

Development and Validation of Physics Diagnostic Test Using Item Response Theory

Nkechi Patricia-Mary Esomonu¹ and Chisom Evangelin Ndubuisi

Department of Educational Foundation, Faculty of Education, Nnamdi Azikiwe University, Awka, Nigeria.

E-mail: npm.esomonu@unizik.edu.ng; nevanchisom@gmail.com +2348026422569; +2349155809849

ABSTRACT

This research aimed to develop a valid and reliable physics diagnostic test using item response theory. The study was conducted in Anambra State, involving 2259 SS2 students and 1800 SS2 students. The Physics Diagnostic Test (PDT) was used for data collection, consisting of 100 item questions from past WAEC and NECO physics questions. The results showed that the test items are multidimensional, with significant differences between the number of items in the Partitioning Subtest (PT) and the Assessment Subtest (AT). The instrument's reliability was confirmed by Yen's Q_3 , with 87% item residual correlations below the absolute value of 0.2. The findings recommend that physics teachers use the PDT to assess students' learning skills for better academic achievement.

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

INTRODUCTION

Information on student competencies in the learning outcomes planned by the curriculum is required for assessing students' learning progress. As a result, physics teachers' present assessment practice is to use open-ended and multiple-choice test items on subject knowledge, and these assessment procedures do not provide insight into students' cognitive skills in problem solving and understanding of physics topics. Furthermore, the formative and summative assessment instruments that teachers use in their classes on a regular basis provide minimal information regarding students' cognitive learning skills [1]. In turn, existing teacher assessments provide little direct and rapid feedback to teachers and students, particularly on learning skills in which pupils are inadequate. As a result, instructors' classroom assessment procedures must be properly integrated with instruction in order to be effective. As a result, instructors' classroom assessment procedures must be well integrated with instruction in order to give relevant and detailed information regarding students' cognitive processes skill strengths and weaknesses. In most cases, test standards for classroom assessments just identify subject requirements, with no explicit attention given to the type of cognitive learning skill that underpins a program. The incapacity of most secondary school teachers to assess diagnostic skills has been frequently recognized in the study of [2]. Although teachers' own classroom assessments can predict students' overall performance, the data do not tell them much about their students' cognitive learning skills in item performance since students tend to focus on recollection to get through the assignment [3]. Most teachers associate diagnostic information with reporting at the individual success level, with limited information elicited from assessment about students' structural knowledge, procedural skills, and capacities [4]. As a result, the goal of this research is to create a valid and reliable physics diagnostic assessment test.

Measurement and evaluation experts have called for diagnostic assessment of how students comprehend mathematical issues and create problem-solving strategies, including the domain of algebra [5]. Diagnostic testing gives information about pupils' strengths and shortcomings in any topic. To better instructional planning, teachers require additional information about the cognitive strengths and weaknesses of specific knowledge and abilities displayed by individual students on assessment [6]. The information received from diagnostic evaluation differs from that provided by existing standardized assessments.

Diagnostic evaluation is intended to guarantee that the cognitive qualities of interest are explicitly targeted throughout the preparation of items and tests. The information generated from diagnostic assessments that

represents students' cognitive strengths and limitations might help teachers lead students in their knowledge and performance in algebraic expressions learning [7]. Diagnostic testing improves the accuracy and reliability of determining students' cognitive attribute skills and can be used to improve both the teaching and learning processes [8]. Thus, the researchers wanted to investigate in this study the diagnostics abilities in physics that physics teachers require to assess physics students, which include: measuring skill, thinking skill, pictorial skill, communication skill, and calculating skill [9].

According to physics specialists and WAEC Chief Examiner reports, the aforementioned competencies are crucial for good learning of the topic [10]. When pupils are lacking in any of the skills listed above, learning the subject becomes challenging and frustrating. Diagnostic evaluation of the aforementioned skills provides vital input to teachers, allowing them to determine what skills students have or have not mastered, as well as how teaching and learning should be modified to the students' needs. Thus, in today's world, the demand for well-balanced classroom assessments that match physics students' 21st century technological advancement skills cannot be overstated. As a result, the researchers used item response theory to create a valid and reliable physics diagnostic test. Despite this, numerous academics have devised diagnostic tests in topics such as mathematics, language, music, and economics [11, 12, 13, 14, 15, 16, 17, 18, 19]. The development and evaluation of diagnostic tests in physics has received little attention from educational academics. In essence, no study has been conducted on the development and validation of physics diagnostic tests, resulting in an educational deficit. Against this backdrop, the current study was created. The goal of this research is to create and validate an IRT-based physics diagnostic test. The study specifically attempted to determine:

- (i) the dimensionality and Local Independence of Physics Diagnostic Test components; and
- (ii) the empirical reliability of the physics diagnostic test.

Question 1: What is the dimensionality and local independence of the Physics Diagnostic Test items?

The empirical dependability of the physics diagnostic test is the second research question.

METHODS

The item distribution in the instrument was based on the Chief Examiner's report in the WAEC report sheets for 2020, 2021, and 2022 (20% of the students' weakness in Physics is in the area of calculation skill, 20% in thinking skill, 20% in communication skill, and 20% in graphic interpretation skill). A table of specifications was used in the construction of the instrument's items. The instrument was shown to one Measurement and Evaluation expert as well as two subject experts. The instrument was improved based on their ideas. The study's PDT scores were analyzed using Kuder Richardson (K-R-20). PDT has a reliability coefficient of 0.89. The gadget was given to SS2 pupils taking physics. The scores acquired from the pupils were analyzed. The research issue was solved using DIMTEST Statistics in DIMPACK software, which is specifically intended for measuring measurement instrument dimensionality. Dimensionality was tested using DIMTEST statistics. A number of more than .05 indicates multidimensionality (Price, 2017). High dependability is indicated by empirical reliability greater than 0.89 [20]. The design of this study is an instrumentation research design. The study was conducted in Anambra State, Nigeria. The population of the study includes 2259 SS2 students (1518 girls and 741 boys) enrolled in physics during the 2021/2022 academic year. The sample for the study consisted of 1800 SS2 students enrolled in physics during the 2021/2022 academic year, with 652 men and 948 females. A multi-stage technique was used to collect the study's sample. At the state level, simple random sampling processes were used to sample four of the state's six educational zones. The third stage involved sampling five public secondary schools from each Local Government Area using a standard random sampling approach. Then, at each of the sampled schools, all SS2 physics students participated. Data was gathered using the Physics Diagnostic Test (PDT). The equipment was designed to meet the needs for developing diagnostic tests.

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?

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

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

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 study's findings revealed that the underlying latent ability of the examinees' responses to the instrument is multidimensional. Furthermore, there is evidence of multidimensionality if the difference in the number of items in the Partitioning Subtest (PT) and the Assessment Subtest (AT) in a test is significant [21]. Furthermore, when Yen's Q3 statistics were used to screen items for local dependence, 88% of the item residual correlations were less than 0.2. This shows that the IRT's local independence premise was not flagrantly violated. According to these statistics, residuals for every pair of elements should be uncorrelated and close to zero. High residual correlations indicate a violation of the local independence assumption, implying that the pair of items has something more in common. High residual correlations indicate a breach of the local independence assumption, implying that the pair of items has more in common with each other than the rest of the item set [19]. The findings matched with the [18] study on the assessment of dimensionality of Osun State unified Mathematics achievement test components, which revealed that they are multidimensional in character. The findings are consistent with the study of [16], who discovered that the test components of the West African Senior Secondary Certificate Examination (WASSCE) mathematics were essentially multidimensional. Furthermore, [14] discovered that fifty (50) 2013 WASSCE items and sixty (60) National Examinations Council (NECO) Geography items violated the assumption of unidimensionality, indicating that there was more than one dimension that accounted for the variation observed in examinees to the geography test items. The preceding finding contradicts the [13] study on the economics quantitative diagnostic test for secondary school pupils based on IRT, which is unidimensional. Similarly, [17] discovered a unidimensional physics diagnostic test for secondary school pupils based on IRT.

The study found that the majority of the test items had standard errors of less than .05, indicating great reliability. The standard error of measurement enables researchers to establish the likely range of the individual's true score. According to [19], standard errors of .05 or below are considered good dependability, while errors greater than .05

are considered low reliability. The result is also consistent with [18], which states that when reliability increases, the standard error of measurement decreases. [16] defines standard error of measurement as a statistical estimate of the degree of random error in evaluating outcomes or scores. This result is comparable to the 0.87 calculated by Adonu (2016), who did an intense study on the development and preliminary validation of an instrument for assessing psychomotor skills in Physics. [20] investigated the development and standardization of an Agricultural Science accomplishment test for senior secondary school students. The dependability value was discovered to be 0.92. [9] discovered that the reliability index for the mathematics achievement test was 0.80. [14] revealed that the reliability index of a 47-item achievement test was 0.81 in a study on the development and validation of a test in integrated science process skills for further education and training learners. Because these reliability indices were regarded high, the current study is also considered to have developed a reliable instrument. The current study instrument's strong reliability index is not surprising given that it was adequately face and content tested prior to administration.

CONCLUSION

The researchers concluded that the final form of the multiple choice test is valid and reliable, suggesting that physics teachers should use it to assess students' learning skills and provide remedial help for better academic achievement, particularly in physics.

REFERENCES

1. Battuaz, M. (2017). *On Wald's test on differential item functioning detection method*. Retrieved from <http://www.docst.edu.ph/index.php>
2. 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
3. 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),234-249.
4. Abedalaziz, N. (2011). Detecting difference using item characteristics curve approaches. *The International Journal of Educational and Psychology Assessment*, 8 (2), 1-15.
5. Brown, A. (2012). *Measurement invariance*. Cambridge: Cambridge University press.
6. 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.
7. 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.
8. 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.
9. Chatterji, M. (2013). *Designing and using tools for educational assessment*. Retrieved from <http://www.columbia.edu/~mb1434/EdAssess.htm>.
10. Ceniza, J.C., & Cereno, D.C. (2012). *Development of mathematic diagnostic test for DORSHS*.
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.

CITE AS: Nkechi Patricia-Mary Esomonu and Chisom Evangelin Ndubuisi (2023). Development and Validation of Physics Diagnostic Test Using Item Response Theory. NEWPORT INTERNATIONAL JOURNAL OF RESEARCH IN EDUCATION (NIJRE) 3(3):10-14.