Journal of Academic Research Vol 29, Issu 1, 2025 "38-56"

Does Test Method Matter? An Investigation of the Effect of a Two-Facet Reading Comprehension Test on Test-Takers' Scores

Dr. Abdulhamid Mustafa El-Murabet Onaiba

Associate Professor of English Applied Linguistics and TESOL, Department of English, School of Language, Libyan Academy for Postgraduate Studies, Misurata Branch, Libya.

a.onaiba@lam.edu.ly

Article information

Keywords:

constructed-response, effect, Multiple-choice, performance, reading

Received 10/01/2025, Accepted 23/01/2025, Available online 02/02/2025

Abstract

The degree to which test-takers perform on a given assessment may vary depending on the method of presentation. This research investigated the impact of test format—constructed-response (CR) and multiple-choice (MC)—on reading comprehension performance among EFL university students. Twenty-four students from a Reading III course participated in both assessments, with the CR test administered first, followed by the MC test a few days later. A paired-sample t-test revealed significant differences in performance, with students typically achieving higher scores on the MC format. To explore response discrepancies, follow-up interviews were conducted with students who provided different answers on the two test formats. The results highlighted several key factors contributing to these differences, including time pressure during the CR test, the presence of answer options in the MC format, and the ability to make educated guesses. While these findings suggest that MC items may be more accessible under timed conditions, the study acknowledges the potential influence of order effects and individual student attributes. Future research should further investigate these factors, employing counterbalanced designs, and exploring the impact of test formats on higher-order cognitive skills.

1. Introduction

Both selected-response (SR) and constructed-response (CR) items are commonly utilized in low-stakes and large-scale educational assessments. SR items are generally regarded as having a more reliable scoring procedure, making them fairer for examinees. Because examinees can respond relatively quickly, a broader range of the content domain can be evaluated, allowing for greater item coverage. An example of a selected-response question is the multiple-choice (MC) item.

MC items are frequently used in assessments and are considered one of the most objective formats available (Kaçar, 2023). These test items primarily focus on information recall. Hughes and Hughes (2020) noted that the main advantages of multiple-choice items are their practicality and reliability. MC items can be scored quickly and accurately, making them cost-effective. The simplicity of scoring also enables prompt score reporting, providing students and teachers with timely feedback on performance. Therefore, assessing language components such

as reading comprehension, listening comprehension, grammar, and vocabulary through selected-response items—specifically MC items—can significantly enhance test validity (Hughes and Hughes, 2020). For a test to be deemed valid, it must also be reliable; thus, assessments based on this type of item are likely trustworthy. All these factors contribute to the appeal of MC items in language-testing classrooms.

Arshad et al. (2020) demonstrated that effective reading test construction requires careful alignment between the assessment's purpose (measuring specific skills at different cognitive levels) and test-taker characteristics, such as background knowledge and proficiency level. These characteristics inform decisions about text selection, length, and item formats, such as MC or CR item tests.

Although MC test formats are considered one of the best-chosen methods for assessing reading comprehension (Arshad et al., 2020; Xie, Singh & Wong, 2024), many practitioners argue that MC items fail to elicit higher levels of cognitive processing and instead engage examinees in a guessing game (Campbell, 1999). Scholars such as Hughes and Hughes (2020), Kaçar (2023), and Cunningham (1998) contend that MC items do not effectively assess higher-order thinking skills. Consequently, using MC items may hinder the achievement of important learning objectives, including creativity, organization of ideas and information, effective presentation and exchange of information, and paraphrasing. Additionally, MC items increase the likelihood of guessing correctly, leading to lower reliability in assessments for struggling students (Cronbach, 1988).

Nevertheless, an increasing body of measurement literature suggests that MC items can assess nearly the same constructs as CR items (Kennedy & Walstad, 1997). Hancock (1994) noted that proponents of the MC format believe that MC items can be designed to evaluate complex understanding; however, creating such MC items is more challenging than developing CR items.

In contrast, many practitioners believe that CR items can elicit meaningful insights. They argue that cognitive processes are crucial (Rashidi, & Safari, 2014; Sarsarabi & Sazegar, 2023). Indeed, it is believed that CR items allow for a wide range of responses generated by the examinee. They are more effective than SR items in minimizing the likelihood of guessing correctly, as the correct answer is not explicitly provided in a CR item; however, this effectiveness relies on a precise scoring system. Furthermore, CR items can reveal the examinees' thought processes, requiring respondents to formulate their answers in their own words. As a result, many large-scale assessments continue to include CR items, despite the relatively high costs associated with scoring and reporting, due to the time and expense involved in evaluating CR items and the inherent subjectivity in scoring (Rashidi, & Safari, 2014).

1.1. The Research Problem

Test-task characteristics significantly impact test takers' exam performance (Fulcher & Harding, 2024). One important aspect of these characteristics is the response format or test type, which influences test takers' performance. In contemporary education, MC tests have become ubiquitous, often replacing constructed response (CR) tests. However, the literature lacks consensus on whether these two test formats are equally effective in measuring students' language abilities or knowledge (Xie et al., 2024). This discrepancy raises concerns regarding using MC and CR tests, particularly in assessing receptive skills, i.e., learners' reading comprehension in our case. This research study aims to illuminate this issue.

1.2. Significance of the Study

The research is significant because it offers valuable insights into the relevant literature. It aims to determine which type of test method—CR or MC —students perform better on. Additionally, this inquiry is important as it may reveal students' preferences for different testing methods in language assessment. Furthermore, this study will contribute to the existing literature, as it is the first conducted within the Libyan educational context, specifically focusing on university-level students majoring in English as a Foreign Language (EFL).

1.3. Purpose of the Study

This research aims to compare the effects of CR and MC items on reading comprehension by creating two equivalent tests for EFL (English as a Foreign Language) university students. The study seeks to identify any impact each type of item has on students' performance in a reading comprehension test. This research is based on the hypothesis that the MC and CR formats measure the underlying construct of reading comprehension differently. Specifically, test takers are expected to perform better on the MC items than on the CR items, utilizing different test-taking strategies. In light of this hypothesis, the following research questions are posed:

1.4. The Research Questions

1) Do students achieve significantly better scores on multiple-choice items than a test of equivalent structured-response items?
2) What reasons do test takers provide for changing their answers, if any, between stem-equivalent constructed-response and multiple-choice items?

2. Previous Studies

Regardless of the current study's context, literature has documented several empirical studies investigating the relationship between test method effect and test-taker performance and scores, i.e., test method on test performance. Researchers have explored various aspects of test takers' performance when the tests used change how they are designed. (Shohamy, 1984; Wolf, 1993; Tsagari, 1994; Currie & Chiramane, 2010; Chehrazad and Ajideh, 2015; Peuker et al., 2013; Karimi et al., 2014; Shahivand et al., 2014; Polat, 2020; Sarsarabi & Sazegar, 2023; Duran and Tufan, 2017; Vasan et al., 2017; Hassani and Maasum, and others).

Shohamy (1984) examined the effects of item format in reading tests and noted that MC tests produced significantly higher scores than stem-equivalent CR tests, with an effect more significant for lower proficiency participants. The researcher suggested that reading comprehension be tested using various methods and that more research be done to determine the most reliable way to measure this attribute. Similarly, Wolf (1993) also found significant differences in scores between equivalent question-answer items in multiple-choice and constructed-response items, favouring the former. She suggested that the multiple-choice and constructed-response formats might measure different abilities and recommended studying differences in difficulty by directly comparing the participant's responses. In addition, Tsagari (1994) studied stem-equivalent items in a test of reading ability and found that multiple-choice items were more comfortable and less discriminating than constructed-response items and produced a markedly different pattern of answering strategies.

Moreover, assessing the same skill, i.e., students' reading comprehension, Karimi et al.'s (2014) study examined whether the type of response format or test procedure impacts test takers' performance. The researchers compared two test formats: MC and short answer. The

exams were given to the same cohort, sixty-four Lahijan Azad University undergraduate English students. Descriptive analyses showed a significant difference between the test takers taking multiple-choice and short-answer formats. The former outperformed the latter. This indicates that two tests with the same material but different formats may not produce measures of the same attribute.

Similarly, Currie and Chiramanee (2010) examined the test method effect of two test formats, MC (multiple-choice) and CR (constructed-response), in English language tests. Two tests with the two formats were administered to compare the results, and the distracters on the MC test were based on incorrect answers from the CR test. The researchers found that students' scores differed significantly, as most students did better in the MC test format than in the CR test format. The study's findings also indicated that test-takers' options selected in MC items did not correspond with their previous answers on CR items, particularly when giving incorrect answers. In much the same vein, Salehi and Sanjareh's (2013) research highlighted the impact of response format on language test performance. Consistent with these findings, the researchers have generally concluded that multiple-choice tests are easier than constructed-response tests when the question content is equivalent.

Chehrazad and Ajideh (2015) also investigated how two distinct answer formats—a multiple-choice test and a multiple-choice cloze—affect Iranian EFL test takers' reading comprehension scores on the pre-intermediate and intermediate levels. The results showed that intermediate test takers outperformed pre-intermediate students on MC reading comprehension tests. This suggests that test-taking strategies can affect test-takers' performance, particularly at higher proficiency levels. In this particular study, the difference between scores is ascribed to the proficiency levels candidates possess rather than to be owed to the test type and format. This suggests that students may need to have attained a particular proficiency level to comprehend the material.

Furthermore, aiming to assess students' reading comprehension abilities, Duran and Tufan (2017) investigated the impact of multiple-choice and open-ended questions on understanding. The study group comprises one hundred thirty-two fourth-graders from Uşak city center's elementary schools. This study primarily employed the experimental design by administering a test of two versions, one with multiple-choice and the other with open-ended questions. The researchers used a four-question interview form for focus group discussions as a secondary data-gathering method. The findings show that multiple-choice test items proved more effective for the pupils than open-ended test items, as most students' ratings on the MCQ test were greater than their scores on the OEQ tests. Following a group interview with fifteen students, it was discovered that all preferred multiple-choice tests over exams with open-ended questions. Because they are accustomed to this type of question and have a chance to choose the correct answer, students approach multiple-choice questions with positivity. There is not a single pupil who feels positively about open-ended inquiries.

Another study comparing language learners' MCQs and OEQs test scores based on the same items prepared for grammar and reading courses was carried out by Polat (2020). The purpose of this study was to compare the results of MCQ and OEQ tests, which were given to 116 students enrolled in an Eskisehir language preparatory school. The same students took four different examinations, each with two types of reading and grammatical questions created for them. The study findings demonstrated a substantial difference in item difficulty and item discrimination levels between the OEQ and MCQ tests. MCQ tests were easier, and students did better in MCQ tests than in OEQ tests in both reading and grammatical assessments. A

similar study was conducted by Lim (2019) to explore whether the role of sub-reading skills changes depending on the test format, i.e., MC vs. OE reading questions. Ninety Chinese ESL students attending a large Midwestern university participated in the study by taking the two sets of exams. The study's findings revealed that the participants performed better on the MC questions than the corresponding OE questions, regardless of the text effect.

Notwithstanding, literature has documented diverse findings from those mentioned above. For instance, Peuker et al. (2013) compared the exam scores of 75 students in an introductory engineering course using MC and CR questions. The study showed minimal variation in scores between the two versions, with MC questions having an average score of 80% and CR questions receiving an average score of 76%.

More distinct variation in scores between MCQs and CRQs in favour of the latter was reported by Vasan et al. (2017). The researchers applied their research to a small group of incoming first-year university students in their preparatory course. The CR test scores increased among test takers on the CR test while remaining as they were on the MC test. For cued responses, many students credited their better performance to switching from superficial learning through easy memorization to deeper learning through conceptual comprehension-in-built responses. The findings are consistent with the hypothesis that using OEQs for assessment, which calls for in-depth knowledge, can improve students' exam performance.

Similarly, Hassani and Maasum (2012) investigated students' reading comprehension in summary writing and open-ended question formats. The researchers grouped test-takers according to their proficiency using the TOEFL reading comprehension test into intermediate and low achievers. The researchers concluded that both groups performed better in the summary test regardless of their level of proficiency. Moreover, Shahivand et al. 's (2014) study concluded that test format impacted examinees' performance and proficiency levels. Intermediate and upper-intermediate levels showed the most deficient performance on the MC test format compared to much better performance on the C-test, short answers, and cloze tests.

Although reviewing pertinent literature in this paper has shown that students' performance and preferences were orientated towards the MCQ test type over the CR counterpart, one may contend that different researchers have demonstrated divergent views and conclusions in the field of understanding the existence and the degree to which the type of test may bring about a positive or negative effect on students' performance and scores. This was also reported by Onaiba and Jannat's (2019) critical review of the pertinent literature. As cited in Onaiba and Jannat (2019, p. 18), Alderson et al. (1995, p. 45) proclaimed that "our understanding of test method effect is still so rudimentary that it is not possible to recommend particular methods for testing particular language abilities." Onaiba and Jannat concluded that this multiplicity of suppositions requires deeper investigations to be implemented to specify empirically the impact of the test method effect. This was also elaborated by Xie et. al. (2024) in their seminal paper entitled: A Systematic Literature Review of the Effect of Test Methods on English Learners' Reading Performance. The authors underscored the complex relationship between test methods and English reading comprehension, involving interactions between language proficiency, response formats, test types, and other variables (also see Arshad, et al., 2020). This could be owed to the claim that there is a limited understanding of the varying difficulty levels associated with the tests of MC and CR items (Sarsarabi and Sazegar, 2023). Consequently, further

research is needed to elucidate this relationship. To this end, this issue is of particular interest to this study.

Moreover, despite these studies, one of the major limitations of the research thus far is that there is no information available on the investigation of the effects of the selected response types, particularly MC and CR tests, as an aspect of the test method facet, on the Libyan EFL test takers' test performance. Hence, this study has gained its role in filling this gap. Besides, it is important to mention that reviewing pertinent literature has beneficially contributed to the choice and selection of the research design and methodology, as explicated in the upcoming section.

3. Methodology

3.1. Research Design

This research aims to determine the effect of deploying CR and MC test items on students' performance on a reading comprehension test. To this end, the study mainly employed a quantitative, experimental, research design. Creswell (2014) defined quantitative research as a research approach that endeavours to explain a phenomenon via gathering numerical data, which is then analyzed using statistical tools. The focus is on examining the relationships between variables and understanding the extent to which changes in one variable may cause changes in the other variable. Indeed, the study attempted to explore whether students achieve better marks in an MC item test than in a stem equivalent CR item test. To fit the purpose, the researcher has deployed a quasi-experimental research design, as this type of experimental research concerns placing participants in a nonrandomized group, purposive sampling (Gravetter & Forzano, 2012).

3.2. Instruments and Procedures

In this study, each participant took two tests of the same stem-equivalent items. The CR test was first administered and scored. So, in the first test, all items were in CR format; in the second, they were in selected response, particularly MC format. The participants' answers in the CR test were then directly compared with their option selection in the MC test to determine in which test item students did better (CR or MC).

The constructed-response (CR) test. This test consisted of twelve items to be answered by the participants, and the total mark was twelve (see Appendix A).

The multiple-choice (MC) tests. In addition to the key answer, in this test, the researcher selected the participants' incorrect answers they provided in the CR test and put them as distracters in the MC test. That is, the distractors in the MC test were based on the common errors from the CR test. The total mark of this test was also 12 (see Appendix B).

It is important to mention here that the participants were aware of the fact that the scores they gained from the two tests would be used and included in their semester assessment rubrics. Therefore, it was a real testing procedure, and knowing this would make students do their best to pass the tests. This will contribute to generating and drawing valid and reliable comparisons of the study's results.

The two tests were designed to test candidates' reading comprehension associated with their Reading III course. The reading passage has been taken from a book about exercises in reading comprehension by Tibbitts (1974).

The interview. This data collection method was utilized in this study because while marking the students' exam papers, the researcher noticed that some students had changed their answers to CR test items. Therefore, interviews were conducted with those students to discover the reasons behind that change and address the second research question. The interview offered

the participants a choice of three reasons: insufficient time, guessing, and the availability of answer choices in the MC format.

During the interviews, participants were asked to comment on their answers in the C/R test. They were also informed to raise the motive behind their option selection in the M/C test for one item where their response in the two tests differed. The researcher constructed all the instruments used to fit the Reading III students' level in the English Department at the Faculty of Education.

3.3. Context and Population

All participants were undergraduate students from the Department of English at the Faculty of Education Misurata University, all from the Reading-Three level.

The participants who attended the CR test were forty-one students, and the students who attended the MC test were thirty students. While the students who attended the constructed response test and did not attend the multiple-choice test were sixteen students, five students attended the multiple-choice test and did not attend the contracted response test. Overall, the students who attended the two tests were twenty-five students. So, the analysis procedures were based on this number of students. In addition, 10 of the 24 students participated in the interviews.

4. Data Analysis and Discussion

4.1. Analysis of Results

Two types of question items (CR and MC) were designed to address the first research question to determine the effect of each type on the students' scores.

The researcher reviewed the two tests before their administration. In addition, the tests were given to a colleague for review and verification. Then the researcher marked students' answer sheets and kept a record of the scores. Each student's total mark, out of twelve in both tests, was calculated and put in the table (see below), and then comparisons between the two attempts were reported using statistical tools.

A paired-sample t-test was conducted to compare marks of the first test (CR items test) and the second test (MC items test). As shown in Table 1 below, there was a statistically significant difference in the scores between the CR test (Mean=8.92, Std. Deviation =3.895) and MC test (Mean =16.25, Std. Deviation =5.396) case: t (11) =- 3.884, p = 0.003.

Table 1: Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean	P-value
Pair 1	CR-test	8.92	12	3.895	1.125	0.003
	MC-test	16.25	12	5.396	1.558	

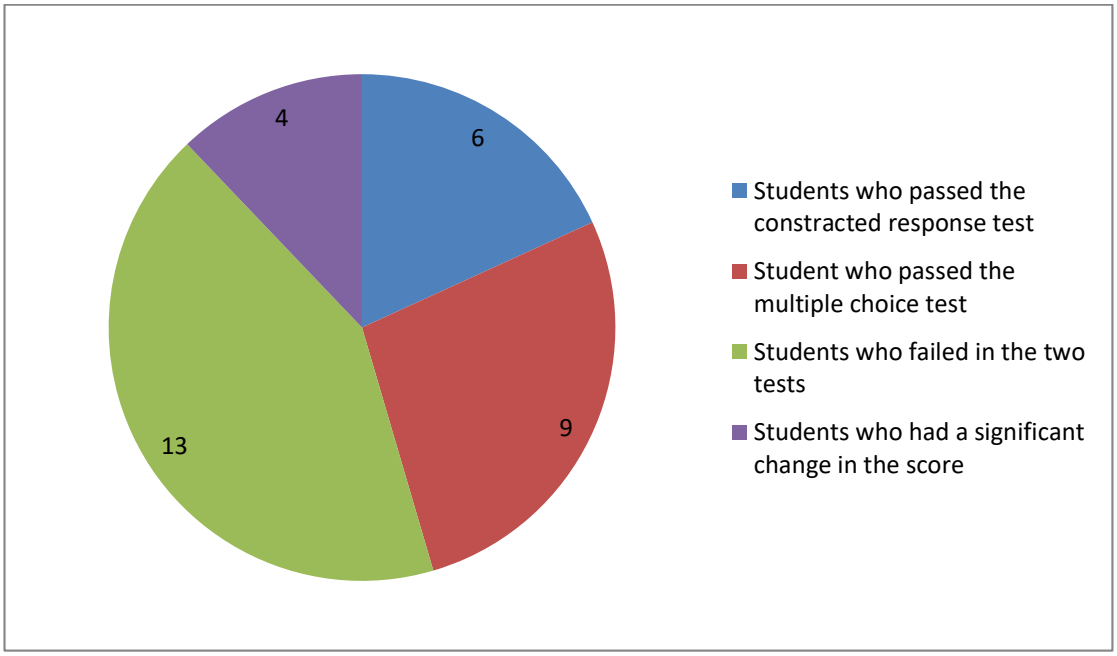
Moreover, Table 2 below visualizes significant changes in students' scores between the two tests. The two exams' results reveal that most students did better in the multiple-choice exam than in the constructed response exam. Scrutinizing the table shows that most students (n=17,

i.e., 68%) did better in the MC test than in the CR test, though some failed it. However, only 32% of students performed better in the CR test than in the MC test; besides, some of them failed this test as well.

Table 2: Students' scores in the two types of reading comprehension test

Students	Constructed Response Test Total mark (12)	MC Test Total mark (12)
1.	2	3
2.	6.5	8
3.	3.5	4
4.	4	5
5.	3	4
6.	5	8
7.	6	5
8.	3.5	9
9.	3.5	4
10.	2	8
11.	6	10
12.	2	3
13.	3.75	5
14.	5.5	4
15.	7.75	4
16.	3.5	1
17.	6	7
18.	5.5	2
19.	5.5	4
20.	4	6
21.	5.5	6
22.	2.5	4
23.	6.5	11
24.	5	4

Moreover, the above table can depict further results, as presented in the following pie chart.



The chart shows that only six students passed the constructed response test, while nine passed the multiple-choice test. Moreover, fourteen students failed the two tests.

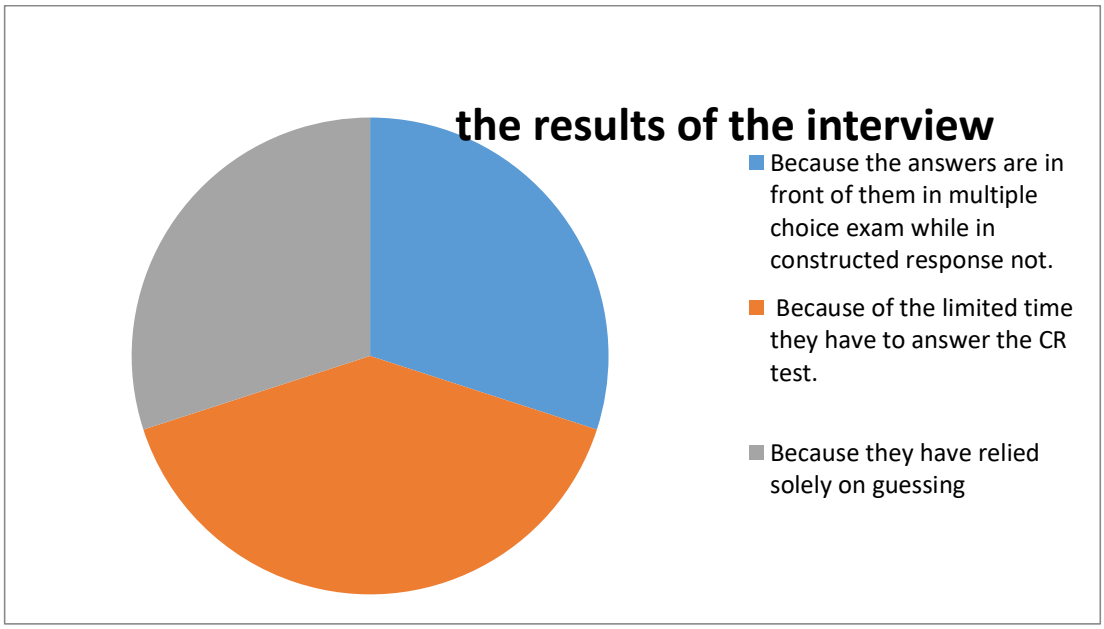
In addition, the above chart shows that some students' scores have changed significantly, getting higher marks in the multiple-choice test. This is also elaborated in Table 2 above (students Nos. 6, 8, 10, 11). The above results suggest that the type of test used does have an effect on the level of scores students gain in tests, particularly in an English reading comprehension test, our case in this study.

As mentioned earlier, some students have changed their answers after marking the MC items' exam papers, as some of them have provided answers that are different from the previous answers in the CR items test. To adequately address the second research question, interviews were carried out with ten students to know the reasons behind that change. The students were given three choices of the reasons behind changing their answers from those in the constructed response test to those in the multiple-choice test.

The three choices were:

- Because the answers are in front of them in the multiple-choice exam while constructed responses are not.
- Because of the limited time, they have to answer the CR test.
- Because they have relied solely on guessing.

The following Pie Chart shows the respondents' choices concerning their reasons for the change.



The above chart shows that 40% of the students changed their answers because of the time. However, 30% of the students changed their answers because they were in front of them in the multiple-choice exam. In addition, 30% of the students relied merely on the chance of guessing to answer the MC items. This result contributed to candidates' opinions when asked about their preferences for any of the test methods being assessed. They confessed their satisfaction with taking the MC test rather than the CR test.

4.2.Discussion

The results of this study evince findings that contribute to our understanding of the "test method effect," the impact of using certain test methods on students' test performance, particularly in a reading comprehension test. The findings also shed light on test takers' views towards taking a particular type of test.

This study's results demonstrate that the type of test format and construction may influence the scores that test-takers gain. For example, EFL students can do well on multiple-choice tests compared to their performance on constructed response-items tests when both tests assess the same skill, reading comprehension.

While some studies (Vasan et al., 2017; Peuker et al., 2013; Hassani and Maasum, 2012) have indicated that the differences in test takers' reading comprehension scores when assessed by various response formats are insignificant, several others, among which the current study, yielded conflicting results (Shohamy, 1984; Currie and Chiramanee, 2010; Karimi et al., 2014; Duran and Tufan, 2017; Lim, 2019; Polat, 2020). The studies demonstrated that the methods employed to assess language ability, reading comprehension in our case, significantly influence performance on language tests. Notably, the results of this study align with Lim's (2019) findings, which compared students' scores across two test facets, MCQs and OEQs, assessing the same skill. The results of Lim as well as the results of the current study revealed that students performed better on the MC test than on the CR test, suggesting that the former was more manageable than the latter.

Moreover, the results of this study concord with Polat's (2020) and Salehi and Sanjareh's (2013) study, which showed language school learners performed significantly better in MCQs

than in OEQs, indicating differences in item difficulty and discrimination levels between the two tests, with MCQ tests being easier.

Another study's findings, reviewed earlier in this paper, which aimed to assess 130 students' reading comprehension abilities, are compatible with this study, that of Duran and Tufan (2017). In particular, MC test items were more effective for students than CR questions, and students' performance on the MC test outperformed their performance on the CR test. Not only was this study's test performance compatible with the results of Duran and Tufan's study results but also this study's interview results echoed their interview results. Both studies indicate that students have positive views and attitudes towards the MCQ format while they negatively perceive the OEQ format. A group interview with some students, further, indicated that students were accustomed to, and hence, preferred the former over the latter.

The results of this study and other studies support the existence of the "test method effect," as candidates' scores differ depending on the type of test they underwent. The results indicate that candidates do better in the MC test format than in the CR test format due to their structured nature, allowing for quick recall of information, and enabling students to complete answering the questions within the allotted time. This advantage cannot be associated with the CR test format. Thus, this may further support the claim that learners' reading comprehension ability can be better assessed via MC rather than CR test items. This presupposition was also maintained by Hughes and Hughes (2020) and Arshad et al. (2020).

Moreover, although this study's findings indicate that the influence of MC format tests versus CR format tests on student performance scores is significant and efficient for evaluating a broad range of materials quickly (Polat, 2020; CR test items demand deeper understanding (Sarsarabi and Sazegar, 2023), preparing students for real-world applications that require critical thinking and problem-solving skills.

5. Conclusion

5.1. Synthesis of Results

This study investigated the effect of test format (constructed-response and multiple-choice) on students' performance in a reading comprehension test. The findings revealed a statistically significant difference in scores, with students performing better on the multiple-choice test. This aligns with several previous studies reporting similar results. Interviews with students provided insights into the reasons for this difference: time constraints, the presence of answer choices, and guessing were identified as key factors influencing performance on the multiple-choice test. Specifically, 40% of interviewed students attributed their answer changes to time constraints on the CR test, 30% cited the availability of choices, and another 30% admitted to guessing on the MC test. Furthermore, most students found the MC test format more appealing than the CR format. However, other factors may also influence performance. Indeed, test-takers' scores may fluctuate between two different test administrations due to individual attributes that were beyond the scope of this research but warrant future investigation.

These findings suggest that test format influences performance in reading comprehension. While multiple-choice items may facilitate quick recall and increase scores under timed conditions, CR items may be more suitable for assessing deeper understanding and critical thinking. This highlights the importance of aligning the test format with the specific learning objectives of the assessment. Therefore, this paper acknowledges that each format has its

strengths and weaknesses; it is not a matter of which is "better" in all situations but rather which is more appropriate for specific assessment goals.

5.2.Limitations and Recommendations for Future Research

Every research study has limitations, and this study is no exception. The small sample size (n = 24) is a key limitation. A larger sample would likely yield more robust and reliable results; consequently, the generalizability of the findings requires further investigation. Moreover, this study focused solely on reading comprehension, using only two sets of tests. Expanding the comparison to other language skills, such as listening, writing, and vocabulary use, could provide richer insights. Therefore, the exclusive focus on reading constitutes another limitation. Several research avenues could be explored in future studies on this topic. First, this research involved statistical comparisons between CR and MC tests regarding their psychometric properties and students' performance. Future studies could compare other test formats (e.g., true-false, matching, fill-in-the-blanks) with MC or CR items to provide a more comprehensive understanding of test method effects. Second, this study primarily addressed lower-level cognitive domain levels. Investigating higher-level cognitive domain levels with different question types in foreign language assessment would be a valuable contribution to the field. This underscores the need for a balanced assessment approach in educational settings.

Associate Professor of English Applied Linguistics and TESOL, Department of English, School of Language, Libyan Academy for Postgraduate Studies, Misurata Branch, Libya

References

Arshad, A., Shakir, A., & Ahmad, M. (2020). A Review on the Principles of a Reading Comprehension Test Construction to assess the Test Takers at different levels. *Psychology and Education*, *57*(8), 1290-1302.

Campbell, J. R. (1999). Cognitive processes elicited by multiple-choice and constructed response questions on an assessment of reading comprehension. Doctoral Dissertation, Temple University. (UMI No. 9938651)

Chehrazad, M. H., & Ajideh, P. (2015). Effects of Different Response Types on Iranian EFL Test Takers' Performance. *Iranian Journal of Applied Language Studies*, *5*(2), 29-50.

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches*. (3rd ed.). SAGE.

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & HI.

Cunningham, G. K. (1998). *Assessment in the classroom: constructing and interpreting texts*. Psychology Press.

Currie, M., & Chiramanee, T. (2010). The effect of the multiple-choice item format on the measurement of knowledge of language structure. *Language Testing*, *27*(4), 471-491.

Duran, E., & Tufan, B. S. (2017). The Effect of open-ended questions and multiple-choice questions on comprehension. *International Journal of Languages' Education and Teaching, 5(1),* 242-254.

Fulcher, G. & Harding, L. (2024) The Routledge Handbook of Language Testing. Routledge.

Gravetter, F. J., & Forzano, L. B. (2012). *Research methods for the behavioral sciences* (4th ed.). Cengage Learning.

Hancock, G. R. (1994). Cognitive complexity and the comparability of multiple-choice and constructed-response test formats. *Journal of Experimental Education, 62*(2).143-157.

Hassani, L., & Maasum, T. N. R. T. M. (2012). A study of students' reading performance in two test formats of summary writing and open-ended questions. *Procedia-Social and Behavioral Sciences*, *69*, 915-923.

Hughes, A. & Hughes, J. (2020). *Testing for language teachers*. Cambridge University Press & Assessment.

Kaçar, K. (2023). Commonly Used Techniques in Testing Foreign Language Skills and Communicative Language Teaching. *Sinerji Uluslararası Alan Eğitimi Araştırmaları Dergisi*, *4*(2), 68-85.

Karimi, F., Tabrizi, H. H., Aval, A. N. & Khorvash, F. (2014). Test Method Facet: Is Response a Facet? *International Journal of Current Life Sciences*, 4(11). 9010-9015.

Kennedy, P. & Walstad, W. B. (1997). Combining multiple-choice and constructed-response test scores: An economist's view. *Applied Measurement in Education,10*(4), 359-375.

Lim, H. (2019). Test format effects: a componential approach to second language reading. *Language Testing in Asia*, *9*, 1-22.

Martinez, M. E. (1999). Cognition and the question of test item format. *Educational Psychologist*, *34*(4), 207-218.

Onaiba, A. E., & Jannat, F. B. (2019). Test Method Effect and Test-Takers Scores: A Critical Review of the Pertinent Literature. *Scientific Journal of Faculty of Education, Misurata University*, 1(14), 3-29

Polat, M. (2020). Analysis of Multiple-Choice versus Open-Ended Questions in Language Tests According to Different Cognitive Domain Levels. *Novitas-ROYAL (Research on Youth and Language)*, *14*(2), 76-96.

Rashidi, N., & Safari, F. (2014). Does the type of multiple-choice item make a difference? The case of testing grammar. *International Journal of Language Testing*, *4*(2), 175-186.

Shohamy, E. (1984). Does the testing method make a difference? The case of reading comprehension. *Language Testing*, *1*(2), 147–170.

Tibbitts, E. T. (1974). Exercises in Reading Comprehension: Huntsmen Offset Printing Pte Ltd, Singapore.

Salehi, M., & Sanjareh, H. B. (2013). The impact of response format on learners' test performance of grammaticality judgment tests. *Journal of Basic and Applied Scientific Research, 3, 2: 1335, 1345*.

Sarsarabi, S. S., & Sazegar, Z. (2023). Investigating Different Kinds of Stems in Multiple-Choice Tests: Interruptive vs. Cumulative. International Journal of Language Testing, 13(2), 170-187.

Shahivand, Z., Paziresh, A., & Raeeszadeh, A. (2014). The effects of test formats on the performances of Iranian EFL students. *Theory and Practice in Language Studies*, *4*(2), 366.

Tsagari, C. (1994). Method effects on testing reading comprehension: How far can we go? Unpublished MA thesis, University of Lancaster, UK.

Vasan, M. C. A., DeFouw, D. O., Holland, B. K., & Vasan, N. S. (2017). Analysis of testing with MC versus open-ended questions: Outcome-based observations in an anatomy course. *Anatomical sciences education*, *11*(3), 254-261.

Wainer, H., & Thissen, D. (1993). Combining multiple-choice and constructed-response test scores: Toward a Marxist theory of test construction. *Applied Measurement in Education*, *6*(2), 103-118.

Walstad, W. B., & Becker, W. E. (1994). Achievement differences on multiple-choice and essay tests in economics. *American Economic Review, Papers and Processing*, *84*(2), 193-196.

Wolf, D. F. (1993). A comparison of assessment tasks used to measure FL reading comprehension. *The Modern Language Journal*, *77*(iv), 473–489.

Xie, S., Singh, C. K. S., & Wong, W. L. (2024). A Systematic Literature Review of the Effect of Test Methods on English Learners' Reading Performance. *Journal of Humanities, Arts and Social Science*, 8(7), 1591-1599

Appendix(A)
Constructed Response Test

Instruction:
Read the text carefully, and then answer the questions.

Students working for their degree at a university are called undergraduates. When they take their degree, we say that they graduate, and then they are called graduates. If they continue studying at university after they have graduated, they are called post-graduates. Full-time university students spend all their time studying. They have no other employment. Their course usually lasts for three or four years. Medical students have to follow a course lasting for six or seven years. Then they graduate as doctors. In Britain, full-time university students have three

terms of about ten weeks each year. During these terms, they go to lectures, or study by themselves. Many students become members of academic societies and sports clubs and take part in their activities. Between the university terms, they have vacations (or holiday periods). Their vacations are long, but of course, they can use them to study at home.

Some universities, like Oxford and Cambridge in England, are residential. This means that during the university terms, the students live in a university college or hostel or in lodgings chosen from an official list. The university and college buildings are often ancient, and amongst them are fine examples of ancient architecture. Other universities are non-residential. Some of the students at these universities can live in a university hostel, but many live at home or in lodgings and have to travel daily to their lectures. Large cities often have universities of this kind. Sometimes the students have to spend quite a lot of time on their journeys, so they cannot join in student activities as easily as students in residential universities can.

Full-time students are also called internal students because they spend all their time at university. There are also external students who cannot attend the university full-time but who are studying for its examinations. They are part-time students. They have to do other work during the day, usually to earn their living, and they study in the evening. External students are often older than full-time internal students. Sometimes they can attend lectures in the evening, but many of them have to study by correspondence. They write at home the work that is set by their tutors. Then they have to post this work to their tutors, and the tutors post corrections and advice back to them.

In some countries,' universities of air' can now help students to gain degrees. After special lectures on radio or television, these students too have to send set work for correction by correspondence. However, for a few weeks each year, they can attend special vacation courses at universities.

For admission to any degree course, a student has passed qualifying examinations. In Britain, there are not enough places for every secondary school student, so these examinations are competitive. This means that only the students with the highest marks can gain admission. However, a few older people are able to gain admission without the full qualifications, if the university thinks that they are suitable candidates.

Now, answer these questions:

1-Why don’t some students have other employment to do?

……

2- How many years does each course last in a British university for full-time students?

……

3-What’s the difference between residential and non-residential universities?

……

4-How can full-time students’ study by correspondence?

……

5- Why do we call full-time students’ internal students?

……

6- In what way(s) external students are different from internal students?

……

7-Who is a part-time student?

……

8- Why is there a qualifying examination in Britain?

……

9- What is an academic society?

……

10-What do we mean by a candidate?

……

11- Why the qualifying examinations are competitive in Britain?

……

12- What do you think is the suitable title for the above passage?

……

Appendix(B)
The multiple-choice test

Read the text carefully, and then answer the questions.

Students working for their degree at a university are called undergraduates. When they take their degree, we say that they graduate, and then they are called graduates. If they continue studying at university after they have graduated, they are called post-graduates. Full-time university students spend all their time studying. They have no other employment. Their course usually lasts for three or four years. Medical students have to follow a course lasting for six or seven years. Then they graduate as doctors. In Britain, full-time university students have three terms of about ten weeks each year. During these terms, they go to lectures, or study by themselves. Many students become members of academic societies and sports clubs and take part in their activities. Between the university terms, they have vacations (or holiday periods). Their vacations are long, but of course, they can use them to study at home.

Some universities, like Oxford and Cambridge in England, are residential. This means that during the university terms, the students live in a university college or hostel or lodgings chosen from an official list. The university and college buildings are often very old, and amongst them are fine examples of ancient architecture. Other universities are non-residential. Some of the students at these universities can live in a university hostel, but many live at home or in lodgings and have to travel daily to their lectures. Large cities often have universities of this kind. Sometimes the students have to spend quite a lot of time on their journeys, so they cannot join in student activities as easily as students in residential universities can.

Full-time students are also called internal students because they spend all their time at university. Some external students cannot attend the university full-time but who are studying for its examinations. They are part-time students. They have to do other work during the day, usually to earn their living, and they study in the evening. External students are often older than full-time internal students. Sometimes they can attend lectures in the evening, but many of them have to study by correspondence. They write at home the work that is set by their tutors. Then they have to post this work to their tutors, and the tutors post corrections and advice back to them.

In some countries,' universities of the air' can now help students to gain degrees. After special lectures on radio or television, these students too have to send set work for correction by correspondence. However, for a few weeks each year, they can attend special vacation courses at universities.

For admission to any degree course, a student has to pass qualifying examinations. In Britain, there are not enough places for every secondary school student, so these examinations are competitive. This means that only the students with the highest marks can gain admission. However, a few older people can gain admission without the full qualifications, if the university thinks that they are suitable candidates.

Choose the appropriate answer for the items below:

1-Full-time students have no other employment to do because:

A. They are always very busy with their study.

B. They do not like to work.

C. They have not graduated yet.
D. They become members of academic societies.

2-Each course for full-time non-medical students lasts for:

A. Six or seven years.
B. Three or four years.
C. Six or seven months.
D. Three or four months.

3-In residential universities:

A. Students have to travel to their lectures.
B. The buildings are very modern.
C. College or hostel or lodging are chosen from an informal list.
D. Students can join activities easily.

4-In non-residential universities:

A. The students live in a hostel or lodging only
B. They do not spend any time on their journeys
C. Students can join activities easily
D. Students have to travel to their lectures

5-The meaning of studying by correspondence is:

A. A course whose students send their work by post to their tutors
B. Studying in the evening
C. Spending all the time studying
D. Passing the qualifying examinations

6-Internal students:

A. Are Part-time students
B. Are Full-time students
C. Spend some of their time in university
D. Are older than full-time students

7- External students:

A. They can attend full-time university programs.
B. Are part-time students**.**
C. Are full-time students**.**
D. They are younger than internal students.

8- Qualifying examination is:

A. It is an exam in which candidates can get high marks easily.
B. Enables universities to reduce the number of new students.
C. Prevents excellent students to be accepted into universities.
D. Makes the admission procedure non-competitive

9-The suitable title for the passage is:

A. Students' degree
B. Qualifying examination
C. Universities
D. Academic Societies

10-The person who is trying to get a university place is:

A. Full-time student
B. Candidate

C. External student
D. Part-time student

11-Large cities have universities of:

A. Residential features
B. Non-residential features
C. Both of the above features
D. None of the above features

12-……. holds a meeting for members interested in a particular university subject.

A. Academic Society
B. University of air
C. Qualifying examination
D. Studying by correspondence

Appendix(C)
The interview

Why did you change the answers from the constructed-response test to the multiple-choice test

- Because the answers are in front of them in the multiple-choice exam while constructed responses are not.
- Because of the limited time, they have to answer the CR test.
- Because they have relied solely on guessing