International Journal of Developing and Emerging Economies
Vol.12, No.2, pp.86-109, 2024
Print ISSN: 2055-608X (Print),
Online ISSN: 2055-6098(Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

Revalidation of Renolds Adaptable Intelligence Test (Rait-Nv) For Use in Nigeria

[1]Uwah, Idongesit Victor (Ph.D) and [2]Bokolo, Faith ((Ph.D)
E-mail: [1]idongesituwah6@gmail.com, [2]faithbokolo2003@gmail.com
[1]Department of Educational Psychology, Guidance and Counselling. University of Port-Harcourt. Rivers State-Nigeria.
[2]Department of Educational Foundations, Isaac Jasper Boro College of Education. Sagbama, Yenegoa, Bayelsa State, Nigeria

doi: https://doi.org/10.37745/ijdee.13/vol12n286109 Published December 15, 2024

Citation: Uwah, I.V. and Bokolo F. (2024) Revalidation of Renolds Adaptable Intelligence Test (Rait-Nv) For Use in Nigeria, *International Journal of Developing and Emerging Economies*, Vol.12, No.2, pp.86-109

Abstract: *The study revalidated the "Renolds Adaptable Intelligence Test (RAIT-NV)" for use in Nigeria. Instrumentation design was used. The researcher purchased and administered four hundred (469) copies of Raynolds Adaptable Intelligence Test-Non-Verbal (RAIT-NV) to a sample drawn across educational institutions, hospitals, corporate organizations and religious bodies. Before then, the researcher had quickly carried out a pilot test using 100 respondents to determine the validity and reliability of RAIT-NV in its original form. Discriminant validation process was used in determining the validity while test retest was used in determining its reliability. For the earlier, the Leibowitz social anxiety scale was administered alongside RAIT-NV. After the scores the researcher correlated the scores and an index of 0.13 was realized testifying the discrepancy in these scales. The test retest also showed a correlation coefficient of 0.86 indicating a high reliability. For determination of reliability, the researcher used Cronbach Alpha, split half method and KR_{20}. In determining the mean differences in the groups, the researcher used ANOVA and t-test in the process. The result showed that thirty (30) and thirty-three (33) items were either too difficult or easy for respondents here in Nigeria. Hence, thirty-two (32) items were considered normal with overall difficulty index for RAIT-NV is 0.55 (55%) in Nigeria. Nineteen (19) items discriminated better. Construct validity showed that there is a significant difference in age while there were non in terms of educational level, ethnic groups and gender. Internal consistency reliability reveals that overall Cronbach of 0.66, KR_{20} of 0.62 for NVA and 0.58 for SEQ and Split-Half reliability index of 0.07 for NVA and 0.18 for SEQ.*

Keywords: revalidation, Renolds adaptable intelligence test, reliability, validity

International Journal of Developing and Emerging Economies
Vol.12, No.2, pp.86-109, 2024
Print ISSN: 2055-608X (Print),
Online ISSN: 2055-6098(Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

INTRODUCTION

It is no doubt that the general mental ability or intelligence level of a student to a great extent determines the extent to which he or she can handle a tasks or get involved in a particular profession. Hence, intelligence is being considered as a variable that may predict career choice. According to the definition given at the Mainstream Science on Intelligence in Gottfredson (1997), intelligence is defined as; A very general mental capability that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience. It is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings" catching on," "making sense" of things, or "figuring out" what to do. Sternberg(2017) stated that **intelligence is the** mental quality that consists of the abilities to learn from experience, adapt to new situations, understand and handle abstract concepts, and use knowledge to manipulate one's environment. From this premise, it could be seen that intelligence involves the ability of an individual to learn from his experience like schooling, life event and other ways which individuals can think of. It may also mean that intelligence involves adaptation to new environment effectively as possible. Intelligence can be understood as a start-up resource of information processing which has to be invested in knowledge in order to enable competencies in a domain

A test or examination whether formal or informal is an assessment process intended to measure student's knowledge, skill, aptitude, physical fitness, or classification in many other topics. A test may be administered verbally, on paper, on a computer, or in a pre-determined area that requires a test taker to demonstrate or perform a set of skills. Tests vary in style, rigor and requirements. A test may be administered formally or informally. An example of an informal test is a reading test administered by a parent to a child. A formal test might be a final examination administered by a teacher in a classroom or an I.Q. test administered by a psychologist in a clinic. Formal testing often results in a grade or a test score. An exam is meant to test a persons' knowledge or willingness to give time to manipulate that subject. In every testing, it either comes in a standardized or non-standardized format.

A test may be developed and administered by an instructor, a clinician, a governing body, or a test provider. In some instances, the developer of the test may not be directly responsible for its administration. For example, Educational Testing Service (ETS), a nonprofit educational testing and assessment organization, develops standardized tests such as the SAT but may not directly be involved in the administration or proctoring of these tests. As with the development and administration of educational tests, the format and level of difficulty of the tests themselves are highly variable and there is no general consensus or invariable standard for test formats and difficulty. Often, the format and difficulty of the test is dependent upon the educational philosophy

of the instructor, subject matter, class size, policy of the educational institution, and requirements of accreditation or governing bodies. In general, tests developed and administered by individual instructors are non-standardized whereas tests developed by testing organizations are standardized. As seen earlier, Intelligence has been defined in many ways. It could be seen as the capacity for logic, understanding, self-awareness, learning, emotional knowledge, reasoning, planning, creativity, critical thinking, as well as problem solving. Like any other construct, it is pertinent to ascertain the level of intelligence of an individual. The process of ascertaining the level of an individual intelligence in numerical values is referred to as intelligence test. Raynolds (1998) argued that intelligence tests are among the oldest devices in the psychologist use and are arguably the most frequently used of tests in the evaluation and general mental assessment. Most times, it is seen that intelligence and aptitude tests are interchangeably used in adult assessment. They are used with adults in predicting a variety of cognitive disorders and in the vocational arena. Since the translation and modification of Alfred Binet's intelligence test for French schoolchildren was introduced in the United States by Lewis Terman (of Stanford University, hence the Stanford-Binet Intelligence Scale), a substantial proliferation of such tests has occurred.

The administration of intelligence test is for a number of reasons. Beres and Perlman(2000) observed that intelligence tests are administered for a variety of reasons including identification(of mental retardation, learning disabilities, other cognitive disorders, giftedness),placement(gifted and other specialized programs), and as a cognitive adjunct to a clinical evaluation. It is noted that intelligence testing, which is useful in clinical and vocational settings, is also a controversial activity, especially with regard to the diagnosis of mild mental retardation among minority cultures in the United States. Used with care and compassion, as a tool toward understanding, such tests can prove invaluable. Used recklessly and with rigidity, they can cause irreparable harm. Extensive technical training is required to master properly the administration of an individual intelligence test (or any individual test for that matter). Even greater sensitivity and training are required to interpret these powerful and controversial devices. Extensive knowledge of statistics, measurement theory, and the existing research literature concerning testing is a prerequisite to using intelligence tests.

According to Stern (2015) intelligence test scores can account for achievement differences in many content areas to a considerable extent. To him, individual's intelligence coefficients as measured arisesfrom a complex interaction between genes and environmental stimulation, foremost schooling. There are a lot of factors that accounts for the intelligence result individuals scores. It could be that teacher's professional competencies have a major impact on how learners exploit their intelligence for learning particular subjects. Stern (2015) also revealed that the complex interaction between genes and environment is also founded on the fact that heritability of intelligence increases during the life span. To understand this very well-established finding, one has to realize that societies which provide access to a broad variety of cognitive activities in

professional as well as in private life enable adults more than children to actively select special environments which fit their genes. People who have found their niche can perfect their competencies by deliberate learning. Although, however, it is beyond any shadow of doubt that in developed societies, genes can explain a huge amount of IQ differences, the continued search for specific genes responsible for the expression of cognitive capabilities has not at all met with much success irrespective of the effort invested in this course. According to Deary (2012), given however, that even for height less than 20% of the variance can be traced back to already identified gene, it is far from surprising that it is almost impossible to track down the genes that are involved in intelligence. Since the development of intelligence testing, various approaches or test has also been developed to test the intellectual abilities of the individual and on of such is the Raynolds Adaptable Intelligence Test (RAIT).

The Raynolds Adaptable Intelligence Test (RAIT-Non-Verbal)

This instrument is the bane of the present study. Derived from the Raynolds Adaptable Intelligence Test (RAIT), this is a rapid, reliable, and valid test of non-verbal intelligence. It was created using the two nonverbal subtests from the RAIT to give an accurate assessment of fluid intelligence. Although the RAIT-NV has a maximum time limit, it remains a power test and not a speeded test. Created for use with individuals who do not speak English, those with hearing impairments, individuals unwilling to communicate verbally, or populations with minimal language capabilities. No reading skill, motor coordination, or visual–motor capabilities are required, reducing the confounds that occur when manipulated objects are used to assess nonverbal intelligence. The RAIT-NV Can be administered individually or in a group format. It may be used in human resource and related industrial settings, schools, juvenile and adult justice systems, and clinical practices. The test is designed to provide continuity across a wide age span. Examined rigorously to be free of gender and ethnic bias, reducing gender and ethnicity as confounds, particularly important for use with English as a second language (ESL) students and adults. Two subtests evaluate fluid intelligence. The Nonverbal Analogies subtest and Sequences subtest use appealing, bright illustrations that are engaging to examinees. The two subtests sum to create the RAIT Nonverbal Intelligence Index (NVII), which is scaled to the familiar IQ metric. The Score Summary Form allows you to track examinees' scores over multiple administrations, to calculate reliable change indexes, and to calculate discrepancy scores. Each subtest has a sample item to facilitate comprehension that may be read by the examinee, read aloud to the examinee, or conveyed using hand gestures, and alternate instructions and additional sample items are provided for special populations. RAIT-NV is standardized on a sample of 2,124 individuals matched to 2010 U.S. Census statistics. Multiple types of scores are provided, including *z* scores, normal curve equivalents, stanines, percentiles, and age equivalents. The validity of RAIT-NV was investigated using individuals from several clinical groups, including intellectual disability, TBI, stroke, dementia, learning disability, hearing impairment, and ADHD. An investigation of RAIT-NV scores' relationship to examinees' occupational industries and job complexity levels revealed

expected patterns. The test is appropriate for use with individuals ages 10 to 75 years who can understand the written or demonstrated directions for subtests and are able to formulate the necessary responses. It is worthy to note that individuals with significant vision problems may perform poorly on the RAIT-NV. This implies that the test can only be administered to individuals with significant fine-motor impairments. The RAIT-NV has two subtests which are timed separately. The two subtests take17 minutes total. The proctor must use a stopwatch to monitor the time limits for each subtest, except when using certain alternate administration instructions. The first page of the RAIT-NV score summary form contains spaces for recording the examinee's demographic and background information. To calculate the raw score for each RAIT-NV subtest, individuals are expected to place the RAIT-NV scoring key on top of the RAIT-NV answer sheet. The transparent scoring key overlays a black box around the correct responses. For a reliable change scores, examiners are to enter the date of testing, age of testing, and T score for Time 1 and Time 2. Then calculate the difference between the two scores. However, for the purpose of this work, the researcher will focus on the revalidation of the RAIT-NV.

As stated earlier, any good test should be reliable and valid. A combination of these is the entire validation process. This will of necessity include test validity as well as test reliability as well as general item analysis. Furthermore, a good item analysis will include determination of the item difficulty, discrimination as well as the item distracter. There are a variety of techniques for performing an item analysis, which is often used, for example, to determine which items will be kept for the final version of a test. Item analysis is used to help "build" reliability and validity are "into" the test from the start. Item analysis can be both qualitative and quantitative. The former focuses on issues related to the content of the test, eg. content validity. The latter primarily includes measurement of item difficulty and item discrimination. Furthermore, Rivera (2007) stated that writing the test item does not produce an item ready to be tested until it is validated. There have been multiple techniques established for reviewing item-objective congruence based on large-scale assessments to small classroom assessments, such as the use of empirical techniques similar to norm-referenced testing, expert judgment used to calculate the index of item-objective congruence, a rating of item-objective match on a 5-point scale conducted by experts, and the use of a matching task. It should also be noted that writing the test item does not produce an item ready to be tested until it is validated. Hambleton and Rogers (1991) provided three features to focus on when reviewing a CRT item's content: (1) item validities, (2) technical quality, and (3) representativeness (p. 18). While these were for CRT tests, the same rules can be applied to multiple-choice question. These three guidelines were based on expert judgment to, "assess the degree to which the sample of items in the test is representative of some defined domain (Hambleton & Rogers, 1991). Haladyna (1994) offered three main characteristics that pertain to item validation: (1) a review of the test item from item development procedures, (2) an analysis of the statistical study of item responses, and (3) a summary of using item response patterns to study specific problems in testing.

Item difficulty: an item's difficulty level is usually measured in terms of the percentage of examinees who answer the item correctly. This percentage is referred to as the item difficulty index, or "p". it is also noted that item difficulty index is one of the most useful, and most frequently reported, item analysis statistics. This measure of the *proportion* of examinees who answered the item correctly; for this reason it is frequently called the *p-value*. As the proportion of examinees who got the item right, the p-value might more properly be called the item easiness index, rather than the item difficulty. It can range between 0.0 and 1.0, with a higher value indicating that a greater proportion of examinees responded to the item correctly, and it was thus an easier item.

Item discrimination: in carrying out item analysis, item discrimination index is a measure of how well an item is able to distinguish between examinees who are knowledgeable and those who are not, or between masters and non-masters. There are actually several ways to compute an item discrimination, but one of the most common is the *point-biserial correlation*. When interpreting the value of a discrimination it is important to be aware that there is a relationship between an item's difficulty index and its discrimination index. If an item has a very high (or very low) p-value, the potential value of the discrimination index will be much less than if the item has a mid-range p-value. In other words, if an item is either very easy or very hard, it is not likely to be very discriminating. A typical CRT, with many high item p-values, may have most item discriminations in the range of 0.0 to 0.3. A useful approach when reviewing a set of item discrimination indexes is to also view each item's p-value at the same time. According to ScorePak (2020), item discrimination is "good" if the index is above .30; "fair" if it is between .10 and .30; and "poor" if it is below .10.

Distracter indices

Also, in carrying out item analysis, distractor analysis addresses the performance of these incorrect response options. Just as the key, or correct response option, must be definitively correct, the distractors must be clearly incorrect (or clearly not the "best" option). In addition to being clearly incorrect, the distractors must also be plausible. That is, the distractors should seem likely or reasonable to an examinee who is not sufficiently knowledgeable in the content area. If a distractor appears so unlikely that almost no examinee will select it, it is not contributing to the performance of the item. In fact, the presence of one or more implausible distractors in a multiple-choice item can make the item artificially far easier than it ought to be.

In a simple approach to distractor analysis, the proportion of examinees who selected each of the response options is examined. For the key, this proportion is equivalent to the item p-value, or difficulty. If the proportions are summed across all of an item's response options they will add up to 1.0, or 100% of the examinees' selections.

According to kpolovie (2016: 337), reliability is the extent which a measuring instrument accurately measures a particular social or psychological trait, construct, characteristics or behavior of the examinees in the best perfect manner without any unsystematic (random) or systematic fluctuation (error) both in the instrument and in the characteristics under investigation. Carmines and Zeller, (1979) also stated that reliability is the degree which an instrument give consistent results. It involves the extent to which a measurement of a phenomenon provides stable and consist result. It is no doubt that determining the reliability of an instrument is equally important in the process of test validation. This is because according to Huck (2007), it assures consistency of a measuring instrument. In education, it is observed that the most frequent method of determining reliability is Cronbach Alpha method. Whitley (2002) opined that there is no standing rule about the internal consistencies but many agree on a coefficient value of .70. Hinton et al. (2004) have suggested four cut-off points for reliability, which includes excellent reliability (0.90 and above), high reliability (0.70-0.90), moderate reliability (0.50-0.70) and low reliability (0.50 and below). (Hinton et al., 2004). Reliability checks involve many approaches. These include; Internal consistency, test-retest, inter-rater reliability, the split-half reliability etc. Huck (2007) stated that developing a valid and reliable instrument usually requires multiple iterations of piloting and testing which can be resource intensive. As earlier stated, it is imperative that developing and being assured of the best result out of any intelligence test demands the determination of both its validity as well as the reliability which is the aim of the present study.

Developing a test suitable for students in one locality does in any way makes it suitable for even identical population but in a different place. In the Nigerian environment, such differences as observed have made it difficult for students to respond adequately to test administered. With all these differences and with the negligence of revalidation by most test users especially in intelligence test, the problem of the study is to revalidate the Raynolds Adaptable Intelligence Test (RAIT-NV) for use in the Nigerian system. This overlooked process has formed the gap that necessitated the present study.

The aim of the study is to Determination of;

1. Difficulty index of RAIT-NV
2. discriminative index of RAIT-NV
3. The construct validity of RAIT-NV using hypothesis testing evidence.
4. Determine the Cronbach Alpha reliability of RAIT-NV
5. Kuder Richardson Formula 20 (KR_{20}).
6. Split- Half method

The following research questions were asked;

1. What is the difficulty index of RAIT-NV?

2. What is the discriminative index of RAIT-NV?
3. What is the construct validity of RAIT-NV from hypothesis-testing evidence (There is no significant difference in the RAIT-NV scores of those from various ethnic group, ages, gender as well as educational level.
4. What is the Cronbach Alpha reliability index of RAIT-NV?
5. What is the reliability index of RAIT-NV using KR_{20}?
6. What is the split reliability index of RAIT-NV?

METHODS

The study used instrumentation design in the study. Instrumentation is the process of constructing research instruments that could be used appropriately in gathering data on the study. The present study used the design because it deals with validation an already standardized instrument for use in Nigeria. The researcher purchased and administered four hundred (469) copies of Raynolds Adaptable Intelligence Test-Non-Verbal (RAIT-NV) to a sample drawn across educational institutions, hospitals, corporate organizations and religious bodies. Before then, the researcher had quickly carried out a pilot test using 100 respondents to determine the validity and reliability of RAIT-NV in its original form. Discriminant validation process was used in determining the validity while test retest was used in determining its reliability. For the earlier, the Leibowitz social anxiety scale was administered alongside RAIT-NV. After the scores the researcher correlated the scores and an index of 0.13 was realized testifying the discrepancy in this scales. The test retest also showed a correlation coefficient of 0.86 indicating a high reliability. The 469 actual samples of the study were drawn using accidental sampling technique. The researcher carefully with the help of three research assistants distributed the instruments to the respondents. Administration of the instrument was done over a period of two (2) months. After the responses, the instruments were collected on the spot and a total of 469 copies were successfully completed while others were either lost or disqualified due to incomplete response. Hence, the total sample size used for the analysis was 469. The validation of the instrument for its usability in Nigeria was done by ascertaining the difficulty as well as discrimination index analysis. For determination of reliability, the researcher used Cronbach Alpha, split half method and KR_{20}. In determining the mean differences in the groups, the researcher used ANOVA and t-test in the process.

RESULT AND DISCUSSION

1. What is the difficulty index of RAIT-NV?

Table 1 show calculation and indices of item difficulty from question 1-95 of RAIT-NV

	ITEMS	REMARKS
1	0.96*	Too easy item
2	0.94*	Too easy item
3	0.91*	Too easy item
4	0.97*	Too easy item
5	0.91*	Too easy item
6	0.79*	Too easy item
7	0.92*	Too easy item
8	0.96*	Too easy item
9	0.79*	Too easy item
10	0.91*	Too easy item
11	0.97*	Too easy item
12	1.00*	Too easy item
13	0.89*	Too easy item
14	0.96*	Too easy item
15	0.84*	Too easy item
16	0.90*	Too easy item
17	0.98*	Too easy item
18	0.94*	Too easy item
19	0.93*	Too easy item
20	0.89*	Too easy item
21	0.68√	Average item
22	0.20**	Average item
23	0.82*	Too easy item
24	0.75*	Too easy item
25	0.06**	Too difficult item
26	0.51√	Average item
27	0.57√	Average item
28	0.54√	Average item
29	0.73*	Too easy item
30	0.79*	Too easy item
31	0.69√	Average item
32	0.33√	Average item
33	0.25**	Too difficult item
34	0.57√	Average item
35	0.60√	Average item
36	0.61√	Average item
37	0.54√	Average item

38	0.10**	Too difficult item
39	0.01**	Too difficult item
40	0.70√	Average item
41	0.14**	Too difficult item
42	0.23**	Too difficult item
43	0.19**	Too difficult item
44	0.07**	Too difficult item
45	0.21**	Too difficult item
46	0.48√	Average item
47	0.15**	Too difficult item
48	0.16**	Too difficult item
49	0.28**	Too difficult item
50	0.00**	Too difficult item
51	0.00**	Too difficult item
52	0.03**	Too difficult item
53	0.80*	Too easy item
54	0.82*	Too easy item
55	0.82*	Too easy item
56	0.76√	Average item
57	0.87*	Too easy item
58	0.91*	Too easy item
59	0.89*	Too easy item
60	1.00**	Too easy item
61	0.77√	Average item
62	0.79√	Average item
63	0.85*	Too easy item
64	0.85*	Too easy item
65	0.72√	Average item
66	0.62√	Average item

67	0.48√	Average item
68	0.53√	Average item
69	0.58√	Average item
70	0.82*	Too easy item
71	0.35√	Average item
72	0.72√	Average item
73	0.64√	Average item
74	0.45√	Average item
75	0.34√	Average item
76	0.24**	Too difficult item
77	0.36√	Average item
78	0.45√	Average item
79	0.36√	Average item
80	0.35√	Average item
81	0.28**	Too difficult item
82	0.24**	Too difficult item
83	0.20**	Too difficult item
84	0.40√	Average item
85	0.61√	Average item
86	0.16**	Too difficult item
87	0.13**	Too difficult item
88	0.32√	Average item
89	0.18**	Too difficult item
90	0.28**	Too difficult item
91	0.17**	Too difficult item

92	0.00**	Too difficult item
93	0.20**	Too difficult item
94	0.14**	Too difficult item
95	0.22**	Too difficult item
∑Indices	**51.96/95**	
Average DIFF	**0.55**	**Average Difficulty**
Percentage	**55%**	

Note: items marked (*) are too difficult items
Items marked (**) are too easy items
Items marked (√) are average, good or moderate difficulty items.
Individual Item difficulty index is gotten by sum of right response divided by total number of students.

Basis for acceptance or rejection of items are only those between 0.30-0.80 i.e 30-80% (Wiersman & Jurs cited in Opara 2016).
(See Appendix 1 for procedure of excel calculation table)

From the table above, it could be seen that the difficulty index for the RAIT-NV is 0.55. from the individual cells, it is seen that 30 items including items 22, 25, 33, 38, 39, 41, 42, 43, 44, 45, 47, 48, 49, 50, 51, 52, 60, 76, 81, 82, 83, 86, 87, 89, 90, 91, 92, 93, 94 and 95 marked '**' were considered too difficult for the students here in Nigeria. Similarly, items 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 23, 24, 29, 30, 53, 54, 55, 57, 58, 59, 63, 64 and 70 marked '*' were considered too easy for people here in Nigeria. Hence, only items 21, 26, 27, 28, 31, 32, 34, 35, 36, 37, 40, 46, 56, 61, 62, 65, 66, 67, 69, 69, 71, 72, 73, 74, 75, 77, 78, 79, 80, 84, 85 and 88 were considered normal for individuals here in Nigeria.

Research Question Two: What is the discrimination index of RAIT-NV

Table 4.2 Discrimination index of RAIT-NV in Nigeria

S/N	DISCRIMINATION INDICES	REMARKS
1	-0.03**	Poor
2	0.02**	Poor
3	0.05**	Poor
4	0.08**	Poor
5	0.09**	Poor
6	0.34√	Good
7	0.10*	Fair
8	0.11*	Fair
9	0.07**	Poor
10	0.04**	Poor
11	0.10*	Fair
12	-0.01**	Poor
13	0.20*	Fair
14	0.14*	Fair
15	0.10*	Fair
16	0.13*	Fair
17	0.07**	Poor
18	0.02**	Poor
19	0.12*	Fair
20	0.25√	Good
21	0.22*	Fair
22	0.02**	Poor
23	-0.12**	Poor
24	-0.19**	Poor
25	-0.07**	Poor
26	0.29√	Good
27	0.27√	Good
28	0.11*	Fair
29	-0.08**	Poor
30	0.09**	Poor
31	0.24*	Fair
32	0.15*	Fair
33	0.12*	Fair
34	0.20*	Fair
35	0.20*	Fair
36	0.57√	Good

37	0.31√	Good
38	-0.05**	Poor
39	0.02**	Poor
40	0.57√	Good
41	-0.11**	Poor
42	0.14*	Fair
43	0.12*	Fair
44	0.05**	Poor
45	-0.09**	Poor
46	-0.09**	Poor
47	-0.04**	Poor
48	0.31√	Good
49	-0.09**	Poor
50	0.00**	Poor
51	0.01**	Poor
52	-0.05**	Poor
53	0.14*	Fair
54	0.06**	Poor
55	0.06**	Poor
56	0.11*	Fair
57	0.18*	Fair
58	-0.04**	Poor
59	0.08**	Poor
60	-0.01**	Poor
61	-0.01**	Poor
62	0.35√	Good
63	0.20*	Fair
64	0.24*	Fair
65	0.23*	Fair
66	0.22*	Fair
67	0.33√	Good
68	-0.02**	Poor
69	0.09**	Poor
70	0.06**	Poor
71	0.38√	Good
72	0.38√	Good
73	-0.12**	Poor
74	0.09**	Poor
75	0.42√	Good

76	0.23*	Fair
77	0.31√	Good
78	0.17*	Fair
79	0.09**	Poor
80	0.34√	Good
81	0.30√	Good
82	-0.06	Poor
83	0.11*	Fair
84	0.01**	Poor
85	0.36√	Good
86	-0.07**	Poor
87	0.23√	Good
88	-0.16**	Poor
89	0.29√	Good
90	-0.03**	Good
91	0.00**	Poor
92	0.15*	Fair
93	-0.06**	Poor
94	-0.16**	Poor
95	-0.05**	Poor

Note: items marked (*) are too difficult items
Items marked (**) are too easy items
Items marked (√) are average, good or moderate difficulty items.

Basis: According to ScorePak (2020), item discrimination is "good" if the index is above .30; "fair" if it is between .10 and .30; and "poor" if it is below .10. Furthermore, negative items depicts that the item may totally measure a different thing entirely as perceived by the student.
(See Appendix 2 for procedure of excel calculation table)

The table indicates items that are poorly discriminating, fairly discrimination as well as those that are poorly discriminating. It is seen that 19 items including items 6, 20, 26, 27, 36, 37, 40, 48, 62, 67, 71, 72, 75, 77, 80, 81, 85, 87 and 89 marked '√' discriminated better. Similarly, 28 items including items 7, 8, 11, 13, 14, 15, 16, 19, 21, 28, 31, 33, 34, 35, 42, 43, 53, 56, 57, 62, 64, 65, 66, 76, 78, 83 and 92 marked '*' were considered too be fairly discriminating items. Finally, 48 items including items 1, 2, 3, 4, 5, 9, 10, 12, 17, 18, 22, 23, 24, 25, 29, 30, 38, 39, 41, 44, 45, 46, 47, 49, 50, 51, 52, 54, 55, 58, 59, 60, 61, 68, 69, 70, 73, 74, 79, 84, 86, 88, 90, 91, 93, 94, and 95 were seen as poor discriminators to be considered normal for individuals here in Nigeria.

Research Question Three: What is the Construct validity of RAIT-NV using hypothesis testing evidence?

a. There is no significant difference in the in the RAIT-NV scores of individuals from the various age groups.

Table 3 one-way ANOVA summary for age group differences between those between 10-15, 16-20, 21-40 and 41-75 years

Years	**N**	**Mean**	**Std.D**
10-15 years	183	46.66	13.11
16-20 years	216	52.68	9.59
21-40 years	47	70.89	11.44
41-75 years	23	51.22	9.56

From the table, individuals within 10-15 years were 183.

Source	Sum of Sq.	Df.	Mean Sq.	F	Sig	Result
BG	22103.06	3	7367.69	57.98	0.000	Significant
WG	59086.69	465	127.06			
Total	81189.76	468				

Between 16-20 years were 216. Between 21-40 years were 47 while 41-75 years of age were 23. The means scores showed a mean of 46.66, 52.68, 70.89 and 51.08 respectively. These show that individuals between 21-40 years had higher performance followed by those between 16-20 years. Followed by those between 41-75 years and finally by those between 10-15 years.

It could also be seen that between group (BG) had 3 degrees of freedom, 22103.06 sum of squares and 7367.69 mean square. Within group has 59086.69 sum of squares, 465 degrees of freedom and 127.06 mean square. The sum of square total is 81189.76 with 468 degrees of freedom. The F ratio obtained is 57.98. Sig value is 0.000 as this is statistically significant at 0.05 alpha level. Since the sig ($P=0.000<0.05$) is higher than the alpha of .05, the null hypothesis of no significant difference in the RAIT-NV scores of individuals from the various age groups is rejected meaning that there is a significant difference in the RAIT-NV scores of individuals from the various age groups.

b. There is no significant difference in the in the RAIT-NV scores of individuals from various educational level.

Table 3 one-way ANOVA summary for educational level differences between jss1-ss3, year 1- final year, masters level as well as Ph.D level.

Educational Level	**N**	**Mean**	**Std. D**
Jss1-ss3	340	50.04	10.64
Year 1- Final	69	48.42	9.67
Masters Level	48	58.85	10.29
Ph.D Level	12	66.17	17.93

Source	Sum of Sq.	Df.	Mean Sq.	F	Sig	Result
BG	6488.22	3	2162.94	18.86	0.000	Significant
WG	53332.96	465	114.69			
Total	59821.78	468				

Table 4.4 show that jss1-ss3 students, year 1- final year students, masters level students as well as Ph.D level students were 340, 69, 48 and 12 respectively. Their mean values were 50.04, 48.42, 58.85 and 66.17 respectively. The mean values has indicated highest performance in the Ph.D group followed by those in the masters level, followed by those in jss1-ss3 and lastly those in year 1- final year. Sum of square between, within and total were 6488.22, 6488.22 and 6488.22 respectively while the mean square between and within are 2162.94 and 114.69. calculated F-ratio is 18.86 while sig value 0.000. therefore, since sig value ($p=0.00001 <0.05$) is less than 0.05 alpha at 3 and 465 degrees of freedom, the null hypothesis is rejected. This indicates that there is actually a significant difference in the RAIT-NV scores of individuals from various educational level.

c. There is no significant difference in the RAIT-NV scores of individuals from various ethnic group.

Table 5 one-way ANOVA summary for ethnic differences between Igbo, Yoruba, Hausa and others.

Ethnic group	**N**	**Mean**	**Std. D**
Igbo	197	53.79	8.08
Yoruba	76	51.19	10.97
Hausa	19	51.57	10.83
Others	177	53.12	10.07

Source	Sum of Sq.	Df.	Mean Sq.	F	Sig	Result
BG	427.64	3	142.55	1.59	0.19	Insignificant
WG	41675.94	465	89.62			
Total	42103.58	468				

Table 4.5 show that Igbo, Yoruba, Hausa and other minority ethnic groups were 197, 76, 19 and 177 respectively. Their mean values were 53.79, 51.19, 51.57 and 53.12 respectively. The mean values has indicated the Igbos performed highest in the test followed by other minorities and by Hausas and finally by the Yorubas. Sum of square between, within and total were 427.64, 41675.94 and 42103.58 respectively while the mean square between and within are 142.55 and 89.62. Calculated F-ratio is 1.59 while sig value 0.19. Therefore, since sig value (p=0.19> 0.05) is higher than 0.05 alpha at 3 and 465 degrees of freedom, the null hypothesis is accepted. This indicates that there is actually no significant difference in the RAIT-NV scores of individuals from various ethnic groups.

d. There is no significant difference in the RAIT-NV scores of male and female respondents.

Table 4.6 Independent t-test analysis of RAIT-NV scores of male and female respondents

Gender	N	Mean	St.d	Df	Alpha	t	Sig.	Result
Male	141	53.57	10.74	467	0.05	0.49	0.65	Insignificant
Female	328	53.11	8.24					

From the table, male respondent were 141 while female respondents were 328. Mean values were 53.57 and 53.11. The mean scores have shown that make respondents scored higher than the females. Calculated t was 0.49 while sig was 0.65. hence, since sig (p=0.65>0.05) is higher than the alpha, the null hypothesis was accepted. This means that there is no significant difference in the RAIT-NV scores of male and female respondents.

Research Question Four: What is the Cronbach Alpha reliability index of RAIT-NV?

	N of Items	Cronbach Alpha	Remarks
NVA	52	0.523	Low Reliability
SEQ	43	0.808	High Reliability
Total	95	0.66	Average Reliability

The table shows that number of items for non-verbal Analogies (NVA) is 52 while sequence (SEQ) is 43. Calculated Cronbach reliability index for NVA is 0.52 which was adjudged as low reliability. On the other hand, reliability index for SEQ is 0.80 which was adjudged to be a high reliability. On the whole, an average reliability index of .66 was realized. This could be seen as an average reliability.

Research Question Five: What is the reliability index of RAIT-NV using KR_{20}?
Table 4.6; internal consistency of RAIT-NV for use in Nigeria using KR_{20}

International Journal of Developing and Emerging Economies
Vol.12, No.2, pp.86-109, 2024
Print ISSN: 2055-608X (Print),
Online ISSN: 2055-6098(Online)
Website: https://www.eajournals.org/
Publication of the European Centre for Research Training and Development -UK

	N	∑X	$\sum X^2$	ŏ	$ŏ^2$	K	∑Pq	KR_{20}
NVA	469	13289	384317	4.07	16.56	52	6.58	0.62
SEQ	469	13519	397531	4.09	16.73	43	7.31	0.58

From the table, it could be deduced that calculation of KR_{20} was done separately for both NVA and SEQ sections of RAIT-NV. The respondent had sum of 13289 and 13519 respectively for NVA and SEQ. Sum of squares were 384317 and 397531 for RAIT-NV. Standard deviation value and the variance was 4.07; 16.56 and 4.09; 16.73 respectively. For the items, NVA had 52 items wile SEQ had 43 items. The sum of the proportions of students who got the items correctly and wrongly was 6.58 and 7.31 respectively. KR_{20} reliability for both items were 0.62 and 0.58

Research Question Five: What is the reliability index of RAIT-NV using Split-Half?
Table 4.7: Split-Half Reliability of RAIT-NV in Nigeria.

		∑X, ∑Y	$\sum X^2, \sum Y^2$	N	∑XY	rh	Rf
NVA	ODD	6498	93616	469	94662	.04	0.07
	EVEN	6823	102969				
SEQ	ODD	6889	104923	469	97743	0.10	0.18
	EVEN	6630	97122				

The table shows that NVA and SEQ had odd and even numbers of 6498, 6498 and 6889, 6630 respectively. Sum of squares for odd items were 93616 for NVA and 104923 for SEQ while even numbers had 102969 and 97122. Sum of the products for both NVA and SEQ were 94662 and 97743. The reliability of the pairs of odds and even are 0.04 and 0.10 while their full reliability is 0.07 and 0.18.

DISCUSSION OF FINDING

Based on finding one, it is shown that difficulty index of the items thirty three (33) items were seen to be too easy in the test. Thirty two (32) items were of average difficulty index while thirty (30) items were considered too difficult for the respondents and having an overall difficulty index of 0.55 (55%) in Nigeria. From the 33 items that were too easy, it could be that the students are very familiar with these set of questions. It could also be that they were simply focused within the time limit of response to the item. On the aspect of the difficult items, it could simply be that the students did not understand these questions are may have been too confused to know the exact answer. In another point of view, it could be that the majority avoided or skipped these items or may have had no time to attempt them thereby seeming unanswered and difficult. It may also be that such easiness index or difficulty level may come as a result of the cultural differences of respondent outside the country and those in Nigeria.

However, it should also be noted that the present standard for adjudging the levels of difficulty may vary and which the present one was only based on that of Wers and Thurs ranging from .30-.80. Therefore, if the researcher had followed any other standard which others may have set will give a total different view of the items with low, average of high difficulty indices. However, it should be noted that items with low difficulty indices (hard items) are also encourage to balance the curriculum content likewise some of high difficulty indices. The analysis also reveals that some of the items are negative. A negative index indicates a mis-match for the items. This means that some of the items may not have measured what it purports to measure and this could only be explained by the fact that some of them use foreign symbols and analogies which are totally unfamiliar with the respondent here in Nigeria resulting in inability to make reasonable choice.

From finding two an based on ScorePak (2020), item discrimination is "good" if the index is above .30; "fair" if it is between .10 and .30; and "poor" if it is below .10. From this premise, it is clear that nineteen (19) items showed good discrimination. Twenty eight (28) items fairly discriminated while forty eight (48) items discriminated poorly. This means that majority of the items of RAIT-NV were not able to distinguish between the high performing students and the lower performing one. This implies that there is no much difference between students who are intelligent and those who are less intelligent. So far, only 19 items showed good discrimination while 28 showed fair discrimination. The outcome of the finding here could only be explained that by the fact that some of the intelligent respondents that responded accurately in the earlier part of the test might have ran out of time thereby making them not to respond to other items which were treated as zero. It could also be that some of the respondents especially those in the school settings bid supervision and were able to dub some right answers from the intelligent students. The finding is not surprising maybe due to the differences in culture and the timing nature of speeded nature of the test. Also, respondents here in Nigeria may not be that familiar with non-verbal or pictorial representations that the test depicts. This could be the reason while the test could not discriminate properly here in Nigeria.

Finding three revealed the construct validity of RAIT-NV via hypothesis testing evidence. First, it was reported that there is a significant difference in the RAIT-NV scores of individuals from the various age groups with those between 21-40 years performing better. This means that individuals performance in the intelligence relative to age. The implication of this is that one should not relate the performance of respondent across ages as this will give an unfair comparism. It has also shown that those between the age range of 21-40 scores higher with those between 11-15 years and those above 40 years scores less. This also signal that intelligence will come more with chronological maturity which falls at the onset, climax at the center and falls again at the later years of life producing a perfect bell-shape. However, there may be some few exceptions to this age relativity of intelligence.

It is also reported that there is a significant difference in the RAIT-NV scores of individuals from various educational level with those within the Ph.D. level scoring high. Clearly, this has suggested the relevance of education in intelligence test. Though there are various aspect of intelligence (the g and f factor etc.), the result here proves that educational level is an important factor in determining the extent to which individual intelligence can go. The finding here is not unexpected to the researcher because he is quite aware that maturity in terms of intelligence is age relative where it can reach climax at the mid-ages and decline in the later years.

Concerning ethnicity, it is revealed that there is an insignificant difference in the RAIT-NV scores of individuals from the various ethnic groups. This finding simply means that the ethnic groups in Nigeria does not play a part in determining the outcome of intelligence test scores. However, the finding here is not surprising to the researcher because ethnic differences does not and cannot determine the intelligence level of individuals. As clearly seen, though there may be differences in the mean of the various groups, such differences are not significant enough to declare one ethnic group superior to others.

It is also reported that there is no significant difference in the RAIT-NV scores of male and female respondents. This finding means that both male and female students did similarly. That is, in items that are tough, both gender experience difficulties vis-avis items that are easy. This finding further buttress the fact as reported by some researchers that gender plays no role in the brain power of individual. Although there are many other findings like that of Abi (2016) showing that gender has significant influence on the academic performance of individuals. This finding may have come along because both male and female respondent put in similar effort in responding to the items of the instrument.

Based on the reliability of RAIT-NV using internal consistency method, Cronbach Alpha reliability index for NVA is 0.52 while that of SEQ is 0.81. On the whole, the overall index of RAIT-NV is 0.66. According to Kpolovie (2014), the higher the reliability of a test, the higher the quality of such a test. Hence, with this values, it is shown that the RAIT-NV has a low reliability index here in Nigeria as a result may not necessarily be a good test in terms of consistency of scores. This low reliability index could be as a result of the inability of the respondents to have similar understanding of the items of the questionnaire. It could also be that some of the respondents are not familiar with the non-verbal representations while some are. This however has created a gap that makes the test inconsistent over time. The result also show that respondent reliability index in the SEQ section was high. The reason for this could be that respondent are a little bit familiar with sequence more that pictorial or non-verbal analogies. However, this section (SEQ) is very much ok for use in Nigeria whenever one is interested in finding out the sequence of things.

It is also reported that $KR_{20\ reliability}$ index is 0.62 for NVA and 0.58 for SEQ. The finding is also in the range of indices as reported by Cronbach. It could also be established from this that some of the items were unfamiliar to the students. Still on the item range as postulated by Kpolovie (2014) it is obvious that the reliability index here as given by KR_{20} is not high for use here in Nigeria.

Finally, Split-Half reliability indicates a poor reliability index of 0.07 and 0.18 for NVA and SEQ respectively. This suggests total rejection or reorganization or modification of these items to suit the respondents here in Nigeria.

CONCLUSIONS

The findings of the study showed that for difficulty index, only 32 items while for discrimination, only 19 items were good items to form the entire RAIT-NV test if we have to adopt the test for Nigerian use. In terms of reliability indices it is seen that all the indices show a relatively average reliability. Hence, for consideration for adoption of RAIT-NV for Nigerian use, there MUST as a matter of fact restructure, modify, add or subtract some of the items of RAIT-NV. Else, it is only 51 items out of the 95 that are qualified to be adapted for use in Nigeria here. In terms of reliability, modification when carried will help in establishing consistency of the instrument over time.

Recommendations

From the findings of the study, it is recommended that;

1. All foreign items adopted to be used in Nigeria MUST as a perquisite requirement pass through an effective item validation process. This means that experts in measurement and evaluation should be up and doing in terms of validating foreign items and making it ready for use n Nigeria.
2. Test users who wish to use the RAIT-NV should be mindful of the items they are to use. Specifically, they should either modify or remove items 22, 25, 33, 38, 39, 41, 42, 43, 44, 45, 47, 48, 49, 50, 51, 52, 60, 76, 81, 82, 83, 86, 87, 89, 90, 91, 92, 93, 94 and 95 since they are too difficult as well as items 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 23, 24, 29, 30, 53, 54, 55, 57, 58, 59, 63, 64 and 70 which are seen to be too easy. However, they are also recommendations that some of those difficult and easy items are also needed in the test to modify the test.
3. In terms of distracter, test users who wish to used RAIT-NV should be wary of items 1, 2, 3, 4, 5, 9, 10, 12, 17, 18, 22, 23, 24, 25, 29, 30, 38, 39, 41, 44, 45, 46, 47, 49, 50, 51, 52, 54, 55, 58, 59, 60, 61, 68, 69, 70, 73, 74, 79, 84, 86, 88, 90, 91, 93, 94, and 95 which are considered not good discriminators. Hence, appropriate adjustment should be made by test users in these items before they are used. Otherwise, they should avoid these items.
4. The items of RAIT-NV should be properly modified in order to have consistency of the items over time.

REFERENCES

Beres, K., A. & Perlman, M., D. (2000) *Handbook of Psychological Assessment (Third Edition),*

Bracken, B. A. (1985). A critical review of the Kauf- man Assessment Battery for Children (K-ABC). School Psychology Review, 14, 21-36.

Brookhart, S. M. (2004). *Educational assessment of students*. Upper Saddle River, NJ: Merrill-Prentice Hall.

Carmines, E. G. & Zeller, R. A. 1979. Reliability and Validity Assessment, Newbury Park, CA, SAGE.

Deary, I., (2012) Intelligence. Annual Review of Psychology 63, 453–482.

Golden, C. J. (1981). The Luria-Nebraska Children' s Battery: Theory and formulation. In: Hund, G. W. and Obrzut, J. E. (Eds.), Neuropsychological Assessment of the School-age Child. New York: Grune and Stratton.

Gottfredson, L.S. (1997). What do we know about intelligence? American Scholar, 15-30.

Haladyna, T.M. (1994). Developing and validating multiple-choice test items.Hillsdale, NJ: Lawrence Erlbaum Associates.

Hambleton , R.K. & Rogers, H.J. (1991). Advances in criterion-referenced measurement in R. Hambleton & J. Zaal (eds.) Advances in educational and psychological testing. Boston: Kluwer Academic Publishers.

Hinton, P. R., Brownlow, C., Mcmurray, I. & Cozens, B. (2004) SPSS explained, East Sussex, England, Routledge Inc.

Huck, S. W. 2007. Reading Statistics and Research, United States of America, Allyn & Bacon.

Inhelder, B., & Piaget, J. (1958). The growth of logical thinking from childhood to adolescence. New York: Basic Books.

Kamphaus, R. W., and Reynolds, C. R. (1987). *Clinical and research applications of the K-ABC*. Circle Pines, MN: American Guidance Service.

Kamphaus, R. W., Beres, K. A., Kaufman, A. S., & Kauf- man, N. L. (1995). The Kaufman Assessment Battery for Children (K-ABC). In C. S. Newmark (Ed.), Major psychological assessment instruments (2nd ed.). Bos- ton: Allyn & Bacon.

Kamphaus, R. W., Beres, K. A., Kaufman, A. S., & Kauf- man, N. L. (1995). The Kaufman Assessment Battery for Children (K-ABC). In C. S. Newmark (Ed.), Major psychological assessment instruments (2nd ed.). Bos- ton: Allyn & Bacon

Kaufman, A. S., & Kaufman, N. L. (1993). Interpretive Manual for Kaufman Adolescent & Adult Intelligence Test. Circle Pines, MN: American Guidance Service, Inc.

Kaufman, A.S. (1990). Assessing adolescent and adult intelligence. Boston: Allyn & Bacon.

Kpolovie P., J. (2014) Test, Measurement and Evaluation in Research (end Ed.)owerri. Springfeild Publishers.

Luria, A. R. (1970). The functional organization of the brain. *Scientific American*, 18, 266–268.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist, 50*, 741-749.

Obringer, S. J. (1988, November). A survey of perceptions by school psychologists of the Stanford-Binet IV. Paper presented at the meeting of the Mid-South Edu- cational Research Association, Louisville, KY.

Orluwene., G.W. (2012) Test Theory and Development Process. Chris-Ron Integrated services

Piaget, J. (1972). Intellectual evolution from ado- lescence to adulthood. Human Development, 15, 1-12.

Reynolds, (1998). Perspectives on bias in mental testing. New York: Plenum.

Rivera, J., E. (2007) Test Item Construction and Validation: Developing A Statewide Assessment For Agricultural Science Education. A Dissertation Presented to the Faculty of the graduate School of Cornell University in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Stern, E. (2015). Intelligence, Prior Knowledge, and Learning. International Encyclopedia of the Social & Behavioral Sciences, *323–328.* doi:10.1016/b978-0-08-097086-8.92017-8

Sternberg, R., (2002). Dynamic Testing. Cambridge, New York.Turkheimer, E., Haley, A., D'Onofrio, B., Waldron, M., Gottesman, I.I., 2003. Socio-economic status modifies heritability of IQ in young children. Psychological Science14 (S), 623–628.

Wechsler, D. (1974). Wechsler intelligence scale for children—revised. New York: Psychological Corporation.

Whitley, B. E. 2002. Principals of Research and Behavioural Science, Boston, McGraw-Hill