

25.10.2025 / Hall-20, Session-6



ANKARA LOCAL TIME

13⁰⁰ : 14⁴⁵

MEETING ID: 815 6475 4742



PASSCODE: 906453

HEAD OF SESSION: Dr. Umar Idris Sarkin BAUCHI

TOPIC TITLE	AUTHORS	AFFILIATION
Relationship Between Parental Income, Occupation And Students' Academic Achievement In Biology Among Secondary Schools In Agala Local Government, Niger State, Nigeria	Mohammed, AMINA Musa Saliha ABUBAKAR Idris, Umar Sarkin BAUCHI(PhD)	Federal University Of Technology NIGERIA
Reexamination Of Archaeological Evidence From Busayra And As-Sela: The Reality Of Wadi 'Abqar	Mohammed WAHEEB	Hashemite University Jordan
Hamacher Based Aggregation With Entropy Measure For Decision Analysis In Interval Valued Q-Rung Orthopair Hesitant Fuzzy Sets	Scholar.Oroosha BOKHARI Dr. Jawad ALI	Kohat University PAKISTAN
Occurrence Of Staphylococcus Aureus On Environmental Fomite, Ife East Osun State, Nigeria	Idris ADEOLA Timothy Oluwafemi OKIKIJESU Esther Ajibola OKUNOLA	Obafemi Awolowo University NIGERIA
Impact Of Green Energy On Socioeconomic Development In Pakistan	Muhammad NOMAN	Kohat University PAKISTAN
A Simplified Pathway To The Development Of Valid And Reliable Teacher-Made Tests: Innovative Assessment Techniques For 21 St Century Teachers	Dr. Umar Idris Sarkin BAUCHI	Federal University NIGERIA
Developing The Modified Ishita Distribution (MID) And Comparing Parameter Estimation Techniques For The NHPP- Ishita And NHPP MID Models	Prof Narinder PUSHKARNA MUSTAFA	University of Delhi INDIA



Scientific Development and Innovative Approach

11th INTERNATIONAL MARMARA SCIENTIFIC RESEARCH AND INNOVATION CONGRESS 24-25 OCTOBER İSTANBUL

Congress Program

Participant Countries:

TÜRKİYE, AZERBAIJAN, ALGERIA, BANGLADESH, IRAN, INDIA, JORDAN, MOROCCO
NIGERIA, INDONESIA, PAKISTAN, ROMANIA, UZBEKISTAN

IMPORTANT, PLEASE READ CAREFULLY

To be able to make a meeting online, login via <https://zoom.us/join> site, enter ID instead of "Meeting ID or Personal Link Name" and solidify the session. The presentation will have 15 minutes (including questions and answers). The Zoom application is free and no need to create an account. The Zoom application can be used without registration. The application works on tablets, phones and PCs. Speakers must be connected to the session 10 minutes before the presentation time. All congress participants can connect live and listen to all sessions. During the session, your camera should be turned on at least %70 of session period. Moderator is responsible for the presentation and scientific discussion (question-answer) section of the session.

TECHNICAL INFORMATION

Make sure your computer has a microphone and is working. You should be able to use screen sharing feature in Zoom. Attendance certificates will be sent to you as pdf at the end of the congress. Moderator is responsible for the presentation and scientific discussion (question-answer) section of the session. **Before you login to Zoom please indicate your name surname and hall number, exp. Hall-1, Fatih KARİPOĞLU**

ÖNEMLİ, DİKKATLE OKUYUNUZ LÜTFEN

Kongremizde Yazım Kurallarına uygun gönderilmiş ve bilim kurulundan geçen bildiriler için online (video konferans sistemi üzerinden) sunum imkanı sağlanmıştır. Sunumlar için 15 dakika (soru ve cevaplar dahil) süre ayrılmıştır. Online sunum yapabilmek için <https://zoom.us/join> sitesi üzerinden giriş yaparak "Meeting ID or Personal Link Name" yerine ID numarasını girerek oturuma katılabilirsiniz. Zoom uygulaması ücretsizdir ve hesap oluşturmaya gerek yoktur. Zoom uygulaması kaydolmadan kullanılabilir. Uygulama tablet, telefon ve PC'lerde çalışıyor. Her oturumdaki sunucular, sunum saatinden 10 dk öncesinde oturuma bağlanmış olmaları gerekmektedir. Tüm kongre katılımcıları canlı bağlanarak tüm oturumları dinleyebilir. Moderatör - oturumdaki sunum ve bilimsel tartışma (soru-cevap) kısmından sorumludur.

TEKNİK BİLGİLER

Bilgisayarınızda mikrofon olduğuna ve çalıştığına emin olun. Zoom'da ekran paylaşma özelliğine kullanabilmelisiniz. Katılım belgeleri kongre sonunda tarafınıza pdf olarak gönderilecektir. Kongre programında yer ve saat değişikliği gibi talepler dikkate alınmayacaktır. Zoom'a giriş yaparken önce lütfen adınızı, soyadınızı ve SALON numaranızı yazınız, Örnek: Hall-1, Fatih KARİPOĞLU

MEETING ID: 815 6475 4742

PASSCODE: 906453

<https://us02web.zoom.us/j/89495290632?pwd=EHs4P6Z2RbLMf1SahC32knJlFqjubb.1>

25.10.2025 / Hall-20, Session-6



ANKARA LOCAL TIME

13⁰⁰ : 14⁴⁵

MEETING ID: 815 6475 4742



PASSCODE: 906453

HEAD OF SESSION: Dr. Umar Idris Sarkin BAUCHI

TOPIC TITLE	AUTHORS	AFFILIATION
Relationship Between Parental Income, Occupation And Students' Academic Achievement In Biology Among Secondary Schools In Agala Local Government, Niger State, Nigeria	Mohammed, AMINA Musa Saliha ABUBAKAR Idris, Umar Sarkin BAUCHI(PhD)	Federal University Of Technology NIGERIA
Reexamination Of Archaeological Evidence From Busayra And As-Sela: The Reality Of Wadi 'Abqar	Mohammed WAHEEB	Hashemite University Jordan
Hamacher Based Aggregation With Entropy Measure For Decision Analysis In Interval Valued Q-Rung Orthopair Hesitant Fuzzy Sets	Scholar.Oroosha BOKHARI Dr. Jawad ALI	Kohat University PAKISTAN
Occurrence Of Staphylococcus Aureus On Environmental Fomite, Ife East Osun State, Nigeria	Idris ADEOLA Timothy Oluwafemi OKIKIJESU Esther Ajibola OKUNOLA	Obafemi Awolowo University NIGERIA
Impact Of Green Energy On Socioeconomic Development In Pakistan	Muhammad NOMAN	Kohat University PAKISTAN
A Simplified Pathway To The Development Of Valid And Reliable Teacher-Made Tests: Innovative Assessment Techniques For 21 St Century Teachers	Dr. Umar Idris Sarkin BAUCHI	Federal University NIGERIA
Developing The Modified Ishita Distribution (MID) And Comparing Parameter Estimation Techniques For The NHPP- Ishita And NHPP MID Models	Prof Narinder PUSHKARNA MUSTAFA	University of Delhi INDIA

**A SIMPLIFIED PATHWAY TO THE DEVELOPMENT OF VALID AND RELIABLE
TEACHER-MADE TESTS: INNOVATIVE ASSESSMENT TECHNIQUES FOR 21ST
CENTURY TEACHERS**

BY

DR. UMAR IDRIS SARKIN BAUCHI

**DEPARTMENT OF SCIENCE EDUCATION, FEDERAL UNIVERSITY OF
TECHNOLOGY MINNA, NIGER STATE.**

Phone: +2348065378781

Email: umar.idris@futminna.edu.ng

**BEING A PAPER PRESENTED AT THE 11TH INTERNATIONAL MARMARA
SCIENTIFIC RESEARCH AND INNOVATION CONGRESS**

24-25 OCTOBER, 2025 AT ISTANBUL

Abstract

Acquisition of effective and innovative assessment knowledge and techniques among 21st century teachers are paramount to the attainment of educational goals of any nation. This underscores the need to educate and equip teachers at all levels of education on the crucial innovative assessment practices expected for the production of accurate and reliable feedback information during and after classroom instruction. The paper explored and discussed tests and some key psychometric properties that teachers should be conversant with in order to improve the quality of teacher-made tests which could generally result in the enhanced classroom assessment and evaluation. The paper further enumerated adoptable practical steps in item construction for school teachers and lecturers capable of improving the validity and reliability of the conventional teacher-made tests using effective and modern techniques in educational testing which could enhance the achievement of educational objectives. The author recommended among others, that teachers and lecturers should be having consistent refresher virtual and physical training opportunities as to be abreast of innovative assessment practices and techniques in learning institutions across the world as to foster best practices in educational testing.

Keywords: Test, Validity Reliability, Item analysis, Item difficulty, Item discrimination, Item distractor indices, Testing Software

2025

0

I. Introduction

Education has played significant roles in the transformation of dynamic world in to force for personal development and massive transformation of our societies in to digital world to meet up with the evolving needs of the modern societies. Teaching and learning activities across all levels of education have been in place for time immemorial across the world with records of tremendous developments in all facets of human life. Education encompasses the systematic process of facilitating learning, acquiring knowledge, skills, values, beliefs and habits that influence the development of a society.

Curricula are planned, developed and implemented to achieve the predetermined national educational goals which are focused on the national development of any given society. Teaching and learning activities are the most significant activities in curriculum implementation which can never be successful without effective assessment and evaluation. Thus, the achievement of teaching and learning objectives can only be realistic through assessment. This underscores the need to give much attention on the quality of assessment activities as to ensure the effective and successful achievement of the set goals in education and by extension, the national development goals.

Significance of Testing

Educational assessment is seen as a systematic process of gathering, interpreting, and using information about students learning outcomes. It is an aggregate of the process of testing and recording students cognitive, psychomotor and affective domains empirical data. One of the tools for assessment is test, which can be defined as a set of problems designed to stimulate certain responses in an individual. It is also explained as a tool for appraising level of acquisition of knowledge and skills. In another words, testing is the process of determining the presence and level of acquisition of certain traits in an individual. Considering the significance of educational testing, it becomes necessary for the teachers and other stakeholders in education to be more concerned with the effective testing as a tool for attaining learning objectives.

II. Types Of Test

Test as a tool for assessment are of different types depending on some attributes as follow:

1. Based on Purpose

The first type we have under the purpose it serves is the formative Test. This is conducted during the instructional process to monitor student's learning and provide ongoing feedback. The goal is to identify areas where students are struggling and to improve teaching and learning in real time. Examples include quizzes, class discussions, and assignments.

Second is the summative test, this is administered at the end of an instructional period, such as the end of a unit, semester, or course; to evaluate what students have learned. Examples include final exams, standardized tests, and end-of-term projects. Summative tests are often used for grading and to determine if educational goals have been met.

The third one is the diagnostic Test. This is any test designed to identify students' strengths and weaknesses in specific areas before instruction begins. They help educators understand students' pre-existing knowledge and skills, allowing for tailored instruction. An example is a pre-test given at the start of a course.

The forth type is the placement test: This is used to place students in the appropriate level of a course or program based on their current abilities. Common examples include language proficiency tests for non-native speakers or math placement tests for college students.

The fifth is the proficiency test: This is used to assess a student's competence in a particular area, regardless of where or how the knowledge or skills were acquired. They are often used for certification or to meet professional standards, such as TOEFL for English language proficiency.

2. Based on Format

i. From the format used in the test construction we have Objective Tests. These tests consist of questions with specific correct answers. They are often multiple-choice, true/false, or matching items, and are easy to score. Objective tests are effective for assessing knowledge and comprehension. Examples include standardized tests like the NECO, WAEC and CBT tests in the tertiary institutions.

The second type is the essay tests. These require students to construct their responses, such as essays or short answers. Scoring can be more complex because it often requires judgment on the

quality of the responses. Subjective tests are useful for assessing higher-order thinking skills, such as analysis, synthesis, and evaluation.

The third one is the Short Answer Tests. These comprise all test in which specific responses are elicited to be filled in the spaces or gaps in the test items. It is objective to some extent as responses are restricted unlike in the essay type.

3. Based on Scope

Based on the scope of the measurement, we have achievement tests: These are tests that measure students' knowledge and skills in specific subjects or areas after instruction. They are commonly used to determine how well students have met learning objectives. Examples include end-of-year assessments and unit tests in subjects like mathematics or science.

The second type is the aptitude test: This is any test that assess a student's potential to learn or succeed in a specific area in the future. They are designed to predict future performance rather than assess what has already been learned. Examples include the SAT, UTME, POST-JAMB, Aptitude Battery etc

4. Based on Standardization

Under this we have standardized tests. These are tests administered and scored in a consistent manner across all test-takers. They are often used for large-scale assessments, such as state or national exams, to ensure comparability of results. Examples include the WAEC, NECO, NABTEB and UTME examinations.

We also have non-standardized tests, otherwise known as teacher-made tests: These tests are designed by individual teachers in schools and can vary in administration and scoring. They are often used in classrooms to assess student learning in a specific context. Examples include teacher-made quizzes or unit tests.

5. Based on Mode of Delivery

On the basis of test are given to the test takers, there is Paper-and-Pencil Tests(PPT)which is the traditional test where students respond to questions on paper. These are still widely used, particularly in the classroom settings.

The second one is the Computer-Based Tests (CBTs). This include tests administered via computers or other digital devices. It is also referred to e-assessment and the current trend globally.They are increasingly common due to the ease of administration, scoring, and the ability to include multimedia elements. Examples include online quizzes and standardized tests like the GRE.

The third one is the oral test: This involves verbal questioning and answering, often used in language assessments or viva voce in academic defenses.

6. Based on the Nature of Items

We have Performance-Based Tests, which are tests that require students to perform a task or create a product that demonstrates their knowledge or skills. Examples include science experiments, art projects, or presentations.

In the class, the second type is the Portfolio-Based Tests: These involve the collection of a student's work over time to show progress and achievement in a particular area. Portfolios are often used in art, writing, and other creative disciplines.

III. Basic Psychometric Properties of Test

In educational testing, there crucial properties of test that are given much concern as to ensure quality in assessment. They include validity, reliability, item difficulty, discrimination and distractor indices. These will be discussed in detail as follow:

I. Test Validity

Validity of a test is simple the extent to which a test assesses what is purported to assess (Adekunle, *et al.*, 2018). In a broader definition by Messick (2020), it is the extent to which theories and empirical evidences support the inference made on test scores. Thus, a valid test is to be a true representation of students' responses in a given task without any interference of unexpected or uncontrolled variables during the test construction and administration,

There are several types of test validity in educational measurement, as highlighted in Idris, (2023), which include:

- 1. Face validity:** This is simply the extent to which a test superficially appears to be rich and accurate. It is expected to be free of technical and grammatical errors which can mislead the test takers to give a wrong response capable of undermining the validity of test
- 2. Content validity:** - This is the extent to which a test instrument adequately covered topics and sub-topics from which the items are generated. This could be achieved by the use of logically constructed table of specification.
- 3. Criterion Related validity:** This is a validity check based on certain predetermined criterion. There are two types of validity under this which are:
 - a. Predictive Validity:** This entails validating a test based on future performance scores of a testees. For example, a student's aptitude scores can be correlated with performance in a given programme or course of study to establish the predictive validity of the aptitude performance. Example is UTME and CGPA scores.
 - b. Concurrent validity:** In this type of validity, student's performance scores are correlated with a closely related scores at the same time to confirm the validity of the target test. Example is version A genetics Test and Version B Genetics Test etc
- 4. Construct Validity:** This is the extent to which a test measures a target trait or construct among test takers. A good test should be testing precise construct it was constructed for, for example, motivation test should be able to assess level of motivation in an individual. Genetics test should only be testing genetics dimension not otherwise. There two types of construct validity
 - a. Convergent:** In this type of validity, a test is expected to corelate with the related test, for example, motivation scale should be able to correlate with interest scale. Likewise, mathematics test should be able to correlate with statistics test or any other related mathematics test.
 - b. Divergent:** In the case of divergent validity, a test is expected no to correlate with unrelated constructs or dimensions. For example, a biologist test should not correlate with Mathematics test; a Motivation test should not relate with aggression test. The more a test differs with unrelated test, the more it assumes divergent or discriminant validity.

II. Test Reliability

Test reliability is simply the extent to which a test produces a consistent or stable result after several administration (Taherdoost, 2016, Taba, 2018). In other words, it is the degree of accuracy of a test. There are several ways in which a tester can establish test reliability as captured by Idris (2023):

- 1. Test retest reliability:** In this type of reliability, a test is administered two times, for instance with two weeks interval on the same students or test takers. The results are eventually correlated using Pearson Product Moment Correlation (PPMC) formula to get the coefficient. 0.00 to 0.50 is a poor reliability index. 0.51 to 0.70 is a moderate reliability index. But 0.71 to 0.99 is high reliability index. The perfect index is 1.00 which is rare.
 - 2. Parallel Form reliability:** In this method of reliability, two sets of test instruments are administered to a testees at the same time, especially using sandwich method if they are many. That is, giving version A to first group and version B to the second group. Later the two versions are swapped to the groups. The scores are finally correlated using PPMC for the indices.
 - 3. Split-Half reliability:** This another method of reliability in which a test is administered once. The tester split all individual scores in to two sets for each student. For example, a student with 10 score, can be 6 and 4, or 5 and 5. The two sets of score are eventually correlated using Spearman Brown Formula to obtain the indices. Same rules also apply.
 - 3. Internal consistency:** Internal consistency is another method of reliability check in which response in test items are observed on how they are correlated.
 - a. Kuder Richardson 20 (KR20):** KR20 is another method of internal consistency in which test is administered once and scores generated are correlated using the formula. The coefficient obtained determines the level of reliability using the rule of thumb early explained.
 - b. Kuder Richardson 21 (KR21):** This an upgrade of KR20 method with little difference in considering item difficulty as in the case of the KR20. The coefficient stands as the others.
- One important note to be taken is that all the previously discussed reliability types are only applicable to test instruments that are dichotomously scored (1 or 0) as in achievement and aptitude tests that are in ratio or interval scales.

c. Cronbach Alpha: This one of the most used method especially on test instruments that have multiple scoring parameters like questionnaire with Likert scale. The formula is applied to the scores and produce internal consistency coefficient that determines the level of reliability of an instrument.

In educational testing, much attention is always paid to the procedures of item construction in order to enhance the achievement of testing objectives. One of the statistical techniques applied by learning institutions, testing organizations and examination bodies is item analysis. This is a statistical technique used to enhance the quality of test items (psychometric strengths) before final administration of the tests to select or reject items (Rezigalla,2022). It is rigorous exercise that can be done manually and with some statistical packages such excel, Test Analysis Programme (TAP) and latest Artificial Intelligence applications. The procedure results to the identification of qualities of test items such as difficulty, discrimination and distraction. The tests that undergo these procedures tend to become more valid and reliable that foster the attainment of educational objectives across the nation.

The following is the discourse on some of the aforementioned concepts:

III. Item Difficulty: This refers to the level of difficulty and ease of test item in a test instrument (Rezigalla,2022, scribd.com, 2025). It is determined by observing the proportion of learners or students who got an item right compared to those who got it wrong and is symbolized as P.

In a manual computation, the teacher or tester should do the following:

- i. The first step is to arrange the students' scores in to a descending order, for instance from the highest to the lowest. E.g. 100, 95, 90, 80, 75, 70, 65, 60, 50, 45,40, 35, 33, 32, 30, 28, 25,22, 20, 10. These are scores of 20 students.
- ii. Then take 25% of the upper scores, and 25% of the lower scores. These will be represented as upper 25% and Lower 25% respectively. In this example, students with scores 100, 95, 90, 80, 75 = U_{25} , while 30,28,25,22,20= L_{25} . By implication all computation will be done using the responses of these sets of students throughout.
- iii. Determine the total number of the students, in this case, 15 students are involved.

iv. You now begin to check item by item using these groups' responses to determine how easy and difficult is each item.

Example:

a. Find the difficulty of an item in a test whose response is 6 students got it right, while 4 students got it wrong: $U_{25} = 4, L_{25} = 2, N = 10$

$$P = \frac{\text{Total number of students that got it right}}{\text{Total no of students attempted the item}}$$

$$P = \frac{U_{25} + L_{25}}{N} = \frac{4 + 2}{10} = 0.60 \text{ (Moderately difficult)}$$

b. Find the difficulty of an item in a test whose response is 3 students got it right, while 7 students got it wrong- $U_{25} = 2, L_{25} = 1, N = 10$

$$P = \frac{\text{Total number of students that got it right}}{\text{Total no of students attempted the item}}$$

$$P = \frac{U_{25} + L_{25}}{N} = \frac{2 + 1}{10} = 0.30 \text{ (30\% got it right, very difficult)}$$

Table 4: Interpretation of Item Difficulty

S/N	Difficulty Value(P)	Interpretation	Remark
-----	---------------------	----------------	--------

1.	1.00	Perfectly easy item, all students got it right, revise	Discard
2.	81-99	Very easy item	Discard
3.	61-80	Easy item	Revise
4.	41-60	Moderately difficult	Accept
5.	21-40	Difficult item	Revise
6.	00-0.20	Very difficult item	Discard

Source: scribd.com (2025)

IV. Item Discrimination

Item discrimination is the extent to which an item differentiates between high and low performing students in a test (Rezigalla, 2022, scribd.com, 2025)

). It is signified by Discrimination Index (DI). A good item should be able to show disparity between higher ability and lower ability students, therefore, should always have positive figures. Whenever, DI figure are found in negative notation, it shows indicates that such item is favouring the low achievers, in other words, more of low achieving student or testees scored it right compared to the high achievers.

Table 5: Item Discrimination Table

S/N	Discrimination Index (DI)	Interpretation	Remark
1.	0.00	Zero discrimination	Discard
2.	-1.00 to -0.60	Questionable item	Discard
3.	-0.59 to -0.20	Not discriminating	Discard
4.	-0.19 to 0.20	Moderately discriminating	Revise and improve
5.	0.21 to 0.60	Highly discriminating	Retain
6.	0.61 to 1.00	Very much discriminating	Retain

Source: scribd.com (2025)

V. Item Distractor Index (IDI)

Item distractor index is the process of determining the extent to which options in an item distracts respondents from lower achieving testees (L₂₅) (Rezigalla, 2022, scribd.com, 2025). Focus is on the lower achieving students as to get much information about the strengths of attraction of the wrong options provided in an item.

Examples

Table 6: Item Distractor Indices

OPTION	A	B	C	D*
L ₂₅ (10)	3/10	4/10	6/10	2/10
IDI	0.30	0.40	0.60	0.20

To calculate the IDI, the lower 25% is usually used. The frequency of the lower achievers selecting an option is divided by the number of the lower achievers e.g. $\frac{L_{25}}{N}$

N

In table 6 above, option A has IDI = 0.30, indicating that about 30% of the students were attracted to the option. Option B was better in IDI which has about 40% attracted to. Option C the highest distractor which has 60% respondents choosing it. The right option has the least percentage which was 0.20 which is a good IDI as 20% were attracted to the option. When an option records zero IDI, it seems it is a terribly bad distractor and quickly be eliminated and replaced. The more the wrong options distract, the more the IDI of such item

IV. Construction of Test Blueprint (Table of Specification)

There are lots of procedures and steps to follow before achieving valid test items

Step i: Assemble all evaluation questions of your objectives of instruction (lesson plan)

Step ii: Delineate the purpose of the test (what are you to assess?)

Step iii: Write down the topics and sub-topics across the periods covered

Step iv: Decide on the no of questions to be developed

Step v: Get the proportion of questions by dividing the number of periods by the total no time covered

Step vi: Multiply the percentages by the desired no of questions to get no of question per topic

Step vii: Construct your table of specification

Step viii: Allocate the questions across the six cognitive domains (Knowledge, comprehension, application, analysis, synthesis& evaluation, Blooms 1954) or use Anderson and Krathwohl (2001) revised taxonomy

Step ix: Start writing the items using the table as a guide and following the guidelines for writing tests

Table 2: Weightage and Number of Items Table

S/N	TOPICS	PERIOD	PERCENTAGE	NO OF QUESTION (40)
1	Reproduction in human	2	$2/15 \times 100 = 13.3$	5.33
2	Fertilization	2	$2/15 \times 100 = 13.3$	5.33
3	Sex determination	4	$4/15 \times 100 = 26.6$	10.64
4	Variation	4	$4/15 \times 100 = 26.6$	10.64
5	Genetic mutation	3	$3/15 \times 100 = 20$	8
Total		15	100	39.94 appx. 40

Table 2 describes the percentage and number of test items to be drafted based on the content learnt. In the first place, topics are listed according to the periods, weeks, hours they are taught. Next is to take each topic or subtopic, divide its period or taken in teaching by the total period spent on all topics. For example, in the item 1 in the table, reproduction in human was taught in two periods, and the total periods is 15. Hence, $2 \text{ divided by } 15 \times 100 = 13.3$, which means we need 13.3% of the test questions to be developed from the concept of genetics. If we move further, 13.3% will be

multiplied by the target number of questions, in this example, 40. This will give rise to 5.33 (5) number of questions. By implication, only five questions will be under concept of reproduction. Same procedure will be followed up to the last topic. In cases of decimals and fractions less than significant figure (1-4), they are ignored, but .5 and above can be rounded-off to get one item arbitrarily.

Table 3: Table of Specification on Genetics

S/N	TOPICS	REM	UND	APPL	ANALS	EVAL	CREAT	TOTAL
1.	Reproduction in human	1	2	1	1	.	.	5
2.	Fertilization	1	1	1	.	1	1	5
3.	Sex determination	2	4	1	2	2	.	11
4.	Variation	2	2	3	.	2	2	11
5	Genetic mutation	2	1	2	.	2	1	8
	TOTAL	11	17	12	3	7	4	40

Table 3 describes table of specification based on Anderson and Krathwohl (2001) taxonomy. It is an extension of the previous computation because the proportion of items planned is strictly used here. The main event in this table is to divide each number of items across the desired level of cognitive domain. For the LOTS, items should be distributed to remembering, understanding and applying. For the HOTS, items should be distributed across analyzing, evaluating and creating. In the above table, all items were distributed across the levels and the total of 40 items as planned were achieved. This gives a test a rich content validity after construction and administration, especially following the other writing procedure.

Hints for Writing Objective Test

- Use the table of specification

- Concentrate on the major concepts
- Consider the difficulty, discrimination and distractor indices
- Clearly write the items without ambiguity
- Use simple English, considering the learners' level
- Never give clues, e.g. a, an, bolding, underlining

Hints for Writing Essay Test

- Items should ask about the target constructs
- Use simple sentences for easy understanding
- Avoid confusing test takers in the items, i.e. no ambiguity
- Don't give too much and too lengthy questions
- Don't ask out of context
- Indicate the marks per item
- Give instructions

Score with strict adherence to marking guide

V. Effective Techniques for Test Administration

After rigorous activities in item construction, there is every to pay much attention on how to administer the test. Administration of test is as vital as the test construction itself as foul play at this stage, could render the whole activity worthless.

The following are some of the guidelines for effective test administration:

1. Date, time and venue of the test should be clearly explained. This will give the students or testees ample time and direction to mentally and physically get ready for the test. Venues that not easily accessible should be avoided at all cost as ensure hitch-free testing

2. As stated in the principles of testing, the students should be aware of the format to be used in the test. This is to enable them get prepared for the expected task in the test. Hiding format of a test is never a security technique for the process but a destructive measure to the testing itself.
3. The environment should be conducive for the testing. There should be enough and comfortable seats, desks, spacious cubicles in CBT, enough lighting, fresh air, toilet facilities, water facilities. The environment should be free of any noise pollution, the examiners should avoid making unnecessary shouts and chatting among them in the hall as doing so distract the thinking of the testees. Any discomfort during testing could negatively affect student's performance which is often called Construct Irrelevant variance (CIV) in educational testing.
4. There should be explicit instruction on the timing of the test which should never be manipulated in the process, as timing of test can also affect students' performance. Also instruction on how to write, how many questions to answer should be clearly stated without any alteration.
5. There should be enough staff to invigilate the testees. Sometimes students could be dubious and cheating once left unguarded. In the era we are, there should be enough CCTV and technologists to monitor the security of testing. It is also important to motivate the invigilators with token and food as to make them have full concentration in securing the testing exercise throughout.
6. There should be proper documentation of examination scripts, number of the issued and returned scripts should be known. Examiners should make sure that scripts are duly signed to avoid loss, illegal replacement and insertions during or after testing.

VI. The Use of Software in Educational Testing

The coming of Information and Communication technology (ICTs) in to educational testing has been yielding positive results over the time. This is more evident with the invention of Artificial Intelligence (AI) which simplified the whole testing activities. This segment of the paper is to provide a hint on some of the AI and non-AI packages that are integrated in to educational assessment to enrich the 21st century teachers with current and innovative practices expected to improve the quality of educational assessment.

i. Artificial Intelligence(AI) for Innovative Assessment

The revolution of artificial intelligence is fast becoming more evident in the contemporary societies and has penetrated all percepts of human endeavour including education. Apart from integration of AIs in to teaching and learning, AIs have become innovative tools for assessment in learning institutions to enhance effective evaluation activities in schools.

There are several AI-Based software that are currently in use in the field of educational testing which include PrepAi, Smallpdf AI Question Generator, Quizbot, Eklavya, Revisely Quiz Generator, Jungle AI and StudyPDF among others.

1. PrepAi: This one of the trending AI application that is used in educational testing. It is used in generating objective test ranging from multiple choice items, matching type items and statement-based items. It operates with texts, PDF, videos to some extent youtube links in item construction. One significant role it plays is that it enhances generating items assessing Higher Thinking Order Skills. In addition, PrepAI enables instant scoring of tests and examinations to facilitate quick reports and placements without delay. PrepAI is suitable for assessment in secondary school and tertiary institutions that syllabus based. The only drawback of this application is the need for consistent subscription for full access.

2. Smallpdf AI Question Generator: This another innovative AI application that is free for basic use and has the ability to converts PDFs, Word files, or images into multiple choice items, true or false and open-ended within a fraction of time. Users can get access to it without any signup as it is in many of the available AIs and so flexible for teachers and students that needs to construct quick quizzes in the classes.

3. Quizbot: This is another astonishing testing application which is has multilingual support, in other words, it can generate items in 55 languages. Quizbot creates items from audio, video, images, and web links. It also supports differentiated assessments based on cognitive level and learning style. The application is suitable for global teams, language tutors, and institutions requiring large pool of item banking in different formats. One of the major concern about the quizbot is that its based on subscription not free.

4. Eklavya: This is another sophisticated testing application that is syllabus-based in item generation and enables mapping questions to specific topics or proportion in a syllabus. The most

astonishing and interesting feature of this application is that it enhances adaptive testing, which is treating an individual based on his or her ability. It has high level of item security by preventing item leaks. Its highly recommendable for tertiary institutions and testing organization such as NECO and WAEC that conduct high stake examinations. It is also accessible based on subscription.

5. Revisely Quiz Generator: This application is another robust type AI which converts notes and diagrams in to well-constructed test items for assessment. It a user-friendly application that works on mobile phones (Androids), laptops and desktops. Learners can generate self-assessment from their notes for daily classroom exercise using free basic version.

6. iJungle AI: Another interesting AI for educational testing is iJungle which produce real world assessments. It is appropriate for assessment in medicine, law and other disciplines that deal with physical and practical situations. iJungle is best for professional certification assessments but can also be adopted in general assessments among learning institutions.

7. StudyPDF: This also a robust application that is accessible by students for general preparation for high stake examinations such as WAEC, NECO, NAPTEB and UTME in Nigeria. It helps in generating test items similar to the target examinations so that students could enhance future performance. One unique feature of the application is that it helps in identifying knowledge gaps or weaknesses among students using performance data which helps in mitigating poor performance in the future examinations.

item difficulty, discrimination and distractor index.

ii. Software That can be Used for Item Analysis

The following are some of the software that are used in conducting item analysis and other psychometric analysis in institutions of learning, banks and testing organizations:

1. Microsoft Excel: This the common excel package used in Microsoft ward. It very user friendly and perform a lot os statistical and graphic tasks. Basic item analysis and inferential statistics can be conducted with excel.

2. Statistical Package for Social Sciences (SPSS) with the "AMOS" Module: This another widely used package in education and commerce and more accessible than R package. It sometimes integrated with AMOS Module, a powerful add-on for conducting advanced factor

analysis and Structural Equation Modeling (SEM) visually, using path diagrams. The package is used for basic analysis such as descriptive statistics, inferential statistics, item-total correlation, difficulty, discrimination and reliability analysis, such as Cronbach's Alpha.

3. **Stata:** This is a statistical software that perform analyses ranging from statistical analysis, data management and visualization to automated reporting. It is widely used in academic researches in learning institutions, banks, and statistics departments in government agencies.

4. **Test Analysis Programme (TAP):** This another simple tool used in item analysis. Its easily accessible and less complex. It analyses item difficulty, discrimination using point biserial correlation and other necessary item information. It used by teacher and lecturers, and students usually adopt it for exercises. It is downloaded free.

5. **ITEMAN:** This is a software designed for conducting Classical Test Theory (CTT) item analysis. Though seem obsolete, but it remains relevant as its functions are now integrated in to more sophisticated software such Winstep and Flexmirt which are good for IRT analyses.

6. **ExamSoft:** ExamSoft: A cloud-based and comprehensive platform common in health science education (medical, nursing, law schools) and high-stakes certification. Its "ExamSoft Strength Report" is a standout feature. It automatically tags every item to learning objectives and provides detailed reports at the student, course, and institutional level, showing where learning gaps exist. For effective security of examinations, the "Exemplify" software downloads the exam to a student's computer and locks it down, preventing cheating among students. ExamSoft is good for metadata and categorization for items. Programs with a strong outcomes-based assessment focus, where linking every test question to a specific competency or objective is critical.

7. **Mplus:** This is analysis package that involves advanced analyses of latent variables. It unifies several statistics such as structural Equation modelling, factor analysis, Bayesian analysis, complex survey data analyses etc.

8. **Python:** This is one of the widely and general purpose software for web development, data science and analytics machine learning and many more. It is an excellent software for the novice analysts but also suitable highly learned.

9. **QuestionMark:** This a software produced by QuestionMark Computing company. It is a robust, enterprise-level platform for high-stakes certification, compliance, and academic assessments. QuestionMark enables multiple authors to collaborate on items with version control and workflow management (draft, review, approved) which improves the test security. It also supports standard

types (MCQ, True/False) as well as complex, interactive questions like hotspots, drag-and-drop, and simulations. It complies with the accessibility standards and usually patronized by organizations, universities, and professional certification bodies that need security, scalability, and a wide variety of item types

10. **Respondus:** This another type of authoring software very popular in the higher education market, often integrated with Learning Management Systems (LMS) like Canvas, Blackboard, and Moodle. It is a powerful Windows-based authoring tool that can create exams offline and then publish them directly to an LMS. Its "Test Bank Network" allows the user to import publisher test banks. Respondus supports the creation of learning activities such as flashcards and games from the same item bank. Recommendable for teachers and lecturers.

11. **TAO (Testing Assisté par Ordinateur):** This an open-source assessment platform designed for next-generation, technology-enhanced items. It is also free to use and modify, which is a major advantage for organizations with technical teams. Deals with Interactive Items such as interactive graphs, ordering tasks, and complex problem-solving scenarios. Developed on open web standards (HTML5, JavaScript), making it highly flexible

12. **R - 'LTM', 'LAVAN', 'PSYCH':** This is a software recommendable for statisticians, psychometricians, and graduate students that possess the basic technical skills and need the sophisticated customizable tools for analysis. It is free but requires programming knowledge in the users. These can perform virtually any psychometric analysis, including Classical Test Theory (CTT), Item Response Theory (IRT), Factor Analysis and Structural Equation Modeling (SEM).

VII. Conclusion

From the previous discussion, it is apparent that psychometric properties of test are indispensable in educational testing and are logical and systematic. The current technologies, AI inclusive, if well integrated in to the testing process and analysis will make it faster and easier to enhance the achievement of effective and quality assessment in learning institutions, provided the basic knowledge and skills are obtained

References

- Adekunle, T. O.; Rafiu, A. O.; & Owolabi, H. O. (2018). Predictive Validity of Students' Final Continuous Assessment Scores in Mathematics and Biology Using Their Performance in National Examinations Council as a Criterion Measure. *FUDMA Journal of Educational Foundations*, 1(1), 148-158.
- Anderson, L. W. & Krathwohl, D. R. (2001). *A taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*: Longman.
- Disha, M. (2019). Validity of a Test: 6Types. <https://www.yourarticlelibrary.com>
- eGyanKosh (2024). *Construction of Evaluation Tools*. <https://egyankosh.ac>.
- Idris, U. S. B. (2023). Effect of mode of Testing, Computer Familiarity and Anxiety on University Students' academic Achievement in Biology in Minna, Nigeria. A PhD Dissertation of the University of Ilorin, Ilorin, Kwara State Nigeria.
- Idris, U. S. B. ; Ezeamagu, M. U.; Aderinto, M. O. & Haliru, Shehu (2017). Improving Content Validity of Teacher-Made Achievement Test: A Way Forward to Ensuring Quality Research and Evaluation. *Journal of the Association of Educational Researchers (ASSEREN)*, 2(1), 167-174
- Kolawole, B. (2010). *Principles of test Construction*. Lagos
- Okwori, R. O., Tafida, A. G. & Idris, U. S. B. (2019). *A book of Reading in Instructional Pedagogy*. Federal University of technology Minna, Niger state.
- Rezigalla, A. A. (2022). Item Analysis: Concepts and Applications. Doi:10.5772.intechopen.100138
- scribd.com (2025). Interpretation of item difficulty. <https://scribd.com>
- Taber, K. S. (2018). The use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, 48(6), 1273-1296.
- Terherdoost, H. (2016). Validity and Reliability of Research instruments: How to Test for Validation of Research Questionnaire/ Survey in Research. <https://www.researchgate.com>