

The Impact of Item Position in Multiple-choice Test on Student Performance at the Basic Education Certificate Examination (BECE) Level

Sam Nii Nmai Ollennu^{1,*}, Y. K. A. Etsey²

¹The West African Examinations Council, Ghana

²Department of Educational Foundations, University of Cape Coast, Ghana

Copyright © 2015 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract The study investigated the impact of item position in multiple-choice test on student performance at the Basic Education Certificate Examination (BECE) level in Ghana. The sample consisted of 810 Junior Secondary School (JSS) Form 3 students selected from 12 different schools. A quasi-experimental design was used. The instrument for the project was a multiple-choice test consisting of 40 items in each of English Language, Mathematics and Science. The items were arranged using the difficulty order to obtain the three treatments i.e. Random (RDM), Easy-to-Hard (ETH) and Hard-to-Easy (HTE). The data collected were subjected to statistical analysis using ANOVA at .05 level of significance. The results of the analysis indicate that for English Language, Mathematics and Science at the BECE level, when item order was altered, the difference in performance was statistically significant. The study led to the conclusion that the proposition of using re-ordering of items of an objective test to curb examination malpractice may not be the best after all especially in English Language, Mathematics and Science at the BECE level. It was therefore recommended that other methods should be investigated for the purpose.

Keywords Multiple-choice, Item-position, Student Performance

1. Introduction

Tests play an important role in determining achievement and certifying attainment. Tests are also used in providing incentives and goals for students, and providing information for decision-making. Thus, in matters of selection for higher education and placement into jobs, tests have been a good choice since they are devoid of most biases that may be termed systematic and provide equal chances to all candidates.

In spite of their highly advantageous use, tests have varied social implications and limitations. Consequently, the use of tests in obtaining facts and data about people has received its fair share of criticisms. Anastasi [4] raised the issue of invasion of privacy of the examinee. She asserted that through a test examinees may be compelled to reveal certain facts and details about themselves which may embarrass them. Certainly, any intelligence, aptitude or achievement test may reveal limitations in skills and knowledge that an individual would rather not disclose. Another limitation put forward by critics is the fact that testing produces rigid grouping practices so far as the test result leads to inflexible classification, in terms of categorization, labelling and grading. Worthen and Spandel [18] argued that classification could be demeaning and insulting and harmful to students who are relentlessly trailed by low test scores. This obviously carries connotations which may cause more harm than any gain that could possibly come from such classifications.

Fear of the demeaning social implications of tests, generate the anxieties that accompany test taking and receiving of test scores. In an attempt to circumvent the negative effects imposed by the social implications of testing, which include impedance to academic progress, forfeiture of professional advancement and promotion, and the stigma of being labelled a non-achiever, many examinees resort to various kinds of examination malpractices during test-taking. The temptation to indulge in this vice is sometimes so strong that candidates who could be classified as well-behaved and would ordinarily not approve of wrong-doing fall prey to it. Yet, the phenomenon is a vice that should not be tolerated since it threatens the moral fibre of the society and can lead to the selecting of misfits into vital and sensitive positions. Unfortunately, in a report presented at the 52nd Annual Council Meeting of The West African Examinations Council (WAEC) in Freetown, Sierra Leone in March, 2004 it was clearly indicated that examination malpractices are on the increase. In recent times, examination malpractice has

assumed a sophisticated technological dimension with the use of the cell phones as a means of transmitting responses to both multiple-choice and essay tests by both voice and text messages.

Examination malpractices are acts that contravene the rules and regulations which govern the conduct of examinations. The act could happen before, during or after the examination. Adeyegbe and Oke [1] defined examination malpractice as ‘an impropriety, an improper conduct to one’s advantage during an examination’ (p.1). Examination malpractice takes different forms which include bringing foreign materials into the examination hall, irregular activities such as stealing, converting or misappropriating other candidates’ scripts, substituting or exchanging worked scripts during or after the examination, seeking or receiving help from non-candidates such as invigilators, teachers or other personalities during the examination. Other forms of malpractices are collusion, impersonation, leakage and insult or assault on the supervisor or invigilator.

One of their recommendations to examining bodies for curbing such practices, Adeyegbe and Oke [1] said, is to “think of administering parallel tests to different students but having the same psychometric properties” (p. 11). According to Anastasi [4] the use of several alternative forms of a test provides a means of reducing the possibility of cheating. Pettijohn and Sacco [12] reported that to prevent cheating on examinations, many professors will mix up the order of multiple-choice test questions from examination to examination without thought of the consequences the change of order may have on student examination performance and perceptions. Carlson and Ostrosky [6] stated that multiple forms of an examination are frequently used as a means of reducing likelihood of cheating in large classes. However, they noted that questions have been raised regarding whether the order of test items influences student performance. Others like Bresnock, Graves and White [5] have expressed similar reservations.

2. Literature Review

In the literature survey, it was discovered that researchers are not unanimous in their findings as to whether or not altering item position in a multiple-choice test would affect performance adversely. MacNicol [10] investigated the effects of changing an “easy-to-hard” arrangement to either hard-to-easy or a random arrangement. He found out that the hard-to-easy arrangement was significantly more difficult than the original easy-to-hard order while the random arrangement was not significantly different. Anastasi [4] argued that different arrangement of items will affect performance. This view is supported by Cacko [6]. Researchers in the Research Division, WAEC, Lagos [17] discovered that different arrangements of items could affect performance adversely or positively depending on the subject in question. Shepard [14] asserted that tiny changes in test format (or arrangement) can make a large difference

in student performance. Recognizing the importance of appropriate arrangement of test items, Sax and Cromack [13] and Ahuman and Clock [2] have advised that tests should be constructed in an easy-to-hard item-difficulty sequence.

Gerow [8] and Allison [3] however found no difference in performance when items were arranged according to a certain order of difficulty or randomly. Soyemi [16] also found no significant differences between easy-to-hard and hard-to-easy arrangement, easy-to-hard and random order; and hard-to-easy and random order.

The fact that there is no consensus among researchers from the literature review indicates that there is a problem and this provided the motivation for the study.

3. Purpose

The purpose of this study was to investigate the impact of item position in multiple-choice test on student performance at the Basic Education Certificate Examination (BECE) level in Ghana. Specifically, the study sought to answer the following research questions: (1) What would be the effect of a change in item order on candidates’ performance in English Language at the BECE level? (2) What would be the effect of a change in item order on candidates’ performance in Mathematics at the BECE level? (3) What would be the effect of a change in item order on candidates’ performance in Science at the BECE level?

4. Method

Design

A quasi-experimental design was adopted. The method was adopted because it was impossible to randomly select candidates from all over Ghana for the study. It was also chosen because it affords the opportunity to generalize over the population of BECE candidates from which the sample was chosen. The major variable in this design is the performance of the BECE candidates. The control group are those who took the Random option and the treatment or programme groups are those who took the Easy-to-Hard and Hard-to-Easy options. The treatment is the re-arrangement of the random order into the Easy-to-Hard and Hard-to-Easy alternatives.

Participants

The target population for the study comprised 308,325 JSS Form 3 students who took the April 2006 Basic Education Certificate Examination (BECE). A sample of 810 students were used who came from 12 different schools, purposively selected to represent (i) public school in an urban area, (ii) private school in an urban area; (iii) public school in a rural area; and (iv) private school in a rural area.

Materials

A multiple-choice test consisting of forty items was

developed in each of English Language, Mathematics and Science. The items were developed by professional test developers to cover all areas in the West African Examinations Council (WAEC) syllabus for the subjects involved. After construction, the items were submitted to experts in the various subject fields for a second look. The structure of all the papers was in consonance with the format and item specifications used by WAEC. In all the papers, examinees were expected to answer all 40 questions and they had 45 minutes to do so which was in accordance with the usual time allocated by WAEC for such tests.

The first type of papers to be crafted was the random ordered ones. They were then piloted in a conveniently chosen public school in Adabraka, a suburb of Accra on 82 pupils who were drawn from two streams of JSS Form 3. The responses were captured on scannable objective answer sheets and machine scored. They were then subjected to item analysis. From the results of the analysis, items which needed modification were re-fixed, although these were very few. The difficulty levels of the items were also determined.

With the results from the item analysis, the items were re-arranged, one from Easy to Hard, and another from Hard to Easy, while keeping the Random order intact. Arranging the items in order of difficulty was straight forward for Mathematics and Science since they both had only one section each. However, for English Language the difficulty order arrangement was on sectional basis. This was because rubrics for one section could not be applicable to another.

Procedure

The test instruments were administered to JSS Form 3 candidates in twelve different schools, four weeks before the final examination was taken at the national level. The tests took two days to administer with the assistance of professional test administrators. The conduct of the tests was strictly according to standards of the West African Examinations Council. Each school took only one of the options of a subject but participated in all subjects. Whether a cohort took the Easy-to-Hard (ETH), Hard-to-Easy (HTE) or Random (RDM) arrangement of a particular subject, was determined by a time-table to ensure fair participation in all the treatments. The drawing of the time table followed the pattern in Table 1:

Table 1. Time Table Pattern for Test Administration

School	English	Mathematics	Science
School 1	RDM	ETH	HTE
School 2	ETH	HTE	RDM
School 3	HTE	RDM	ETH

Data Analysis

The examination scores were first standardized before analysis. The reasons for this were to make up for variances in the sample sizes from the different schools, cater for variance in conditions and facilities from one school to the other, make up for missing values for candidates who may

not have been able to complete all the items in a test and provide for other unknown factors which may not have been noticed.

The data collected for each of English Language, Mathematics and Science were separately subjected to one-way analysis of variance (ANOVA) at the .05 level of significance. The independent variable was item order with three levels as Random order (RDM), Easy-to-Hard order (ETH) and Hard-to-Easy order (HTE). The dependent variable was the scores of the tests. A preliminary test for homogeneity of variance was performed to ascertain if population variances were equal. The post hoc test used was Dunnett C since homogeneity of variance assumption was not met.

5. Results

Research Question 1

What would be the effect of a change in item order on candidate's performance in English Language at the BECE level?

The results for the analysis of scores for English Language are given in Tables 2 and 3.

Table 2 gives the descriptive statistics for the three levels of the item order.

Table 2. Descriptive Statistics for Performance in English Language

Order	N	Mean	Std. Deviation
Random (RDM)	75	33.0	3.09
Easy-to-Hard (ETH)	158	18.7	6.42
Hard-to-Easy (HTE)	39	19.8	6.74
Total	272	22.8	8.52

Table 3 gives the results of the ANOVA for English Language.

Table 3. One-way ANOVA for Performance in English Language

Source	Sum of Squares	df	Mean Square	F	Sig
Between groups	10773	2	5386.6	162.7	.001
Within groups	8904	269	33.1		
Total	19677	271			

The preliminary homogeneity of variance test at .05 level of significance showed that population variances were not equal. In Table 3, the one-way ANOVA for English Language showed a statistically significant result, $F(2, 269) = 162.7, p = .001$.

The results of the Dunnett C multiple comparisons post hoc test indicated that at the .05 level of significance,

- (i) there was significant difference in performance between the Random ($M=33.0, SD=3.09$) and Easy-to-Hard ($M=18.7, SD=6.42$) treatments; Performance was better in the random order treatment.

- (ii) there was significant difference in performance between the Random (M=33.0, SD=3.09) and Hard-to-Easy (M=19.8, SD=6.74) treatments; Performance was better in the random order treatment.
- (iii) there was no significant difference in performance between the Easy-to-Hard (M=18.7.0, SD=6.42) and Hard-to-Easy (M=19.8, SD=6.74) treatments.

Thus an answer to Research Question 1 is that the effect of a change in item order on candidates' performance at the BECE level was significant with regard to English Language.

Research Question 2

What would be the effect of a change in item order on candidate's performance in Mathematics at the BECE level?

The results for the analysis of scores for English Language are given in Tables 4 and 5.

Table 4 gives the descriptive statistics for the three levels of the item order.

Table 4. Descriptive Statistics for Performance in Mathematics

Order	N	Mean	Std. Deviation
Random	249	15.2	7.95
Easy-to-Hard	291	18.3	8.08
Hard-to-Easy	64	10.4	3.85
Total	604	16.2	8.06

Table 5 gives the results of the ANOVA for Mathematics.

Table 5. One-way ANOVA for Performance in Mathematics

Source	Sum of Squares	df	Mean Square	F	Sig
Between groups	3677	2	1838.6	32.1	.000
Within groups	35528	601	59.1		
Total	39205	603			

The preliminary homogeneity of variance test at .05 level of significance showed that population variances were not equal. In Table 5, the one-way ANOVA for Mathematics showed a statistically significant result, $F(2, 601) = 32.1, p = .000$.

The results of the Dunnett C multiple comparisons post hoc test indicated that at the .05 level of significance,

- (i) there was significant difference in performance between the Random (M=15.2, SD=7.95) and Easy-to-Hard (M=18.3, SD=8.08) treatments; Performance was better in the Easy-to-Hard order treatment.
- (ii) there was significant difference in performance between the Random (M=15.2, SD=7.95) and Hard-to-Easy (M=10.4, SD=3.85) treatments; Performance was better in the random order treatment.
- (iii) there was significant difference in performance between the Easy-to-Hard (M=18.3, SD=8.08) and Hard-to-Easy (M=10.4, SD=3.85) treatments. Performance was better in Easy-to-Hard order

treatments.

Thus an answer to Research Question 2 is that the effect of a change in item order on candidates' performance at the BECE level was significant with regard to Mathematics.

Research Question 3

What would be the effect of a change in item order on candidate's performance in Science at the BECE level?

The results for the analysis of scores for Science are given in Tables 6 and 7.

Table 6 gives the descriptive statistics for the three levels of the item order.

Table 6. Descriptive Statistics for Performance in Science

Order	N	Mean	Std. Deviation
Random	134	15.4	5.21
Easy-to-Hard	35	13.5	4.29
Hard-to-Easy	295	16.7	7.83
Total	464	16.1	7.00

Table 7 gives the results of the ANOVA for Science.

Table 7. One-way ANOVA for Performance in Science

Source	Sum of Squares	df	Mean Square	F	Sig
Between groups	414	2	207.2	4.3	.014
Within groups	22263	461	48.3		
Total	22677	463			

The preliminary homogeneity of variance test at .05 level of significance showed that population variances were not equal. In Table 7, the one-way ANOVA for Science showed a statistically significant result, $F(2, 461) = 32.1, p = .014$.

The results of the Dunnett C multiple comparisons post hoc test indicated that at the .05 level of significance,

- (i) there was no significant difference in performance between the Random (M=15.4, SD=5.21) and Easy-to-Hard (M=13.5, SD=4.29) treatments;
- (ii) there was no significant difference in performance between the Random (M=15.4, SD=5.21) and Hard-to-Easy (M=16.7, SD=7.83) treatments;
- (iii) there was significant difference in performance between the Easy-to-Hard (M=13.5, SD=4.29) and Hard-to-Easy (M=16.7, SD=7.83) treatments. Performance was better in Hard-to-Easy order treatments.

Thus an answer to Research Question 3 is that the effect of a change in item order on candidates' performance at the BECE level was significant with regard to Science.

6. Discussion

These results generally disagree with the findings of researchers like Gerow [8], Allison [3], Soyemi [16], Perlini,

Lind and Mumbo [11] and Laffittee [9] who found no difference in performance when items were arranged according to a certain order of difficulty or randomly. The differences in the findings of these researchers and the findings of the present study may be due to the vast difference in the levels of education in which the researches were conducted. Whilst they used university and second cycle school students, this study used basic education school students.

The results of this study however agree with those of educational measurement experts like Shepard [14], MacNicol [10], Sax and Cromack [13], Skinner [15] and Ahuman and Clock [2] who asserted that changes in test format (or arrangement) can make a large difference in students' performance. The findings of this investigation also agree with that of a study by the Research Division of WAEC, Lagos [17]. They found significant differences in the performance of students in Agricultural Science, Biology, Economics, English Language and Mathematics at the second cycle level when items were re-ordered.

Since this study has shown that altering arrangement according to difficulty level does affect performance, then if performance is poor at any time the item order could be a contributory factor. Thus an answer to the research question "Would a change in item order affect candidates' performance in anyway?" would be in the affirmative. The present study has shown that there was indeed statistically significant difference in performance when the positions of the items were altered.

7. Conclusions and Recommendations

The main thrust of this study was to investigate the impact of change of item order in a multiple-choice test at the level of the Basic Education Certificate Examination. From the study, it was discovered that the change of item order impacts performance in English Language, Mathematics and Science.

The study has shown that the proposition of using re-ordering of format of a test to curb examination malpractice may not be the best especially in English Language, Mathematics and Science at the BECE level. Should an examining body decide to pursue the option of administering different versions of multiple-choice tests to curb the spate of collusion in examinations, it must do it with great caution. This is because in employing ordering method as in this study, some candidates could be seriously disadvantaged. Other methods such as equivalent or alternate forms of tests should be investigated for tackling the vice which is assuming alarming rates and sophistication with the introduction of mobile phones into the equation.

Learning institutions which depend heavily on multiple-choice tests for summative assessment should also take interest in the search for an appropriate method to curb examination collusion. They should support their staff to undertake research projects that would find solutions for the

irregularity.

Research units of examining bodies should not give up on the use of the re-ordering method but find appropriate adjustments to neutralize the impact of the difference in performance to facilitate the use of re-ordering of multiple-choice tests to arrest examination malpractices and increase the integrity of the certificates they issue.

Since the West African Examinations Council (WAEC) is a major examining body in Africa, it should champion further research into the use of different orders of a multiple-choice test with the view to finding appropriate adjustments which could neutralize the impact of the difference in performance. This should be done before WAEC embarks on the use of this method. In the event that appropriate adjustments are discovered, requisite training must be given to subject officers in the application of the method. For security reasons, setters may submit their items in a randomized order and subject officers should have the sole responsibility of applying the treatment to develop the different forms.

The research involved English Language, Mathematics and Science. Further study should be conducted using other subjects to give more knowledge and understanding for generalization of the findings at this level of assessment.

WAEC should consider commissioning a similar study into subjects of the West African Senior School Certificate Examination (WASSCE) to give the findings a universal application since the vice of collusion in examination is quite rampant at that level of education. It will be beneficial if the study could be conducted in each member country of WAEC and the results brought together to present a wider picture and offer a much better understanding of the problem.

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, A. (1976). *Psychological testing*, New York: Macmillan Press Ltd.
- [5] Bresnock, A. E., Graves, P. E., & White, N. (1989). Multiple-choice testing: question and response position. *Journal of Economic Education*, (Summer 1989), 239 – 244
- [6] 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.
- [7] 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.
- [8] Gerow, J. R. (1980). Performance on achievement tests as a function of the order of item difficulty. *Teaching of Psychology*, 7, 93 – 96.
- [9] Laffitte, R. G. (1984). Effects of item order on achievement test scores and students' perceptions of test difficulty. *Teaching of Psychology*, 11(4), 212 - 214.
- [10] MacNicol, K. (1956). Effects of varying order of item difficulty in an unspeeeded verbal test. Unpublished manuscript. Educational Testing Service, Princeton, New Jersey.
- [11] 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.
- [12] 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.
- [13] 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.
- [14] Shepard, L. A. (1997). The challenges of assessing young children appropriately. In Katheleen M. Cauley (12th ed.). *Educational Psychology*. Sheffield: Dubuque Inc.
- [15] 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.
- [16] Soyemi, M. O. (1980). Effect of item position on performance on multiple-choice tests. Unpublished M.Ed. dissertation, University of Jos.
- [17] WAEC (1993). The effects of item position on performance in multiple choice tests. Research Report, Research Division, WAEC, Lagos.
- [18] Worthen, B. R., & Spandel, V. (1991). Putting the standardized test debate in perspective. Association For Supervision And Curriculum Development, Educational Leadership (Feb. 1991 ed).