African Journal of Interdisciplinary Studies *Vol. 9, pp 55-62, December 2016, ISSN: 0855 – 9724*

Andrews Cobbinah
Department of Education and Psychology, University of Cape Coast- Cape Coast, Ghana. Tel: +233242576092
Email: andycobbina30@yahoo.com

Items' sequencing on difficulty level and students' achievement in mathematics test in Central Region of Ghana

Abstract

This study investigated items' sequencing on difficulty level and students' achievement in Mathematics test in Cape Coast Metropolis of Central Region, Ghana. An attempt was made to find out whether item difficulty index and item ordering of the W.A.E.C multiple choice test items has influence on Senior High Students achievement in Core Mathematics. The quasi-experimental design was adopted for the study. Purposive sampling technique was used to select six (6) schools comprised 250 form two students from six intact classes in the Central Region of Ghana. Forty Core Mathematics multiple choice items were adapted from W.A.E.C past questions. Item difficulty index was carried out; a one and a two way analysis of variances were used to test the first and the second hypotheses respectively. The reliability of the instrument was established by the use of Kuder-Richardson formula 20 and it was 0.78.The result established among others that the effect of change of item order on Senior High School students' achievement in Core Mathematics was significant. Based on these findings it was suggested that W.A.E.C. should consider only one format to be used in their examinations or equivalent/parallel test and again teachers and examination officers should be educated on the psychometric properties of test and the effect it has on item ordering

Key words*: Item ordering, student achievement, multiple –choice, item difficulty*

Introduction

Mathematics has long been thought of as a subject for only those with special talents. It is not surprising because Mathematics learning has been a problem to many over the world, not our times only but from the time immemorial. This is because the difficulty of Mathematics has been expressed in many forms of learning Mathematics. For example, there is no confidence in learning Mathematics, Mathematics is for a privilege few, and Mathematics as necessary evil. Such feelings or emotions associated with Mathematics learning difficulties contribute to a large extent to the poor participation and performance in Mathematics at higher levels by students with the more acute at higher levels Otuo Serebour (2013). But across nations, this attitude is now changing. Mathematics has been accepted to no longer be for few, but for all (Mcllracth & Huitt, 1995)

In Ghana, it is required of a student to pass in four core subjects, namely; English Language, Social Studies, Integrated Science and Core Mathematics in addition to two elective subjects to guarantee a Senior High School placement. Core Mathematics happens to be one of the core subjects and this has trigged the urge of student at the Basic school level to do everything possible to pass Mathematics. Many students find Mathematics a very difficult subject and not able to apply Mathematics ideas in situations of daily life. The question then is, is it not possible that teachers test practices could contribute to the students' achievements? A variety of tests are used in education but the use of multiple choice tests is the trend all over the world. Test plays an important role in giving feedback to stakeholders in education on various aspects of our life. Therefore the quality of test has always been a hot issue since long; consequently Gajja et al., (2014) indicated that multiple-choice questions are frequently used to assess students in different educational streams for their objectivity and wide reach of coverage in less time. However, the multiple-choice questions to be used must be of quality which depends upon its difficulty index, discrimination index and distracter efficiency. (Zaman et-al., 2009), opined that item difficulty is not affected by the sequence of items in a test. Zaman et al., (2009) have investigated the effect of item sequence in test on the difficulty level of an item by investigating the idea that discouraging result on preceding item may affect the performance of the student on the subsequent item. This study has concluded that there is no such effect on the item difficulty. This view was supported by Allison (1984) who also found no difference in performance when items were arranged according to a certain order of difficulty. Soyemi (1980), 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. However, Shepard (1994) asserted that tiny changes in test format (or arrangement) can make a large difference in students' performance. This view was supported by researchers in the research division of W.A.E.C; Lagos (1993) who discovered that different arrangement of items could affect performance adversely depending on the subject area.

A study by Moe (2011), on the topic 'gender differences do not mean genetic difference: Externalizing improvement performance in mental rotation' found that, male outperformed their female counterparts as a result of item difficulty which had to do with time allocation. Moe opined that when the time aloud is doubled females solved more items but the same was true for males. However when additional time is provided the extra items solved are the difficult ones and as difficulties increase males performed better than the female. Moe therefore concluded that what makes the difference is the belief that failures or difficulties are dependants on genetic factors.

In item analysis two approaches are widely used: Namely classical Test Theory and item response theory. The classical test theory utilizes two main statistics – item difficulty and discrimination. But of most important to this paper is the item difficulty, also referred to as the p-value which is the percentage of students who answered the item correctly . The range is from 0% to 100%, the higher the value, the easier the item. P-values above 0.90 are very easy items and might be a concept not worth testing. P-values below 0.20 indicate difficult items and should be reviewed for possible confusing language or the contents need re-instruction. Optimum difficulty level is 0.50 for maximum discrimination between high and low achievers. Test constructors, including classroom instructors, often begin a test with a few relatively easy items with the intent of minimizing the impact of examinees' test anxiety on their performance. Indeed, this practice is suggested in a number of textbooks on measurement. Specifying the difficulty of the first few items is a very simple approach to item sequencing' within a test. More complete specifications might include item sequences based on monotonically increasing difficulty (easy-to-hard), monotonically decreasing difficulty (hard-to-easy), "spiraling" difficulty (random assignment) of items to positions within the test. Duzel and Heinze (2002), investigated the extent to which item sequencing affects also in random-mixed design of recognition memory. Duzel and Heinze concluded that, in random-mixed design, brain activity can be affected by the sequence in which different types of items are presented.

Statement of the Problem

The over dependence on the use of objective tests in classroom practice has affected students learning and the form of subject- matter knowledge to be acquired by them. Indeed, factors like well stocked libraries, availability of adequate standard Mathematics text books, students and teacher ratio, number of periods on the time table and the like have found to be contributing factor to Mathematics achievement (Anamuah-Mensah, 2002) .It is still imperative for researchers to consider the contribution of item sequencing and difficulty indices which are also possible contributors to students' low achievement in Mathematics. However, questions have been raised regarding whether the order of test items influences student performances. Perhaps these arguments seem unending because, it is seen in most of our secondary schools that this approach of sequencing(e.g. easy-to- hard) is not adhered to and so many do not see the effect it has on students' performances. In University of Ilorin- Nigeria for instance, due to the increasing student population at the undergraduate level, most of their examinations are done using Computer Based Test (CBT) which involves the use of multiple test items. In this, test items are arranged randomly without sequencing either from easy-to-hard or hard –to- easy and the results from this type of examination is used for students' certification. This means that item sequencing is not a factor considered in this case, but others such as cheating and large student population seem to matter. Our limited knowledge of item sequencing on difficulty level of students' performance is a problem that should be addressed. Item analysis allows us to observe the characteristics of a particular item and can be used to ensure that questions are of an appropriate standard for inclusion in a test. There seemed to be conflicting results on the positioning of items in a test and the impact it has on students' achievement in Mathematics. Though available information suggests that there have been some findings on the positioning of items and their effect on students performance, this has not been conclusive and also not much has been done in senior high schools in the Central Region of Ghana. A comprehensive knowledge of the construction of a good test item and sequencing can enable us to create more effective test besides standardizing the existing tests.

Purpose of the Study

This study was aimed at investigating at items**'** sequencing on difficulty level and students' achievement in Mathematics in Cape Coast Metropolis of central region of Ghana. Two hypotheses were formulated.

Hypotheses

The following null hypotheses were formulated to guide the study;

1. $H_{O\,:}$ There is no significant effect of the mode of sequencing on the difficulty level of W.A.E.C multiple-choice test item and achievement of students in Core Mathematics.

2. H_0: There is no significant effect of the mode of sequencing on the difficulty level of W.A.E.C multiple choice test item and achievement in Core Mathematics of students with respect to gender

Methodology

Research Design

The quasi-experimental design was used in this study. This is a non-equivalence or nonrandomized experimental research Ary et al., (2002). According to Ary et al., quasi-experimental design is applicable in situation whereby subjects cannot completely assign to a group. This study adopted the design because school authority may not allow the already grouped classes to be disrupted as they will prefer research to be conducted on the subjects as they appeared in the natural form.

Population

The population of the study was made up of all public senior high schools students in the Cape Coast metropolis of the Central Region of Ghana. There are 10 public senior high schools in the Metropolis made up of 5 singled sex and 5 mixed schools. The average age of the students in this population is 16.6 years. The target population comprised only senior high school form two students.

Sample and Sampling Technique

Purposively sampling technique was used to select six schools two of which are girls and four boys schools. Purposive sampling because the schools had to meet certain criteria such as being at the same stage of completion as far as the coverage of the form two SSCE syllabus was concerned. Secondly, their performances in Mathematics were checked to be almost the same using their past Mathematics examination records. Through random sampling technique six intact classes were selected consisting of 250 students.

Instrumentation

The instrument for the study was a standardised test developed by WAEC which was adapted. Only the Multiple Choice Core Mathematics paper was used since it is the one for which the psychometric properties being studied could be obtained. To find the item sequence effect on difficulty level, this research work was driven by the theory that good and bad items behave differently. Classical approach was adopted due to its simplicity. To analyze the items the difficulty (D) level was judged using the guidelines suggested by Courville (2004), who has given the following four guidelines for the interpretations for D values. 1. If D = 0.40: no item revision necessary; 2. If D = 0.39: little to no item revision is needed; 3. If D = 0.29: item revision is necessary; and 4. If D = 0.19: either the item should be completely revised or eliminated. A test of 60 items was adapted by the researcher out of which 40 were used for data collection based on the difficulty analysis criteria. The researcher has 12 years' experience of teaching of Mathematics in public secondary level with involvement of six other relevant teachers. Bloom taxonomy was used as framework for test construction. Total items in the test were 40 with 10 items from knowledge domain, 10 from comprehension and 10 each from application and analysis domain. The content validity of the items had been established by the examination body (WAEC), while the reliability of the instrument was subjected to a trial test and found to be 0.78.

Data Collection Procedure

Three forms of the Multiple- Choice items were obtained as test formats A, B and C in order of random arrangement, easy-to-hard and hard-to-easy items respectively. The only treatment done was the development of three different forms of the test formats. The three forms of items were administered within the period of six weeks. Each school responded to all the three formats to ensure uniformity. In each intact class, students were randomly selected to response to different versions of the test formats.

Data Analysis Technique

Data was analysed by the use of a means, standard deviation, one -way ANOVA and two-way ANOVA. A preliminary test for homogeneity of variance was done to ensure if population variances were equal. The post hoc test used was Scheffe since homogeneity of variances assumption was not met.

Results

Data collected were used in answering the two hypotheses. The alpha value for failing to reject and rejecting the hypotheses were 0.01 and 0.05 for hypotheses 2 and 1 respectively. This was because the test of homogeneity of variances or the error among the scores was significant. In other words the test of homogeneity of variances was violated.

Hypothesis 1:

There is no significant effect of the mode of sequencing on the difficulty level of W.A.E.C multiple-choice test item and achievement of students in Core Mathematics..

The results for the analysis of scores for students' Mathematics performances are given in Table 1 and 2. Table 1 gives the descriptive statistics for the three levels of the item order.

Table 1: Descriptive Statistics for Performance of students in Core Mathematics

Order	N	Mean	Std. Deviation
Random (RDM)	250	65.2400	12.83281
Hard-to-Easy (HTE)	250	52.1960	12.86404
Easy-to-Hard (ETH)	250	59.7800	11.15911
Total	750	59.0720	13.40915

Table 1 represents the mean and the standard deviation of the scores for the three formats. The result indicates a better performance of students in the random formats, means score (65.24) followed by easy- to –hard, mean score (59.78) and the hard-to-easy having the lowest mean score (52.20).

Table 2 gives the results of the ANOVA analysis

Table 2: One -way Analysis of Variance for Students Core Mathematics Scores

Source	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	21456.216	2	10728.108	70.783	.000
Within Groups	113217.896	747	151.563		
Total	134674.112	749			

Since the test of homogeneity was violated, a robust test of equality of means was conducted which gave the Welch and Brown-Forsythe to be significant (i.e. 0.000). Hence the acceptance of the ANOVA results. In Table 2, the one –way ANOVA showed a significant results, F (2,747) = 70.78, P = 0.000.

The result of the Scheffe multiple comparison post hoc tests indicated that at 0.05 level of significance as presented in table 3.

Table 3: Scheffe Multiple Comparison Post –hoc test

	Format type	N	Subset for alpha = 0.05		
			1	2	3
Scheffe[a]	B	250	52.1960		
	C	250		59.7800	
	A	250			65.2400
	Sig.		1.000	1.000	1.000

(i) There was significant difference in performance between”spiraling”(random) order (M= 65.24, SD= 12.83) and monotonically increasing (easy-to-hard) order (M 59.78, SD= 11.16) treatments. Which is in favour of random order treatment?

(ii) There was significant difference in performance between random order (M= 65.24, SD= 12.83) and monotonically decreasing (hard-to-easy) order (M =52.20, SD= 12.86) treatments. Which is in favour of “spiraling” (random) order treatment?

(iii) There was significant difference in performance between monotonically decreasing (hard-to-easy) order (M= 52.20, SD= 12.86) and monotonically increasing (easy-to-hard) order (M 59.78, SD= 11.16) treatments. This is in favour of monotonically increasing (easy-to-hard) order treatment.

These results above showed that the effect of change of item order on Senior High School students’ achievement in Core Mathematics was significant.

Hypothesis 2: There is no significant effect of the mode of sequencing on the difficulty level of W.A.E.C multiple choice test item and achievement in Core Mathematics of students with respect to gender.

The results for the analysis of scores for students’ mathematics performances are given in Table 4 and 5.

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

Table 4: Descriptive statistics in terms of gender for the three levels of the item order.

Order	Gender	Mean	Std. Deviation	N
	Male	67.0076	12.86468	131
Random(RDM)	Female	63.2941	12.56532	119
	Total	65.2400	12.83281	250
	Male	53.3969	11.06601	131
Hard-to-Easy (HTE)	Female	50.8739	14.52325	119
	Total	52.1960	12.86404	250
	Male	60.0763	11.36230	131
Easy-to-Hard (ETH)	Female	59.4538	10.96972	119
	Total	59.7800	11.15911	250
	Male	60.1603	13.01029	393
Total	Female	57.8739	13.75379	357
	Total	59.0720	13.40915	750

Table 4 shows the descriptive statistics for the three main test formats with respect to gender. Males students had a mean score in the three test formats in this order of increment (RDM =67.0, ETH= 60.07 and HTE = 53.40 respectively) ; While the females had a mean score in this order of increment (RDM =63.29, ETH= 59.45 and HTE = 50.87 respectively). This result indicates that males' students had a mean gained over the females counterparts in all the formats with the highest mean gained registered in the RDM for both sex.

Again the Levene's test of equality of error variance was not significant (i.e $p<.05$), hence a more stringent significance level of 0.01 was used for the analysis of the results. Therefore the main and interaction p-value should be greater than 0.01. Table 5 gives the results of a two-way ANOVA analysis. From Table 5, a two-way analysis of variance was conducted. The interaction effect between gender and item order was significant, $F(2,744) = 1.007$, $p = 0.366$. There was a statistically significant main effect for gender, $F(2,744)= 6.500$, $p = 0.11$. These results showed that gender has role to play as far student achievement in Mathematics and the order of items are concern. It also indicated that male students generally performed better no matter the order of items than their female counterpart.

Table 5: A two- way analysis of variance for Students Core Mathematics Scores

Source	Type III Sum of Squares	Df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	22737.213[a]	5	4547.443	30.225	.000	.169
Intercept	2606247.340	1	2606247.340	17322.688	.000	.959
Item order	21335.674	2	10667.837	70.905	.000	.160
Gender	977.884	1	977.884	6.500	.011	.009
Item order*Gender	303.114	2	151.557	1.007	.366	.003
Error	111936.899	744	150.453			
Total	2751800.000	750				
Corrected Total	134674.112	749				

a. R Squared = .169 (Adjusted R Squared = .163)

Discussion

The results generally disagree with the findings of Zaman et al., (2009), they opined that item difficulty is not affected by the sequence of items in a test. Again Zaman, et al, (2009) have investigated the effect of item sequence in test on the difficulty level of an item by investigating the idea that discouraging results on preceding item may affect the performance of the student on the following item. Their study concluded that there is no such effect on the item difficulty. This view was supported Allison (1984) who also found no difference in performance when items were arranged according to a certain order of difficulty or randomness. Perhaps their finding may be different from this current study because of different environmental and also different educational levels. Instead this finding agrees with the following researchers Shepard and WAEC Research Division Lagos. Shepard (1994) asserted that tiny changes in test format (or arrangement) can make a large difference in students' performance. This view was supported by researchers in the research division of W.A.E.C; Lagos (1993) discovered that different arrangement of items could affect performance adversely depending on the subject area .It was also evidence that gender could affect students achievement in Mathematics when the items are arranged in different orders. As seen in the scores of students that male students consistently outperformed their female counterparts in the three formats of the test. This finding seems to agree with the finding by Moe (2011), on the topic 'gender differences do not mean genetic difference: Externalizing improvements performance in mental rotation' found that, male outperformed their female counterparts as results of item difficulty which has to do with time allocation. Moe opined that when the time aloud is doubled, females solve more items but the same is true for males. However when additional time is provided, the extra items solved are the difficult ones and as difficulties increase males performed better than the females. Moe therefore concluded that what makes the difference is the belief that failures or difficulties are dependant on genetic factors. Particularly when one looks at the performance of females in the easy-to hard formats of the test (Male, SD=10.97, Female, SD=11.16). There was a significant main effect of the order of item but not significant main effect on gender. This means that males and females differ in scores in terms of the different formats of the test, but not in the same order format. There was also an interaction effect of gender on the three forms of test and that choosing different versions of a test to be used gender should be considered.

Conclusion

The present study has indicated that there was statistically significant difference in the achievement of students when the items orders were change. Again the study found that on the same item format both sexes performed almost the same, but found out that males and females students perform differently on different forms of the test. The results of this study should be seen as an indication to teachers, item writers, subject officers, psychometrists and examination bodies that students' achievement in Mathematics could be enhanced positively by adopting the use of random arrangement, since it was seen as the best format where students' performances showed very good results. Re-ordering of format or different versions of the test to curtail the incidence of examination malpractice by the West African Examination Council should be pursued with ultimate care, since some students will be at disadvantage.

Recommendations

Based on the conclusion drawn from this study, the following recommendations were made:

(i) The W.A.E.C. should consider only one format to be used in their examinations or equivalent/parallel test

(ii) Teachers, examination officers should be educated on the psychometric properties of tests and the effects they have on item ordering

(iii) This finding should be adopted by stakeholders as it may solve the problem of comparability of standard, certification or quality faced by them.

(iv) The West African Examination Council and other examination bodies should champion a further research into the use of different orders of multiple-choice test in order to finding appropriate adjustment which will automatically adjust the impact of the difference in achievements when different test formats are used.

References

Allison, D. E. (1984). Test anxiety, stress and intelligence-test performance. *Measurement and Evaluation in Guidance, 16, 211-217*

Anamuah-Mensah Committee Report (2002). *Meeting the challenges of the Education in the 21st century, Review report of the President's Committee on Educational Reforms in Ghana*: Author

Ary, D. , Jacobs,L.C & Razavieh, A. (2002*). Introduction to Research in Education. (*6th Ed.).Indiana : Indiana University.

Courville, T.G., (2004). *An empirical comparison of item response theory and classical test theory item/person statistics*, Unpublished PhD thesis, Texas A&M University, Texas.

Duzel, E. & Heinze, H. J. (2002). The effect of item sequence on brain activity during recognition Memory. *Brain Res, Cognition Res, 13(1), 115 – 127*

Gajjar, S., Sharma, R., Kumar, P. & Rana, M. (2014). Item and test analysis to identify quality Multiple-choice questions for assessing medical students of Ahmadabad, Gujarat. *India J Community Med, 39(1), 17 – 20*

Soyemi, M. O. (1980). Effect of item position on performance on multiple –choice test Unpublished M.Ed. dissertation, University of Jos, Nigeria.

Mcllrath,D.& Huitt, W. (1995).The teaching- learning process. A discussion of models. *Educational Psychology Interative*. Valdosta, GA: Valdosta State University.

Moe, A. (2011). Gender differences does not mean genetic differences: Externalizing Improves performance in mental rotation. Retrieved on February 4, 2015 from http://wwwelsevier. com/locate.lindif

Otuo-Serebour, S. K. (2013). Factors, teachers, students and parents perceived as influencing Junior high school students' mathematics learning. Unpublished MPhil. Thesis, University of Cape Coast, Cape Coast.

WAEC (1993).*The effects of item position on performance in multiple choice test,* Research Report, Research Division, WAEC, Lagos

Zaman, A., Kayani, M.M. & Farooq, R.A. (2009). Developing and analyzing a multiple choice test in physics for Pakistani secondary schools, In Proceedings Western Australian Institute for Educational Research Forum 2009, www.waier.org.au/forums/2009