A Meta-Analytic Assessment of Standard Fixing at Senior Secondary School Certificate Examination by West African Examinations Council, Nigeria

Fausiyah Adenike Zubayr[1] and Abdul-Wahab Ibrahim[2]

[1] *Department of Educational Foundations and Counselling, Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria*
[2] *Department of Education, Faculty of Education, Sule Lamido University, Kafin Hausa, Jigawa State, Nigeria*

Abstract: This study analysed the empirical differences in standard setting procedures in public examinations and determined the differences in the standard setting procedures used for core Senior School Certificate Examination (SSCE) subjects of Mathematics, English Language and Biology by West African Examinations Council (WAEC). It determined the differences in the general standards of SSCE core subjects and in the proportion of the grades in relation to the effect size. These were with a view to evolving a scientific stable procedure in public examinations. The ex-post facto research design was adopted for the study. The sample consisted of 30 secondary data of empirical studies purposively selected using the internet and hand searching of journals, articles and papers presented at the WAEC monthly and annual seminars. Data collected was analysed using the One-Way Analysis of Variance (ANOVA) statistical method. The result showed that a statistically significant difference existed in the standard setting procedures used for core subjects of SSCE. Also, there was a significant difference in the standard setting procedures used for the core SSCE subjects of Mathematics, English Language and Biology with regard to student performance in public examinations. Further, there was a significant difference in the standards between SSCE core subjects in the proportions of the grades in relation to the effect size. The study concluded that the Stanine grading system currently in use for reporting the SSCE is adequate for the purpose of identifying candidates and level of their attainment in SSCE subjects.

Keywords: Standard setting procedures, meta analytic method, cut-off scores, grade awards, norm-referenced procedures

1. INTRODUCTION

In Nigeria, evaluation agencies, which also act as examining bodies, are tasked with maintaining a common standard in the development and administration of public examinations. According to Faleye and Afolabi (2005), evaluation agencies were set up to promote education, to co-ordinate educational programmes, and to control and monitor the quality of education in educational institutions, the essence of which is the organization of public examinations so as to provide uniform standards to all test takers, irrespective of the type or method of instruction they have received. Some of these examination bodies in Nigeria include the West African Examinations Council (WAEC), the National Examinations Council (NECO), the Joint Admission and Matriculation Board (JAMB), and the National Business and Technical Examinations Board (NABTEB). A closer look at the operations of these boards reveals that some of them perform similar functions. WAEC, NECO and NABTEB, for instance, all conduct secondary school

graduate certification, although in the case of NABTEB, the examination is intended for graduates of Nigerian Technical and Vocational Colleges.

The assemblage of subject examinations conducted by these examining bodies is known as the Senior Secondary School Certificate Examination (SSCE) and serves as an end-of-course evaluation for all secondary school graduates. The purpose of this examination is to ascertain to what degree students in a particular course have achieved the course or educational objectives (Offor, 2001). In view of the economic and social importance attached to senior secondary school certificates, and the opportunities for higher education for those who posses such certificates, the awarding of this certificate is one of the most important events in the Nigerian academic calendar. Thus, much is expected from certificate examining and awarding bodies in terms of ensuring that the spirit and focus of the examinations is not misplaced.

E-mail address: zufausiyah@gmail.com, *edutestsexpert2013@gmail.com*

On standard fixing at SSCE by WAEC the first consideration is to note that subjects for SSCE have different features, each having its own realm of meaning and forms of knowledge distinct from others. The SSCE is a public examination and from every indication a norm-referenced test. Zubayr (2014) has asserted that most public examinations (including SSCE) are norm-referenced tests and that the raw scores from such examinations are almost meaningless. Usually therefore, raw scores need further treatment before they are ready for consumption. This treatment is usually referred to as standardization. Furthermore, the 9 numerical scales (A1 to F9) which is used to report performance in the SSCE has been in use for a very long time and was used for reporting the school certificate and GCE 'O' Level examinations prior 1988. The use of this scale for SSCE is seems adequate more so in terms of value as there is not much difference between the GCE ordinary Level and SSCE except for the beefing up of the contents of the syllabus. The SSCE performs the same function as GCE 'O' Level; for example, preparation for world of work and tertiary education (Zubayr, 2014).

According to Kalgo (2005), the current practice of awarding grade shows some inconsistencies and followed no laid down principle. The approach followed either a norm or criterion reference principle. The minimum boundary score for each grade and the percentage of students falling into each group varied from year to year and subject; the boundary scores for each grade is also observed to vary by subject. Ukwuegbu's (2004) assertion about standard and grade award in the SSCE confirms Kalgo (2005), when he opined that WAEC appears to be using yesterday's tool for today's work as the reporting format for maintaining standard in examination scores seems to be falling out of place.

Likewise, Adeyegbe and Daramola (2004) reported a no significant difference between some preceding subject grade boundaries so as to need a sub-classification of the grades. For example, there was consistently insignificant difference between grades A1 and B2 in almost all the subjects considered. It was also found that grade B3 did not belong to A group; there was no difference between B3 and C4, and sometimes, C5 and C6 belonged to the same group. The study also observed that determination of subject grades was highly subjective and unstatistical because most of the required statistics of performance were hardly produced for award meetings. It could be argued that grades awarded in a manner reported upon above could lead to wrong placement of students and create in their minds a wrong impression of their results and ability in the subjects. When such students are placed in a course which demands high intellectual abilities, they may not be able to cope. The public then starts to point accusing fingers, while in reality; the examinations wrongly graded or measured their abilities. The simplest

conclusion to a lay man is that there is a fall in standard of education. As Hirst and Peters (1970) cited by Zubayr (2014) have tried to explain, achievement in one discipline must be recognised as radically different from those in any other. This is because within a subject, the objective to be attained and knowledge to be acquired are such radical differences which are neither equitable with nor reducible to that of any other discipline.

In spite of these discrepancies in the results generated by the standard procedure by WAEC, and the relevance of the application of norm referencing in the grading process, Ukwuegbu (2004) had suggested that criteria referencing should be used for grading the SSCE, on the strength that it is more objective and could withstand the test of time better than the norm referenced procedure currently in use. It should be noted nonetheless that the essence of WAEC's activity as an examining body is to ascertain whether curriculum objectives derivable from the syllabus are being or have been achieved after the students have gone through it. There are two major ways to achieve this purpose viz assessing through norm-referencing and criterion-referencing. The former has been in practice since inception and is widely adopted largely by many examination bodies. The latter is equally useful assessment. However, one has to state all the criteria to be measured at the onset and a value placed on each criterion. Furthermore, the measuring instruments i.e. test or examination question paper must ensure that each of the criteria is taken care of. It therefore means that the teaching/learning situation must accommodate the attainment of the criteria set before testing. If a situation where such could be met presents itself, then criterion referencing and consequently fixing of grades/marks for different performance before the examination is attempted can be possible. Even then it is hardly possible to fix the same criteria for different subjects, in which case the issue of different marks for grades is still likely to prevail across subject areas. As long as it is to identify or label, then if enough effort is made to scrutinize the raw marks to ensure standard, norm referencing, the type in practice or criterion referencing which may be suggested as an alternative will be applicable for achievement testing (Ukwuegbu, 2004).

However, Akeju (2001) had actually warned that any model that presumes a given standard of candidates to be in a given range for a fluid educational population as in Anglophone West Africa is logically invalid. Yet the model on which WAEC's scale is built is being violated still apart from the above in that the critical grades of B2, C6 and E8 are fixed at award meetings, without any reference to standardized raw scores corresponding to standardized scores of 75, 60 and 30. It is not clear whether this is clone before the grades on which candidate performance are based are awarded. This was perhaps why Yoloye et al., (2001) asserted that WAEC

used certain formulae for deciding the mark ranges corresponding to the various grades but the officials by and large did not know the basis of the formulae.

Although Soriyan (2002) had given some indications of the statistical bases of the formulae, his claims are somehow at variance with information obtained on it by Yoloye et al., (2001) from the United Kingdom (U.K.) examination boards. Yet, the establishment of a concrete standard base for scales is necessary for the purpose of grading. Thus, the WAEC's nine point scale grades seem to be doing this. It would however, appear that it does not have a standard setting basis and hence a need to have a rethink about its continued use. It is pertinent to note that even the U.K. bodies from which the numerical grade was inherited had since 1975 abandoned it for letter grades which are based on certain criteria as outlined by Yoloye et al., (2001), hence his study. Contextually, meta-analysis is a collection of systematic techniques for resolving apparent contradictions in research findings. Meta-analysts translate results from different studies to a common metric and statistically explore relations between study characteristics and findings. When the treatment effect or effect size is consistent from one study to another, meta-analysis can be used to identify this common effect. When the effect varies from one study to another, it may be used to identify the reason for the variation (Hambleton and Pitoniak, 2006).

According to Ferdous & Plake (2007), meta-analysis can also be seen as a statistical technique in which the results of two or more studies are combined in order to improve the reliability of the results. The validity of a hypothesis cannot be based on the results of a single study, because results typically, vary from one study to another. Rather, a mechanism is needed to synthesize data across studies. Narrative reviews had been used for this purpose, but the narrative review is largely subjective different experts can come to different conclusion and becomes impossibly difficult when there are more than a few studies involved Meta-analysis, contrast, applies objective formulas and can be used with any number of studies. Therefore, a meta-analysis also allows us to make the best use of all the information that we have gathered in our systematic review by increasing the power of the analysis. By statistically combining the results of similar studies, we can improve the precision of our estimates of treatment effect, and assess whether treatment effects are similar in similar situations. The decision about whether or not the results of individual studies are similar enough to be combined in a meta-analysis is essential to the validity of the result (Cizek, 2006). Therefore, this study was conducted to investigate empirical differences in standard setting procedures on some WAEC examination questions focusing on core subjects of mathematics, English language and Biology. Specifically, it was conducted to determine the difference

in the standard setting procedures used for core SSCE subjects of Mathematics, English Language and Biology; and examine the difference in the standards between SSCE core subjects in the proportion of the grades in relation to the effect size. To achieve the objectives of the study, the following research question and research hypotheses were raised:

Research Question

1. What is the difference in the standard setting procedures used for core subjects of SSCE Mathematics, English Language, and Biology?

Research Hypotheses

1. There is no significant difference in the standard setting procedures used for core subjects of SSCE with regard to student performance in public examinations.

2. There is no significant difference in the standards between SSCE core subjects in the proportion of the grades in relation to the effect size.

2. METHODOLOGY

The study adopted the ex-post-facto research design. According to Cooper and Schindler (2001), ex-post-facto research design seeks to find out factors that are associated with certain occurrences, outcomes, conditions or type of behaviours by analysis of past events or of already existing conditions. Given the importance of standard setting procedures and the potential for scrutiny on theoretical and psychometric grounds, meta-analytic procedures was used to determine the effects of different procedural modifications of the cut scores/cut off scores method as used by WAEC on relevant outcomes of the judgment process. Therefore, using meta-analytic techniques, the researcher provided a quantitative synthesis of past research that evaluated the impact of common standard setting procedural modifications of the cut scores /cut off scores method since the goal of the study was to evaluate the systematic effects of these procedural modifications, alone and in combination, on the cutoff scores that resulted from the procedure and the degree of consensus among the judges on what that cutoff score should be. Thus, in this study a meta-analytic method was used to investigate standards between and within core subjects in Ordinary Level subjects of the SSCE.

All studies on SSCE Ordinary Level subjects comprised the target population of this study. The sample for the study was 30 studies based on core subjects in Ordinary Level subjects of the SSCE. The necessity of conducting a study examining the issue of standard setting procedures in the SSCE Mathematics, English Language and Biology is best understood when one takes

into account the fact that every year the test is taken by more than one million senior secondary school class 3(SS III) students and the inferences made on the basis of the test grades as a measure of standard by WAEC are a crucial factor in determining the admission of the students in the University Matriculation Examination (UME) programme to gain admission into the University in the country.

The sample procedure was judgmental and choosing a study depended on whether the study reported significance level of its results or whether it is possible to convert the statistics used by primary researcher into appropriate effect sizes. The study was summarised in terms of author and year, sample size, statistical methods used, level of significance research coverage and indications for comparison. All research studies used were published in the last ten years. Studies on standard fixing and grade award exercises at the Ordinary Level subjects of the SSCE by WAEC was collected from many sources within the published professional literature, commissioned projects, and papers presented at the WAEC monthly and annual seminars. The assistance of the Coordinator of WAEC Research Division, Lagos, Nigeria, was sought for and obtained to get easy access to relevant WAEC published papers and commissioned projects especially on standard settings since inception of SSCE in 1988.The 30 studies used were chosen on the basis of author and year, sample size, statistical methods used, level of significance, research coverage and research prediction. The data collected were analysed by using Hunter and Schmidts (1990) psychometric meta-analytic method. Statistical techniques such as descriptive statistics like means, standard deviation, as well as One-Way Analysis of Variance (ANOVA) were used to test the hypotheses postulated in this study. All hypotheses were tested at 0.05 level of significance.

3. RESULTS

Research Question: What is the difference in the standard setting procedures used for core subjects of SSCE Mathematics, English Language, and Biology? In order to resolve this research question, the marks for the critical grades from subject to subject and year to year were first obtained before subjecting these scores to grade points and baseline performance levels in each of the core subjects of SSCE Mathematics, English Language, and Biology for the 2011 to 2013 examination periods. The results are presented in Table 1.

Table 1: Reversed/Modified Grade Point and Baseline Levels in the Selected Subjects 2011-2013

Years of examination	Subjects	Maximum Marks Obtainable	2011 Stanine Grades									2012 Stanine Grades									2013 Stanine Grades								
			1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9	1	2	3	4	5	6	7	8	9
	Biology	200	151	140	117	110	102	87	74	61	0	153	141	116	108	100	83	72	60	0	152	139	114	105	97	80	69	58	0
	Mathematics	150	119	110	93	87	81	70	54	38	0	111	102	88	82	77	67	52	37	0	115	107	91	85	80	69	54	39	0
	English Language	200	149	140	123	117	111	100	90	80	0	149	140	123	117	111	100	90	80	0	150	140	121	114	108	95	85	75	0

From Table 1, it can be seen that there was a variation in the scores that approximated the grades of the different subjects. For instance grade A1 in 2011 were as follows 151 (75.5%) for Biology, 119 (79.33%) for Mathematics, and 149 (74.5%) for English Language. The variation assumed a similar trend for the other grades in 2012. Similar variation levels were also recorded for 2013 across the three subjects. However, variation in the scores that approximated the grades were relatively stable for each subject across the years for instance, A1 had the following baseline marks for Biology: 151 in 2011, 153 in 2012 and 152 in 2013. The stability in the scores was recorded for the other grades as well. In particular, the baseline performance levels for English Language was an outlier for grade point B2 which recorded140 in 2011, 140 in 2012 and 140 in 2013.

Consequently, the marks range of the nine grades for each of the three selected subjects were compared across the years 2011 to 2013 and the extent of variation from the stipulated deviation of ± 3 for fixing the grades of the SSCE were computed using a deviation index. The results are presented in Tables 2, 3, and 4, for Biology, Mathematics, and English Language respectively.

Table 2: Marks Range and Deviation Indices for Biology 2011-2013

Stanine	Marks Range			Deviation Indices %
Grades	**2011**	**2012**	**2013**	
1	151-200	153-200	152-200	0
2	140-150	141-152	139-151	0
3	117-139	116-140	114-138	0
4	110-116	108-115	105-113	66.7
5	102-1-9	100-107	97-104	66.7
6	87-101	83-99	80-96	133.3
7	74-86	72-82	69-79	66.7
8	61-73	60-71	58-68	0
9	0-60	0-59	0-57	0

Table 3: Marks Range and Deviation Indices for Mathematics 2011-2013

Stanine	Marks Range			Deviation Indices %
Grades	**2011**	**2012**	**2013**	
1	119-150	111-150	115-150	33.3
2	110-118	103-110	107-114	133.3
3	93-109	88-102	91-106	66.7
4	87-92	82-87	85-90	66.7
5	81-86	77-81	80-84	33.3
6	70-80	67-76	69-79	0
7	54-69	53-66	54-68	0
8	38-53	37-51	38-53	0
9	0-37	0-36	0-37	0

Table 4: Marks Range and Deviation Index for English Language (2011-2013)

Stanine	Marks Range			Deviation Indices %
Grades	**2011**	**2012**	**2013**	
1	149-200	149-200	150-200	0
2	140-148	140-148	14-149	0
3	123-139	123-139	121-139	0
4	117-122	117-122	114-120	0
5	111-116	111-116	108-113	0
6	110-110	110-110	90-107	66.7
7	90-99	90-99	85-94	66.7
8	80-89	80-89	75-84	66.7
9	0-79	0-79	0-74	0

It is apparent from Table 2 that the marks ranges were not stable across the three years especially for the C4, C5, C6 and D7 grades. The recorded deviation indices were 66.7, 133.3 and 66.7 for the C4, C5, C6 and D7 grades respectively.

Table 3 shows that the marks deviation for Mathematics was remarkable for the B2 grade where it recorded a deviation index of 133.3. Apart from an index of 66.7 for each of the B3and C4 grades and an index of 33.3 each for grades A1 and C5, the marks for the other grades were very stable. Marks stability in mathematics

is very important in that it is one of the three compulsory subjects for every candidate in the SSCE. The situation for English Language, the second compulsory subject is presented in Table 4.

From Table 4, English Language enjoyed a relatively high level of marks stability especially for grades A1 and C5. The few deviations recorded for the subject were in grades C6 to E8 with a deviation index of 66.7 each. This implies that there is a statistically significant difference in the standard setting procedures used for core subjects of Senior School Certificate Examination (SSCE) Mathematics, English Language, and Biology.

Research Hypotheses

Hypothesis One: There is no significant difference in the standard setting procedures used for core subjects of Senior School Certificate Examination (SSCE) with regard to student performance in public examinations. The hypothesis on the significant difference in the standard setting procedures used for core subjects of Senior School Certificate Examination (SSCE) of Mathematics, English and Biology to be explained by student performance in public examinations was tested using analysis of variance (ANOVA). The results are presented in Table 5.

Table 5: One-Way ANOVA on Difference in the Standards Setting Procedure used for Subjects with regard to Student Performance in Public Examinations.

ANOVA					
Model	**Sum of Squares**	**Df**	**Mean Square**	**F**	**Sig.**
Between Groups	228.05	2	114.03	9.760*	.000
Within Groups	315.44	27	11.683		
Total	543.49	29			

*Significant, p<0.05

Table 6: Proportion of Grades 1 & 2, 3-6, 7-9 for years 2011-2013

Biology	Grades 1 & 2	Grades 3- 6	Grades 7-9
2011	1.09%	29.53%	71.47%
2012	0.10%	21.46%	78.54%
2013	0.04%	13.22%	86.71%
Expected % on Stanine Reversed stanine (WAEC)	6%	64%	30%
Expected % on WAEC Scale Reversed stanine	5%	65%	30%

Table 7: Proportion of Grades 1&2, 3-6, 7-9 for years 2011-2013

Mathematics	Grades 1 & 2	Grades 3- 6	Grades 7-9
2011	1.60%	23.20%	76.80%
2012	0.62%	13.65%	86.35%
2013	0.96%	18.14%	81.86%
Expected % on Stanine Reversed stanine WAEC	6%	64%	30%
Expected % on WAEC Scale	5%	65%	30%

The analysis of variance results in Table 5 show that the calculated *F-value* of 9.760 is greater than the critical value in the *F-table* (3.27) at 0.05 percent level of significance. *The F-value* was obtained by dividing the explained mean square (114.03) and residual mean squares (11.683) which is significant at 5% significant level. In this case, the null hypothesis is not sustained. Therefore, the null hypothesis is rejected while the alternative hypothesis is accepted. Hence, this implies that there is a statistically significant difference in the standard setting procedures used for core subjects of Senior School Certificate Examination (SSCE) Mathematics, English Language, and Biology with regard to student achievement/performance in public examinations.

Hypothesis Two: There is no significant difference in the standards between SSCE core subjects in the proportion of the grades in relation to the effect size. To test this hypothesis, the West African Examinations Council (WAEC) model was adopted which fixes the percentage of the candidates' population in each of the grades 1-9. Hence, expected percentages in grades 1-9 for years 2011-2013 SSCE in the core subjects were used to determine the proportion of grades in relation to the effect size. The results are presented in Tables 6, 7 and 8 for Biology, Mathematics and English Language respectively.

Table 6 presents the proportion of grades in relation to the effect size. It could be seen from Table 10 that the expected % of proportion of grades by candidates on WAEC's Scale is for grades 1-9 is equal to that of Stanine respectively. However, from the Table, the proportion of grades 1-9 earned by candidates from 2011-2013 deviated from the Stanine and also from the WAEC's laid down expected grading system. For instance, in WAEC's system, about 5% of the candidates

should score between A1 and B2, about 65% should score between B3 and C6, while about 30% should score between D7 and F9. But surprisingly, this expectation was not attained in any of the core SSCE subjects studied for years 2011-2013. For Biology, 1% out of expected 5% scored A1 & B2 in 2011 as compared to less than 1% of candidates scored A1 and B2 in 2012 and 2013 respectively. This pattern was followed by 29%, 21%, and 13% who obtained B3 and C6, out of the WAEC's expected 65%. Further, more grades 7-9, were obtained by candidates as against WAEC's fixed 30% grading system across the years 2011-2013. This implies that the proportion of grades 7-9 obtained over years are

statistically significant than the proportion of grades 1-6 due to effect size.

Table 7 shows that, for Mathematics, about 2% out of expected 5% scored A1 & B2 in 2011 as compared to less than 1% of candidates who scored A1 and B2 in 2012 and 2013 respectively. This pattern was followed by 23%, 13%, and 18% of candidates who obtained B3 and C6, out of the WAEC's expected 65%. Further, more grades 7-9, were obtained by candidates as against WAEC's fixed 30% grading system across the years 2011-2013.

Table 8: Proportion of Grades 1&2, 3-6, 7-9 for yeas 2011-2013

English Language	Grades 1 & 2	Grades 3- 6	Grades 7-9
2011	0.10%	13.3%	86.7%
2012	0.24%	17.02%	81.09%
2013	0.08%	16.58%	83.42%
Expected % on Stanine	6%	64%	30%
Expected % on WAEC Scale	5%	65%	30%

Table 9: One-Way ANOVA on Difference in the Standards Between SSCE core Subjects in the Proportion of the Grades in Relation to the Effect Size

ANOVA					
Model	Sum of Squares	Df	Mean Square	F	Sig.
Between Groups	182.6	2	91.30	6.367*	.000
Within Groups	387.1	27	14.34		
Total	569.7	29			

*Significant, p<0.05

Table 8 shows that, for English Language, less than 1% out of expected 5% scored A1 & B2 in 2011, 2012 and 2013 respectively. This pattern was followed by 23%, 17%, and 16% of candidates who obtained B3 and C6, out of the WAEC's expected 65%. Further, more grades 7-9, were obtained by candidates as against WAEC's fixed 30% grading system across the years 2011-2013. This implies that the proportion of grades 7-9 obtained over years are statistically significant than the proportion of grades 1-6 due to effect size. Further analysis of the data to determine significant difference in the standards between SSCE core subjects in the proportion of the grades in relation to the effect size using One-Way ANOVA reveals that the $F_{calculated}$ 6.37 is greater than $F_{critical} = 3.35$, which is significant at 0.05 level of significance. This led to the rejection of the null hypothesis, thus, there is a significant difference in the standards between SSCE core subjects in the proportion of the grades in relation to the effect size. The results of the analysis are presented in Table 9.

4. DISCUSSION

The findings of the first objective of this study show that there is a statistically significant difference in the standard setting procedures used for core subjects of Senior School Certificate Examination (SSCE) Mathematics, English Language, and Biology. Also, a variation exists in the scores that approximated the grades of the different subjects. For instance grade A1 in 2011 were as follows 151 (75.5%) for Biology, 119 (79.33%) for Mathematics, and 149 (74.5%) for English Language. These findings imply that the marks for the critical grades differ from subject to subject and year to year culminating in the disparity in standards between subjects. A grade C in one subject may therefore not be equivalent to a grade C in another subject. These findings are supported by Adeyegbe and Daramola (2004), and Kalgo (2005) who in their separate studies discovered that achievement in one discipline must be recognised as radically different from those in any other. This is because within a subject, the objective to be attained and knowledge to be acquired are such radical differences which are neither equitable with nor reducible to that of any other discipline. Similarly, Soriyan (2002) stated that

the SSCE as public examinations is norm-referenced. The award procedures should arguably take cognizance of the general level of performance of candidates in each of the subject. The levels of performance could therefore vary from one subject to the other and thus the marks that equate the same grades for different subjects are likely to vary from subject to subject. In a clearer term, grade fixing is to indicate whether a performance is good and the extent of "goodness".

However, the findings contrasted with Akeju (2001) who believed that any model that presumes a given standard of candidates to be in a given range for a fluid educational population as in Anglophone West Africa is logically invalid. Yet the model on which WAEC's scale is built is being violated still apart from the above in that the critical grades of B2, C6 and E8 are fixed at award meetings, without any reference to standardized raw scores corresponding to standardized scores of 75, 60 and 30. It is not clear whether this is done before the grades on which candidate performance are based are awarded. This was perhaps why Yoloye et al., (2001) stated that WAEC used certain formulae for deciding the mark ranges corresponding to the various grades without the officials knowing the basis of the formulae. Generally, performance can be good in many ways depending on the subject and the level of attainment which the test experts in that subject considered good. However, the aims and objectives of the syllabus or curriculum should always be put in perspective. Also, since the students who have attempted the examinations from which marks have been generated were of varying cognitive abilities, the norm is often resorted to by WAEC/NECO in setting standards for SSCE. Thus, the use of Stanine (A1-F9) in reporting performance in SSCE by the examination bodies is considered appropriate.

The results of the first hypothesis of this study showed that there was a significant difference in the standard setting procedures used for core subjects of Senior School Certificate Examination (SSCE) with regard to student performance in public examinations. These results were corroborated by Hirst and Peters (1970) as cited by Zubayr (2014) who stated that, achievement in one discipline must be recognised as radically different from those in others. This is because within a subject, the objective to be attained and knowledge to be acquired are different; they are neither equitable with nor reducible to that of any other discipline. Also, on standard fixing at SSCE by WAEC, the first consideration is to note that subjects for the SSCE have different features, each having its own realm of meaning and forms of knowledge distinct from others. The SSCE is a public examination and from every indication a norm-referenced test. Bello (2005) has asserted that most public examinations (including SSCE) are norm-referenced tests and that the raw scores from

such examinations are almost meaningless. Usually therefore, raw scores need further treatment before they are ready for consumption. This treatment is usually referred to as standardization. Furthermore, Adeyegbe (2005), the 9 numerical scales (A1 to F9) which is used to report performance in the SSCE is not new in the history of WAEC. It had been in use for a very long time and was used for reporting the school certificate and GCE 'O' Level examinations prior 1988. The use of this scale for SSCE is seemed adequate more so in terms of value as there is not much difference between the GCE ordinary Level and SSCE except for the beefing up of the contents of the syllabus. The SSCE performs the same function as GCE 'O' Level; for example, preparation for world of work and tertiary education.

Further, the results of the second hypothesis showed that there was a significant difference in the standards between SSCE core subjects in the proportion of the grades in relation to the effect size. This implies that the proportion of grades 7-9 obtained over years are statistically significant than the proportion of grades 1-6 due to effect size. This finding is not surprising as it is consistent with Adeyegbe and Daramola (2004) which asserted that the standard fixing and grade award procedures for the SSCE followed the laid down patterns which are most often rigorous, creative and numerical. Zubayr (2014) noted that after the conduct of the examination, the first major step towards standard fixing and grade award for the SSCE is the conduct of preliminary coordination meetings where the draft marking schemes prepared at the item moderation stage are finalised. As Kalgo (2009) observed, it is at these meetings that the question paper, in its final form, is closely considered to ensure that the final marking schemes takes into account all problems – ambiguity of rubrics or questions, typographical errors or printer devil, unlabelled diagram, etc. which could possibly place the candidate at a disadvantage, however minor. Effort are also made to ensure that the final marking are scheme exhaustive of possible points that could be raised answers to the questions, and that the marks are adequate well spread to reflect weighting of each question or question. It is also at these meetings that the Chief Examiners or Team Leaders standardize their use of the fir marking scheme by marking dummy scripts to ensure unit in marking. It is only when the markings have, become uniform or standardized that they in turn ensure the standardization of the markings of their assistant examiners against theirs (Chief Examiners'/Team Leaders'). It is when all these have been done (again using dummy, scripts) the real marking of live scripts of candidates commence.

Soriyan (2002) stated that during marking, the chief examiners compile their subject reports in which they highlight, among other things, the standard of the

question paper, the standard or level of candidates' performance and the comparability or otherwise of these standards with past ones. These help them in determining what marks to suggest or recommend in their representative to the Council for the three critical grades (2, 6 and 8). The three critical grades serve as determinants for the other six grades on the nine-point grading scale Al to F9. Grade 6 is the least of the credits (C4 to C6) and serves as a line of demarcation between the 'good' candidates Al   C6 and the not so good candidates D7- F9. The A2 as reported by Ukwuegbu (2005) is an elite grade, the base grade for determining the excellent candidates. The B2   and the median of Al (excellent) and B3 (good) is interpreted as 'very good'. The E8 is also amedian, between (pass) and F9 (fail). It divides the weak candidates into two categories - those with bare or ordinary pass and those with very poor result.

In spite of these discrepancies in the results generated by the standard procedure by WAEC, and the relevance of the application of norm referencing in the grading process, this researcher believes suggests that criteria referencing should be used for grading the SSCE, on the strength that it is more objective and could withstand the test of time better than the norm referenced procedure currently in use. It should be noted nonetheless that the essence of WAEC's activity as an examining body is to ascertain whether curriculum objectives derivable from the syllabus are being or have been achieved after the students have gone through it. There are two major ways to achieve this purpose viz assessing through norm-referencing and criterion-referencing. The former has been in practice since inception and is widely adopted largely by many examination bodies. The latter is equally useful assessment. However, one has to state all the criteria to be measured at the onset and a value placed on each criterion. Furthermore, the measuring instruments i.e. test or examination question papers, must ensure that each of the criteria is taken care of. It therefore means that the teaching/learning situation should accommodate the attainment of the criteria set before testing. If a situation where such could be met presents itself, then criterion referencing and consequently fixing of grades/marks for different performance before the examination is attempted can be possible. Even then, it is hardly possible to fix the same criteria for different subjects, in which case the issue of different marks for grades is still likely to prevail across subject areas. As long as it is to identify or label, norm referencing (the type in practice) or criterion referencing (which may be suggested as an alternative) will be applicable for achievement testing, if effort is made to scrutinize the raw scores to ensure standard.

5. CONCLUSION AND RECOMMENDATIONS

From the findings of this study, it can be concluded that a statistically significant difference existed in the standard setting procedures used for core subjects of Senior School Certificate Examination (SSCE) Mathematics, English Language, and Biology. Also, there was a significant difference in the standards between SSCE core subjects in the proportion of the grades in relation to the effect size. Therefore, the pattern of standard fixing and grade awards for the core SSCE subjects followed the norm-referenced procedures as the marks for the critical grades B2, C6, and E8 differed from subjects to subjects and from year to year. The Stanine grading system currently in use for reporting the SSCE is barely adequate for the purpose of identifying candidates and level of their attainment. The marks range for then different grades in each of the subjects investigated appeared not consistent for the core SSCE subjects for the three year period. Based on the findings of this study, the following recommendations were made: the score deviations for each grade of every examination paper should be pegged at +3 from that of the proceeding year and should be strictly adhered to by every grade award committee; standards within and between subjects should regularly be monitored in order to improve the examinations as a whole; attempts should be made by subject experts to set questions of comparable difficulty and having a good spread of the syllabus from year to year. This will make for stability and consistency in the grades awarded; a new system of Ordinary Level grading based on proportion of the marks instead of, as at present, on proportion of the marks in comparison with the proportion of the entry be considered; and the Councils of the examination bodies should set up a task force consisting of officers in Research, Test Development and Test Administration Divisions to have a critical review of the current grade award procedures and make necessary recommendations.

REFERENCES

Adeyegbe, S. O. & Daramola, S. O. (2004). Pre and post item analysis of objective papers of the WAEC GCE O/L Examinations. *WAEC Research Report*, 8(3), 1-17.

Akeju, D. (2001). Statistical approaches to subject grade awards in the GCE O/L Examinations. *WAEC Research Report*, 3(10), 28-41.

Bello, M. A. (2005). A study of the pattern of standard fixing and grade award for the WAEC'S SSCE. *WAEC Research Report*, 6(9), 22-38.

Cizek, G.J. (2006). Standard setting. In S.M. Downing and T. M. Haladyna (Eds.). *Handbook of test development*. Mahwah, NJ: Lawrence Erlbaum Associates, Publishers.

Cooper, S. and Schindler, T. (2001). A generalized examinee-centered method for setting standards on achievement tests. *Applied Measurement in Education*, 14, 343-366.

Faleye, B. A. & Afolabi, E. R. I. (2005). Pre-dictive validity of the Osun State Junior Secondary Certificate Examination. *Electronic Journal of Research in Educational Psychology*, 5(3), 131-144.

Ferdous, G. & Plake, J. (2007). Making the cut-scores in school districts: Alternative methods of setting cutscores. *Educational Assessment*, 6(4), 291-304.

Hambleton, R. K. and Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Eds.). *Educational Measurement*. Westport, CT: American Council on Education and Praeger Publishers.

Hirts, P. H. & Peters, R. S. (1970). *The logic of education*. London: Routledge and Kegan Paul.

Hunter, J. & Schmidt, F. (1990). *Methods of meta-analysis: Correcting error and bias in findings*. Beverly Hills CA: Sage.

Kalgo, F. A. (2009). The debate over standards and the uses of testing. *British Journal of Educational Studies*, 36(1), 198-209.

Kirsch, I. Montgomery, G., & Sapirstein, G. (1995). Hypnosis as an adjunct to cognitive-behavioural psychotherapy: A meta-analysis. *Journal of Consulting and Clinical Psychology,* 63, 214-220.

Lan, S. P. (1992). *The design and evaluation of educational assessment and accountability systems*. Colorado: University of Colorado.

McCullough, M. E. (1999). *Religion and spirituality in rehabilitation psychology*. London: Harper & Row Inc.

Offor, G. (2001). *The content validity of WAEC question papers for the SSSCE in Ghana from 1997-2000*. Accra: WAEC

Soriyan, H.A. (2002). Information on Systems development in Nigeria Software Company: Research Methodological and Assessment form the Health Care Sector's Perspective Electronic. *Journal on Information Systems in Developing Countries,* 5, 234-242.

Ukwuegbu, N. (2004). A multi-stage dominant profile method for setting standards on complex performance assessments. *Applied Measurement in Education*, 8(1), 57- 83.

Yoloye, E. A. Bajah, S. T. Falayajo, W. Obemeata, J. O. And Chacko, I. (2001). Standards *of WAEC examinations: Evaluation reports*. Ibadan: University of Ibadan Press.

Zubayr, F. A. (2014). *A meta-analytic assessment of empirical differences in standard setting procedures in public examinations.* Unpublished Ph.D Thesis, Obafemi Awolowo University, Ile-Ife, Nigeria.