Canadian Journal of Educational and Social Studies
Vol. 3(4), 2023, pp. 120-132

Development of Attitude Towards Assessment Test for Secondary School Students in Delta State

Andrew Moyioritse Megbele & John Nwanibeze Odili & Patrick Uzor Osadebe

[1,2,3] Measurement and Evaluation, Delta State University, Abraka, Nigeria
Correspondence: Andrew Moyioritse Megbele, Measurement and Evaluation, Delta State University, Abraka, Nigeria
Email: megbele.andrew@delsu.edu.ng

DOI: 10.53103/cjess.v3i4.162

Abstract

This study developed Attitude Towards Assessment Test using the rating scale model of the Item Response Theory. Three research questions guided the study. An instrumentation research design was adopted. The population consisted of secondary school students in Delta State, Nigeria. The sample size comprised 1,000 students, selected through simple random and non-probability cluster sampling techniques. The researcher developed the test, which comprises a total of 60 items. The Person separation reliability index was used to answer research question 1 and 2 on convergent and discriminant validities respectively. The Categorical Confirmatory Factor Analysis (CCFA) was used to answer research question 3 on construct validity. The Statistical Package for Social Sciences (SPSS) version 26 was used to analyse the data The findings of this study revealed that the Attitude Towards Assessment Test has an acceptable convergent validity; that the Attitude Towards Assessment Test has an acceptable discriminant validity; the instrument has a good construct validity. The study recommended that the Attitude Towards Assessment Test should be used to assess secondary school students' attitude towards assessment.

Keywords: Rating Scale Model, IRT, Attitude Towards Assessment Test, Secondary School Students Introduction

Introduction

Education is widely acknowledged as a crucial tool for fostering social, political, scientific, and technological progress. Consequently, no society can afford to disregard the importance of educating its citizens, as neglecting this aspect could impede overall development at a slow pace (Jessa, 2017). Education, in its broad sense, encompasses all the processes through which a child or young adult nurtures their skills, attitudes, and other behavioral traits that hold significance within their respective societies. It involves purposeful instruction aimed at preparing the younger generation for a life that is both beneficial to themselves and to the community they are part of.

The continuous assessment mode of evaluation currently used in the educational system of Nigeria was introduced to take care of the lapses that characterized the traditional

mode of assessment. Some of these lapses are inherent in the domain of assessment. The focus of the traditional mode of assessment was on the cognitive domain of learning, such that at the end of a particular term or school session, students are given a set of questions aimed at assessing the extent to which they have learnt a particular school subject. This system of assessment was inherited from the colonial educational system, which based assessment only on the cognitive domain at the expense of the affective and psychomotor domains (Osadebe & Jessa, 2018).

The exclusive focus on the cognitive domain also meant that the instrument used for the assessment was limited to multiple-choice and essay test questions. No room was given to such assessment tools as observation, checklist and rating scales., which are instruments for measuring the affective and psychomotor domains of learning, covered by the continuous assessment mode of assessment.

Assessing the extent to which students have obtained learning in the cognitive, affective and psychomotor domains makes learning to be comprehensive. This is because, students are not only expected to understand the contents of what is being learnt, they are also expected to appreciate learning and practise what was learnt. Behaviour change (learning) can only be validated when students are able to practise what they have learnt. It is therefore important that the three domains of learning should be systematically and comprehensively assessed during and after learning. The focus of this study is on the affective domain of learning.

The affective domain plays a significant role in the exploration and adaptation of human interests, attitudes, values, and appreciation. Quantifying affective learning outcomes through traditional testing methods is challenging, as they rely more on qualitative self-reflection. The taxonomy of the affective domain consists of five levels: receiving, responding, valuing, organization, and characterization, ranging from the lowest to the highest level. To assess changes in affective learning, this taxonomy was applied to written self-evaluations.

Receiving refers to the awareness of the need and willingness to attentively listen and pay attention, such as respectfully listening to others and remembering names of newly introduced people. Responding involves actively participating in learning and can include compliance, willingness, or satisfaction in response. Examples include engaging in class discussions, asking questions to enhance understanding, and following safety protocols.

Valuing is the ability to judge the worth or value of something, including objects, phenomena, behaviors, or information, and expressing it clearly. It ranges from simple acceptance to a more complex state of commitment. Learners' internalized values are often reflected in their explicit and identifiable behaviors. Examples include expressing convictions about the democratic process, respecting diversity, addressing value conflicts, proposing social improvement plans, and fulfilling commitments.

Organization involves comparing and classifying values, resolving conflicts

between them, and creating a unique value system that emphasizes comparison, relevance, and integration. Examples include recognizing the need for a balance between freedom and responsibility, understanding the importance of system planning in problem-solving, accepting ethical standards, creating life plans aligned with abilities, interests, and beliefs, effectively prioritizing time for personal, organizational, and family needs.

Characterization refers to the establishment of a value system that guides learner behavior consistently and predictably, becoming the most important aspect of the learner's identity. Teaching objectives focus on individual, social, and emotional patterns that learners adjust. Examples include working independently, collaborating in group activities, using objective methods to solve problems, practicing professional ethics, being open to modifying beliefs and changing behavior based on new evidence, and valuing people beyond superficial characteristics.

In line with the above taxonomies of the affective domain of learning, various attributes can be assessed. These attributes include (but not limited to) attitude towards education, academic self-efficacy, academic motivation, career aspiration and test anxiety. No single assessment tool can be used to assess all these behaviours. Hence, the focus of this study is on attitude. Attitude is defined as a psychological tendency that is expressed by evaluating a particular entity with some degree of favour or disfavour'. Inherent in this definition is the idea that reporting an attitude involves the expression of an evaluative judgement about a stimulus object. In other words, reporting an attitude involves making a decision concerning liking and disliking, approving and disapproving or favouring and disfavouring a particular issue, object or person.

The principle of consistency is an underlying assumption regarding the relationship between attitudes and behavior. It suggests that we expect a person's behavior to align with their held attitudes, as people are considered rational and strive to behave consistently. The strength of an attitude often predicts behavior, with stronger attitudes having a greater likelihood of influencing behavior. Assessing attitudes as a human behavior is complex and cannot be done in isolation; it requires consideration of the individual's disposition and the specific object or context to which the attitude relates.

In the context of this study, assessment refers to a series of measures used to evaluate a complex attribute of an individual or group. It involves gathering and interpreting information to determine the level of student achievement in learning goals. Assessments help identify individual strengths and weaknesses, enabling educators to provide tailored academic support. Within the classroom, assessments serve the purpose of improving teaching and learning. They provide feedback on instructional effectiveness and allow students to gauge their progress. Classroom assessment has two primary functions: to determine the success of learning and to clarify teachers' expectations of students. Therefore, students are likely to benefit from assessment outcomes, but the extent of their benefit may depend on their attitude towards assessment.

It has been observed that a mere mention of the word "examination" most times appears scary to students irrespective of their level of academic brilliancy. Different reactions can be elicited from the school environment during this tense period. No more partying, sickness cautiously creep in, students begin to look for ways to complete their lesson notes and cover gaps in learning. Attitude towards assessment therefore, is a predictor of students' reaction towards assessment.

As important as students' attitude towards assessment on the outcome of assessment, a search through the literature revealed that to the best knowledge of the researcher, no assessment tool exists for the measurement of students' attitude towards assessment, particularly in the context of students in Delta State Nigeria. This is the crux of the study, to develop a rating scale that can be used to measure secondary school students' attitude towards assessment in Delta State, Nigeria. In developing the rating scale, the researcher wishes to use Item Response Theory rating scale polytomous model in the calibration of the instrument.

Item Response Theory (IRT) is widely recognized as the primary psychometric approach used for constructing, scoring, and analyzing assessments. Its numerous advantages have made it the preferred method for large-scale assessments. Although IRT is often regarded as a singular approach, it is actually a family of models that continues to expand. Within Item Response Theory, two main models are utilized.

The first model is focused on multiple-choice items, which typically present respondents with 4 to 5 options but offer only two possible scores: either correct or incorrect. Examples of this model include True/False or Yes/No items, commonly found in surveys or inventories. In contrast, the second model, known as polytomous models, applies to items with more than two possible scores. Likert-type items, where respondents rate on a scale of 1 to 5, and partial credit items, where scores on an essay can range from 0 to 5 points, are common examples of polytomous models. IRT's versatility allows for the analysis of different types of items, accommodating both dichotomous and polytomous scoring systems. By employing IRT, assessments can leverage its flexibility to yield valuable insights into respondents' abilities or attributes.

Items used in psychological testing contexts such as attitude, interest, personality surveys, inventories, or questionnaires often have a consistent structural form across all of the items for a given measurement instrument. Often this is a Likert or similar type of rating scale format where people are asked to respond to an item using a pre-defined set of responses and where the same set of response alternatives is applied to all the items in the test. A typical example of such a response format is where people are asked to respond to each item on a test using a four-point scale ranging from Strongly Disagree to Strongly Agree.

The rating scale model (RSM) is uniquely suited to the above type of response format. The fundamental assumption of the model, which distinguishes it from other

polytomous IRT models, is that the use of a single rating response format across items in a test implies that the response format is intended to function in the same way (that is, be consistently applied) across all items, since the categories have a constant definition. The implied intent is reflected in the model by the exposition of a common parametric form for all the items using a given rating scale. The result is that, in the RSM, any given category on the response scale is modelled to be the same "size" across all items, although within an item on the scale the different response categories need not be the same size as each other. The RSM is a special case of Rasch's (as cited in Famogbiyele, 2017) polytomous model in that it begins with the postulate that a single dimension is modelled in contrast to Rasch's general model which is multidimensional in terms of both the person parameter and with respect to item response categories.

In practical applications, the graded response model and the generalized partial credit model are commonly utilized polytomous IRT models. These models are specifically designed to analyze nominal and ordinal variables, such as rating scales, as opposed to the traditional linear factor analytic (FA) approach which is more suitable for continuous variables. Researchers like Maydeu-Olivares, Cai, and Hernández (2011) have indicated that IRT models exhibit better fits to polytomous ordinal data compared to FA models, primarily due to their ability to accommodate a higher number of parameters.

In addition to assessing overall model fit and item parameters, IRT modeling generates other valuable statistics, such as item information and person-level model fit, which are instrumental in the development of measurement instruments. These statistics aid in evaluating the quality and effectiveness of the instrument.

Despite the promising potential of polytomous IRT models in instrument development, their performance has not been thoroughly evaluated, particularly within the Nigerian context. The extent to which these models can effectively support the development of measurement instruments in this specific setting remains an area that requires further investigation. Hence, the need for the present study, which is aimed at applying the rating scale Item Response Theory model to develop Attitude Towards Assessment Test.

Research Questions

The study was guided by the following research questions:

1. What is the convergent validity of the Attitude Towards Assessment Test?
2. What is the discriminant validity of the Attitude Towards Assessment Test?
3. What is the construct validity of the Attitude Towards Assessment Test?

Methods

Designs

This study adopted an instrumentation research design. This design is suitable for the study because the intention of the researcher was geared towards the development of attitude towards assessment test for secondary school students in Delta State.

Participants Selection

The sample size comprised 1,000 students. This sample size is based on the recommendation of Linacre (1994), who stated that larger sample sizes are needed for more complex models. A total of 40 students in each local government area of the state was selected to make a total of 1,000 students. This was done through simple random and non-probability cluster sampling techniques. In this case, the schools in each Local Government Area of the state were treated as clusters, such that the researcher randomly selected one school in each Local Government Area to make a total of 25 schools. This was done through a simple random sampling technique of the balloting method. Using this procedure, the researcher wrote the names of all the schools in each local government area on pieces of paper, fold and pour them into a container. He then shuffled them and picked one piece of paper from the container. Schools picked from this process were the selected school in that Local Government Area. This was done for all Local Government Areas until all the 25 schools (one for each Local Government Area) were selected. The above procedure produced 25 clusters, one for each Local Government Area. For each cluster, the researcher randomly selected one classroom out of the various classrooms in the secondary schools. All the students in the selected classroom were used for the study.

Development of Instrument

The researcher developed an Attitude Towards Assessment Test (ATAT). The test comprises a total of 60 items which are distributed according to the components of attitude, such as Cognitive component (having 20 items), Affective component (20 items) and Behavioural components (20 items). Some of the items were phrased in negative forms and were represented with letter **(N)** at the end of the statement. The instrument was structured on a 4-point Likert-type scale, ranging from 1 for strongly disagree to 4 for strongly agree. The expected score for the instrument is between a minimum of 60 and a maximum of 240. The negative items were reverse coded before the item analysis.

After the test development, the test items were subjected to scrutiny by the researcher's supervisors, who are experts in Measurement and Evaluation and two other experts in Guidance and Counselling. This is to gain their insight as to the appropriate matching of the items to the domains and construct areas arrangement. The experts were

required to critically review the test and made some suggestions for the improvement of the test items.

In developing the ATAT, the researcher relied on the recommendation of McCoach et al. (2013). The steps included:

Step 1: Specification of the purpose of the instrument
Step 2: Confirmation that there are no existing instruments that adequately served similar purpose
Step 3: Description of the constructs and provision of preliminary conceptual definitions
Step 4: Development of operational definitions
Step 5: Selection of a scaling technique
Step 6: Matching of items back to the dimensions/constructs, ensuring adequate content representation on each dimension
Step 7: Conducting of a judgmental review of items
Step 8: Development of directions for responding; creation of final version of survey
Step 9: Gathering of pilot data from a sample that is as representative as possible of target population
Step 10: Analysis of pilot data
Step 11: Revision of the instrument based on the initial pilot data analyses
Step 12: Preparation of a test manual or a manuscript

The ATAT was administered to the students directly by the researcher, with the help of 5 research assistants, who were trained on the purpose of the study. The research team visited the schools personally before the testing date to make their intention known to the principal or head of the school and to obtain permission. The research assistants were briefed on the purpose of the study and how to approach testees. The data were collected on the spot from the respondents.

Data Analysis

The Person separation reliability index was used to answer research question 1 and 2 on convergent and discriminant validities respectively. The Categorical Confirmatory Factor Analysis (CCFA) was used to answer research question 3, 6 and 7 on construct validity, evidence of unidimensionality and local independence respectively. The Statistical Package for Social Sciences (SPSS) version 26 was used to analyse the data to answer the research questions 1, 2, 3, 6 and 7.

The Rasch Rating Scale Model was used to answer research 4, which assessed person and item reliability, item statistics and ordering of response categories. Two fit

indices including the infit and outfit mean square (MNSQ) statistics was used to answer research question 5. The data that were used to answer research questions 4 and 5 were analysed with the aid of the Jmetrik IRT software.

Results

Research Question 1: What is the convergent validity of the Attitude Towards Assessment Test?

Table 1: Convergent validity of the attitude towards assessment test

Component	No of Items	*N*	Mean	*SD*	Convergent validity	
					Range of Correlation	Scaling Success (Percent)
Total Score	60	1,000	3.03	0.89	-	-
Cognitive	20	1,000	3.15	0.86	0.304-0.515	75(15/20)
Affective	20	1,000	2.79	0.95	0.304-0.538	95(19/20)
Behavioural	20	1,000	3.16	0.85	0.300-0.492	95(19/20)

Table 1 shows the convergent validity of the Attitude Towards Assessment Test. From the result, majority of the items (15 out of 20 items for cognitive component; 19 out of 20 items for affective component; and 19 out of 20 items for behavioural component) show a good convergent validity because the correlation of all the items within the scales are within the range of 0.300 and 0.700, which indicate that the scales have good convergence validity and are statistically significant at 0.05 level of significance (See appendix).

Research Question 2: What is the discriminant validity of the Attitude Towards Assessment Test?

Table 2: Discriminant validity of the attitude towards assessment test

Component	No of Items	*n*	Mean	*SD*	Discriminant validity	
					Range of Correlation	Scaling Success (Percent)
Total Score	60	1,000	3.03	0.89	-	-
Cognitive	20	1,000	3.15	0.86	0.000-0.296	25(5/20)
Affective	20	1,000	2.79	0.95	0.000-0.298	5(1/20)
Behavioural	20	1,000	3.16	0.85	0.000-0.298	5(1/20)

Table 2 shows the discriminant validity of the Attitude Towards Assessment Test. From the result, majority of the items (5 out of 20 items for cognitive component; 1 out of 20 items for affective component; and 1 out of 20 items for behavioural component) show a good discriminant validity because the correlation of all the items within the scales are

below 0.300, which indicate that the scales have good discriminant validity.

Research Question 3: What is the construct validity of the Attitude Towards Assessment Test?

Table 3: Factor loadings, item selection and communalities of the attitude towards assessment test

S/N	Item No	Item	F_1	F_2	F_3	h^2
1	BC58	When taking tests, I take every question seriously	0.671			0.490
2	BC60	I read and try to understand the instructions before starting to answer the questions	0.654			0.513
3	BC55	I will encourage students to prepare very well for their examination	0.652			0.486
4	BC50	I prepare very well for every examination	0.635			0.489
5	BC47	I often work very hard to achieve success during assessment	0.623			0.401
6	BC59	When taking tests, I often pay attention to the details	0.611			0.412
7	BC51	I will attend an exam preparatory class if it is available	0.580			0.372
8	BC56	I always take my school examination seriously	0.533			0.396
9	AC28	Assessment is one of my favourite activities in school		0.652		0.441
10	AC24	I would not mind writing an examination every day		0.650		0.428
11	AC23	I like assessment		0.607		0.440
12	AC30	Assessment is always fun for me		0.582		0.352
13	AC26	I am always excited when I think of an examination		0.581		0.370
14	AC25	Assessment interests me		0.576		0.435
15	AC27	I look forward to the next examination		0.446		0.326
16	CC11	Classwork is a kind of assessment			0.768	0.598
17	CC10	Assignment is a kind of assessment			0.668	0.481
18	CC12	Assessment is also known as an examination			0.664	0.452
19	CC14	Assessment carried out by WAEC, NECO or JAMB are known as external assessment or examination			0.651	0.454
20	CC13	Assessment carried out by teachers are known as internal assessment or examination			0.533	0.327

Table 3 shows the factor loadings, item selection and communalities of the Attitude Towards Assessment Test, which was carried out to examine the construct validity of the Attitude Towards Assessment Test. The result shows that as hypothesized, the instrument has three factors namely Cognitive, affective and behavioural. Items on these three components were appropriately loaded in their respective factors with their associated communalities. From the result, behavioural component had factor loadings ranging from 0.533 to 0.671; affective components had factor loadings ranging from 0.446 to 0.652; while cognitive component had factor loadings ranging from 0.533 to 0.768. The result further showed that out of the overall 60 items in the scale, a total of 40 items were removed due to their weakness to the construct. A total of 20 items were retained (8 items for behavioural component; 7 items for affective component; and 5 items for cognitive component). The result implies that the instrument has a good construct validity.

Discussion

The first finding shows that the Attitude Towards Assessment Test has an acceptable convergent validity. The finding revealed that majority of the items (15 out of 20 items for cognitive component; 19 out of 20 items for affective component; and 19 out of 20 items for behavioural component) show a good convergent validity because the correlation of all the items within the scales are within the range of 0.300 and 0.700, which indicate that the scales have good convergent validity and are statistically significant at 0.05 level of significance. The finding implies that the three components of the Attitude Towards Assessment Test capture a common construct.

This finding is consistent with the study conducted by Jafari et al. (2012) who employed the Rasch rating scale model to reassess the psychometric properties of the Persian version of the PedsQLTM 4.0 Generic Core Scales in school children. They discovered that the instrument exhibited satisfactory internal consistency and demonstrated excellent convergent and discriminant validity. The observation of convergent validity in the instrument aligns with the perspective of Campbell and Fiske (as cited in French & Vo, 2020), who argue that "at the current stage of psychological progress, the crucial requirement is the demonstration of some convergence." Nunnally (as cited in Eichenbaum, et al., 2019) also suggests that highly correlated measures indicate high convergence, while measures with near-zero correlations imply weak or no convergence (p. 91).

Research evidence indicates that the actual levels of convergent validity in organizational research display considerable variation. Although the construct validity of measures and the convergent validity of alternatives may be less developed in the early stages of studying a phenomenon (Nunnally, as cited in Eichenbaum, et al., 2019), inadequate convergent validity at any stage can impact the magnitudes and interpretability

of research findings.

The second finding revealed that the Attitude Towards Assessment Test has an acceptable discriminant validity. The finding showed that majority of the items (5 out of 20 items for cognitive component; 1 out of 20 items for affective component; and 1 out of 20 items for behavioural component) show a good discriminant validity because the correlation of all the items within the scales are below 0.300, which indicate that the scales have good discriminant validity. This means that the test is able to account for more variance in the observed variables associated with. If this is not the case, then the validity of the individual indicators and of the construct would have been questionable.

The importance of accurately assessing discriminant validity has been repeatedly emphasized across multiple disciplines (Henseler, et al., 2015). For example, the discipline of personality psychology has critically evaluated the proliferation of theoretically indistinguishable constructs on multiple occasions. Additionally, Harter and Schmidt (as cited in French & Vo, 2020) discuss the importance of discriminant validity in terms of needing to accurately determine if constructs that are argued to be conceptually different are indeed empirically distinct so as to ensure that newer constructs are indeed new and not simply renaming of behaviours that have been previously and thoroughly studied. Lastly, the accurate application of discriminant validity tests within empirical studies has also been assessed and critically questioned (e.g., Farrell, 2010; Shiue et al., 2011).

This finding is consistent with the study conducted by Jafari et al. (2012), where they utilized the Rasch rating scale model to reevaluate the psychometric properties of the Persian version of the PedsQLTM 4.0 Generic Core Scales in school children. They discovered that the instrument demonstrated satisfactory internal consistency and exhibited excellent convergent and discriminant validity. The finding aligns with Bove et al. (2009), who emphasized that the absence of discriminant validity poses significant challenges for subsequent analysis. However, it contrasts with the findings of Patterson and Ahmed (as cited in Eichenbaum, et al., 2019), who reported results that lacked discriminant validity, leading to interpretational difficulties in three out of the 21 hypotheses examined.

The third finding showed that the instrument has three factors namely Cognitive, affective and behavioural. Items on these three components were appropriately loaded in their respective factors with their associated communalities. From the result, behavioural component had factor loadings ranging from 0.533 to 0.671; affective components had factor loadings ranging from 0.446 to 0.652; while cognitive component had factor loadings ranging from 0.533 to 0.768. The result further showed that out of the overall 60 items in the scale, a total of 40 items were removed due to their weakness to the construct. A total of 20 items were retained (8 items for behavioural component; 7 items for affective component; and 5 items for cognitive component). The result implies that the instrument has a good construct validity.

The above finding is in line with Gregory and Noto (2018), who developed an

attitudinal instrument development to assess the cognitive, affective, and behavioural domains of teacher attitudes toward teaching all students. Their study revealed the test is valid and reliable and can be used to measure the attitudes toward teaching all students. The finding also agrees with Seker (2011), who developed a questionnaire to assess student attitudes towards school by describing the factors that affect these attitudes and found that the instrument is suitable in the light of the findings of his study, and its scales are thought to affect students' attitudes towards school.

Conclusion and Recommendations

Based on the findings of the study, it was concluded that the Attitude Towards Assessment Test developed by the researcher had a good convergent, discriminate and construct validities for all the components that made up the test. Based on the findings of the study, the researcher recommended the following:

1. The Attitude Towards Assessment Test should be used to assess secondary school students' attitude towards assessment
2. The test can be used by teachers prior to any examination
3. The test can be used by examination bodies for the assessment of students in the affective domain
4. Other researchers who wish to carry out a study on students' attitude towards examination can use the test

References

Bove, L.L., Pervan, S.J., Beatty, S.E., & Shiu, E. (2009). Service worker role in encouraging customer organizational citizenship behaviours. *Journal of Business Research,* 62 (7), 698-705.

Eichenbaum, A., Marcus, D., & French, B. F. (2019). Item response theory analysis of the psychopathic personality inventory-revised. *Assessment* 26, 1046–1058. https://doi.org/10.1177/1073191117715729

Famogbiyele, O.T. (2017). Nigerian senior secondary school students' attitude and motivation towards English Language learning: Case study of Northern and Southern geopolitical zones. *Asian Journal of Educational Research, 5*(4), 9-29.

Farrell, A. M. (2010). Insufficient discriminant validity: A comment on Bove, Pervan, Beatty, and Shiu (2009). *Journal of Business Research,* 63, 324–327.

French, B. F., & Vo, T. T. (2020). Differential item functioning of a truancy assessment. *J. Psychoeducational Assess.* 38 (5), 642–648. https://doi.org/10.1177/0734282919863215

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science,* 43, 115–135.

Jafari, P., Bagheri, Z., Ayatollahi, S. M. T., & Soltani, Z. (2012). Using Rasch rating scale

model to reassess the psychometric properties of the Persian version of the PedsQLTM 4.0 Generic Core Scales in school children. *Health and Quality of Life Outcomes,* 10(27), 1-11.

Jessa, M. O. (2017). Availability, adequacy and utilization of educational facilities and school administrators' effectiveness in colleges of education in Delta State.

Maydeu-Olivares, A., Cai, L., & Hernández, A. (2011). comparing the fit of item response theory and factor analysis models. Struct. Equation Model. *A Multidisciplinary J.* 18(3), 333–356. https://doi.org/10.1080/10705511.2011.581993

Osadebe, P. U., & Jessa, M. O. (2018). Development of social studies achievement test for assessment of secondary school students. *European Journal of Open Education and E-Learning Studies*, 3(1), 104-124.

Penfield, R. D. (2014). An NCME instructional module on polytomous item response theory models. *Educ. Meas. Issues Pract*. *33*(1), 36–48. https://doi.org/10.1111/emip.12023

Shiue, Edward, Simon J. Pervan, Liliana L. Bove, & Sharon E. Beatty (2011). Reflections on discriminant validity: Reexamining the Bove et al. (2009) findings. *Journal of Business Research,* 64 (5), 497–500.

Yambi, T. A. C. (2018). Assessment and evaluation in education. Retrieved from https://www.researchgate.net/publication/342918149_ASSESSMENT_AND_EVALUATION_IN_EDUCATION#fullTextFileContent