AFRICAN JOURNAL OF THEORY AND PRACTICE OF EDUCATIONAL ASSESSMENT (AJTPEA)

VOL 8(1), 2019: 48 – 62 Copyright © Educational Assessment and Research Network in Africa (EARNiA). ISSN 2545-5834
http://www.earnia.org; info@earnia.org; earnia13@gmail.com

DEVELOPMENT AND APPLICATION OF GEOGRAPHY ATTITUDE SCALE USING RASCH MODEL IN SOUTH EAST NIGERIA.

Chikodiri C. Ikwueze, Ursula N. Akanwa, Alphonsus O. Ovute & Madu A. Onuabuchi

ABSTRACT

The study was designed to develop an instrument that will be used to assess students' attitude to geography using the Rasch model. The instrumentation design was adopted for the study. The sample was 850 students from 78 secondary schools using multistage sampling. An instrument called Geography Attitude Scale (GAS) was validated using face, factorial and test-of-the-assumption-of-unidimensional validation. Based on these, 38 items were discarded while the 42 items that met the set criteria formed the final instrument. Data were analysed using Rasch rating scale model software WINSTEPS (version 3:81) to answer the research questions and test hypothesis at 0.05 level of significance. The Rasch model expectation fits were met by 40 out of 42 items. The item threshold showed that 20 items were with positive values while 22 items had negative measures. Item and person separation index revealed good spread of items and persons from high difficulty to low difficulty level of endorsement and into higher and lower levels of attitude respectively. It was recommended that instrument GAS should be used to assess attitude of students towards geography in secondary schools.

Key words: Development, Geography, Attitude, Rasch Model.

INTRODUCTION

Everyday life endeavours present to us challenges that task our ingenuity. In each of these challenges, we are faced with a number of alternatives from which to make a choice. Assessment is a very important part of learning; it involves the process of making a judgement on measurement of the worth of an entity. Oluwatayo (2006) described assessment as a human operation always carried out to find the worth, value or credibility of an action, or operation. Hopkin (2006) acknowledged that effective practice of

assessment is an important means of improving student's attainment in a subject. There are many ways of conducting an assessment. It may be oral, aural, practical, or observational (Miller, 2005). It can also be carried out in one-to-one, small or large group setting, or on a computer. According to Bloom (1956), assessment can be carried out in the three domains of learning, that is, cognitive domain, which focuses on knowledge and intellectual development of skills; psychomotor domain, which includes manipulative and motor skills and the affective domain which addresses attitude and value. Affective domain portrays the emotional area of learning mirrored by the beliefs, values, attitudes, and behaviours of learners. Affective learning is concerned with how learners feel while they are learning, as well as with how learning experiences are internalized to guide the learner's attitudes, opinions, and behaviour in the future.

Affective domain deals with students' attitude, emotions and feelings. Affective behaviour deals with attitudes, interest, emotional adjustment, values, beliefs, social relations, habits, and overall lifestyle (Federal Republic of Nigeria (2013). Attitude has been defined by different authors from different fields of study. However, the elements that constitute these definitions are similar and complementary. Guven (2006) defined attitude as the predisposition to respond in a generally favourable or unfavourable manner with respect to the object of the attitude. He further said that attitude is an emotional and mental readiness or a preliminary tendency that is built on emotion, knowledge, experience or motivation on any subject, social topic or event. Kind, Jones, and Barmby (2007) viewed attitude as having cognitive (knowledge, belief, and ideas); affective (feeling, like, dislike) and behavioural (tendency, towards an action) components.

The attitude of an individual towards a thing makes the person to form judgement as to whether the object is good or bad, healthy or unhealthy, attractive or unattractive, valuable or invaluable (Crano & Prislin, 2006,). Attitude towards geography is an affective behaviour that has been studied over the years in relation to geography achievement. It can be defined as a positive or negative emotional disposition towards geography as a school course.

A genuine attitude to school subjects equally reveals students' affection towards the study of the subjects and their supremacy in the subject. Attitude towards a subject is not always perfect. Guven (2006) identified explicit and implicit attitude as two different ways in which a person's attitude can be measured. Explicit attitudes occur consciously in a man out of diverse exposures in life. On the other hand, implicit attitudes are attitude that develop unconsciously in an individual. The attitude of students towards a school subject, like geography, is measured at the conscious level.

Geography is the study of places, environments, and spaces of the Earth's surface and their interrelations. It attempts to respond to the question of why things remain the way

they are. The current academic fields of geography is embedded in old practices concerned with the attributes of places, in particular their natural environments and people, together with the inter relationship between the two (Oformata, 2008)). Its different identity was first introduced and named some 2000 years ago by the Greeks. The terms 'geo' and 'graphia' where merged to literally mean 'earth' and 'description' respectively. Geography is a field of science devoted to the study of the location of people as well as activities within the earth, and the motive behind their distribution.

Geography is useful to all the students, whether they are likely to continue with it at tertiary levels or they intend to stop at secondary school level. It prepares students with geographical knowledge and skills to become functional and socially useful. Geography is a conspicuous and active science and or social science course that concerns the study of mankind and his natural surroundings (Akintade, 2012). Geography hence assists our youths to understand the worth of their environment and its growing innate raw materials. As a discipline, it deals with numerous towards related areas, expressively, and it is intellectually stimulating. It shows relationship with different related school courses. It also imbues in the learners the necessity to cherish and create a feeling of obligation in respect to their own community. Geography is a vast but an exciting subject which is connected to other subjects such as history, government, physics, mathematics, economics, biology and agriculture. It therefore, requires a great deal of effort and diligence to learn and enjoy it at the senior secondary school level. However, there has been a conspicuous decline in the population of students that offer geography at the senior secondary school classes in south east Nigeria. According to Aydin (2009), many students see geography as a hard subject which requires a combination of quantitative data and skill. Majority of students do not see the relevance of geography lessons. Teacher's attitude and association with students has a very important effect on the student's attitude towards geography.

Another relevant challenge based on the present study is how teachers can conduct assessment on attitude that would be reliable and valid, knowing that attitude involves latent traits which can only be described but cannot be measured directly. A way of solving the problem is to develop a good instrument for measuring students' attitude. Instrument development is one of the essential processes of educational measurement and evaluation. In developing an instrument, Gall, Borg and Gall (2007) have identified objectivity, reliability and validity as criteria that can be used to judge whether a test or scale is of sufficient quality for use in educational research. An objective test or scale score is one that is not distorted by the biases of the individuals who administered and scored it. Rasch (1960) contends that objectivity requires that the measure assigned to an attribute/construct be independent of the observer. A reliable measure is one that yields consistent results while a valid instrument measures what it is intended to measure. In education, a number of instruments are used to assess behaviour. They include:

intelligence tests, personality tests, achievement tests, aptitude tests, interest inventories, attitude inventories, behavioural procedures and neuropsychological tests. The practical relevance of these instruments depends on the level of reliability, validity, difficulty, and discrimination indices they possess. To develop and use these tests, two theories required for characterizing the results are often involved. These are the Classical Test Theory (CTT), and the Item Response Theory (IRT).

The classical test theory is a component of similar psychometric theories that predict the outcome of psychological testing such as the item facility or the competency of testees. Usually, the aim of CTT is to understand and boost the reliability of psychological tests. Under the CTT framework, item analysis consists of calculating difficulty and discrimination indices for each item. According to Thorndike (1977), CTT views the score X that a student receives in an examination as the total of two observable components, genuine\true score and error\trial score). This is anchored on the presumption that a testee has an observed score and a true score (Crocker and Algina, 1986). Research shows that CTT has a number of limitations that hinder its usefulness as a foundation for modern testing. According to Wiberg (2004), CTT can be disputed since the true score is not a complete characteristic of a test-taker as it relies on the composition of the test. These limitations have made the CTT framework unfit for proper establishment of the psychometric properties of tests or surveys. This calls for a better framework which the items response offers. According to Nworgu (2010), test development techniques based on latent trait theories which include Item Response Theory (IRT) represent a more recent development in test theories.

Item Response Theory (IRT) is a psychometric practice stressing the truth that a person's answer to a given test item is influenced by the attributes of the person and the standard of the item. IRT contains many methods for getting feedback about individuals, items and tests. Therefore, the essential form of IRT asserts that an individual's response to an item is affected by the individual's ability level and the item's difficulty level. Three models have been identified in IRT as a result of variations in the parameters that are used to describe them. The models are– one parameter Rasch model, two-parameter and three-parameter models (Baker, 2001).

The Rasch model of IRT is a mathematical expression relating the likelihood of the outcome when an individual tries a single item to peculiarities of the person and the item. It is among the three models of the item response theory (latent trait measurement) advanced by psychometricians as a new measurement system to address the limitations of CT measurement. According to this model, a person's agreement to dichotomously scored items depends on the level of the trait and the item facility. The underlying theory behind the one parameter logistic model is how test takers at different levels of ability for a particular trait should response to an item.

Baker (2001) stated that there are only two technical properties of a Rasch model that are used to describe it. These are difficulty parameter designated by (b) and the discrimination parameter designated by (a).

According to Embretson and Reise (2000), the mathematical expression is presented as:

$$P(X_{is}=1|\ \theta_s,\beta_i) = \frac{e^{(\theta_s-\beta_i)}}{1+e^{(\theta_s-\beta_i)}}$$

Where:

X represents response (X) given by respondents to item i

represents person's ability.

represents item or task difficulty.

X = 1 denotes correctly answering the question.

E represents the base of the natural logarithm or Euler's number, 2.7183. Therefore, $P(X_{is}=1|\ \theta_s,\beta_i)$ shows the probability (P) of success upon i. The straight bar in the equation reveals a "dependent" probability, that is, the likelihood that the subjects will rightly agree to the item depends on the respondent's trait level ($_s$) as well as the item's difficult (Bi). Several applications of Rasch model have been made in the area of educational and psychological research. The Rasch-calibrated scale is particularly helpful in the advancement of item reserves from which items and or measure of noted difficulty or attribute can be withdrawn to develop further assessment instrument that are related. It is as well extremely important for application in (1) ability testing (2) measurement of attitude (3) person fit (4) computerized adaptive testing and (5) advancement in the evaluation of fundamental psychometric features of items in a test using feedback gotten from the item properties. This enables the test writers pick items that mirror the suitable scope of trait levels. When Rasch analysis is used for attitude measurement, Saed, Hind and Mutasem (2013) noted that negatively worded items usually outfitted the acceptable range of fit statistics and most of the positively worded items fall within the acceptable range.

Another major issue in educational planning and scoring is the effective assessment of the cognitive, psychomotor and affective behaviours of students in a school subject like geography. For high precision in planning and scoring the performance of a child, a meaningful ability estimate is important. This can be accomplished by using a psychometric model to give the examinees' an ability estimate which is independent of recent individual facet elements such as judges, task and items. Only few instruments concerning attitude of students to school subjects have been developed using the Rasch model; and none in geography. This means that literature is still scarce in this area, especially in Nigeria. It is against this background that the researcher decided to develop an instrument using Rasch model that will be used in assessing students' attitude to geography.

The educational system is mainly concerned with assessing the child's intellect, knowledge and ability to reason. Most evaluation programmes use one form of achievement test or the other and neglect the psychomotor and affective domains. Attitude, as one of the attributes of affective behaviour, has no visible attribute. This is to say that, they are abstract notions which are difficult to assess and measure directly. Due to this handicap, geography educators worry over getting an effective measure or instrument for assessing affective behaviour of students of geography. Literature shows that no proper measure has been applied in assessing the attitude of students especially in geography. Omare, cited in Philips (2011), identified lack of a proper measure for assessing the affective behaviour of students such as attitude.

CTT does not give a true assessment of student's affective outcomes, although it is the most commonly used psychometric measure to determine the quality of instruments for objective assessment of students' performance. The affective domain traits are mixed up as if they have the same meaning and the same influence on students' learning. In affective assessment, there is need that data should be analysed as the result will help the students in choosing a fulfilling career in future.

To ensure that assessment of students' performance is done objectively, validly and reliably, proper assessment method must be used. Previous researches conducted in Nigeria used diverse geography attitude assessment scales and the data got from the scales usually analysed traditionally without taking into account the importance of every item, (Okpala, 2000).Most researchers in Nigeria, still use the classical analysis instead of IRT analysis(Okpala, 2000). Although the Rasch model created by George Rasch has turned out to be well-known in many countries, the model does not gain sufficient attention from Nigeria researchers. There is, therefore, a need to further investigation of the application of Rasch measurement model in assessing students' attitude to geography. Hence, this study is on the development and application of geography attitude scale using Rasch model in South East, Nigeria. The purpose of this study was to develop and apply Geography Altitude Scale (GAS) using Rasch model. The study was designed to:

1. ascertain how the items of Geography Attitude Scale (GAS) fit the Rasch model
2. estimate the item threshold parameter of Geography Attitude Scale.
3. determine the item and person separation indices.

Research Questions

The following research questions guided this study

1. What are the items of Geography Attitude Scale that fit the Rasch models?
2. What are the item threshold parameter estimates of Geography Attitude Scale?
3. What are the item and person separation indices of Geography Attitude Scale?

Research Hypothesis

The hypothesis below was tested at 0.05 level of significance.

Ho_1: There is no significant fit of the items in the Geography Attitude Scale (GAS) to the Rasch model.

Methods

In carrying out the research, an instrumentation design was employed. All public secondary schools in south-east, Nigeria represented the demography of the study. Using a stratified proportionate random sampling method, a sample of eight hundred and fifty (850) students were selected from a total of 24,921 in78 out of the 1,104 senior secondary schools in three of the five states in the south east geo-political zone of Nigeria. The sample size is 3.4% of the total population. The students were from mixed and single sex senior secondary schools in Abia, Anambra and Enugu states. The instrument for the study, which was Geography Attitude Scale (GAS), was given to three experts in measurement and evaluation and one in geography department at Michael Opara University of Agriculture, Umudike for face validation. Factor analysis was used to obtain the construct validity. The ones that were factorally pure were chosen to form the 42 items in the instrument that was used for the study. Thus, 38 items were discarded because they did not meet up with the minimal approval value of .350.

From the percentage of variance explained by the first factor which was much higher than the second factor and subsequent ones, the items of the instrument were found to be unidimensional and by implication, locally independent. The validated instrument was trial-tested using 100 students from Kogi State. The reliability of the instrument calculated using Cronbach Alpha was 0.775. Rasch rating model software WINSTEPS was used to obtain Rasch person and item internal consistency of 0.93 and 0.95 respectively. Research questions and hypothesis were analysed using Rasch rating scale model software WINSTEP (Version 3: 81).

RESULT

Research Question 1:

What are the items of GAS that fit the Rasch model?

Table 1: Rasch Fitting Models; Infit and Outfit Mean Square (MNSQ) of the Items

Item Number	Sample Size (n)	Infit MNSQ	Outfit MNSQ
1.	850	.78	.76
2.	850	.98	.98
3.	850	.82	.83
4.	850	1.01	1.03
5.	850	.91	.90
6.	850	**1.52**	**1.52**
7.	850	.88	.89
8.	850	1.02	1.00
9.	850	1.25	1.24
10.	850	1.07	1.06
11.	850	1.11	1.12
12.	850	.94	.94
13.	850	.22	.22
14.	850	.78	.80
15.	850	1.03	1.02
16.	850	1.22	1.26
17.	850	1.06	1.03
18.	850	**1.51**	**1.51**
19.	850	1.33	1.32
20.	850	.94	.90
21.	850	.92	.91
22.	850	.99	.99
23.	850	1.07	1.08
24.	850	.94	.93
25.	850	.93	.93
26.	850	1.05	1.04
27.	850	1.02	1.00
28.	850	.99	.99
29.	850	.88	.86
30.	850	1.23	1.29
31.	850	1.12	1.11
32.	850	1.05	1.05
33.	850	.84	.83
34.	850	.84	.84
35.	850	.94	.94
36.	850	1.26	1.24
37.	850	.88	.87
38.	850	.82	.80
39.	850	.95	.94
40.	850	.93	.91
41.	850	.95	.95
42.	850	1.16	1.15
Overall infit and outfit mean		.99	.99

Table 1 shows the result of the fit statistic of the items of the GAS. Using the range of .6 to 1.4, the result showed that almost all the items examined had their mean square fit statistic values between .6 and 1.4. It is only items 6 and 18 that have mean square infit/outfit values above 1.4; this mean that the 2 items are under-fitting. The two items are seen to be non-fitting and discarded. The fit statistic of the items, apart from items 6 and 18, indicates that all the items are reliable and valid since they all fall within the range of fit that is regarded valid.

Hypothesis One

There will be no significant fit of the items in the Geography Attitude Scale (GAS) to the Rasch model. The infit and outfit mean square was used to test whether there is fit between attitude of students towards Geography and Rasch model. The results are shown in Table 1.

Table 1 revealed that two items (item 18 and 6) representing 76% were statistically significant and did not fit the model. The two items are seen to be misfitting and are to be dropped. Forty items representing 95.2% were not significant. They therefore fit the model.

Research Question 2

What are the item threshold parameters of Geography Attitude Scale?

Table 2: Threshold Estimates of the Items Descending Order and Model S tandard Error (SE).

Item Number	Item Measure (Threshold)	Model STDD Error
9	.41	.04
22	.30	.04
36	.29	.04
5	.27	.04
23	.25	.04
18	.25	.04
16	.23	.04
19	.21	.04
14	.20	.04
37	.20	.04
2	.20	.04
42	.18	.04
25	.12	.04
41	.11	.04
24	.02	.04
29	.02	.04
34	.02	.04
21	.02	.04
33	.00	.04
30	-.03	.04
38	-.04	.04
35	-.04	.04
20	-.07	.04
3	-.09	.05
40	-.09	.05
10	-0.11	.05
12	-.1	.05
7	-.13	.05
15	-.14	.05
6	-.14	.05
38	-.14	.05
39	-.15	.05
26	-.15	.05
31	-.16	.05
17	-.18	.05
32	-.23	.05
4	-.25	.05
8	-.28	.05
27	-.29	.05
11	-.37	.05
1	-.37	.05
Overall mean	.00	.04
SD	.20	.00

Table 2 shows the result of the threshold estimates for items called item measures. Items with positive measures represent items that are difficult to endorse while negative item measures signify items that are easy to endorse. The results revealed that the items range in difficulty from -.37 for item 1 to .41 for item 9. The accompanying standard errors are from .04 to .05. Twenty items (48%) are fairly difficult to endorse as the range is from .00 to .41.These are items with positive values, while 22 items (52%) were easy to endorse as they have negative measures that range from -.37 to -.03.

Research Question 3

What are the person's parameter estimates of GAS?

Table 3: Person and Item Separation Index:

Category	Real separation	Model separation
Persons	3.33	3.61
Items	4.21	4.32

Table 3 shows the 'Real' as well as 'Model' separation estimates. 'Real' represents 'worst case estimates' while 'model' can be taken as '.best case estimate' with the correct separation falling somewhere in-between (Edkins & Roy, 2011). Separation measures that are less than 1 are not good for decision-making, they show that the person of the students have not been separated into levels of the trait and attitude towards geography. Separation measures for subjects in the sample fall within 3.33 to 3.61, thus showing enough spread of subjects. In addition, item separation index ranged from 4.21 to 4.32. This means that there is a good spread of the items into levels of attitude.

DISCUSSION

The purpose of the first research question was to establish the construct validity of the instrument

(GAS) and to show how the data collected matched the Rasch modelled expectation. The criterion for fit to the model enables a test developer to identify and delete misfitting items. Green and Franthon (2002) posited that misfitting items are simply discarded because they add noise to the data when developing an instrument.

According to Wright and Linacre's (1994), the yardstick for appropriate INFIT and OUTFIT mean square range for rating scale fit amongst the items examined had their mean square fit statistic values within the specified range. The range is from 0.6 to 1.4. However, two items had their values above 1.4and were declared under fitting. Therefore, the 2 items were discarded. This brought down the number of the items of the

GAS to 40. Since 95% of the items fit Rasch model expectation, the model fits the data, meaning that the items of the instrument are valid. This is in agreement with Brentari and Golia (2008) where one item failed to fit Rasch modelled expectations.

The null hypothesis which states that there is no fit of the items in the GAS to the Rasch model was accepted for only two items and rejected for forty items. This means that there was a significant fit between the items and the one-parameter logistic model. This is in agreement with the findings of Metu (2015) who also used infit and outfit mean square (MNSQ) of the items to check for item fit. She discovered very low non-fitting items.

The intent of second research question was to estimate the item difficulty of the items. Since the item parameter which is the difficulty or threshold parameter of an item shows how difficult it is to agree with a statement or to identify any category in the ordinal rating scale, it means that a student needs to have a higher level of attitude towards geography in order to agree strongly with an item whose threshold value is high. Normally, item threshold values range from -4 to +3.Ranges beyond this are rarely seen (Harris, 1991).

Twenty-two (22) items (52%) with negative item measures mean that they are very easy to endorse. Twenty (20) items (48%) with positive measures are fairly difficulty to endorse. This shows that there are as much fairly difficult and easy to endorse items in the scale. The result is however not in agreement with the study carried out by Marais, Styles and Andrich (2011) where there was no item with negative values and where the threshold parameter range in difficulty from 1.93 to 3.34 (all highly positive items).

Research Question 3 tried to find out how persons who responded to the instrument were able to align the items into a hierarchy from high difficulty to low difficulty level of endorsement and whether items of the GAS were able to separate persons (students) into high level and low level of attitude of the participants. Person separation is employed to separate subjects; the Person Separation Index (PSI) provides evidence of the capacity of the scale to discriminate between differing levels of attitude towards geography. Low person separation with a suitable person sample means that the instrument may not be good enough to differentiate between high and low performers; more items may be required. According to Green (1996), Rasch models positioned items and persons on the same scale along a continuum. When lower numbers are found below the scale, it implies redundancy in items and less variability between subjects in connection to the assessed traits. Separation estimate for respondents in the sample indicated sufficient spread. This does not agree with the study by Brentari & Golia (2008) whose PSI was .83 and another study by Marais, Styles and Andrich (2011) whose PSI was also below 1.00 (i.e. .91).This shows that there are no clear separation of persons into high and low levels of the trait.

CONCLUSION

It is safe to conclude that GAS is a valid and reliable instrument, since only 2 misfitting items out of 42 items were discarded, making the items of the GAS at 32.

RECOMMENDATIONS

The following recommendations emanate from the study.

1. Psychometricians, test developers, teachers and other persons involved in any kind of measurement should be trained in item response theory framework. This will enable the advantage of the framework and its overall essence to be appreciated and popularized in our local situation.
2. The instrument, GAS, should be used to assess students' attitude towards geography in all high public and private school settings, especially at senior secondary level. This is the main purpose of this research which is to develop an attitude scale for assessing attitude to geography.
3. According to the results of the study, the application of IRT might involve the two and three parameter models of attitude of students towards geography and other school subjects.
4. Given the obvious advantages of IRT over other popular measurement framework, the government should encourage examination bodies like WAEC, NECO, NABTEB to adopt this measurement framework. This will ultimately surmount the measurement problems we frequently encounter in Nigeria.
5. Secondary school teachers in Nigeria should be oriented on the use of IRT for psychometric analysis of their examinations. This way the quality of test items in such school will get more refined and measurement problems associated with the present framework will get obliterated.

REFERENCES

Akintade, B. O. (2012). Considering the determinants of selecting geography as a discipline: The case of senior secondary school students in Ilorin, Nigeria. Ozean Journal of Social Sciences, 5(1), 202.

Aydin, F. (2009). Secondary school students' opinions about geography course: A qualitative study. Archives of Applied Science Research, 4(2), 105-110.

Baker, B. F. (2001). The basics of item response theory. (2nd ed.). ERIC clearinghouse on assessment and evaluation. University of Maryland College Parkton.

Bloom, B. S. (1956). Taxonomy of educational objectives. New York: David McKay Company Inc.

Brentari, E. & Golia, S. (2008). Unidimentional it in the Rasch model. How to detect and interpret it. *Statistical*, 01; 67(3), 258-261.

Crano, W. D. Prislin, R. (2006). Attitude and persuasion. Annual Review of Psychology, 57, 345-374.

Crocker, L. & Algina, J. (2008). Introduction to classical and modern test theory. New York: Harcourt Brace.

Edkins, R. & Rolal, K. (2011). Psychometric properties of the death anxiety scale (DAS) among terminally ill cancer patients. Journal of Psychology oncrology, 29(4), 317-356.

Embreton, S. E. & Reise, S. P. (2000). Item response theory for psychology, Mahwah. New Jesey: Lawrence Erlbaum Associates.

Federal Republic of Nigeria (2013). National policy on education, Lagos. NERC Press.

Gall M. D. Borg, W. R. & Gall, J. P. (2007). Educational Research: An Introduction (8th ed.). New York: Longman.

Green, K. E. (1996). Application of the Rasch model to evaluation of survey data quality. New Directions Evaluation, 70, 81-92.

Green, K. E. & Franton, C. G. (2002). Survey development and validation with Rasch model. A paper presented at the international conference on questionnaire development, evaluation and testing. Charleston, S.C. November 14-17, 2002.

Gruijter, D. N. & Kamp, J. T. V. (2001). Statistics test theory for education and psychology. Boston Kluwer.

Guven, B. (2006). Secondary geography course' attitude scale development study. Kastamonu Educational Journal, 14(2), 527-536.

Harris, D. J. (1991). A comparison of Angoff's design I and design II for vertical equating using traditional and IRT methodology. Journal of Educational Measurement, 28, 221-235.

Hopkin, J. (2006). "Assessment for learning in geography teaching geography, 25(1), 42-43.

Kind, P. M., Jones, K. & Barmby, P. (2007). Developing attitude towards science measures. International Journal of Science Education, 29(7), 351-352.

Marais, L. Styles, I. & Andrich, D. (2011). Report on Rasch analysis of the Academic Perception of the Teaching Environment (APOTE) for 2010. Pearson Psychometric Laboratory Graduate School of Education. The University of Western Australia.

Metu, I. C. (2015). Development and standardization of teachers' job satisfaction scale using the Rasch model in Enugu state. Unpublished UNN Thesis.

Miller, M. (2005). Teaching and learning in affective domain. In M. Drey (Ed.). Emerging perspectives on learning, teaching and technology. Retrieved October 14, 2009 from http;\\projects.coe.ugo.edu.\ep\tt\.

Nworgu, B. G. (2010). Differential item functioning: A critical issue in regional quality assurance. A lead paper presented at 29th conference of the association for educational assessment in Africa in Nairobi.

Oformata, G. E. K. (2008). Geographic thought: 2nd edition. Enugu: Jamoe Enterprises.

Okpala, J. (2000). The image of school geography in Nigeria. Past, present and future. 186 International Research in Geographical and Environmental Education.

Oluwatayo, O. C. (2006). Education measurement for teachers. Lagos: Thomas Nelson (Nig.) Ltd.

Philips, J. M. (2011). Roles of goal Orientation, ability, need for achievement, locus of control in the self-efficacy and goal-setting process. Journal of Applied Psychology, 82, 792-802

Rasch, G. (1960). Probabilistic models for intelligence and attainment tests (expanded edition). Chicago, IL: British Journal of Mathematics and Statistical Psychology, 19, 49-57.

Saed, S. Hind, H & Mutasem, A. (2013). Validation of a scale of attitude toward science across countries using Rasch model: Findings from Timss. Journal of Baltic Science Education, 12(5), 692-702.

Thorndike, R.L. (1977). Measurement and evaluation in psychology and education. New York.

Wiberg, M. (2004) Test in the Swidish driving-license test (No. 50) Umea: Kluwer Academic Publications.

Wright, B. D. & Linacre, J. (1994). Reasonable mean-square fit values. Rasch measurement Transaction, 8(3), 370.