IOSR Journal of Research & Method in Education (IOSR-JRME)
e-ISSN: 2320–7388,p-ISSN: 2320–737X Volume 6, Issue 2 Ver. II (Mar. - Apr. 2016), PP 42-48
www.iosrjournals.org

Measuring Differential Frequency of Option Response Patterns in Four-Five Options Multiple-Choice Test Item among Undergraduate Students in Nigeria

Abdul-Wahab IBRAHIM (Phd)
Department of Education, Faculty of Education, Sule Lamido University, Kafin-Hausa, Jigawa State, Nigeria

Abstract: *Multiple-choice items are a widely used item format in tests of achievement, knowledge, and ability. Given the high-stakes associated with the use of multiple-choice test in schools, colleges and universities, it is important to collect differential frequency of option selection evidence to support such usage of test scores. Adopting a survey design, the sample consisted of an intact class of 117 part 2 undergraduate students who registered for EDU 203: Introduction to Research Methods, in the First Semester (Harmattan) of 2015/2016 session. Data on 117 students were obtained from 2 alternative test formats of 20-item each. The arrangement of correct responses (key) and incorrect responses (distracters) was randomized throughout the test by a scheme which allowed each position an equal number of choices. The correct choice for each item appeared in a different position on each of the forms. Data collected were analysed using Chi-Square(X^2) and One-Way Analysis of Variance (ANOVA) statistics. Results reveal a significant influence of item positional preference bias on item difficulty in multiple-choice test. Also, there exists a significant influence of correct response choice location on the difficulty of multiple-choice test items. However, item option length has no significant effect on response bias in multiple-choice tests. It is concluded that the position of the most plausible distracters more logically accounts for any significant response patterns than thus a position preference.*
Key words: *Differential frequency, response pattern, item difficulty, key and distracters,*

I. Introduction

Multiple-choice objective tests normally have options ranging from three to five with one of the options being the key. They have been found to be effective in measuring ability in different subjects and at different levels in schools and colleges (Ibrahim, 2010). However, as in all fixed-choice objective tests, multiple-choice test score have been influenced by contaminants such as response set, response changing and guessing, among other testee-related factors (Afolabi, 2012). Incidentally, the above mentioned influences are substantially related to the technical standard and psychometric adequacy of the test. Where the test is not well prepared, students may earn undeserved scores that tend to misrepresent their actual academic ability through testwiseness (Afolabi & Ibrahim, 2009).

In selecting the correct alternatives, authors (e.g. Markus, 2001; Penfield & Jimmy, 2008; DeMars, 2010) reported a tendency among test takers to favour certain response position to the neglect of others. Their findings suggest that the difficulty level of a multiple-choice test is influenced by the position to which the correct choice has been assigned. On the other hand, other authors (e.g. Rodriguez, 2005; Osterlind & Everson, 2009; Herrmann-Abell & DeBoer, 2011) contend that multiple-choice tests are relatively free from position response bias.

Similarly, Klassen (2006) and Liu & Boone (2006) found that for five-choice items, those items having right answers or correct keys in the fourth position are the most difficult; those with right answers (keys) in the second and third positions are the easiest; and the first and fifth positions are of equal moderate difficulty. Their results for four-choice items show that the third position is most difficult, and that difficulty increases from the first through the third position, and decreases with the fourth. This finding is interpreted as agreeing with the results for five-choice items (Ibrahim, 2010) in that the next to last position was always found to be most difficult. Even though this position factor was relatively small, it was found to be statistically significant.

In spite of these inconsistent findings, Herrmann-Abell & DeBoer (2011) contend that an understanding of the material presented in a question is not the only factor at work in the selection of the correct answer. They believe that a subject does not always select an alternative on the basis of information alone, but is influenced to some degree by these positional factors. Essentially, their hypothesis is that since both the first and last choice in a list are more outstanding than the middle three or two as the case may be, these inner choices become less noticeable and are less likely to be selected. Another possible explanation they offer is that a person going through the list without making a choice, is more likely to select the last choice rather than going through the list for a second time. This, however, is not sufficient to explain why the penultimate position was always found to be more difficult than the rest.

Conceptual problem included definitional differences that led to divergent views about what evidence would demonstrate a positional response cum correct response location on the difficulty level of multiple-choice (MC) questions. For example, Wesman (2000) argued that internal consistency of a "bias score" constituted evidence of positional bias but Larkins and Shaver (2001) evaluated the differential frequency of option selection.

According to Green (2000) positional response bias may depend on degree of testwiseness or on guessing strategy. He argued that positional guessing was a testwise response to biases in the positioning of keyed responses. If keyed responses are evenly distributed, however, then a positional guessing strategy is unsupported. Hence, positional responding, guessing strategy, and testwiseness are related concepts; but in previous research, these constructs have been studied separately.

This claim is consistent with Afolabi's (2000) position that if change made were not due to impulse or cheating or were as a result of testwiseness, informed guessing, or a new insight to the requirements of the item, then testees with superior ability are most likely to gain more from any changes made in their initial answers. Of course, they would also be expected to make the last changes since they ought to be able to identify the key to most of the items of an achievement test at the first attempt.

In tests on learning and study skills, it has been asserted that learning skills students do not read all the alternatives on multiple-choice exams before selecting an answer or correct key. Failure to read all options would result in a response bias for early options (Herrmann-Abell & DeBoer, 2011). Of course, other factors such as primacy effect or deliberate guessing strategy could also result in a response bias for early options. In fact, deliberately choosing early options is part of testwiseness. DeMars (2010) posited that if a test was designed that each item of it contained alternatives chosen deliberately so that they ranged from completely wrong to almost right, and if one of the alternatives for each item were "correct answer not given", it seems reasonable that one might find a response set such that some people would consistently tend to check the correct answer whereas others might consistently avoid it. Thus a less obvious factor that influences test result is response bias, a consistent tendency to follow a certain pattern in responding to test items.

As a corollary to the above, the present study attempts to examine the selection by testees of one response position more often than others in a multiple-choice objective test, determine relationship between the position to which the key and distracters have been assigned and find out the influence of item option length on response bias in multiple-choice test specifically among undergraduate students in Nigeria. The positioning of keyed responses and students responses set are variables capable of influencing performance in multiple-choice tests. Since multiple-choice tests are used in Nigeria at every level of the educational system, the value of early option keys will assist test constructors and examination bodies in the choice of the appropriate number of options for each level of the school systems. Not only this, but also the result will assist item writers to optimize the difficulty level of test items, vis-à-vis selecting suitable number of distracters appropriate for Primary, Junior Secondary, Senior Secondary and Tertiary level students in the country (Nigeria). Thus, the outcome of the study will serve as a guide to the classroom/subject teachers throughout the federation on the need to be familiar with procedures involved in the randomization of correct choices (keys) on equal basis among the alternatives/options, in developing Teacher Made Tests for use on their students. Thus, three hypotheses meant to guide the direction of this study were postulated as follows:

(i) There is no significant influence of item positional preference bias on item difficulty in multiple-choice tests.
(ii) There is no significant influence of correct response choice location on the difficulty of multiple-choice test items.
(iii) That item option length has no significant effect on response bias in multiple-choice tests.

II. Methodology

This investigation adopts survey research design. A survey research design allows for the collection of quantifiable data from a sample to explain a particular phenomenon (Upadhya and Singh, 2008).

All undergraduate students who registered for a compulsory course in Introduction to Research Methods during the First Semester (Harmattan) of 2015/2016 session in the Faculty of Education of Sule Lamido University, Kafin Hausa, Jigawa State, Nigeria, constituted the target population for the study. There were a total of 117 undergraduates who registered for the course during the session. The sample consisted of an intact class of 117 part 2 undergraduate students who registered for EDU 203 in First Semester (Harmattan) of 2015/2016 session. Thus, the entire population was therefore used and no sampling was carried out. Also, the choice of the EDU 203 is appropriate in this study due to its Faculty status as a compulsory course aimed at introducing the basic procedures involved in educational research as well as the procedures involved in the development, administration and reporting of research of acceptable standard. Also, as a Faculty course, all undergraduate students of different majors must register for the course to be allowed to earn a Bachelor Degree in Education.

A 40-item multiple-choice achievement test designated "Undergraduate Achievement Test (UAT)" was developed by the researcher and used as instrument for the study. The instrument is a 40-4/5 option multiple-choice test items that was developed using the course (EDU 203: Introduction to Research Methods) content. Subjects were instructed to circle or tick the correct and most appropriate option. The subjects were duly informed not to leave unattended any of the 40 items, but to attempt all items on the special answer sheet provided by shading appropriate letter-alternatively per item. The content validity of the instrument was established by the researcher using Kuder and Richardson Formula 21 reliability method. The Kuder - Richardson Formula 21 method was preferred because of the desire to determine the internal consistency of the instrument for data collection. The Kuder-Richardson reliability co-efficient of the instrument is 0.72 and interval consistency of 0.84. Its mean ($\bar{x}$) difficulty index was 0.52 with a standard deviation of 0.24. The item discrimination indices have a mean ($\bar{x}$) value of 0.28 and a standard deviation of 0.19. All these values were acceptable and appropriately high for a study of human behaviour due to its complexity.

To control for response patterns in the arrangement of the alternatives (A-E), an approximately equal number of correct choices were randomly scattered among the options. In each, the number of questions keyed A, B, C, D and E was the same, and their order was randomized. Alternate forms order was developed for counter-balancing the order of alternatives for each question. Test items were all multiple-choice items and consisted of an incomplete statement which the examinee could complete correctly by selecting one of the four-five phrases following it. Below is a sample of the type of questions used in the test:

1. Which of the following best describes quantitative research?
A. The collection of non-numerical data
B. An attempt to confirm the researcher's hypotheses
C. Research that is exploratory
D. Research that attempts to generate a new theory

2. A variable that is presumed to cause a change in another variable is called a(n): ....
A. Categorical variable
B. Dependent variable
C. Independent variable
D. Intervening variable
E. Confounding variable

Moreover, the position that each alternative was assigned for each item was determined by a scheme contrived by Afolabi and Ibrahim (2009) for arranging and randomizing correct choices and distracters. All 24 permutations of the numbers one through four were assigned positions in the test in accordance with the order of their appearance as determined by a table of random numbers. Alternatives for each item were rearranged to conform to the sequence number of the permutation assigned to it. The scheme was reapplied to determine the positions of alternatives for each of the remaining items. Each permutation was used once before any one was repeated.

The number-rights score was used in grading the test scripts. The test was not speeded and contains no instruction against guessing. Since the test condition was formal, the papers (i.e. scripts) of subjects who omitted items or who responded more than once to an item were discarded. Approximately 117 answer sheets from that test were analysed and a difficulty, defined as the percentage of subjects failing an item were computed and rearranged in order of increasing difficulty. This was done to create a uniform bias throughout the test and to increase the probability of eliciting a position preference as the items becomes more difficult.

The researcher personally administered the instrument who doubled as the Course Instructor/Lecturer of EDU 203: Introduction to Research Methods. The hard copies of the instrument were administered on the subjects as part of Continuous Assessment for the course during the First Semester (Harmattan) 2015/2016 session. The confidentiality of the subjects' responses was assured while respondents were allowed enough time to complete the test. The subjects were instructed to read each item very carefully and then to mark on their answer sheets that alternative which they have decided is most correct. They were told not to omit any items especially when in doubt. Subjects were also informed that they would have sufficient time to complete all the items. Such a procedure provides uniform response set thereby minimizing individual differences in responding. Hence, the test condition was formal. After the completion of responses by the subjects, the instrument copies were collected immediately from the subjects as they were leaving the Hall. Also, the researcher cross-checked the nature of subjects' responses on the items, as some respondents may have skipped some of the items to correctly and fully respond to such items. The respondents were also thanked for their co-operation. Frequency counts, mean ($\bar{x}$) scores and standard deviations were first computed using Chi-Square (X^2) and One-Way Analysis of Variance (ANOVA) statistical techniques were used to test the three hypotheses in the study. All hypotheses formulated were tested at 0.05 level of significance.

III. Results

Hypothesis One: This states that there is no significant influence of item positional preference bias on item difficulty in multiple-choice tests. The results of the analysis are presented in Table 1 for 4 – option multiple-choice and Table 2 for 5 – option multiple-choice.

Table 1: Chi-square analysis of influence of positional preference on item difficult of 4-option multiple-choice test

Position preference	Item Difficulty						Total	X^2	p
	.34	.41	.53	.62	.73	.84			
A	10	-	-	-	8	8	26	36.04*	.05
B	-	5	8	6	-	-	19		
C	6	5	5	8	8	6	38		
D	6	6	7	5	5	5	34		
Total	22	16	20	19	21	19	117		

***Significant; p<0.05**

Table 1 describes the number of testees who selected options A, B, C or D for an item corresponding to item difficult range of 0.3 to 0.8. Thus, the latter positions C and D with 38 and 34 frequencies respectively were preferred than the early positions A and B. Further, more persons selected options of items having difficult value of 0.3 (n=22); 0.7(n=21) and 0.5 (n=20) than those selected options with difficult value of 0.4 (n-16); 0.6 (n=19) and 0.8 (n = 19). To determine whether preference for particular item option is influenced by the difficult of the item, a chi-square analysis yielded a value of $X^2 = 36.04$, which is significant at p<. 05. That is, in 4-option multiple choice test, the preference by testees of option positions is significantly influenced by the difficult of test items.

Table 2: Chi-square analysis of influence of positional preference on item difficult of 5-option multiple-choice test

Position preference	Item Difficulty						Total	X^2	p
	.31	.44	.52	.64	.72	.83			
A	8	6	6	5	5	-	30	42.15*	<.05
B	-	6	5	-	6	5	22		
C	-	8	6	5	8	-	27		
D	-	-	5	5	-	5	20		
E	5	-	-	8	5	5	18		
Total	13	20	22	23	24	15	117		

***Significant; p<0.05**

Table 2 presents the results for 5 – option multiple-choice. In Table 2, there appears to be a reducing tendency by testees to select options placed at the first positions (e.g. A, B) than those placed at last positions (e.g. D, E). For example, while 30 testees selected option A, the number reduced from 22 and 20 selecting B and D respective to 18 selecting E; the value for C may be an outlier. Also, fewer testees selected the options corresponding to low difficulty (n = 13 when p = 0.3) or high difficulty (n = 15 when p = 0.8). Chi-square analysis of the results yielded a value of $X^2 = 42.15$, p< 0.5. These results suggested an association between option preferences in multiple-choice and item difficult. In particular, testees tend to select earlier options in difficult items.

Hypothesis Two: This hypothesis states that there is no significant influence of correct response choice location on difficulty of multiple-choice test items. The results of chi-square analysis of the data on the hypothesis are presented in Tables 3 and 4.

Table 3: Chi-square analysis of influence of correct response choice location on difficulty of 4-option multiple-choice

Item Difficult Levels	Correct Choice (Keys)						
	A	**B**	**C**	**D**	**Total**	**X^2**	**p**
.34	6	6	6	5	23	33.98*	<.05
.41	-	8	5	6	19		
.53	5	8	9	-	22		
.62	7	5	-	8	20		
.73	-	6	7	-	13		
.84	5	-	6	9	20		
Total	23	33	33	28	117		

***Significant; p<0.05**

From table 3, 42 testees selected the correct choices for items of relatively low difficulty of 0.3 and 0.4 compared with 33 testees who selected the correct choices corresponding to items of relatively high difficulty (p = 0.7 and p = 0.8). Further, the number of testees selecting the correct edge options (i.e. A and D) were less the same (n=51) as the 66 that selected the middle correct options (B and C). However, 56 testees selected the correct options A and B, while 61 testees selected the correct options C and D. A Chi-square (X^2) analysis of the results yielded 33.98, which is significant at p<.05. Thus, the null hypothesis is disconfirmed; that is, there is a significant influence of correct response choice location on the difficulty of 4 – multiple-choice test. Table 4 presents the results for 5-option multiple-choice.

Table 4: Chi-square analysis of influence of correct response choice location on difficulty of 5 – multiple-choice test

Item Difficult Levels	Correct choice (Keys)							
	A	B	C	D	E	**Total**	**X^2**	**p**
.31	5	6	-	7	5	23	35.23*	<.05
.44	7	5	-	5	5	22		
.52	-	5	9	5	8	27		
.64	-	7	-	5	-	12		
.72	6	6	-	-	5	17		
.83	5	-	6	5	-	16		
Total	23	29	15	27	23	117		

***Significant; p<0.05**

From Table 4, in the 5 – option multiple-choice test, 72 testees selected correct option for items of relatively low difficult of p = 0.3, 0.4 and 0.5 compared with 45 testees who selected correct options corresponding to items of relatively high difficulty (p = 0.6 to 0.8). Also, more choices were made of the second (n = 29) and middle (n = 27) positions than of the first (n=23), third (n=15) and last (n=23) positions. A Chi-square (X^2) analysis of the results yielded 35.23, which is significant at p<.05. Thus, the null hypothesis is disconfirmed; that is, there is a significant influence of correct response choice location on difficulty level of 5 – option multiple-choice test.

Hypothesis Three: This hypothesis states that item option length has no significant effect on response bias in multiple-choice test. Table 5 presents the descriptive data of the analysis while Table 6 presents the One-Way Analysis of Variance (ANOVA) comparison of the mean values.

Table 5: Mean ($\bar{x}$) response bias due to item option length

Option length		Correctchoice	position			
		A	**B**	**C**	**D**	**E**
	N	33	30	35	19	-
4 – Option	$\bar{x}$	23.24	26.36	29.36	21.6	-
	S	2.22	3.18	2.12	2.10	-
	N	27	22	30	25	13
5 – Option	$\bar{x}$	24.64	24.69	26.58	25.55	24.13
	S	2.27	2.60	2.51	2.29	2.31

From Table 5, in the 4 – option multiple-choice, the middle options B and C have higher mean scores ($\bar{x}\text{b} = 26.77; \bar{x}\text{c} = 29.36$) than the edge options (A and D) with option C having the highest ($\bar{x} = 29.36$) mean valueand option D, the lowest ($\bar{x} = 21.69$) mean value. Similarly, in the 5-option items, there appears to be a tendency for testees to be biased in favour of the middle options (B, C and D) which have greater mean values

($\bar{x} = 24.69$; $\bar{x} = 26.58$); and $\bar{x} = 25.55$) than the edge options A and E ($\bar{x} = 24.64$; and $\bar{x} = 24.13$) with the most middle option C having the greater mean score of $\bar{x} = 26.58$. In both 4- option and 5-option formats, more testees showed greater bias for the middle correct positions, n= 35 and n= 30 for option C respectively. Further, Tables 6 and 7 present ANOVA results which reveal a not significant F=0.414, p>.05 between response bias in 4 –option item formats.

Table 6: One-way Analysis of Variance (ANOVA) on response bias due to item option length of 4-option items

Sources of Variation	Sum of Squares	Degrees of Freedom	Means of Squares	f-ratio	P
Between Groups	66.1	3	22.03		
				0.414	>.05
				(ns)	
Within Groups (Error)	6018.5	113	53.26		
Total	6084.6	116			

Ns= not significant; df = 3 and 113; critical F = 2.68

Similarly, Table 7 presents the results for 5-option format. A comparison of the mean values revealed F=0.502, p>.05, which suggests that response bias is not influenced by test item length. Thus, in both 4 – option and 5 – option formats, item option length has no significant influence on response bias in multiple-choice tests.

Table 7: One-way Analysis of Variance (ANOVA) on response bias due to item option length of 5 – option items

Sources of Variation	Sum of Squares	Degrees of Freedom	Means of Squares	F-ratio	p
Between Groups	128.5	4	32.13		
				0.502	>.05
				(ns)	
Within Groups (Error)	7163.9	112	63.96		
Total	7292.4	116			

ns= not significant; df = 4 and 112; critical F = 2.44

IV. Discussion

From the above, the result of the analysis of the first hypothesis reveals that there is a significant influence of item positional preference bias on item difficulty in multiple-choice tests. That is, in 4-option multiple choice test, the preference by testees of option position is significantly influence by the difficulty of test items. Also an option preference in 5-option multiple-choice test is significantly influenced by the difficulty of test items. It follows from this that the difficulty level of a multiple-choice test items is influenced somewhat by the placement of the correct choice. These findings agree with those of Markus (2001); Penfield & Jimmy (2008); DeMars (2010); Rodriguez (2005) and Green (2000). These scholars in their separate studies found that the difficulty level of a multiple-choice test item is influenced by the position to which the correct choice has been assigned. They concluded that response sets are most apparent when items become more ambiguous or when they increase in level of difficulty.

Another finding of this study indicates that there is a significant influence of correct response choice location on difficulty level of multiple-choice tests. That is, a significant influence of correct response choice location exists on the difficulty level of 4-option multiple-choice test. Also there is a significant influence of correct response choice location on the difficulty level of 5-option multiple-choice test. The finding lends support to the postulation that difficulty is a function of correct choice placement (Herrmann-Abell & DeBoer, 2011). In other words, that position response sets are negligible and certainly not a significant source of invalidity accounts for any significant response bias than does a position preference. Hence the effect of correct response location on the difficulty level of multiple-choice questions exists.

This researcher is of the opinion that a person who has arrived at the correct choice is, perhaps more likely to select the last choice than to go through the list of choices another time. The possible confusion of thought resulting from perusal of several wrong answers certainly cannot be a factor at play in this case, since the last choice will be subjected to more proactive effects than the penultimate. But proactive and retroactive influences are not sole determiners since in five-choice items, these would be strongest for the third rather than the fourth position, nor do they explain the greater difficulty of position one, when compared with two and three, in five-choice questions. That is, positional response bias could be an extraneous source of variance in test scores depending on the degree of individual testee's testwiseness.

Another finding of this study shows that item option length has no significant influence on response bias in multiple-choice tests. This finding is consistent with the findings of Osterlind & Everson (2009) and

Afolabi & Ibrahim (2009) who found a set favouring the longest option on difficulty level of multiple-choice tests. The finding also corroborates Larkins and Shaver (2001) result when they reported that test consisting of extra material with cueing words or specific determiners did effectively lesson the difficulty of the test. However, the validities and internal consistencies for tests consisting of items with or without cues were not significantly different.

In the aforementioned studies, it was generally found that test consisting of items with complete sentence stems were less difficult or produced higher mean scores than tests containing items with incomplete sentence stems. The completeness or incomplete sentence stems did not seem to have a great positive or negative effect on the internal consistency or validity of the multiple-choice tests. Although, there was equal number of correct answers in each position throughout the test, the fact that an answer is in a certain position on one item may influence the test taker to respond in a certain way on successive items. It should be understood that the arrangement of alternatives in a randomized fashion does not prevent an individual's response from being influenced by his previous responses. Actually, the variance in results may be due to precisely such a set in the individual produced by the sequence effect.

V. Conclusion And Recommendations

Based on the findings obtained from the study, it can be concluded that tendency exists amongst examinees to favour certain positions among the alternatives in preference to others. Not only this, but also there does appear to be a tendency for examinees not to respond in a random manner as they tend to avoid repeating the same response position on consecutive items (for example, BBB or CCC). Hence, examinees tend to use backward-series response sequence (e.g. EDC) more often than would be expected by chance; there does not appear to be a corresponding phenomenon for forward series. Therefore, the following are recommended based on the findings of this study namely: (i) any analysis of learning achievement as measured by a large number of such questions administered to a large number of subjects should include consideration of choice-placement elements since this affects the results obtained; (ii) despite the fact that the present study had dealt exclusively with achievement testing, it may very well be true that other types of test employing this kind of question e.g. personality and attitude inventories show findings due in some measure to this position factor; and (iii) teachers should be familiar with the procedure involved in item analysis vis-à-vis test construction.

References

[1]. Afolabi, E. R. I. (2000). The changing of initial answers in true-false test: Discrepancy in sex, age and ability groups. IJOTRE, 5(1), 1-7.
[2]. Afolabi, E. R. I. (2012). Scoring of essay and objective tests. In E.R.I. Afolabi & O. O. Dibu- Ojerinde (Eds.), Educational Tests & Measurement. Ile-Ife, Nigeria: Obafemi Awolowo University Press.
[3]. Afolabi, E. R. I. & Ibrahim, A. (2009). Effects of positional response bias and correct response location on the difficulty level of multiple-choice questions. Unpublished Manuscript.
[4]. DeMars, C. E. (2010). Scoring multiple choice items: A comparison of IRT and classical polytomous and dichotomous methods. Center for Assessment and Research Studies, Harrisonburg: James Madison University.
[5]. Green, K. (2000). Multiply-choice and true-false reliability and validity compared. Journal of Experimental & Educational Review, 48, 42-54.
[6]. Herrmann-Abell, C. F. & DeBoer, G. E. (2011). Using distractor-driven standards-based multiple-choice assessments and Rasch modeling to investigate hierarchies of chemistry misconceptions and detect structural problems with individual items. Chemistry Education Research and Practice, 12, 184–192.
[7]. Ibrahim, A. (2010). The effect of positional response bias on the difficulty level of multiple- choice questions in English language. Lagos Education Review- A Journal of Studies in Education, 2(12), 165-182.
[8]. Klassen S. (2006). Contextual assessment in science education: Background, issues, and policy. Science Education, 90, 820-851.
[9]. Larkins, A. and Shaver, R. (2001). Multiple-choice models: The distractors are also part of the item. Journal of Educational Measurement, 26, 161-176.
[10]. Liu, X. and Boone W. J. (2006). Introduction to Rasch measurement in science education. In Liu, X. and Boone, W. J. (Eds.), Applications of Rasch measurement in science education. Maple Grove: Minnesota, JAM Press.
[11]. Markus, K. A. (2001). The converse of inequality argument against tests of statistical significance. Psychological Methods, 6, 147-160.
[12]. Osterlind, S. J. & Everson, H. T. (2009). Differential item functioning (Series: Quantitative applications in the social sciences; 161). California: SAGE Publications Inc.
[13]. Penfield, R. D. & Jimmy, T. (2008). A new response model for multiple-choice items. Paper presented at the annual meeting of the National Council on Measurement in Education, New York.
[14]. Rodriguez, M. C. (2005). Three options are optimal for multiple-choice items: A meta- analysis of 80 years of research. Educational Measurement: Issues and Practice, 10, 3-13.
[15]. Upadhya, B. & Singh, Y. K. (2008). Advanced educational psychology. New Delhi: APH Publishing Corporation.
[16]. Wesman, M. H. (2000). Measurement essentials. Wilmington: Delaware. Wide Range Inc.

Biography

Dr Abdul-Wahab IBRAHIM, is a Psychometrician and Full-Time Lecturer at Department of Education, Faculty of Education, Sule Lamido University, Kafin-Hausa, Jigawa State, Nigeria.