Journal of Education and Practice
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.8, No.17, 2017

Could different Items arrangements affect 10th grade students' performance in multiple-choice tests in Maths, Science, and English language final exams

Abdelnaser Sanad Alakayleh, Directorate of Education Amman II

Ministry of Education, Amman – Jordan

Abstract

The present study investigated the effect of the order of the subjects and tests of mathematics, science and English language tests on the performance and achievement of the tenth grade students in Jordan. The study sample consisted of 764 students selected from twenty regular government schools. The study adopted quasi-experimental design. The study instrument used was a multi-choice type of 48 subjects for mathematics, science and English. The test items were arranged according to their difficulty for three processors (RDM), Easy -to-Hard (ETH) and Hard-to-Easy (THE). The data collected were analyzed statistically using the analysis of mono-variance at the statistical significance level of 0.01.

The results of the analysis indicated that for mathematics, science and English courses, and in the order of items, the change in performance was statistically and morally significant, and it was found that the proposal to rearrange the test items for final achievement tests to control the penetration and misbehavior of the test rules and controls may not be optimal, The study of the English language, which did not show significant differences and suggests the use of other methods such as random model and parallel and split- half tests and further studies of other levels of study.

Keywords: Multiple-choice, Item-position, Student Performance

Introduction:

The tests play an important role in the assessment of the student or the examinee and determine the level of his performance on the course being tested for it, Sabah, S., & Hammouri, H (16) , Haladyna, T. M., Downing, S. M., & Rodriguez, M. C (11) as it is conducted to provide the examinees to benefit them in the decision-making of their future in education or career and career trend, and is the best choice because it is mostly devoid of bias and give the candidates equal opportunities equally among all the participants in the test.

Despite their high positive uses, they have social and educational applications and determinants, as their use shows facts, data and specific details of the subjects examined. Anastasi (4), Cacko, I (7) pointed to the violation of the privacy of the subjects and concluded that during the test, the examinees may be forced to show certain facts and details. A test of collection, performance, intelligence, or diagnosis may reveal the limited skills and knowledge that the examinee wishes to show, as well as criticism of the methods of conducting the test that lead to an inelastic or consistent classifications.

Worthen and Spandel (24) noted that the ratings resulting from the tests may reduce low-grade students on the test, and that the fear of the social and occupational applications of the test generates concern with the test and its consequences thereafter, and to try to circumvent the negative effects imposed by the social constraints of the test and included hinder academic progress and the loss of professional progress and brand unrealized.

Many of the examinees resort to different types of negligence or penetration or misconduct during the test, which is sometimes strong to explain and change the personal understanding and the classification of the examinees in a negative and unreal, and this in turn threatens the values and concepts of the community, which may lead to mis-selection in critical and important situations.

Sam Nii Nmai Ollennu (17) had pointed out that Unfortunately, the annual report of the 25th West African Testing Council in 2004 in Freetown showed that malpractice, penetration and test violations are on the

rise. Currently, the situation is worsened by the use of cellular phones, loudspeakers and SMS through the Bluetooth-enabled application during the test.

The annual report issued by the Jordanian Ministry of Education for 2016 , since 2012, the Jordanian Ministry of Education has adopted a rigorous approach to applying the standards and procedures in a strong and rigorous manner to those who have been tested in the general secondary school certificate, in order to reduce the malformation and Non-real inflation of students' scores in the years prior to 2012 which did not reflect the real level of students In the category of students with high grades of the rate of more than 95%, and some procedures went to a greater intensity, where the entry of pens and phones, Installation of cellular sensors and jammers in exam room, and the results of these procedures quickly and remarkably reduced the rate of distortion in the test scores significantly and stabilized the degrees of students within the normal distribution, after its application for four consecutive years from 2012-2016.

The study of Adeyegbe and Oke (1) referred to different forms of penetration of the test and its rules and misconduct of the subjects such as theft and the exchange of paper examined with another or ask for help from another person such as teachers, as well as collusion and plagiarism or leak or insult and assault on the supervisors of the test, By applying parallel tests to students with the same psychometric characteristics. This view is reinforced by Anastasi's (4) observation that the use of different test models reduces the likelihood of cheating on subjects.

Pettijohn (15) and Cacko (7) added that many teachers and examiners give mixed tests between multiple choice tests and other forms of items or in order to reduce fraud and sources during testing. Carlson and Ostrosky (9) suggest that multiple test models can be used for a little fraud. Large and wide rows However, Baghaei, P., & Amrahi, N (6), the question raised is: Is the arrangement of or test items affect the performance of examinees or not?.

Literature Review:

The researchers did not find a single or unified position on the question of whether the order of the or the items and the change in location in the test affects the performance of the examinees or students, Anastasi (4) indicated that the performance of the subjects to the order of the test items, as well as recent studies that indicated such The effect of WAEC, Lagos, Ahuman and Clock (2), Skinner (22) , Shepard (21) , Soyemi (23) , Sax and Cromack (20) , and others, some of which agreed on the existence of the effect and differed each other on the absence of the effect of the order of the or their placement in the tests.

This remains controversial among the researchers. It also affects the number of items, the quality of the subjects, their social environment, the subject of the test, the number of subjects and the test instructions and their quality, whether it is (RDM), Easy -to-Hard (ETH) and Hard-to-Easy (HTE).. However, to ensure neutrality and distance from bias, Is the best way to test because it gives all tested equal opportunity with all other examinees When applying random assumptions to the study community and its sample, which some researchers have mentioned in their studies.

The study objective:

The main objective of this study was to detect and investigate and determine whether there was an effect on the position and order of the or items in the Mathematics, Science and English Language final multiple-choice test on the performance of the examinee. Based on this objective, the study came in an attempt to answer the following questions:

1. What would be the effect of a change in item order on examinees performance in Mathematics?

2 - What would be the effect of a change in item order on examinees performance in Science?

3 - What would be the effect of a change in item order on examinees performance in English Language?

Methodology:

The study was based on a quasi-experimental design because random selection of all 10th grade students in Jordan was not possible. The possibility of generalization was another reason for selection. The independent variable of the study was determined by the order of the or items and the dependent variable in the performance of the subjects. An officer who was made up of the subjects who made the random selection while

the treatment was the program groups that were distributed among them is Easy -to-Hard (ETH) and Hard-to-Easy (HTE). We mean here by processing the rearrangement of the randomization of the organization among the three easy models for the Examinee. random alternating between them.

population and Study Sample:

The study population consisted of all students of the 10th grade for the academic year 2015/2016 (94495) According to the statistical report of the Jordanian Ministry of Education. The sample consisted of (764) student randomly chosen among the twenty basic schools of formal education in Jordan.

Study instrument:

The study instrument consisted of a multiple-choice type of (45) items for each course of English language, mathematics and science, which were developed by teachers and specialists with long experience in teaching these courses and their questions from the different directorates of education in Jordan. Second, items of the three tests were presented to a group of arbitrators from the experts of measurement and evaluation to present their opinions in the three and the three test subjects. Some of the were deleted and some were added and some of them were also modified to suit the good test specifications and the psychometric characteristics of the arbitrators. Where the items of tests stabilized on (42) item.

The specified time was the quota time has been chosen in schools set at (45) minutes for conducting and applying the test for each of the three courses test. The first model (ETH) was given to a sample of (64) students from two different divisions of the same 10th grade. After that, the response papers were collected from the subjects and were automatically corrected and their difficulty and discrimination values were calculated. The values of the reliability indicators for the three tests were also calculated as other Psychomertric Properties besides the validity of an arbitrators using the statistical package for Social Sciences SPSS. The value of the reliability index for the mathematics test (0.84) and science (0.87), while the English language (0.83) They may be acceptable for such tests in the study.

After analyzing the results of the items, the items were rearranged (E-H), (H-E), while the random order model was retained as it was for the math and science decision makers based on the difficulty of the items. This is because one section's metrics cannot be applied to another.

Procedures:

The study instrument has applied a set tests on the subjects of the students in twenty different basic school and the official day for final tests by trying to ensure that the affected subjects of any time changes to a test In order does not constitute a factor of anxiety that may lead to change the test stream results, have been applied standards official test issued by the Ministry of Education of Jordan in terms of time and the way and the form and content of the test and thoroughness of the decision, where each school took only one option for the subject with its participation of all subjects, and in a step to ensure neutrality and distance from bias in the treatment, each group of respondents took the specific order of the subject of the three models (E-H), (H-E), (RDM).

Data analysis:

The sampling differential volume between the different schools involved in the study paid researcher to remedy this discrepancy through calibration of test scores before the analysis and to provide the study by differing characteristics and cases from one school to another, and to compensate for lost values for students who did not complete all the test items as well as other unknown factors that may be observed during the application, correction and analysis of the test results.

After the calibration processthe data and test results of the three courses of mathematics, science and English were subjected to the analysis of the one-way ANOVA at the level of statistical significance ($\alpha = 0.01$), The independent variable was the order of the test items at the three levels (E-H), (H-E) , (random), while the dependent variable is the test scores, and to ensure equal variances study population researcher has conducted a preliminary test to examine the homogeneity of variance using the test Dunnett C comparisons deliberately multi-dimensional universe assumptions of homogeneity of variance is uncertain. Hahne, J (10), DeMars, C (9).

Results:

Research Question 1: What would be the effect of a change in item order on examinees performance in final mathematics test?.The results of the analysis of the mathematics test scores are as shown in Table 1.2

Table1. Descriptive Statistics for Performance in Mathematics

Order	N	Mean	S.D
Random (RDM)	**54**	**28.0**	**2.08**
Easy-to-Hard (ETH)	**132**	**13.5**	**5.65**
Hard-to-Easy (THE)	**38**	**15.2**	**5.78**
Total	**224**	**19**	**7.13**

Table 2. One-way ANOVA for Performance in Mathematics

Source	Sum of Squares	df	Mean Square	F	Sig
Between groups	**9884**	**2**	**4942**	**145.39**	**0.001**
Within groups	**7512**	**221**	**33.99**		
Total	**17396**	**223**			

In Table 2, the analysis of the One way ANOVA of the mathematics test shows significant statistical results of F (2,221)= 145.39 at p = 0.001. While the preliminary homogeneity of variance test at .01 level of significance showed that the differences of individuals examined at the level of statistical significance were not equal.

The results of the Dunnett C multiple comparisons post hoc test indicated that at the 0.01 level of significance

1) There were significant differences in performance between the use of the RDM(M=28, S.D = 2.08) and the model (E-H) (M=13.5, S.D= 5.65) for the RDM in the treatment.

2) There were significant differences in performance between the use of RDM(M=28, S.D = 2.08) and the model (H-E) (M= 15.2, S.D = 5.78) for the RDM in the treatment.

3) No significant differences in performance between the use of the model (H-E) (M=15.2, S.D =, 5.78) and the model (E-H) (M=13.5, S.D =, 5.65) in the treatment.

Therefore, the answer to the research question1: What would be the effect of a change in item order on examinees performance in final mathematics test?. Were significant for the mathematics test.

research question 2: What effect the order of the items on the performance of the subjects in the final science test?

The results of the analysis of the science test scores are as shown in Table 3.4

Table3. Descriptive Statistics for Performance Science

Order	N	Mean	S.D
Random (RDM)	**22 8**	**14.1**	**8.61**
Easy-to-Hard (ETH)	**261**	**16.2**	**9.12**
Hard-to-Easy (THE)	**49**	**8.7**	**4.48**
Total	**572**	**13.82**	**9.2**

Table 4. One-way ANOVA for Performance in Science

Source	Sum of Squares	df	Mean Square	F	Sig
Between groups	**3254**	**2**	**1627**	**29.4**	**0.001**
Within groups	**31482**	**569**	**55.33**		
Total	**34736**	**571**			

In the table 3 and 4, the analysis of the One way ANOVA of the science test shows significant statistical results of F (2,569)= 29.4 at p = 0.001. While the preliminary homogeneity of variance test at .01 level of significance showed that the differences of individuals examined at the level of statistical significance were not equal.

The results of the Dunnett C multiple comparisons post hoc test indicated that at the 0.01 level of significance

1) There were significant differences in performance between the use of RDM(M=14.1, S.D =, 8.61) and the model (E-H) (M= 16.2, S.D =, 9.12) for the RDM in the treatment.

2) There were significant differences in performance between the use of RDM(M=14.1, S.D =, 8.61) and the model (H-E) (M=8.7, S.D =, 4.48) for the RDM in the treatment.

3) No significant differences in performance between the use of the model (H - E) (M=8.7, S.D =, 4.48) and the model (E-H) (M=16.2, S.D =, 9.12) for the model (E-H) in the treatment.

Therefore, the answer research question 2 is: What affect the order of the or items on the performance of the subjects in the final science test? Were significant for the science test.

The research question 3 : What would be the effect of a change in item order on examinees performance in Final English Language test?? The results of the analysis of the science test scores are as shown in Table 5.6

Table 5. Descriptive Statistics for Performance English Language

Order	N	Mean	S.D
Random (RDM)	**65**	**18.6**	**6.11**
Easy-to-Hard (ETH)	**128**	**13.5**	**5.18**
Hard-to-Easy (THE)	**272**	**15.1**	**8.72**
Total	**465**	**15.3**	**8.13**

Table 6. One-way ANOVA for Performance in English Language

Source	Sum of Squares	df	Mean Square	F	Sig
Between groups	**512**	**2**	**256.1**	**5.93**	**0.01**
Within groups	**19942**	**462**	**43.16**		
Total	**20454**	**464**			

In table 5 and 6, The analysis of the One way ANOVA of the English language final test showed significant statistical results of F (2,462)= 29.4 at p = 0.01, While The preliminary homogeneity of variance test at .01 level of significance showed that the differences of individuals examined at the level of statistical significance were not equal.

The results of the Dunnett C multiple comparisons post hoc test indicated that at the 0.05 level of significance

1) There were significant differences in performance between the use of RDM(18.6M =, 6.11S.D =) and the model (E-H) (M=13.5, S.D =, 5.18) for the RDM in the treatment.

2) There were significant differences in performance between the use of RDM(18.6M =, 6.11S.D =) and the model (H-E) (M=15.1, S.D =, 8.72) for the RDM in the treatment.

3) No significant differences in performance between the use of the model (H - E) (M=15.1, S.D =, 8.72) and the model (E-H) (M=13.5, S.D =, 5.18) for the model (E-H) in the treatment.

Therefore, the answer to the question of the third study is: What affect the order of the or items on the performance of the subjects in the final test of the English language? Was significant for the English language test.

Discussion:

The results were agreed with the results of WAEC (7), Ahuman Clock, Skinner, Shepard, Sam Nii Nmai Olennu (17), While some researchers believe that the order of the items in the test changed or changed its shape The results of the study differed with the results of the current study with some results of studies such as Laffittee (12), Allison (3), Soyemi, Perlini (14), Lind and Mumbo. Which did not find significant differences in the performance of the subjects in the order of the test items depending on the difficulty or randomness, and may be the reason for this difference to the different levels of education that applied the researchers in the studies mentioned the test them, WEngin-Demir, C(23) , Öztürk, D. Ve Uçar, S (13) the different environments coming from them and the different curricula they learned from different stages of study From basic education to higher education in various institutes and universities.

In view of the results of the English language final exam, which showed that there were no statistical and significance differences between the RDM on the one hand and the two models (ETH), (THE) on the other hand, it may be due to the fact that English is a second language for the subjects, The results of the national test conducted by the Jordanian Ministry of Education annually 2016 for some grades of the basic stage, including the tenth grade, show the weakness in the performance of students especially in the basic skills of writing and expression in this course, which shows the results in different models, methods and methods of providing tests, and therefore it has been a bad background for students and parents It is not surprising that these results appear in the analysis of the English language final test.

Conclusion and Recommendations:

The main objective of the present study was to identify and determine whether there was an effect on the level of multiple choice tests on the level of students 'performance in the final test of Mathematics, Science and English. The results showed that changing the order of the cages and items affects students' And the use of the order of and items and rearrangement in order to control malpractice and neglect and break the rules of the standard test during the test may not be the best method in the tests of achievement because the They may be used to such behavior with the complicity or leniency of some teachers or examiners.

It is therefore necessary to be cautious when rearranging the and test items to ensure the desired results of the rearrangement Attali, Y. & Bar-Hillel, M (5) , and perhaps the use of the RDM away from the order of and test items be a good choice between the models in addition to the use of parallel , split-half or equivalent tests to ensure the highest degree of impartiality and performance best , and the current study recommends through its results researchers to conduct further studies on other courses and classes at different levels of study to approve or oppose the results of the current study and other studies that examine the impact of the order of and items of test on The final examinees of the infected.

REFERENCES:

1. Adeyegbe, S. O. A., & Oke, M. G. (1994, September). The new and widening dimension of examination malpractices and the effect on the integrity of educational credentials in the West African sub-region. Compilation of Papers Presented at the 12th Annual Conference of the Association for Educational Assessment in Africa (AEAA).
2. Ahuman, S. W., & Clock, N. D. (1971). Item difficulty level and sequence effects in multiple-choice achievement tests. Journal of Educational Measurement, 9 (Summer), 105-11.
3. Allison, D. E. (1984). Test anxiety, stress, and intelligence-test performance, Measurement and Evaluation in Guidance, 16, 211 – 217.
4. Anastasi (4), A. (1976). Psychological testing, New York: Macmillan Press Ltd.
5. Attali, Y. & Bar-Hillel, M. (2003). Guess where: The position of correct answers in multiple-choice test items as a psychometric variable. Journal of Educational Measurement, 40, 109-128.
6. Baghaei, P., & Amrahi, N. (2011). The effects of the number of options on the psychometric characteristics of multiple choice items. Psychological Test and Assessment Modeling, 53, 192-211.
7. Cacko, I (1993). Preparation of good objective test items as a step toward obtaining valid assessment of students' achievement at the SSSCE. Articles of WAEC Monthly Seminar, Accra, March 1993 ed., 87 – 92.

8. Carlson, J. L., & Ostrosky, A. L., (1992). Item sequence and student performance on multiple-choice exams: Further evidence. Journal of Economic Education, 23 (3), 232 -235.
9. DeMars, C. (2010). Item response theory: Understanding statistics measurement. New
10. Hahne, J. (2008). Analyzing position effects within reasoning items using the LLTM for structurally incomplete data. Psychology Science Quarterly, 50, 379-390.
11. Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiplechoice item-writing guidelines for classroom assessment. Applied Measurement in Education, 15, 309-334.
12. Laffittee, R. G. (1984). Effects of item order on achievement test scores and students' perceptions of test difficulty. Teaching of Psychology, 11(4), 212 - 214.
13. Öztürk, D. Ve Uçar, S., (2010). The Determination and Comparison of Factors Affecting 4th And 8th Grade Students Mathematics And Science Achievement by Means of TIMMS Data in Taiwan and Turkey.
14. Perlini, A. H., Lind, D. L., & Mumbo, B. D (1998). Context effects on examinations: The effects of time, item order and item difficulty, Canadian Psychology, 39 (4), 299 – 307.
15. Pettijohn II, T. F., & Sacco, M. F. (2001). Multiple-choice exam order influence on student performance, completion time and perceptions. Journal of Instructional Psychology, 34 (3), 142 –149.
16. Sabah, S., & Hammouri, H. (2010). Does subject matter matter? Estimating the impact of instructional practices and resources on student achievement in science and mathematics: Findings from TIMSS 2007. Evaluation & Research in Education, 23(4), 287-299.
17. Sam Nii Nmai Ollennu1,* , Y. K. A. Etsey2.(2015). The Impact of Item Position in Multiple-choice Test on Student Performance at the Basic Education Certificate Examination (BECE) Level, Universal Journal of Educational Research 3(10): 718-723.
18. Sax, G., & Cromack, T. A. (1966). The effects of various forms of item arrangements on test performance. Journal of Educational Measurement, 3 (Winter), 309-11.
19. Shepard, L. A. (1997). The challenges of assessing young children appropriately. In Katheleen M. Cauley (12th ed.). Educational Psychology. Shefield: Dubuque Inc.
20. Skinner, B. F. (1999). When the going gets tough, the tough gets going: effects of item difficulty on multiple-choice test performance. North American Journal of Psychology, 1 (1), 79 - 82.
21. Soyemi, M. O. (1980). Effect of item position on performance on multiple-choice tests. Unpublished M.Ed. dissertation, University of Jos.
22. WAEC (1993). The effects of item position on performance in multiple choice tests. Research Report, Research Division, WAEC, Lagos.
23. WEngin-Demir, C. (2009). Factors influencing the academic achievement of the Turkish urban poor. International Journal of Educational Development, 29(1), 17-29. right, B. D., & Stone, M. H. (1979). Best test design. Chicago, IL: Mesa Press.
24. Worthen, B. R., & Spandel, V. (1991). Putting the standardized test debate in perspective. Association For Supervision And Curriculum Development, Educational Leadership (Feb. 1991 ed).