

DETERMINATION OF ABERRANT EXAMINEES' RESPONSES OF 2018 AND 2019 MAY/JUNE BIOLOGY OBJECTIVE QUESTIONS OF NATIONAL BUSINESS AND TECHNICAL CERTIFICATE EXAMINATION

OHIOZUA, E. N. & OMOROGIUWA, O. K.

nnebuio.go.ohiozua@uniben.edu, kingsley.omorogiuwa@uniben.edu

Abstract

The study investigated the determination of aberrant examinees' responses of 2018 and 2019 May/June Biology multiple choice test questions of the National Business and Technical Certificate Examination (NABTCE) in Nigeria. Six research questions were raised, and two hypotheses were formulated and tested at 0.05 level of significance. This research adopted the survey research design of the ex post facto. The population for the study was 57357 Biology students responses which consisted of 28121 and 29254 for 2018 and 2019 students' scored responses of the NABTCE respectively. The sample size for the study was 13533. This comprised of 7086 and 6447 examinees' responses for the 2018 and 2019 NABTCE Biology. A multistage sampling technique was adopted for the study, in which a stratified sampling technique, simple random sampling technique and a cluster sampling was applied for effective selection. The instrument used was the NABTCE May/June Biology multiple choice test questions for 2018 and 2019 made up of fifty items with four options each. The instruments were assumed to be valid and reliable as it was a test constructed by a public examination body. The Perfit which is a non parametric Item Response Theory (IRT) statistic was used to determine aberrant examinees' responses. Examinees' responses whose HT coefficients were less than or equal to .25 and .15 were flagged as aberrant for the 2018 and 2019 respectively. Sample with percentages of aberrancy of $\leq 5\%$ for low level, moderate level 6% - 15% and high level 16% - 25%. The research questions one and two were answered using frequency counts and percentages, research questions three and four were answered using Cronbach Alpha statistics while hypotheses one and two were tested using Chi-Square test of goodness of fit. The findings of this study revealed that the levels of the aberrant examinees' responses in the 2018 and 2019 NABTCE Biology multiple choice test examination were moderate. It was also revealed that there was a slight insignificant improvement in the internal consistency reliabilities of the NABTCE Biology for 2018 and 2019 multiple choice test after screening the data of aberrant examinees. There was also a significant difference in the frequency of occurrence between the aberrant examinees and the non aberrant examinees for the NABTCE Biology for 2018 and 2019 multiple choice test. Based on these findings, the researcher recommended amongst others, that the National Business and Technical Examination Board should continue in their good work for constructing Biology multiple choice items having low aberrant examinees' responses.

Psychometricians, test researchers and counsellors should be aware that the impact of aberrancy may be strong enough to cause misclassification of ability or construct.

Keywords: Aberrant Examinees' Responses, Internal Consistency Reliability, Aberrant Examinees, Non Aberrant Examinees

Introduction

Examination is a tool or instrument designed by an examining body, faculty, department or school to gather information about students' progress or performances as related to the content taught. It is an instrument administered at any speculated time designed by the institution. These examinations are organized at all levels consistent with the National Policies and Goals, and in line with emerging international trends. An examinee is a candidate who is exposed to testing or exposed to an examination. The responses supplied by these examinees are used to evaluate the level of presence or absence of the measured traits or constructs.

The evaluation of human behavior is an essential part in research. It deals with measuring the responses from testing precisely, clearly and correctly depending on some assumptions. This involves expressing the observed data mathematically by using the appropriate statistics. The aim of measuring is to make an evaluation. Evaluation leads to decision making about the examinees depending on the results from the assessment. The correctness and appropriateness of the decision depends on the appropriateness of measurement tool, the measurement results and the evaluation process. In line with this, there is the need to use standard measurement tools in assessing students' learning outcomes.

The measuring instrument used in psychology and education are tests, rating scales, checklists, questionnaire, observation, inventories among others. A test must be reliable, valid, objective, comprehensive, readable and standard for it to be good enough to get appropriate responses from examinees. These qualities of reliability and validity are the intrinsic components of a test which is precisely termed the psychometric properties of a test. Psychometric properties reveal information concerning a test adequacy, relevance and usefulness. It is concerned about the goodness of the test to eliciting appropriate responses from the examinees. According to Stephen (2021), psychometric test are assessments of people to measure psychological attributes such as personality or intelligence. A good test must have three fundamental properties of reliability, validity and norming. However there are two broad types of psychometric properties that a test needs to possess in order to be considered a good measure of a particular construct, they are reliability and validity.

Reliability is the extent to which the scores gotten from a test, or its parallel form are consistent over time. According to Heale and Twycross (2015), reliability is the extent to which measurements are repeatable when different people perform the measurement on different occasion, and under different condition, supposedly with

alternative instruments which measure the construct or skill.

A reliable scale is one that makes measurements without an error. Although it is impossible to make such measurements without an error because of the factors that necessitate the data collecting process and the data analysis process. However, an insignificant error is of high necessity to increasing the reliability of the measurement. To attain this, it is required that the sources of these errors are determined and then controlled for. The sources of these errors might be from the measuring tool (random) or a source other than the measurement tool (non random). Though reliability is important for a study, it is not sufficient unless combined with validity. To this effect, it can be stated that, all valid are reliable but, not all reliable are valid. In line with this, some researchers like Oyeniyi, et al (2016), Wayne (2014) and others have identified that though a reliable data does not necessarily mean a valid data, but a valid data means a reliable data. This explains the necessity of validity added to reliability rather than reliability alone. This is because a valid test explains how good a test is for a particular purpose whereas reliability ensures trustworthiness of test scores.

Validity can be referred to a test potential to measure what it claims to measure. It is the test capacity to measure what it claims to measure. Validity is a crucial part of a reliable test that indicates whether the test measures what it is expected to measure in the examinee. It is important to say that high validity of a test guarantees that, the items remain firmly connected with the test's intended core interest. Validity depends substantially on the sample of participants such as age, gender, language and culture.

Even when the validity and reliability of a data set have been proven good, such data set may still contain some mis fitting responses supplied by examinees, who are termed aberrant examinees. These responses may not match with the examinee's ability or construct level. An aberrant examinee is identified by his responses. A response straying away from the expected is said to be aberrant, mis fitting or unexpected. Also, when the expected score of an examinee do not match with the observed score, it is referred to as aberrant or misfit responses. When other factors other than the latent trait or psychological construct being measured affect an examinee's response pattern, the examinee's response behavior becomes aberrant or unexpected. Responses of such kind would result in test scores not adequately reflecting the level of the latent trait of the examinee. There are various behaviors that are indicative as aberrant responses. Meijer (2001) cited by Jia et al (2019) proposed some examinees behaviors that are related to aberrant respondents such as, slipping or carelessness, guessing, cheating, alignment error (when items on a test do not align with the curriculum), and deficiency of sub-abilities (such as, ability in critical thinking and problem solving).

Aberrant responses may come in as a result of instrumental errors, examinee's disposition or environmental distraction. A respondent's level of

motivation, poor understanding of the instructions, in attentive reading of the items, inability to respond sincerely and ignoring some response categories may cause testee/examinee to provide response patterns that are inconsistent with the underlying model of the trait. This can lead to biased examination results and erroneous decisions about individuals. Significantly, inaccurate estimation negatively affects individuals and organizations, and may lead to erroneous conclusion about test validity (Makalesi 2021). The type of aberrant response produced by an examinee is a function of the item or the examinee. When difficult items are answered incorrectly by a low ability examinee, such is a function of the item but, when easy items are answered incorrectly by a high ability examinee then, such is a function of the examinee.

These aberrant responses though hidden are part of the responses been supplied by the examinees. Although these parameters cannot be determined by subjecting test responses to the statistics of validity and reliability, because such statistics are not capable of identifying examinees whose responses are aberrant. In most study conducted, such responses are not been sorted out from the entire responses, before statistical analysis are done. This therefore implies that for every large or small data set not screened, may contain some percentages of aberrant responses. These aberrant responses are all together included in the data analysis process in significant or insignificant degrees. To ensure the truthfulness of test scores, it becomes necessary to identify statistical procedures for detecting aberrances. These statistics operate by detecting responses that are a mis match or are aberrant.

These aberrant responses are identifiers to the respondents, as aberrant. The responses can be detected in the design of the test or in the irregular behavior of the examinee(s). However examining the test score or response patterns of examinees that deviates, can be done through person-fit analysis in Item Response Theory. Item response theory (IRT) is a statistical method which is based on the responses supplied by examinees to interpret his or her ability level in line with the item characteristics. The classical estimate of ability in IRT is highly sensitive to response disturbances. These disturbances termed aberrant, may jeopardize measurement accuracy among respondents and invalidate the use of IRT. As mentioned by Jia et al (2019), in IRT framework, aberrant responses can be addressed through (i) methods based on response times (RTs), (ii) methods without response times, such as person fit analysis, (iii) weight robust estimation method which reduce the influence of aberrant responses on ability estimation, (Magis, 2014).

Person fit statistics is concerned with detecting whether an examinee's item score pattern is unusual or unexpected, when compared to the item score patterns in a group of test takers (group-based analysis) or alternatively when compared with a prespecified IRT model (model-based analysis). Existing person fit statistics includes parametric and non parametric statistics in IRT. Parametric IRT person fit

statistics assess the fit of a response pattern relative to a given IRT model such as the 3 PLM. They are used to estimate item and ability parameters for calculating person fit indices and then, classify respondents as fitting (non aberrant) or mis fitting (aberrant) respondents. This statistics evaluate the misfit of an observed response pattern with an IRT model by calculating the response probabilities using respondents' ability and item parameters (Mousavi & Cui 2020). In the non parametric IRT statistics, the probability of a correct item is defined as an ordinal function of the latent ability. Here the examinee's ability coefficient is compared with a cutoff value generated by applying algorithm to determine if he is an aberrant or non aberrant examinee.

Biology is among the several subjects which National Business and Technical Examination Board (NABTEB) conducts examinations on. It is one of the science subjects taught in senior secondary schools, which attracts patronage of both science and non science students. For this reason, biology has a very high enrolment of students in the external examination West African Examination Council (WAEC, 2015). Biology test is either constructed as objective/multiple choice, essay and practical or a combination of any. Multiple choice test is a test in which the examinee is expected to choose the correct answer from the given options. It is made up of two parts, the stem and the options. The correct answer among the options is called the key, while the other alternatives are called distracters. The number of alternatives used in multiple choice test vary, but in most cases, four options are used.

Oluwatimilehin (2021) investigated the prevalence of aberrant response behavior in Mathematic test. The researcher analyzed the data using descriptive and inferential statistics. Matrix Laboratory (MATLAB) programme was used to compute, aberrance values for the response of each student to MAT. The result showed that there was a high prevalence of aberrant behavior in Mathematics test among the sampled students. The researcher concluded that aberrant response behavior was prevalence in Mathematics test among secondary school students.

Arias, et al (2020) investigated on, A Little Garbage in, lots of Garbage out in assessing the impact of careless responding in personality survey data.. Between 4.4% and 10% were identified as careless insufficient effort (CIE) cases. The complete sample revealed that all the theoretical models obtained an unacceptable fit, while the clean sample fitted all theoretical models satisfactorily. The result also proved that a limited amount of Careless insufficient effort respondents data can lead to a drastic deterioration of the fit of the theoretical models, produce large amount of spurious variances, raise serious doubts about the dimensionality and internal structure of the data and reduce reliability with which the trait scores of all surveyed are estimated.

Marasi, et al (2019) studied to observe if participant carelessness is a substantial problem with survey data, that significantly influences or biases the results of statistical analysis on correlation, t-test and simple linear regressions.

Based on the findings, participant carelessness does not appear to be a substantial problem or demonstrate levels of bias in the result. There are two significant differences between the full and careful samples with the t-test and the regression comparisons of fit statistics demonstrate the careful samples to have a weak improvement over the full sample. However, none indicated bias. In the study carried out by Schneider, May and Stone (2018) to examine the impact of careless responding in internet-based quality of life assessments. Their study found that about 7.4% of the participants were classified as careless responders, and that, no substantial differences in the reliability PROMIS measures between careless and non careless responder groups.

Hormann and Schudlik (2021) investigated a study on the identification and compensation of aberrant response patterns. Based on this approach, a full sample was compared with a cleaned sample to examine whether aberrant responses have significantly biased the results. They concluded that the differences were negligible, which is a proof for the data quality.

Aberrant responses have negative impact on item parameter estimation, examinee's evaluation and other statistical analysis. The effects of aberrant data can have very undesirable consequences in the interpretation and replication of test results. Though, aberrant responses are not completely unexpected in practice particularly when an assessment instrument is administered to a large sample of examinees. However, a high magnitude of occurrence is an issue of concern, as it is capable of negatively limiting data quality and thereby affecting the conclusion derived from such data. Given the importance of the National Business and Technical Certificate Examination as large scale examination, used for certification purpose, there is need to empirically investigate for aberrant respondents before its negative impact becomes consequential. The study therefore seeks to determine the level of aberrant respondents for 2018 and 2019 May/June NABTCE Biology objective questions.

Research Questions

To guide this study, six research questions were raised for the study.

1. What is the level of the aberrant examinees' responses in the NABTCE 2018 Biology multiple choice test examination?
2. What is the level of the aberrant examinees' responses in the NABTCE 2019 Biology multiple choice test examination?
3. What is the internal consistency reliability of the NABTCE 2018 Biology multiple choice test items, before and after screening the data of aberrant examinees?
4. What is the internal consistency reliability of the NABTCE 2019 Biology multiple choice test items, before and after screening the data of aberrant examinees?
5. Is there a difference between the non aberrant examinees and the aberrant

- examinees in the NABTCE 2018 Biology multiple choice test examination?
6. Is there a difference between the non aberrant examinees and the aberrant examinees in the NABTCE 2019 Biology multiple choice test examination?

Hypotheses

Research questions five and six were hypothesized for the study.

1. There is no significant difference between the non aberrant examinees' responses and the aberrant examinees in the candidates of the NABTCE 2018 Biology multiple choice test examination.
2. There is no significant difference between the non aberrant examinees and the aberrant examinees in the candidates of the NABTCE 2019 Biology multiple choice test examination.

Methodology

The researcher adopted a survey research design. The population of the study was 57375 examinees who sat for the NABTCE May/June Biology objective test questions. This comprised of all the responses of SSIII Biology students in Nigeria, who sat for the NABTCE 2018 and 2019 Biology objective test questions and are 28121 and 29254. The sample for the study was 13533 examinees which comprised of 7086 and 6447 examinees' responses for 2018 and 2019 respectively. A multistage sampling technique was employed for the study and, the selection was done in three stages. First stage: stratified sampling was adopted to stratify Nigeria into the 6 geopolitical zones. Secondly the simple random sampling techniques was used to select 2 states each from each geopolitical zones by balloting with replacement. By so doing, 2 states were randomly picked from each geopolitical zones for the respective years (2018 and 2019) to achieve a total of 24 states. Thirdly a cluster sampling was adopted for the study. This involved using the entire examinees' responses from the two states, that were randomly selected from each of the geopolitical zone. This led to a total of 13533 examinees' responses selected from the entire population. The instruments that were used for the study are the NABTCE May/June Biology Objective Test Questions for 2018 and 2019. Each of the instruments comprised of 50 questions with 4 options. The NABTEB Biology objective test instruments were considered valid and reliable by the researcher because it is a standard examination Board. The researcher collected the responses of the 2018 and 2019 NABTCE Biology objective test items in the form of students by item matrix. Correct responses were coded as 1, while incorrect responses were coded as 0, in a Micro excel spreadsheet. The IRT perfit Ht coefficient is a non parametric statistics propounded by Sijtsma in 1986 and was used to detect aberrant examineess in the NABTEB 2018 and 2019. Biology Objective Test data set. The entire students' responses for the analysis was fed into the Perfit package in R programming. The Ht coefficient is the examinees' coefficients. This was estimated

for every examinees using an algorithm. The cutoff value for identifying aberrant or misfitting respondents employed a bootstrap resampling procedure for approximating the sample distribution of the PFS based on generated fitting response patterns. Hence the Ht cutoff value for the 2018 and 2019 Biology sampled responses were 0.245 and 0.152 respectively. Examinees whose Ht coefficients were less than or equal to 0.25 and 0.15 were considered as aberrant respondents for the 2018 and 2019 respectively.

Research questions one and two were answered using descriptive statistics. A sample with percentages of aberrancy of $\leq 20\%$ was classified as low level, moderate level 21% - 40% and high level 41%, and above. Research question three and four were answered using Cronbach Alpha Statistics to determine the internal consistency correlation of the items in the test, before and after screening the data of aberrant examinees for the 2018, 2019 and 2020. Hypotheses one and two were tested using the Chi-Square test of goodness of fit statistics. All the hypotheses were tested at an alpha (α) level of 0.05 level of significance.

Research Question One: What is the level of Aberrant examinees' responses in the candidates of the NABTCE 2018 Biology Multiple Choice test examination?

Table 1: Distribution of Aberrant Examinees' Responses of the NABTCE 2018 Biology Multiple Choice items

Status	Frequency	Percent
Aberrant	714	10.1
Non Aberrant	6372	89.9
Total	7086	100.0

Table 1 represents the distribution of examinees who sat for NABTCE 2018 Biology Multiple Choice test. The study sample consisted of 7086 examinees for the 2018 Biology NABTCE. Aberrant examinees were determined using a cutoff value of 0.25. Examinee whose HT coefficient was less than 0.25 were flagged as aberrant. The table above shows that 714 examines representing 10.1% were classified as aberrant, while 6372 representing 89.9% of examinees were classified as non aberrant. This implies that the level of aberrant examinees is low

Research Question Two: What is the level of Aberrant examinees' responses in the candidates of the NABTCE 2019 Biology Multiple Choice test examination?

Table 2: Distribution of Aberrant Examinees' Responses for the NABTCE 2019 Biology Multiple Choice item

Status	Frequency	Percent
Aberrant	858	12.9
Non Aberrant	5589	87.1
Total	6447	100.0

Table 2 represent the distribution of examinees who sat for NABTCE 2019 Biology Multiple Choice test. The study sample consisted of 6447 examinees for the 2019 Biology NABTCE. Aberrant examinees were determined using a cutoff value of 0.15. Examinee whose HT coefficient was less than 0.15 were flagged as aberrant. The table above shows that 858 examines representing 12.9% were classified as aberrant, while 5889 representing 87.1% of examinees were classified as non aberrant. This also implies that the level of aberrant examinees is low.

Research Question Three: What is the Internal Consistency Reliability of the NABTCE 2018 Multiple Choice Test Items before and after Screening the Data of Aberrant Examinees?

Table 3: Cronbach Alpha Coefficient before and after Screening the Data of NABTCE 2018 Biology Multiple Choice Items

Before Screening		After Screening	
	No. of Items	Cronbach's Alpha	No. of Items
.775	50	.795	50

The internal consistency reliability of the instrument was established firstly using the sample containing both the aberrant and non aberrant examinees (7086 Biology examinees) and secondly using only the non aberrant examinees (6372). The data were analysed using Cronbach's alpha statistics and it yielded a reliability coefficient of .775 and .795 respectively. This implies that the internal consistency of the items is good and acceptable, although with a slight increment in the non aberrant sample.

Research Question Four: What is the Internal Consistency Reliability of the NABTCE 2019 Multiple Choice Test Items before and after Screening the Data of Aberrant Examinees?

Table 4: Cronbach Alpha Coefficient before and Screening the Data of NABTCE 2019 Biology Multiple Choice Items

Before Screening		After Screening	
Cronbach's Alpha	No. of Items	Cronbach's Alpha	No. of Items
.803	50	.813	50

The internal consistency reliability of the instrument was established firstly using the sample containing both the aberrant and non-aberrant examinees (6447 Biology examinees) and secondly using only the non-aberrant examinees (5589). The data were analysed using Cronbach's alpha statistics and it yielded a reliability coefficient of .803 and .813 respectively. This implies that the internal consistency of the items is good and acceptable, although with a better improvement in the non-aberrant sample.

Hypothesis One: There is no significant difference between the non-aberrant examinees and the aberrant examinees of the NABTCE 2018 Biology multiple choice test examination

Table 5: Chi-Square test of Goodness of fit

	Status
Chi-Square	4.514 ^a
Df	1
Asymp. Sig	.000

Table 5 shows a Chi-Square value of 4.514 and a p value of .000. Testing at an alpha value of 0.05, the p value is less than the alpha value so the null hypothesis which states that there is no significant difference between the non-aberrant examinees and the aberrant examinees of the NABTCE 2018 Biology multiple choice test examination is rejected. Consequently there is a significant difference between the non-aberrant examinees and the aberrant examinees of the NABTCE 2018 Biology multiple choice test examination.

Hypothesis Two: There is no significant difference between the non aberrant examinees and the aberrant examinees of the NABTCE 2019 Biology multiple choice test examination

Table 6: Chi-Square test of Goodness of fit

	Status
Chi-Square	3.472 ^a
Df	1
Asymp. Sig	.000

Table 6 shows a Chi-Square value of 3.472 and a p value of .000. Testing at an alpha value of 0.05, the p value is less than the alpha value so the null hypothesis which states that there is no significant difference between the non aberrant examinees and the aberrant examinees of the NABTCE 2019 Biology multiple choice test examination is rejected. Consequently, there is a significant difference between the non aberrant examinees and the aberrant examinees of the NABTCE 2019 Biology multiple choice test examination.

Discussion of Findings

Research questions one and two addressed the level of aberrant examinees' responses using data from 2018 and 2019. The Perfit statistics distinguished between aberrant and non-aberrant examinees. The findings revealed that there is a low percentage of aberrant examinees among those who sat for the NABTCE 2018 and 2019 Biology multiple choice test examinations. The reason for this could be because every examinee would ensure that he or she respond to the items as accurate as possible. Although, the result from 2019 NABTCE Biology multiple choice test revealed the higher level of aberrant examinees' responses with percentage of 12.9 as against that of 2018 with percentage of 10.1. The reason for this high increase may be due to the effects of Covid 19 pandemic, when all schools were shut down to reduce the risk of the spread of the disease, and so probably students were not so much serious with their academic work. The results of this findings collaborated with the findings of Schneider et al (2018) in their study on Careless Responding in Internet-based Quality of Life Assessment. The researchers used the Patient Reported Outcomes Measurement Information System questionnaire, out of which 148 (7.4%) participants were classified as careless responders from the entire sample of 2000. This study is also in line with the study conducted by Arias, et al (2020) in a study on A Little Garbage in, Lots of Garbage out, in Assessing the Impact of Careless responders in Personality Survey. The result of the study showed that the factor mixture model detected inconsistencies of response to items with different semantic polarity identified between 4.4% to 10% of careless or insufficient effort respondents. Although the result of Oluwatimilehin (2021), who in her study on Aberrant Response Behavior in Mathematics Test Among Secondary School Students, revealed a high percentage of examinees who responded aberrantly. 91.7% representing 1552 students were considered to be aberrant responders while 140 students were non-aberrant.

Based on the collective results, the researcher agrees that survey data set contains some levels of response patterns that are aberrant. This although could be of varying levels, which may be as a result of intrinsic and extrinsic factors that may affect the examinee in the course of testing. This result supports the IRT three parameter logistic model and the four parameter logistic models with the theoretical claim of examinees whose response patterns are incongruence due to guessing or slipping factors which is termed the “c” and “d” parameter respectively.

The implication of this finding is that some examinees when responding to survey items especially in large scale items can produce responses that are not their true ability match. The pattern of responses produced may boost their performances without increase in ability or may decrease their performance without a decrease in ability. In whichever case aberrancy is presented, performance is affected and the psychometric properties of the scale or the purpose for which the scale was developed may be affected. Also, that aberrancy can be estimated using Perfit statistic of the non-parametric IRT.

Research questions three and four indicated that there is no statistical differences in the reliabilities before and after screening the data of those who responded aberrantly. The differences in the reliabilities before and after screening the data were not up to .1, hence no significant statistical difference. The result of this study may have appeared this way, since the levels of those who responded aberrantly were low for 2018 and 2019. The findings of this study also aligns with that of Schneider et al (2018) in their study on Careless Responding in Internet-based Quality of Life Assessment. Part of their findings was that there was no substantial differences in the reliability of PROMIS measures between careless and non-careless respondents. This was also supported by the work done by Kam (2019) on Careless Responding threatens factorial analysis results and construct validity of personality measure. They stated that the claims made by some researcher that, careless responding distort factor loading pattern and may also threaten convergent validity received limited support. This study is contrary with the study conducted by Maraci, Wall and Brewer (2019), who studied to observe if participant carelessness is a substantial problem that significantly influences or biases the results of statistical analysis. In their findings, they suggested that the individual reliability and the RIR approaches were not entirely fundamentally similar. Also, Arias, et al (2020) carried out a study on A Little Garbage in, Lots of Garbage out in Assessing the Impact of Careless responders in Personality Survey. Their findings was that it raised serious doubts about the dimensionality and internal structure of the data and reduced the reliability with which the trait scores of the survey are estimated. The implication of this study is that aberrant respondents have a very small impact on the internal consistency reliability. Based on this finding, the researcher agreed that the reliabilities of the data before and after screening the data set of aberrant examinees for 2018 and 2019 NABTCE Biology multiple choice items are statistically similar.

Hypothesis one and two addressed the differences between the non aberrant examinees and the aberrant examinees for the 2018 and 2019 NABTCE Biology. The results for the 2018 and 2019 indicated that there is a significant difference between the aberrant examinees and the non aberrant examinees for the NABTCE Biology multiple choice test items spread across the years. The result from testing these hypotheses may have appeared this way because of the pattern of responses between the aberrant examinees and the non aberrant examinees for those who sat for the NABTCE Biology multiple choice test examinations for the two years, and also because the number of those who responded aberrantly compared to those whose responses were non aberrant are greatly different. To support this result is the findings of Maraci, Wall and Brewer (2019) who showed that there were two significant differences between the full and careful samples. In support of this view, is the findings of Schneider et al (2018) who found out that careless responding meaningfully and significantly affected the correlations in Quality of Life (QoL) between medical and disabled groups. This was not so in the findings of Hormann and Schudlik (2021) who concluded that the differences were negligible with a proof

of data quality. This was not in agreement with this study because a secondary data were used. This may be one of the reasons why the results were contrary.

Conclusion

Based on the research findings, the researcher concludes that the 2018 and 2019 NABTCE Biology multiple choice test contained aberrant examinees in low levels and therefore may lead to invalidation of psychometric properties. This study also concludes that the internal consistency reliabilities obtained before and after screening the National Business and Technical Certificate Examination (NABTCE) Biology multiple choice test of aberrant examinees had close values, with the reliabilities after screening the data of aberrant examinees responses, having a slight improvement over the reliabilities before screening. Also with respect to comparing the differences that existed between the aberrant and non aberrant samples of the National Business and Technical Certificate Examination (NABTCE) Biology multiple choice items for 2018 and 2019, there were significant differences between the aberrant and the non aberrant examinees.

Recommendations

Based on the findings and conclusion reached in this study, the researcher therefore recommends as follows;

1. That the National Business and Technical Examination Board has credit for constructing Biology multiple choice items for 2018 and 2019 having low aberrant examinees' responses.
2. The study has proved the occurrences of aberrant examinees responses in survey test. Psychometricians, test researchers and counsellors should be aware of this fact because its impact may be strong enough to cause misclassification of ability or construct.
3. Students should be well informed about the implications of responding aberrantly during testing as this can impact on them as well as the Nation at large.
4. Teachers should also educate their students from time to time on the dangers of responding aberrantly, so that with constant reminders examinees may gradually put off such practices.

References

- Arias, V. B. Garrido, L. E., Jenaro, C., Martinez-Molina, A. & Arias, B. (2020). A Little Garbage in, Lots of Garbage out: Assessing the Impact of Careless Responding in Personality Survey Data. *Behaviour Research Methods*. Advanced Online Publication <https://doi.org/10.3758/s13428-020-01401-8>.
- Heale, R. & Twycross, A. (2015). A Validity and Reliability in Quantitative Studies. *Evid Based Nurs* 18 (3), 66-70
- Hormann, H. & Schudlik, K. (2021). Identification and Compensation of Aberrant Response Patterns. Quality of Comfort Assessments in Human Subject Trails. Issn 1434-8454 ISRN DLR-FB-2021-05 German Aerospace Center (DLR) Institute of Aerospace Medicine Hamburg/ Germany
- Jai, B., Zang, X. & Zhu, Z. (2019). A Short note on Aberrant Response Bias in Item Response Theory. *Frontiers in Psychology*, Edited by Holmes Finch. <https://doi.org/10.3389/fpsyg.2019.00043> www.frontiersin.org.
- Lobindo-wood, G. and Haber, J. (2013). *Nursing Research: Methods and Critical Appraiser for Evidence Best Practice 98th ed*. St Louis, Missouri: Elsevier and Mosby Publications.
- Magis, D. (2013). A note on the Item Information Function of the four Parameter Logistic Model. *Applied Psychological Measurement*, 37 (4), 304-315
- Marasi, S., Wall, A. & Brewer, K. (2019). Participant Carelessness: Is it a Substantial Problem with Survey Data? *The Electronic Journal of Business Research Methods*, 17(1), 1-16 www.ejbrm.com
- Makalesi, A. (2021). Impact of Aberrant Responses on Item Response-Based Model Estimations. *Kastamonu Education Journal*, 29 (5), 1024-1033. Doi: 10.24106/kefdergi.836241
- Mousavi, A. & Cui, A. (2020) The Effect of Person Misfit on item Parameter Estimation and Classification Accuracy: A Simulation Study. *Educ Sci* doi: 10.3370/educsci10110324
- Oluwatimilehin, T. R. (2021). Aberrant Response Behaviour in Mathematics Test Among Secondary School Students in South Western Nigeria. (Doctorate Dissertation): Obafemi Awolowo University

- Oyeniya, J. O., Abiodun, A. J., Moses, L. C., Obamiro, J. K. & Osibanjo, A. O. (2016). Research Methodology with Simplified Step-by-Step Application of SPSS Package. Lagos, Pumark Nigeria Limited
- Schneider, S., May, M. & Stone, A. A. (2018). Careless Responding in Internet-based Quality of Life Assessments. *Qual Life Res.* 27(4) 1077-1088
<https://doi.org/10.1007/s11136-017-1767-2>
- Stephen, M. N. (2021). What is a Psychometric Test. Assess.com>psychometric
- Wayne, L. D. (2014). Terrorism, Homeland Security and Risk Assessment via Research Proposal (3rd Ed) Bloomington: Xhbris, LLC.
- Wilson, J. (2010). Essentials of Business Research: A Guide to doing your Research Project, SAGE Publication.