The Impact of Escape Alternative Position Change in Multiple-Choice Test on the Psychometric Properties of a Test and it's Items Parameters

Dr. Iyad Mohammed Hamadneh.
Faculty of Educational Sciences – Al al-Bayt University

Abstract
This study aimed at investigating the impact changing of escape alternative position in multiple-choice test on the psychometric properties of a test and it's items parameters (difficulty, discrimination & guessing), and estimation of examinee ability. To achieve the study objectives, a 4-alternative multiple choice type achievement test consisting of 39 items in psychological and educational measurement and evaluation was constructed. The test had four different forms according to the escape alternative position. The statistical software Bilog - MG3 was used to analyze the responses of the total sample (1521) examinees according to the three parameter logistic model (3PLM) of item response theory (IRT). The results of the study revealed that there were statistically significant differences in the means of item parameters (difficulty & discrimination) between the first and the fourth forms, due to the changing of escape alternative position; the differences were in favor of the four form. The results also showed there were statistically significant differences in the means of examinee ability between the first and the fourth forms, due to the changing of escape alternative position; the differences were in favor of the fourth form, and there were no statistically significant differences between other forms. The results also revealed there were statistically significant differences among criterion validity coefficients in favor of form four of the test. Finally, significant differences were noticed in the value of empirical reliability coefficients in favor of the fourth form.
Keywords: Multiple-Choice Tests, Escape Alternatives, Item Response Theory**.**

1.Introduction

Multiple-choice Items are a mainstay of achievement testing, and the most flexibility of measurement tools, and the benefits of this type of items in overcoming the problem of correction great responses from examinees (Liu, Lee & Linn, 2011).

Item multiple choice include two parts, "or more presentations the phrase" sometimes called "text or Stem", followed by a number of suggested answers called "Alternatives or Responses or Options". And function of the presentations ferries offer the function to be performed, or the question you want to answer, or to identify the problem to be solved. The alternatives are viable include one response is "Correct Response", and the rest of the alternatives they called "Task or Distractors," and function to provide solutions or answers, seem plausible to examine who does not know the correct answer (Allam, 2011; Aiken, 2003).

There is no doubt, that the aim of the analysis of the test items are to assess the effectiveness of building the items, so the items that must be excluded and items that must be kept analysis; because the test quality depends mainly on the quality of items, which can be achieved by psychometric properties of these items. And more to determine, if the structure of a item multiple-choice in its two parts stem or whether alternatives have an impact on the quality of, and then on the psychometric properties related to or included in the test (Aiken, 2003; Gregory, 2005). This topic has received a number of research studies that asked about the impact of the product defect for violation of the rules of the formulation of a Multiple-choice Items on the psychometric properties of the test and items and the expected impact on her performance examinee (DiBattista, Sinnige & Fortuna, 2014).

And many of the authors in the field of educational measurement and evaluation (Abu Fouda & BaniYounes, 2012; Downing, 2005; Nitko, 2001) see that in spite of the diversity of the rules of the formulation of multiple choice in the number and content of the items, but the building is considered an art and pursuant creative at the same time. In the same vein, there are alternatives used when contagious test finds it difficult to appropriate generation number of alternatives or dispersants, and focus on quantity rather than quality, which do not add knowledge of something new, so-called "Escape Alternatives" such as: "All of the above", "All that said", " none of the above", " neither of these", "cannot be determined", "b + c", and other, and called Escaping; because contagious test escapes to put it among the alternatives when it is not in his mind a suitable replacement, on the other hand escapes examinee for selected when it is not really knows the answer (Al-Nabhan, 2004).

In addition, many researchers in the same field (DiBattista, Sinnige & Fortuna, 2014; Martínez, Moreno, Martín & Trigo, 2009) have confirmed to avoid the use of escape alternatives "is not one of the above" or "none of the above" unless there was the correct answer is not challenged by one, and to avoid the use of Escape alternatives "all of the above" or "all of the" especially if they are item of the kind that requires the best

answer; as the selection or exclusion of such alternatives is linked to the selection or exclusion of other alternatives, When the first and second alternate correct, here is likely to choose alternative examinee "all of the above," apart from the other alternatives, as examinee often do not need to read or meditate out. In the same way are chosen alternative "none of the above" if the first and second alternatives sinful. In general, be careful when using such alternatives, but if he does not have to be used must each be true and some are false (AL-Nabhan, 2004; Al-Yacoub, 2000;Yacoub & Abu foodah, 2012; Downing, 2005).

It should be noted that the Item Response Theory (IRT) helps to make a lot of solutions to the problems related to the construction of testing and development, and the great assumption of this theory states that an examinee's responses to different items in a test are statistically independent. For this assumption to be true, an examinee's performance on one item must not affect, either for better or for worst, his or her responses to any other items in the test (Hambleton, Swaminathan & Rogers, 1991).

1.2 Questions of the Study
The current study seeks to answer the following questions:
1- Do estimations accuracy of item parameters (difficulty, discrimination, and guessing) using (3PLM) of (IRT) vary depending on the changing of escape alternative position in the four test forms?
2- Do estimations accuracy of examinee's ability using (3PLM) of (IRT) varies depending on the changing of escape alternative position in the four test forms?
3- Do criterion validity coefficients and the empirical reliability coefficients derived from the concepts of (IRT) change depending on the changing of escape alternative position in the four test forms?

1.3 Aims of the study
This study aimed to detect the effect of changing the position of escape alternatives "all of the above, all of the, otherwise, none of the above, cannot be determined" contained in the multiple-choice test with four alternatives in four forms (the first alternative, the second alternative, the third Alternative, the fourth alternative), which measures the examinee achievement in educational measurement and evaluation- on the psychometric properties of a test and it's items parameters (difficulty, discrimination & guessing), and estimation of examinee ability, according to the three- Parameters logistic model (3PLM) based on the concepts of (IRT).

1.4 Importance of the study
The importance of the current study in the following aspects:
First: Theoretical: The importance of this study, the importance of the topic that it considered, and of multiple choice test items alternatives, specifically escape alternatives, and the impact of changing its position on the psychometric properties of the test and it's items; Due to the wide spread of this type of tests. As expected, this study provides a theoretical framework to verify the psychometric properties of the test and its items in light of changing escape alternative position, according to scientific principles based on previous studies and the findings and recommendations, those on the Item Response Theory (IRT).
Second: In practical terms: Its importance lies in that it adopted real data, obtained from the educational attitudes and realistic, and try to detect the effect of changing escape alternative position on the psychometric properties of a test and it's items parameters, and especially as it has been the practice to be escape alternative another alternative position regardless of the number of alternatives. What if we put this alternative as an alternative to the first (A) or the second alternative (B) or a third alternative (C) and thus, does this action affect on the psychometric properties of a test and it's items parameters?. It is hoped and will reach as a result of the mechanism of the present study, the results lead to a deeper understanding of the escape alternatives.

1.5 Operational definitions
 •**Escape Alternative**: Is the alternative that does not add any new information on the level of a item individual alternatives range "all of the above, all of the, otherwise, none of the above, cannot be determined" and chosen by the largest number of examinees who do not really know the answer.
 •**Escape Alternative position**: Four forms of multiple choice test items with four alternatives were formed, which were the scrape alternatives distributed in (10) items as follows: the first form: escape alternative is the first one, the second form: escape alternative is the second one, the third form: escape alternative is a third one, the fourth form: escape alternative as usual the last one.

1.6 Limitations of the Study
1-The study sample was limited to students of the Faculty of Educational Sciences at Al al-Bayt University enrolled in the course of psychological and educational measurement and evaluation principles during the year 2013/2014, which would limit the dissemination of the results of the study outside their community.
2-The study was limited to the achievement test tool in psychological and educational measurement and

evaluation, and the items of multiple choices with four alternatives.
3-The study was limited to the use of three Parameters logistic model (3PLM) of Item Response Theory (IRT), and was used statistical software specialist available (Bilog - MG3).

2. Review of Related Literature
2.1 Theoretical Literature
Item Response Theory (IRT) has a set of assumptions, namely: Unidimensionality , and Local Items independence, Monotonicity, and Non-Speededness (Embretson & Reise, 2000). As a result of this theory is a set of models known as Latent Trait Models, and aims to identify the relationship between the performance of examinee in the test and the attribute that lies behind this performance and interpreted. These models differ in the number of item-estimated by Parameters, with a Three-Parameter Logistic Model, 3PLM overall shape of the models logistics; it includes three possible for item parameters, namely: the difficulty, discrimination, and guessing, respectively.

The item and test information functions play key roles in (IRT). Through these, it is possible to ascertain the standard errors of measurement of each item at a given level of ability θ. In contrast, the standard error of measurement obtained through classical methods is an aggregate quantity over the entire range of ability (Gruijter & kamp, 2005). Also, the use of a large sample of examinee guaranteed to get great accuracy in Parameter discrimination estimates (Crocker & Algina, 1986; Hattie, 1984).

On the other hand, the application of IRT involves two separate steps: a first one to estimate item parameters and to ensure that they match the desired characteristics in terms of psychometric properties and test requirements, and a second step to locate examinees into the latent trait scale (Chernyshenko, Stark, Chan, Drasgow & Williams, 2001).

2.2 Empirical Studies
Knowles & Welch (1992) study aimed to compare item difficulty and item discrimination indices in multiple-choice tests contain escape alternative "none-of-the-above". A meta-analysis of the difficulty and discrimination of the "none-of-the-above" test option was conducted with 12 articles (20 effect sizes) for difficulty and 7 studies (11 effect sizes) for discrimination. Findings indicate that using the NOTA option does not result in items of lesser quality. The results also indicated that contain the item of this alternative gives the difficulty and discrimination no less in quality for items that do not contain such an alternative.

Crehan, Haladyna & Brewer (1993) Conducted study aimed to a comparison of three versus four options; and the use of the inclusive "none of these" option versus a content option. 48 items were used. Each item was written in four versions: (1) four options without "none of these"; (2) four options with "none of these"; (3) three options without "neither of these"; and (4) three options with "neither of these". Item analysis and test analysis comparing the manipulated item versions were conducted using item response theory. The three-option format was found to be less difficult than the four-option format, and the use of the "none of these" option resulted in more difficult items. There was no difference in discrimination between three- and four-option formats, a possible argument in favor of a three-option format. However, the study did indicate observed differences in the reliability favoring the four-option format

The results of Taylor research (2005), this research examined 2 elements of multiple-choice test construction, balancing the key and optimal number of options. In Experiment 1 the 3 conditions included a balanced key, overrepresentation of a and b responses, and overrepresentation of c and d responses. The results showed that error-patterns were independent of the key, reflecting selection of the most plausible foil. Experiment 2 examined the optimal number of options. The comparison of a 3-option to a 4-option test showed that the 3-option test retained reliability, while allowing for a greater sampling of information

Whereas Meyers, Murphy, Goodman & Turhan (2012) study aimed to identify the impact of item position change on item Parameters difficulty and discrimination and guessing, according to three Parameter logistic model of IRT. The results showed that the values of the parameters (difficulty, discrimination, guessing) significantly affected by item position change, the change was not statistically significant change sample size.

The study of Weinstien & Roediger (2012), which aimed to identify the impact of item order on test performance assessment, the results showed optimism over the performance assessment examinees on the item ordered from easy to difficult, as the difficulty of the individual changed over the distribution of the item differences.

As indicated Results DiBattista, Sinnige & Fortuna (2014) study which aimed to assess the effects of using the escape alternative "none of these" in (40) Items multiple-choice test in public knowledge, given to college students, that the use of this alternative increases the difficulty of the item, and reduces discrimination. The researchers recommended that alternative "none of these" should not be used in multiple choice tests.

3. Methodology of the Study
3.1 Participants
The study sample consisted of (1521) students in Faculty of Educational Sciences at Al al-Bayt University, and who have studied the course of psychological and educational measurement and evaluation principles, and enrolled in (16) Division of the course itself during the first, second and summer seasons, respectively during the year 2013/2014.

3.2 Instruments
The researcher prepared an achievement test of multiple choices in psychological and educational measurement and evaluation. Where possible the formulation of a 60 multiple choice items. including four alternatives, one correct answer, has been taken into account in the formulation of the technical foundations in the writing of this kind of item (Gronlund & Linn, 1990) The test was revised by Arabic language and measurement and evaluation instructors in order to validate its content and items, after the presentation to a panel of judges and taking their suggestions, the number of test items primary as orally 53 item, and after the survey sample answers during the initial experimentation process and the study of items discriminations, 3 items has been deleted, so the test consists of (50) items cover behavioral area to be measured.

3.3 Procedures
The researcher randomly select (10) items only, representing 20% of the component test items in its final form (50) items, a items number (7, 9, 16, 17, 22, 23, 29, 32, 39, 46) respectively. It has been replaced by more choices by examinees in previous applications of this test on the item itself one of the escape alternatives possible "all of the above, all of the, none of the above, cannot be determined"; for the purposes of diversifying escape alternative and highlight it, knowing that it does not represent the correct answer in any of them, and this information certainly did not know any of the examinees.
For the purpose of achieving the objectives of the study was the formation of four tests forms, without changing the embedded item, and without a change in the order of item or in the number of its alternatives, as follows:
A) - the first form: the escape alternative in ten selected item is the first alternative.
B) - the second form: the escape alternative in ten selected item is the second alternative.
C) - the third form: the escape alternative in ten selected item is a third alternative.
D) - The fourth form: the escape alternative in ten selected item is the fourth (last one)
Then, papers were prepared instructions for testing and form answer sheet (correction key) for each of the four test forms.

3.4 Data Collection and Analysis
To achieve the purpose of the study, the researcher applying randomly test forms on the study sample totaling (1521) students, where each student on a item form of these forms answered, has supervised the researcher himself to all the application procedures, was sort each form separately, after that has securities correct procedures introduced data into computer memory, and used software statistical (SPSS) to verify the assumption of a Unidimensionality, as used statistical software (Bilog - MG3) for the detection of matching study data to (3PLM) used in the current study, as well as to investigate the effect of changing escape alternative position on the psychometric properties of a test and it's items parameters:
First, check the assumption of Unidimensionality
One of the main requirement or assumptions of Latent Trait Models (LTM) is the Unidimensional character of the measures. To check this assumption in each form of the four tests forms, using the statistical software SPSS. A factor analysis in a Principle Components Analysis (PCA). The table (1) shows the results of factor analysis method: Eigenvalues, Explained Variance for the first and the second factors, and the percentage of first Eigenvalues to the second Eigenvalues in the four test forms.
Table 1: Eigenvalues and Explained Variance% for the first and second factor and the percentage of first Eigenvalues to the second Eigenvalues in the four test forms

Test form	Factors	Eigenvalues	Explained Variance%	the percentage of first Eigenvalue to the second Eigenvalue
The first	The first	9.834	%20.038	2.061
	The second	4.771	%7.910	
The second	The first	9.827	%21.302	2.226
	The second	4.414	%8.229	
The third	The first	9.838	%21.154	2.214
	The second	4.443	%8.546	
The forth	The first	9.893	%20.618	2.141
	The second	4.620	%7,860	

Can be seen from Table (1), that the Eigenvalue of the first factor is high value when compared with the second factor Eigenvalue, in the four test forms, meaning that the first factor explained more than twice of the second factor Eigenvalue. Add to that, it is noted that the percentage of first Eigenvalues to the second Eigenvalues greater than (2) in all forms. Also, the explained variance for the first factor is higher than (20%) for each form. These are Indicators about "Unidimensionality" in the four test forms (Hulin, Drasgow & Parson, 1983; Hattie, 1984).

Second, the assumption of Local Independence:

Because the assumption of Unidimensionality equivalent of the assumption of local independence, the researcher -Precisely - only to verify the assumption of Unidimensionality to deduce the check positivist assumption of local independence (Embretson & Reise, 2000).

Third: the assumption of Non - Speededness:

Was to ensure that the examinees failure to answer the test items due to the decline in their abilities, and not to the effect of speed factor in the answer, and that by giving adequate time for them to answer the test items, in addition to that he did not complain of any examinee of time constraints and inadequate during the application test forms.

Fourth: (Goodness -of -Fit -test):

In order to verify person-fit and item-fit to the expectations of (3PLM) was used software (Bilog - MG3) to analyze the data for each of the four test forms, and produced the results of the analysis when using the software for the first time on the raw data, and through statistical Chi-square (χ^2) at the significance level ($\alpha = 0.01$) that there is person-fit in all examinees of the first and fourth test forms, the non-conformity of responses (18) They were examination of the study sample to the expectations of 3PLM, (15) examinees in the second form, and (3) examinees of the third part of the test form, where the amount of probability to each less than 0.01 (Fit probability <0.01). While the amount of error in estimating the ability of some other great, and this is what pamper him unable program in the standard error of the abilities account, by giving value (999.000 *) as an indicator of that, so it has been deleted their responses and to keep the responses (1503 examinees.

With respect to examine item- fit to the expectations of (3PLM), was re-analysis using software (Bilog - MG3) after deleting examinees non-conforming to 3PLM, where the results of the analysis showed in the second time, through statistical Chi-square (χ^2) at the significance level ($\alpha = 0.01$) that not matching (11) items in the four test forms to expectations of 3PLM, a item Nos. (2, 6, 15, 17, 24, 27, 35, 38, 44, 46, 47), and was including two items with escape alternatives, as the value of the item- fit probability for each of them less than 0.01

After deleting all the common items in the four test forms of (11) Item non-conforming to the expectations of (3PLM) in the previous phase, and then re-analysis for the third time to get the final estimates for each of item Parameters and examinees abilities on the model used. Thus, the test became a component of (39) Items for each form, including (8) items containing escaping alternatives, a item Nos. (7, 9, 16, 22, 23, 29, 32, 39).

3.5 Statistical Treatment

To answer the study questions, was used software for statistical (Bilog - MG3) (Zimowski, Mislevy & Back, 1996) to estimate item parameters (difficulty, discrimination, and guessing) for forms of the four test using (3PLM), and used1- Way ANOVA, and M statistic (Hakstain & Whalen, 1976); to detect the statistically differences among empirical reliability coefficients in the test forms. As used statistical V (Hays, 1980); to detect the statistically differences among criterion validity coefficients in the test forms.

4. Results and Discussion

4.1 First, the results relating to estimate item Parameters (difficulty, discrimination, guessing) using (3PLM) of (IRT)according to the changing of escape alternative position in each form of the four test forms. It has been estimated these parameters using the statistical software Bilog - MG3 which gives accurate estimates by re successive operations of appreciation. The table (2) shows item parameters estimates (difficulty b_i, discrimination a_i, guessing C_i) of the final test as adopted 40 items for each form (after deleting item non-conforming), as the means and standard deviations of these parameters account.

Table 2:Item parameters (difficulty, discrimination, guessing) for the four test forms according to (3PLM), and the means and standard deviations of these parameters

Form No / parameters	The first (Alternative escape: the first Alternative)			The second (Alternative escape: the second Alternative))			The third) escape Alternative : the third Alternative (			The fourth (Escape Alternative the fourth Alternative)		
	b_i	a_i	C_i	b_i	a_i	C_i	b_i	a_i	C_i	b_i	a_i	C_i
1	0.31	1.376	0.495	0.228	1.632	0.423	0.168	2.793	0.188	0.141	2.421	0.408
3	0.248	1.822	0.282	0.185	2.007	0.218	0.096	2.061	0.183	-0.05	1.972	0.204
4	-0.839	1.391	0.49	0.245	1.641	0.281	-0.091	1.49	0.379	-0.796	1.42	0.267
5	1.537	3.374	0.173	-1.137	2.62	0.372	0.233	1.773	0.266	-0.141	4.771	0.356
7	1.317	1.376	0.421	1.152	2.283	0.482	1.081	0.987	0.366	0.655	3.156	0.372
8	-2.308	1.56	0.504	-0.015	3.008	0.334	0.783	1.575	0.262	-2.04	1.716	0.325
9	1.115	1.883	0.485	0.924	2.118	0.263	0.753	0.965	0.499	-0.838	3.226	0.25
10	-0.044	2.244	0.306	1.315	2.893	0.146	0.681	2.464	0.193	0.802	2.337	0.132
11	-1.081	2.721	0.446	0.192	1.527	0.402	0.044	3.651	0.351	-0.667	1.687	0.39
12	-0.861	2.82	0.421	-0.186	2.041	0.354	1.291	2.903	0.138	0.198	2.988	0.334
13	0.631	2.985	0.251	-0.496	2.671	0.302	-0.133	3.03	0.196	-0.857	1.055	0.287
14	0.338	2.965	0.405	-1.201	1.988	0.461	0.255	1.661	0.407	0.175	1.263	0.446
16	1.09	1.823	0.332	0.853	2.225	0.438	0.681	0.771	0.198	0.353	3.112	0.423
18	-0.892	1.435	0.466	-0.785	1.468	0.466	0.849	2.151	0.25	-0.315	2.807	0.451
19	0.626	1.662	0.483	-0.918	2.95	0.355	0.791	1.026	0.375	1.044	1.697	0.337
20	-0.292	3.121	0.306	0.688	3.941	0.273	-1.231	1.913	0.395	1.313	2.216	0.261
21	0.122	1.845	0.397	-1.001	0.771	0.443	-0.823	1.383	0.432	-0.294	3.332	0.237
22	1.497	1.836	0.357	1.121	1.206	0.304	1.219	1.003	0.401	0.57	2.913	0.182
23	1.449	2.214	0.153	1.748	2.457	0.394	1.791	1.251	0.293	1.168	3.156	0.25
25	-0.316	1.955	0.358	0.566	1.038	0.409	-0.885	2.964	0.313	0.35	3.131	0.27
26	1.618	1.851	0.375	1.31	2.187	0.278	0.609	3.45	0.249	-1.707	1.662	0.491
28	-1.776	1.453	0.509	0.326	1.554	0.431	0.125	2.24	0.407	0.017	1.113	0.463
29	1.428	1.156	0.196	1.294	1.433	0.415	0.947	0.853	0.249	0.538	2.556	0.261
30	-0.668	2.541	0.403	0.391	2.725	0.293	0.173	1.055	0.31	-1.623	2.545	0.436
31	0.229	1.351	0.478	1.251	4.188	0.198	0.14	1.362	0.361	1.463	1.067	0.261
32	1.119	0.912	0.344	1.085	1.126	0.441	0.776	0.592	0.158	0.297	2.136	0.284
33	-0.929	2.651	0.432	1.151	2.282	0.162	-0.726	1.945	0.41	-1.162	2.493	0.361
34	0.441	0.993	0.478	1.402	0.896	0.274	-0.671	0.826	0.493	1.338	2.494	0.166
36	0.17	2.661	0.356	-1.128	2.901	0.394	-0.637	3.599	0.5	0.159	2.326	0.402
37	0.044	2.871	0.333	0.227	2.195	0.294	-1.602	2.67	0.43	0.482	2.054	0.273
39	1.612	3.756	0.19	1.402	3.36	0.146	0.405	1.022	0.461	-0.034	4.385	0.498
40	0.368	3.667	0.227	0.493	1.542	0.281	0.925	1.755	0.338	0.326	1.022	0.439
41	-0.515	3.264	0.309	0.924	1.842	0.28	0.22	1.826	0.301	-0.125	2.214	0.26
42	0.394	3.221	0.271	-0.115	2.006	0.305	-0.936	3.435	0.404	1.156	1.902	0.188
43	1.39	2.211	0.242	1.136	1.968	0.207	1.414	3.817	0.17	0.54	2.552	0.196
45	1.862	1.411	0.346	0.012	1.156	0.383	0.057	2.112	0.39	-0.481	3.271	0.207
48	-1.302	0.912	0.509	0.423	2.718	0.217	0.9	2.066	0.284	0.418	2.913	0.191
49	1.282	2.623	0.248	0.368	3.541	0.236	-0.114	2.15	0.286	0.941	1.782	0.229
50	-1.303	1.591	0.504	-1.151	2.302	0.414	1.141	1.51	0.17	0.7	3.291	0.139
mean	**0.235**	**2.181**	**0.363**	**0.194**	**2.233**	**0.322**	**0.173**	**2.286**	**0.319**	**0.082**	**2.523**	**0.498**
Standard deviation	**1.127**	**0.697**	**0.103**	**0.941**	**0.786**	**0.098**	**0.907**	**0.817**	**0.096**	**0.834**	**0.924**	**0.112**

Can be seen from Table 2 that the mean items difficulty Parameters values in the first form was the top (the most difficult), according to (IRT). Based on the foregoing, it is clear that there were differences in the means of items parameters (difficulty & discrimination) between the first and the fourth forms, due to the changing of escape alternative position; the differences were in favor of the four form. With regard to guessing parameter, the results indicated the same table, the highest mean guessing Parameter (0.354) was the first form, a form in which the escape alternative first alternative; This is consistent with the general trend of item difficult Parameter to be guessing difficult for a item parameter, means of items guessing, in general, were close in the four test forms.

To know the statistically significant differences in the means of item parameters (difficulty, discrimination &guessing) depending on the change the location of escape alternative in four forms, has been used (1-Way ANOVA), as shown in Table (3).

Table 3: 1-Way ANOVA of the differences between the means in items parameters estimations (difficulty, discrimination, guessing) in the four forms.

Parameters Items	Source of variation	Sum of Squares	Degree of Freedom	Mean of Sum Square	Statistical Value F	Statistical Significance
Difficulty	Between groups (forms)	0.0008	3	0.00027	4.354	0.037
	Within group(error)	0.093	1499	0.00006		
	Total	0.0938	1502			
discrimination	between groups(forms)	3.004	3	1.0013	1.41	0.042
	Within group(error)	1.07	1499	0.00071		
	Total	4.074	1502			
Guessing	between groups(forms)	6.101	3	2.0336	2.991	0.126
	Within group(error)	1.022	1499	0.00068		
	Total	7.123	1502			

Can be seen from Table (3) a statistically significant differences at the level of significance ($\alpha = 0.05$) between the means of items difficulty Parameters due to the changing of escape alternative position in the four test forms. As it turns out that there is a statistically significant differences at the level of significance ($\alpha = 0.05$) between the means of items discrimination Parameters due to the changing of escape alternative position in the four test forms.

In order to determine statistically significant differences positions, has been used scheffe test (Scheffe),as shown in Table (4).

Table 4: The results of scheffe test

Parameter	Forms No	The first	The second	The third	The forth
Difficulty	The first	_			
	The second	0.02	_		
	The third	0.03	0.01	_	
	The forth	0.08*	0.05	0.04	_
Discrimination	The first	_			
	The second	0.01	_		
	The third	0.03	0.02	_	
	The forth	0.11*	0.09	0.08	_

• Statistically significant by scheffe test ($\alpha = 0.05$)

Shown in Table (4), that there were statistically significant differences in the means of item parameters (difficulty & discrimination) between the first and the fourth forms, due to the changing of escape alternative position; the differences were in favor of the four form, in which the escaping alternative(D).

Can be interpreted this result as when you put the escape alternative as a first choice (not uncommon among examinee) (not the last alternative), this may lead to dealing with them is an actor, especially when you put the first alternative, since previous studies indicate the importance of starting to give the same effective alternatives high item and natural form of the usual unexamined; examinee in order to raise self-confidence and ensure stimulating and exciting and reduce the level of concern has to end the test. In the case of the alternative escaping is a non-sequential (the first alternative, the second alternative, a third alternative), this may lead to a feeling of frustration examinee and not wanting to end the test or termination of a non-actor (Taylor, 2005).

Perhaps that penetrate the researcher to the rules of preparation achievement tests in the event of the use of alternatives escaping, which is that in the event of the use of alternatives escaping must be some right and some wrong (Al-Nabhan, 2004; Downing, 2005; Yacoub & Abu foodah, 2012) may be more difficult Item in the four test forms in addition to changing the alternative escaping position, where all the items that included escaping alternatives in the current study, the alternative escaping which did not represent any of the correct answers. This result and caught up with all of the tools used in previous studies (Knowles & Welch, 1992;

Tollefson, 1987; Weinstien & Roediger, 2012), in which the alternative Escaping is the last option, regardless of the number of alternatives. As the results of the current study, caught up with (Crehan et al., 1993) and the study of (Knowles & Welch, 1992) which indicated that the use of escape alternative   make the item more difficult. As the result differed with the findings of the study (DiBattista et al., 2014) which recommended against the use of escape alternatives in multiple-choice items.

4.2 Second, the results relating to estimate examinee's ability using (3PLM) of (IRT) according to the changing of escape alternative position in each form of the four test forms. To find out the result of this question, examinee's ability has been estimated using Marginal Maximum Likelihood estimation (MML) through software Bilog - MG3,as shown in Table (5).

Table 5: Means and standard deviations of the examinee's ability estimates depending on changing escape alternative position in four forms

Forms No	Mean	standard deviations
The first	0.019	0.811
The second	0.012	0.856
The third	0.01	0.83
The forth	0.0008	0.897

Notes from the results contained in Table (5)that there are morphological differences between the means of examinee's ability estimates depending on the change escaping alternative position in four forms. To find out the statistical significance of the differences was used (1- Way ANOVA),as shown in Table (6).

Table 6: 1- Way ANOVA for the differences between the examinee's ability estimates in four forms.

Source of variation	Sum Square	Freedom	Mean of Sum Square	statistical value F	Statistical significance
between groups(forms)	0.041	3	0.0136		
Within group(error)	1.248	1499	0.00083	16.386	0
Total	10289	1502			

Can be seen from Table (6) that there are statistically significant differences at the level of significance ($\alpha = 0.05$) between the examinee's ability estimates in four forms due to the difference in change escaping alternative position. In order to determine statistically significant differences positions have been used scheffe test (Scheffe), as shown in Table (7).

Table 7: The results of scheffe test

form	The first	The second	The third	The fourth
The first	_			
The second	0.0031	_		
The third	0.0033	0.003	_	
The fourth	0.0535*	0.0047	0.0042	_

*Statistically significant by scheffe test ($\alpha = 0.05$)

The results of the analysis that show in the table (7) there are statistically significant differences at the level of significance ($\alpha = 0.05$) between the examinee's ability estimates due to the difference in change escaping alternative position between the first and fourth forms, in favor of the fourth form, and there were no statistically significant differences between other forms.

This is consistent with what indicated by Hambleton et al., (1991) that the accuracy examinee's ability estimates increases the closer the means of item difficult estimates. This can be explained this result in the light of the psychological characteristics of examinees, noting a number of examinees after meeting them and specifically who answered the first form (alternative escaping : The first alternative A) that the level of concern have risen, and are less concentrated in the process of the answer, and the form made them feel frustrated; the belief that this alternative is the correct answer, and the examiner is trying to hide by changing the location, and they cannot get the grades they expect. In the fourth form (alternative Escaping: the last alternative D) may substitute escaping position in the last alternatives (as usual) probably removes the doubt to the examinees that the examiner is trying to hide the correct answer, but was prepared test items natural image usual (alternative Escaping D), which may have a positive impact in improving the concentration level of examinees, and increase their confidence in the answer, and the continuation of their motivation.

4.3 Third, the results relating criterion validity coefficients and the empirical reliability coefficients depending on the changing of escape alternative position in the four test forms, according (3PLM) to the (IRT). Used statistical software (SPSS) to find criterion validity coefficients for each form of the four test forms;

by calculating Pearson correlation coefficient between the examinees scores on each form and their final scores in the course of psychological and educational measurement and evaluation principles, as shown in Table (8).

Table 8 : Criterion validity coefficients for each form of the four test forms

Form	The first	The second	The third	The forth
criterion validity coefficients	0.8	0.85	0.89	0.91

Shown in Table (8) that correlation coefficients between examinees performance on the psychological and educational measurement and evaluation course test, and the final score in the course (criterion), all of which were positive and high; which means that the covariance ratio between the test forms were high, as the lower value (0.80) for the first form, the highest value (0.91) for the fourth form.

To find out whether there is a statistically significant differences at the level of significance ($\alpha = 0.05$) between criterion validity coefficients due to change alternative escaping position in four forms, it has been the statistical use (V), which follows a Chi-square (χ^2) at the significance level ($\alpha = 0.05$) with degrees of freedom (df = 3), as shown in Table (9).

Table 9: Results of the analysis that relate with criterion validity coefficients for the four forms

Form	criterion validity coefficients	Value of V	χ^2	) df(
The first	0.08			
The second	0.85	9.42	7.93	3
The third	0.89			
The forth	0.91			

Notes from Table (9) that V value greater than Chi-square (χ^2) value; meaning there are statistically significant differences at the level of significance ($\alpha = 0.05$) between criterion validity coefficients. To detect significance differences Z test was used (Feldt, 1980), as shown in the table(10)

Table 10: The results of bilateral comparisons between criterion validity coefficients s for the four forms

Form	correlation transactions	Z fisher	value Z Calculated	Critical
The first	0.8	1.149	0.51	1.96
The second	0.85	1.19		
The first	0.8	1.149	1.34	1.96
The third	0.89	1.214		
The first	0.8	1.149	2.42*	1.96
The forth	0.91	1.433		
The second	0.85	1.19	0.86	1.96
The third	0.89	1.214		
The second	0.85	1.19	1.88	1.96
The forth	0.91	1.433		
The third	0.89	1.484	1.06	1.96
The forth	0.91	1.433		

Notes from the table (10)that there were statistically significant differences among criterion validity coefficients in favor of form four of the test, while there were no statistically significant differences between the rest of the forms.

The table (11) shows the empirical reliability coefficients depending on the changing of escape alternative position in the four test forms, according (3PLM) to the (IRT).

Table 11: Empirical reliability coefficients in the four test forms

Form	The first	The second	The third	The forth
Empirical Reliability Coefficients	0.875	0.893	0.911	0.946

Notes from the table (11) that the fourth form (alternative Escaping: D) has consistently higher than the rest of the other forms, the least of the first form (alternative Escaping: The first alternative A).

To detect the sign of the differences between empirical reliability coefficients, M statistical were use, which follows the distribution of Chi-square (χ^2) at the significance level ($\alpha = 0.05$), and degrees of freedom (df = 3), where he revealed statistically significant differences among empirical reliability coefficients at the significance level ($\alpha = 0.05$), where the statistical value calculated M (3.69) is less than the value of Chi-square (χ^2) (8.87) when the degrees of freedom (df = 3), these differences were in favor of the fourth form.

The researcher attributed the reason for this may be that the examinees were reaching the correct answer because of high ability, not because of guessing, which were relatively few in this form. It also can be

interpreted in the light of this result is referred to Allam (2011) about the factors affecting the reliability of the test, most notably the homogeneity of the group. It was found that the fourth form had a high discriminatory ability; thus the opportunity influenced by testing the stability least in this form, unlike the first form. Based on the findings of the current study, this form can be the best, in terms of the psychological properties of the items parameters and the best in estimating examinee's ability.

5. Recommendations

In light of the current study findings, the researcher recommends the following:

1-Necessity to have Escape alternative the final one of item alternatives, and when preparing achievement tests with multiple choice with four alternatives, not to be favored escape alternative in another location within the item alternatives, especially as an first alternative. And conducting similar studies about multiple-choice items psychological properties, which includes escape alternatives, to include a different number of alternatives are three alternatives or five alternatives, and includes also the largest item in other university courses, using other latent trait models (1PLM, 2PLM) of IRT, in order to consolidate the results of the current study to ensure the possibility of their wider dissemination.

References

Abu Fouda, B., & BaniYounes, N. (2012). Achievement tests (1ST Ed.).   Amman: Dar ALmasirah for Publication and Distribution.

Aiken, L. (2003). Psychological testing and assessment (11th Ed.). Boston: Person Education Group.

Allam, S. (2011). Educational and psychological measurement and evaluation, fundamentals and applications and contemporary guidance. Cairo: Dar Al-Fiker Al-Arabi for Publication and Distribution.

AL-Nabhan, M. (2004). Measurement Fundamentals in the behavioral sciences. Amman: Dar Alshorouq for publication and distribution.

Alyacoub, I. (2000). The effect of different number of alternatives and "none of the above" alternative on the psychometric properties of multiple choice test. Journal of Arab Educational,16 (2), 91- 116.

Chernyshenko, O., Stark, S., Chan, K., Drasgow, F., & Williams, B. (2001). Fitting item response theory models to two personality inventories: Issues and insights. Multivariate Behavioral Research, 36(4), 523-562.

Crehan, K., Haladyna, T., & Brewer, B. (1993). Using of an inclusive option and the optimal number of options for multiple-choice items. Educational and Psychological Measurement, 53, 241- 247.

Crocker, L., & Algina, J. (1986). Introduction to classical and modern test theory. New York: CBS College Publishing.

DiBattista, D., Sinnige, E., Fortuna, G. (2014). The "None of the above" option in multiple-choice testing: An experimental study. Journal of Experimental Education, 82(2), 168- 183.

Downing, S. (2005). The effects of violating standard item writing principles on tests and students: the consequences of using flawed test items on achievement examinations in media education. Advanced in Health Sciences Education Theory and Practice, 10(2), 133- 143.

Embretson, S., & Reise, S. (2000). Item response theory for psychologists. New Jersey: Lawrence Erlbaum Associates.

Feldt, L. (1980). A test of the hypothesis that Cronbach's alpha reliability coefficient is the same for two tests administered to the same sample. Psychometrika, 45, 99- 104.

Gregory, R. (2005). Psychological testing. Boston: Person Education Group.

Gronlund, N., Linn, R. (1990). Measurement and evaluation teaching (6th Ed.). New York: Macmillan Publishing Company.

Gruijter, D., & Kamp, L. (2005). Statistical test theory for education and psychology. Retrieved December 30, 2005 from: www.leidenuniv.nl /~ gruijterdnmde.

Hakstain, A., & Whalen, E. (1976). A k-sample significance test for independent alpha coefficients. Psychometrika, 41(2), 219- 231.

Hambleton, R. (1989). Principles and selected applications of item response theory. New York: Macmillan Publishing Company.

Hambleton, R., Swaminathan, H., & Rogers, J. (1991). Fundamentals of Item Response Theory. Newbury Park California: Stage publications.

Hattie, J. (1984). An empirical study of various indices for determining unidimensionality. Multivariate Behavioral Research, 19(1), 49-78.

Hays, W. (1981). Statistics for the social sciences (3rd Ed.). New York: Holt, Rinehart and Winson, Inc.

Hulin, C., Drasgow, F., & Parson, K. (1983). Item response theory: Applications to psychological measurement. Homewood, Illinois: Dow Jones – Irwin.

Knowles, S., & Welch, C. (1992). A meta-analytic review of item discrimination and difficulty in multiple-choice items using "none of the above". Educational and Psychological Measurement, 52, 570- 577.

Lin, L., Lee, H., & Linn, M. (2011). An investigation of explanation multiple choice items in science assessment. Educational Assessment, 16(3), 164- 184.

Martínez, R., Moreno, R., Martín, I., & Trigo, M. (2009). Evaluation of five guidelines for option development in multiple–choice item writing. Psycothema, 21(2), 326- 330.

Meyers, J., Murphy, S., Goodman, J., & Turhan, A. (2012). The impact of item position change on item parameters and common equating results under the 3PL model. Annual Meeting of the National Council on Measurement in Education, Vancouver: Canada, 1- 4.

Nitko, A. (2001). Educational assessment of students (3rd Ed.). New Jersey: Prentice-Hall, Inc.

Taylor, A. (2005). Violating convention wisdom in multiple-choice construction. College Student Journal, 39(1), 11- 30.

Tollefson, N. (1987). A comparison of the item difficulty and item discrimination of multiple-choice items using the "none of the above" and one correct options. Educational and Psychological Measurement, 47, 377- 382.

Weinstein, Y., & Roediger, L. (2012). The effect of question order on evaluations of test performance: how does the bias evolve?. Psychonomic Society, 40(1), 727- 735.

Yacoub, I., & Abu foodah, B. (2012). Violating multiple choice items writing rules on the psychometric properties of the test and it's items. Journal of Damascus University, 28 (1), 419- 443.

Zimowski, M., Mislevy, R., & Back, D. (1996). Bilog-MG3: Multiple- Group IRT analysis & test Maintenance for binary items. Chicago: Scientific Software.

The IISTE is a pioneer in the Open-Access hosting service and academic event management. The aim of the firm is Accelerating Global Knowledge Sharing.

More information about the firm can be found on the homepage:
http://www.iiste.org

CALL FOR JOURNAL PAPERS

There are more than 30 peer-reviewed academic journals hosted under the hosting platform.

Prospective authors of journals can find the submission instruction on the following page: http://www.iiste.org/journals/ All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Paper version of the journals is also available upon request of readers and authors.

MORE RESOURCES

Book publication information: http://www.iiste.org/book/

Academic conference: http://www.iiste.org/conference/upcoming-conferences-call-for-paper/

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar