Investigating the Effect of Changing Answers in Multiple-Choice Questions of Reading and Listening Tests in EFL Context

Parviz Ajhideh
Professor of TEFL, Department of English Language and Literature,
Faculty of Persian & Foreign Languages, Tabriz University, Tabriz, Iran
E-mail: parvizaj@gmail.com

Solmaz Saeedi Manesh*(Corresponding Author)
PhD Candidate of TEFL, Department of English Language and Literature,
Faculty of Persian & Foreign Languages, Tabriz University, Tabriz, Iran
E-mail: s_saidi_m@yahoo.com

Abstract
This study investigates whether the practice of answer changing on multiple-choice questions (MCQs) is beneficial for Iranian students’ overall test performance in EFL context. Several studies have shown that answer changing in multiple choice examinations is generally beneficial for improving scores. On the other hand, some studies have indicated that answer changing in multiple choice examinations results in an increased number of wrong answers rather than improved scores. In this study, two types of listening and reading comprehension multiple-choice tests were used in two different proficiency levels. The test was administered to 269 EFL students, both males and females, in Kish Language Institution, Tabriz branch. Results show that in both proficiency levels among males and females, answer-changing leads to increased number of wrong answers. Findings suggest that we should encourage students not to change their initial answers unless they have scrutinized their initial answers precisely and there are more plausible alternatives. The findings were compatible with the traditional perception ‘go with your first response’.

Keywords: Multiple-Choice Questions, Answer Changing, Reading Comprehension, Listening Comprehension, Proficiency Level, and Gender
Introduction

Nowadays multiple choice questions (MCQs) are one of the most common objective tests that are used in evaluating the performance and progress of students and estimating different educational learning systems. Multiple choice items consist of a stem, the correct answer, keyed alternative, and distractors. The stem is the beginning part of the item that presents the item as a problem to be solved, a question asked of the respondent, or an incomplete statement to be completed, as well as any other relevant information. The options are the possible answers that the examinee can choose from, with the correct answer called the key and the incorrect answers, distractors. Only one answer can be keyed as correct. A correct answer earns a set number of points toward the total mark, and an incorrect answer earns nothing. However, tests may also penalize students for incorrect answers to discourage guessing.

There are several advantages to multiple choice tests. If item writers are well trained and items are quality-assured, it can be a very effective assessment technique. Multiple choice tests often require less time to administer for a given amount of material than would tests requiring written responses. On the other hand, the most serious disadvantage is the limited types of knowledge that can be assessed by multiple choice tests. Multiple choice tests are best adapted for testing well-defined or lower-order skills. Problem-solving and higher-order reasoning skills are better assessed through short-answer and essay tests. However, multiple choice tests are often used, not because of the type of knowledge being assessed, but because they are more affordable for testing a large number of students in terms of time and budget.

Another disadvantage of multiple choice examinations is that a student who is incapable of answering a particular question can simply select a random answer and still have a chance of receiving a mark for it. If a student randomly guesses an answer, there is usually for him/her a 25 percent chance of getting it correct on a 4 answer choice question. It is common practice for students with no time left to give all remaining questions random answers in the hope that they will get at least some of them right. A system of neutralizing the effects of random selection is formula scoring, in which a score is proportionally reduced based on the number of incorrect responses and the number of possible choices. Despite being challenged, the format remains popular due to its practicality, reliability, and cost effectiveness.

Based on the obtained results, these types of evaluating exams improve learning in students, diagnose learners’ problems, and reduce learners’ weaknesses and educational programs’ deficiencies. So factors that have an impact on the application of these tests and their results are significant to consider. Educators agree that the skills students need to have to do well on a test are dependent on the skill area assessed and the exam question format (Scruggs & Marsing, 1988; Burdess, 1991; Foster et al., 1999).
A frequently used format in large-scale, high-stakes English language proficiency exams is the multiple-choice question (MCQ). At the end of the test, some of the students, if they have enough time, review the questions and their selected answers, and if they have doubt about them, they may change their previously selected answers. One may have heard the sentence that goes: “Don’t change the answer; the preselected answer in MCQs is commonly right.” This is based on the assumption that one’s brain knows the right answer before one starts to think. When one starts to think and pay attention to possibilities and other choices, one gets confused and gets far from the right answer. If the above sentences are right, it is better not to change the preselected answer until one has reasonable and sufficient reasons. Apparently with changing the answer, three possibilities are in hand:

1- Changing the wrong answer to the right one that leads to a higher score,
2- Changing the right answer to the wrong one that leads to a lower score,
3- Changing the wrong answer to another wrong one that makes no change in the resulting score.

Most of the teachers in preparation classes advise learners to go with their first response. The question arises as to whether this is a research-based practice (Torrence, 1986), or simply anecdotal advice or conventional wisdom on the part of the teacher (Al-Hamly & Coombe, 2005). Most skills-based instructional programs that focus on test-taking strategies tend to give ‘common sense’ suggestions rather than empirically verifiable strategies to students (Stough, 1993).

Several researchers have examined the answer-changing behaviors of students taking objective tests. First studies in literature showed that students get higher scores with changing the wrong answers to the right ones (Lehman, 1928; Mathews, 1929; Jarrett 1946), but later studies have led to doubt in this idea and have implicated other factors such as the level of students’ knowledge, the level of the difficulty of the tests and the precise amount of these changes (Reile & Briggs, 1952; Bath, 1967). Most of the research conducted is aimed at testing the accuracy of ‘first impressions’ in test-taking. Mostly, it is believed that ‘one should not change answers on objective tests because initial reactions to test questions are intuitively more accurate than subsequent responses’ (Benjamin, Cavell, & Shallenberger, 1984; Hanna, 1989).

The first empirical study on answer changing dates back to 1929, when Mathews investigated answer-changing behaviors of college level students in educational psychology courses and found that more than 53% of the answers changed on MCQs were from Wrong answer to Right answer (WR), approximately 22 percent were from the Right answer to a Wrong answer (RW) and the remaining changes were from Wrong answers to other Wrong answers (WW). The basic finding of this first study was that for every point lost, roughly 2 to 3 points were
gained. Findings since have noted a remarkable consistency of results by several later researchers (Bath, 1967; Foote & Belinky, 1972; Mueller & Wasser, 1977; Videler & Hansen, 1980; Benjamin et al., 1984; Geiger, 1991).

Since 1929, a lot of studies have been published concerning answer-changing behaviors on objective tests. The most consistent findings in these studies have indicated that there is nothing inherently wrong with changing first answers. In fact, empirical evidence uniformly indicates that only a small percentage of items are actually changed, although most test-takers are answer changers. The results showed that most of these changes are from Wrong to Right answers. As a result, most answer changing leads to a higher score. The data across twenty separate studies indicate that the percentage of “Right to Wrong” changes is 20.2%, whereas the percentage of “Wrong to Right” changes is 57.8%, nearly triple (Benjamin et al., 1984). Skinner (1983) found that 51% changed answers from Wrong to Right (WR), 26.3% changed answers from Right to Wrong (RW), 22.3% changed answers from Wrong to Wrong (WW). Female participants made more changes (double), although male participants made 54% successful changes.

Previous studies have investigated a number of different variables and their relationship to answer changing. Researchers have examined individual characteristics such as gender. For example, Reile and Briggs (1952) found that females changed answers more often, but overall gained fewer points than male students. Bath (1967), however, found that females gained more points than males, and several studies have found no gender relation to switching behavior or to net gain (Copeland, 1972; Mueller & Shwedel, 1975; Geiger, 1991). Reiling and Taylor (1972) reported that gains were made from changing answers but final grade, gender, and analytical questions had no impact on answer-changing gains. Penfield and Mercer (1980) found gains from answer changing although there was no gender difference. Also high-scoring participants made more changes than low-scoring ones.

Some studies investigated item difficulty factor in changing answers. Videler and Hansen (1980) found that changes were more likely to be made on difficult rather than easy items. Most participants changed answers from Wrong to Right (WR) but relationship between item difficulty and direction of answer change was insignificant. In another study, Jacobs (1972) concluded that students changed more answers on low to moderate difficulty items. In a related area of research, Pressley and Ghatala (1988), Pressley, Ghatala, Woloshyn, and Pirie (1990) and Zakay and Glicksohn (1992) examined student confidence on multiple-choice questions. They found that students often believed that there was a good chance that their initial answer was correct when, in fact, it was incorrect.

Wagner (1998) conducted a study in science content area to investigate participants’ learning styles. He found that participants changed 2.02 answers and
gained an average of .83 points as a result of their changes. The more impulsive participants changed more answers and gained more points than the more reflective ones. Field-dependence/independence variable was not significantly related to answer-changing behavior. So, changing answers on tests was beneficial but the benefit differed depending on the participant’s level of impulsivity/reflectivity. Regarding test taker’s ethnicity, Payne (1984) found that black participants made more answer changing as they had higher test anxiety, but he could not find any significant difference in gender variable.

Researchers have examined other possible effects on answer-changing behavior such as the differences due to item type (Smith, White, & Coop, 1979; Geiger, 1991), the cognitive styles of students (Friedman & Cook, 1995), students’ perceptions of answer changes (Mathews, 1929; Skinner, 1983; Geiger, 1991; Prinsell, Ramsey, & Ramsey, 1994), and answer changing as a function of test anxiety. Green (1981) found that high test anxious students made more answer changes than low test anxious students, and all students benefited from answer changing. Researchers have found that there exists a positive relationship between advantages from answer changing and overall student performance (Friedman & Cook, 1995). Test wiseness is another area that has received empirical attention. Millman, Bishop, and Ebel (1975) have indicated that the careful changing of one’s original answer selection is a basic aspect of being test wise. Green (1981) also examined item difficulty. Her findings suggest that more answers were changed on difficult rather than on easy items by students who experienced both high and low levels of test anxiety. Pascale (1974) investigated the effect of answer changing in different proficiency levels within different genders; in his study, males performed better and all participants improved their scores when they changed answers. However, no effect of proficiency level on test score was reported.

The issue of answer changing and students’ performance on tests in content areas has received little attention in the field of language learning, in general, and English as a second language (ESL) or English as foreign language (EFL), in particular. The present study aims to investigate test-takers’ performance when taking listening and reading tests. The primary purpose of the present study is to determine the effect of answer changing on overall test scores.

**Research questions**

The research questions of the present study are:

RQ 1: Is there any relationship between students’ proficiency level and the quantity and quality of answer changes made?

RQ 2: Is there any relationship between students’ gender and the quantity and quality of answer changes made?
Research hypothesis

The research hypotheses in the present study are:

AH 1: There is a relationship between students’ proficiency level and the quantity and quality of answer changes made.

NH 1: There is no relationship between students’ proficiency level and the quantity and quality of answer changes made.

AH 2: There is a relationship between students’ gender and the quantity and quality of answer changes made.

NH 2: There is no relationship between students’ gender and the quantity and quality of answer changes made.

Methodology

Participants

A total of 269 EFL adult students (111 males, 158 females) ranging in age from 17 to 35 at Kish Language Institution, Tabriz Branch, participated in this study (see Table 1). The participants were enrolled in General English classes in two different proficiency levels (132 elementary level and 137 advanced level) based on placement test (see Table 2).

<table>
<thead>
<tr>
<th>Institute</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>158</td>
<td>111</td>
<td>269</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Institute</th>
<th>Elementary</th>
<th>Advanced</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>132</td>
<td>137</td>
<td>269</td>
</tr>
</tbody>
</table>

Instruments

To gather the data needed for the study two types of tests were used; a) Listening Test and b) Reading Test. The test was designed in form of MCQs (Multiple-Choice Questions) with just one right answer for each question. The difficulty levels of the tests were in accordance with the proficiency level of the subjects.

Procedures

Participants were selected randomly based on their proficiency level from different classes which were held in the Institution meanwhile. They were also asked to take
part in an oral placement test to ensure that their proficiency level is appropriate. Data were collected in two separate administrations for two different proficiency levels over two days. First, Listening Test was administered for each group, using a recorded type of listening comprehension test followed by 10 multiple-choice question items. Participants were asked to read the question prior to playing the recording. The recording was played twice, so the participants had a chance to change their initial answers. The allocated time for completing the listening part was 20 minutes. Next, reading comprehension test was administered, which included a general topic passage followed by 10 multiple-choice question items. The allocated time for completing the reading part was 10 minutes.

Scoring

The answer sheets were hand-marked and were examined for erasures indicating answer changes. A count of answer changes was made and recorded, and then each answer change on each answer sheet was further examined and classified according to the direction of the change. These directions were identified as: Right to Wrong (R–W), Wrong to Right (W–R), and Wrong to Wrong (W–W). Recurrent changing of answers was defined, if a participant had changed answers for a question twice or more.

Results

The results found that of the 269 students who took part in this study, 198 (73.6%) had changed their preselected answers, and 71 (26.3%) had not changed their answers. Students made 549 answer changes (ACs) which is 10.2% of actual test answers which were 5380 answers.

RQ 1: Is there any relationship between students’ proficiency level and the quantity and quality of answer changes made?

Table 3 shows changing the answers based on the proficiency level of students. The direction of changes in answers was analyzed and the review of results showed that 269 students had changed 549 answers in total. As Table 3 shows, 369 questions - 67% of them - which were reviewed had correct answers and the students had changed the right answers to wrong ones, and the remaining 180 questions - the remaining 33% - had wrong preselected answers. Of the total 549 reviewed MCQs, 107 (20%) answers were wrong, which had been changed to correct ones, 369 (67%) were correct, which had been changed to wrong ones, and 73 (13%) were wrong, which had been changed to another wrong answer. Therefore, the greatest percentage of answer changes was from Right to Wrong in both proficiency levels. In advanced level, 141 (57%) answers were changed from Right to Wrong, 57 (23%) from Wrong to Right, and 50 (20%) from Wrong to Wrong. On the other hand, in elementary level, 228 (76%) answers were changed from Right to Wrong,
50 (16%) from Wrong to Right, and 23 (8%) from Wrong to Wrong. The total number of changes done by advanced students was 248 (9.0%), while the total number of changes done by elementary students was 301 (11.4%). It indicates that elementary students tend to change more answers than do advanced students.

Table 3. Number and direction of answer changes based on proficiency level

<table>
<thead>
<tr>
<th></th>
<th>Right-Wrong</th>
<th>Wrong-Right</th>
<th>Wrong-Wrong</th>
<th>Total ACs</th>
<th>Total As</th>
</tr>
</thead>
<tbody>
<tr>
<td>Advanced</td>
<td>141(57%)</td>
<td>57(23%)</td>
<td>50(20%)</td>
<td>248</td>
<td>2740</td>
</tr>
<tr>
<td>Elementary</td>
<td>228(76%)</td>
<td>50(16%)</td>
<td>23(8%)</td>
<td>301</td>
<td>2640</td>
</tr>
<tr>
<td>Total</td>
<td>369(67%)</td>
<td>107(20%)</td>
<td>73(13%)</td>
<td>549</td>
<td>5380</td>
</tr>
</tbody>
</table>

RQ 2: Is there any relationship between students’ gender and the quantity and quality of answer changes made?

Table 4 shows the changing of the answers based on the gender of the students. The changes in the answers were analyzed and the results showed that 549 answers had been changed in total. Considering the direction of answer changes, as Table 3 shows, 368 (67%) questions had correct answers and the students had changed the right answers to the wrong ones, 103 (19%) had wrong answers and the students had changed the wrong answers to the right ones, and 78 (14%) had wrong answers, which had been changed to another wrong answer. Therefore, the greatest percentage of answer changes was from Right to Wrong. Among males, 131 (66%) answers were changed from Right to Wrong, 39 (20%) from Wrong to Right, and 27 (14%) from Wrong to Wrong. On the other hand, among females, 237 (67%) answers were changed from Right to Wrong, 64 (18%) from Wrong to Right, and 51 (15%) from Wrong to Wrong.

Another point is the total number of changes which were done by males and females. As Table 4 shows, males had changed 197 (8.8%) answers, while the total number of answer changes by females was 352 (11.1%). Previous studies have in general shown that females make more total answer changes than males, so the results of this study are consistent with other studies in the literature (Reile & Briggs, 1952; Bath, 1967; Foote & Belinky, 1972; Skinner, 1983).

Table 4. Number and direction of answer changes based on gender

<table>
<thead>
<tr>
<th></th>
<th>Right-Wrong</th>
<th>Wrong-Right</th>
<th>Wrong-Wrong</th>
<th>Total ACs</th>
<th>Total As</th>
</tr>
</thead>
<tbody>
<tr>
<td>Males</td>
<td>131(66%)</td>
<td>39(20%)</td>
<td>27(14%)</td>
<td>197</td>
<td>2220</td>
</tr>
<tr>
<td>Females</td>
<td>237(67%)</td>
<td>64(18%)</td>
<td>51(15%)</td>
<td>352</td>
<td>3160</td>
</tr>
<tr>
<td>Total</td>
<td>368(67%)</td>
<td>103(19%)</td>
<td>78(14%)</td>
<td>549</td>
<td>5380</td>
</tr>
</tbody>
</table>
Discussion

The results of the current study showed that around 73% of the students had changed their preselected answers, which are comparable with other studies (between 40 and 90% with the mean of 55%). The direction of the changes for students in our study was different from other studies. In most of the studies, about 50% had corrected their wrong answers, 25% had changed their correct answers to wrong ones, and 25% had changed their wrong answers to other wrong ones, whilst in our study it was different. In other words, in most studies, there was a high chance for correcting wrong answers to correct ones, whereas, in our study, they were about 67% correct to wrong changes, 19% wrong to correct changes, and 14% wrong to wrong changes.

Several surveys studied the opinion of students about reviewing answers to MCQs and explained that most of the students expressed that first selected answer would be correct. A few surveys studied the feelings and the understandings of the students about reviewing answers to MCQs, and mostly confirmed the above conclusion that first selected answer would be correct. In 1991, Geiger studied the feelings and the functions of the students and declared that most of them believed that changing the preselected answers led to a decrease in their final scores.

Considering gender factor, Reile and Briggs (1952) showed that females changed their preselected answers more than males and with a meaningful difference, but the total enhancement of their score is less than males. In contrast, the study carried out by Bath (1967) showed that females got a better enhancement than males. Yet, in another study by Geiger (1991), it was shown that females rendered such a behavior more than males did, with a meaningful difference, but no enhancement in their final score was found in comparison with males. The findings of the present study was in line with these studies, since here too females changed more answers than males, but the direction of the most changes made by both males and females was from Right to Wrong. In other words, neither males nor females could enhance their final scores by changing their preselected answers.

Considering proficiency factor, elementary students tended to change more answers than did advanced students. In a study by Al-Hamly and Coombe (2005), the new concept of anxiety in the Test Session, its impact on their behaviors, and the corresponding results were analyzed. The results showed that there was not a meaningful correlation between anxiety and changing the preselected answer. But, when the students were divided into two groups with mild and severe anxiety, it was shown that the group with less anxiety had less change in the preselected answers, as well. Therefore, in higher proficiency levels, due to lower rates of anxiety, participants tend to change less answers and, therefore, they have higher rate of confidence in their preselected answers. However, more studies are needed to clarify the issue.
In the current study, no meaningful change was seen, neither regarding the gender impact on changing the preselected answer, nor the proficiency level impact. Also, most of the studies did not show meaningful change based on gender (Copeland, 1972; Mueller & Shwedel, 1975; Geiger, 1991; Reiling & Taylor, 2003; Al-Hamly & Coombe, 2005).

The studied sample in our survey was similar in changes of number and gender to other studies, but in terms of quality of changes a considerable difference was found. In most studies, around half of the changes had led to an increase in final scores, but, in the current study, more than half of the participants got lower scores due to their changing the preselected answers, which is not acceptable, a fact that asks for more studies to be carried out in this regard.

Recommendations for further research

It should be acknowledged that the present study is just an attempt to shed light on the answer-changing behavior and practices of EFL learners in EFL setting. However, more focused researches are needed before definitive conclusions can be made. For example, future studies might investigate the effects of student test anxiety level on answer changing. An investigation into the anxiety level of students and how this affects answer changing would be of interest to educators in the field of EFL/ESL. More studies are also needed to investigate other variables such as the difficulty level of multiple-choice questions, which might affect answer changing. Researchers might also investigate the relationship between test preparation skills and other personal variables such as students’ various cognitive styles and their possible impact on answer changing behaviors.

Conclusion

The present study found that changing answers on a multiple-choice questions test was quite detrimental as most changes made were from Right to Wrong, which resulted in losing scores. These findings suggest that we should encourage students not to change their initial answers unless they have scrutinized their initial answers carefully to see if there were more plausible alternatives. The findings were compatible with the traditional perception which holds: ‘go with your first response’.

References


Authors Biography

**Professor Parviz Ajhideh** was born in Ardebil, Iran, in 1957. He received his B.A. degree in English Literature from Allameh Tabataba’i University, Tehran, Iran, in 1979, his M.A. degree in English Teaching from Tarbiat Modares University, Tehran, Iran, in 1989, and his Ph.D. degree in English Teaching from Allameh Tabataba’i University, Tehran, Iran, in 2003. He is an academic member in the Department of English at the University of Tabriz, Iran. His main research interests include Testing, ESP, Reading, and Translation. He has published many articles in national and international journals.

**Solmaz Saecidi Manesh** was born in Tabriz, Iran, in 1988. She received her B.A. degree in English Literature and her M.A. degree in English Teaching from the University of Tabriz, Tabriz, Iran. She is currently a Ph.D. candidate in English Teaching at the University of Tabriz, Tabriz, Iran.