International Journal of Humanities Social Sciences and Education (IJHSSE)
Volume 3, Issue 1, January 2016, PP 24-30
ISSN 2349-0373 (Print) & ISSN 2349-0381 (Online)
www.arcjournals.org

The Effect of Guessing on the Parameters of Abilities and Items in Multiple-Choice Tests

[1]ImanPakneJad, [2]Mohammadreza Falsafinejad

[1]PhD's Student of Measurement and Assessment in Tehran University of Alameh Taba Tabaei and the Faculty Member of Farhangian University
[2]Faculty member of Alameh Tabataei University
[1]e.paknejad@cfu.ac.ir

Abstract: *This paper aims to study the effects of guessing on the total score, ability, questionnaire items (coefficient of difficulty, coefficient of identification), and the reliability coefficient in multiple-choice tests on a sample of 200 students (holding Bachelor's degree) in Farhangian University, Khuzetan. A 24-item test on the students' educational progress in statistics was used as an empirical study with a sample randomly selected. Data were analyzed by Bilog-mg and SPSS. Results showed that there was a significant difference ($p<0.01$) between the total score of the group forbidden to guess the answers of questions (experimental group) and the group allowed to guess the answers. However, no significant difference was observed between the coefficients of difficulty and the reliability of the experimental (not allowed to guess) and control groups (allowed to guess).*

Keywords: *Supposition, Ability, Questions, Coefficient of Difficulty, Coefficient of Reliability, Multiple-Choice Test.*

1. INTRODUCTION

To find human features such as thought and mental attributes and sentimental features, accurate measurement tools are required. Tests are, thus, an important tool used to study people's mental and educational features. Among written tests, substantial tests, especially multiple-choice tests, are used more because of easy scoring system, accuracy, concreteness, covering more content than other tests (Seyf, 2011).

Based on multiple-choice test results, major decisions such as attracting students from university, selecting talented students of education centers for smarts, and employing employees and workers for organizations, are made. Multiple choice tests is popular among written tests and major decisions are made based on them, experts do not though agree on its application as a reliable measurement tool.

Experts have different opinions because multiple-choice tests cannot asses educational progress, force learners to memorize unimportant contents, do not help understanding the relationship between educational contents and guessing. Psychometric consultors have tried for years to use standardized tests to measure the students' educational ability and progress in general and especial areas (William & Amini, 201.2

Multiple-choice tests have been always challenged for subjects usually guess the answers. There are generally two models of guessing. In the first model, called blind guess and wild guess, subjects have no idea about options and randomly select a choice. In the second model, called informed guessing and educated guessing, subjects are believed to use a logical reason or a relative knowledge to guess the correct answer by rejecting other choices.

Guessing is a major factor in multiple-choice tests. Guessing is to try to find answers or judge about a thing regardless of being certain about its trueness (Hornby, 2001).

Between 2 to 5 options are, thus, considered for multiple-choice tests, especially for the educational progress tests. Four-choice tests are though more popular, in which on option is correct and the others are digressive (2011).

In multiple-choice tests, not knowing the answer, subjects can find the correct answer just by guessing. However, in personality or attitude test in which there is not a correct answer, guessing makes no problem. In performance or educational progress tests, this makes some problems because

subjects, in case of not being aware of the subject, may find the correct answer by a probability of 25 percent (Allen and Yen, 1995).

Examining a strong group's answers, questions which have been guessed by subjects can be identified. When the number of choices relating to a question is equal in a strong group, this is an indicator of guessing (Delavar & Zahrakar, 2013). Guessing is very important in assessing the multiple questions. Guessed answers increase the variance error of test scores and lower the reliability (Yoella, Joachim, &Oded, 2002).

Guessing first raise the measurement error and the probable answers. Italso results in a structural error variance and it is considered as a major threat for the construct validity. Secondly, it causes errors and weakens the correlation between answers (Rogers, 1999, quoted by Messiki, 1995) (Chien 2007).

Two methods are used to eliminate the effect of guessing in multiple-choice question. First, increasing the number of options to 5, 6, 7, and more so that people lack of the required knowledge cannot guess the answers. However, it is difficult to keep the attraction of test and avoid the subjects' deviation, in educational progress tests in particular, by considering more choices. Second, using formulaic scores (i.e. subtracting a percent $S= R- \frac{W}{N-1}$ of correct answers because of incorrect answers using , where s is the corrected answers or corrected scores, R is the number of questions subject have answered correctly, W is the number of questions with wrong answers, and N is the number of the choices of questions. According to question-answer theory, guessing parameter is different for any question. In the classic theory, guessing for all questions regarding the number of choices is equal. It is virtually supposed that all choices are equal in attraction (Ayala, 2009).

A study conducted in Nigeria aiming to study the effect of guessing on the scores of a sample of 20 students. Results were consistent with Lord's insights (1977). Lord observed that the classic theory of testing cannot provide a good framework for the reliability tests (William &Amini, 2012).

Richard F. Burton reviewed three empirical research. Results showed that ordinary scoring had stronger effects on test reliability than negative scoring (Morteza Hejri, Khabaz MafiNejad, Jalili, 2014).

In a study, one group were allowed to guess answers, but the other group were not allowed. In the first group, negative scores raised the test reliability. However, no change was observed in the other group by using negative scores (Morteza Hejri, Khabaz MafiNejad, Jalili, 2014).

In another study on students from the Gazvin University, negative scores lowered students' scores and lead them to be rejected in the interested subject. Also, a significant correlation was also observed between students' score before and after applying negative scores. The analysis of students' performance also disclosed that there were other factors than guessing affecting the selection of choices (Gholami, H. Mojdehi Panah, & t. Derakhshan, 2013).

Michael Kane and James Moloney (1978) conducted a study to examine the effect of question features and zero – one scoring and the answer – until – correct methods (AUC). Results revealed that scoring rules requires different question features to minimize the effect of guessing on the test reliability (Kane & Moloney, 1978).

An important question is that if informing subjects of the negative scores for wrong answers (i.e. unless you do not guess if do not know the answer, you will lose points) or informing subjects of the inexistence of negative scores for wrong answers have any effect on the subjects' ability, the total score, and other parameters such as the coefficient of difficulty or the coefficient of identification? This paper aims to study the effect of not guessing on the ability and the parameters of multiple-choice tests. Therefore, the following hypothesis will be examined as follows.

1.1. Research Hypotheses

- There is a significant difference between the educational performance of students allowed to guess answers and those were not allowed.
- There is a significant difference between the ability of students allowed to guess answers and those were not allowed.
- There is a significant difference between the difficulty of questions answered by students allowed to guess answers and the questions answered by those were not allowed.

- There is a significant difference between the coefficient of identification of questions answered by students allowed to guess answers and the questions answered by those were not allowed.
- There is a significant difference between the coefficient of reliability of questions answered by students allowed to guess answers and the questions answered by those were not allowed.

1.2. Research Methodology

This is an empirical practical study with a sample randomly selected for the control and experimental group. Here, a two-group plan was used with an experimental intervention and a post-test.

1.3. Research Population

The statistical population included all male educational sciences students-teachers in Farhangical University, Northern Khuzestan, in the academic year 2013-2014. They were aged between 19 and 23 employed by the Education System at the very beginning of the entrance to the University.

1.4. Research Sample

The research sample included 200 students randomly clustered from Farhangian Centers and divided into two groups of experimental and control.

1.5. Research Tool

The research tool is a teacher-made 24-multiple-choice question test on statistic regarding the major headings of statistic at the end of a semester. The test was developed based on a two-dimensional table. As it was designed based on the content table and the education purposes of an education term and considering the comments of a number of professors teaching in high education centers for statistics, the test was found to be sufficiently valid. The research reliability was calculated by bilog-mg at 0.7. Both experimental and control groups were provided by same questions. However, a guideline on questions and the way of answering and the negative scores had been considered for the experimental group not allowing them to guess the answers. The structure was also explained orally at test session. Negative scores were not considered though for the control group and they were allowed to guess the answers.

1.6. Test Findings

Data were descriptively and inferentially analyzed by SPSS AND Bilog-mg.

Table1. *Descriptive statistic indexes relating to students' education performance*

Groups	Statistical indexes	Numbers	Mean	Standard Deviation
Experimental group (without guessing)		101	10.64	3.36
Control group (with guessing)		99	12.43	3.41

Regarding the above table, the control group mean (or the group allowed to guess answers) (12.43) was greater than the experimental group mean (10.64). The standard deviation were equal though.

Table2. *Descriptive statistic indexes relating to students' ability*

Groups	Statistical indexes	Numbers	Mean	Standard Deviation
Experimental group (without guessing)		101	0.0031	0.77
Control group (with guessing)		99	0.032	0.77

Regarding the above table, the experimental group ability (0.0031) was greater than the control group mean (0.032). The standard deviation were equal though.

Table3. *Descriptive statistic indexes relating to the difficulty of test questions*

Groups	Statistical indexes	Numbers	Mean	Standard Deviation
Experimental group (without guessing)		23	0.22	1.32
Control group (with guessing)		23	0.30	1.43

As both groups answered a test, results by Bilog-mg showed that a question eliminated and the number of questions were reduced from 24 to 23.

Regarding the above table, the difficulty of questions for the experimental group (0.22) was greater than questions considered for the control group (0.30). The standard deviation of the experimental group (1.32) was smaller than the control group (1.43).

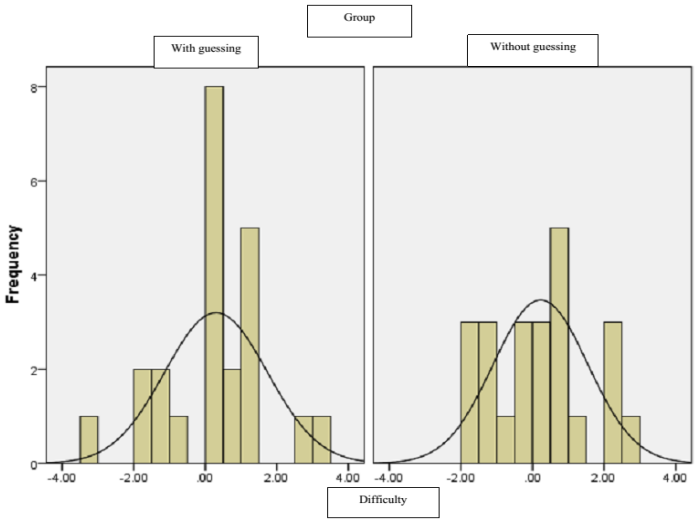

Fig1. *Coefficients of difficulty for both groups*

Diagram 3 shows that the distribution of the scores of difficulty coefficients are normal and equal for both groups.

Table4. *Descriptive statistic indexes relating to the coefficients of identification*

Groups	Statistical indexes	Numbers	Mean	Standard Deviation
Experimental group (without guessing)		23	0.50	1.42
Control group (with guessing)		23	0.51	1.56

Regarding the above table, the mean and standard deviation of both experimental (not allowed to guess) and control groups (allowed to guess) were relatively equal.

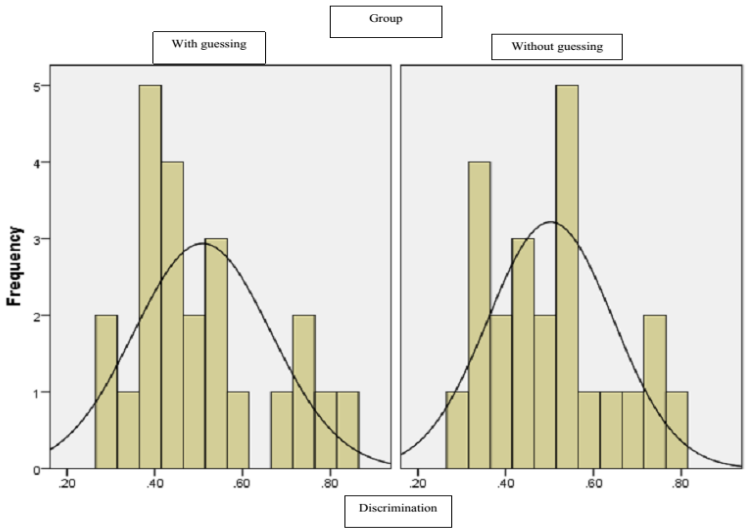

Fig2. *Coefficients of Dicrimination for both groups*

Diagram 4 shows that the distribution of the scores of Dicrimination coefficients are normal and equal for both groups.

2. INFERENTIAL FINDINGS BASED ON RESEARCH HYPOTHESES

First Hypothesis: There is a significant difference between the educational performance of students allowed to guess answers and those were not allowed.

Table5. *Statistical findings relating to the first hypothesis*

Groups	Statistical indexes	Numbers	Mean	Standard Deviation	Degree of Freedom	Observed T	Sig.
Experimental group (without guessing)		101	10.64	0.33	198	3.74	P*<0. 01
Control group (with guessing)		99	12.43	0.34			

Regarding the above table, t is 3.74 which is significant at p<0.01. It can be, thus, concluded that there is a significant difference between the education performance (test total score) of the experimental group (not allowed to guess) and the control group (allowed to guess). As the control group's education performance is higher, the first hypothesis is confirmed.

Second Hypothesis: There is a significant difference between the ability of students allowed to guess answers and those were not allowed.

Table6. *Statistical findings relating to the second hypothesis*

Groups	Statistical indexes	Numbers	Mean	Standard Deviation	Degree of Freedom	Observed T	Sig.
Experimental group (without guessing)		101	0.0031	0.08	198	-0.26	P<0.05
Control group (with guessing)		99	0.032	0.08			

Regarding the above table, t is -0.26 which is not significant at p<0.05. It can be, thus, concluded that there is not a significant difference between the ability of the experimental group (not allowed to guess) and the control group (allowed to guess). The second hypothesis is, therefore, rejected.

Third Hypothesis: There is a significant difference between the difficulty of questions answered by students allowed to guess answers and the questions answered by those were not allowed.

Table7. *Statistical findings relating to the third hypothesis*

Groups	Statistical indexes	Numbers	Mean	Standard Deviation	Degree of Freedom	Observed T	Sig.
Experimental group (without guessing)		23	0.22	0.28	44	0.19	P<0.05
Control group (with guessing)		23	0.30	0.30			

Regarding the above table, t is 0.19 which is not significant at p<0.05. It can be, thus, concluded that there is not a significant difference between the difficulty of tests for the experimental group (not allowed to guess) and for the control group (allowed to guess). The third hypothesis is, therefore, rejected.

Forth Hypothesis: There is a significant difference between the coefficient of identification of questions answered by students allowed to guess answers and the questions answered by those were not allowed.

Table8. *Statistical findings relating to the forth hypothesis*

Groups	Statistical indexes	Numbers	Mean	Standard Deviation	Degree of Freedom	Observed T	Sig.
Experimental group (without guessing)		23	0.50	0.03	44	0.15	P<0.05
Control group (with guessing)		23	0.51	0.03			

Regarding the above table, t is 0.15 which is not significant at p<0.05. It can be, thus, concluded that there is not a significant difference between the coefficient of identification of questions answered bythe experimental group (not allowed to guess) and by the control group (allowed to guess). The fourth hypothesis is, therefore, rejected.

Fifth Hypothesis: There is a significant difference between the coefficient of reliability of questions answered by students allowed to guess answers and the questions answered by those were not allowed.

Table9. *Statistical findings relating to the fifth hypothesis*

Groups	Statistical indexes	Reliability	Zr	Standard Deviation	Observed z	Sig.
Experimental group (without guessing)		0.7	0.867	44	0	P<0.05
Control group (with guessing)		0.7	0.867			

Regarding the above table, as the coefficients of reliability of the questions provided for the experimental group (not allowed to guess) and the control group (allowed to guess) are equal (0.7), their Fisher's zs are also equal (0.867). The significance of the observed z is zero. It is, thus, concluded that there is no significant difference between the test reliability of both groups. The fifth hypothesis is, therefore, rejected.

3. RESULTS

This paper aims to study the effects of guessing on the total score (educational performance), ability, coefficient of difficulty and the coefficient of identification, and the coefficient of reliability in multiple-choice tests. Results showed that:

- There is a significant difference between the total score and the educational performance of the students not allowed to guess the answers (experimental group) and those not allowed (control group) at the probability level of less than 0.01. The observed-t is 3.74 which is significant at $P<0.01$.
- There is no significant difference between the ability of students not allowed to guess the answers (experimental group) and those not allowed (control group), because the observed-t is -0.26 which is not significant at $P<0.05$.
- There is no significant difference between the difficulty of questions provided for students not allowed to guess the answers (experimental group) and for those not allowed (control group), because the observed-t is 0.19 which is not significant at $P<0.05$.
- There is no significant difference between the identification parameter of questions provided for students not allowed to guess the answers (experimental group) and for those not allowed (control group), because the observed-t is 0.15 which is not significant at $P<0.05$.
- There is no significant difference between the reliability of questions provided for students not allowed to guess the answers (experimental group) and for those not allowed (control group), because the observed-z is 0 which is not significant at $P<0.05$.

4. DISCUSSION

This paper aims to study if guessing or not guessing answers have any effect on reliability, the coefficient of difficulty and the subjects' ability. As this subject has not been widely studied, it is not possible to accurately comment on this research results and other studies. Findings also showed that using or not using guessing or applying formulaic scores only affected the total score and not highly affect ability and the question parameters. This is in consistence with some findings reported by Gholami et al. (A. Gholami, H. Mojdehipanah, M. Drakhshan, 2013). Our results, however, do not agree with studies conducted by Richard F. Burton (2002) (Morteza Hejri & Khabaz Mafinejad, 2014). This may be because of using different samples, the number of test questions, and different analysis methods.

Suggestion: to find more accurate information, a study is recommended to be conducted on university entrance tests, both for Bachelor's and Master's degrees, with a larger samples and more questions to see how much guessing affect and to see that if it does not significantly affect ability parameters, why should be applied in such tests.

REFERENCES

Ayala, R., J, de. (2009). the theory and practice of item response theory . London: the Guilford ress New York London.

Chien, C. Y. (2007). The Effect of Guessing on Assessing Dimensionality in Multiple -Choice Tests: A Monte Carlo Study with Application. (PhD), Pittsburgh.

Hornby, A. S. (2001). Oxford advanced learner's dictionary for current English London . Oxford university press.

Kane, M., & Moloney, J. (1978). The Effect of Guessing on Item Reliability under Answer -Until-Correct Scoring. Applied psychology measurement, 2, 41-49.

William, J., ubulom, & Amini, c. M. (2012). The effect of guessing on the test scores . Mathematical theory and modeling, 2, 6.

Yoella, Bereby, meyer, Joachim, m., & Oded, M., flasher. (2002). prospect theory analysis of guessing in multiple choice test. Journal of behavioral decision marking, 5, 14.

Alen, A. J., Yen, A., Vandi, (1995), an introduction into the measurement theories (A. Delavar, Trans), Tehran, Samt.

Delavar, A., Zahrakar, K., (2013), assessing and measuring in psychology and educational sciences, Tehran, Arasbaran.

Seyf, A., A., (2011), measuring, assessing and appraising education, Tehran, Doran.

Gholami, A., Mojdehipanah, H., and Derakhshan, 2013, the effect of negative score on test results: reporting a case, Journal of Medical Education and Development, 1, 49-53.

Mortezaei, S, Khabaz Mafinejad, M., (2014), guessing multiple-choice questions: challenges and approaches, Journal of Training in Medical Sciences, 7, 594-604.