PSYCHOMETRIC QUALITIES OF JUNIOR SECONDARY SCHOOL CERTIFICATE EXAMINATION QUESTIONS OF BASIC SCIENCE AND TECHNOLOGY

Ifeanyi I Adonu (Ph. D)[1]; **[2] Ugwu Matthias N[2]; Ezema Blessing Chizoba[3] & Dialoke Christian E[4].

[1]Dept. of EDUPSY, FCE Eha-Amufu

[2]Dept. of Chemistry, FCE Eha-Amufu

[3]Govt. Tech. College, Nsukka.

[4]Dept of Int. Science, FCE Eha-Amufu

****Corresponding Author's Email: mnugwu@gmail.com**

Abstract

The purpose of the study was to analyze some Psychometric qualities of Junior Secondary School Certificate Examination (JSSCE) of Basic Science and Technology (BST) questions used in 2021 Basic Education Certificate Exams (BECE) conducted by NECO. Specifically the study evaluated the reliability, difficulty, discrimination, distractor indices of 2021 BST in BECE. The negligence of psychometric qualities assessment of teacher-made tests and standard tests by practicing teachers and the absolute importance of item qualities evaluation motivated the researchers to carry out this study. The design of the study was instrumentation research design and the study area was Nsukka Education zone of Enugu State. The population of the study was all 901 Junior Secondary Students (JSS 3) in Science Vocational and Technical Colleges in the Education Zone in 2021/2022 academic session. The instrument for the study was BST questions of BECE 2021. The study was guided by four research questions that were analyzed using Correlation, frequency, percentages, under Classical Test Theory (CTT) framework. The major findings of the study showed that BECE 2021 basic science and technology objective items have inadequate difficulty indices (only 27 (45%) out of 60 items within the acceptance range), poor discrimination indices (38 items ie 80% of the items are poor with only 12 items ie 20% that are with fairly good discrimination. But the instrument in question has adequate distractor indices (154 options ie 85.6% are good distractors) and is highly reliable (0.86). The results of the study will provide an opportunity to change the way MCQs are developed and used in educational assessment. Based on some desirable qualities of 2021 BST of BECE as revealed by the findings it is recommended that practicing teachers frequently evaluate teacher-made tests vis-à-vis standard exam questions to enthrone standard internal assessments.

Keywords: *Psychometric, BST, BECE, CTT, MCQ*

Introduction

Education is the most fundamental and vital aspect of social inquiry and the harnessing of human resources. It is commonly known that assessment drives learning. Properly and carefully constructed tests by educators will enhance educational functions. Increasingly multiple choice questions (MCQs) are used for assessing students' performance. Therefore, there is a growing concern about the quality of the tests that are used for assessment. A wide range of content and objectives for a large number of students are assessed through MCQs. Few concede that it is time-consuming and challenging to develop MCQs for measuring higher cognitive skills (Vyas, 2016). But Walsh (2006) opined that MCQs focus on recall rather than comprehension, application and analysis of course-related information. The Basic Education Certificate Examination (BECE) is a prerequisite for students in the ninth year of their basic education class and the third year of junior secondary school. The BECE is an external examination conducted by the Ministry of Education of each state, normally conducted in June/July yearly, the exact date may vary from state to state. The curriculum for these also lasts three years and leads to a trade/craftsmanship certificate. Just like other examinations in Nigeria with a grading system, BECE has its grading system. To pass any subject one must have a grade of at least pass and above. However, the study will focus only on multiple choice test items of Basic science and Technology in BECE. Basic Education Certificate Examinations (BECE) occupy a

prominent position in the educational life of Nigerian child in vocational/technical colleges. It is the first public certification examination taken by students in Nigeria (Osuafor, 2013). Basic Education Certificate Examination (BECE) is taken by students who have completed their first three years of secondary education. Junior secondary education is the education that a child receives immediately after primary education. The objectives of this level of education as spelt out in the National Policy on Education (NPE, 2013), are to :(a). Develop patriotic young people equipped to contribute to social development and the performance of their civic responsibility (b). Provide children with diverse basic knowledge and skills for entrepreneurship and educational advancement(c). Inspire national consciousness and harmonious co-existence irrespective of differences in endowment, religion, color, ethnic, and socio-economic background (d). Inculcate values and raise morally upright individuals capable of independent thinking, and who appreciate the dignity of labor.

In pursuit of these objectives, the Junior Secondary Education curriculum was designed to provide children with training through the following subjects: English language, one Nigerian Language (one of Hausa, Igbo, Yoruba, Edo, Ibibio, and Efik), Mathematics, Religion and National Values, Basic Science and Technology, Pre-Vocational Studies, French, Cultural and Creative Arts, Business Studies, and Arabic. Among these 10 subjects, Basic Science and Technology (BST) is of interest to this study. Omi (2013) argued that BST is the only subject at BECE level that avails students the capacity to: explain events in nature; develop in an all-round manner covering cognitive, affective and psychomotor skills; possess the ability of reasoning and thinking logically; evolve global skills in which scientific know-how can utilize information gathering from nature to solve the problems of mankind, and stir instinct during experiments. These are in addition to the fact that science and technology are tools that catapults nations to economic development, self-sufficiency, and sustainability. Hence, the nation being mindful of the importance of science and technology and their contribution towards technological advancement, intentionally included BST as one of the major and compulsory subjects in the junior school system. In testing students during the BECE examination, which is a condition needed for a student to be certified to have completed junior secondary education, the National Examinations Council administers 60 multiple-choice questions (MCQs) covering four main components of BST namely Basic Science, Physical and Health Education, Basic Technology and Computer Studies.

The Basic Education Certificate Examination (BECE) was established in April 1999 by the then Head of State, General Abdulsalami Abubakar. Its core mandate includes the general control of the conduct entrance examination to unity schools, Conduct of BECE for National technical schools, and conduct of external Senior Secondary School Certificate Examinations in Nigeria without prejudice to the existing powers and functions of the West African Examination Council; conducting a Standard National Assessment of Educational Performance at junior and senior secondary school levels; conducting research leading to national improvement of testing and examination procedures at junior and senior secondary school levels (NECO 2000).

For the tests conducted by BECE to meet international standards, a lot of activities take place behind the scene. One of such activities is the determination of the psychometric properties of the tests used in a process referred to as the item analysis. Item analysis is the process in which the validity, reliability, discrimination, difficulty, and distraction indices of individual items of a test are examined as they affect the whole test. Item analysis was defined as a procedure which provides a test developer with the opportunity to examine both test item and the answers supplied by the testees to find out whether they are of adequate quality and if the content coverage is sufficient for the test they are developed for. Furthermore, this analysis of test items can be computed or derived at both qualitatively and quantitatively in order to evaluate test items to determine their usefulness and functionality. Hassan (2017) stated that item analysis is mainly carried out on already administered test or pilot test, with the hope that the test will be corrected for future use. Item analysis is done on items/questions of a test after it has been administered. Sabri (2013) advanced that the upside to the practice of item analysis is that it allows for the observation of the item characteristics and the improvement of the quality of the test. This means that item analysis helps the test developer to examine each specific item to appraise their quality concerning difficulty, discrimination, and distractor indices.

Item difficulty for dichotomous items under CTT assumption is referred to as P-value, the population of test takers getting the item correct while for the polychromous items it is simply the average score. The difficulty index shows the percentage of students (population of students) who answered an item correctly that has been consistently used in measurement is somewhat of a reverse definition. That is the larger the index, the easier the item, the smaller the index, the more difficult the item (Chandrika,

Kishan, Sajita, Permi, 2016). For example, if an item has a difficulty index (D) of 0.90, this means a very easy test item while a D of 0.020 is a very difficult item. The difficulty index can range from 0 to 1. Effective MCQs possess appropriate difficulty level and the ability to discriminate between performers and non-performers (Terrant, Knierim, and Hayes 2006). An item analysis enables identifying the quality of MCQs based on the difficulty index (D), discriminating index (DI) and distractor efficiency (DE). Clues and errors influence the ease or difficulty of an item. Average item difficulty lies between 30 and 70 %. High-quality MCQs require well-written options, with effective distractors. Items on teacher-generated-tests therefore need to be studied (Owen, 1987). Poorly and ineffectively answered items can be removed from further use thereby making the test more effective.

Item discrimination is a measure of whether an item discriminated between students who know the material well and students who don't. Item discrimination reveals the degree to which a test differentiates between the higher ability students and the lower-ability students. The higher-ability students are those with high scores and answered an item correctly while the lower-ability students are those with lower scores and answered an item correctly. For an item to be a good discriminator, it must surely discriminate between the bright and dull student's scores. That is, the bright students will score an item correctly more than the dull students**.**

The distractor index is a measure of how well the incorrect options in an item prevent the 'not-so-bright' test-takers from picking the right option (also called the key) (Kinancee, Nkem, Orluwene 2017). A distractor is an incorrect or wrong or alternative while the key is the correct option. Chauhan, Chauhan, Chauhan and Vaza (2015) demonstrated that good distracters appeal to a higher proportion of low-achieving test-takers than they do to the bright ones, in that way, they result in negative statistics. The distracter index ranges from -1.00 to +1.00. A negative distracter value indicates that more of the students in the lower ability group chose it than the upper ability group. This indicates that the distracter is functional/effective. A positive distracter value indicates that more of the students in the upper-ability group chose it more than those in the lower-ability group. This indicates that the distracter is nonfunctional/ineffective

Nworgu, (2015) and Harbour-Peters (1999) demonstrated the following approach for calculating the difficulty index/item facility (p), Discrimination index (d) and Distractor Indices (D.I).

Difficulty index p = $\frac{U+L}{2N}$ $\frac{U+L}{2N}$ The range of an ideal item is 0.5. An acceptable p-value is 0.30 to 0.70.

Discrimination Index (d) which answers the question: Does the item distinguish between the bright and the dull students? It is given by $d = \frac{U-L}{N} d = \frac{U-L}{N}$. An ideal item should have a d-value of +1.00 but could range from +0.30 to +1.00.

Distractor Index (DI) $= \frac{L-U}{N} = \frac{L-U}{N}$. DI depends on the number of options. 0.25 is ideal for 5 options, 0.33 for 4 options and 0.5 for 3 options. DI could range from -1 to +1

Where,

U = the number of examinees in the upper one-third of the group who passed the item.

L = the number of examinees in the lower one-third of the group who passed the item.

N = the total number of examinees in either the upper or lower one-third of the group who took the test.

Bichi; Embong; Talib and Salleh (2019) noted that there are two widely used approaches to item analysis. These are Classical Test Theory (CTT) and Item Response Theory (IRT). CTT, a measurement framework that assumes Total Score = True Score + Error Score for a testee, uses two main statistics: the item difficulty index (the proportion of examinees that responded correctly to the item) and the discrimination index (the point-performance on individual items and total test scores).

IRT, a measurement framework that models the ability of the testee to the probability of his responding correctly to an item. IRT however, refers to two key things, both item statistics and examinee's capability with the postulation about the existence of a relationship between his score on a single item and eventual overall performance in the test. In BECE, both CTT and IRT are used to analyze the items of the various tests they conduct. To do this, a pilot study of the items that are intended for use in future examinations is done through a process known as trial testing. This enables the examination body to take a critical look at individual items of the test with the view of making sure that only those of them that meet expected standards are used.

However, there is a gap created by the lack of information from an independent observation standpoint, concerning the quality of the items used for testing candidates in Nigeria as the standard Exam bodies always keep their item quality as a classified information.

Statement of the Problem

It is tantamount to overemphasizing the obvious, that anybody that calls him or herself a teacher need to be conversant with how to evaluate the psychometric qualities of instruments so as to always utilize items with desirable qualities. There exist two theories of measurement for calculating various item qualities. They are Classical Test Theory (CTT) and Item Response Theory (IRT). IRT is a modern theory with its own merits but has this demerit of complex theoretical assumptions which makes it almost unrealistic for use in analysis for educationally less developed areas. But CTT has simple assumptions that can be utilized by all and sundry for needed frequent assessment of item qualities by all teachers. Hence teachers need to adopt CTT to frequently evaluate the qualities of tests from standard exam bodies. This way they will get used to the intricases and qualities of standard items which would be unconsciously extended to all their teacher-made tests for enhanced assessments.

Furthermore, some researchers hold the opinion that the quality of items is the cause of the unstable nature of the performance of students in key subjects like BST in BECE while others hold the opinion that it is due to students' inabilities. Some researchers have asserted that there is a probability of the existence of items that could be said to possess elements of technical inadequacies in the tests developed and administered by examination bodies like BECE (Oluwatayo, 2012). Likewise, there are some opinion that sometimes students' failure in key subjects BST in BECE is because of the faults inherent in the psychometric properties of the test, not just due to their inabilities.

It is therefore not out of proportion for an independent item analysis to be carried out with a view of ascertaining whether or not the items used in public examinations by BECE are of the expected quality. Hence, this is also what informs the need for this study.

Purpose of the Study

The purpose of the study was to conduct a Psychometric analysis of BST test administered in the 2021 BECE. Specifically, the study sought to:

1. estimate the reliability of the Basic Science and Technology test items used for the 2021 Basic Education Certificate Examination
2. determine the difficulty indices of the Basic Science and Technology test items used for the 2021 Basic Education Certificate Examination.
3. determine the discrimination indices of the Basic Science and Technology test items used for the 2021 Basic Education Certificate Examination.
4. determine the distractor indices of the Basic Science and Technology test items used for the 2021 Basic Education Certificate Examination.

Research Questions

The following research questions guided the study.

1. What is the reliability of the test items of Basic Science and Technology BECE 2021.
2. What are the difficulty indices of objective test items of Basic Science and Technology BECE 2021
3. What are the discrimination power of the objective test items of Basic Science and Technology BECE 2021
4. What are the distractor indices of the items in the Basic Science and Technology BST 2021 Objective test items?

Methodology

The study adopted an instrumentation design. The area of study is Nsukka education zone of Enugu state. The zone is made of three Local Government Areas. They are Nsukka, Igbo-Etiti and Uzo-Uwani Local Government Areas, all in Enugu state, The population of the study was all JSS III students of the nine (9) public Science, Technical and Vocational Secondary Schools in Nsukka Education zone. Statistics obtained from the Science, Technical and Vocational Schools Management Board (STVSMB) Nsukka zone, revealed that there were nine hundred and one (901) JSS III students in Nsukka zone in 2021/2022 session and this makes the total population of the study. The sample for the study was 150 JSS III students. Simple random sampling technique was used to select five (5) secondary schools out of nine (9) public schools in the area. The simple random sampling technique

was also used to select 30 students each in the 5 schools bringing the total sampled students to 150. The instrument used for the study was a 60-item multiple choice test consisting of the science and technology components of the Basic Science and Technology test administered by NECO in the 2021 BECE. The instrument was not subjected to validation because NECO being a standard examination body subjected it to validation before administation. The reliability was not established before the study since BECE was subjected to reliability analysis before administration and secondly this is one major thrust of the study. The direct delivery method was employed to ensure high percentage of the 60-items questions and answer scripts return. The questions were given to the examinees and collected immediately by the researcher and assistants after their time allocation has elapsed. Data was analyzed using quantitative procedures. Each examinee score on the items was ascertained and documented. Based on Classical Test Theory (CTT), the research questions were answered using descriptive statistics such as percentages, correlation, proportion, frequency etc.

Results

The results of the study are presented in line with the research questions

Research Question 1. What is the reliability of the test items of Basic Science and Technology BECE 2021?

Table 1: Testees scores separated into odd and even scores-with their ranks and rank differences

Students (total) (n)	X(odd) (total)	y(even) (total)	Total scores (T)	Ranks R_x	Ranks R_y	d	d^2
150							$\sum \mathbf{d^2}$=135376.96

Table 1 shows the scores of the students both odd (x) items, even (y)items, ranks of odd $(R_x)(R_x)$ and even $(R_y)(R_y)$ items of the test obtained from **150** students. The difference between the two ranks of odd and even items (d) for each testee and square of the rank difference d^2 (See Appendix 1 – Sampled sheet of testees scores separated into odd and even scores with d and $d^{2)}$. The reliability coefficient of the half test is **0.76** while the reliability of the whole test is **0.86**.

Research Question 2: What are the difficulty indices of objective test items of Basic Science and Technology BECE 2021

Table 2: Summary of the difficulty indices of BECE Basic Science and Technology objective test items of 2021

Range of difficulty	**Frequency and percentage of items**	**Remark**
0-0.20	13(21.7%)	Very Difficult
0.21-0.30	18(30%)	Difficult
0.31-0.70	27(45.0%)	Moderately difficult
0.71-1.0	2(3.3%)	Very Easy

In Table 2 that is BECE 2021 Basic science and Technology objective test had 18 items that were difficult which represented 30%, 27 items were moderately difficult; this represented 45% of the items,13 items were very difficult; this represented 21.7% of the items while 2 items were very easy which represented 3.3% of the items.

Research Question 3: What is the discrimination power of the objective test items of Basic Science and Technology BECE 2021

Table 3: Summary of discrimination indices of BECE Basic Science and Technology objective test items of 2021

Range of discrimination	Frequency and percentage of items	Remark
0.30- 1.0	1(1.7%)	Excellent
0.20-0.29	11(18.3%)	Marginal (fair)
0- 0.19	38(63.3%)	Poor
-1- 0.1Negative value	10(16.7%)	Bad

Table 3 BECE 2021 had 10(16.7%) item as bad (lower ability students tend to get the items correctly), 38(63.3%) of the items were poor, (few upper ability testees get them correct) 11(18.3%) of the items were marginal while 1(1.7%) of the items was excellent item that needed neither improvement nor revision as they discriminate perfectly well.

Research Question 4: What are the distractor indices of the items in the Basic Science and Technology BST 2021 Objective test items?

Table 4: Summary of distractor indices of BECE Basic Science and Technology objective test items of 2021

Values	Frequency and percentage of distractors	Distractor
Negative values (**-1 to -0.1**)	111(61.7%)	Functional distracters
Positive values(0.3 to 1)	43(23.9%)	Non-functional distractors
Zero values(**0**)	26(14.4%)	Functional distracters

Table 4. BECE 2021 had 60 items (MCQs), a, b,c and d with a total of 240 options. 60 options are correct and the remaining 180 options are distractors. Out of 180 options,111(61.7%) options had negative values which were functional distractors, (Since Distractor indices ranges between -1 and +1 and maximum limit for functional distractor for 4 option test is 0.33). 43(23.9%) of the options(+values) were non-functional and 26(14.4%) of the options (0 values) were also functional distractors.

Discussion

From table 1, the split half method was used to determine the reliability of the test items based on Spearman-Brown prophecy formula. A reliability coefficient of 0 means no reliability and 1.0 means perfect reliability. Generally, if the reliability of a standardized test is above 0.80, it is said to have a very good reliability and if it is below 0.50, it would not be considered a very reliable test (Gronlund, 1976). From table 1, reliability coefficient of the half test was 0.76 while that of the whole test was 0.86, showing that the test items have a very good reliability coefficient.

The results from Table 3, was based on the categorizations of Ebel and Frisbie as documented in Orluwene (2015), that items with coefficients greater than 0.30 are very good, between 0.2 and 0.29 are fairly good; between 0.20 and 0.29 are marginal and needs some revision or eliminated; below 0.19 are poor and need major revision; items with a negative discrimination index are bad and should be eliminated. Hence; in BECE 2021, 10 items were bad and should be eliminated or replaced. It can be seen from table 3, that in terms of discrimination indices, that 38 items were poor and they need to be reviewed and tried again and 11 items were marginal items, could be considered fairly good items, and may be retained. Only 1 item was excellent and need to be retained. It is only 12 items that may have passed the discrimination indices criteria and represent 20% of the total number of items that constituted the test

Harbour-Peters (1999) and Nworgu (2015) posits that distractor indices range from -1 where high ability students choose the option while no low ability student choose the option through 0 to +1 where the converse is the case. From table 4, out of 180 distractors, 111 options which represented 61.7% of the total distractors have negative discrimination value and therefore are functional distractors. On the other hand 43 options which represented 23.9% of the distractors have positive value that range from 0.3 to +1 and are functional distracters. 26 options which represented 14.4% of the distractors (0.3-1) value are ineffective distractors and therefore should be removed. In line with Millman & Greene, 1993, Distracters that are never or infrequently selected should be revised or eliminated.

Conclusions

Based on the findings of the study, it was concluded that BECE 2021 basic science and technology objective items have inadequate difficulty indices (only 27 (45%) out of 60 items within the acceptance range), poor discrimination indices (38 items ie 80% of the items are poor with only 12 items ie 20% that are with fairly good discrimination. But the instrument in question has adequate distractor indices (154 options ie 85.6% are good distractors) and highly reliable (0.86). This reliability coefficient is excellent. The reasons for some unstable nature of the instrument could be due to some other factors such as students' attitude, low morale on the part of teachers and the poor state of school infrastructure. The results of this study will help to change the way MCQs are developed and used in the educational assessment. It will provide the basis for modifying the assessment strategy in the curriculum

Recommendations

Recommendation for this study is based on the results of the findings which are as follows:

1. Every practicing teacher should be conversant with evaluation of psychometric qualities of his/her instrument administered during examinations so as to ensure that their instruments are such that will enthrone objectivity in measurement during teaching for sustainable development.
2. Both teacher made tests and exam body tests should always be comprehensively subjected to item qualities evaluation so as to jettison bad quality items and distractors.
3. Just as teachers are being monitored on writing of lesson plans and other activities, stake holders in education should ensure that psychometric qualities of teacher made items should be monitored and comparable with that of standard exam bodies
4. School teachers at all levels of the secondary school should imbibe the strategy of incorporating national examination bodies' like BECE past questions into internal evaluation processes to familiarize their students with the standards they will be facing in their final examination.
5. School proprietors and the government should endeavor to sponsor teachers to attend item generation workshops to be up-to-date with how to generate test items that meet the criteria for final examinations

References

Bichi A. A; Embong, R; Talib, R; & Salleh, S. I (2019). Comparative Analysis of Classical Test Theory and Item Response Theory using Chemistry Test Data, International. J Eng Adv Technol. 8(5c);1260–6.

Chandrika, R; Kishan-Prasad, H.L; Sajitha, K; Harish-Permi, J.S (2016). Item Analysis of Multiple Choice Questions: Assessing an Assessment tool in Medical Students. Int J Educ Psychol Res. 2(4), 201–4.

Chauhan, P., Chauhan, G. R., Chauhan, B. R., Vaza, J. V. (2015). Relationship between Difficulty Index and Distracter Effectiveness in Single Best-answer Stem type Multiple Choice Questions. Int J Anat Res. 3(4), 1607–10.

Federal Republic of Nigeria (2013). National Policy on Education. NERDC Press Lagos.

Gronlund, N E (1976). Measurement and Evaluation in Teaching, _Third Edition. Reliability and other Desired Characteristics. Macmillan. London.

Harbour- Peters, VFA (1999). Noteworthy Points on Measurement and Evaluation. Item Analysis. Snaap Press Limited. Enugu. Nigeria.

Hassan, S. HR.(2017). Use of Item Analysis to Improve the Quality of Single Best answer Multiple choice Question in Summative Assessment of Undergraduate Medical Students in Malaysia. Educ Med J. 9(3):33–43.

NECO (2001) Establishment Act, Minna, Nigeria

Nworgu, B. G. (2006). Educational Measurment and Evaluation. Theory and Practice, Second Edition.Item Analysis. University Trust Publishers. Nigeria.

Omi, M. (2013). Teaching Methodology in Basic Science and Technology Classes in South-West Nigeria. Asian J Educ e-Learning. 4(1):206–14.

Orluwene, G. W. IBN. (2015). The Influence of Educational Systems on the Academic

Performance of JSCE Students in Rivers State. J Educ Pract. 6(6):85–9.

Osuafor, A. M. OJN (2013). Challenges Encountered by non-science Teachers in Teaching Basic Science and Technology in the Nigerian Universal Basic Education (UBE) Curriculum. African J Teach Educ. 3(3):1–8.

Owen, S V. FR. (1987). What's Wrong with three_Option Multiple Choice Items? Educ Psychol Meas. 47(2):513-22.

Sabri, S.(2013). Item Analysis of Student Comprehensive Test for Research in Teaching Beginner String Ensemble using Model Based Teaching among Music Students in Public Universities. Int J Educ Res. 1(12):1–14.

Tarrant, M., Knierim, A., Hayes, SK WJ.(2006) . The Frequency of Item Writing Flaws in Multiple-Choice Questions used in High Stakes Nursing Assessments. Nurse Educ Today. 26(8):662–71.

Vyas, R SA.(2016). Multiple choice Questions: A Literature Review on the Optimal Number of Options. Natl Med J India. 21(3):130–3.

Walsh, CM SL. (2006). Measuring Critical Thinking. Nurse Educ. 31(4):159-62.

APPENDIX

(Students score separated into each testees even and odd score alongside ranks and rank difference d for calculation of split half reliability using Spearmann correlation)

Students	x(odd)	y(even)	Total scores	Ranks R_x	Ranks R_y	d	d^2
1	22	20	42	1	3	-2	4
2	21	21	42	2	1.5	0.5	0.25
3	20	19	39	3	4.5	-1.5	2.25
4	18	21	39	8	1.5	6.5	42.25
5	18	18	36	8	7.5	0.5	0.25
6	19	16	35	4.5	27	-22.5	506.25
7	17	18	35	16	7.5	8.5	72.25
8	18	17	35	8	13.5	-5.5	30.25
9	17	17	34	16	13.5	2.5	6.25
10	16	18	34	29	7.5	21.5	462.25
11	16	18	34	29	7.5	21.5	462.25
12	19	15	34	4.5	49	-44.5	1980.25
13	17	17	34	16	13.5	2.5	6.25
14	16	17	33	29	13.5	15.5	240.25
15	17	16	33	16	27	-11	121
16	17	16	33	16	27	-11	121
17	17	16	33	16	27	-11	121
18	17	16	33	16	27	-11	121
19	16	17	33	29	13.5	15.5	240.25
20	18	15	33	8	49	-41	1681
21	18	15	33	8	49	-41	1681
22	14	19	33	68.9	4.5	64.4	4147.36
23	16	16	32	29	27	2	4
24	16	16	32	29	27	2	4
25	16	16	32	29	27	2	4
26	15	17	32	46.5	13.5	33	1089
27	15	17	32	46.5	13.5	33	1089
28	15	17	32	46.5	13.5	33	1089
29	17	15	32	16	49	-33	1089
….							…
150							135376.96

The reliability of the test was calculated using Spearman-Brown prophecy formula but we need to first solve for the reliability coefficient of the half test which is given below

$$r_{½,½}\ r_{½,½} = 1 - \frac{6\sum d^2}{n^3 - n} = 1 - \frac{6 \times 135376.96}{150^3 - 150} = 0.761 - \frac{6\sum d^2}{n^3 - n} = 1 - \frac{6 \times 135376.96}{150^3 - 150} = 0.76$$

Where $r_{½,½}\ r_{½,½}$ = the reliability coefficient of the half test (Spearman Rank Correlation Coefficient), $\sum d^2 \sum d^2$ = the sum of the square of the difference between ranks of odd (x) and even (y) number scores of the students and 'n' is the total number of students that took the test.

The reliability coefficient of the whole test ($r_{11}\ r_{11}$) can then be calculated using the formula below (Spearman-Brown prophecy).

$$r_{11} = \frac{2(r_{½,½})}{1 + r_{½,½}} = \frac{2(0.76)}{1 + 0.76} = 0.86$$