An Open Access Journal
ISSN: 10683844

Multicultural Education

Research Article

Homepage: www.MC-caddogap.com

NEXUS BETWEEN TEST-ANXIETY AND RESPONSE ABERRANCE PATTERNS IN MATHEMATICS TEST ITEMS

Ayanwale, Musa Adekunle

Department of Science and Technology Education, Faculty of Education, University of Johannesburg, Johannesburg, Auckland Park, 6000, South Africa.

A B S T R A C T

Examinees who skip questions or have prior knowledge of the test content will generate an inaccurate and invalid test score. Assurance that test scores reflect the examinee's abilities is challenging for test administrators. Thus, this study examined students' response aberrance in the unified mathematics test based on their test anxiety and mathematics ability. The research design used was a cross-sectional descriptive survey. Through a multi-stage sampling procedure, 550 subjects were sampled. Data were collected using the Mathematics Achievement Test (MAT) and Test Anxiety Inventory, which had reliability coefficients of 0.84 and 0.82. Using the person fit package implemented in R programming language version 4.0.1 software, response aberrance (W* and B*) indices for each student were computed. A 2 x 2 x 2 factorial multivariate analysis of variance (MANOVA) was used to analyse the data descriptively and inferentially. Results showed that high school students have a response aberrance in mathematics with a mean of 0.96 and 4.83, which are greater than 0.50 for both aberrance indices. The study found a significant interaction between student ability and test anxiety, with a significant main effect on student ability and test anxiety. The results indicated a significant MANOVA, and a follow-up analysis of the variance test showed a statistically significant difference in the response aberrance as measured by W* and B* on students' ability, with low ability having the highest mean scores. The study also indicated a significant difference in the response aberrance as measured by B* on test anxiety, with high anxiety having higher mean scores than low anxiety. The study concluded that high school students' ability and test anxiety affect the severity of response aberration in Mathematics.

A R T I C L E I N F O

Keywords:
Test anxiety, aberrance response, within ability (W*), beyond ability (B*), mathematics achievement test.

Article History:
Received: 24th Jun 2022
Accepted: 19th Jul 2022
Published: 26th Aug 2022

© 2020 The authors. Published by CADDO GAP Press USA. This is an open access article under the Creative Commons Attribution Non-Commercial 4.0

1. INTRODUCTION

Learners are exposed to new ideas through curriculum outlines. Testing and other assessment procedures are crucial to assessing mastery of these outlines. Over the years, multiple-choice tests have increasingly been used to grade student mastery. Kim and Moses (2018) found that students developed response patterns rather than responding sequentially to keys. It is possible for these response patterns to accommodate aberrations in the field of testing. Aberrance refers to students who demonstrate inconsistent knowledge of some items in an examination compared to how they perform on other items (Jia et al., 2019). Questions that are more difficult are answered correctly, but those that are easier are not. In test-taking behaviour, aberrance refers to responses that differ from the statistical models used to predict normal behaviour (Meijer et al., 2008). Several factors may contribute to aberrances, such as irregular testing, poor preparation, illness, lack of motivation, guessing,

differential preparation, and cheating (Meijer & Sijtsma, 2001). An abnormal pattern of test responses could indicate an irregular testing process. Occasionally, some students receive assistance with some questions without receiving assistance with others. Several modes of instruction may have been used. Modes of assistance and modes of independent work are both included here. Using Caveon (2005) bimodal aberrance statistics, one can assess the test-taking ability levels of some examinees based on their "within ability aberrance" (W*) and "beyond ability aberrance" (B*). Within ability aberrance occurs when an examinee wrongly answers an item on an easy items, and beyond ability aberrance occurs when they correctly answer a difficult items (Adediwura & Agunbiade, 2015; Jia et al., 2019)

Several person-fit indicators can be used to identify abnormal response patterns (Karabatsos, 2003). In many of these techniques, the observed response patterns are compared to the expected outcomes as defined by a particular model. Based on Meijer and Sijtsma (2001), when observed item response patterns do not conform to what is implied by the model, person misfit is observed. Nevertheless, Meijer and Sijtsma (2001) argues that traditional test theory and item response theory define expectation differently in response patterns and measure deviation from that expectation differently.

Aberrant Response Patterns

Consider a scenario in which M examinees take a test with N items. It is possible to summarize test data by forming a matrix with M rows and N columns of 1s and 0s representing correct and incorrect answers, respectively. There are two methods for examining the responses matrix: either column vectors that illustrate the responses of M examinees item by item or row vectors that illustrate the responses to N items by individual examinees (Jia et al., 2019). When examining the responses of groups of examinees to a particular item, the first approach often considers how items differ based on their ethnicity (e.g. race or gender) (Hulin et al., 1983 cited in Chen, 2004). In addition, an item response theory (IRT) or non-parametric IRT model can be used to examine an examinee's observed response vector to the N items, comparing that response vector to the expected response vector (Ackerman, 2010; Bichi & Talib, 2018; Cai et al., 2016). Low-ability examinees who miss many easy questions show deviations from the expected pattern. It is impossible to measure a person's ability based on an aberrant response pattern. According to this study, each examinee's response vector is examined as part of the second approach. Research has discussed many types of aberrant response patterns. Based on Wright and Phillips (1979) cited in Adediwura and Agunbiade (2015) classification, there are four types. A "fumbling" pattern indicates difficulty getting started in a test since incorrect responses cluster at the beginning of the test. Secondly, there is the "sleeping" pattern, in which examinees miss the easiest questions by sleeping during the easy section of the test. Third, the "guessing" pattern results in the hardest question being correctly answered by chance. The method of generating a "constant response set" involves selecting the same option for all items; say, B. "Plodding" is the last anomalous response pattern. The individual who responds in this manner works slowly and cautiously on each item and does not proceed to the next until the present item has been answered correctly (Lee et al., 2014). IRT models assume that responses are stochastic, so presenting items in increasing difficulty will result in an overly ideal response pattern (Kim & Moses, 2018; Seock-Ho, 2004).

An examinee's true ability can be inaccurately estimated when they suffer from person-related disturbances (Smith, 2000). Among the disturbances are "start-up test anxiety", "copying from another person", "guessing to complete or randomly," "disinterest or boredom," "sleeping," "illness," "fatigue" (usually introduced by long tests) and "external distractions". Evidence suggests that several kinds of behaviour might explain some of the differences between correct and incorrect responses (Lee et al., 2014). The behaviour is linked to a limited understanding of machine-scored tests, resulting in incorrect alignment of responses. This behaviour is also related to an individual's ability to interpret easy test items with "extreme creativity." Individuals with high levels of intelligence may perceive easy items as too trivial or simplistic to be true, so they interpret them with deeper understandings and eventually choose the incorrect answer. A third behaviour involves cheating caused by "exposed tests" or "lack of vigilance". Occasionally, low-ability examinees may peek at their neighbours' examination script during a test to get a heads-up about their test. For example, each cheater will generate a response pattern with a combination of locks of nearly correct responses and blocks of nearly incorrect responses. An example of a test with 10 items of increasing difficulty is shown in Table 1 to illustrate the aberrant response patterns described above. A number-correct score may not be adequate to measure an individual's trait level when different test behaviours result in the same overall score. In this study, aberrant responses are examined, which is related to test anxiety, ability, and gender, respectively.

Table 1. Sample aberrant responsible Patterns on a 10-items test.

Test Behaviour	*Response Pattern Easy➡ Hard*	*Total Score*	*Interpretation*
Fumbling/Start-up Test Anxiety/Sleeping	**000**|1110|110	5	Not sure how to get started.
Lucky guessing	111|0100|00**1**	5	The ability to obtain some items beyond what is expected.
Plodding	111|11**00**|**000**	5	Reached items with high accuracy.
Extreme Creativity	**00**1|1110|100	5	Considering easy items"too simple to be true".
Cheating(Copying or Pre-Knowledge)	110|0000|**111**	5	Answering hard items by copying from neighbours.
Alignment Errors	111|001|010	5	A mismatch between responses and items, such as skipping one item.
Deficient Ability	111|**0000**|110	5	Medium difficulty items contain unfamiliar content.

Student achievement is affected by several factors, including motivation, quality of school instruction, subject matter knowledge, family socioeconomic status, and school factors. However, test anxiety, gender, and ability level are equally important. Based on Adolf (2011), implied tasks or lessons were used to systematically observe suspected educational or psychological traits or attributes. The test aims to determine a student's skill, achievement, aptitude, attitude, interest, etc., through a series of exercises or questions. As a result, tests have emerged as potent anxiety-producing stimuli, and test anxiety has become a widespread problem in our society (Spielberger & Vagg, 1995 cited in Duraku, 2017).

Test anxiety and response aberrance

Today's society is oriented toward testing (Karabatsos, 2003; Kim & Moses, 2018). Test scores play a major role in shaping our educational lives in many ways. Exam results determine admission to higher education institution in Nigeria. Consequently, today's society has a significant problem with test anxiety, which is a major source of anxiety due to tests (Wang, 2003). Test anxiety is a reaction characterised by cognitive and affective reactions that are likely to negatively impact performance on a test or during an evaluation situation (DordiNejad et al., 2011; Duraku, 2017). According to literature studies on test anxiety, it is a psychological problem that interferes with the educational process and the mental health of American adolescents (Jia et al., 2019; Meijer & Sijtsma, 2001; Wigfield & Eccles, 1989) as well as high school students (Wang, 2003). Academic performance is negatively affected by cognitive test anxiety (Cassady & Johnson, 2002). Whenever people of all ages have to evaluate, assess, and grade their abilities, achievements, or interests, Cohen (2004) warns that performance anxiety can adversely affect them. Anxiety can also contribute to underachieving since it impairs test performance (Alemu, 2020; Wigfield & Eccles, 1989). There is evidence that test anxiety impacts learning and performance in developed countries (Alemu, 2020; DordiNejad et al., 2011). In school and college, test anxiety negatively influences achievement (Alemu, 2020; DordiNejad et al., 2011). The number of studies of this kind in Nigeria is relatively low. Due to this reason, high school students were selected. A growing body of research has been conducted on test anxiety in recent decades. Only a few studies have examined the appropriateness of examinees' response patterns when faced with test anxiety instead of their total test scores (Chen, 2004). During an examination, high-test-anxious candidates can get distracted, have difficulty understanding instructions, and have trouble recalling information (Duraku, 2017). Therefore, these symptoms may impair an examinee's ability to solve problems, negatively affecting his or her overall test performance (Duraku, 2017). Examinee performance at the item level is determined by person-fit or aberrance assessment (Kim & Moses, 2018). Test anxiety seems to be the most disruptive factor in school and other settings where testing is administered based on the above assertions (Alemu, 2020). Therefore, this study investigated the effects of test anxiety on students' response patterns in consideration of the possible consequences of test anxiety on performance. It is also important to consider the student's gender.

Gender and response aberrance

As a cultural construct, gender distinguishes the roles, behaviours, mental and emotional characteristics of males and females. In academics and other areas of human resource development, the concept of gender does

not support or suggest a dominant position for males over females. However, it places a high value on equity and equality in recognizing, developing, and utilising the skills and abilities of both sexes. Females and males may experience different levels of test anxiety, which could explain gender differences in test anxiety. Nonetheless, females seem to be more emotional, causing them to score higher on test anxiety. In contrast to their female counterparts, male individuals exhibit a significantly lower cognitive test-anxiety score than female individuals because of their high level of risk-taking behaviours; however, this may not imply higher or lower academic performance. It is likely that gender has a significant influence on students' anxiety levels, which could lead to an aberration in the response pattern, leading to academic failure (Chen, 2004). The number of male aberrant responses is also consistently higher than the number of female aberrant responses. By randomly gridding responses or systematically marking them, Jing (2004) examined if students might lack motivation for the test based on National Assessment of Educational Progress (NAEP) data. Male students are more likely to mark their responses systematically than female students are.

Considering this literature, Schmitt et al. (1999) hypothesised that misfit measures might be associated with gender. A study was conducted to analyse the gender, race, and aberrance indexes. In both cognitive and personality tests, men produced more irregular responses, which indicates they are more prone to producing irregular responses. Most associations are eliminated when conscientiousness is taken into account. The motivation for taking the test was irrelevant. It is possible that carelessness is the cause of male misfits.

Essentially, academic ability is valuable to the learning process. The most common form of measurement seems to be examinations and continuous assessments. There was a correlation between individual differences in academic performance or ability and individual differences in intelligence and personality (Alemu, 2020; DordiNejad et al., 2011). Student item responses in the classroom may be affected by the working climate in a positive or negative way (Duraku, 2017). When you respond abnormally, you may also have an aberrant response pattern. A test can be difficult for some examinees to complete because they have difficulty responding to the items. It is also possible for some examinees to have trouble starting the test, which can result in them entering incorrect responses from the start. When shading multiple-choice questions on the answer sheet, it is also possible for examinees to make mistakes. The response pattern of the examinees in secondary school Mathematics provides information that cannot be derived from the total score. Response aberrations can lead to the misuse of information (Jia et al., 2019; Wang et al., 2019).

Some students' response patterns are aberrant for reasons that research has not fully explained. Meijer and Sijtsma (2001); Wang et al. (2019) suggested conducting more studies involving background variables to gain additional information about the misfit type. An examination of test anxiety levels and gender was conducted in the present study. The overarching research questions and hypotheses are: What are the patterns of distribution W* and B* scores in mathematics achievement tests? What is the frequency of aberrant responses in the Mathematics Achievement Test (MAT) among high school students? There is no significant main effect of student ability on response aberrance as measured by W* and B*, and there is no significant main effect of test anxiety on response aberrance as measured by W* and B*.

2. MATERIALS AND METHODS

Research design

A cross-sectional descriptive survey research design was adopted in this study. For this study, ten government-owned high schools were selected in one State in Nigeria.

Participants

The three existing senatorial districts in Ondo State, Nigeria, were stratified. A purposive sampling technique was used to select the sample for the study. Specifically, the district with a high number of government-owned high schools was chosen for the study. A total of ten schools from the chosen district were also randomly selected, and to give the study a more robust approach, a class of intact students in grade 11 in the mathematics subject were used as participants. Thus, five hundred and fifty students (57.7% males and 44.3% females) from grade 11 were sampled during the 2021/2022 academic year, with a mean age of 14.6 years old. More so, throughout their educational journeys, all of the participating students were taught primarily in English as the language of academic instruction. Because of this, there was an expectation of good proficiency in reading and mathematics for the participants.

Measures

A structured questionnaire measuring the Test Anxiety Inventory (TAI) adapted from Wang (2003) was used for data collection. TAI is a self-report scale that measures test anxiety as a situation-specific personality trait. Consequently, it is chosen from various instruments that measure test anxiety. The results of a TAI assessment reflect a long-term, inherent personality characteristic. Respondents were asked to describe how they are generally affected by anxiety symptoms before, during, and after examinations using a 20-items questionnaire. Examples of items on the scale included: 'Before or during exam, I find myself thinking about how much brighter some of the other students are'; 'Worrying about how well I will do interferes with my preparation and performance on tests'; 'I never seem to be fully prepared to take tests'; 'I have a hollow, uneasy feeling before taking a test'; 'If my score is low, my parents will be very disappointed.'; 'When taking a test, my emotional feelings interfere with my concentration'; 'My anxiety about tests makes me want to avoid preparing fully, and this just makes me more anxious'; 'I do not feel confident and mentally relaxed before a test' and 'I often fell physically panicky when I have to take a really important test'. Likert scale was used for this instrument, where a score of 1 equals strongly disagree, a score of 2 equals strongly disagree, a score of 3 equals agree, and a score of 4 equals strongly agree. Responses to all questions are scored 1 through 4 except for items 1 and 10. These items are scored in reverse order of 4, 3, 2, and 1. A score of 20 to 80 was obtained on the TAI statements. In general, a low score (20-39) indicates that the participants are not anxious about test taking, while a high score (40-80) indicates that they are anxious about it.

In addition, the second instrument was based on the questions on the unified grade 11 examination in Ondo State. It is believed that the Examination Department of the Ondo State Ministry of Education validated and standardised the questions. As a result, the test items were already valid and reliable. Students' mathematics skills are assessed using this instrument. The mathematics achievement test consists of 50 multiple-choice items with five alternatives (A-E). The optical mark recorder sheets scored students' responses in two ways: either correctly, which is coded as 1, or incorrectly, which is coded as 0. The total score was calculated to determine the student's ability. There were 50 possible scores. The low scores on the test (1-24) indicate that the participants had low abilities, whereas high scores on the scale (25-50) indicate that they had high abilities.

Validity and reliability of the instrument

The research instrument used for data collection in this study was an adapted questionnaire. Thus, there is a need to validate the instrument and ensure its validity and reliability. The researcher ensured the content validity of the questionnaire by having it carefully examined by experts in psychology and measurement and evaluation. The researcher created the questionnaire using a Google form, and 20 (12 males and 8 females) panellists examined the essentiality and non-essentiality of the statements on the scale between 10-30th April 2022. Responses gathered from the 20 panellists were subjected to a content validity ratio (CVR) and, in turn, established a content validity index (CVI) (Lawshe, 1969) of 0.64. Also, the responses of the sampled students outside the study were subjected to a reliability analysis to determine the instrument's internal consistency. The ordinal alpha reliability coefficient of 0.82 was returned and implemented using the psych package of R programming language version 4.0.1. Also, the Kuder Richardson 20 (Kr20) reliability estimate for the mathematics achievement test yielded 0.84 implemented using the 'validateR' package of R programming language version 4.0.1 software.

Data collection procedures

The data was collected from students in government-owned high schools in Ondo central senatorial district, Nigeria, between 5-30th June 2022. Students in grade 11 were intactly used, and they responded to the questionnaire and mathematics achievement test in the English language. Respondents were provided with printed instruments, and four research assistants assisted with onsite administration. Prior to participating in the study, all participants were informed about its purpose and consented.

Data analysis

The independent and dependent variables were analysed using descriptive statistics to determine central tendencies (mean) and dispersion (standard deviation). Students' ability (low vs high) and their test anxiety (low vs high) were independent variables, while aberrance responses B* (beyond-ability) and W* (within-ability) were dependent variables. Inferential statistics using the 2 × 2 x 2 factorial MANOVA were generated to determine the differences in the participants' test anxiety and ability vis-à-vis their B* (Beyond-ability) and W* (Within-ability) in the mathematics achievement test. Using a MANOVA, the assumptions of multivariate normality, homogeneity of variance, and the quality of the variance error of the dependent variables were examined. Due to the two dependent variables examined in this study, MANOVA was considered suitable for analysing the data. MANOVAs are suitable for analyses involving several factors (each with two or more levels) and multiple dependent variables, as Grice and Iwasaki (2007) explained. Analysis of the data generated from the two instruments was done using SPSS, version 28, and R, version 4.0.1, a programming language for statistical computing.

Ethical Declarations

The study obtained informed consent from all participants by completing a consent form. Respondents had no requirement to provide their names, so all data collected was anonymous.

3. RESULTS

What is the distribution pattern of W and B* scores in the mathematics achievement test?*

Shapiro-Wilk test and normal P-P plot were used to examine the multivariate normality assumption. Firstly, the Shapiro-Wilk test was performed on the dependent variables Beyond-ability and Within-ability to determine their univariate normality. Because the dataset was less than 2,000, Shapiro-Wilk was used to determine the normality of the distribution. According to Shapiro-Wilk test, data for B* and W* have normally distributions (w = 0.64, p = 0.00 and w = 0.72, p = 0.01). Furthermore, a normal P-P plot of regression-standardised residuals (see figures 1 and 2) confirmed Shapiro-Wilk test results.

Furthermore, the MANOVA assumptions were tested using the Mahalanobis distance, which gave a maximum value of 9.36. The value of 9.36 was less than the critical value of 11.69 in the Mahalanobis distance; consequently, it was assumed that the multivariate normality assumption was met. Furthermore, the Box's M test value of 22.38 was associated with a p-value of 0.07, considered non-significant (Huberty & Petoskey, 2000). As a result, the MANOVA assumed that the covariance matrices between the groups were equal. Aside from that, it also demonstrated that the error variance across groups was the same. A linear assumption was also established. The Pearson product-moment correlation (r = 0.39), which is low, shows that the dependent variables correlate with one another. Though, there was a significant positive correlation between the variables (p = 0.01), indicating the MANOVA is suitable (see table 3).

Figure 1: Normal P-P Plot for Beyond-ability/surprise index

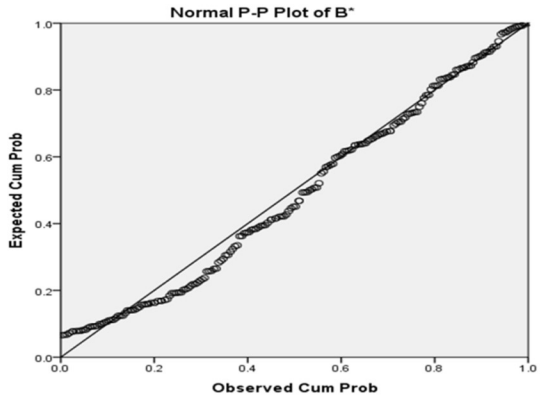
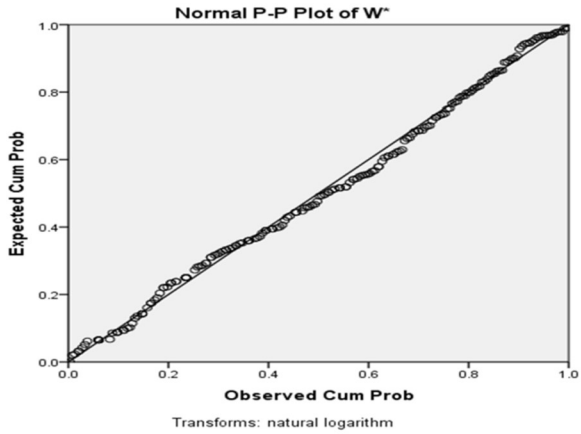


Figure 2: Normal P-P Plot for Within-ability/Caution index



The implication of figures 1 and 2 was that data collected for this study were normally distributed since most of the data points were clustered along the diagonal line. If the data points stray from the line in an obvious non-linear fashion, then the data set are not normally distributed. However, one of the variables was not normal (W*), hence, transformed using natural logarithm. Therefore, the researcher concluded, based on the information depicted by normal P-P plot of regression standardised residual, that the data did not substantially violate normality assumption.

Also, Box's M test analysis was conducted to ascertain that the data did not violate the assumption of homogeneity of variance (see table 2).

Table 2. Box's test of equality of covariance matrices

Box's M	*F*	*df1*	*df2*	*Sig.*
22.38	3.23	21	10373	0.07

Table 2 remarked that there was no significance difference. This indicates that the assumption of homogeneity of variance-covariance was not in any way violated, if otherwise, there is a significant difference ($p < 0.05$). Another one is the linearity assumption; this was also assessed using a correlation matrix between the dependent variables (W* and B*). This provides evidence of linearity but the correlation coefficient must not be too high that is (ryy< 0.85) (see table 3).

Table 3. Pearson correlation coefficients between W* and B*

Dependent Variables		*W* (Within-ability)*	*B* (Beyond-ability)*
W (Within-ability)*	*Pearson Correlation*	1	0.39**
	Sig. (2-tailed)		0.01
	N	550	550
B (Beyond-ability)*	Pearson Correlation	0.39**	1
	Sig. (2-tailed)	0.01	
	N	550	550

** Correlation is significant at the 0.05 level (2-tailed).

Table 3 revealed that the two dependent variables were linearly related. The correlation coefficient was low (0.39), which was substantially less than the 0.85 cutoffs. Nevertheless, it was statistically significant. Therefore, the data did not violate the assumption of linearity. Hence, the dataset satisfied the assumptions underlying the statistical tool, used for the analysis of data.

What is the frequency of aberrance response in the Mathematics Achievement Test (MAT) among Ondo State secondary school students?

Students' responses to Mathematics Achievement Test were scored and subjected to analysis. Person Fit package implemented in R programming language version 4.0.1 was used to estimate students' aberrance index scores. These include two aberrance indices; within-ability or caution index (W*) and beyond-ability or surprise index (B*). This was estimated for each examinee to measure their response aberrance to the Mathematics Achievement Test. It was depicted from the results that larger percentage of examinees had aberrance value higher than the benchmark of 0.5. This is according to D'costa (2013), which remarked that value above 0.5 reveals that there is evident of aberrance in the response pattern (see table 4 and figure 3).

Table 4. Descriptive statistics of examinees' response aberrance

Variables	*N*	*Minimum*	*Maximum*	*Mean*	*Std. Deviation*
W (Within-ability)*	550	0.04	4.99	0.96	0.87
B (Beyond-ability)*	550	0.54	13.57	4.83	2.86

Figure 3. Pattern of response aberrance in mathematics achievement test

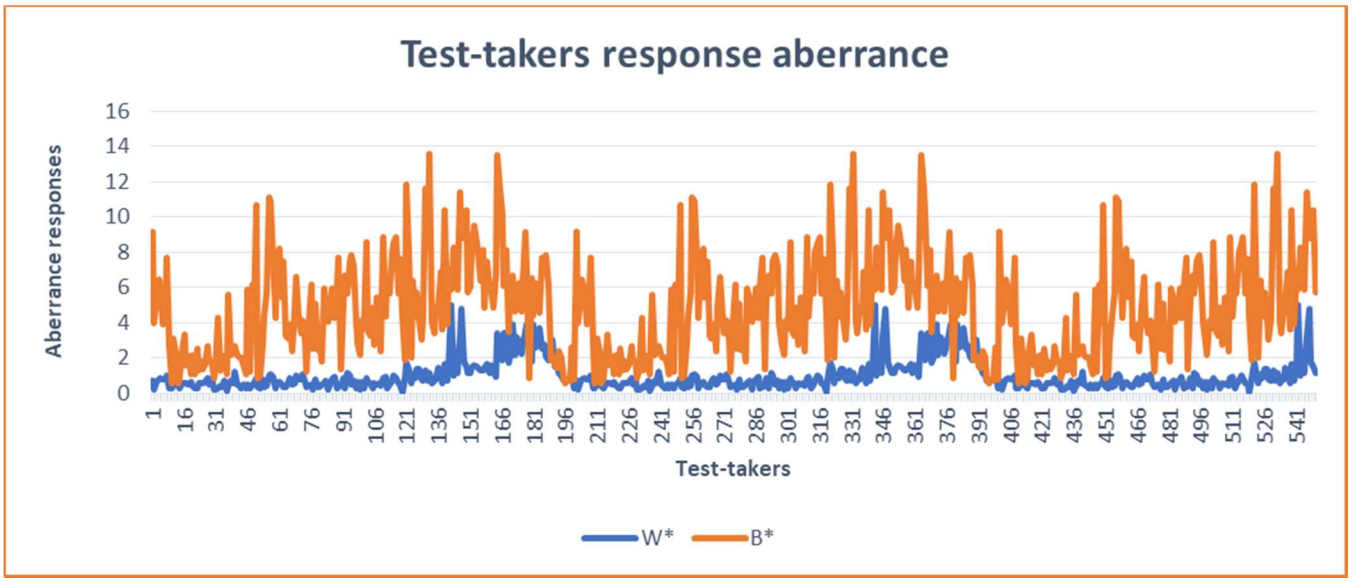


Table 4 and figure 3 remarked that within-ability (W*) had a mean value of 0.96 and a standard deviation of 0.87, while beyond-ability (B*) had a mean value of 4.83 and a standard deviation of 2.86. Also, it was revealed that the mean for beyond-ability (B*) had a higher value than the mean for within-ability (W*). However, the mean value for beyond ability (B*) and the within ability (W*) aberrance indices were above the 0.5 cutoffs. A larger percentage of examinees got the difficult/hard items correctly, while some examinees missed the easy items. The implication is that the examinees' pattern of responses deviated substantially from the expected. Consequently, the researcher concluded that there is evidence of aberrance in the examinees' pattern of response to the mathematics achievement test.

There is no significant main effect of student ability on response aberrance as measured by W and B**

Data obtained were analysed using MANOVA to test for significance. In conducting MANOVA, the overall multivariate hypothesis must be tested first. Importantly, it is established that when two independent variables are present and the test statistics for the multivariate are unequal, Wilks' lambda distribution cannot be used to assess significance; instead, Pillai's trace should be used (Tabachnick & Fidell, 2007). However, Wilks' Lamda value is regarded as one of the most reported statistics in the literature (Pallant, 2011). In this study, the significance was therefore tested with Pillai's trace. Thus, if the null hypothesis is not rejected (that is no significant difference), it is ideal to stop the interpretation of the analysis at this point and remark that

the independent variables have no significant difference from the two dependent variables. Nonetheless, if it is evident that the multivariate test is significant, the researcher will then continue to assess which independent variables influence the dependent variable. This was achieved by conducting a series of univariate analyses of variance on individual dependent variables. Table 5 presented the overall multivariate hypothesis as follows.

Table 5: Summary of multivariate tests

Effect		*Value*	*F*	*Hypothesis df*	*Error df*	*Sig.*	*Partial Squared*
Intercept	Pillai's Trace	0.713	677.333b	2	545	0.01	0.727
	Wilks' Lambda	0.287	677.333b	2	545	0.01	0.727
	Hotelling's Trace	2.486	677.333b	2	545	0.01	0.727
	Roy's Largest Root	2.486	677.333b	2	545	0.01	0.727
Ability	Pillai's Trace	0.168	54.918b	2	545	0.01	0.146
	Wilks' Lambda	0.832	54.918b	2	545	0.01	0.146
	Hotelling's Trace	0.202	54.918b	2	545	0.01	0.146
	Roy's Largest Root	0.202	54.918b	2	545	0.01	0.146
Anxiety	Pillai's Trace	0.076	22.518b	2	545	0.01	0.078
	Wilks' Lambda	0.924	22.518b	2	545	0.01	0.078
	Hotelling's Trace	0.083	22.518b	2	545	0.01	0.078
	Roy's Largest Root	0.083	22.518b	2	545	0.01	0.078
*Ability * Anxiety*	Pillai's Trace	0.026	7.309b	2	545	0.01	0.014
	Wilks' Lambda	0.974	7.309b	2	545	0.01	0.014
	Hotelling's Trace	0.027	7.309b	2	545	0.01	0.014
	Roy's Largest Root	0.027	7.309b	2	545	0.01	0.014

a. Design: Intercept + ability + test anxiety + Ability * Test anxiety
b. Exact statistic

As reported in Table 5, there was a significant interaction between the students ability (low and high) and test anxiety (low and high) of the participants (Pillai's trace = 0.026, $F_{(2,545)} = 7.30$; $p < 0.05$, $\eta 2 = 0.014$). This implies that 1.4% of the variance in the combined dependent variables (DVs) of within-ability (W*) and beyond-ability (B*) could be explained by the interactive effects of students ability and test anxiety. Therefore, the results showed minimal effects.

To further reveal the significant main effects of the students ability and test anxiety of the participants, the results in Table 5 show a significant main effect for students ability (Pillai's trace = 0.168, $F_{(2,545)} = 54.918$; $p < 0.05$, $\eta 2 = 0.146$) and test anxiety (Pillai's trace = 0.076, $F_{(2,545)} = 22.510$; $p < 0.05$, $\eta 2 = 0.078$). This result could be seen as having a total of 14.6% and 7.8% variance in combined dependent variables (DVs) of within-ability (W*) and beyond-ability (B*), which could be explained by their ability and test anxiety, respectively. The obtained degree of variance in the combined dependent variables (DVs) of aberrance response was indicative of the gender influence on the mathematics test (Schmitt et al., 1999). Subsequently, a univariate analysis was conducted as a follow-up protocol to ascertain the differences between students ability (low and high) and test anxiety (low and high) across the dependent variables (see table 6).

Table 6. Tests of Between Subjects Effects

Source	*Dependent Variable*	*Type III Sum of Squares*	*df*	*Mean Square*	*F*	*Sig.*	*Partial Squared*
Corrected Model	W*	74.04a	3	24.68	39.08	0.01	0.177
	B*	340.17b	3	113.39	14.92	0.01	0.076
Intercept	W*	506.06	1	506.06	801.20	0.01	0.595
	B*	8150.94	1	8150.94	1072.34	0.01	0.663
Ability	W*	69.01	1	69.01	109.25	0.01	0.167
	B*	181.68	1	181.68	23.90	0.01	0.042
Anxiety	W*	7.72	1	7.72	12.22	0.01	0.022
	B*	116.79	1	116.80	15.37	0.01	0.027
*Ability * Anxiety*	W*	7.18	1	7.19	11.38	0.01	0.020
	B*	0.92	1	0.92	0.12	0.73	0.000
Error	W*	344.87	546	0.63			
	B*	4150.18	546	7.60			
Total	W*	929.81	550				
	B*	17302.5	550				
Corrected Total	W*	418.91	549				
	B*	4490.34	549				

R Squared = .237 (Adjusted R Squared = 0.210)
R Squared = .098 (Adjusted R Squared = 0.065)

It is noteworthy to state that for univariate analysis, Bonferroni adjustment reduces the likelihood of making a Type 1 error when multiple analyses are performed (Pallant, 2011). Thus, by dividing the original alpha level of 0.05 by 2 (number of dependent variables), a new alpha level of 0.025 is used. Table 6 revealed a statistically significant difference in the response aberrance displayed by the examinees as measured by W* ($F_{(1,546)} = 109.25$, $p < 0.025$, $\eta 2 = 0.167$), and B* ($F_{(1,546)} = 23.90$, $p < 0.025$, $\eta 2 = 0.042$) based on students' ability. Also, the Adjusted R Squared value of 0.210 and 0.065 indicated that examinees ability accounted for about 21.0% and 6.5% of the variation observed in the combined DVs of W* and B* of examinees in Mathematics respectively. Furthermore, figures 4a and 4b presented plot of estimated marginal mean for W* and B* for examinees' ability and test anxiety.

Figure 4a. Simple plot on interaction effects between student ability and test anxiety on within-ability (W*)

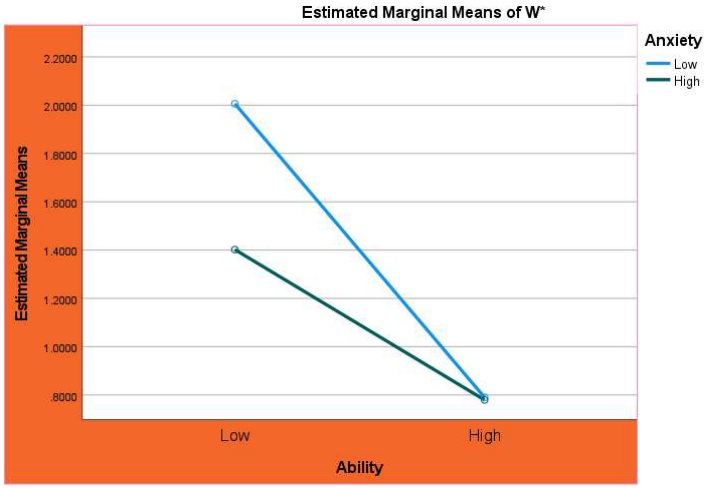
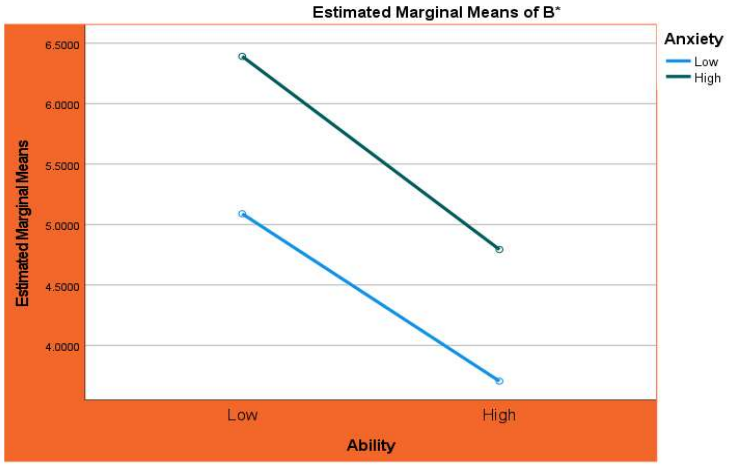


Figure 4b. Simple plot on interaction effects between student ability and test anxiety on within-ability (W*)



Figures 4a and 4b showed that the estimated marginal means of within-ability (W*) for examinees with low and high ability was (M =1.704 and M = 0.785), while estimated marginal means of beyond-ability (B*) values for examinees with low and high ability was (M = 5.741 and M = 4.249). This indicated that for within-ability (W*) examinees with low ability exhibited higher response aberrance. Meanwhile, in all, the W* and B* mean values were above 0.5 benchmark. More so, it was clear that the response aberrance of low ability examinees was more than that of high ability examinees as measured by W*. This indicated that the examinees missed some of the questions within the confine of their ability level. Also, the response aberrance as measured by B* for low ability examinees was more than that of high ability examinees which depicts that some examinees responded to some items that are beyond their ability level.

There is no significant main effect of test anxiety on response aberrance as measured by W and B**

Table 6 also remarked a statistically significant difference in the response aberrance displayed by the examinees as measured by W* ($F_{(1,546)} = 12.22$, $p < 0.025$, $\eta 2 = 0.022$), and B* ($F_{(1,546)} = 15.37$, $p < 0.025$, $\eta 2 = 0.027$) based on text anxiety. Also, the Adjusted R Squared value of 0.210 and 0.065 indicated that examinees ability accounted for about 21.0% and 6.5% of the variation observed in the combined DVs of W* and B* of examinees in Mathematics respectively. This implies that exam fever or tension before the examination contributed immensely to the aberrance response pattern of the examinees. Furthermore, figures 4a and 4b showed that the estimated marginal means of within-ability (W*) for examinees with low and high test anxiety was (M =1.398 and M = 1.091), while estimated marginal means of beyond-ability (B*) values for examinees with low and high test anxiety was (M = 4.397 and M = 5.593). This showed that for within-ability (W*) examinees with low test anxiety displayed greater response aberrance. Meanwhile, in all, the W* and B* mean values were above 0.5 benchmark. Also, it was evident from figure 4a that the response aberrance of examinees with low test anxiety was more compare to high examinees with test anxiety as measured by W*. This indicated that the examinees missed some of the easy items due to their level of anxiety.

There is no significant interaction effect of student ability and test anxiety on response aberrance as measured by W and B**

More so, the response aberrance as measured by B* for examinees with high test anxiety was higher than low test anxiety examinees, which depicts that examinees responded correctly to some difficulty items that are beyond their ability. Furthermore, a sloping plot was used to obtain a better understanding and clarity of the overall interaction pattern and to determine whether a significant relationship existed between participants' ability (low vs high) and test anxiety (low vs high). The graphs in Figure 4b do not show parallel slopes, while Figure 4a shows interactive slopes. As a result, only the within-ability (W*) had significant

interactions across students' ability and test anxiety. As shown in Figure 4b, there was no interaction between low-ability and high-ability participants regarding their test anxiety.

4. DISCUSSIONS

The importance of providing fair and accurate scores to examinees cannot be overstated. A test score can often be spuriously high or low due to aberrant responses, but this is often unavoidable in reality. First, the distribution pattern of the students' aberrance to the test showed that it followed a normal distribution. This allows us to choose a parametric method to analyse their responses. In addition, there is evidence of a positive relationship between beyond-ability and within-ability for the student, which infers those students are getting correctly items at a difficulty level above their ability and getting wrong items that are within the ability range that is the easy items. This position laid credence to the work of Kim and Moses (2016); Chen (2004), that this could mean there are some assistance received by the students in the course of responding to the items, and they are not worried about getting any item wrong, even if they are the easiest of the pack. Also, on the basis of simulated data sets, Huang (2002); D'Costa (1993b) reported that W* and B* appear to be positively correlated, and there is a curvilinear relationship between these two indices and the ability ratio. In contrast, using real data, D'Costa (1993a) found that the W* index is negatively correlated with the total score, while the B* index is positively correlated with the total score and W* and B* are negatively correlated with each other.

A higher level of beyond-ability (B*) further indicates the likeliness of students having assistance while responding to the test items, which distorted the actual ability level of the students. It was also discovered that students with a lower levels of anxiety have better mean scores on the test, which indicates that anxiety would have a negative relationship with student achievement on the test. This is according to the submission of Spielberger & Vagg (1995); Chen (2004) that students should be tested in a conducive environment when they are not going through any negative physical, mental or psychological situations. However, test anxiety has negatively impacted both school and college achievement according to (Alpert & Haber, 1960; Hill & Sarason, 1966; Wigfield & Eccles, 1989). Findings also showed that most of the lower achievers have higher mean scores than higher achievers, which explains high level of beyond-ability frequencies and further buttresses the possibility of assistance from external individuals outside of the original test takers. This argues the submission of Chen (2004) that low ability students demonstrated higher response aberrance than high ability students as measured by W*. The poor responds of low ability students to easy items according to the finding of Chen (2004) may be caused by their ability, poor preparation and plodding. It was also observed that lower achievers have higher mean value for both within-ability and beyond-ability students. More so, female examinees missed some questions within the confines of their ability level. Also, the response aberrance as measured by B* for low ability examinees was more than that of high ability examinees, which depict that some examinees responded correctly to some items beyond their ability level. This position disagree with Jing (2004) that males produced more irregular responses for both cognitive and personality tests, which indicates that males produce more irregular response patterns compared to their female counterpart.

5. CONCLUSION

From this study, it has been revealed that rendering assistance for students during assessment can only distort the student's true ability but can be revealed through statistical analysis. In addition, anxiety can negatively affect the level of student achievement, which will pass wrong impression about the level of student ability. There should be awareness on how to reduce the student anxiety before the exam so as not to have negative impact on their response pattern.

6. ACKNOWLEDGMENT

The author is very grateful to all the students and teachers who participated in this study. Also, the peer reviewers are acknowledged for their time and effort in improving the quality of published article.

References

1. Adolf Obalum, N. (2011). Test development: A pragmatic approach to item generation. A Paper Presented at a Capacity Building Workshop, Minna, Nigeria.
2. Ackerman, T. A. (2010). The Theory and Practice of Item Response Theory by de Ayala, R. J. Journal of Educational Measurement, 47(4), 471–476. https://doi.org/10.1111/j.1745-3984.2010.00124.x
3. Adediwura, A. A., & Agunbiade, V. O. (2015). The Prevalence and Effect of Mathematical Ability on Response Aberrance among Secondary School Students in Osun State. American International Journal of Contemporary Research, 5(4). www.aijcrnet.com
4. Alemu, B. (2020). The relationship between test anxiety and academic achievement of grade ten students of Shirka Woreda, Oromia Regional State, Ethiopia. African Educational Research Journal, 8(3), 540–550. https://doi.org/10.30918/aerj.83.20.097
5. Bichi, A. A., & Talib, R. (2018). Item Response Theory: An Introduction to Latent Trait Models to Test and Item Development. International Journal of Evaluation and Research in Education (IJERE), 7(2), 142. https://doi.org/10.11591/ijere.v7i2.12900
6. Cai, L., Choi, K., Hansen, M., & Harrell, L. (2016). Item Response Theory. In Annual Review of Statistics and Its Application (Vol. 3, pp. 297–321). Annual Reviews Inc. https://doi.org/10.1146/annurev-statistics-041715-033702
7. Cassady, J. C., & Johnson, R. E. (2002). Cognitive test anxiety and academic performance. Contemporary Educational Psychology, 27(2), 270–295. https://doi.org/10.1006/ceps.2001.1094
8. Caveon. (2005). Texas Assessment of Knowledge and Skills Test Administration. Pilot Analysis of the Spring. https://www.google.com/search?q=Caveon+(2005).Texas+Assessment+of+Knowledge+and+Skills+Test+Administration.+Pilot+Analysis+of+the+Spring.&oq=Caveon+(2005).Texas+Assessment+of+Knowledge+and+Skills+Test+Administration.+Pilot+Analysis+of+the+Spring.&aqs=chro
9. Chen, J. (2004). Effect of Test Anxiety, Time Pressure and Ability on Response Aberrance. Unpublished Phd Dissertation. Ohio State University.
10. Cohen, A. (2004). Test anxiety and its effect on the personality of students with learning disabilities. Retrieved April 8, 2014, from http//www.thefreelibrary.com
11. D'Costa, A. G. (2013). The validity of W, B and Sato Caution Indexes. Paper presented at the Seventh International Objective Measurement Conference, Atlanta.
12. DordiNejad, F. G., Hakimi, H., Ashouri, M., Dehghani, M., Zeinali, Z., Daghighi, M. S., & Bahrami, N. (2011). On the relationship between test anxiety and academic performance. Procedia - Social and Behavioral Sciences, 15, 3774–3778. https://doi.org/10.1016/j.sbspro.2011.04.372
13. Duraku, Z. H. (2017). Factors Influencing Test Anxiety Among University Students. The European Journal of Social & Behavioural Sciences, 18(1), 69–78. https://doi.org/10.15405/ejsbs.206
14. Grice, J. W., & Iwasaki, M. (2007). A truly multivariate approach to MANOVA. Applied Multivariate Research, 12 (3), 199–226. https://doi.org/10.22329/amr.v12i3.660
15. Hill, K. T. & Wigfield, A. (1984). Test anxiety: A major educational problem and what can be done about it. Elementary School Journal, 85, 105-126.
16. Huang, T. W. (2002). The Power of the Beyond-ability Surprise Index and the Within-ability Concern Index for Detecting Person/Item Aberrances: A Monte Carlo Study. Unpublished doctoral dissertation, The Ohio State University.
17. Huberty, C. J., & Petoskey, M. D. (2000). Multivariate analysis of variance and covariance. In Tinsley, Howard EA, Steven D. Brown, (Eds.), handbook of applied multivariate statistics and mathematical modeling (pp. 183–208). Academic Press
18. Hulin, C. L., Drasgow, F., & Parsons, C. K. (1983). Item Response Theory. Homewood, IL: Dow Jonse-Irwin.
19. Jia, B., Zhang, X., & Zhu, Z. (2019). A short note on aberrant responses bias in item response theory. Frontiers in Psychology, 10(JAN), 43. https://doi.org/10.3389/fpsyg.2019.00043
20. Karabatsos, G. (2003). Comparing the Aberrant Response Detection Performance of Thirty-Six Person-Fit Statistics. Applied Measurement in Education, 16(4), 277–298. https://doi.org/10.1207/S15324818AME1604_2
21. Kim, S., & Moses, T. (2018). The Impact of Aberrant Responses and Detection in Forced-Choice Noncognitive Assessment. ETS Research Report Series, 2018(1), 1–15. https://doi.org/10.1002/ets2.12222
22. Lee, P., Stark, S., & Chernyshenko, O. S. (2014). Detecting aberrant responding on unidimensional Pairwise

preference tests: An application of lz based on the Zinnes–Griggs ideal point IRT model. Applied Psychological Measurement, 38(5), 391–403. https://doi.org/10.1177/0146621614526636

23. Meijer, R. R., Egberink, I. J. L., Emons, W. H. M., & Sijtsma, K. (2008). Detection and validation of unscalable item score patterns using item response theory: An illustration with harter's self-perception profile for children. Journal of Personality Assessment, 90(3), 227–238. https://doi.org/10.1080/00223890701884921
24. Meijer, R. R., & Sijtsma, K. (2001). Methodology review: Evaluating person fit. In Applied Psychological Measurement (Vol. 25, Issue 2, pp. 107–135). https://doi.org/10.1177/01466210122031957
25. Schmitt, N., Chan, D., Sacco, J. M., McFarland, L. A., & Jennings, D. (1999). Correlates of person fit and effect of person fit on test validity. Applied Psychological Measurement, 23(1), 41–53. https://doi.org/10.1177/01466219922031176
26. Seock-Ho, K. B. F. B. (2004). Item Response Theory: Parameter Estimation Techniques. In Biometrics 50(3). Marcel Dekker. https://doi.org/10.2307/2532822
27. Smith, R. M. (2000). Fit analysis in latent trait measurement models. Journal of Applied Measurement, 1(2), 199-218.
28. Spielberger, C. D., & Vagg, P. R. (1995). Test Anxiety: A Transactional Process Model. In C.D. Spielberger, & P. R. Vagg (Eds.), Test Anxiety: Theory, Assessment, and Treatment (pp. 3-14). Washington DC: Taylor & Francis.
29. Wang, C. (2003). The reliability and validation of Chinese version Test Anxiety Inventory in College Students. Chinese Journal of Clinical Psychology, 11(1), 69-70.
30. Wang, C., Weiss, D. J., & Su, S. (2019). Modeling response time and responses in multidimensional health measurement. Frontiers in Psychology, 10(JAN). https://doi.org/10.3389/fpsyg.2019.00051
31. Wigfield, A., & Eccles, J. S. (1989). Test Anxiety in Elementary and Secondary School Students. Educational Psychologist, 24(2), 159–183. https://doi.org/10.1207/s15326985ep2402_3
32. Wright, G. N., & Phillips, L. D. (1979). Personality and probabilistic thinking: An exploratory study. British Journal of Psychology, 70(2), 295–303. https://doi.org/10.1111/j.2044-8295.1979.tb01686.x