OPEN ACCESS
Journal of Advances in Education and Philosophy
Abbreviated Key Title: J Adv Educ Philos
ISSN 2523-2665 (Print) |ISSN 2523-2223 (Online)
Scholars Middle East Publishers, Dubai, United Arab Emirates
Journal homepage: https://saudijournals.com

Original Research Article

Item Difficulty as a Source of Variability in Student Achievement in the West African Secondary School Certificate Examination (WASSCE): Application of Generalizability Theory

Andrews Cobbinah[1], Ruth Keziah Annan-Brew[1], Frank Quansah[2*]

[1]Department of Education and Psychology, University of Cape Coast, Cape Coast, Ghana
[2]Department of Educational Foundations, University of Education, Winneba, Ghana

DOI: 10.36348/jaep.2022.v06i03.001 | **Received:** 18.01.2022 | **Accepted:** 21.02.2022 | **Published:** 06.03.2022

***Corresponding author:** Frank Quansah
Department of Educational Foundations, University of Education, Winneba, Ghana

Abstract

Students' achievement in core subjects in the West African Secondary School Certificate Examination (WASSCE) has been a subject of concern for stakeholders due to poor performance. The purpose of this study is to assess whether item difficulty is a significant source of variability to the measurement of students' achievement in WASSCE in Ghana, using the generalizability theory. The study had three specific objectives: (1) to examine the sources of variability (e.g., item) to students' achievement in WASSCE in Ghana, (2) to assess the dependability of students' responses in relation to their ability, (3) evaluate how many items are sufficient to provide an optimum measure of student achievement. The one-facet crossed random design was adopted as the research design. Data were obtained from students' achievement in 2015 WASSCE in the four subjects, namely, Mathematics, Science, Social Studies and English Language. The generalized analysis of variance (GENOVA) was used for the analysis, conducting both G-study and D-study. The analysis revealed that the findings from this study showed that item difficulty had minimal effect on the variability in students' achievement in English Language and Mathematics. The result further revealed that item difficulty had a significant effect on variability in students' achievement in Science and Social Studies. Only the English Language test showed relatively low reliability. The study recommended the item difficulty and content structure of the Science and Social Studies multiple-choice test should be evaluated by WAEC officials and their examiners.

Keywords: Item difficulty, student achievement, generalizability theory, English Language, Mathematics, Science, Social Studies.

INTRODUCTION

Performance of students in the West African Secondary School Certificate Examination (WASSCE) (West African Examination Council [WAEC], 2014, 2015, 2016) and Basic Education Certificate Examination (BECE) (Adamoah & Acquah, 2016; Amuzu, Ankalibazuk, & Abdulai, 2017; Nugba, Quansah, Ankomah, Tsey, & Ankoma-Sey, 2021; Mills & Mereku, 2016) in Ghana has been a concern for stakeholders in education. Particularly for the Senior High School (SHS) students, the poor performance has grave consequences on the student's life in terms of having access to a tertiary institution; this marks an important point of their lives in building a career for themselves (Ansah, 2017). While exploring the factors accounting for the poor performance of students in high stakes examinations has been on the agenda of the government of Ghana and its stakeholders, the search has led to inconclusive findings and observations. This is to say that the possible causes of low student achievement in Ghana are varied and include teacher and student absenteeism, lateness on the part of both students and teachers, insufficient teaching and learning resources, poor parental involvement in school activities, poverty, teaching experiences of teachers, teacher subject matter knowledge, teacher pedagogical skills, among others (see Ankoma-Sey, Asamoah, Quansah, & Aheto, 2019; Ansah, Quansah, & Nugba, 2020; Amaah, Ato-Davies, Kpeyibor, Semordey, & Oppong, 2019; Fletcher, 2018; Gyan, Mabefam, & Baffoe, 2014; Nugba, 2021; Mohammed, Asare-Bediako, & Aboagye, 2016; Quansah, 2021). With all these factors, one thing is common; that is, the probable factors identified as the possible causes of poor students' performance are largely external (i.e., factors

outside the examination context) and do not include issues with the test/item itself.

Statistical test theory has stipulated that the quality of a test or an item plays a significant role in the outcome of the test (Lord & Novick, 1968). The item difficulty, item discrimination, and distractor function are critical when developing a multiple-choice test (Nitko, 2004). Item difficulty- which is defined as the proportion of examinees who had an item right- has been extensively discussed in the literature as a pointer to the variability in students' academic performance (Huck & Bower, 1972; Nitko, 2004). Despite this, the relationship between item difficulty and academic achievement is still inclusive. Whereas several studies have found no link between item difficulty and academic performance (e.g., Di-Battista *et al*., 2009; Skinner, 1999; Vander Schee, 2013), other scholars have established that there is a link between item difficulty and students' academic achievement (e.g., Brenner, 1964; Caldwell & Pate, 2013; MacNicol, 1960; Hambleton & Traub, 1974). Based on the latter scholars' view, there appears to be a wide range of factors that affect item difficulty and consequently, the performance of the test takers. Prominent among these factors include ambiguity, misalignment between what was taught and test content, time allocated for the test, items not matching the class-level of the examinees, not using age-appropriate words in item development, unclear distractors, unfamiliar item context, among others (Nitko, 2004).

The purpose of this study is to assess whether item difficulty is a significant source of variability to the measurement of students' achievement in WASSCE in Ghana, using the generalizability theory. The generalizability theory is a psychometric theory that provides a conceptual view for assessing the dependability of scores obtained from a test (Cronbach, Rajaratnam, & Gleser, 1963; Shavelson, Webb, & Rowley, 1989). This study is based on the premise that the variability in students' achievement in WASSCE should not be explained by any other factor except for the ability of the examinee (Huck & Bower, 1972). Therefore, the general achievement of students on WASSCE would be confounded if the item difficulty parameter contributes a significant measurement error to the achievement construct. The study, which operated within the framework of generalizability theory, posited that the difficulty of an item contributed to the students' performance in three specific ways: (1) the sources of variability (e.g., item) in students' achievement in WASSCE in Ghana, (2) the dependability of students' responses in relation to their ability, (3) the number of items sufficient to provide an optimum measure of student achievement.

This research is the first of its kind in Ghana, using generalizability theory to understand the role of item difficulty in explaining the sources of variability in students' achievement in WASSCE core subjects. Previous studies which attempted to investigate the effect of item difficulty on the item and test performance generally used experimental design and such studies did not use tests that were high stakes (Brenner, 1964; Caldwell & Pate, 2013; Di-Battista *et al*., 2009; Huck & Bower, 1972). Due to the experimental design used by the earlier studies, the scholars ended up with small sample sizes and focused on a single subject area. This study used secondary data on student scores in WASSCE for the four core subjects written by the examinees. The outcome of the study provides insight into the role of item difficulty in explaining student achievement in WASSCE, to inform appropriate policies and practices of the West African Examination Council (WAEC) and all relevant stakeholders in education.

Research Questions:

The following research questions were formulated to guide the study

1. What are the sources of variability (e.g., item) in students' achievement in WASSCE in Ghana?
2. What is the dependability of students' responses in relation to their ability?
3. How many items are sufficient to provide an optimum measure of student achievement?

METHODS AND MATERIALS

Design

The research employed the one-facet crossed random design (Brennan, 2001). The single facet was the item. The person (student) facet was not considered as a major facet in the design (Shavelson & Webb, 1991). The major sources of variability were person (i.e., students serving as the measurement object) symbolized as *p*, and items labelled as *i*. Therefore, the one-facet crossed design was symbolized by *p x i*. The interaction source of variability was also symbolised by *p x i*, *e* (person-by-item interaction/residual). The expected mean square for the sources of variations is shown in T:

Table 1: ANOVA Table for a Person-by-Item Design

Source of variation	Sum of square	df	Mean Square	Expected Mean Square
Person (p)	SS_p	n_p-1	$MS_p = SS_p / df_p$	$n_p\sigma_p^2 + \sigma_{pi,e}^2$
Item (i)	SS_i	n_i-1	$MS_i = SS_i / df_i$	$n_p\sigma_i^2 + \sigma_{pi,e}^2$
pi,e	$SS_{pi,e}$	$(n_p-1)(n_i-1)$	$MS_{pi,e} = SS_{pi,e} / df_{pi,e}$	$\sigma_{pi,e}^2$

Data Source and Participants

Data were obtained from the WAEC on students' responses to the multiple-choice section of the four papers, namely, Mathematics, Science, Social Studies and English Language. The Mathematics, Science, and Social Studies subjects had 50 multiple-choice items each with 2566, 2487, and 2519 students respectively sitting for the papers. The Englsih Language test had 80-items which was administered to 2257 students. The data covered senior high school students who sat for the 2015 WASSCE. These four subjects were chosen because they form the requirement for the progression of students to the next level of their academic life. The data was obtained after ethical clearance has been granted by the Institutional Review Board (IRB), University of Cape Coast, Ghana.

Data Analysis

The data were processed in the R-studio environment and analysed using the Generalized Analysis of Variance (GENOVA) within the generalizability theory framework. Both the generalizability study (G-study) and decision study (D-study) were performed. The G-study involved the computations on the mean squares and variances components, providing much information about the sources of variability (Shavelson & Webb, 1991). This research design yielded three variance components: (1) variances due to person, (2) variability due to item difficulty, and (3) variations due to item-by-person, residuals. The estimates from the G-study were used for the D-study to estimate the reliability coefficients and their associated errors as well as optimizations (Sudweeks *et al*., 2005). The reliability coefficients were interpreted based on Creswell (2012) who indicated that for a reliability coefficient to be judged as sufficient, it should be greater than or equal to .84, especially when such coefficient is used to make high stake decision.

RESULTS

Sources of variability (e.g., item) in students' achievement in WASSCE in Ghana

The study explored the sources of variability in students' achievement in WASSCE in Ghana. Two sources of variability were considered, namely, items and residuals (i.e., items interacting with other variables, and other unexplained factors). The person (i.e., student) facet was not considered as a facet but carefully guided the dispersion of scores. Hence, emphasis was not placed on the person variability. Table 2 presents the details of the results.

Table 2: Sources of Variability in Students' Achievement in WASSCE in Ghana

Subject	Sources of variability	df	Sum of square	Variance	Percent
Mathematics	Person (*p*)	2565	4990	0.035	14.6
	Item (*i*)	49	2912	0.023	9.6
	p x i, e	125685	22934	0.182	75.8
Science	Person (*p*)	2486	2041	0.016	33.5
	Item (*i*)	49	1501	0.012	25.7
	p x i, e	12181	23821	0.020	40.8
Social Studies	Person (*p*)	2518	1162	0.009	28.3
	Item (*i*)	49	650	0.005	16.8
	p x i, e	123382	21230	0.017	54.9
English Language	Person (*p*)	2256	2174	0.009	4.3
	Item (*i*)	79	1410	0.008	3.5
	p x i, e	17822	36395	0.204	92.2

The results shown in Table 2 revealed that variability of students' responses to the multiple-choice items was largely explained by the residuals (i.e., person-by-item interaction, plus other systematic and unsystematic factors). For the Mathematics subject, the residual explained about 75.8% in the variations in students' achievement (Table 2). About 54.9% of the variability in Social Studies were explained by some unknown systematic and unsystematic variables. Similarly, these systematic and unsystematic factors accounted for 92.2% of the variability in students' achievement in the English Language. It appears that it was only Science subject which the residuals explained less than 50% of the variances in the students' achievement (40.8%).

Whereas item difficulty explained very little of the variances in students' achievement in WASSCE in some subjects, the story was different for other subjects. Taking the English Language subject, for instance, item difficulty accounted for 3.5% in the variation in students' achievement in the subject, followed by Mathematics subject with item difficulty accounting for 9.6% variances in students' achievement. For Social Studies and Science subjects, the variations in students' achievement explained by item difficulty were quite high. While item difficulty explained about 16.8% achievement variability in Social Studies, item difficulty in Science subject accounted for 25.7% variability in students' achievement.

Dependability of students' responses to the items in relation to their ability

The reliability coefficients of the multiple-choice test in their current forms (i.e., 50-items for Mathematics, Science, Social studies, and 80-items for the English Language) were examined. Table 3 presents the detailed results.

Table 3: Reliability Coefficients of the Test

Subject	*g*-coefficient	Relative error	Phi-coefficient	Absolute error
Mathematics	0.91	0.060	0.90	0.064
Science	0.98	0.020	0.96	0.026
Social Studies	0.96	0.019	0.95	0.021
English Language	0.79	0.051	0.78	0.051

The results from the analysis showed that the Science test with 50-items yielded the highest reliability (r_g=0.98, r_{phi}=0.96) compared to the other tests (Table 3). This was followed by the Social Studies, with 50-items, which also had a reliability coefficient greater than 0.90 (r_g=0.96, r_{phi}=0.95). Similarly, the English Language test with 80-items also yielded an appreciable level of reliability (r_g=0.91, r_{phi}=0.90). The reliability of the English Language test (r_g=0.79, r_{phi}=0.78) did not show sufficient or optimum reliability, however.

How many items are sufficient to provide an optimum measure of student achievement

Optimization was further conducted to understand how many items should constitute a test that can adequately measure students' achievement in the subject area. The details of the results are shown in Table 4.

Table 4: Optimizations

Mathematics	**G-study**		**Option 1**		**Option 2**		**Option 3**		**Option 4**		**Option 5**	
	Lev.	**Univ.**	**Lev.**	**Univ.**	**Lev.**	**Univ.**	**Lev.**	**Univ.**	**Lev.**	**Univ.**	**Lev.**	**Univ.**
P	2566	INF	2566	INF	2566	INF	2566	INF	2566	INF	2566	INF
I	50	INF	20	INF	30	INF	40	INF	60	INF	70	INF
g	0.91		0.79		0.85		0.89		0.92		0.93	
phi	0.90		0.77		0.84		0.87		0.91		0.92	
Science	G-study		Option 1		Option 2		Option 3		Option 4		Option 5	
	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.
P	2566	INF	2566	INF	2566	INF	2566	INF	2566	INF	2566	INF
I	50	INF	20	INF	30	INF	40	INF	60	INF	70	INF
g	0.98		0.94		0.96		0.97		0.98		0.98	
phi	0.96		0.91		0.94		0.95		0.97		0.97	
Social Studies	G-study		Option 1		Option 2		Option 3		Option 4		Option 5	
	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.
P	2519	INF	2566	INF	2566	INF	2566	INF	2566	INF	2566	INF
I	50	INF	20	INF	30	INF	40	INF	60	INF	70	INF
g	0.96		0.91		0.94		0.95		0.97		0.97	
phi	0.95		0.89		0.92		0.94		0.96		0.97	
English	G-study		Option 1		Option 2		Option 3		Option 4		Option 5	
	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.	Lev.	Univ.
P	2257	INF	2566	INF	2566	INF	2566	INF	2566	INF	2566	INF
I	80	INF	100	INF	50	INF	60	INF	70	INF	90	INF
g	0.79		0.84		0.70		0.74		0.76		0.81	
phi	0.78		0.84		0.69		0.73		0.76		0.80	

The results outlined in Table 4 revealed that the 50-item test in Mathematics, Science, and Social Studies yielded an optimum level of reliability in students' achievement. The reliability estimates for both relative and absolute parameters were greater than 0.50. With 50-items, Mathematics had a reliability estimate of .90, Science had .98, and Social Studies had .96. Although optimization with items more than 50 yielded a slightly higher level of reliability, it is not prudent to use these number of items for three related reasons: (1) the reliability estimate of .90 and above is adequate based on recommendation from literature (Creswell, 2012), (2) the difference between the reliability estimates of a 50-item test and a test with more than 50 items is very little, (3) considering the little change in reliability estimates, it will not be economical (e.g., paying item developers for more items, using more resources in printing the items, etc) to go for more items. In the case of the English Language, using the current form with 80 items could not yield sufficient

reliability. Based on the optimization, a 100 item English Language test would yield a test with a sufficient level of reliability.

DISCUSSION

The findings from this study showed that item difficulty had minimal effect on the variability in students' achievement in English Language and Mathematics. For these two subjects, the results further revealed that variability in students' achievement was largely explained by some systematic and unsystematic factors which were not explored in this study. Indeed, previous studies in Ghana have stressed that poor performance in subjects like Mathematics is due to poor teacher pedagogy, subject matter knowledge, negative attitude towards teaching and learning of the subject, teaching experiences, and low student-textbook ratio (Ankoma-Sey *et al*., 2019; Nugba *et al*., 2021; Ansah *et al*., 2020; Quansah, 2021). The findings of this study that factors other than item difficulty in English Language test contributed greatly to students' achievement, is also reflected in a study by Mosha (2014) which revealed that factors such as unqualified English teachers, large class size, poverty, and limited support from home played a significant role in students' achievement in the subject. The challenges confronting the teaching of Mathematics and the English Language is real and as such, it is not surprising that other factors outside the test itself contributed to the variances in students' achievement.

The result, additionally, revealed that item difficulty had a significant effect on variability in students' achievement in Science and Social Studies. This effect was much greater in Science achievement as compared to Social Studies achievement. This implies that, although other systematic and unsystematic variables played a role in the variability in students' achievement in these two subjects, the difficulty of the items contributed significantly to errors in the measurement of the construct. The result suggests that some of the items in these subjects may be above their ability and not because the students were taught and did not learn. This result is supported by the research conducted by Tatar, Tüysüz, Tosun, and İlhan (2016) which cited the Science curriculum as the major cause of poor achievement in the subject, indicating that the science curriculum may be quite difficult or has been structured in a way which makes the subject a difficult one. Also, Boakye and Ampiah (2017) intimated that some teachers were not able to complete the Science syllabus and, as a result, students may find it difficult to answer questions in the areas which were not covered by the teachers.

The findings also showed that the English Language test had low reliability, indicating that the measurement errors were quite high. Further inspection of the results revealed that the original 80-item test was not enough to yield optimum dependability in terms of measuring the ability (i.e., true score) of the test takers. This may be explained by the fact that the 80 items were not enough to cover other aspects of the English Language ability of the examinees. This is supported by the various reliability theory where the number of items plays a crucial role in determining the extent of reliability of the measurement (Nitko, 2004). Thus, increasing the number of items would be a great way to improve the reliability of the estimates (Nitko & Brookhart, 2014; Quansah, 2017). In an optimization test, it was found that between 100 test items in English Language would yield optimum reliability greater than .80. This also has implications for WEC examiners regarding the test content coverage; the test content needs to be expanded to cover other areas of English Language skills and competencies.

CONCLUSION AND RECOMMENDATIONS

This study highlighted the role of item difficulty in explaining the variability in students' achievement in Mathematics, Science, English Language, and Social Studies subjects in WASSCE in Ghana. The outcome of the research showed that item difficulty played little role in students' variations in achievements in Mathematics and English Language subjects. The trend of results for Science and Social Studies subjects were quite different, with item difficulty accounting for relatively moderate to high variances in students' achievements. For instance, item difficulty accounted for more than one-fourth of the total variability in students' achievement in Science. In as much as, item difficulty is very tough to eliminate entirely from a test with about 50-items, its effect can be reduced to the barest minimum. The results have implications on the review of item content and structure, especially in Science subject, by experts to ensure that items match the ability of examinees. This can be done by conducting pilot testing in schools, covering a wider geographical location.

The English Language items yielded a relatively low level of reliability, and this was explained by the fact that the items were not adequate to fully capture students' proficiencies in the subject. It is recommended to the WAEC and their examiners to increase the number of multiple-choice items for the English Language paper to 100. It is further suggested that the item structure of Mathematics, Social Studies, and Science should be maintained by WAEC. This notwithstanding, several studies of this nature need to be conducted to validate this present research. This study was carried out using a single year group data (i.e., 2016) and thus, generalizing the findings to items in other years may be limited, especially when different items are used for different year groups. It is suggested that a similar study should be conducted to cover several years, say, 10 years. Also, future studies should conduct similar research using some selected elective subjects (like Physics, Biology, Elective Mathematics,

among others) to also understand how item difficulty functions in such subjects.

CONSENT

As per international standard or university standard, respondents' written consent has been collected and preserved by the authors. Ethical issues such as confidentiality, anonymity, data de-identification, data validation, among others, were strictly observed in the conduct of this study

ETHICAL APPROVAL

Ethical approval was obtained from the Institutional Review Board (IRB) of the University of Cape Coast, Cape Coast, Ghana.

AUTHOR CONTRIBUTION

AC and RKA conceptualized the idea. FQ performed the analyses. All authors prepared, read and approved the final manuscript.

Competing Interest: Authors have declared that no competing interests exist.

REFERENCES

- Adamoah, Y. K., & Acquah, J. (2016). Determinants of students' performance in basic education certificate examination (BECE) in the Upper East Region of Ghana: A case study of Kassena-Nankana West District (KNWD). *American Journal of Research Communication, 4*(10), 91- 107.
- Amaah, E. O., Ato-Davies, A., Kpeyibor, P. F., Semordey, C. R., & Oppong, J. D. (2019). poor academic performance of students in Ghana: Appraisal of Wamfie Circuit junior high schools in Dormaa East District. *Education Journal, 2*, 67-76. DOI: 10.31058/j.edu.2019.24001
- Amuzu, S., Ankalibazuk, A., & Abdulai, S. I. (2017). Low performance of pupils in BECE: A case study of Sagnarigu District in Northern Region, Ghana. *International Journal of Advance Research in Science, Engineering and Technology, 4*(7), 4176-4184.
- Ankoma-Sey, V. R., Asamoah, D., Quansah, F., & Aheto, S. K. (2019). Factors affecting junior high school pupils' performance in mathematics in Cape Coast Metropolis, Ghana. *Staff and Educational Development International, 24*(3), 120-129.
- Ansah A. (2017). *A descriptive assessment of higher education access, participation, equity, and disparity in Ghana*. Retrieved from http://sgo.sagepub.com/conten
- Ansah, J. K., Quansah, F., & Nugba, R. M. (2020). 'Mathematics Achievement in Crisis': Modelling the influence of teacher knowledge and experience in senior high schools in Ghana. *Open Education Studies, 2*, 265-276. https://doi.org/10.1515/edu-2020-0129
- Boakye, C., & Ampiah, J. G. (2017). Challenges and Solutions: The Experiences of Newly Qualified Science Teachers. *SAGE Open*. https://doi.org/10.1177/2158244017706710
- Brennan, R. L. (2001b). *Generalizability theory*. New York: Springer-Verlag.
- Brenner, M. H. (1964). Test difficulty, reliability, and discrimination as functions of item difficulty order. *Journal of Applied Psychology*, *48*, 98–100.
- Caldwell, D. J., & Pate, A. N. (2013). Effects of question formats on student and item performance. *American Journal of Pharmaceutical Education, 77*(4), 1-5. DOI: https://doi.org/10.5688/ajpe77471
- Creswell, J. W. (2012). *Educational research: Planning, conducting and evaluating quantitative and qualitative research* (4th ed.). Boston, MA: Pearson Education, Inc.
- Cronbach, L. J., Rajaratnam, N., & Gleser, G. C. (1963). Theory of generalizability: A liberalization of reliability theory. *British Journal of Statistical Psychology, 16*(2), 137–163.
- Fletcher, J. (2018). Performance in Mathematics and Science in basic schools in Ghana. *Academic Discourse: An International Journal, 10*(1), 1-18.
- Gyan, C., Mabefam, M. G., & Baffoe, M. (2014). Push out or drop out? Taking a critical look at the poor and drop-out of students of the JSS/JHS programme in Ghana. *Academic Journal of Interdisciplinary Studies, 3*(1), 409-420.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamental of item response theory*. Newbury Park, CA: Sage.
- Huck, S. W., & Bowers, N. D. (1972). Item difficulty level and sequence effects in multiple-choice achievement tests. *Journal of Educational Measurement*, *9*(2), 105-111.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores.* Reading, Massachusetts: Addison Wesley.
- MacNicol, K. (1960). Effects of varying order of test item difficulty in an unspeeded verbal test. Unpublished manuscript, Educational Testing Service, Princeton, NJ.
- Mills, E. D., & Mereku, D. K. (2016). Students' performance on the Ghanaian junior high school mathematics national minimum standards in the Effutu Municipality. *African Journal of Educational Studies in Mathematics and Sciences, 12*, 25-34.
- Mohammed, F., Asare-Bediako, M., & Aboagye, P. (2016). Factors contributing to the low academic achievement of students at the basic education certificate examination in some public junior high schools of a political district in the Western Region of Ghana: Students' and teachers' perspective. *International Journal of Mathematics and Statistics Studies, 4*(3), 1-10.

- Mosha, M. A. (2014). Factors affecting students' performance in English Language in Zanzibar rural and urban secondary schools. *Journal of Education and Practice, 5*(35), 64-76. https://www.iiste.org/Journals/index.php/JEP/article/viewFile/17455/17714
- Nitko, A. J. (2004). *Educational Assessment of Students* (4th ed). Ohio: Merrill Prentice Hall
- Nitko, J. A., & Brookhart, S. M. (2014). *Educational assessment of students* (6th ed.). Harlow: Pearson Education Limited.
- Nugba, R. M. (2021). The status of educational input factors in junior high schools in the Central Region of Ghana. *American Journal of Education and Learning, 6*(1), 50-63.
- Nugba, R. M., Quansah, F., Ankomah, F., Tsey, E. E., & Ankoma-Sey, V. R. (2021). A trend analysis of junior high school pupils' performance in the Basic Education Certificate Examination (BECE) in Ghana. *International Journal of Elementary Education, 10*(3), 79-86. doi: 10.11648/j.ijeedu.20211003.15
- Quansah, F. (2017). The use of Cronbach Alpha reliability estimate in research among students in public universities in Ghana. *African Journal of Teacher Education, 6*(1), 56-64.
- Quansah, F. (2021). Teaching and Learning of Mathematics in Lower Primary in Schools in Ghana. In *Developing Mathematical Literacy in the Context of the Fourth Industrial Revolution* (pp. 136-154). IGI Global.
- Shavelson, R. J., & Webb, N. M. (1991). *Generalizability theory: A primer*. Newbury Park, CA: Sage.
- Shavelson, R. J., Webb, N. M., & Rowley, G. L. (1989). Generalizability theory. *American Psychologist, 44*(6), 922–932.
- Skinner, N. F. (1999). When the going gets tough, the tough get going: Effects of order of item difficulty on multiple choice test performance. *The North American Journal of Psychology*, *1*, 79–82.
- Sudweeks, R. R., Reeve, S., & Bradshaw, W. S. (2005). A comparison of generalizability theory and many-facet Rasch measurement in an analysis of college sophomore writing. *Assessing Writing, 9*, 239-261.
- Tatar, E., Tüysüz, C., Tosun, C., & İlhan, N. (2016). Investigation of factors affecting students' science achievement According to student science teachers. *International Journal of Instruction, 9*(2), 153-166. DOI: 10.12973/iji.2016.9211a
- Vander Schee, B. A. (2013). Test item order, level of difficulty, and student performance in marketing education. *Journal of Education for Business, 88*(1), 36-42.
- West African Examinations Council. (2014). *Chief examiners report for 2014*. Retrieved from https://www.waecgh.org/Portals/0/PDF/General%20Resume2014.pdf
- West African Examinations Council. (2015). *Chief examiners report for 2015*. Retrieved from https://www.waecgh.org/Portals/0/PDF/General%20Resume%20W15.pdf20Resume2014.pdf
- West African Examinations Council. (2016). *Chief examiners report for 2016*. Retrieved from https://www.waecgh.org/Portals/0/PDF/General%20Resume%20W16.pdf