IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Examination of the Distracter Efficiency and Reliability Coefficient of Teacher Made Achievement Test of Senior High Schools in Ghana

BY
COBBINAH ANDREWS
UNIVERSITY OF CAPE COAST
DEPARTMENT OF EDUCATION AND PSYCHOLOGY
CAPE COAST- GHANA

&

BENJAMIN SARQUAH
BRAKWA SENIOR HIGH AND TECHNICAL SCHOOL
DEPARTMENT OF MATHEMATICS
ODOBEN DISTRICT- GHANA

Abstract

There have been many criticisms about the low performance of students in mathematics by the examinations conducted by the West African Secondary School Certificate Examination (WASSCE). Comparing the performance of students in mathematics in the district to their performance in the WASSCE leaves some discrepancies. This study, therefore, examined the distracter efficiency and reliability coefficient of teacher made achievement test of senior high schools in Ghana

. The research design adopted for the study was a descriptive survey. Purposive sampling was used to select the school for the study. The data was gathered from an adopted teacher-made test developed from the mathematics syllabus of the Ghana Education Service by the teachers. The fifty items were randomly sampled from the previous (2018-2021) end-of-term mathematics test of Brakwa Senior High and Technical School teachers. Kuder Richardson version 20 (KR20) statistics was used to check for the reliability of the study instrument and percentages for the functional distractors and non-functional distractors. The finding revealed that out of fifty (50) items, 38 were good, and 12 were poor as indicated by the distracter efficiency. It could be concluded from the findings that generally the distractor and reliability analyses result of the teacher-made test of Brakwa Senior High and Technical in mathematics was better. It was recommended that heads of the mathematics department of Brakwa Senior High and Technical School should organise more in-service training on item analysis for mathematics teachers to upgrade and improve their test construction skills.

Keywords: item psychometric property, Mathematics achievement, school-based assessment, teacher-made test

INTRODUCTION

Assessment plays a vital role in teaching and learning if students' successes are to be attained and improved. Globally, student learning assessments have become a major focus of educational literature in recent years (Shavelson, Carley & Webb,1990). Additionally, educational attainment and test results are sometimes used to validate the performance of the government, instructors, students, other educational stakeholders, and the worldwide community (Konadu,2015). Education systems worldwide use assessment systems for selecting their best and most deserving students for promotion to the next educational level and grade, school performance monitoring forms bases for resources allocation (Pongi,2004).

In Ghana, the link between good assessment practices and instructional improvement has been conducted by (Oduro-Okyire & Partey,2014). Quagrain and Arhin (2017), did some studies on the quality of items and instruments in the Central Region of Ghana which was an indication of values/importance placed on assessments by researchers.

In Ghana, the education system is such that, mathematics is given all the necessary importance in the curriculum and all educational policies from basic level to secondary level. Sa'ad, Adamu, and Sadiq (2014), explain that mathematics helps the individual acquire a basic computational skill that fosters desires and the ability to be accurate in problem-solving and also prepares the mind of the individual to appreciate and understand further mathematics.

Assessment plays a crucial role in education and other fields of studies which includes selection and placement of students (employees). However, for the result from the assessment to be acceptable, it must be valid and reliable. American Education Research Association (AERA), the American Psychological Association (APA), and the National Council on Measurements in Education (ACME) define validity as "the degree to which evidence and theory support the interpretation of test scores implied by suggested uses of the tests." (p. 9). Again, validity refers to the adequacy or suitability of the interpretation and use of the outcomes of student evaluations. According to Taale and Ngman-Wara (2015), reliability (dependability) refers to the degree to which an assessment produces consistent outcomes across identical periods, examiners, and contexts. In other words, it refers to the consistency with which a test or instrument measures whatever it measures throughout time.

It is generally believed that examination malpractices (Rajier, 2011) and wrong assessment instruments are due to poorly constructed items (Amadahe & Gyimah, 2016). They are among the ways that can affect the quality of the test results, which makes it the test maker difficult to judge the performance of students for decision making. Quaigrain and Arhin (2017) examined the relationship between test items, test quality, difficulty index (p-value), discrimination index, and distractor efficiency. The test's internal consistency reliability was 0.77 when KR 20 was used.

The average efficiency of a distractor was 55.04% (SD 24.09 %). In a study to determine the relationship between item difficulty index and item distracter effectiveness in a Single Best-Answer Stem Type Multiple Choice Questions (Chauhan, Girish, Bhoomika, Vasa & Rathod, 2015) used one hundred and twenty (120) students of the first year of Medical College in India. It was found out that items with two functioning distracters were more difficult than others, followed by items with three functioning distracters.

The literature review on the study of assessment of students in Ghana indicated that apart from the study of Quaigrain and Arhin (2017), that took into consideration the study of the analysis of test items and instruments, the rest of the studies were foreign-based and also which mostly focused on formative assessment practices and this establishes a gap in the literature that needs to be filled.

In addition, according to the school's minutes captured in 2019, for the past four academic years, students' performance in mathematics in the school-Based Assessment of Brakwa Senior High and Technical School has been high compared to their performance in West Africa Secondary School Certificate Examination (SBA's documents in the schools) The percentages of WASSCE result passes in mathematics for the Brakwa Senior High and Technical Schools showed inconsistency and low performances from the academic years of 2015/16 to 2018/19. In the 2015/16 academic year, out of seventy-five (75) students registered, only 16 (21%) passed with A1 to C6. In the subsequent academic year, 2016/17, the core

mathematics performance of students came down from 21% to 3.8% out of fifty-three (53) students registered only two students passed with A1 to C6

This implies that students' performance in mathematics in the SBA's was always higher than the external examination. This was evident in the student's report cards, class exercises and home assignments, and end-of-term minutes for the school. This situation led to parents and school authorities doubting the teacher's credibility of the school's SBA's This present study, therefore, seeks to investigate the psychometric properties (specifically the distractors analysis and the reliability) of teacher-made mathematics achievement tests of senior high schools in Asikuma Odoben district**.**

CLASSICAL TEST THEORY AND ITEM RESPONSE THEORY

The theories underpinning the study were the classical test theory and item response theory In classical test theory, a non-functioning distractor is defined as an option with either a response frequency of <5% or more. A positive discriminating power; on the other hand; a functioning distracter is an option with a response frequency greater or equal to five (5) % of the total number of examinees. The strength of distracter efficiency of an item depends on the functional distracter it has. Gajjar, Sharma, and Rana (2014) recommended that Distracter Efficiency (DE) is determined for each item based on several Non-Functional Distracters (NFD) and ranges from zero (0) to one (1). If an item contains three, two, one, or nil Non-Functional Distracters (NFD), then Distracter Efficiency (DE) will be 0%,33.3%,66.6, and 100%, respectively.

In Item Response Theory, also known as the Latent Trait Theory. The theory explains how examinees or individuals at different ability levels have performed on the item. Most of the application of the theory assumes that a single latent trait account for a response to an item on the test.

The test composer uses item response theory to postulate that responses to an item can be explained by a latent attribute that is less numerous than the test items. The hypothesis is based on a mathematical model of how examinees with varying skill levels for the attribute should respond to an item. This enables one to compare examinees who took separate tests and apply item analysis results to examinees with different ability levels than the group utilised for the item analysis. These theories are relevant to the study because to estimate the distracter effectiveness and the reliability of the items will depend on the person's ability to get an item correct.

RESEARCH QUESTIONS

The following research questions guided the study:

1. What was the distractor effectiveness of end-of-term mathematics examination items?
2. What was the reliability coefficient of end-of-term mathematics examination items?

MATERIALS AND METHODS

Research Design

A descriptive survey design was used for the study. The descriptive survey design is a scientific method which involves observing and describing the behaviour of a subject without influencing it in any way. It aims at documenting aspects of a situation as it naturally occurs (Amedahe & Gyimah, 2003). This design describes and documents the state and nature of core mathematics items and instruments used to judge students' performance. Descriptive statistics such as means, frequency, and percentages were used to analyse item distracter efficiency. Good distractor appeals to a higher proportion of low achieving students than high achieving students, resulting in negative statistics.

Population.

The study population included all students and teachers at Brakwa Senior High and Technical High Schools. As of 2019, the total population of teaching staff and students was 1,421 (1421) comprising sixty-five (65) teachers and one thousand three hundred and fifty-six (1356) students in this group

Sample and Sampling Procedure

. Purposive sampling was used to select the Brakwa Senior High and Technical School in one of the districts. Purposive sampling was again used to select all form two students for the study. This is because the form two students' mathematics test scores provide immediate feedback to the researcher in the study. In addition, the form two students were more experienced by way of having covered more content areas than the form one students. The sample size for the study consisted of all 229 students in the form two classes of Brakwa Senior High and Technical. The sampling of the two hundred and twenty-nine (229) students was based on the fact that the researchers wanted to involve the all students in form two studies in the study.

Instrumentation

The fifty items were randomly sampled from the previous end-of-term mathematics test of Brakwa Senior High and Technical School teachers. The items were sampled from the mathematics domain of statistics, variation, algebra, and percentages because these were the topics the teachers fully covered for the semester due to the covid-19 break. This test was an adopted teacher-made test developed from the core mathematics syllabus of the Ghana Education Service by the teachers.

The test items consisted of twenty-eight percent (28%) statistics, twenty-eight percent (28%) variation, twenty-two percent (22%) algebra, and twenty-two percent (22%) percentages which were guided by a table of specifications. The test items were designed to measure knowledge, comprehension, application, analysis, synthesis, and evaluation. The test was dichotomously scored, with the correct answer being one and zero for the wrong answer respectively.

Additionally, the data were subjected to content validity testing to verify how accurately the items were evaluated following the study's objectives, research questions, and variables of interest.

Under normal circumstances, it would have been appropriate to pilot test the study instrument on other senior high school students to establish its reliability. However, the instrument was not pilot tested in the study because the instruments consisted of teacher-made items already used to assess the students. Again, one of the objectives of the study was to assess the reliability of mathematics teacher-made tests, therefore, checking for the reliability of the instruments would be a duplication of effort and time-wasting

Data Collection Procedure

The researchers use two weeks for the data collection Ethical approval from the University of Cape Coast's Ethical Clearance Review Board was sought. In pursuance of ethical issues, the right to privacy, voluntary participation, no harm to participants, anonymity, and confidentiality were held highly esteemed.

Data Analysis Procedure

Research questions one and two sought to explore the distracter efficiency and reliability coefficient respectively of end-of-term mathematics test items. The researchers used percentages for the distracter efficiency and Kuder Richardson version twenty (KR20) to estimate the reliability coefficients of end of term mathematics assessment instrument. KR20 was used because the items were dichotomously scored.

RESULTS

The analysis was done based on the research questions below. The median score was 16. The median score was slightly less than the mean (17), implying that the test score is almost normally distributed.

Criteria for Determining Quality of Items Based on Distracter Efficiency and Reliability of Instruments.

According to (Gajjar, Sharma, & Rana, 2014), the strength of distracter efficiency of an item depends on the number of non-functional distracters. This means the lesser the number of non-functional distracters in an item, the greater the Distracter Efficiency. They recommended that Distracter Efficiency (DE) is determined for each item based on several Non-Functional Distracters (NFD). Distracter Efficiency ranges from zero (0) to one hundred (100), but in terms of probability ranges from zero (0) to one (1). If an item contains three, two, one, or nil Non-Functional Distracters (NFD), then the size of Distracter Efficiency

(DE) will correspond to 0%, 33.3%, 66.6, and 100%, respectively. Abel's (1965) classification of reliability coefficients, very high ≤ 1, High ≤ .8 Intermediate (Moderate) ≤ .6, Low≤ .4 Very low ≤ .2

Research Question One: What is the distractor effectiveness of end-of-term mathematics examination items?

This research question aimed to identify how effective the distractors of options of end-of-term core mathematics examination items of Brakwa Senior High and Technical School functions.

Table 1: ***Distracter Frequency as per distracter Effectiveness***

Distracter Effectiveness	**Frequency**	**Percentage**
≥ 5%N	111	73.51%
< 5%N	40	26.49%
Total	**151**	**100%**

Table 1 present the number of effective distracters against the non-effective distracters. The table indicates that 111options which constitute 73.51% had an effective distracter as they were options that were selected by at least 5% of the total number of examinees. Forty (40) items that constituted non-effective distracters assumed 26.49%. This means that those options were selected by less than 5% of the total number of examinees.

In all, out of two hundred (200) options of fifty end of Core mathematics items, there were 49 (24.5) items that had keys. Although it should have been fifty (50) keys, there was no key in item thirty-three, thus reducing the total number of keys to forty-nine (49). However, the total number of distracters was one hundred and fifty-one (151) as indicated in the table and constituted 75.5% of the two hundred (200) options

Table 2: ***Frequency Distribution Based on the number of Functioning Distractors***

Function Distracter per item	Frequency (Percentage)
0	1(2%)
1	11(22%)
2	15(30%)
3	23(46%)
Total	**50(100%)**

Mean = 2.25, S.D = 1.11

Table 2 indicates the number of functional distracters per item in the set of fifty items. Out of fifty (50) items, twenty-three items (46%) had three (3) functional distracters, meaning it contains zero (0) non-functional distracters. Therefore, Distracter Efficiency is 100% (Gajjar, Sharma & Kumar,2014).

Fifteen (15) items had two functional distracters, meaning it contains one non-functional distracter and therefore has a Distracter Efficiency of 66%. Eleven items (22%) contain one functional distracter, which means it had two non-functional distracters and therefore have a Distracter Efficiency of 33%. One item (2%) had zero distracter functioning, meaning it had three non-functional distracters. There is no key in all four options, and that item is item thirty-three. The eleven items (22%) that contain two non-functional distracters or had the Distracter Efficiency of 33% were items 2, 5, 8, 19, 22, 24, 30, 32, 36, 37, 39, 46 and 48. It means these items had weak options that were not attractive to the uninformed students. Item 33 had 0% distracter efficiency as it has a null functioning distracter.

Research Question Two: What is the reliability coefficient of end-of-term mathematics examination items?

This research question aims to estimate the reliability over time of end-of-term mathematics examination instruments of Brakwa Senior High and Technical school

Amedahe and Gyimah (2016) defined reliability as consistency of assessing results overtime on a population of individuals. Kuder Richardson version 20 was used to estimate for the reliability of the end of term mathematics instrument because it is the most suitable for Multiple-Choice Items. Kuder Richrdson version 20 is computed by formula,

$\mathbf{KR_{20}}=\frac{K}{K-1}\left(1-\frac{\Sigma pq}{\sigma^2 X}\right)$, where K is the number of items on the test, $\sigma^2 X$ is the total test variance, and Σpq is the total variance of item

Table 3: ***Analysis of Reliability Coefficient of the end of Mathematics an Instrument using the Kuder Richardson Version 20 Method.***

Item	Item Difficulty(P)	q=1-p	Item variance(pq)
1	0.53	0.47	0.2491
2	0.49	0.51	0.2499
3	0.45	0.55	0.2475
4	0.33	0.67	0.2211
5	0.18	0.82	0.1476
6	0.47	0.53	0.2491
7	0.28	0.72	0.2016
8	0.38	0.62	0.2356
9	0.34	0.66	0.2244
10	0.33	0.67	0.2211
11	0.49	0.51	0.2499
12	0.36	0.64	0.2304
13	0.31	0.69	0.2139
14	0.31	0.69	0.2139
15	0.29	0.71	0.2059
16	0.59	0.41	0.2419
17	0.57	0.43	0.2451
18	0.25	0.75	0.1875
19	0.29	0.71	0.2059
20	0.39	0.61	0.2319
21	0.42	0.58	0.2436
22	0.28	0.72	0.2016
23	0.17	0.83	0.1441
24	0.29	0.71	0.2059
25	0.32	0.68	0.2176
26	0.27	0.63	0.1971
27	0.20	0.80	0.1600
28	0.40	0.60	0.2400
29	0.42	0.58	0.2436
30	0.26	0.74	0.1924
31	0.48	0.59	0.2419
32	0.38	0.62	0.2356
33	*	*	*
34	0.21	0.79	0.2059
35	0.35	0.65	0.2275
36	0.39	0.61	0.2379
37	0.29	0.71	0.2059
38	0.41	0.59	0.2419
39	0.33	0.67	0.2211
40	0.45	0.55	0.2475

41	0.25	0.75	0.1875
42	0.24	0.76	0.1924
43	0.2	0.8	0.1600
44	0.26	0.74	0.1924
45	0.17	0.83	0.1411
46	0.19	0.81	0.1539
47	0.24	0.76	0.1824
48	0.12	0.88	0.1056
49	0.15	0.85	0.1275
50	0.38	0.62	0.2356
Total			Equal

ẍ(Mean)=$\frac{\Sigma fx}{\Sigma f} = \frac{3920}{50} = \mathbf{17.12}$

$\sigma^2 X$(Total score variance) $\boldsymbol{d} = \frac{7585.269}{228}$=**33.269.**

Calculation parameters:

K=50, Mean=17.12, Σpq=11.611 and $\sigma^2 X = 33.269$

$\mathbf{KR_{20}}=\frac{K}{K-1}\left(1-\frac{\Sigma pq}{\sigma^2 X}\right)$

$\mathbf{KR_{20}}=\frac{50}{50-1}\left(1-\frac{11.611}{33.262}\right) = \mathbf{0.664}.$

Therefore, the reliability coefficient of the end of core mathematics instrument using the Kuder Richardson version 20 method is 0.664. which is moderate.

DISCUSSION

The various estimation of item distractor parameters and reliability coefficient of the mathematics instrument by the results of the study are discussed below: The results of research question one indicates that out of two hundred options,111options which constitute 73.51%, had an effective distracter as they were options with response frequency of or greater than 5% of the respondents, forty (40) items which constituted non-effective distracters assumed 26.49%. This means there were more functioning distractors than non-functioning ones.

For the results of several functional distracters per item, the majority of the items, 23 (46%), had three (3) functional distracters, meaning it contains zero (0) non-functional distracters; therefore, Distracter Efficiency is 100%.

The present findings corroborate with Uddin and Mullick (2020) concerning the number of functional distracters per item with three functioning distracters (which implies that these items have a distracter efficiency of 100%). However, in this study majority of the items, twenty-three (46%), had three (3) functional distracters per item, which means it had a distracter efficiency of 100%.

The study found the mean item distracter efficiency of 2.25 and a standard deviation of 1.11 which did not support the previous study by Quaigrain and Arhin (2017). They found out that the mean distracter efficiency was 55.04 and a standard deviation of 24.09. This means that distracters in this current study were more homogeneous and more attractive to the un-informed students than the current study. Regarding the total number of functional distracters against the non-functional distracters, this study showed more quality. For instance, in this study, 111 options which constitute 73.51%, had an effective distracter against forty (40) items that constituted non-effective distracters constituting 26.49%. Mehta and Mokhasi (2014) found a total of fifty items that had 150 distractors. Among these, 53(35.3%) were non-functional distractors, 38(18.6%) were functional distracters. Comparing Mehta and Mokhasi's findings to the current one, it becomes clear

that the current findings have more had more effective distracters than theirs. These differences in the findings probably may be due to the nature of the items and the students 'characteristics which were not the same.

The results of research question two also showed that the reliability coefficient of the end of term mathematics test of Brakwa Senior High and Technical School using the Kuder Richardson version 20 method was 0. 664. This means the end-of-term mathematics instrument of Brakwa Senior High and Technical School is highly reliable.

The study of Quaigrain and Arhin (2017) results on the reliability of the assessment instrument were highly reliable. However, the reliability coefficient of the previous study (0.77) is higher than this current study (0.644) using the same (Kuder Richardson version 20 method) of estimating the reliability coefficient. This difference in the reliability coefficients may be due to different sample sizes in both studies, the quality of the items as well as the differences in students' characteristics.

CONCLUSIONS

It could be concluded from the findings that generally the distractor and reliability analyses result of the teacher-made test of Brakwa Senior High and Technical in mathematics was moderate which is better. Notwithstanding, efforts need to be put in place for teachers to study the structure of the WASSCE test items so they can construct their test items to match the WASSCE standard to reduce the failure rate of students in the WASSCE mathematics examination. Though the test items were of moderate quality they were not probably constructed to the standard of the WASSCE test items hence the discrepancies between the SBA and the WASSCE results. One limitation of the study was the test administration challenges in the sense that the study made use of all the form two students. Mobilising all the form two students to administer a test was difficult because most teachers were not prepared to sacrifice their period for this exercise.

RECOMMENDATIONS

1. The Asikuma Odoben Brakwa District Educational Directorate should train core mathematics teachers on item analysis and construction skills to improve their item construction skills.
2. Head of the mathematics department of Brakwa Senior High and Technical School should organise more in-service training on item analysis for mathematics teachers to upgrade and improve their test construction skills. to enable them to develop test items of high distractor efficiency.
3. Teachers should deliberately study the structure of the WASSCE test items so they can construct their test items comparable to the WASSCE one.
4. Though the reliability coefficient of the mathematics test items was high, efforts should be made by the teachers to maintain and also improve on it.

REFERENCES

Akyina, O. k, & George Oduro-Kyire, G. (2019). Formative assessment practices of senior high school teachers in Ashanti Region Municipality of Ghana. *British Journals of Education*,*3*(7),27-38

Amedahe, F. k, & Asamoah- Gyimah, K. (2016). *Introduction to Measurements and Evaluation*. Cape Coast. Hampton Press,

Amedahe, F. K., & Asamoah, E. G. (2003). *Introduction to educational research.* Accra: Paramount Press.

Donald, G. (1990). *Personal management*: *Theory and practice*. London: D. P. Publication Ltd.

Ebel, R.L. (1965). Measuring Educational Achievement. Englewood Cliffs. NJ:

Gajjar, S., Sharma, R., Kumar, P., & Rana, M. (2014). Item and test analysis to identify quality multiple-choice questions (MCQs) from an assessment of medical students of Ahmedabad, Gujarat. *Indian journal of community medicine: official publication of Indian Association of Preventive & Social Medicine*, *39*(1), 17-28

Konadu, K. (2016). *Assessment of teaching and learning conditions in basic schools in Offinso North District* Unpublished Doctoral dissertation. Department of Education and Psychology, University of Cape Coast

Mehta, G., & Mokhasi, V. (2014). Item analysis of the multiple-choice questions-an assessment of the assessment tools. *Int J Health Sci Res*, *4*(7), 197-202.

Oduro-Okyire, G & Partey, D. N. L., (2014). The effect of assessment on classroom learning in the Ashanti Mampong Municipality of the Ashanti Region of Ghana.*The International Journals of Humanities and Social Studies*,*2*(7),150-155.

Ponji, V. (2004*). Marking the switch from assessment for ranking towards assessment for learning challenges facing the small Island State of the Pacific*. Paper presented at the Third Conference of Commonwealth Examination and Accreditation Bodies (ACEAB),Nadi, Fiji.

Quaigrain, K., & Arhin, A. K. (2017). Using reliability and item analysis to evaluate a teacher-developed test in educational measurement and evaluation. *Cogent Education*, *4*(1), 103-113.

Ragier, J. (2011). Why is academic success important? Applied Science and *Technology Scholarship. Saskatchewan: Social Board Association).*

Sa'ad, U. T., Adamu, A., & Sadiq, A. M. (2014). The causes of poor performance in mathematics among public senior secondary school students in Azare Metropolis of Bauchi State, Nigeria. *IOSR Journal of Research & Method in Education*, *4*(6), 32-40.

Shavelson, R., Carley., N &Webb (1990). *Indicators of science achievement: Option for a powerful policy instrument*. Phi Delta Kappan,*71*(9),692- 697.

Taale, K. D (2015) & Ngman-Wara, E (2015). Assessment in Basic School. University of Education Winneba. University Press.

Uddin, M. K., & Mullick, T. T. (2020). Evaluation of Multiple-Choice Questions in Secondary School Certificate Examination in Bangladesh. *International Journal*, *1*(1), 9-19.

.