https://www.eejournals.org Open Access
EURASIAN EXPERIMENT JOURNAL OF SCIENTIFIC AND APPLIED RESEARCH (EEJSAR) ISSN: 2992-4146
©EEJSAR Publications Volume 4 Issue 2 2023

Development and Validation of Physics Diagnostic Test Using Item Response Theory

Nkechi Patricia-Mary Esomonu[1] and Chisom Evangelin Ndubuisi[2]

Department of Educational Foundation, Faculty of Education, Nnamdi Azikiwe University, Awka, Nigeria.
E-mail: npm.esomonu@unizik.edu.ng; Phone: +2348026422569
E-mail: nevanchisom@gmail.com Phone: +2349155809849

ABSTRACT

The current assessment practice by physics teachers is the used of open-ended and multiple choice test items on content mastery and these assessment techniques does not provide insight into students' cognitive skills in problem solving and their understanding of physics concepts. This is because the assessments are designed for ranking, predicting and sorting that usually lack the details needed to target specific learning skills for improvement. This research is therefore aimed at study how to develop a valid and reliable physics diagnostic test using item response theory. The study was conducted in Anambra State. Two research questions were formulated to guide the study. An instrumentation research design is for this study. The population of the study consisted of 2259 SS2 students offering physics. The sample for the study consisted of 1800 SS2 students (852 males and 948 females). The instrument for data collection was Physics Diagnostic Test (PDT). Physics Diagnostic Test (PDT) consisted of 100 item questions culled from past WAEC and NECO physics questions. The instrument was subjected to validation. The reliability coefficient for PDT was .89 using Kuder Richardson (K-R-20). The research question one was answered using DIMTEST Statistics in DIMPACK software specifically designed for dimensionality assessment of measurement instrument. Research question two was answered using information criteria statistics in **MPLUS** software. Analysis of Covariance (ANCOVA) was used to test the hypotheses at $p<0.05$ level of significance. The result of the study revealed that the Physics multiple choice test items is multidimensional since p-value is <.05 level of significance. Furthermore, the difference between the number of items in Partitioning Subtest (PT) and the Assessment Subtest (AT) in a test is significant suggesting evidence of multidimensionality. Using Yen's Q3 to screen items for local dependence, 87% item residual correlations were below absolute value of 0.2. This indicates that the local independence assumption of the IRT was not grossly violated. The empirical of the instrument is .86. This showed that the instrument is reliable. The findings enabled the researchers to conclude that the final form of the multiple choice test is valid and reliable. Thus, it was recommended among other things that physics diagnostic test should be used by physics teachers in schools. Particularly in assessment of learning skills students are deficient in so that remedial help can be provided by teachers to them for better academic achievement.

Keywords: Development, validation, physics, diagnostic test, IRT

INTRODUCTION

In the assessment of students' learning progress, information about student skills in the learning outcomes intended in the curriculum is necessary. Thus, the current assessment practice by physics teachers is the used of open-ended and multiple choice test items on content mastery and these assessment techniques does not provide insight into

students' cognitive skills in problem solving and their understanding of physics concepts [1-4]. Furthermore, the formative and summative assessments instruments that teachers regularly administer in their classrooms provide limited information about students' cognitive learning skills [5-7]. Implicatively, the current assessments used by teachers provide no direct and immediate feedback to the teachers and students particularly on learning skills students are deficient in. Hence, teachers' classroom assessment practices have to be well integrated with instruction in order to provide valid and detailed information about students' strengths and weaknesses of the cognitive processes skill. Normally, the test specifications for assessments in classroom only specify content requirements and no explicit consideration is given to the type of cognitive learning skill that underlies a curriculum. The inability of most teachers to assess diagnostic skills of secondary school students has been widely reported in the study of [8-9]. Although teachers can predict the overall performance of the students through their own classroom assessment, the results do not tell them much about their students' cognitive learning skills in item performance as students tend to focus on recall to get through the task [10-14]. Most teachers associate diagnostic information with reporting at the individual achievement level with limited information of students' structural knowledge, procedural skills and abilities elicited from assessment [15-17]. Hence the aim of this study to develop a valid and reliable physics diagnostic assessment test.

Diagnostic assessment has been to advocated by measurement and evaluation experts the ways students interpret mathematical problems and construct strategies in problem solving, including the domain of algebra [2]. Diagnostic assessment provides information about the strengths and weaknesses of students in any subject. Teachers need more information about the cognitive strengths and weaknesses of specific knowledge and skills individual student demonstrated on assessment to improve their instructional planning. The information obtained from diagnostic assessment is different from what the current standardized assessment provides. Diagnostic assessment is design to ensures that the cognitive attributes of interest are explicitly targeted during items and test development. The information that reflects students' cognitive strengths and weaknesses elicited from diagnostic assessment is able to guide teachers to help the students in their knowledge and performance in the algebraic expressions learning. Diagnostic assessment helps to increase the accuracy and reliability of the determination of the students' cognitive attribute skills, and can be utilized to improve both the teaching and the learning processes. Thus, the diagnostics skills in physics the researchers sought to investigate in this study that physics teachers need to assess physics students include: measurement skill, thinking skill, graphical skill, communication skill, and calculation skill [18-21].

The aforementioned skills as mentioned by physics experts and WAEC Chief Examiner reports are critical for effective understanding of the subject [14]. When students are deficient in any of the skill learning the subject become difficult and frustration. Diagnostic assessment of the above skills provides valuable feedback to teachers, which could help them identify what skills students have or have not mastered as well as to decide how teaching and learning needs to be adapted to the students' needs. Thus, there need for well blended classroom assessment that meet 21st century technological advancement skills of physics students cannot be over state in the cotemporary time. Hence, the researchers sought to develop a valid and reliable physics diagnostic test using item response theory.

Although, various researchers have developed diagnostic tests across different subjects in education like mathematics, language, music, and economics [15-21]. Educational researchers have paid little attention to the development and validation of diagnostic test in physics. In essence, the development and validation of physics diagnostic test has not been researched on, thereby constituting an educational gap. It is against this background that this present study was conceived.

Purpose of the study: The purpose of this study is to develop and validate a physics diagnostic test using IRT. Specifically, the study sought to determine the: (i) dimensionality and Local Independence of items of the Physics Diagnostic Test; (ii) empirical reliability of the physics diagnostic test.

Research question 1: What is the dimensionality and Local independence of the items of the Physics Diagnostic Test**?**

Research question 2: What is the empirical reliability of the physics diagnostic test?

Methods

The design of this study is an instrumentation research design. The study was conducted in Anambra State, Nigeria. The population of the study comprised 2259 (1518 females and 741 males) SS2 students offering physics in 2021/2022 academic session. The sample for the study consisted of 1800 SS2 students offering physics in 2021/2022 academic session comprising 652 males and 948 females. The sample for the study was drawn through multi-stage procedure. At the state level, simple random sampling techniques was used to sample four educational zones out of six educational zones in the state. At the second stage, a simple random technique was used to sample two local government areas from each educational zone. At the third stage, a simple random sampling technique was used to

sample five public secondary schools from each Local Government Area. Then, all the SS2 students offering physics in the sampled schools were involved. Instrument for data collection was Physics Diagnostic Test (PDT). The instrument was constructed based on the guidelines for development of diagnostic test. The distribution of the items in the instrument was based on the report by the Chief Examiner in the WAEC report sheets of 2020, 2021, and 2022 (20% of the students' weakness in Physics is in the area of calculation skill, 20%, thinking skill, 20% in communication skill, 20% on graphic interpretation skill). In constructing the items of the instrument, a table of specifications were used. The instrument was presented to one expert in Measurement and Evaluation as well as two subject experts. Based on their suggestions the instrument was be improved upon. PDT scores from the study was analysed using Kuder Richardson (K-R-20). The reliability coefficient of PDT is 0.89. The instrument was administered to the SS2 students offering physics. The scores obtained from the students was subjected to analysis. The research question was answered using DIMTEST Statistics in **DIMPACK software** specifically designed for dimensionality assessment of measurement instrument. DIMTEST statistics was used to test dimensionality. A value greater than .05 implies multidimensionality [12]. Empirical reliability above 0.89 indicates high reliability [11].

RESULTS

Research question 1: What is the dimensionality and Local independence of the items of the Physics Diagnostic Test**?**

To answer the research question one, two basic assumptions of item response theory- dimensionality and local independence were examined.The dimensionality assumption was investigated using DIMTEST statistics.

Table 1: Dimtest Statistics of Physics Diagnostic multiple choice test items

TL	TGbar	T	AT	PT	P-value
18.1871	2.9306	14.1880	27	73	0.0000

The result in Table 1 above indicated that Physics Diagnostic multiple choice test items is multidimensional since p-value is <.05 level of significance. Using Yen's Q3 statistics to screen items for local dependence, 88% item residual correlations were below absolute value of 0.2. This indicates that the local independence assumption of the IRT was not grossly violated.

Research question 2: What is the empirical reliability of the physics diagnostic test?

Item	SE	Item	SE	Item	SE	Item	SE	Item	SE
1	.01	21	.31	41	.67	61	.01	81	.01
2	.21	22	.01	42	.00	62	.46	82	.47
3	.02	23	.36	43	.68	63	.01	83	.00
4	.31	24	.02	44	.00	64	.42	84	.02
5	.03	25	.11	45	.01	65	.01	85	.33
6	.66	26	.00	46	.71	66	.00	86	.03
7	.01	27	.46	47	.01	67	.81	87	.00
8	.15	28	.00	48	.04	68	.01	88	.71
9	.00	29	.07	49	.03	69	.04	89	.69
10	.22	30	.01	50	.31	70	.28	90	.01
11	.14	31	.02	51	.02	71	.02	91	.11
12	.03	32	.63	52	.68	72	.38	92	.00
13	.01	33	.00	53	.00	73	.02	93	.01
14	.02	34	.01	54	.01	74	.08	94	.81
15	.61	35	..06	55	.38	75	.04	95	.00
16	.00	36	.00	56	.00	76	.02	96	.03
17	.20	37	.67	57	.00	77	.61	97	.52
18	.00	38	.01	58	.01	78	.00	98	.01
19	.01	39	.02	59	.60	79	.00	99	.02
20	.30	40	02	60	.00	80	.01	100	.36

Table 2: The empirical reliability of the physics diagnostic multiple choice test items

The average SEM=.379
Empirical reliability= 1-(.379)²
Empirical reliability =.86
Table 2 showed that 40 items had standard error of measurement above .05 and were rejected. The empirical reliability of the instrument is .86. This showed that the instrument is reliable.

DISCUSSION

The findings of the study indicated that the underlying latent ability of examinees responses to the instrument is multidimensional. Furthermore, if the difference between the number of items in Partitioning Subtest (PT) and the Assessment Subtest (AT) in a test is significant there is evidence of multidimensionality [12-16]. Furthermore, using Yen's Q3 statistics to screen items for local dependence, 88% item residual correlations were below absolute value of 0.2. This indicates that the local independence assumption of the IRT was not grossly violated. Based on this statistics, residuals for any pair of items should be uncorrelated, and generally close to zero. Residual correlations that are high indicate a violation of the local independence assumption, and this suggests that the pair of items have something more in common than the rest of the item set have in common with each other [11-15].
The finding agreed with the study of [12] study on assessment of dimensionality of Osun State unified Mathematics achievement test items is multidimensional in nature. Findings is line with the work of Oguoma, Metibemu and [14] on dimensionality assumption test on 2014 Mathematics achievement items of West African Senior Secondary Certificate Examination (WASSCE) concluded that the test items of WASSCE mathematics were inherently multidimensional in nature. Furthermore, [17] also found that fifty (50) items of 2013 WASSCE and sixty (60) items of National Examinations Council (NECO) Geography respectively violated assumption of unidimensionality and that there was more than one dimension that accounted for the variation observed in examinees to the geography test items. The above finding is contradiction with [20] study on the economics quantitative diagnostic test for secondary school students based on IRT is unidimensional. Similarly, [20] also found physics diagnostic test for secondary school students based on IRT to be unidimensional.
The study revealed that most of the test items had standard error below .05 which indicated high reliability. The standard error of measurement allows researchers to determine the probable range within which the individual's true score fall. The result is in agreement with [12] that standard error of .05 and below is described as high reliability, while error .05 is described as low reliability. The result is also in agreement with [13] that if reliability increases, the standard error of measurement becomes smaller. According to [11], standard error of measurement is a statistical estimate of the amount of random error in the assessment of results or scores. This value is similar to value of reliability coefficients calculated by [10], who conducted an intensive study on development and preliminary validation of an instrument for assessment of psychomotor skills in Physics which was found at 0.87. [12], conducted a research on development and standardization of Agricultural Science achievement test for senior secondary school. The reliability value was found at 0.92. [21], found the reliability index for mathematics achievement test at 0.80. [14], reported a study on development and validation of a test in integrated science process skills for further education and training learners the reliability index of 47 items achievement test to be 0.81. These values of reliability indices were considered high reliability thus the present study is equally considered to have developed a reliable instrument. The high reliability index calculated for the present study instrument is not surprising because the instrument was adequately face and content validated before administration.

CONCLUSION

The findings enabled the researchers to conclude that the final form of the multiple choice test is valid and reliable. Hence it was suggested that physics diagnostic test should be used by physics teachers in schools. Particularly in assessment of learning skills students are deficient in so that remedial help can be provided by teachers to them for better academic achievement.

REFERENCES

1. Abedalaziz, N. (2011). Detecting difference using item characteristics curve approaches. The *International Journal of Educational and Psychology Assessment, 8* (2), 1-15.
2. Adedoyin, O. O. (2010). Investigating the invariance of person parameter estimates based on classical test and item response theory. *Botswana Journal Education Science, 23*(4),234249.
3. Adonu, I.I. (2014). Psychometrics analysis of WAEC and NECO practical physics test using partial credit model. (*Unpublished Ph. D Dissertation). University of Nigeria Nsukka.*
4. Anyanwale, M. A., Isaac-Oloniyo, M. & Abayomi, F. R. (2020). Dimensionality assessment of binary response test items. A non-parametric approach of Bayesian item response theory measurement. *International Journal of Evaluation and Research in Education, 9*(2), 385-393.

5. Azizian, M. and Abedi, M. R. (2016). Construction and standardization of reading level diagnostic test for third grade primary school children. *Iranian Journal of Psychiatry and Clinical Psychology, 11*(4), 379-387.
6. Battuaz, M. (2017). *On Wald's test on differential item functioning detection method. Retrieved from http//www. doscst.edu.ph/index.ph*
7. Brown, A. (2012). *Measurement invariance.* Cambridge: Cambridge University press.
8. Chatterji, M. (2013). *Designing and using tools for educational assessment. Retrieved fromhttp://www.columbia.edu/~mb1434/EdAssess.htm.*
9. Ceniza, J.C., & Cereno, D.C. (2012). *Development of mathematic diagnostic test for DORSHS.*
10. Dadughan, S.I. (2015). Development and calibration of primary school mathematics diagnostic test based on item response theory. *(Ph. D Dissertation), University of Nigeria, Nsukka*
11. Esomonu, N. P. M. & Erutujiro, G. (2021). Development and validation of geography diagnostic test using item response theory. *Journal of Humanities and Social Science, 26* (11), 1-9.
12. Esomonu, N.P.M. & Eleje, L.I. (2013). Diagnostic quantitative economics skill test for secondary schools: Development and validation using item response theory. *Journal of Education and Practice, 8(22),*110-125.
13. Kazeni, M. M. (2005). Development and validation of a test of integrated science process skills for the further education and training learners. *(Unpublished M. Ed. Thesis), University of Pretoria, South Africa.*
14. Latun, O.S. (2011). Development and validation of diagnostic in physics for secondary school students in Limpopo Province of South Africa using item response theory. *(Unpublished M.Ed Thesis), University of Pretoria South Africa.*
15. Meredith, D. G., Joyce, P. G. & Walter, R B. (2017). *Educational research: An introduction (8th ed.). United State of America: Pearson Press*
16. Obinne, A.D.E. (2013). *Test item validity: Item response theory perspective for Nigeria. Retrieved from www.emergingresource.org.*
17. Oguoma, R. O., Metibemu, C.C. & Okoye, M.A. (2016). An assessment of the dimensionality of 2014 West African secondary school examination mathematics objective test scores in Imo State, Nigeria. *African Journal Theory Practice. Education Assessment, 4,* 18-33.
18. Okereke, S. C. (2008). Development and preliminary validation of an instrument for the identification of mathematically gifted pupils in Ebonyi State. *(Unpublished Ph.D Thesis), University of Nigeria Nsukka.*
19. Okwilagwe, M. A. & Ogunrinde, E.A. (2017). Assessment of unidimensionality and local independence of WAEC and NECO 2013 Geography Achievement Tests. *African Journal Theory Practice. Education Assessment,5,*31-44.
20. Onah, F. E. (2006). Development and standardization of agricultural science achievement test for senior secondary schools in Enugu State. *(Unpublished Ph. D Thesis), University of Nigeria Nsukka.*
21. Young, B.A. (2014). *Development and validation test in English Language for secondary school in Kisumu Municipality using item response theory. (Unpublished M. Ed Thesis), University of Kassel, Wizenhausen, Germany.*

Nkechi Patricia-Mary Esomonu and Chisom Evangelin Ndubuisi (2023). Development and Validation of Physics Diagnostic Test Using Item Response Theory. *Eurasian Experiment Journal of Scientific and Applied Research* 4(2):31-35