-proficient students. There were significant differences among the scores elicited by the three different test methods, short answer questions were the most difficult.

Alderson (2000,p.60) has reviewed a range of factors that might affect reading comprehension performance. He mentions that many aspects of text including "linguistic features of text, text type, organization, genre and text topic" might influence the performance of readers. He points out that any variable that has an impact on reading process or product goes the validity of our test under question.

According to McNamara (2000), one point in test design is the way in which candidates will be required to interact with the test materials, particularly the response format. He mentions two broad approaches to understanding the relation of test method to test content. The first sees method as an aspect of content, and raises issues of authenticity; the second, treats method independently of content, and allows more obviously inauthentic test response format.

Kobayashi, M. (2002) investigated the effects of method on reading comprehension test performance. The main purpose of his research was to investigate the effect of factors other than language ability on reading comprehension test performance. The two main variables were: text organization and response format. The main finding was that text types and response format not only had significant effects on reading comprehension separately, but they also interacted with each other.

Farhady,etal (2003,p.140) also mention some factors that affect test takers performance. According to the authors, knowing these factors and having some information about them are necessary because it helps test developer to be more careful in controlling them. These factors, according to them, are: the effect of testees, the effect of test factors, the structure of the test, the effect of administration factors, and the influence of scoring factors.

Brantmeier, C. (2005) conducted a research about effects of readers' knowledge, text type, and test type on L1 and L2 reading comprehension in Spanish. This study examined how a reader's subject knowledge and type of test (written recall, sentence completion, and multiple – choice) affect first language (L1) and second language (L2) reading comprehension. According to the author, analysis of covariance revealed that subject knowledge related significantly to reading comprehension. The author also found that there was no overall positive effect of text type on L1 and L2 reading comprehension as measured by recall, sentence completion, and multiple – choice tests .

Fletcher (2006) identified three issues that are practically important for the measurement of reading comprehension, namely: the nature of the text, how reading comprehension is assessed, and individual differences.

Alshumaimeri (2011) investigated the effects of different reading methods on the comprehension performance of Saudi EFL 10th grade male students. The participants were asked to read three comparable passages in three different ways (oral, silent and subvocalizing). Results reveal a significant difference between oral reading and subvocalization, and between oral reading and silent reading. He states that oral reading had the greatest effect on comprehension performance among the three reading methods examined.

Akhondi and Malayeri (2011) investigated if different response formats (test methods) measure reading comprehension of expository text differently. 48 students received an expository passage having descriptive rhetorical structure followed by three response formats, namely, incomplete outline, graphic organizer, and summary writing. Results indicated that high-achievers outperformed intermediate- and low-achievers across the three response formats. Moreover, graphic organizer appeared as the most difficult task as the respondents achieved the lowest score in this task.

Test Taking Strategy

Test taking strategies are strategies used by individuals in certain type of test. There are a lot of recommended test – taking strategies. In second language research, it is significant not only to examine learner's final products – the results of testing- but also to bear in mind the process how they think, choose and produce the appropriate answers in testing situation. The goal of this research is to identify such process and the effect of test method (multiple-choice and free-response tests) on testee's performance in English reading tests.

Present study

This study tries to investigate the effect of two tests formats (multiple-choice and free-response formats) on Iranian high-ability learners' reading comprehension performance studying English as a Foreign Language.

The question is whether these two kinds of test methods are testing the mental construct involved in reading comprehension to the same degree? Are the test taking strategies employed by test takers in answering reading comprehension questions equivalent in each of these two methods? So two null hypotheses were generated:

Hypothesis A: There would be no significant difference between scores obtained on the free – response and multiple – choice reading comprehension tests.

Hypothesis B: The test taking strategies employed by test takers in answering reading comprehension questions in free – response and multiple – choice formats would be equivalent in that they would be measuring the same construct.

To confirm or disconfirm these two hypotheses, two types of data were collected from students : data about their performance on two reading tests and data about test taking strategies by means of self – report checklist .

Methodology

Materials

The two selected reading tests were adapted from Tzagari, C. (1997), who had used them for his study for Greek students. He had taken two passages from Educational Testing Service (1987a , quoted Tzagari, C.1997). They both consist of the same passage and are followed by the same 6 items in two different formats : free – response and multiple – choice formats . The reading items were designed to measure the same construct in both tests .

Subjects

Forty undergraduate students (both male and female), who were studying English literature, in English Department of Rasht Azad University took part in this study. All Subjects were high-ability Iranian students studying English as a Foreign Language. The range of their age was from 21 to 23 and all were in the last year of their studies.

Instruments

Instrument consists of three parts :

(1) Instruction : This part of the test includes instructions on how test taker should take the test . It is the same for both formats.

(2) Checklist of Test Taking Strategies: A checklist for identifying and then comparing the process used by test – takers, were provided. The checklist for each format was different in number. For free-response test, the checklist devised consisted of 9 strategies but the checklist used in multiple-choice format consisted of 12 strategies. In fact three strategies were specific to multiple-choice format.

(3) The Reading Comprehension Test.

Procedure

In order to receive feedback about the clarity of instruction, the time needed for test taking, and testing procedures itself, it was significant that the instruments be administered in a trial administration session. The pilot test was administered in Tarbiat Moallem University. The result of this administration was used for the modification of the final version of this instrument . Then this instrument was administered with the following sequence: 40 students received multiple- choice format in the first week of the study. Two weeks later this group of students received another format, I mean free-response format . Students were told that they had 20 minutes for multiple- choice test and 30 minutes for performing in free- response test. This time limitation was necessary so as to create real testing conditions and providing suitable data.

Scoring systems

In order to compare the scores obtained from two methods of testing, the same scoring method for two tests was used. All items were scored as right or wrong. In order to be sure about the reliability in free- response format intra – rater reliability was used, and all items were corrected on two separate occasions with two week intervals. The reliability was estimated between these two sets of ratings as scores obtained from parallel tests. The correlation coefficient between two sets of ratings (Spearman rank – order coefficient) was computed as 0.91 which is high enough to be relied on. Multiple – choice items were also double – checked for any possible inaccuracy in scoring.

Data Analysis and Discussion

Quantitative Analysis

or all six questions (refer to appendix) of both free – response and multiple – choice tests , item facility / difficulty and item discrimination were calculated. All statistics are given in the following tables :

Table 1. Multiple – choice test.

Questions	Item facility	Item difficulty	Item discrimination
1	0.75	0.25	0.49
2	0.60	0.40	0.21
3	0.82	0.18	0.59
4	0.80	0.20	0.45
5	0.52	0.48	0.62
6	0.63	0.27	0.45

Table2. Free – response test.

Questions	Item facility	Item difficulty	Item discrimination
1	0.60	0.40	0.63
2	0.79	0.21	0.57
3	0.89	0.11	0.32
4	0.52	0.48	0.60
5	0.22	0.78	0.49
6	0.78	0.22	0.49

As it can be seen from the above tables , items intended to measure the same trait did not have the same difficulty levels . For example items one, four and five were more difficult in Free – response test than in multiple – choice’ test , whereas items two , three and six had higher item facility when they appeared in free – response test . Regarding item discrimination , items one , two , four and six when appeared in Free – response format discriminated the students better than multiple – choice format. But items three and five had more discrimination indices when they appeared in multiple – choice format. The easiest item was item three and the most difficult item was item five. What is interesting is that it was the same for both formats. But what is more interesting is that when these two items appeared in multiple – choice format they had greater discrimination. By comparison of item facilities/ difficulties and discrimination indices for both versions it became clear that multiple – choice format was easier than free – response but free – response format discriminated better than the multiple – choice one.

On the whole, it can be said that the variation observed between the two tests, purporting to test the same skills, provides evidence of a possible method effect which disconfirms Hypothesis A. This assumption needs to be further supported by looking at data processing.

Table 3. The overall frequency of strategy use of multiple-choice (M) and free-response(F) format for all six questions .

	Question 1		Question2		Question3		Question4		Question5		Question6	
	M	F	M	F	M	F	M	F	M	F	M	F
Bk Knowledge	0	0	0	0	0	1	0	0	2	1	1	1
Key words	0	3	0	0	0	1	0	2	0	0	0	4
The whole	16	16	4	11	4	10	8	17	8	10	7	4
Summarize	0	0	0	0	0	0	0	0	0	0	0	0
Turn Back	0	0	4	1	0	1	0	0	2	11	0	0
Match text	2	2	4	4	4	1	3	3	5	2	2	1
Order	1	4	6	0	10	8	5	6	5	6	2	2
Locate	6	15	8	15	1	18	14	12	5	10	8	10
Other	0		0		0		0		0		0	
Guess	5		4		3		4		1		2	
Match stem	1		1		13		1		1		0	
Eliminate	9		9		5		5		12		18	

Qualitative Analysis of The Data

This section is a analysis of the data collected by means of the checklists. In fact the frequency of strategy use in the two formats was analyzed for all six items, as it is shown in table three(next page).

Question one: As it can be seen from the table three no students used Background Knowledge, Summarize, Turn Back and Other strategies and they selected other strategies with varying frequencies with the exception of The whole and Match text strategies which were selected equally. The whole strategy is the most common strategy used by students. It seems that there is a method effect which is reflected in the selection of strategy use in each format for example the Key words, Order and Locate strategies are used differently by students, the more interesting point is that fifteen times students chose the three specific strategies of multiple – choice format that further supports the difference in the way students have processed this item .

Question two: The strategies that were not chosen by students, were Background Knowledge , Key words, Summarize and Other strategies . The most popular strategy for this question in both format was Locate but there was significant difference between them. When the question appeared in free-response format The whole and Order strategies were used more but when it appeared in multiple – choice format the students tended to use other strategies more. This distinction in using strategies can be the evidence for the test method effect. Match text strategy was selected by the same number of students. Three specific strategies of multiple-choice format were used fourteen times by the students .

Question three: Match stem strategy was the most common selected strategy in multiple-choice format. The most common strategy in free-response format was Locate strategy but only with one occurrence in multiple-choice format. The other popular strategy, in free – response format, was The whole strategy and in multiple – choice format was the Order strategy. Only one occurrence of Background Knowledge , Key words , and Turn Back was reported. Generally three specific strategies of multiple-choice format(Guess, Match stem and Elimination) were used 21 times by the students. It shows the considerable effect of test method on test performance. Also the frequency of strategy use by students varied significantly, again we can come to conclusion that this item has not been processed in similar way by students in both format .

Question four: The whole strategy was the most popular strategy selected by the students in free- response format. In multiple – choice format the Locate strategy was the most popular one. Three multiple-choice specific strategies were selected ten times by students. So it can be said that there is evidence about method effect in the way students processed this item . The variance of the frequency of strategy use for Key words , Match text, and Order strategies in both format was approximately the same. The Background, Summarize , Turn back and Other strategies were not chosen by any one in both formats.

Question five: The Turn back strategy was the most common strategy in free- response format (eleven times), but only two times were used in multiple – choice format. The whole and the Locate strategies were used equally in free response format (ten times), but the number of occurrence of them in multiple – choice format was eight and five.

The three multiple-choice specific strategies were used fourteen times . In fact Elimination strategy was the most common used strategy in multiple – choice format. The key word and Summarize and Other strategies were not chosen by any one. The distinction in using strategies, again can be the evidence for the test method effect.

Question six: The Locate strategy had the highest number of occurrence in free- response format (ten times), but the Elimination strategy was the most popular one in multiple – choice format (eighteen times). It shows a considerable effect of test method on the cognitive process of this item by students .Other strategies were also used differently. the Summarize , Other and Turn Back strategies were chosen by no one.

It should be mentioned that the Summarize and Other strategies were never chosen by students in both formats . Regarding the Summarize strategy ,the students may know it as a learning strategy and use it during learning rather than test taking session ; and about Other strategy , some reasons can be guessed. For example the students might be under the influence of those strategies which their names were explicitly mentioned in the checklist or they thought that those are more valid and more common in test taking, so they forced themselves to choose one of them. Or perhaps they had become aware that they were using another strategy except those mentioned in checklists, but they didn't have enough writing ability for explaining about it.

Conclusion

The analysis of statistical data about students' performance on the free- response and multiple – choice tests indicated that there was no evidence to support Hypothesis A. The difference of students performance between the two test designed to test the same skills indicated that the students did not perform equally well on the two tests. In addition to the above, the discrimination indices calculated for the items over both formats showed that two reading tests with identical content but different formats could discriminate students differently. This was more investigated by analyzing data obtained from the checklists of test taking strategies. The frequency of strategies chosen for the same questions in both formats was not similar. This disconfirm Hypothesis B. It can be concluded that students have behaved differently with the experience of taking free-response and multiple – choice reading items which it confirms the assumption that these two formats engage different test taking processes by the examinees .

Pedagogical Implication

These findings have important implications for language testing and consequently for language teaching. Since tests are administered in order to obtain information about students and making judgment about them, it is necessary to know precisely what is being tested. Thus the method effect as well as the trait under examination needs detailed specification.

References

- Akhondi,M., & Aziz Malayeri, F. (2011). Assessing reading comprehension of expository text across different response formats. *Iranian Journal of Applied Language Studies*, 3(1), pp. 1-26.
- Alshumaimeri, Y. (2011). The effects of reading method on the comprehension performance of Saudi EFL students. *International Electronic Journal of Elementary Education*, 2011, 4(1), pp. 185-195.

- Alderson, J.C. (2000). *Assessing Reading*. Cambridge: Cambridge University Press.
- Bachman, L.F. (1990). *Fundamental Consideration in Language Testing*. Oxford: Oxford University Press.
- Bachman, L.F. and Palmer, A.S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.
- Brantmeier, C. "Effects of Reader's Knowledge, Text Type, and Test Type on L1 and L2 Reading Comprehension in Spanish". *The Modern Language Journal* 89, (2005) 37-53.
- Brown, J.D. (1996). *Testing in Language Programs*. New Jersey: Prentice Hall Regents.
- Farhady, H., Jafarpur and Birjandi (2003). *Testing Language Skills From Theory to Practice*. Tehran: SAMT Publication.
- Fletcher, J.M. (2006). Measuring reading comprehension. *Scientific Studies of Reading*, 10(3), pp. 323-330.
- Graesser, A.C., Hoffman, N.L., & Clark, L.F. (1980). Structural components of reading time. *Journal of Verbal Learning and Verbal Behavior*, 19, 135-151.
- Hughes, A. (1989). *Testing for Language Teachers*. Cambridge: Cambridge University Press.
- Kintsch, W. & Yarbrough, J.C. (1982). Role of rhetorical structure in text comprehension. *Journal of Educational Psychology*. 74. 828-834.
- Kobayashi, M. "Method Effects on Reading Comprehension Test Performance: Text organization and response format". *Language Testing* 19(2002), 193 – 220.
- Lewkowicz, J.A. (1983). *Method effect on testing reading comprehension: A comparison of three methods*. Unpublished MA Thesis. University of Lancaster.
- Liu, Jianda. (1998). *The Effect of Test Methods on Testing Reading*. *Foreign Language Teaching and Research*, 2, 48-52.
- Mc Namara, T. (2000). *Language Testing*. Oxford: Oxford University Press.
- Mousavi, S.A. (1999). *A Dictionary of Language Testing*. Tehran: Rahnama Press.
- Reder, L.M. & Anderson, J.R. (1980). A comparison of texts and their summaries: Memorial consequences. *Journal of Verbal Learning and Verbal Behavior*, 19, 121-134.
- Shohamy, E. "Does the Testing Method Make a Difference? The Case of Reading Comprehension". *Language Testing* 1(1984), 147-170.
- Tsagari, C. (1997). *Testing Reading Comprehension: How fair is Our Choice of the Method?* "Paper presented at 19th Annual Language Testing Research Colloquium, March 6-9, 1997.
- Weir, C.J. (1990). *Communicative Language Testing*. London: Prentice Hall International (UK) Ltd.