The Impact of Correction for Guessing Formula on
MC and Yes/No Vocabulary Tests' Scores

Abdollah Baradaran
Islamic Azad University Tehran Central Branch
Sdeideh Ahanghari[*]
Shokouh Rashvand Semiari
Islamic Azad University, Tabriz Branch

A standard correction for random guessing (cfg) formula on multiple-choice and Yes/No examinations was examined retrospectively in the scores of the intermediate female EFL learners in an English language school. The correction was a weighting formula for points awarded for correct answers, incorrect answers, and unanswered questions so that the expected value of the increase in test score due to guessing was zero. The researcher compared uncorrected and corrected scores on examinations using multiple-choice and Yes/No formats. These short-answer formats eliminated or at least greatly reduced the potential for guessing the correct answer. The expectation for students to improve their grade by guessing on multiple-choice and Yes/No format examinations is well known. The researcher examined a method for correcting for random guessing (cfg) " no knowledge" on multiple- choice and Yes/No vocabulary examinations by comparing application and non-application of correction for guessing (cfg) formula on scores on these examinations. It was done to determine whether the test takers really knew the correct answer, or they had resorted to a kind of guessing. This study represented a unique opportunity to compare scores from multiple-choice and Yes/No examinations in a setting in which students were given the same number of questions in each of the two format types

[*] Corresponding author. E-mail: s_ahangari@yahoo.com

testing their knowledge over the same subject matter. The results of this study indicated that the significant differences were highlighted between the subjects' scores when cfg formula was applied and when it was not.

Key*words*: Correction for Guessing (CFG), Yes/No Questions, Multiple-Choice Questions

As far as multiple-choice tests are concerned, there has always been concern over the fact that *guessing* has an impact on the scores of these tests. Some educators therefore tended to discourage students from all guessing considering it dishonest. "However it became indispensable parts of mass testing and found to have some virtues like broader coverage of instructional topics, accuracy of scoring, and provision of statistical feedback at the item level"(Frary, 1982, p. 338). Thus, neither admonishment against guessing, nor avoidance of such tests was a satisfactory approach to resolve the matter. Frary (1982, p. 339) suggests the use of a "scoring formula" which corrects for purely random guessing. The conventional correction formula subtracts a fraction of the wrong answers from the number of right scores. Cross, & Frary (1977, p. 319) outline the reasons why either method of correction for guessing is likely to be undesirable in a typical academic setting:

- Very few examinees will be so ignorant or so slow that they will fail to attempt or be completely unable to eliminate a single wrong choice on any substantial proportion of questions. The effort of correction for guessing is therefore wasted to a large extent. The few who legitimately should omit substantial proportions of questions under formula scoring will be so low in achievement that very low scores will result regardless of whether totally random guessing is suppressed.
- The admonishment not to guess in the absence of information may be interpreted differently by each examinee and thus may introduce score variance associated with personality or background factors. This phenomenon has been confirmed in numerous published

studies. Other published studies have shown that when students do omit questions under conventional correction for guessing instructions, they are, on average, able to choose significantly more correct answers to these questions than under chance expectation.

- Individuals may choose to disregard the instructions because, on average, correction for guessing does not penalize for random guessing but only removes the score gain expected from completely random guessing. In fact, if a student's knowledge is inadequate for obtaining a needed score, the best strategy for that student is to guess on all questions, hoping that luck in the short term will be favorable. Because this action is at odds with the instructions not to guess randomly among all choices, the instructor is placed in the questionable position of giving directions, which some students may ignore to their benefits.

The correction for guessing formula (cfg) applied to the raw scores of the multiple choice (MC) and Yes/No items is widely used in the field of language testing: Yes/No format, since, as Jones (2006, p.14) states, "it eliminates or at least considerably reduces the opportunity for guessing the correct answer", and MC format because the possibility for students to improve their grades by guessing on such format is well known. The aim of this correction is to take into board the fact that subjects have a good chance to obtain the correct response by guessing, in which case the accounted credit fails to reflect their real knowledge. The final score, therefore, results in an overestimation of what is intended to be measured. The main motive behind this formula covers the following points:

● There have been so many suggestions that guessing is a component of error variance.

● What is guessing and its close cousins, response bias, in Yes/No and in multiple-choice?

● The fundamental issue in differential weighting is to increase reliable variance and reduce the effect of guessing.

The theoretical model behind the transformation from raw scores (numbers of correct response) into corrected scores (numbers of items really known by the participants) rests on some assumptions proposed by Frary (1988) as follows: "When a participant comes to a specific recognition item (like MC and Yes/No), either he has the knowledge to answer it correctly or he does not. There is nothing in between. It is either yes or no" (p. 36). This is therefore called an "all-or-nothing" or discrete model." As also stated by Cureton (1966), "If a participant has the knowledge, he will get the correct answer, but if he does not have the knowledge, he will guess. Each incorrect response is the result of a random guess among all options given" (p. 4)

It has been noted by a variety of writers that these assumptions are generally invalid and that the correction may be an overcorrection, e.g. Malcolm (1968, p.16). The first assumption, for example, cannot be possibly true, since even on items that a participant is not completely sure of, he probably knows something. In that case, "if he carefully examines the item, he can perhaps eliminate some of the incorrect response options and thus increase the likelihood of getting the item correct" (Dennis, 2000, p. 2).

"If a participant gets to an item and doesn't make any response, the correction for guessing formula is impotent to act" (Rowley & Traub, 1977, p. 18). While it could be arbitrarily called an error, since he didn't respond, yet there's no basis that he didn't know it. After all, it's possible that he simply skipped the item on the first encounter, but forgot to go back to it. Omitted items, therefore, cannot be assumed to be wrong and cannot be subtracted from the correct since the omissions are being considered wrongs. Thus, omissions play no role in using the correction for guessing formula. If one doesn't respond, he can't get any credit for it, either (Talento-Miller 2009, p. 6).

 (Frary, 1982, p. 348) defines "the scoring formula" or "correction for guessing" as:

FS= R-W/(C-1, in which,

FS= corrected or Formula score

R= number of items answered right

W= number of items answered wrong

C= number of choices per item

(Mousavi, 1999, p.69) quotes, "a correction for guessing is usually applied where test takers do not have sufficient time to complete all items on the test and where they have been instructed that there will be a penalty for guessing." He puts forward the formula as below:

Score= Right- wrong/ n–1

n refers to the number of alternatives for an item. The formula, therefore, might apply to various selection type items as follows:

Yes/No items: S= R– w/2-1   or   S= R–W

Multiple-choice items:

Three alternatives S= R-W/2 // Four alternatives S= R-W/3 // and so forth.

It has been attempted in this paper to determine whether the correction for guessing formula (cfg), which might be used for multiple-choice and Yes/No items, can lead to different results; that is, different test scores once the formula is applied and once it is not applied. To fulfill such a purpose, the following question was raised:

Does the correction for guessing formula (cfg) have any significant impact on the results of different test formats such as MC and Yes/No tests?

Method

Participants

The subject pool of this study consisted of 90 students, from among them 60 functioning as the main subjects. They were all students of an English language school. The ultimate subjects were chosen on the basis of their language proficiency scores. That is, those who scored one SD above and below the mean. The rationale behind selecting the aforementioned students was to do study with more proficient students. The students' age range was 16~20. They were all female EFL learners at the intermediate level. Sex was therefore the controlled variable held constant to neutralize the

probable effect on the outcome. The effect of correction for guessing was investigated after the completion of the tests and awarding the grades.

Instrumentation

An English proficiency test: It was used to evaluate the learners' proficiency in English language to ensure their homogeneity. The test comprised 80 items to be answered in 90 minutes and it was divided into three sections; the first section was the structure and written expression consisting of two separate parts. The first part was identifying the correct options, including 20 items and the second part was the written expression including incorrect words or phrases consisting of 10 items. The subjects were asked to choose the synonyms for the underlined words or phrases. The third section was reading comprehension. This part consisted of some reading passages each followed by some MC reading comprehension questions. Here the students had to read the passages carefully and select the correct answers to the questions, this section consisted of 20 items.

The vocabulary test: This test was used to evaluate the learners' vocabulary knowledge in English. It consisted of two separate test formats. The first part consisted of MC items. The subjects were asked to identify the correct options, including 40 items. The second part consisted of 40 Yes/No vocabulary items in which the students were asked to indicate the words the meanings of which are known to them by saying yes/no.

Procedure

To test the proposed hypothesis and to indicate the relationship between the mean scores of students, that is, in the presence/absence of the cfg formula, the following procedures were met to produce acceptable results:

- Phase 1, the homogeneity procedure: English proficiency test was administered to the 90 subjects. Test results were then calculated and recorded in a table. After calculating IF and ID indexes, removing bad items and rescoring the

proficiency test, 60 students were chosen as the subject of the study.

- Phase 2, the main study: There were 60 subjects taking part in the vocabulary test including MC and Yes/No items. Each test format included 40 items. The students' papers were once corrected with the application of the cfg formula and once again without its application. Then the means, standard deviations, and variances in either case – the presence/absence of the cfg formula- were calculated. A t-test was also performed to answer the question about the statistical significance of differences between the means.

Design

In this study, attempts were made to equate the subjects as much as possible by random assignment. The design of study was post-test only, group design (Quasi-experimental). The students were exposed to two different test formats. They were informed of penalty beforehand in order to obtain some omitted responses. The gained calculated means were compared by performing a t-test to see if the difference between the mean scores was significant.

Data Collection and Data Analysis

The purpose of this study was to determine the impact of the cfg formula on MC and Yes/No test scores. The process of data collection started with administering a proficiency test identified as a baseline for ensuring subjects' homogeneity. Their papers were thereafter scored once with the application of the cfg formula and once again without it. To collect the relevant data, the following statistical computations were made:
- IF and ID indexes of the proficiency test.
- The reliability of the proficiency test through K-R21 formula.
- The means and standard deviations of the scores on the proficiency test.
- A t-test to guarantee the homogeneity of two groups.
- A t-test to check the difference between the means.

Results and Discussion

This study was concerned with one main research question: Does the correction for guessing formula (cfg) have any significant impact on the results of MC and Yes/No vocabulary tests?

Concerning the above research question, the following null hypothesis was proposed: the correction for guessing formula (cfg) has no impact on MC and Yes/No vocabulary test scores. To collect the relevant data, the following steps were met by the researcher:

To minimize the individual differences among the subjects and to ensure their homogeneity, a proficiency test was developed. It was administered to 90 female EFL learners at an English language school. Thereafter, IF (total, upper, and lower) and ID indexes of the proficiency test items were calculated. The next step was to discard bad items and rescore the test. In fact, those items with $25 \leq IF \leq 75$ and $0.34 \leq ID \leq +1$ were retained and the rest were simply removed from the test (that is, items number 2,3,4,7,8,9,10,11,14,15,17,20,25,26,27,28,30,34,38,41, and 59). Thereafter, the scores of each part of the proficiency test were calculated separately.

Then the means, SDs, and reliabilities of each of the aforementioned parts calculated. The results are shown as follows:

Table 1
Descriptive Statistics and Reliability of Grammar, Vocabulary, and Reading Section

KR-20	SD	Mean	Sum	N	Grammar score
0.94	4.24	14.68	1322	90	

KR-20	SD	Mean	Sum	N	Vocabulary score
0.84	3.39	9.08	818	90	

KR-20	SD	Mean	Sum	N	Reading Score
0.65	2.59	4.83	435	90	

After ensuring the reliability of the proficiency test, the researcher chose 60 subjects. It's worth mentioning that among all the testees, only those whose scores were within one standard deviation above and below the mean were selected as the subjects of this study. Table 2 and 3 show the results:

Table 2
Sort Proficiency Descriptive Statistics

Mean±1 SD	SD	Mean	Sum	N	Sort
43.47 20.69	11.39	32.08	2888	90	proficiency

Table 3
 Ability Groups

N	The group to which the cfg formula applied	N	The same group to which the cfg formula was not applied
60	Cfg	60	#

The next step was to administer the newly developed MC and Yes/No vocabulary tests.

The null hypothesis for the main research question was that there is no significant difference between MC and Yes/No vocabulary test scores when the correction for guessing formula is applied and when it is not applied. To test this hypothesis, the MC and Yes/No test scores of the groups under study were once computed with the application of the cfg formula and once again without it. Table 4 demonstrates the means, and standard deviations of the pairs.

Table 4
Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Mc/cfg not used	29.7167	60	7.06169	.91166
	Mc/cfg used	27.6667	60	6.85607	.88511
Pair 2	yes,no/cfg not used	30.5667	60	4.88292	.63038
	yes,no/cfg used	28.2000	60	5.51669	.71220

Then, in order to see whether the difference between the means is statistically significant or not, a t-test was conducted between the pairs. It is represented as follows:

Table 5
Paired Sample T-Test

		Mean	*SD*	*t*	*df*	*Sig.(two tailed)*
Pair 1	*Mc/cfg not used - Mc/cfg used*	*2.0500*	*1.06445*	*14.918*	*59*	*.000*
Pair 2	*yes,no/cfg not used - yes,no/cfg used*	*2.3667*	*1.02456*	*17.893*	*59*	*.000*

As shown in the table, for multiple choice questions (t $_{(59)}$= 14091,p=.000) and for the yes/no questions it is (t $_{(59)}$ = 17.89 , p= .000). So in both cases of MC and Yes/No tests, since P value is less than .05, one may conclude that the difference between the means was significant at the 0.05 level. The null hypothesis was therefore rejected.

Discussion

There exists a significant difference between the scores of the subjects under study in the presence/absence of the cfg formula. In fact, there is a significant difference between the mean scores of MC and Yes/No test scores when the cfg was applied compared with that when it was not applied. One can thus conclude that when correction for guessing formula (cfg) is applied, the MC and Yes/No test scores of female Iranian learners decrease.

Diamond, & Evans (1973: 184), reported on the need for specific instructions to be given to students about guessing to allow examinations with correction for guessing to retain reliability. Students must be informed that a correction for guessing will be applied and must be shown the effect of guessing without knowledge or even with partial knowledge (the ability to

eliminate one or more incorrect answers) as well as the potential benefits of partial knowledge.

Lord (1975: 9) argues that formula scoring will always improve reliability provided the student leaves at least one question unanswered. Intuitively, this can be understood as removing some random guessing component from the score and, thus, focusing on the student's actual knowledge. This advantage of formula scoring has been empirically supported and discussed in several recent studies. Thus the use of formula scoring (corrected multiple-choice examinations) not only results in increased validity but also saves faculty time that would have to be spent grading Yes/No examinations. Increasing the number of questions that are included in the multiple-choice examinations would potentially result in greater reliability. This addition would not increase faculty time spent in grading but would require additional time in test preparation.

The findings of this research have been both supported and rejected by several researchers. There are many situations in which the cfg doesn't fulfill its intended function (Kasten, 1982: 839). For example, Lord (1975: 11) claimed that the correction for guessing formula (cfg) can only be used when test instructions clearly state that the examinees are to omit items only when they feel that they would have to guess randomly. Diamond and Evans (1973: 187) stated the following points:

- The model is mathematically weak in that candidates do not guess when they have no information (they have in fact partial knowledge). In fact, as also confirmed by Budescu (2008: 21), the underlying cognitive model ignores the case of partial information.
- The correction for guessing formula (cfg) has the most impact on low scoring candidates who are least likely to understand the instructions to omit if they do not know.
- The formula score is correlated with risk taking. To prove this claim, we can include items with no correct answer to measure the propensity for risk taking.

Students who are lower in risk taking are penalized more by formula scoring than those who are prone to take risks on objective

tests. As also stated by Miller (2009:11), low-ability examinees usually omit the responses while high-ability ones resort to guessing in such cases. Moreover, one of the greatest problems of correction for guessing formula (cfg) is the influence of omitted items on the corrected scores. For instance, assume a condition in which two examinees get the same score (eg, 38/50). For the remaining 12 which neither really knows, one omits 4 and the other does not. For one we have 38 correct and 12 wrong. For the first, the corrected score would be 38-12/4=35 (number of options would be 5 in this example). But for the second, we have 8 wrong answers, the corrected score would be 38-8/4=36. In this case, there is a differential correction made and the examinees are put in different final score positions…one 35 and the other 36. So, it would stand to reason that the instructor would treat them differently; that is, perhaps assign different grades based on their different corrected scores. Besides, if an examinee doesn't answer an item, it does not mean that he did not know the answer. He may simply skip over that item. He may provide the right answer if s/he comes back to that item.

Budescu (2008:23) also proposed the following demerits:

- People are notoriously mis-calibrated with regard to their own level of knowledge.
- The various options are rarely (if ever) equally attractive.

Paul (2007: 21) believes that although "Elimination scoring", where students are asked to eliminate the incorrect responses and "Inclusion scoring", in which students are asked to choose the smallest subset of answers that includes the right answer are found to be more reliable than correction for guessing formula, they tend to add confusion for test takers and produce inconsistent results. However, among excuses for using correction for guessing formula (cfg), the followings are worth heeding:

- To get one's score closer to the truth.
- To discourage random guessing.

Yet, no-one could support the hypothesis that testees' accounted credit, while cfg does not exist, may be interpreted as their full knowledge. Choppin (1988, p. 384) points out that correction for guessing addresses three concerns: "(1) guessing

introduces a random factor into test scores that adversely lowers reliability and validity, (2) expected correct guesses inflate estimation of students' abilities, (3) the inflation from guessing can be an unfair advantage for students who guess frequently when compared to students with equal ability who do not guess. Applying the correction for guessing reduces the advantage for students who guess frequently"

Mousavi (1999, p.70) states the problems with the cfg formula as below:

1) It assumes that all the wrong answers are due to guessing, but this is not the necessarily so. Subjects may be misinformed. The guessing formula treats such subjects harshly.

2) It assumes that where guessing has occurred, there is an equal chance for each option to be chosen. However, this is not so, since in some items, subjects may have eliminated some of the distracters but still guess the answer. For these subjects, the correction formula is an underestimate.

3) This guessing correction applies to subjects on average. In any individual case, it may well be wrong.

Conclusions

As already stated by Harvey (2004: 3), all models of detection and discrimination have at least two psychological components or processes: the sensory process (which transforms physical stimulation into internal sensations) and a decision process (which decides on responses based on the output of the sensory process). Good knowledge of vocabulary has just received the attention it always needed. It is important for anyone who wants to use the language. Kitao & Kitao (1998: 6) stated that the vocabulary knowledge can be divided into four types: active speaking vocabulary, passive listening vocabulary, passive reading vocabulary, and active writing vocabulary. It's the test constructor's task to assess the relative importance of these skills at the various levels and to devise as accurate means of measuring the students' knowledge of the meaning of certain words, as well as patterns and collocations in which they occur. Such a test may assess their active/productive or passive/receptive vocabularies. The main

research question is concerned with the existence of a significant difference between the means of MC and Yes/No test scores when the cfg is applied and when this formula is not applied.

After the administration of the main vocabulary test, including Yes/No and MC (the included vocabularies were all selected from the subjects' study book- Headway-Intermediate.) and the correction of answer sheets, the data was gathered and analyzed. It is emphasized that subjects were homogenized through taking an already standardized proficiency examination before the main vocabulary test of study. Thus it could be concluded that the participating subjects were significantly different in their scores when the correction for guessing formula was adopted. Obviously, the proposed null hypothesis was rejected. It should be reminded that the researcher had no access to male participants and just had a limited access to female participants. Regarding the fact that the bigger the sample, the better it reflects the population characteristics, a future investigation might be conducted with gender as its dependent variable and also with a greater population to further explore the reality of the issue.

Implications

The findings of this study may advance our realization of the vital role of the correction formula- cfg in this study- in tests scores and their relevant interpretations. Theoretically, they contribute to the development of modified vocabulary test formats and different correction procedures. Practically, they directly influence the EFL teachers' judgment of learners. They even influence the learners' emotional and rational viewpoints of themselves as well as their foreign language vocabulary knowledge (Bachman & Palmer, 1996: 18).

Theoretical Implications

Theoretically speaking, the findings of this research are consistent with following assumptions:

- When the participant comes to specific recognition items (like MC and Yes/No), he either has the knowledge to

answer it correctly or he doesn't. There is nothing in between. It is either yes or no (this is therefore called an all-or-nothing or discrete model).

- If the participant has the knowledge, he will get the correct answer, but if s/he doesn't have the knowledge, he will guess.

- Each incorrect response was the result of a random guess among all options given. It is however difficult to draw this conclusion because the matter is more complex than it may first appear. As one of the most important theoretical implications of this study, it appears that the phrase "correction for guessing" should be discussed in more detail. In other words, unless a reference correction is identified, the phrase is meaningless. The question to be raised here is: what is the correction? The reference correction is essential, since its existence assures true knowledge of testees. Unfortunately, such a reference "correction" has not been well developed yet. Correction is the procedure used to show the results of language tests. Those results are often reported as numbers or scores. Since test scores are commonly used to assist the researcher/teacher in making decisions about individuals, the correction formulae (the criteria by which test takers' responses are evaluated and the procedures followed to arrive at a score) we are essentially determining how to quantify test takers' responses. Correction is thus the essential step to arrive at a measure in addition to any qualitative and descriptive information obtained from the test takers' responses. Deciding what type of correction formulae to use is therefore very important (Kasten 1982: 840). In some cases, considerations of correction may influence the specific tasks or intended responses included in the test. If, for example, we had limited responses or writing test tasks but still needed a test that could be corrected/scored/ fairly quickly, we might decide to avoid multiple-choice items. The reason is that these items typically require considerable resources to write. In such

cases, limited production tasks should be included. For instance, we can include completion or short answer items. These items that can be written with fewer resources can still be scored quickly and efficiently using a scoring key. Also maximum benefit will only be accomplished through an awareness of students' true ability and knowledge. In other words, teachers' awareness of the students' true knowledge might save class time. Yet, other teachers who lack the awareness might either use an exaggerated long or short time for the tests they employ. The results of such a case might be students' fatigue, anxiety and even hatred and fear toward language testing. Utilizing appropriate time for each test in the classroom, is therefore of great importance. It enables the language teachers (1) to motivate students to be exposed to natural language testing situations, (2) to save time by giving sufficient time to learners to perform well on the tests.

Moreover, teachers can manipulate the exam duration from short to long, getting students informed of the formula which is going to be employed and warning them about the penalties.

Clearly, if we don't inform testees of the use of the cfg formula, it's the same as hiding the truth about how their test scores will be handled. Such practice would be regarded as unethical behavior. Normally, the instructions will inform examinees that they should not engage in complete random guessing on an item.

Therefore, as a theoretical impact of the present study, a suggestion of modification of the correction procedure and also test format is proposed so that authorities consider this issue more seriously. Furthermore, test designers and methodologists may use the findings of this study to consciously select the relevant correction formula according to the learners' proficiency levels. Students and EFL learners too can benefit from the findings of this study by determining their own weaknesses and their areas of difficulties. They will be also aware of their own true knowledge, not something they believe to know. As the general findings of this study suggest, the correction for guessing formula can influence teachers' judgment of their students' true ability/knowledge. It also

helps teachers determine the appropriate level of education for each testee. During the study, several research questions crossed the researcher's mind. Some of them are suggested hereunder wishing that a reader might find one of them interesting enough to pursue:

- Will a similar result be obtained with male participants being the only sex involved in the same study?
- To what degree will the results change if other correction formulae are utilized?
- If participants have prior knowledge of the vocabulary test, will the correction for guessing formula really work?
- Will the results change if the subjects are chosen from another proficiency level?
- To what extent will the result change if other variables such as age or alternation of test content are also taken on board?
- Will the results change if the number of participants and groups differ?

In the end, it is the researcher's heartful wish that the findings of this research would pave the way for enhancement of English language testing employing appropriate correction formulae and determining students' true knowledge of the subject being tested.

The Authors

Abdollah Baradaran is Vice-Chancellor and Research Deputy of Islamic Azad University, Central Tehran Branch. He has 22 Years of academic teaching experience and also heads the Graduate English Department of the same university. His major research interest is computer-assisted language learning.

Saeideh Ahangari is an assistant professor in TEFL at the Department of English Language, Tabriz Branch, Islamic Azad University. She obtained her PhD degree from Islamic Azad University/ Science and Research Branch. She has an M.A in Teaching English from the University of Tabriz. Her main interests are in the area of Task-based language Teaching, Language testing, Systemic functional linguistics and CALL and their interface with

the issues in English Language Teaching. She has published and presented papers in international conferences and journals.

Shokouh Rashvand is a PhD student of Islamic Azad University, Tabriz branch. Her main research interest is in the area of English language testing.

References

Bachman, L.F. & Palmer, A.S. (1996). *Language testing in practice: Designing and developing useful language tests.* London: Oxford University Press.

Budescu, D.V. (2008). A decision theoretical perspective on psychometrics: Analyzing Test-taking behavior. The internet. http://www.yahoo.com, 21-23.

Cross, L.H., & Frary, R.B. (1977). An empirical test of Lord's theoretical results regarding formula scoring of multiple-choice tests. Journal of Educational Measurements *14*, 313-321.

Cureton, E.E. (1966). The Correction For Guessing. The journal of Experimental Education. *34* (4). The Internet. http://www.google.com,4.

Dennis, R.(2000). Let's talk about the "correction for guessing" formula. The Internet.http://WWW.Yahoo.com, 1-5.

Diamond, J., & Evans, W. (1973). The correction for guessing. Review of Educational Research *43*, 181-191.

Frary, R. B. (1988). Formula scoring of multiple-choice tests (correction for guessing).No. 3 in the series; Instructional Topics in Educational Measurement, B. S. Plake, Editor. Educational Measurement: Issues and Practices, *7*(2), 33-38.

Frary, R.B. (1982). A simulation study of reliability and validity of multiple-choice test scores under six response-scoring modes. Journal of Educational Statistics *7*, 333-351.

Harvey, L.O. (2004). Detection sensitivity and response bias. The internet. http://www.yahoo.com, 3.

Jones, A.C. (2006). Correcting for Guessing Increases Validity in Multiple-Choice Examinations in an Oral and Maxillofacial Pathology Course. The internet. http://www.yahoo.com, 12-16.

Kasten, G. (1982). Correction for guessing. Evaluation Review.*6* (6), 837-841. The Internet. http:// www.yahoo.com

Kitao,K.&Kitao S.K.(1998). Test Design. Retrieved October 12, 2009: http://ilc2.doshisha.ac.jp/users/kkitao/library/article/test/design.htm, 4-6.

Lord, F.M. (1975). Formula scoring and number-right scoring. Journal of Educational Measurement *12*, 7-12.

Malcolm, J.S. (1968). The penalty for not guessing. Journal of Educational Measurement. (2). The Internet. http://www.google.com, 16-22.

Miller, E.T. (2009). Tactics and guessing. The internet. The official GMAT Blog.htm, 4-12.

Mousavi, S.A. (1999). *A Dictionary of Language Testing*. Tehran: Rahnama Publications, 68-70.

Paul, J. (2007). Improving educational assessment by incorporating confidence measurement, analysis of self-awareness, and performance evaluation: The computer-based alternative assessment (CBAA) project. The internet. http://www.yahoo.com, 16-22.

Prihoda, T.J. & Pincjard, N. & McMahan, C.A. & Jones, A.C. (2006). Correction for guessing increases validity in multiple-choice examinations in an oral and maxillofacial pathology course. Journal of Dental Education *70* (4), 378-386.

Rowley and Traub (1977). Formula scoring, number right scoring, and test-taking strategy. Journal of Educational Measurment *14* (1), 15 – 22.

Talento-Miller, E. (2009). Tactics & Guessing. The internet.http://www.yahoo.com, 4-12.

تاثیر فرمول cfg بر آزمونهای لغت چهارگزینه ای و بله / خیر

عبداله برادران

دانشگاه آزاد آسلامی واحد تهران مرکز

سعیده آهنگری

شکوه راشوند سمیاری

دانشگاه آزاد آسلامی واحد تبریز

در این تحقیق تاثیر فرمول استانداردcfg بر آزمونهای چهارگزینه ای و بله/خیر زبان آموزان خانم سطح متوسط در یک موسسه زبان انگلیسی بررسی گردیده است. روش تصحیح اوراق استفاده از یک فرمول بارم بندی است که بارم پاسخ های درست – پاسخ های نادرست و سئوالات بی جواب را بگونه ای در نظر گرفته که افزایش نمرات از روی حدس/گمان تقریبا" به صفر رسیده است.

نمرات زبان آموزان یکبار با استفاده از فرمول cfg و یکبار بدون استفاده از فرمول فوق در دو شکل چهارگزینه ای و بله/خیر مورد بررسی قرار گرفته است. انتخاب آزمونهای چهارگزینه ای و بله/خیر انتخاب جواب از روی حدس/گمان را مرتفع یا حداقل بطرز قابل ملاحظه ای کاهش می دهد.چرا که احتمال افزایش نمره با در نظر گرفتن فاکتور حدس همواره در این دو نوع آزمون مشهود بوده است. محقق روش تصحیح با در نظر گرفتن فاکتور حدس = عدم دانش کافی= در آزمونهای لغت چهارگزینه ای و بله/خیر را یکبار با استفاده و یکبار بدون استفاده از فرمولcfg مورد بررسی قرارداده است تا مشخص نماید زبان آموزان واقعا" به جواب سئوالات واقف بوده و یا از روی حدس و گمان به پاسخ صحیح دست یافته اند. این تحقیق فرصتی استثنایی برای مقایسه نمرات آزمونهای چهارگزینه ای و بله/خیر در شرایطی که زبان آموزان تعداد سئوالاتی برابر با مفادی یکسان در هر دو شکل آزمون در دسترس داشته اند ایجاد نموده است.

نتایج این تحقیق حاکی از آن است که تفاوت قابل ملاحظه ای در نمرات زبان آموزان با استفاده از این فرمول و بدون استفاده از آن وجود داشته است.

کلیدواژه ها: فرمول cfg ، سئوالات بله/خیر، سئوالات چهارگزینه ای