



AGE DIFFERENTIALS IN CALIBRATED ITEMS OF WAEC ENGLISH LANGUAGE OBJECTIVE TEST TAKEN BY STUDENTS IN NIGERIA

ISAAC OFEM UBI AND ESTHER CHINENYE UDEMBA

(Received 4, June 2021 Revision Accepted 5, July 2021)

ABSTRACT

The purpose of this study was to explore the influence of age differentials on item difficulty and discrimination indices of West African School Certificate (WAEC) English Language Objective test for May/June 2014 taken by students in Nigeria. The study area was the southern Education Zone of Cross River State. The design used for the study was ex-post facto, justified by the fact that the variable of the study had occurred before being studied and that no manipulation of subjects of study was involved. The 2014 May/June English objective test items were used as they were. The instrument (made up of all the fifty items of the test) was administered to a sample of 100 students selected through accidental sampling procedure. Results of the study indicate that item response to the test depended, among other things, on the age brackets of the students. The older students seem to respond more correctly to items than their younger counterparts. The study recommends more proactive steps in admission of age compliant candidates for terminal examinations in Secondary Schools.

KEYWORDS: Age Differentials, Calibrated Items, item difficulty indices, item discrimination indices.
(Word count, 187)

INTRODUCTION

Test as an instrument in the hands of any teacher is very important. This is because it is a very relevant tool in educational measurement and has received an endless interest in education. It is the tool that enables teachers to place judgment, make decision, check performance, get response style or picture and determine students' ability.

To measure the learning and teaching of any school subject, the instrument of measurement must be planned, evaluated and tested to ensure it meets the qualities of reliability, validity and usability. Each students' response and

Performance are measured or evaluated on three (3) components; teaching, learning and practice. A study carried out by Umoinyang, Asim, Akwa and Basse (2004) support this performance measurement and evaluation. They opined that "the development of individual attributes at school involve teaching, learning and measurement, to check performance." Without measurement, performance levels of students, which are a function of their individual abilities, cannot be determined. Without measurement, test item appraisal will be impossible, meaning that students' scores in tests will be full of error scores. Also Nenty (2004) admitted that learners' abilities depend on the teacher's instrument of

Isaac Ofem Ubi, Department of Educational Foundations, University of Calabar, Nigeria

Esther Chinenye Udemba, Department of Educational Management and Business Studies
Federal University, Oye-Ekiti

testing, which is a function of the goodness of fit of its construction. One of the major steps in test construction is appraisal of the test through item analysis. This according to Joshua (2005) is “the act of testing the test items to verify whether each item is serving the purpose of testing. It helps in improving the item, and the skill of the teacher in test construction.”

Item analysis involves methods used to evaluate items on a test, both qualitatively and quantitatively, for the purpose of evaluating the quality of items. The goal is to help its developers to improve the instrument by revising or discarding items that do not meet a minimally acceptable standard. Repeatedly used characteristics are the item difficulty (p-value), and discrimination (d-value).

“An item level of difficulty is a factor that affects an individual’s probability of responding in a particular way, similarly, a test item that has a high difficulty index is easier to be answered correctly than an item that has a low difficulty index” (Ubi, Basse & Joshua, 2009). Ijeoma (2001) reported Thorndike and Hagen (1977) that in IRT analysis, trait level and item difficulty are basically connected to each other. In fact, item difficulty is conceived in terms of trait level. Specifically, a difficult item requires a relatively high trait level in order to answer it correctly, but an easy item requires only a low trait level to be answered correctly. Students might need to have a high level of the item ability in order to have a good chance of answering it correctly. They added that “the connection between trait level and difficulty of an item needs a close study on the item construction, before justifications are made.

This paper is focused on two characteristics of an item (item difficulty, p-value & item discrimination d-value). The impact of age at school certificate entry point on these two characteristics using data from West African Senior Secondary Certificate (WASSC) English Language objective test is its focus. Several studies have been carried out on best school age brackets at different levels of schooling. The study of Hambleton and Swaminathan (1991) indicates that “a higher age at entry into an examination is associated with a higher annual GPA. The impact, according to them, tends to diminish with time.”

Anderson (1994) concludes that students whose performances are better in high schools are related to their age, and their achievement varies as their ages differ. The study also revealed that students’ performance in

examinations largely depended on the item strength and students’ background ages. Their study sample of 500 students of different age brackets showed that the older students performed better with 67% and were able to answer the difficult questions than the younger students. The study recommended that tasks should be given to individuals according to their ages and maturity.

In agreement to this assertion, Long (2002) opined that there are maturational constraints on learning and the level of attainment is contingent upon the age at which the learner begins, and suggested that “a sensitive period occurs in learning. Learning that takes place during the ages of 4 -14 is successful while learning after these period is limited.”

Just as the items on a test might differ in terms of their difficulties, the items on a test might also differ in terms of the degree to which they can differentiate individuals who have high trait levels (bright students) from individuals who have low trait levels (dull students). This item characteristic is called item discrimination. According to Xinmng and Yiufai (2004), any item in a test that cannot discriminate the bright from the dull is a bad item. Their study on internal and external factors affecting item discrimination showed that that “77% of the items that discriminates badly were caused by internal factors (sex & age) of the students, while 13% were caused by external factors (environment & teacher) of the students.” Kelly and Linacre (2002), studied item discrimination using Rasch model, and found that the older students were able to read the words with higher vocabularies better than the younger pupil. They noted that the discrimination was basically as a result of the tasks and the ages of the pupils. Other researchers like Gallagher and Delisi (2004) and Halpern (2002) claimed that students’ achievement in Mathematics is basically due to physiological or cognitive difference in individuals and level of materiality.

The main problem of this study is the seeming variations in student response patterns with regards to sex differences and age brackets. Even when difficulty and discrimination indices are averagely acceptable for objective test items of external examinations, there are evidences of discrepancy when the respondents are separated in terms of sex and age brackets. This is not supposed to be so if examination bodies consider these norms while preparing the test items. Literature available, for now and reviewed here, show that earlier researchers have to great

extent estimated item characteristics like item difficulty, discrimination and option distraction. Literature on students' characteristics like sex differences and age brackets with regards to item characteristics seems to be scanty. This represented the gap filled by the present study. The study will be of immense importance to test experts, teachers as front liners of school based assessment, school counselors and administrators.

METHODOLOGY

The area of this study was Calabar Education zone, located in Cross River State, Nigeria.

Ex-post facto design was adopted for the study on the basis that the variables under study had taken place before the researcher embarked on the study and the independent variables cannot directly be manipulated. The study population was made up of all Senior Secondary School three (SS3) Students in Public Schools, in the Seven Local Education Authorities of the zone. As at the time of this research, the target population was 4,674 students in 77 public schools (Source; Planning Research and Statistics, State Secondary Education Board - February, 2014 Academic Session). The sample for the study was made up of 100 students chosen through accidental sampling procedure. This sample size was used for purpose of convenience, since a study involving item analysis would normally involve so many sub-analysis before summaries are prepared. Accidental sampling procedure was adopted to allow only students who were willing to participate in the study to be members of the sample.

The instrument used for collecting data was an objective test on English Language adapted from the West African Senior Secondary School Certificate Examination (WASSCE) for 2013/2014 May/June. The adapted instrument was divided into two parts. Part one consisted of bio-data of the students which elicited information on sex, age, name of school, while part 2 consisted of the 50 items on English Language

RESULTS

Research question one

To what extent are students' age brackets determinants of item response indices in terms of item difficulty?

In order to answer this research question, an item analysis was first carried out to determine the general item difficulty indices for the 50 items. This was done using the person item –matrix. The sample of 100 students was divided into different age bracket of 13years - 14years (28 students), 15years –16years (62 students) and 17years and above (10 students) and the P-values for each item in each age bracket were calculated using;

No or correct responses

No of students

The percentage of items with “High”, “Average” and “Low” difficulty indices were calculated and presented accordingly. See Tables 1 and 2

Results in Tables 1 and 2 show that among the students of age bracket 13years-14years, 18 items (36%) had high difficulty indices. The items were 1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18.

The remaining 32 items (64%) had low difficulty indices (items no. 19 -50). Results for students in age bracket 15years –16years show that 33 items (66%) had high difficulty indices (item no. 1 -33), while 17 items (item no 34-50) had low difficulty indices. Results for age bracket 17years and above were slightly different as up to 39 items representing 78% of the items (items no. 1-36, 40, 41, 42) were found to have high difficulty indices, while only 11 of the items representing 22% (item no 37,38,39,43,44,45,46,47,48,49,50) were found to have low difficulty indices.

These results imply that more of the older students responded correctly to the items, while most of the younger students responded wrongly to the items. Put some other way, most of the items were easier for them which made their p-values high. But the items were difficult for the younger students which made most of their p-value low. It was therefore concluded that test items in the WASSCE 2013/2014 May/June English language objective examination were easier for the older students than they were for the younger students.

Table 1: Summary of item difficulty (p-value) according to students' age brackets

Items	Number of Right Responses			p -value		
	13-14yrs(N=28)	15-16yrs(N=62)	17 – Above(N=10)	13 -14yrs	15 – 16yrs	17– Above
1	25	60	10	0.89*	0.97 *	1.00
2	16	60	10	0.57*	0.97 *	1.00
3	19	59	10	0.68*	0.95 *	1.00
4	20	58	10	0.71*	0.94 *	1.00
5	22	58	8	0.79*	0.94 *	0.80
6	21	57	10	0.75*	0.92 *	1.00
7	19	60	10	0.68*	0.97 *	1.00
8	16	51	7	0.57*	0.82 *	0.70
9	23	50	8	0.82*	0.80 *	0.80
10	20	42	8	0.71*	0.68 *	0.80
11	17	42	6	0.61*	0.68 *	0.60 *
12	20	55	8	0.71*	0.89 *	0.80 *
13	22	48	8	0.79*	0.84 *	0.80 *
14	19	49	8	0.68*	0.79 *	0.80 *
15	16	51	8	0.57*	0.82 *	0.80 *
16	19	35	9	0.68*	0.56 *	0.90 *
17	9	53	7	0.32	0.85 *	0.70 *
18	19	52	7	0.68*	0.84 *	0.70*
19	6	47	7	0.21	0.75*	0.70 *
20	7	50	7	0.25	0.81 *	0.70 *
21	8	50	9	0.29	0.81 *	0.90 *
22	5	49	8	0.18	0.79 *	0.80 *
23	10	47	8	0.36	0.76 *	0.80 *
24	6	28	7	0.21	0.77 *	0.70 *
25	8	49	7	0.29	0.79 *	0.70 *
26	6	46	7	0.21	0.74 *	0.70 *
27	8	8	8	0.29	0.71 *	0.80 *
28	4	45	7	0.14	0.73 *	0.70 *
29	7	38	9	0.25	0.61 *	0.90 *
30	5	35	7	0.18	0.56 *	0.70 *
31	8	33	7	0.29	0.53 *	0.70 *
32	11	32	8	0.39	0.52*	0.80 *
33	9	30	7	0.32	0.52 *	0.70 *
34	8	20	6	0.29	0.32	0.60 *
35	4	27	5	0.14	0.44	0.50 *
36	3	28	5	0.11	0.45	0.50 *
37	1	26	4	0.04	0.42	0.40
38	10	24	3	0.36	0.39	0.30
39	9	29	2	0.32	0.48	0.20
40	7	23	6	0.25	0.37	0.60*
41	0	18	9	0.00	0.29	0.90*
42	0	19	5	0.00	0.31	0.50*
43	8	25	4	0.29	0.40	0.40
44	4	22	3	0.14	0.35	0.30
45	2	30	4	0.07	0.48	0.40
46	3	26	5	0.17	0.42	0.50*
47	1	1	2	0.04	0.02	0.20
48	1	12	1	0.04	0.19	0.10
49	0	4	0	0.00	0.06	0.00
50	1	1	2	0.04	0.02	0.10

* High difficulty indices (High p-value) Ranging from 0.50 -1. 00

Item without asterisks indicate low difficult indices(Low p-value) ranging from 0.00 –0.49

Table 2: Summary of simple percentage for item difficulty indices according to students' age brackets

S/No	Age Brackets	Difficulty		
		Dimensions	No of items	%
1.	13years – 14years	High	18	32
		Low	32	64
		Total	50	100
2.	15years – 16 years	High	33	66
		Low	17	34
		Total	50	100
3.	17years and above	High	39	78
		Low	10	22
		Total	50	100
4.	Overall	High	30	58
		Low	21	42
		Total	50	100

RESEARCH QUESTION TWO

To what extent are students' age brackets determinants of item response indices in terms of item discrimination?

To answer this research question, the student's responses were arranged in a Person-Item Matrix Table and were separated according to their different age brackets. For each age bracket, the students were further divided into 'High', 'Average' and 'Low' (33.3% for each group). The discrimination indices using 'High' and 'Low' groups were then computed. (See results in Tables 3a, 3b & 3c). The percentages of the number of items with high and low item discrimination indices were calculated and presenting in Table 4.

Results in Tables 3a, 3b and 3c for students within the ages of 13years - 14years, 43 items (items no. 2,3,4,5,6,8,9,-19,21-25,27,29,30,31,32,34 -50) were having high discrimination indices and only 7 items (items no 1,7,20,22,26,28,33) had low discrimination indices. Also for ages 15years -16years 35 items (items no 8,10-28,31-45) had high discrimination

indices, while 15 items (items no 1,2,3,4,5,6,7,9,29,30,46,47,48,49,50) had low discrimination indices. Result for students of 17years and above showed that 36 items (items no 1-11, 14,15,22, 23,24,25,26,27,29,30,31,32,33-50) discriminated highly, while 14 of the items (items;12,13, 16,17,18,19,20,21,,28,39,48,49,50) discriminated lowly.

Percentage of item discrimination indices Table 4 indicate that for 13years -14years (extracted from Table 3a) 14% of the 50 items discriminated highly while 86% discriminated lowly. For ages 15years -16years, it was discovered that 70% of the items discriminated highly, while 30% was of low discrimination level. Students of 17years and above had 62% of the items discriminating highly, while 38% discriminated lowly. From the overall results in Table 4, it was observed that items in the WASSCE English language objective test discriminated the bright student from the dull students better for older students that they did for younger students.

Table 3a: Age and Item discrimination indices (*d-values*) for 13-to-14years (n=28)

Items	Number of students in the upper and lower groups that got the item correctly		d-value
	High (Bright) U = 9	Low (Dull) L = 9	
1	9	3	0.67*
2	6	4	0.44
3	7	5	0.22
4	9	5	0.44
5	6	3	0.33
6	7	4	0.33
7	8	3	0.56*
8	4	3	0.11
9	6	6	0.00
10	6	5	0.11
11	6	5	0.11
12	6	5	0.11
13	5	6	-0.11
14	6	4	0.22
15	9	3	-0.22
16	3	6	-0.33
17	5	4	-0.11
18	7	9	-0.22
19	5	2	0.22
20	9	3	0.67*
21	6	3	0.33
22	9	2	0.79*
23	8	20	-0.33
24	5	3	0.22
25	3	2	0.11
26	8	1	0.77*
27	3	5	-0.22
28	9	2	0.77*
29	3	4	-0.11
30	2	5	-0.33
31	2	4	-0.22
32	3	2	0.11
33	6	1	0.55*
34	2	8	0.11
35	0	2	-0.22
36	1	0	0.11
37	0	1	-0.11
38	6	3	0.33
39	4	2	0.22
40	4	2	0.22
41	0	0	0.00
42	0	0	0.00
43	5	2	0.33
44	3	1	0.22
45	2	0	0.22
46	2	0	0.33
47	1	0	0.11
48	0	0	0.00
49	0	0	0.00
50	1	0	0.11

*High discrimination indices (High d-value) ranging from 0.50- 1.00

Item without asterisk are Low discrimination indices (Low d-value) ranging from 0.00 –0.49

Table 3b: Age and Item discrimination (d-value) for 15--16 year (n = 62)

Items	Number of student in the upper and lower group that got the item correct		d – value
	High (Bright) U=20	Low (Dull) L = 20	
1	20	20	0.00
2	20	20	0.00
3	19	18	0.05
4	20	16	0.20
5	19	18	0.05
6	19	18	0.05
7	20	19	0.05
8	20	10	0.50*
9	19	18	0.05
10	20	9	0.55*
11	14	2	0.60*
12	19	0	0.70*
13	18	0	0.90*
14	20	4	0.80*
15	20	3	0.85*
16	17	4	0.65*
17	20	3	0.85*
18	19	3	0.80*
19	18	2	0.80*
20	15	2	0.65*
21	16	2	0.70*
22	15	4	0.55*
23	16	3	0.65*
24	18	3	0.75*
25	18	3	0.75*
26	15	4	0.55*
27	16	5	0.55*
28	15	4	0.55*
29	16	9	-0.35
30	16	8	-0.40
31	19	7	0.60*
32	19	7	0.60*
33	17	6	0.55*
34	15	3	0.60*
35	16	1	0.75*
36	19	8	0.55*
37	20	1	0.95*
38	18	2	0.80 *
39	18	8	0.50 *
40	16	4	0.60 *
41	17	3	0.70*1
42	18	4	0.70*
43	18	4	0.70*
44	16	2	0.70*
45	17	2	0.75*
46	12	11	0.05
47	1	0	0.05
48	1	0	0.05
49	2	1	0.15
50	0	0	0.00

*High discrimination indices (High d-value) ranging from 0.50- 1.00

Item without asterisk Low discrimination indices (Low d-value) ranging from 0.00 – 0.49

Table 3c: Age and Item Discrimination indices (*d-values*) for 17yrs and above (n=10)

Items	Number of students in the upper and lower group that got the item correct		d – value
	High (Bright) U = 3	Low (Dull) L = 3	
1	3	1	0.67*
2	3	1	0.67*
3	2	0	0.67*
4	3	1	0.67*
5	3	1	0.67*
6	3	1	0.67*
7	3	1	0.67*
8	3	1	0.67*
9	3	1	0.67*
10	3	1	0.67*
11	3	1	0.67*
12	3	3	0.00
13	3	3	0.00
14	3	1	0.67*
15	3	1	0.67*
16	3	2	0.33
17	2	1	0.33
18	2	1	0.33
19	1	2	-0.33
20	3	3	0.00
21	3	2	0.33
22	3	1	0.66*
23	1	3	-0.66
24	3	0	1.00*
25	3	1	0.67*
26	3	1	0.67*
27	3	0	1.00*
28	3	2	0.33
29	3	3	0.67*
30	3	1	0.67*
31	2	1	0.33
32	3	1	0.67*
33	3	2	0.67*
34	3	1	0.67*
35	3	1	0.67*
36	2	0	0.67*
37	3	1	0.66*
38	2	0	0.67*
39	0	0	0.00
40	3	0	1.00*
41	2	0	0.67*
42	3	0	1.00*
43	2	2	0.67*
44	2	0	0.67*
45	3	0	1.00*
46	3	1	0.67*
47	2	0	0.67*
48	0	1	0.33
49	0	0	0.00
50	0	1	-0.33

*High discrimination indices (High d-value) ranging from 0.50- 1.00

Item without asterisk Low discrimination indices (Low d-value) ranging from 0.00 – 0.49

Table 4: Summary of simple percentage for item discrimination indices according to students' age bracket

S/No.	Age Brackets	Discrimination Dimensions	No of items	%
1	13years – 14years	High	7	14
		Low	43	86
		Total	50	100
2	15years– 16 years	High	35	70
		Low	15	30
		Total	50	100
3	17years and above	High	36	68
		Low	14	28
		Total	50	100
4	Overall	High	26	52
		Low	24	48
		Total	50	100

DISCUSSION OF FINDINGS

The findings of this research reveal that student's age influenced their responses to test items in terms of item difficulty. Older students chose more correct options in the test than younger students. The finding indicates that the older the students were the better their responses to Mathematics items at the school certificate examination. This finding is in agreement with that of Udemba (2018) who at different study areas and examinations revealed that students' performance at secondary schools depended upon their age brackets. According to them, when students who are far apart from each other in ages are evaluated together their responses differ in to the advantage of the older ones.

The findings of this study is also in line with Akwa (2008) who confirmed that under aged students in the university perform lower than their counterparts of higher age brackets, and supports Nenty (2004) who concluded that, allowing under aged students (below 18years) into the university education can affect their social-educational performance. The finding however contradicts that of Long (2002), who concluded that, irrespective of the age of students, their response/performance in examinations depended basically on the students' level of readiness, problem solving ability, interest, teaching methods and hereditary trend. These conflicting findings are not surprising in the face of modern trends in education. Children whose ages are as low as two to three years are sent to primary school thus

making it possible for them to be ripe for school certificate examination at 14 to 15 years of age.

The findings of the study in terms of item discrimination indicate that ages of the students influenced the discrimination indices of the items in WASSCE English language objective test. This study explains that the higher the age brackets of students the better the power of the items to discriminate between the bright and dull students.

This finding supports that of Xinmng and Yiufai (2004) who concluded that any item of a test that cannot discriminate the bright from the dull is a bad item. They used a written examination with internal factors (ages and sex) of students, and reported that 77% of the items that discriminates badly were caused by these internal factors of the students while 13% were caused by external factors (environment, test, teacher) of the student.

CONCLUSION

This study used some SS three students who were about to write the main Senior Certificate examination as sample. The items of the test were calibrated using the responses of the students into item difficulty indices (p-values) and item discrimination indices (d-values). Results of the calibrations were arranged according to age brackets of the students, and the results show that students' age differentials affected item characteristics like difficulty and discrimination indices. Based on the results, the study concludes that students of older ages had less difficulty in tackling the English Language

objective items than their younger counterparts. It was also concluded that items of tests discriminate better with older students than with their younger counterparts.

Based on these results the study recommends that school administrators and examination bodies should make policies that will help the system. School administrators should step up their age requirements for admission into junior secondary one (JSS 1), and senior secondary one (SSS 1). Also examination bodies should consider a good age limit for entrants into the WASSCE before enrolling the students for the examination. This will enable the students get to more mature ages before they are ripe for school certificate examination. Teachers and parents should identify students' abilities, thereby taking care of individual difference and pay more attention to the under aged students by arranging for extra-mural classes and tutorials in English language. The researchers are of the opinion that more studies in this area are needed in other subject areas.

REFERENCES

- Anderson, W.S.,1994. Basic learning theory for teachers. In C.E. Skinner (Ed) Educational Psychology pp 229 -416. New Delhi. Prentice –Hall.
- Collier, V. P., 1999. Age and rate of acquisition purpose. *Tesol Quarterly*, 30 (2), 21-35.
- Halpern, A., 2002. Sex difference in cognitive ability 2nd Edition. Hilldale N.J: Erlbaum.
- Hambleton, R. R., and Swaminathan, H and Rogers, H. J., 1991. *Fundamentals of item response Theory*. New Burry Park, Sage Press.
- Ijeoma G, M., 2001. Testwiseness, Guessing tendency and performance in Mathematics among secondary school students in Imo. Unpublished Ph.d Thesis, Faculty of Education, University of Calabar.
- Kelley, T., Linacre, J.,2002. Item discrimination indices. In Rasch Measurement

Transactions, 2002 16: 3 p3-4, www.rasch.org/rmt163a.htm

- Long, M., 1990. Maturation constraints on Language development. *Studies in Second language acquisition*, 3(6), 56-68.
- Nenty H. J., 2004. From Classical test theory (CTT) to item Response theory (IRT): Introduction to a desirable transition. Owerri, Nigeria. In O. A. Atermikhe and J. C. Adewale (Eds.) *Issues in Educational Measurement and Evaluation in Nigeria (371-384)*. Ibadan: Institute of Education, University of Ibadan Nigeria.
- Thorndike, R. L., and Hagen, E. P., 1977. *Measurement and evaluation in psychology and education* 5th ed. New York: MacMillan.
- Ubi, I. O., Bassey, S. W. and Joshua, M. T., 2009. Demographic factors in Mathematics performance in University Matriculation Examination (UME) in Nigeria. *African Journal of Interdisciplinary Studies*, 2, 12-16.
- Udemba, E. C., 2017. Analysis of Students' characteristics and item response indices in an English Language objective test among secondary school students in Southern Education Zone of Cross River State, Nigeria. An Unpublished M.Ed Thesis, University of Calabar, Nigeria.
- Umoinyang, I. E., Asim, A. E., Akwa, A. M., and Bassey, S. W., 2004. *Principles and Techniques of educational assessment and evaluation*. Calabar: Helimoassociates.
- Xinming, A. and Yiufa Y., 2004. An examination of item-parameter effects (DIF) and Gender DIF. *Applied Measurement in Education*, 14, 73–90.