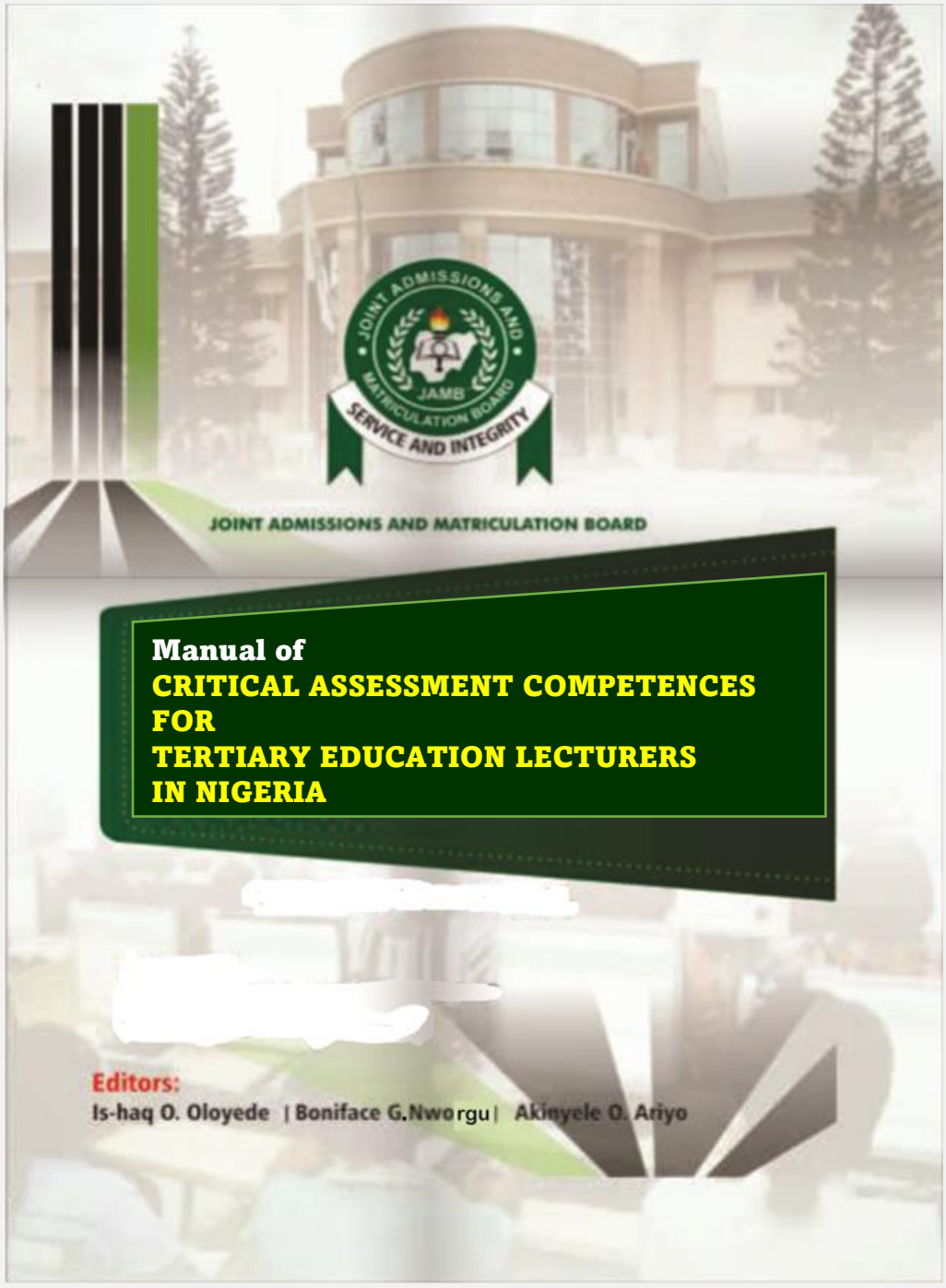

JOINT ADMISSIONS AND MATRICULATION BOARD
JAMB
SERVICE AND INTEGRITY
JOINT ADMISSIONS AND MATRICULATION BOARD
Manual of
CRITICAL ASSESSMENT COMPETENCES
FOR
TERTIARY EDUCATION LECTURERS
IN NIGERIA
Editors:
Is-haq O. Oloyede | Boniface G.Nworgu | Akinyele O. Ariyo

**ISBN: 978-978-797-935-8**
© First Edition 2022
Revised Edition 2023
Published by:
JOINT ADMISSIONS AND MATRICULATION BOARD
Tel: +2348036087987, +2348034292924
Email: jambxmetrics@gmail.com



## INTRODUCTION TO THE REVISED EDITION

Today, the field of psychometrics, measurement and evaluation is growing at a fast pace globally owing to its enormous value in solving educational assessment problems. My experience as a lecturer in the University for several years has exposed me to the fact that theories and techniques of psychological measurement, which include the measurement of knowledge, abilities, attitudes, and personality traits should be applied in all areas of assessment especially in the higher institutions of learning. Psychometrics deals with design and interpretation of tests that measure psychological values such as aptitude, ability, personality, memory, happiness, and intelligence. It is a well-known fact that there is inadequate knowledge of the theory and practice of educational assessment techniques among lecturers in higher institutions in Nigeria, which is rather unfortunate.

More saddening is the fact that agencies and institutions abound with mandates to engage in educational assessment but without any thought being given to their ability and competence based on any formal or even informal training in Assessment. Institutes, Academics and Professional bodies award membership, fellowship and diplomas purportedly based on educational/skill-assessment yet without any prior training for the educational assessors. No wonder that quite often, the validity and reliability of the test are not only suspect but laughable.

No examination in Nigeria is misunderstood by the populace as the Unified Tertiary Matriculation Examination (UTME) - a cohort-based norm-reference test, consistently being misconstrued as achievement rather than ranking test even by some educated elites.

This manual, which addresses this knowledge gap, targets squarely the critical assessment competences required by this all-important group of stakeholders in higher education. The manual is organised into ten inter-connected modules.

Module One deals with conceptual and policy issues in assessment, meaning, process, and types of assessment (assessment of, for, and as learning), national policy framework on assessment and implications. Module Two examines issues referring to the typology of educational measurement - different tests and what they measure. Module Three addresses developing credible types of examinations, curriculum/syllabus, objective levels, coverage (test blueprint), alignment of learning outcomes, activities, and assessment, while Module Four addresses techniques for constructing achievement tests, focusing on types of

questions/items (essay, objective, practical, oral, portfolio), and refining test items (item qualities).

In Module Five, issues of quality and confidence (validity, reliability, objectivity, fairness, and usability) are the focus, just as Module Six discusses test administration and scoring (environment/supervision, guiding against cheating, testing procedures - regular and CBT, scoring rubrics, marking scheme, moderation of scores/grades, etc). Module Seven deals with issues on reporting within the feedback process as well as the ethical issues involved in this aspect of assessment.

On the other hand, Module Eight focuses on understanding test scores in relation to measurement error and comparability of scores (standardisation of scores). Module Nine explores comprehensively the gamut of e-Testing, highlighting its aspects and dimensions and their strengths and drawbacks.

Finally, Module Ten prescribes reconciliation of testees with tests through application of alternative psychometric approaches/principles as a therapy for examination malpractices. It explains different types of alternative psychometric approaches and explores how they can effectively curb the menace of examination malpractices.

Two national workshops with the participants from tertiary institutions and a few secondary school teachers from all the zones of the federation were held at both the University of Ibadan and Manpower Development Institute, Dutse, Jigawa State to present and analyse the manual. This revised edition incorporates the lessons from the workshops. The workshops also convinced the Board to make the manual freely available to all by securing the removal of all copyright restrictions from the work. This has been achieved. In addition, not less than seventeen (17) credible national educational assessment establishments/bodies have, as a result of the inspiration from JAMB, and this manual, subscribed to the establishment of the Professional Assessment Academy of Nigeria for the training of the educational assessors. An independent team to midwife the establishment of the non-Governmental Academy is being constituted. This again, is an immediate outcome of this manual.

This Manual, which is a guide for lecturers in the higher educational institutions, particularly those without formal educational assessment training but find themselves setting and grading educational tests at various levels and other circumstantial educational assessors, is purposefully pedagogical to

engender critical insights into assessment processes and strategies for achieving validity, reliability, and fairness in testing. The manual will also serve as a reference model for lecturers who are not in the field of educational assessment but are eager to learn how to administer quality assessment to their students.

I, therefore, strongly recommend this manual to all lovers of quality educational assessment, with the hope that it would bridge some obvious gaps in contemporary educational assessment in Nigeria.

***Professor Is-haq O. Oloyede, CON, OFR, FNAL (Regd. Teacher)***
Registrar, Joint Admissions and Matriculation Board

## ACKNOWLEDGMENTS

The production of this manual could not have been possible without the support of the Honourable Minister of Education, Mallam Adamu Adamu, and the entire Board members of the Joint Admissions and Matriculation Board, under the leadership of Dr. Emmanuel Ndukwe.

Special appreciation goes to all the members of the Committee of Experts/ authors in Measurement and Evaluation constituted by the Board for their services and contributions to this manual. The members are Professor B. G. Nworgu, University of Nigeria, Nsukka, Professors Gbenga J. Adewale and Henry O. Owolabi of the Universities of Ibadan and Ilorin respectively. Others are Professors Monday Joshua (University of Calabar), C. A. Ugodulunwa (Alex Ekwueme Federal University, Ndufu-Alike, Ikwo); Muhammad I. Yakasai (Bayero University, Kano), Abubakar Hamman-Tukur (University of Maiduguri, Maiduguri), R. A. Lawal (University of Ilorin, Ilorin), and O. A. Afemikhe (University of Benin, Benin). They also include Dr. (Mrs) Christiana Agi of the Federal University of Agriculture, Makurdi.

Our sincere appreciation also goes to the Chairman of the Committee, Prof. B. G. Nworgu, University of Nigeria, Nsukka, who along with Professor Is-haq O. Oloyede and Dr. Akinyele O. Ariyo edited the manual.

Deserving of Board's gratitude is the Secretariat of the Committee ably headed by Dr. Akinyele O. Ariyo and supported by the following:

(a) Dr. Mrs. Bosede O. Amoo, MNIM
(b) Mrs. Chidinma B. Ihedoro
(c) Mr. Eniola Oyebanji
(d) Mrs. Ijeoma H. Onyekwere
(e) Mr. Suleiman Habeeb
(f) Mr. Qasim Egbewole
(g) Mr. Mubarak Oladipo Abisoye

*Professor Is-haq O. Oloyede, CON, OFR, FNAL*
Registrar, JAMB

## LIST OF CONTRIBUTORS

1. Professor B. G. Nworgu, University of Nigeria, Nsukka.
2. Professor O. A. Afemikhe, University of Benin, Benin.
3. Professor R. A. Lawal, University of Ilorin, Ilorin.
4. Professor Gbenga. J. Adewale, University of Ibadan, Ibadan.
5. Professor Muhammad I. Yakasai, Bayero University, Kano.
6. Professor Abubakar Hamman-Tukur, University of Maiduguri, Maiduguri.
7. Professor (Mrs.) C. A. Ugodulunwa, Alex Ekwueme Federal University, Ndufu Alike Ikwo.
8. Professor Henry. O. Owolabi, University of Ilorin, Ilorin.
9. Professor Monday Joshua, University of Calabar, Calabar.
10. Dr. (Mrs.) Christiana Agi, Federal University of Agriculture, Makurdi.
11. Professor Is-haq O. Oloyede, Registrar, Joint Admissions and Matriculation Board, Bwari-Abuja
12. Dr. Akinyele O. Ariyo, Joint Admissions and Matriculation Board, Bwari-Abuja.

## TABLE OF CONTENTS


**INTRODUCTION TO REVISED EDITION** iii
**ACKNOWLEDGMENTS** vi
**LIST OF CONTRIBUTORS** vii
**TABLE OF CONTENTS** viii

**MODULE ONE**
**CONCEPTUAL AND POLICY ISSUES IN ASSESSMENT** **1**
1.0 Module Overview 1
Unit 1.1: Concept of Assessment 1
1.1.1 Expected Learning Outcomes 1
1.1.2 Assessment 1
1.1.3 Purpose of Assessment 3
1.1.4 Assessment Techniques 4
Activity 1.1 6
1.1.5 ***Summary*** 6
Unit 1.2: Assessment of Learning 7
1.2.1 Expected Learning Outcomes 7
1.2.2 Concept of Assessment of Learning 7
1.2.3 Teachers' Role in Assessment of Learning 7
1.2.4 Planning Assessment of Learning 8
Why am I Assessing? 8
Activity 1.2 9
What am I Assessing? 9
What Assessment Method should I use? 10
How can I ensure Quality in this Assessment Process? 11
*Validity* 11
*Reliability* 11
*Reference Points* 12
*Record-Keeping* 12
How can I use the Information from this Assessment? 13
Activity 1.3 13
Feedback 13
Differentiating 13
Reporting 14
Activity 1.4 14
1.2.5 ***Summary*** 14
Unit 1.3: Assessment for Learning 15
1.3.1 Expected Learning Outcomes 15

1.3.2 Uses of Assessment for Learning 15
1.3.3 Teachers' Role in Assessment for Learning 16
1.3.4 Planning Assessment for Learning 16
Why am I Assessing? 16
What am I Assessing? 17
What Assessment Method should I use? 17
How can I ensure Quality in this Assessment Process? 18
*Validity* *18*
*Reliability* *18*
*Reference Points* *19*
*Record-Keeping* *19*
How can I use the Information from this Assessment? 20
*Feedback to Students* *20*
*Differentiating Learning* *20*
*Reporting* *20*
Activity 1.5 21
1.3.5 ***Summary*** 21
Unit 1.4: Assessment as Learning 21
1.4.1 Expected Learning Outcomes 21
1.4.2 Concept of Assessment as Learning 22
1.4.3 Teachers' Role in Assessment as Learning 24
1.4.4 Planning Assessment as Learning 25
What am I Assessing? 25
What Assessment Method should I use? 25
How can I ensure Quality in this Assessment Process? 26
*Validity* *26*
*Reliability* *26*
*Reference Points* *26*
*Record-Keeping* *27*
How can I use the Information from this Assessment? 27
*Feedback* *27*
*Differentiating Learning* *27*
*Reporting* *28*
Activity 1.6 28
1.4.5 ***Summary*** 28
Unit 1.5: Integration of Types of Assessment 29
1.5.1 Expected Learning Outcomes 29
1.5.2 Integration of Assessment Types 29
Activity 1.7 31
1.5.3 ***Summary*** 31
Unit 1.6: National Policy Assessment Framework 31

1.6.1 Expected Learning Outcomes 31
1.6.2 Provision for Assessment in National Policy on Education 31
1.6.3 Certification Issues in the National Policy on Education 32
Activity 1.8 33
1.6.4 ***Summary*** 33
References 34

**MODULE TWO**
**TYPOLOGY OF EDUCATIONAL MEASUREMENT 36**
2.0 Module Overview 36
Unit 2.1: Classification of Tests (Generally) 37
2.1.1 Introduction 37
2.1.2 Expected Learning Outcomes 37
2.1.3 Bases of Classification 38
Activity 2.1 41
2.1.4 ***Summary*** 41
Unit 2.2: Classification of Tests Based on What They Measure (Test Purpose) 42
2.2.1 Introduction 42
2.2.2 Expected Learning Outcomes 42
2.2.3 Kinds of Test Based on What They Measure 42
Activity 2.2 45
2.2.4 ***Summary*** 45
Unit 2.3: Classification of Tests Based on Method of Interpreting the Scores 45
2.3.1 Introduction 45
2.3.2 Expected Learning Outcomes 46
2.3.3 Test Classification Based on Methods of Interpreting the Scores 46
Activity 2.3 48
2.3.4 ***Summary*** 48
Unit 2.4: Classification of Tests Based on Use of the Results 49
2.4.1 Introduction 49
2.4.2 Expected Learning Outcomes 49
2.4.3 Test Types Based on Use of Test Results 49
Activity 2.4 52
2.4.4 ***Summary*** 52
Unit 2.5: Classification of Tests Based on Method of Preparation (Who Constructs?) 53
2.5.1 Introduction 53
2.5.2 Expected Learning Outcomes 53
2.5.3 Test Types Based on Method of Preparation (Who Constructs) 54

Activity 2.5 56
2.5.4 ***Summary*** 56
Unit 2.6: Methods of Assessment in the Affective and Psychomotor Domains 56
2.6.1 Introduction 56
2.6.2 Expected Learning Outcomes 57
2.6.3 Assessment of Learning Outcomes in the Affective Domains 57
2.6.4 Classification of Instrument for Assessing Non-Cognitive Learning Outcomes 61
2.6.5 Methods of Assessment in the Psychomotor Domain 61
Activity 2.6 62
2.6.6 ***Summary*** 62
Bibliography 63
References 63

**MODULE THREE**
**CONSTRUCTING GOOD ACHIEVEMENT TESTS** **65**
3.0 Module Overview 65
Unit 3.1: The Concept and Purpose of Achievement Test 65
3.1.0 Introduction 65
3.1.1 Expected Learning Outcomes 65
3.1.2 Concept of Achievement Test 65
Activity 3.1 66
3.1.3 Purpose of Achievement Test 67
Activity 3.2 68
3.1.4 ***Summary*** 68
Unit 3.2: Constructing Essay and Objective Test Items 68
3.2.0 Introduction 68
3.2.1 Expected Learning Outcomes 68
3.2.2 Types of Essay Tests 69
3.2.3 Types of Objective Tests 70
Activity 3.3 72
Activity 3.4 73
Activity 3.5 74
3.2.4 Constructing Essay and Objective Test Items 74
Activity 3.6 82
3.2.5 ***Summary*** 82
Unit 3.3: Designing Oral Assessment 83
3.3.0 Introduction 83
3.3.1 Expected Learning Outcomes 83
3.3.2 Meaning of Oral Assessment 83

3.3.3 Importance of Oral Assessment 84
Activity 3.7 85
3.3.4 How to Design a Good Oral Assessment 86
Activity 3.8 86
3.3.5 ***Summary*** 86
Unit 3.4: Designing a Practical Test 87
3.4.0 Introduction 87
3.4.1 Expected Learning Outcomes 87
3.4.2 Concept of Practical Test 87
3.4.3 Types of Practical Tests 89
Activity 3.9 90
3.4.4 How to Design a Simple Practical Test 90
3.4.5 Development of a Performance Rating Form 91
3.4.6 Guidelines for Writing Instructions for Practical Tests 91
3.4.7 ***Summary*** 92
Unit 3.5: Designing a Portfolio 92
3.5.0 Introduction 92
3.5.1 Expected Learning Outcomes . . . 91
3.5.2 Meaning and Purpose of Portfolio 93
3.5.3 Types of Portfolios 94
3.5.4 Procedure and Guidelines for Designing Portfolio 95
Activity 3.10 96
3.5.5 ***Summary*** 96
Unit 3.6: Refining Achievement Test Item Qualities 97
3.6.0 Introduction 97
3.6.1 Expected Learning Outcomes 97
3.6.2 Meaning of Item Vetting 97
3.6.3 Item Vetting Process 98
Activity 3.11 100
3.6.4 ***Summary*** 100
References 100

**MODULE FOUR**
**ISSUES OF QUALITY AND CONFIDENCE IN ASSESSMENT 103**
4.0 Module Overview 103
Unit 4.1: Validity 103
4.1.0 Introduction 103
4.1.1. Expected Learning Outcomes 104
4.1.2 Concept of Validity 104
4.1.3 Kinds of Validity 105
Activity 4.1 108

4.1.4 Factors Affecting Validity 110
4.1.5 How to Ensure the Validity of a Test 111
Activity 4.2 111
4.1.6 ***Summary*** 111
Unit 4.2: Reliability 112
4.2.0 Introduction 112
4.2.1 Expected Learning Outcomes 113
4.2.2 The Concept of Reliability 113
4.2.3 Types of Reliability Estimates 114
Activity 4.3 121
4.2.4 Factors affecting the Reliability of a Test 121
Activity 4:4 122
4.2.5 ***Summary*** 122
Activity 4.5 123
Unit 4.3: Objectivity 124
4.3.0 Introduction 124
4.3.1 Expected Learning Outcomes 124
4.3.2 Meaning of Objectivity 125
4.3.3 Importance of Test Objectivity 126
4.3.4 Factors Affecting Objectivity of a Test 127
4.3.5 Ways of Ensuring Objectivity of a Test 127
Activity 4.6 128
4.3.6 ***Summary*** 128
Unit 4:4: Fairness 128
4.4.0 Introduction 128
4.4.1 Expected Learning Outcomes 129
4.4.2 Meaning of Fairness 129
4.4.3 Factors that affect Fairness 130
4.4.4. Ways of Achieving Fairness in Educational Assessments 130
Activity 4.7 131
4.4.5 ***Summary*** 131
Unit 4.5: Test Usability 131
4.5.0 Introduction 131
4.5.1 Expected Learning Outcomes 132
4.5.2 Meaning of Usability 132
4.5.3 Factors That Determine Usability of a Test 132
4.5.4 Ways of Ensuring that a Test is Usable 133
Activity 4.8 134
4.5.5 ***Summary*** 134
References 134

**MODULE FIVE**
**TEST ADMINISTRATION AND SCORING** **136**
5.0 Module Overview 136
Unit 5.1: Testing Procedures 136
5.1.0 Introduction 137
5.1.1 Expected Learning Outcomes 137
5.1.2 Moderation of Tests 137
5.1.3 Types of Custodians 138
5.1.4 Conditionalities for Selecting Appropriate Custodian Points 139
5.1.5 Qualities and Functions of Examination Coordinators 140
5.1.6 Appropriate Channels of Test Distribution and Retrieval 140
5.1.7 Rights of the Examinees 142
5.1.8 Strategies for Combating Examination Malpractice 143
Activity 5.1 144
5.1.9 ***Summary*** 144
Unit 5.2: Scoring Rubrics 145
5.2.0 Introduction 145
5.2.1 Expected Learning Outcomes 145
5.2.2 Scoring Rubrics: What are they? 145
5.2.3 Types of Scoring Rubric 146
5.2.4 Functions of Scoring Rubrics 147
5.2.5 Scoring Rubric Structure 147
5.2.6 How to Create a Scoring Rubric 148
Step 1: Define Your Goal 148
Step 2: Choose a Rubric Type 148
Step 3: Determine Your Criteria 148
Step 4: Create Your Performance Levels 149
Step 5: Write Descriptors for Each Level of Your Rubric 149
Step 6: Revise Your Rubric 149
5.2.7 An Example of Scoring Rubrics for Analytic Dimension 149
Activity 5.2 150
5.2.8 Good Side of a Scoring Rubric 150
5.2.9 Challenges in the Use of Scoring 150
Activity 5.3 151
5.2.10 ***Summary*** 151
Unit 5.3: Marking Scheme 152
5.3.0 Introduction 152
5.3.1 Expected Learning Outcomes 152
5.3.2 Concept of Marking Scheme and its Uses 152
5.3.3 Designing Marking Schemes 153
5.3.4 How to Improve Marking of Essay Questions 155

Activity 5.4 155
5.3.5 ***Summary*** 156
Unit 5.4: Scoring Techniques 156
5.4.0 Introduction 156
5.4.1 Expected Learning Outcomes 156
5.4.2 Scoring Procedures 157
5.4.3 Criteria for Scoring Tests 159
5.4.4 Moderation of Scores 160
Activity 5.5 162
5.4.5 ***Summary*** 162
Bibliography 162
References 163

**MODULE SIX**
**UNDERSTANDING TEST SCORES** **164**
6.0 Module Overview 164
Unit 6.1: Nature of Psychological Tests, Meaning and Types of Test Scores 164
6.1.0 Introduction 164
6.1.1. Expected Learning Outcomes 165
6.1.2 What is a Psychological Test? 165
Activity 6.1 165
Activity 6.2 166
6. 1.3 Characteristics of a Psychological Test 167
6.1.4 Qualities of a Good Psychological Test 167
Activity 6.3 167
6.1.5 Broad Areas of Application 168
6.1.6 What is a Test Score? 168
6.1.7 Types of Test Score 168
6.1.8 ***Summary*** 170
Unit 6.2: Model of Test Score and Standard Error of Measurement 170
6.2.0 Introduction 170
6.2.1 Expected Learning Outcomes 171
6.2.2 Basic Model of a Test Score 171
6.2.3 Observed Score, True Score and Error Score 172
Activity 6.4 172
6.2.4 Reliability of Test Scores 172
6.2.5 Standard Error of Measurement (SEM) 174
6.2.6 Interpreting Test Scores using Score Band 174
6.2.7 ***Summary*** 175
Unit 6.3: Frameworks for Interpreting Test Scores 175

6.3.0 Introduction 175
6.3.1 Expected Learning Outcomes 176
6.3.2 Criterion-Referenced Interpretation 176
6.3.3 Norm-Referenced Interpretation 176
Activity 6.5 178
6.3.4 ***Summary*** 178
Unit 6.4: Comparing Scores from Different Tests (or Examinations) 178
6.4.0 Introduction 178
6.4.1 Expected Learning Outcomes 179
6.4.2 What is Score Standardisation and Why Standardise Scores? 179
6.4.3 Standard Scores 180
Activity 6.6 181
Activity 6.7 182
Activity 6.8 183
6.4.4 ***Summary*** 183
References 183

**MODULE SEVEN**
**REPORTING AND FEEDBACK IN ASSESSMENT 185**
7.0 Module Overview 185
Unit 7.1: Rationale for Assessment Reporting 185
7.1.0 Expected Learning Outcomes 185
7.1.1. Concept of Assessment 185
7.1.2 Concept of Reporting in Assessment 185
7.1.3 Rationale for Reporting 187
7.1.4 Benefits of Reporting 188
Activity 7.1 188
Activity 7.2 188
7.1.5 ***Summary*** 188
Unit 7.2: Examination Reporting in Tertiary Educational Institutions 189
7.2.1 Expected Learning Outcomes 189
7.2.2. Types of Reporting 189
7.2.3 Grade Point Average 190
7.2.3.1: Stages in Obtaining Students' Cumulative Grade Point Average 190
Activity 7.3 192
7.2.4 ***Summary*** 192
Unit 7.3: Feedback Functions of Reporting 192
7.3.1 Expected Learning Outcomes 192
7.3.2 Types of Report 192
Activity 7.4 194

7.3.3 Types of Feedback 194
Activity 7.5 195
Activity 7.6 195
7.3.4 ***Summary*** 196
Unit 7.4: Ethical Issues in Reporting 196
7.4.1 Expected Learning Outcomes 196
7.4.2 Ethical Issues and Educational Quality/Standards 196
Activity 7.7 198
7.4.3 ***Summary*** 198
References 198

**MODULE EIGHT**
**e-TESTING** **200**
8.0 Module Overview 200
Unit 8.1: Understanding e-Testing 200
8.1.0 Introduction 201
8.1.1 Expected Learning Outcomes 201
8.1.2 The Concept of e-Testing 201
Activity 8.1 202
8.1.3 Distinction between e-Testing and e-Assessment 202
8.1.4 Components of e-Testing 203
Activity 8.2 203
8.1.5 E-assessment and Online Assessment 204
Activity 8.3 204
8.1.6 Online Assessment of Learning and Assessment of Online Learning 204
Activity 8.4 205
8.1.7 ***Summary*** 205
Unit 8.2: Processes in e-Testing 205
8.2.0 Introduction 205
8.2.1 Expected Learning Outcomes 205
8.2.2. Cycle of e-Assessment 206
Activity 8.5 207
8.2.3 E-Testing System 207
8.2.4 Characteristics of e-Testing System 208
Activity 8.6 210
8.2.5 Support in e-Testing 210
Activity 8.7 211
8.2.6 ***Summary*** 211
Unit 8.3: Online Assessment Models 212
8.3.0 Introduction 212

8.3.1 Expected Learning Outcomes 212
8.3.2 Types of Online Assessment Delivery Modes 212
Activity 8.8 214
8.3.3. Online Assessment Delivery Models 215
Activity 8.9 221
8.3.4 ***Summary*** 221
Unit 8.4: Computer-Based Testing (CBT) 221
8.4.0 Introduction 221
8.4.1 Expected Learning Outcomes 222
8.4.2 Meaning of Computer-Based Test 222
8.4.3 Types of CBT 223
8.4.4 CBT Modes 223
Activity 8.10 224
8.4.5 Strengths of CBT 224
8.4.6 Challenges in e-Testing 225
Activity 8.11 226
8.4.7 Test Security in CBT 227
Activity 8.12 229
8.4.8 ***Summary*** 229
References 229

**MODULE NINE**
**CREDIBILITY ISSUES IN EXAMINATIONS 231**
9.0 Module Overview 231
Unit 9.1: Meaning of Credible Examination 231
9.1.0 Introduction 231
9.1.1 Expected Learning Outcomes 231
9.1.2 What is a Credible Examination? 232
9.1.3 Meaning of Impact, Curricular and Instructional Validity 232
9.1.4 Importance and Characteristics of Credible Examinations 233
Activity 9.1 Class Discussion 233
9.1.5 ***Summary*** 233
Unit 9.2: Credibility Issues in Preparation of Examinations 234
9.2.0 Introduction 234
9.2.1 Expected Learning Outcomes 234
9.2.2 Preparatory Measures for Credible Examination by Lecturers 234
9.2.3 Preparatory Measures for Credible Examinations by Institutions 235
Activity 9.2 Discuss in Groups and Report 236
9.2.4 ***Summary*** 236
Unit 9.3: Credibility Issues in Administration and Scoring 236
9.3.0 Introduction 236

9.3.1 Expected Learning Outcomes 236
9.3.2 Functions of the Invigilators 237
9.3.3 Forms of Examination Malpractice by Students 237
9.3.4 Ensuring Credibility of Examinations through Scoring 238
Activity 9.3: Group Discussion 239
Unit 9.4: Credibility Issues in Reporting of Examinations 239
9.4.0 Introduction 239
9.4.1 Expected Learning Outcomes 239
9.4.2 Importance of Accurate Reporting of Examination Scores 239
Activity 9.4 240
9.4.3 ***Summary*** 240
References 241

**MODULE TEN**
**RECONCILING TESTEES WITH THE TESTS: A THERAPY AGAINST EXAMINATION MALPRACTICE 242**
10.0 Module Overview 242
10.1 Psychometric Principles that can demystify Tests or Examinations 243
10.1.1 Integrated Approach to Assessment 243
10.1.2 Mainstreaming Tests into the Learning-Teaching Process 244
10.1.3 Designing Tests that cater for different abilities and individual differences 245
10.1.4 Democratisation of Test Development Process 245
10.1.5 Engaging Students as Markers and Drafters of the Marking Scheme 246
10.1.6 Using the Marking Scheme as Instructional Resource 247
10.1.7 Open Book Examinations (OBE) 248
10.2 ***Summary*** 249
References 250

**INDEX 252**

# MODULE ONE

# CONCEPTUAL AND POLICY ISSUES IN ASSESSMENT

**by**

**O. A. Afemikhe**

## 1.0 Module Overview

Do you know that in education, teaching and learning are very paramount? Without them the educational process would be a nullity. While the teacher facilitates learning, the learners are expected to learn. Learners and teachers, as part of the educational process, are involved in collecting information for decision-making about themselves, the curriculum, educational policies, among others. In this Module, we are going to look at how this is done, and the players involved in it. In addition, we will examine how the information collected is used as part of the teaching-learning process. Similarly, this Module shall focus on the concept of assessment, purposes of assessment, the assessment process and assessment types and how each can be implemented vis-a-vis the assessment policy in Nigeria.

## Unit 1.1: Concept of Assessment

### 1.1.1 Expected Learning Outcomes

At the end of this Unit, you are expected to be able to:

- Explain the concept of assessment
- Outline the purpose of assessment

### 1.1.2 Assessment

What is assessment? Can you remember the many decisions that you have to make as a lecturer? How did you go about making those decisions? Remember that the decisions are quite many, including what

to teach, how to teach, grading of students, and textbooks to recommend. To make these decisions, we require good quality information. The process of collecting information for decision making is what we refer to as assessment. There are different types of decisions relating to the learners, teachers, school, curriculum, educational policy and indeed everything about education. As teachers, we need to be familiar with the different processes of collecting information for decision making and their relevance to the decision to be made.

Assessment plays a very important role in education. Can you remember how teachers normally taught in Nigerian schools and applied assessment? Many of them believed that the purpose of classroom assessment was to provide evidence that the learners had mastered the content. Thus, they taught and tested the learners based on what was taught and went on to teach the next unit. This looked good as it met societal expectations. Societal expectations are changing, and goals and objectives of school are equally changing. Therefore, new approaches to assessment must be applied. In the early stages of education, acquisition of basic skills was seen as of interest to only a few. With mass education in vogue, and jobs being in short supply, would focusing on basic skills be enough for school learners? It will not be. This is because the present society requires that graduates be proficient in complex skills of critical thinking, problem-solving, collaboration, and effective communication, in addition to the technical skills, to meet demanding societal, economic, and technological challenges. Therefore, assessment must change in its purposes and strategies.

Our understanding of learning is also changing. It is no longer thought of as accumulation of knowledge that can be ordered in a sequence to be taught and reinforced. In learning, students are expected to construct understanding and attempt to connect new information to what they already know, so that coherence can be built into the knowledge base of the learner. We have to motivate learners to learn successfully. Thus, assessment should provide learners' evidence of progress and

achievement, rather than the failure and defeat associated with being compared to more successful peers (Stiggins, 2001). These changes in societal expectations should make us to rethink what we teach, what our students learn, and especially how we apply classroom assessment practices.

Can you think of the different ways of obtaining information for use by teachers? How do teachers use the information collected as part of the teaching-learning process? Different levels of assessment are discernible. We have assessments at the classroom level, assessment as part of external examinations, national assessments, and international assessments. In this module, our focus is on assessment conducted as part of classroom interactions. Do you know that classroom assessments have for a long time been classified as **formative** and **summative**? Formative assessment is collecting information for use as part of teaching and learning. This information is normally collected and used as feedback in ensuring that learners learn. The usual questions and assignments used during teaching fall in this category. While formative assessment involves collecting information to be used in driving teaching, summative assessment involves collection of information at the end of instruction to determine whether the teaching has been well executed.

### 1.1.3 Purpose of Assessment

Assessment as a process requires that we are familiar with how information is collected and used as part of classroom interactions. Classroom assessment is used for various purposes: assessment for learning, assessment as learning, and assessment of learning. These purposes have been linked to summative and formative assessment. The formative assessment has been further classified as **assessment for learning** and **assessment as learning**. Summative assessment is seen as assessment of learning. These different types would be examined in this module. The role of the teacher in each of these purposes is not the

same; teachers play different roles. The different purposes call for different planning, and different quality issues are normally raised. The most important part of assessment is the interpretation and use of the information that is gleaned for its intended purpose(s). Have you ever asked yourself questions you must answer in your assessment responsibilities? In planning, developing, and using assessment methods that are "fit for purpose," teachers ask themselves the following questions, as shown in Figure 1.1.

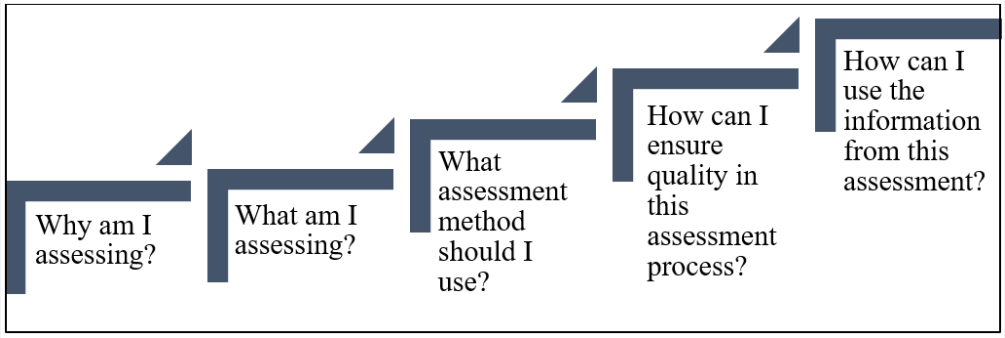



Figure 1.1: Assessment - Related Questions
(Adapted from Manitoba Education and Education, 2003)

These questions would now be used in developing Units, particularly with respect to the different types of assessment. The next unit will focus on assessment of learning.

### 1.1.4 Assessment Techniques

Do you know how information for assessments is collected? Table 1.1 contains some methods that we use in this endeavour (Afemikhe, 2014). We have provided different methods and have indicated a description of each. Look at them carefully and ask yourself how each is constructed. Try to rationalise how each can be used. We are not going to provide details here.

*Table 1.1: Assessment Methods*

| **Method** | **Description** |
|---|---|
| Questioning | Asking questions in class that are focused on eliciting information about the understanding displayed by the students. |
| Observation | Systematically looking at students as they process ideas to obtain information about their actions and behaviours. |
| Homework | Assignments given to students to execute at home to elicit their understanding of what was taught. |
| Learning Conversations or Interviews | Investigative discussions with students about their understanding and confusions, and how they are related to their previous knowledge. |
| Demonstrations, Presentations | This is provision of opportunities for students to show their learning in oral and media performances and exhibitions. |
| Quizzes, Tests, Examinations | Opportunities for students to manifest their learning through written response usually in internal or external examinations. |
| Rich assessment tasks | Involves complex tasks that encourage students to show connections that they are making among concepts they are learning. |
| Computer-Based Assessments | Systematic and adaptive software applications connected to curriculum outcomes using the computer and in some cases some internet-based facilities. |
| Simulations, Docudramas | Simulated or role-playing tasks that encourage students to show connections that they are making among concepts they are learning |
| Learning logs | With this, students make descriptions of the process they go through in their learning. This is very much |

| Method | Description |
|---|---|
| | like logbooks kept by students during industrial training. |
| Projects and Investigations | Opportunities for students to show connections in their learning through investigation and production of reports or artifacts. It normally involves use of research skills. |

***Activity 1.1***

1. Explain the concept of assessment. What are the information from it used for?
2. What changes in educational expectations have further given greater impetus to looking at the purposes of assessment?
3. Conduct a library search on learning logs and projects to show the richness of information that can emerge from their use by the classroom teacher.
4. Give practical examples of formative and summative assessments in your typical teaching activity.

### 1.1.5 Summary

1. Classroom assessment is used for various purposes: assessment for learning, assessment as learning, and assessment of learning.
2. Teachers ask the following questions in implementing assessments:
    a. Why am I assessing?
    b. What am I assessing?
    c. What assessment method should I use?
    d. How can I ensure quality in this assessment process?
    e. How can I use this information from the assessment?
3. Different types of assessment methods were examined.

## Unit 1.2: Assessment of Learning

### 1.2.1 Expected Learning Outcomes

At the end of this Unit, you are expected to be able to:

- explain the concept of assessment of learning;
- state teachers' role in assessment of learning;
- state why assessment of learning is being carried out;
- identify what is being assessed;
- specify assessment methods to be used, and
- indicate how the assessment information can be used.

### 1.2.2 Concept of Assessment of Learning

In assessment of learning, we focus on strategies designed to confirm what students have learnt and can demonstrate, to ascertain whether they have met curriculum/course outcomes or the goals of whatever they were being exposed to. Do you know that it also involves certifying the proficiency of the learners and making decisions about the students' future programmes or placements? Usually, it is designed to provide evidence of achievement by learners to parents, other teachers, the students themselves, and sometimes to outside groups (e.g., employers, other educational institutions). The information collected could be for institutional use or use by agencies external to the school. Assessment of learning is the assessment that indicates that learning has happened, and it is credible and defensible. This type of assessment includes the end-of-course examination, promotion examination and certification examination at the end of the programme. It falls within the purview of summative assessment.

### 1.2.3 Teachers' Role in Assessment of Learning

According to Earl and Katz (2006), the onus lies on the teacher to provide credible assessment of learning. This is because the consequences of assessment of learning are often far-reaching and affect

students greatly, and that is why they may be seen as high-stakes. Teachers have the responsibility of reporting student learning accurately and fairly, and this must be based on evidence obtained from a variety of contexts and applications. Therefore, for effective assessment of learning as teachers, you need to provide the following:

1. A rationale for undertaking a particular assessment of learning at a particular point in time;
2. A clear description of the intended learning;
3. A process that makes it possible for students to demonstrate their competence and skill;
4. Arrangement of alternative mechanisms for assessing the same outcomes;
5. Public and defensible reference points for making judgements;
6. Transparent approaches to interpretation;
7. Descriptions of the assessment process; and
8. Strategies for recourse in the event of disagreement about the decisions.

With your help as a teacher, your students can look forward to assessment of learning tasks as occasions to show their competence, as well as the depth and breadth of their learning. This would be motivational. Can you briefly rationalise why this is the case?

### 1.2.4 Planning Assessment of Learning

Remember the questions posed in Figure 1.1. We are going to apply them in planning the assessment of learning. Each of them will be closely examined.

***Why am I Assessing?***

The purpose of assessment of learning is to measure, certify, and report the level of students' learning, so that reasonable decisions can be made about them, the school, or even national performance of learners. It is

very important particularly for promotion in schools and at the end of programme examination.

***Activity 1.2***

| How can teachers utilise information from assessment of learning? |
|---|

You may need to think about this question and provide the answer.

- There are many potential users of the information derived from assessment of learning. Can you list some of them? Below are some users and what they may use the information for:
- Teachers can use the information to communicate with parents about their children's proficiency and progress apart from using it to take decisions about placement, among others;
- Parents can use the results for making educational and vocational decisions for their children;
- Students can use the information to size themselves up, and take informed educational and vocational decisions;
- Potential employers can use the information to make decisions about hiring or firing;
- Post-secondary institutions can use the information for selection and/or placement purposes;
- Principals, heads of departments, administrators, and teachers can use the information to review the curriculum offered in school.

***What am I Assessing?***

We have noted that assessment of learning focuses on proficiency of the learners. This proficiency is based on the curriculum. Therefore, in assessment of learning, the interest is in important curricular areas. We need to be aware of the nature and complexity of the intended learning and this has to be factored into the assessment. Thus, we have to collect

and interpret information about how well the learners have progressed in the important curricular areas. Realise that genuine learning for understanding is much more than just recognition or recall of facts or algorithms, and so assessment of learning tasks must be structured in a way that enables students to demonstrate the complexity of their understanding. Current thinking in the knowledge domain of the discipline are important as they provide the students with mileposts against which to apply key concepts, knowledge, skills, values, and attitudes in ways that are authentic. These fall in the three major domains of learning, namely the cognitive, affective and the psychomotor. You must be aware of these recent developments and focus your teaching on them, and consequently direct the assessment of learning to them too.

***What Assessment Method should I use?***

Can we really prescribe a method or methods for assessment of learning? Doing that would be simplistic because there can be variation in the intended curriculum outcomes and the continuum of learning that is required to reach the outcomes. To this extent, what is being assessed is an important consideration in this regard. Whatever methods you use must allow all students to show their understanding and produce sufficient information to support credible and defensible statements about the nature and quality of their learning, so that others can use the results in appropriate ways.

Assessment of learning uses methods that may include not only tests and examinations, but also a rich variety of products and demonstrations of learning - portfolios, exhibitions, performances, presentations, simulations, multimedia projects, and a variety of other written, oral, and visual methods. A combination of these methods can provide very dependable results if you use them well. Therefore, you should be conversant with different assessment methods from which you can draw on in your assessment responsibilities.

***How can I ensure Quality in this Assessment process?***

The quality of the assessment process is very important. For you to make quality decisions you have to conduct assessment of learning very carefully. This is to enable you collect information upon which decisions that are of the highest quality could be made. Since assessment of learning is normally summative, and for it to produce defensible and accurate descriptions of student competence in relation to defined outcomes and, occasionally, in relation to other students' assessment results, the quality of the assessments must be properly ensured. Your decisions on students' proficiency should be based on a rigorous, valid, reliable, and equitable process of assessment and evaluation.

*Validity*

Does your assessment activity measure exactly what you intended to measure, as reflected in the objectives of your instruction? As a teacher, you have to collect information or evidence based on the content taught. As statements about students' proficiency in wide areas of study would normally be made, assessment of learning tasks must reflect the key knowledge, concepts, skills, and dispositions set out in the curriculum. Whatever statements and inferences that emerge must be upheld by the evidence collected. This is how validity evidence is ensured.

*Reliability*

The reliability of the assessment of learning depends on how accurate, consistent, fair, and free from bias and distortion the assessment is. As teachers, you should attempt to:

1. obtain enough information about the learning of the particular student to make a definitive statement;
2. give all students an equal chance to show their learning with respect to the information collected;
3. ensure that another teacher will arrive at the same conclusion; and
4. ensure that any other teacher will make the same decision based on the information at any other time.

*Reference Points*

An important question in assessment of learning is how to interpret the assessment information collected? You need to be familiar with the reference points for assessment of learning, which are the learning outcomes as identified in the curriculum that make up the course of study. Therefore, you as the teacher must include assessment tasks that measure these learning outcomes that are of interest. Thereafter, a student's performance is interpreted and reported in relation to these learning outcomes. In some situations, where selection decisions need to be made for limited positions (e.g., university entrance, scholarships, employment opportunities), assessment of learning results are used to rank students. In such situations, what is being measured needs to be clear, and the way it is being measured needs to be transparent to anyone who might use the assessment results. This is to build confidence into the assessments.

*Record-Keeping*

As a teacher, you can use any strategy for assessment of learning. The records, however, provide details about the quality of the measurement. You should provide detailed records of the various components of the assessment of learning, clearly indicating a description of what each component measures, with what accuracy and against what criteria and reference points; and should include supporting evidence related to the outcomes as a justification.

With detailed records kept, you are in an excellent position to provide meaningful reports to parents and others. Try to avoid a symbolic representation of a student's accomplishments (e.g., a letter grade or percentage) as this may be inadequate. Your reports to parents and others should identify the intended learning that the report covers, the assessment methods used to gather the supporting information, and the criteria used to make whatever judgment that was made.

***How can I use the information from this assessment?***

***Activity 1.3***

> What can assessment of learning information be used for by the teachers? Think about this question and write out your thoughts on a sheet of paper.

We now present two important uses of assessment of learning information by the teacher.

*Feedback to Students*

Feedback is very important and must be provided by the teacher in a timely manner. If timely information is provided to students on their marks through comments that indicate their level of success, it can help them to make decisions about their future learning efforts. In this way, proficiency can be enhanced.

*Differentiating Learning*

In assessment of learning, there should be differentiation in assessment itself. You need to know the students well and be able to utilize assessments that take into considerations the students' physical characteristics. Through differentiation in assessment of learning you have to put in place the necessary accommodations that allow students to make the particular learning of interest visible. This you can do through multiple forms of assessment as this offers multiple pathways for making student learning transparent to the teacher. You should realise that curriculum outcomes, might be demonstrated in different ways such as visual, oral, dramatic, or written representations. When writing is not an explicit component of the outcome, students who have difficulties with written language, for example, would then have the same opportunity to demonstrate their learning as other students.

You may not use assessment of learning to differentiate instruction or resources; it, however, has a profound effect on the placement and

promotion of students. In this way, it would have effect on the nature and differentiation of the future instruction and programming that students receive. You must, therefore, ensure that assessment results are accurate and detailed enough to allow for plausible recommendations.

*Reporting*

How do you report students' proficiency? This can be done in varied ways. The audience is an important consideration, and the teacher should provide all of the information necessary for them to make reasoned decisions. Whatever form of reporting is applied, it should be honest, fair, and provide sufficient details so that it can be clearly understood. Traditional reporting, which relies only on a student's average score, provides little information about that student's skill development or knowledge. You can introduce a parent-student-teacher conference or forum. Through this forum, you are able to provide to parents a great deal of information, and consequently reinforce students' responsibility for their learning.

***Activity 1.4***

1. Explain the role of the teacher in assessment of learning?
2. Brainstorm a situation in which assessment for learning is the only possibility.
3. Indicate which assessment methods can be used in collecting the relevant information for decision making.
4. How can assessment of learning information be used?

### 1.2.5 Summary

1. Assessment of learning involves finding out what the students' competences are; this is used for selection and certification.
2. The teacher plays a very important role in it.

3. Information can be gathered in different ways and can be used to provide feedback to the students, parents and other stakeholders for decision making.

## Unit 1.3: Assessment for Learning

### 1.3.1 Expected Learning Outcomes

At the end of this Unit, you are expected to be able to:

- explain the concept of assessment of learning;
- state teachers' role in assessment for learning;
- state why assessment for learning is being carried out;
- identify what is being assessed;
- specify assessment methods to be used; and
- indicate how the assessment information can be used.

### 1.3.2 Uses of Assessment for Learning

Assessment for learning is intended to provide information to enhance learning when the teaching-learning process is still going on. It can take place throughout the learning process. You use it to make each student's understanding visible. With the visibility, you as the teacher can decide what can be done to help students progress in their studies. Your students will learn in individual and idiosyncratic ways, yet, at the same time, there are predictable patterns of connections and preconceptions that some of the students may experience as they move along. In assessment for learning, the teachers use assessment as an investigative tool to find out as much as they can about what their students know and can do, and what confusions, preconceptions, or gaps they might have.

Through the wide variety of information that you collect about your students' learning processes, you obtain a basis for determining what they need to do next to improve on their learning. Assessment for learning provides the basis for providing descriptive feedback for

students and deciding on groupings, instructional strategies, and resources that further enhance their learning.

### 1.3.3 Teachers' Role in Assessment for Learning

Remember that assessment for learning occurs throughout the learning process. It is interactive, and the role of the teachers, according to Earl and Katz (2006), include:

1. aligning instruction with the targeted outcomes;
2. identifying particular learning needs of students or groups;
3. selecting and adapting materials and resources;
4. creating differentiated teaching strategies and learning opportunities for helping individual students move forward in their learning; and
5. providing immediate feedback and direction to students.

You can also use assessment for learning to enhance students' motivation and commitment to learning. Your commitment to learning as the focus of assessment can change the classroom culture to one that promotes student success. It makes very visible what students believe to be true, and you can use that information to help students who are not progressing well to move forward in manageable, efficient, and respectful ways.

### 1.3.4 Planning Assessment for Learning

***Why am I Assessing?***

The intent of assessment for learning is to enhance student learning. Teachers would, therefore, use assessment for learning to uncover what students believe to be true and to learn more about the connection students are making, their prior knowledge, preconceptions, gaps, and learning styles. With this information, you can structure and differentiate instruction and learning opportunities to reinforce and build on productive learning and to challenge beliefs or ideas that are creating

problems or inhibiting the next stage of learning. You have to use this information to provide your students with descriptive feedback that will further their learning progress.

***What am I Assessing?***

The curriculum specifications serve as your starting point in deciding what to assess, and to focus on why and how students gain their understanding. Through assessment for learning you carry out assessment of the curriculum outcomes that comprise the intended learning. You should create assessments that will expose students' thinking and skills in relation to the intended learning, and the common preconceptions. In this way, understanding by the students is enhanced.

***What assessment method should I use?***

Whatever assessment method you are to use should give information that will be useful for planning and executing your teaching. You can use:

a. Focused observations
b. Questioning
c. Conversations
d. Quizzes
e. Computer-based assessments
f. Learning logs, or
g. Whatever other methods would provide the needed information.

Therefore, each time you plan an assessment for learning, you have to think about what information the assessment is designed to expose and must decide which assessment approaches are most likely to give detailed information about what each student is thinking and learning.

You have to use methods that incorporate a variety of ways for students to demonstrate their learning. For example, opportunities for students to complete tasks orally or through visual representation are important for those who are struggling with reading or learning English.

***How can I ensure Quality in this Assessment process?***

When is assessment for learning of high quality? This, you must consider taking into considerations what the assessment information is to be used for. You should be able to use it to make decisions about students' learning with enough specificity to be able to provide descriptive feedback, and to design the next stage of learning. These requirements you have to look out for.

*Validity*

Assessment for learning is valid if it provides information that can shed light on students' understanding of the ideas that are contained in the learning outcomes, and in the effectiveness of the choices and the guidance that the teacher provides for the next stage of learning. You can judge the validity of your assessment processes by monitoring how well the assessment shows the progress of students' learning along the continuum of the curriculum.

*Reliability*

Your assessment for learning should focus on the nature of students' thinking and learning at any given point in time; and since it is used to determine the next phase of teaching and learning, reliability depends on the accuracy and consistency of teachers' descriptions of the learning. You need to be sure that you are actually getting a clear picture of how the students are thinking and what it is that they understand or find confusing. Thus, a single assessment may not be sufficient to produce detailed insights into students' learning. You have to use a range of assessments in different modes (e.g., oral, visual, active, written), and do them at different times to develop an unfolding picture of the student's progress and development. You should continuously look for evidence and descriptions of each student's way of understanding the concepts.

Working with other teachers is one of the best ways for you to gain reliable insights into how students are thinking. In this way, you share your views about students' work and the nature and quality of the

learning in relation to curriculum outcomes. Consequently, you and the other teachers become consistent and coherent in your descriptive accounts of the students, and they can feel more confident about the final decisions and next steps in teaching.

*Reference Points*

The reference point for you in assessment for learning is the curriculum learning outcomes generally. However, for some students, learning outcomes of an individualised learning plan are the reference points for assessment for learning. These should serve as guides in providing feedback and in planning instruction. You have to specify the learning expectations or objectives that are clear and detailed, with exemplars and criteria that differentiate the quality and the changes along the learning continuum. Through these, you are enabled to accurately consider each student's work in relation to these expectations.

*Record-Keeping*

Record-keeping is an important part of ensuring quality in assessment for learning and you should consider in all the situations that you conduct assessment. You have to keep detailed notes, not for making comparative judgments among the students, but to provide each student with individualised descriptive feedback that will further help student's learning. If you keep good records, you would be able to show whether their work is on track. When it is not, you would have to raise questions about the instruction and ways it could be adjusted. The records you keep in assessment for learning is on documenting individual student's learning and annotating it in relation to the continuum of learning. Through it, you would be able to identify groups of students with similar learning patterns so that instruction can be efficiently differentiated. The records that you keep need to be based on the curriculum learning outcomes, and you have to give detailed accounts of students' accomplishments in relation to these outcomes, with evidence to support these accounts.

***How can I use the information from this assessment?***

As we did in Unit 1.2, we will present two uses of assessment for learning information.

*Feedback to Students*

The feedback that you provide has to be descriptive if it is to be successful. For your students to learn from your assessment, you have to provide specific, detailed feedback, and direction to each student to guide his or her learning. Feedback for learning is part of the teaching process. It comes after the initial instruction takes place, when information is provided about the way that the student has processed and interpreted the original material. It serves as a vital link between the teacher's assessments of a student's learning and the action following that assessment. To be successful, feedback needs to be immediate and should identify the way forward. The feedback that you provide should not simply tell learners whether their answers are right or wrong, or simply provide evaluative feedback in the form of grades and short, non-specific comments of praise or censure. It should give specific direction on the way forward.

*Differentiating Learning*

Through assessment for learning, you should be able to provide information about what students already know and can do, so that, as the teacher, you can design the most appropriate next steps in instruction. The focus on assessment for learning is continually making comparisons between the curriculum expectations and the continuum of learning for individual students, adjusting their instruction, grouping practices, and resources. In this way, each student can then receive the material, support, and guidance that he or she needs to progress, without experiencing unnecessary confusion and frustration.

*Reporting*

Your reporting in assessment for learning is based on open, frequent, and ongoing communication with students about progress in learning. The

methods that the teacher uses to ensure ongoing progress must be specified, and ways that students, teachers, and parents might help move learning forward with minimal misunderstanding and confusion for the student must be clear. The reports might focus on a single outcome, but more often, on a series or cluster of outcomes. Reporting should take into account what learning is expected, provide good models of what students can achieve, identify strategies for supporting students, and checking whether the gaps in learning have been closed.

***Activity 1.5***

1. Outline the main features of assessment for learning.
2. Explain the specific roles of the teacher and the learner.
3. Show specific and practical ways of how the information from assessment for learning can be utilised in enhancing learning.

### 1.3.5 Summary

1. Assessment for learning involves collecting information for use by the teacher to direct instruction.
2. It provides the basis for providing descriptive feedback for students and deciding on groupings, instructional strategies, and resources.
3. The focus of assessment for learning is learning to uncover what students believe to be true and to learn more about the connection students are making, their prior knowledge, preconceptions, gaps, and learning styles.

## Unit 1.4: Assessment as Learning

### 1.4.1 Expected Learning Outcomes

At the end of this Unit, you are expected to be able to:

- explain the concept of assessment as learning;
- state teachers' role in it;
- explain why the assessment is being carried out,

- state what is being assessed;
- specify assessment methods to be used; and
- indicate how the assessment information can be used.

### 1.4.2 Concept of Assessment as Learning

In assessment as learning, you focus on what students would do as a result of assessment and emphasise assessment as a process of metacognition (knowledge of one's own thought processes) for students. With assessment as learning, it is believed that learning is not just a matter of transferring ideas from someone who is knowledgeable to someone who is not. Rather, it is seen as an active process of cognitive restructuring that occurs when individuals interact with new ideas. Within this view of assessment as learning, students are the critical connectors between assessment and learning. Do you know that the students have to be actively engaged in creating their own understanding and therefore, they must learn to be critical assessors who have to make sense of assessment information, and relate it to prior knowledge, and use it for new learning? This is the regulatory process in metacognition as indicated by Earl and Katz (2006). That is, students become adept at personally monitoring what they are learning and use what they discover from the monitoring to making adjustments, adaptations, and even major changes in their thinking. Assessment as learning has a warrant as it is based on research about how learning happens and is characterised by students reflecting on their own learning and adjusting so that they achieve deeper understanding.

What is metacognition? We need to look at this briefly as it plays a very important role in assessment as learning. Metacognition is a process of reflecting on one's thinking as a human being. Costa (1983) describes metacognitive thinking as an internal continuous cycle of students' feedback and adjustment, conversation which involves monitoring their own understanding, predicting their performance, deciding what else

they need to know, organising and reorganising ideas, checking for consistency among different pieces of information, and drawing analogies that help them advance their learning. Through it, students become more efficient and begin to internalise the process of standing outside their own learning and considering it against a range of criteria, not just the teacher's judgement about quality or accuracy. Through it, students are able to monitor their learning and along the way, make corrections, and develop a habit of mind for continually reviewing and challenging what they know. The dimensions of metacognition, according to Brown (1987), involve:

1. Knowledge of Cognition: This covers
    - learners knowing themselves and the factors that influence their performance;
    - learning strategies which they can apply in their learning; and
    - being able to clearly indicate when and why they are to use a particular strategy.
2. Regulation of Cognition: This includes
    - Planning: setting goals and activating relevant background knowledge;
    - Regulation: monitoring and self-assessment; and
    - Evaluation: appraising the products and regulatory processes of learning.

Your students become productive learners when they see that the results of their work are part of critical and constructive decision-making. You should encourage your students to internalise the ability of needing-to-know and challenging-of-assumptions as habits of mind. This can assist in achieving the ultimate goal in assessment as students acquire the skills and the habits of mind to be cognitively aware of increasing independence. In assessment as learning, you have to focus on the explicit students' capacity over time to be their own best assessors. Your

role in this regard begins with presenting and modelling external, structured opportunities for students to assess themselves.

### 1.4.3 Teachers' Role in Assessment as Learning

We have indicated that learners occupy a central position in assessment, as learning does not mean that the teachers' responsibilities are diminished. As a teacher, you will have to design the teaching and assessment in such a way that allows all students to think about and monitor their own learning. This is because learners are capable of becoming adaptable and possess an appreciation of what high-quality work is. The learners can develop with the teachers' assistance, the evaluative skills necessary for them to compare with some level of objectivity, the quality of what they are learning. To become independent learners, students must progressively develop sophisticated combinations of skills, attitudes, and dispositions such as self-monitoring and evaluation. To become metacognitively aware requires modelling and teaching on the part of the teacher, and practice on the part of students.

Your role in promoting the development of independent learners is achieved through modelling and teaching the skills of self-assessment which will guide your students in setting goals and monitoring their progress. This you can achieve by providing examples and models of good practice, and quality work with which students can develop clear criteria of good practice. You can also guide them in developing internal feedback or self-monitoring mechanisms to validate and question their own thinking. To surmount the ambiguity and uncertainty that is inevitable in learning, the teacher has to provide regular and challenging opportunities to practice the skills of self-assessment. Your duties extend to monitoring students' metacognitive processes and their learning, as well as creating an environment where it is safe for students to take chances.

### 1.4.4. Planning Assessment as Learning

***What am I Assessing?***

Assessment as learning requires rich and detailed information about how students are progressing in developing the habits of mind and skills to monitor, challenge, and adjust their own learning. The teacher has to facilitate this. It is expected that students have to apply these skills. To this end, they have to learn to monitor and challenge their own understanding, predict the outcomes of their current level of understanding, make reasoned decisions about their progress and difficulties, decide what else they need to know, organise, and reorganise ideas. In addition, your learners should check for consistency among different pieces of information, draw analogies that help them advance their understanding, and set personal goals. The consequence is that the teachers are interested in how students understand concepts, and in how they use metacognitive analysis to make adjustments to their understanding. You should monitor students' goal-setting process and their thinking about their learning, and the strategies they use to support or challenge, adjust, and advance their learning.

***What Assessment Method should I use?***

Which assessment methods should you use in implementing assessment as learning? The focus should be on methods that elicit detailed information both about students' learning and about their metacognitive processes. Hence, the teacher should use a range of methods in assessment as learning.

Even though many assessment methods have the potential to encourage reflection and review, what matters in assessment as learning is that the methods allow students to consider their own learning in relation to models, exemplars, criteria, rubrics, frameworks, and checklists that provide images of successful learning.

***How can I ensure Quality in this Assessment Process?***

You should establish high quality assessment process by ensuring that students have the right tools and are accumulating the evidence needed to make reasonable decisions about what it is that they understand or find confusing, and what else they need to do to deepen their understanding. This is because quality in assessment as learning depends on how well the assessment engages students in considering and challenging their thinking, and in making judgements about their views and understanding.

*Validity*

What is proficient learning? Your students need to know this and the various steps that they need to take to attain the desired expertise. Thus, provide them with clear criteria and varied examples of what good work looks like, as well as opportunities to compare teacher feedback and advice about what to do next.

*Reliability*

What does reliability connote in assessment as learning? Do you know it is related to consistency and confidence in students' self-reflection, self-monitoring, and self-adjustment? You should assist the students to develop the skills to make consistent and reliable interpretations of their learning. This you can do through engaging students in the metacognitive processes. You need to scaffold students' understanding; providing criteria, exemplars, and resources to help them analyse their own work. The students should be taught the necessary skills to think about their own learning in relation to their prior understanding and the curricular learning outcomes; as well as how to gather evidence about how well they are learning.

*Reference Points*

For purposes of interpretation, look at a blend of curricular expectations and each student’s understanding at an earlier point in time. Help the

students to internalise the ability to compare their own learning over time with descriptions and examples of expected learning.

*Record-Keeping*

Who are the key players here? From our discussion so far, students are the key players in all the components of assessment as learning. They should, therefore, develop skills and attitudes that allow them to keep systematic records of their learning. The record should focus on reflections and insights as they occur. If well kept, their individual records become the evidence of their progress in learning and in becoming independent learners.

***How can I use the information from this Assessment?***

*Feedback*

On the part of the students, they use assessment as learning to gain knowledge about their progress, show milestones of success that are worthy of celebration, adjust their goals, make choices about what they need to do next to move their learning forward, and advocate for themselves. As a teacher, you have to provide guidance and direction for students to develop independent learning through detailed and relevant feedback. When students encounter new information, they filter it through their existing beliefs and ideas and those of their community and culture. They compare the new information to the beliefs and ideas held by the people around them. Hence, peers, family, and community members are also important players.

*Differentiating Learning*

Assessment as learning provides the conditions under which students and teachers can discuss what the students are learning. What it means to do it well, what the alternatives might be for each student to advance his or her learning, what personal goals have been reached, and what more challenging goals can be set are important here.

*Reporting*

Reporting in assessment as learning is the responsibility of students, who must learn to articulate and defend the nature and quality of their learning. When students reflect on their own learning and must communicate it to others, they are intensifying their understanding about a topic, their own learning strengths, and the areas in which they need to develop further. The students need to have been deeply involved in assessment as learning throughout the instructional process and be able to provide their parents with evidence of their learning. The evidence needs to include an analysis of their learning progress and what they need to do to move it forward.

***Activity 1.6***

1. Generate a situation in which assessment as learning can be used in the classroom and indicate the role of the students and of the teachers in it.
2. What are the activities of students and teachers in assessment as learning?

### 1.4.5 Summary

1. Assessment as learning focuses on students and emphasises assessment as a process of metacognition (knowledge of one's own thought processes) for students.
2. Metacognition involves knowledge and regulation of cognition.
3. In assessment as learning, the role of teachers includes designing instruction and assessment that allow all students to think about and monitor their own learning.
4. Reporting is the responsibility of the students.

## Unit 1.5: Integration of Types of Assessment

### 1.5.1 Expected Learning Outcomes

At the end of this unit, you are expected to be able to:

- explain the concept of integration of assessment types;
- rationalise the type of assessment to be given greatest priority; and
- specify benefits of integration of assessment types.

### 1.5.2 Integration of Assessment Types

We have classified assessment into three types. Do you think this implies that we use them independently in teaching and learning?

The answer is definitely no!

What is of importance in all assessments is students acquiring the desired knowledge, skills, and values, and being able to use them as part of their life opportunities. Can you reflect on what happens in your institution with respect to assessment? You will notice that most teachers use assessment of learning as they hardly use the assessment information to drive teaching and learning. You could reflect on what benefits would accrue if all the three types of assessment are used in the classroom. The ability to do this could be an innovation worthy of pursuit.

Notwithstanding the weaknesses of the different approaches, using them in combination and in synergy can be very beneficial as the limitations inherent in the different types can be overcome through their combined use.

Afemikhe (2019) indicated that assessment of learning is seen at various stages of teaching and learning process, but its use has not been too systematic in many classrooms; the use has been informal and implicit. As aptly indicated by Earl and Katz (2006) and Earl (2013), integration of all assessment types can be used to meet the demands of an information and knowledge-based economy, which requires students to

work towards higher order thinking, autonomy, and self-management. A range of formative and summative assessment approaches that complement each other, provide an adequate level of challenge, and support to each student is, therefore desirable.

An integrated approach to the use of assessment proposed by Siarova, Sternadel, & Masidlauskaite, (2017)

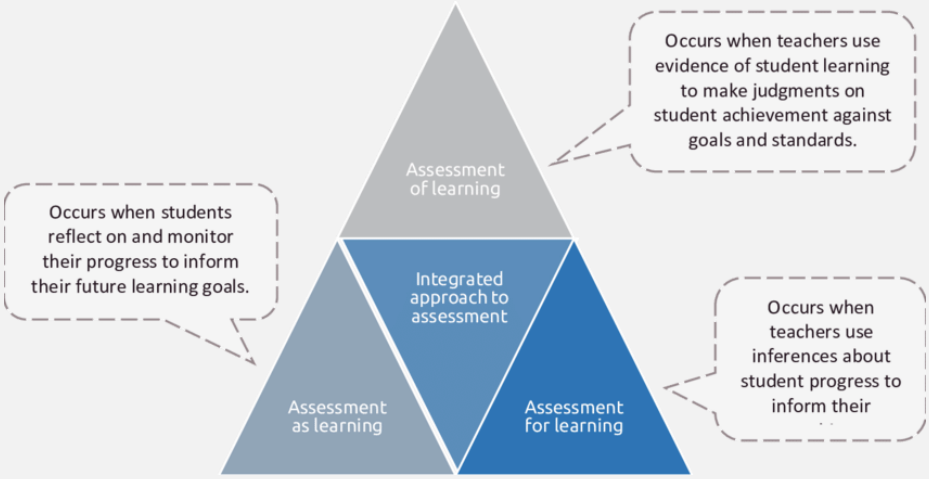



***Figure 1.2: Integrated Vision of Assessment***
***Source: Siarova et al., (2017)***

Learners with high learning needs benefit from a range of different teaching approaches, which utilise diverse assessment approaches. Through this, there is a potential to better inform and summarise the learning, as well as capture the multi-dimensionality of the learning process, as concluded by Bourke and Mentis (2014). Integration of assessment is relevant for linking assessments to specific functions with prepositions 'of, 'for' and 'as', without considering their integrated nature, may divert attention away from the key construct that 'assessment is learning' (Siarova, Sternadel, & Masidlauskaite, 2017).

As teachers, we should attempt to use all the three types by giving assessment the highest learning priority, followed by assessment for

learning and finally assessment of learning. In this way, the students will develop independence early, hence the burden on the teacher becomes reduced.

***Activity 1.7***

1. How can a teacher combine assessment information from assessment of learning, assessment as learning and assessment for learning in the classroom?

### 1.5.3 Summary

1. Using integrated assessment types can be very beneficial as the limitations inherent in the different types can be overcome through their combined use.
2. A range of formative and summative assessment approaches that complement each other provides an adequate level of challenge and support to each student.

## Unit 1.6: National Policy Assessment Framework

### 1.6.1 Expected Learning Outcomes

At the end of this Unit, you are expected to be able to:

- identify the provisions for assessment in the National Policy on Education;
- explain certification issues in the National Policy on Education;

### 1.6.2 Provisions for Assessment in National Policy on Education

In this section, we will look at the current policy as contained in the National Policy on Education (Federal Republic of Nigeria, 2013). Earlier policies indicated that primary schools were to issue testimonials on completion of primary level education. The present policy is silent on this. As part of the Basic Education section of the policy, it is clearly indicated that Continuous Assessment (CA) and school examinations

would be distributed in the ratio 40:60. Do you know what CA is? Just a quick reminder, CA is the teachers' collection of information for use in updating the judgment of the learners under his/her care. Which type of assessment is this, based on what we have discussed so far? It can either be assessment as learning or assessment for learning. This is because they are designed to ensure that learning is effectively executed. The school examination, on the other hand, can be seen to be assessment of learning, particularly if used for purposes of promotion to the next class.

Do you know at what point that certificates are to be awarded in formal education in Nigeria? The National Policy has clearly indicated that this should be on completion of 9-year Basic Education. The certificate to be awarded is the Basic Education Certificate Examination (BECE or Junior School Arabic and Islamic Certificate Education (JAISC). These certificates are to be awarded by State Governments and Federal Capital Territory. Experience shows that CA is used as a component of the scores for the award of the certificates. The Basic Education Certificate Examination (BECE) and Junior School Arabic and Islamic Certificate Examination (JAISCE) are indeed types of assessment of learning.

### 1.6.3 Certification Issues in the National Policy on Education

The National Policy on Education has been mentioned in this document. Do you have a copy? You need to own one and read it so as to familiarise yourself with other issues apart from assessment that we are now discussing. In section 42 of the policy, certification issues after completion of Senior Secondary School are outlined to include:

a. Certificates to be awarded at the end of secondary education: West African Senior School Certificate (WASSC), Senior School Certificate (SSC), National Technical Certificate (NTC), National Business Certificate (NBC), National Vocational Certificate (NVC), Senior Arabic and Islamic Studies Certificate (SAISC), etc. All these are based on continuous assessment and a national examination.

b. Statutory public examination bodies shall be responsible for conducting Post-Basic Education Certificate Examinations (PBECE) in order to ensure uniform standards at this level.
c. Examination ethics shall be strictly adhered to in the conduct of all examinations.

In Section 43, it is clearly stated that transition from post-basic education level to tertiary education shall be through the appropriate selection mechanisms. These examinations are examples of assessment of learning as through them certificates are awarded, or selection made. Note that the examination bodies that conduct these examinations also require schools to submit Continuous Assessment (C.A.) scores.

Can you reflect on how assessment is carried out at the tertiary levels? Take a notebook and jot down what assessment practices are in vogue.

Note that as with other levels of education, C.A. also constitutes part of end of semester examination. The examinations are cumulatively used for certification. To this extent, they are assessment of learning. The C.A.s that are used to ensure that the learners learn are, therefore, seen as assessment for learning, or assessment as learning, which we had earlier categorised as formative assessment.

***Activity 1.8***

1. Read the National Policy on Education and identify sections in which assessment-related issues are presented.
2. What are inherent weaknesses in the implementation of CA in your institution?

### 1.6.4. Summary

1. Assessment is part of the issues presented in the National Policy on Education.

2. These issues include assessment for learning, assessment of learning, and assessment as learning.
3. Continuous Assessment is to be combined with examination scores in the ratio of 40:60 for only Junior Secondary Schools. The proportion varies at other levels based on the accreditation agency.

**References**


Afemikhe, O. A. (2004). Introduction to Education Measurement and Evaluation. Lagos: Amfitops Books.

Afemikhe, O. A. (2019). *Innovations in Learning-Oriented Assessment Practices*. Keynote address presented at the AEAA conference in Abuja.

Bourke, R., & Mentis, M. (2014). An assessment framework for inclusive education: Integrating assessment approaches. *Assessment in Education: Principles, Policy, and Practice*, 21(4), 384-397.

Brown, A. (1987). Metacognition, executive control, self-regulation, and other more mysterious mechanisms. In F. Weinert & R. Kluwe (Eds), *Metacognition, motivation, and understanding*. Mahwah, NJ: Erlbaum.

Crossouard, B. (2011). Using formative assessment to support complex learning in conditions of social adversity. *Assessment in Education: Principles Policy and Practice* (1), 59-12.

Earl, L. M. (2013). Assessment as learning: Using classroom assessment to maximize student learning. Thousand Oaks: Corwin Press.

Earl, L. M., Katz, S. (2006). Rethinking classroom assessment with Purpose in mind. Retrieved from https://digitalcollection.gov.mb.ca/awweb/pdfopener?smd=1&did=12503&md=1

Federal Republic of Nigeria (2013). *National Policy on Education*.

Lagos: Nigerian Educational Research and Development Council.

Manitoba Education, Citizenship and Youth (2003). *Independent Together: Supporting the Multilevel Learning Community*. Winnipeg, MB: Manitoba Education, Citizenship and Youth.

Stiggins, R.J. (2001). *Leadership for Excellence in Assessment: A powerful new school district planning guide*. Portland. OR: Assessment Training Institute.

Siarova, H., Sternadel, D., Masidlauskaite, R. (2017). Assessment practices for 21st century learning: review of evidence, *NESET II report.* Luxembourg: Publications Office of the European Union, doi: 10.2766/71491.

William, D., Lee, C., Flarrison, C., Black, P. (2004). Teachers developing assessment for learning: Impact on student achievement, *Assessment in Education: Principles, Policy & Practice*, 11(1), 49-65.

## MODULE TWO

## TYPOLOGY OF EDUCATIONAL MEASUREMENT

**by**

**Monday Joshua**

### 2.0 Module Overview

Assessment of educational outcomes, whether at classroom level, school level, state, regional or national levels, makes use of various instruments. The most common, and perhaps the easiest instrument of measurement in the cognitive domain, is test. This is so mainly because tests have uniformity of scales, and calculable reliability and validity indices. Test itself is defined as an instrument for systematic measure of a sample of behaviour, or a systematic procedure for observing a person's behaviour, and describing such with the aid of a numerical score or scale. Thus, a test is seen as an instrument that measures behaviour. How does it measure? It measures behaviour systematically, and it cannot measure ALL the possible behaviours at the same time in one given instrument. So, a given test measures only a sample of possible behaviours (Brown, 1983; Joshua, 1998, 2005).

Just as a measuring tape is the instrument in the hands of a carpenter or a tailor for gathering data for decision making about his/her clients, a test is a measuring instrument in the hands of a teacher or assessor for gathering data for decision making about his/her learners (clients). Unlike in the direct measurement done by a carpenter or tailor, the measurement done by the teacher is indirect; that is, a person is not measured as in measuring a table or the waistline of a man, rather we infer a person's characteristics (e. g. mathematical adding ability) from his/her responses to test items (questions). A test contains a sample of all possible items (questions). The application of test (the instrument) produces data (scores) for decision making. There are, as expected,

different types of tests, and there are different classifications of tests. The criteria or bases of classification are many and different. This module examines the different bases of classifications of tests, and the types of tests under such classifications. This is done in five (5) units. Unit 1 explores the different bases of classification of tests and the types of tests under each, while the details of the classifications of tests under four of the bases are examined in Units 2, 3, 4 and 5. Each Unit shows its objectives, the content, some activity exercises, and a summary.

Thus, this Module is organised under the following units:

i. Classification of tests (generally).
ii. Classification of tests based on what they measure (Test Purpose)
iii. Classification of tests based on method of interpreting the scores.
iv. Classification of tests based on use in classroom instruction.
v. Classification of tests based on method of preparation (Who constructs?)

## Unit 2.1: Classification of Tests (Generally)

### 2.1.1 Introduction

It is difficult to answer the question "how are tests classified?" straight away without further enquiry like "on what basis?" Therefore, to give proper answer to such question, it is important to first specify the basis/bases on which the classification is to be done. Thus, tests for assessment can be classified in many ways; and some of the bases of such classifications, and the types of tests under such are presented in this Unit.

### 2.1.2 Expected Learning Outcomes

At the completion of studying this Unit, you should be able to:

- correctly explain the reason for first specifying the basis/bases for classification of tests before such classification;

- mention at least seven bases on which tests can be classified;
- explain the different types/classifications of tests under each basis mentioned;
- construct at least two items to differentiate the types of tests under some specific bases of classification of tests.

### 2.1.3 Bases of Classification

Generally, tests can be classified on the following bases:

***Method of Administration***

Tests can be classified on the basis of how the given test is administered to the testee(s). On this basis, we have ***Individual tests*** and ***Group tests***. Individual test is administered on one-on-one approach, i.e., one person at a time. The individual has no partner or competitor while taking the tests. The results from such tests are used mainly to understand and describe the person (testee). Counsellors, psychologists, and teachers (under certain circumstances) do administer individual tests. With the advent of technology, many types of tests are being administered individually, many times at the convenience of the test taker. On the other hand, group test is administered to a group of individuals at the same time, in the same accommodation and under the same physical and climatic conditions. The results from such tests are usually used to describe and classify the group members. Most of the tests given in school settings are group tests.

***Type of Response Demanded***

On this basis, tests are classified as either *Free-Response items tests or Alternative Response items tests.*

In a Free-Response items test, the testee is required by the examiner to supply the expected answers to the questions.

For the Alternative Response items test, the testee is required by the examiner to select the expected answers from the groups of alternatives or options given by the examiner.

Example:
**Free-Response Type** - Which river is the longest in Africa? ----------
**Alternative-Response Type** - Which of the following is the longest river in Africa?
A. River Niger B. River Nile C. River Benue D. River Limpopo

For the first item, the testee freely supplies the expected answer; but for the second question, the testee selects the expected answer from the alternatives given by the examiner.

***Time Required for The Test***
On this basis, tests are classified as either *Speed tests* or *Power tests*.
Speed tests have a large number of relatively simpler items but with stringent time limits, with an individual's score in it being an index of speed of response.

Power tests, on the other hand, have items of varying difficulty but with time allocation that allows completion of all items. Here, individual's score reflects his/her maximum performance under ample time condition.

***Type of Learning Outcomes***
On this basis, tests are classified as *Mastery tests* or *Survey tests.*
Mastery tests measure degree or extent of mastery of a limited set of specific learning outcomes. Survey tests, on the other hand, measure a learner's general level of achievement over a broad range of outcomes.

### *Type of Performance Displayed*

On this basis, tests are classified as either *Maximal (*or *Maximum) Performance tests* or *Typical Performance tests.*
In a Maximal Performance test, a testee attempts to make the best or highest possible score. The goal is to measure or identify the upper limit or maximum of the testee's ability (or whatever trait that is being measured). On the other hand, Typical Maximum test seeks to measure the individual's usual or habitual behavior, performance, or reaction. The goal is to measure or identify what is typical of the individual in given situations, while questions that require or contain right and wrong answers do measure maximum performance. Questions that require levels/degrees of agreement (e.g., Likert Attitude Scale using *strongly agree, agree, disagree, and strongly disagree),* importance (e.g., *not very important, not important, important, not very important),* frequency of response (e.g., *always, often, rarely, never*),etc., measure typical performance.

### *Types of Scoring*

On this basis, tests are classified as either Objective tests or Subjective tests. An Objective test is one which produces the same (near same) score by many equally competent scorers; while a subjective test is one in which scores awarded by other equally competent scorers vary owing to influence of other factors not relevant to the purpose of the measurement. Such factors could be testee's handwriting, scorer's knowledge/opinion/judgement of the testee, etc.

### *What the Test Measures (Test Purpose)*

On this basis, tests are classified as *Achievement tests, Aptitude tests, Ability tests* or *Intelligence tests*. Elaboration on these is done in a subsequent Unit.

***Method of Interpreting Scores***

On this basis, tests are classified as either *Norm-Referenced tests* or *Criterion-Referenced tests*. Elaboration on these is done in a subsequent Unit.

***Use in Classroom Instruction***

On this basis, tests are classified as *Placement tests, Diagnostic tests, Formative tests,* or *Summative tests*. Elaboration on these is done in a subsequent Unit.

***Method of Preparation/Who Constructs?***

On this basis, tests are classified as either *Teacher-Made tests* or *Standardised tests*. Elaboration on these is done in a subsequent Unit.

***Activity 2.1***

1. Discuss (in groups) the practical applications of different types of tests identified in this Unit.
2. Construct two items to differentiate classification of tests on the bases of:
   a. Type of response required,
   b. Types of Scoring, and
   c. Types of Performance displayed.
3. Discuss the challenges of using each type of tests (as per the bases identified) in our school.

## 2.1.4 Summary

In this Unit, you have learnt that:

a. Classification of tests cannot be done anyhow,
b. There is need to identify or specify the basis/bases for such classification.

c. Ten of such bases have been identified and highlighted in this Unit, and under each basis, the types or classifications of tests have been outlined, and examples given in some cases.

## Unit 2.2: Classification of Tests Based on what they Measure (Test Purpose)

### 2.2.1 Introduction

Educational tests are constructed to measure different things, and at different times in the course of the teaching-learning process. In other words, tests are constructed to serve different purposes. So, educational tests can be classified on the basis of what they are designed to measure. On this basis, four (4) major types of tests are identified and discussed in this Unit. These are Achievement tests, Aptitude tests, Ability tests and Intelligence tests. Each has its own importance, frame of reference and other peculiarities.

### 2.2.2 Expected Learning Outcomes

At the end of this Unit, you should be able to:

- explain the similarities and differences among the four types of tests (on the basis of what they measure).
- identify practical situations where each of these test types will be more appropriate than the others.

### 2.2.3 Kinds of Test Based on What they Measure

As indicated in the previous Unit, tests, on the basis of what they measure, are classified as *Achievement tests, Aptitude tests, Ability tests or Intelligence tests.*

***Achievement Test***

An achievement test measures learning that has taken place, i.e., what has been achieved, after a certain specific teaching-learning process has taken place. It is tied to specific school subjects or well-defined

curriculum (curricula). It measures the present level of knowledge the learner has gained (or achieved) after going through a specified content/curriculum. Thus, its frame of reference is the present. Most of the tests given to pupils/students in schools and by examination bodies are examples of achievement tests. Achievement test scores are often used in educational system to determine the level of instruction for which a student is prepared. High achievement scores usually indicate a mastery of the grade-level material, and the readiness for advanced instruction. Low achievement scores equally indicate low-level mastery, and therefore, the need for mediation or repeating the teaching-learning process. Achievement tests have also been used to assess or determine proficiency of learners of well-defined knowledge and/or skill.

***Aptitude Test***

Aptitude is an inborn talent or capacity to do something in future. Aptitude test measures a person's ability or potential to acquire certain skills or behaviours, if given appropriate training. It is a systematic means of testing a job, candidate's abilities to perform specific tasks, and react to a range of different situations. Though given at the present, aptitude test captures a person's propensity or likelihood of doing something, mastering a skill, or benefitting from an educational/training programme in the future. Thus, the frame of reference for aptitude test is the future performance. Hence, aptitude test predicts future performance. Aptitude tests are used extensively for selection purposes: for example, selection of applicants for jobs or admission offers in secondary and tertiary educational institutions and other professional schools (e.g., verbal and quantitative aptitude tests).

***Ability Test***

An ability test measures the power (or ability) to perform a certain present task. It is a test to check an individual's ability to perform a certain kind of job. It usually tests the individual for certain skills or whether he is capable of acquiring those skills in the course of the

present job or through training. Its frame of reference is the present (though it could also be used to predict future performance). While achievement test measures well-defined specific learning experience, ability test measures more general/broad learning experiences such as problem solving, communication skill, critical thinking, and manual dexterity. Ability tests are used to assess individual differences in skills and knowledge.

***Intelligence Test***

An intelligence test is a test of general mental ability or general intellectual level. It measures an individual's capacity to think abstractly, to adapt to new situations, or to integrate new experiences. It consists of series of tasks designed to measure capacity to make abstractions, to learn, and to deal with novel situations and ideas. Intelligence tests are psychological tests designed to measure a variety of mental functions such as reasoning, comprehension adjudgment. The goal of intelligence tests is to obtain an idea of the person's intellectual potential. The test centres around a set of stimuli designed to yield a score based on the test maker's model or concept of intelligence. Intelligence tests are often given as a battery of tests. They usually yield a single score known as “intelligence quotient (IQ). Some test experts see intelligence test as a kind of ability test, while others see it as a kind of aptitude test. Thus, its frame of reference is both present and future.

It is to be noted that the line(s) delineating each of these test types from others is/are very thin, and all of these can be used to serve same or similar purposes, for example, prediction of students' performance in examinations.

***Activity 2.2***

1. Explain the similarities and differences among the four types of tests (on the basis of what they measure).
2. Identify practical situations where each of these test types will be more appropriate than the others.

### 2.2.4 Summary

At the end of this Unit, you have learnt that:
When tests are classified on the basis of what they measure (or test purpose), four test types readily emerge; achievement tests, aptitude tests, ability tests, and intelligence tests. Achievement tests measure knowledge, skills or learning that are presumed to have been acquired (or achieved) after a prescribed teaching-learning process. Aptitude tests measure the capacity or potential to learn new things subsequently. Ability tests measure general endowment at any given time for a particular purpose. Intelligence tests measure general mental ability to think abstractly, integrate novel experiences or adapt to new situations. The term of reference for achievement tests is the present, for ability tests the present, for aptitude tests the future, and for intelligence tests both the present and the future.

The line separating these test types is very thin.

## Unit 2.3: Classification of Tests based on Method of Interpreting the Scores

### 2.3.1 Introduction

The end product from testing is scores; that is, application of test on the examinees produces scores. Scores constitute raw data for taking decisions, and they do not make meaning until they are interpreted. Thus, tests can be classified based on the method of interpreting the scores. There are two major types of tests based on the method of

interpreting test scores, namely: *Norm-Referenced tests and Criterion-Referenced tests*.

### 2.3.2 Expected Learning Outcomes

At the completion of studying this Unit, you should be able to:

- compare and contrast the characteristics of Norm-referenced tests and Criterion-referenced tests.
- identify in the Nigerian setting, the practical testing situations that call for each of Norm-referenced tests and Criterion-referenced tests.

### 2.3.3 Test Classification based on Methods of Interpreting the Scores

The two major types of tests based on the method of interpreting test scores are Norm-Referenced tests and Criterion-Referenced tests.

***Norm-Referenced Tests***

A norm-referenced test is one that produces scores, and an individual's score derives meaning from the general performance of the group members taking the test with him/her. In it, an individual's score is interpreted with reference to the performance of the group members taking the test. That is, an individual's score only makes meaning when you compare it with scores of other members of the group/class. Norm has to do with what is peculiar or typical of a group. Norm-referenced implies that a particular behavior or performance is being referred to, or compared with, the typical or peculiar behavior/performance (norm) of the group in focus. Thus, norm-referenced interpretation of test results is in terms of how a learner performs relative to his/her class/course mates. In interpreting results from norm-referenced test, we are not describing number-right or percentage of test items a student scores right, rather we describe the proportion or percentage of the students in his/her norm group that he/she has surpassed. We are describing the standing of the individual based on his/her group performance. Norm- referenced tests are designed primarily to rank the learners/testees in order of

achievement or performance, from high to low, or vice versa; and an individual's performance is reflected in his/her rank in the group/class. Another approach in handling scores from norm-referenced tests is to determine the mean and standard deviation of the scores of the class; and an individual's performance is seen as the distance of his/her own score from the mean of the class, or how many standard deviations is his/her score from the mean of the class. Tests that serve selection purposes are mainly Norm-referenced tests. Examples include Unified Tertiary Matriculation Examination (UTME), Post-UTME and other examinations/tests conducted by examination bodies in Nigeria.

***Criterion-Referenced Tests***

A Criterion-referenced test is one in which an individual's score is interpreted with reference to a pre-determined standard of performance called the "criterion". That is, an individual's score only makes meaning when you compare it with a standard earlier defined, without reference to the performance of other members of his/her class or group. Instead of ranking the testees in the order of their performance on the test within a group, Criterion-referenced test, instead, enables one to describe specifically what an individual can do in terms of the criterion or standard performance expected, without recourse to what other members of the group can or cannot do. The criterion can be derived by identifying the specific behavior, as stated in the behavioural objectives, that the learner will demonstrate to represent mastery of the material/skill learnt. This is usually represented by cut score, or criterion score or cut-off point, sometimes pre-determined before the test is taken, or determined based on the spread and general state/nationwide performance on the given test. An examinee passes the test if his/her score reaches or exceeds cut score/cut-off point and fails if it does not. The concern in criterion-referenced testing is to determine whether the learners have learnt or mastered a certain content or skill before graduation or progressing to another level or content area/topic. It is not about how an individual performs with reference to his/her group

members, but it is whether an individual's performance has met the expected or desired standard set prior to taking the test. Most of the tests that prepare or qualify examinees for licensing are Criterion-referenced tests, e.g., Teachers' Registration Council of Nigeria (TRCN), Nigeria Institute of Management (NIM), Nigerian Medical Association (NMA), and Institute of Chartered Accountant of Nigeria (ICAN).

***Activity 2.3***

1. Compare and contrast the characteristics of Norm-referenced tests and Criterion-referenced tests.
2. Identify in the Nigerian setting the practical testing situations that call for each of Norm-referenced tests and Criterion-referenced

### *2.3.4 Summary*

In this Unit, you have learnt that:

Tests can be classified based on how scores are being interpreted. On this basis, the two types of tests are Norm-referenced tests and Criterion-referenced tests.

In Norm-referenced test, an individual's score has meaning only when it is compared to the scores or performance of other members of his/her group. He/she can be a star or otherwise, all depending on the performance of his/her group members.

In Criterion-referenced test, an individual's score has meaning only when it reaches or exceeds the desired or expected standard or criterion represented by a criterion score or cut-off point.

Norm-referenced tests identify people in their standings or ranks within their group, while Criterion-referenced tests identify people in a group that have mastered the content/skill taught before being certified to practice or move to subsequent level.

## Unit 2.4: Classification of Tests Based on Use of the Results

### 2.4.1 Introduction

Tests serve many and different roles prior to and during classroom instruction. Different tests are used for different purposes. Generally, test results are used for taking definite decisions, and these decisions also dictate the time in a typical teaching-learning process when these tests are administered. When tests are classified based on the use of their scores in classroom instruction, four types of tests are identified. These are Placement tests, Formative tests, Diagnostic tests, and Summative tests.

### 2.4.2 Expected Learning Outcomes

At the completion of studying this Unit, you should be able to:

- compare and contrast the characteristics of each of placement tests, formative tests, diagnostic tests, and summative tests.
- identify in the Nigerian setting the practical testing situations that call for each of these four test types.
- identify the advantages and disadvantages of each of these four test types.
- discuss the role of the teacher and that of the learners in maximising the benefits of each of the four test types.

### 2.4.3 Test Types Based on Use of Test Results

The types of tests under the classification on the basis of using test scores to enhance classroom instruction are Placement tests, Formative tests, Diagnostic tests, and Summative tests.

***Placement Test***

A placement test is designed for use in determining learner's entry performance or behavior prior to commencement or furtherance of the teaching learning process. It seeks to determine the knowledge and

skills that the learner(s) already possess(es) in readiness to commence or benefit from the instructional process. The results from Placement test indicate how much of what is to be taught that the learner(s) already know(s). Such information equips the teacher with information about the learner's readiness to begin and benefit from current instruction. It also indicates the extent to which some or all of the students have already achieved the objectives of the lesson being planned for them. The goal of placement testing is to determine the position in the instructional sequence and the mode of instruction that is most likely to be beneficial to the learner. It serves a useful purpose of helping the teacher or school administrator to properly place each learner at the appropriate point in an instructional sequence, or in class or programme that best suits the learner's aptitude, interest and, perhaps, aspirations. Placement test is usually referred to as pre-test as it is usually given prior to the commencement of teaching-learning process or programme implementation.

***Formative Test***

A formative test is usually undertaken during the process of teaching or programme implementation, and the result of it is used as feedback or input to improve the teaching-learning process or improve the programme implementation. It is used to monitor learning progress during instruction, and to provide continuous feedback for the improvement of the instructional process to realise the goals of the process. It provides information on the areas of strengths and weaknesses in students' learning, as well as those aspects of the instructional techniques and materials that are not adequate. Classroom work, assignments, bi-weekly or monthly tests/quizzes, teacher's observations or ratings are forms or examples of formative tests. Generally, formative test provides feedback to the learners, the teacher, the curriculum developer, the sponsor/financier, the organisation/ government that is to decide on the adoption or otherwise of the programme being implemented/experimented, and to all other

stakeholders concerned. It is periodic, that is, carried out at various points during the course of instruction, resulting in a number of assessments that are not used for grading. It fits into modern assessment mode referred to as assessment for learning, as against the traditional mode of assessment of learning. It enhances the quality of teaching and learning and encourages individualised instruction.

***Diagnostic Test***

A diagnostic test is usually administered to unravel any persistent or recurring learning difficulties that are left unresolved by the standard corrective prescriptions of formative assessment. It involves the use of specially prepared diagnostic items as well as some observational techniques, and sometimes the services of remedial, psychological, or medical specialists to identify typical or underlying problems impeding learning, and the necessary therapies to be applied. The goal of a diagnostic test is not to measure achievement over a wide area of content. Rather, it is often centred on a limited content area. Such a content or segment of instruction is usually tested with items that have relatively low level of difficulty. The intention is to identify as much as possible learners that have or exhibit problems in learning or undergoing certain aspects of instruction or learning experiences. The main aim in diagnostic testing is to determine possible causes of learning problems, and to formulate plans for remedial action.

***Summative Test***

A summative test is usually administered at the end of an instructional process, or at the end of term/year, or the end of programme, or the end of curriculum implementation to provide data for some terminal decisions and value judgments. It is designed to determine the extent to which the earlier specified objectives of the lesson/course/ programme/curriculum have been achieved. A summative test samples a wide area of content and makes use of test items with a wide range of item difficulties. The essence is to make sure that learners of varying abilities and aptitudes have the opportunity of performing maximally,

and to determine what an individual has gained from a lesson/subject/course/programme. The results are used to assign grades, to determine mastery, for certification, for promotion/demotion, for adoption/continuation/dropping of the programme/curriculum, for awards. The results may also be used to judge the appropriateness of the course/programme objectives, the effectiveness of the instruction and the propriety of learning outcomes. It fits into traditional assessment mode referred to as assessment of learning, as against a relatively modern one called assessment for learning. End-of- term/year examinations and all other examinations from examination bodies and government/private agencies that end up in award of grades, certificates or prizes are forms or examples of summative tests.

***Activity 2.4***

1. Compare and contrast the characteristics of placement tests, formative tests, diagnostic tests, and summative tests.
2. Identify in Nigerian setting the practical testing situations that call for each of these four test types.
3. Identify the advantages and disadvantages of each of these four test types.
4. Discuss the role of the teacher and that of the learners in maximising the benefits of each of the four test types.

### 2.4.4 Summary

In this Unit, you have learnt that:
Tests can be classified based on the use to which the results will be put in classroom instruction. On this basis, the four types of tests are Placement tests, Formative tests, Diagnostic tests, and Summative tests. Placement tests are used to determine prerequisite skills, degree of mastery of previous course objectives, and/or best mode of learning/teaching. Formative tests are used to determine learning progress, provide feedback to reinforce learning, and correct learning

errors. Diagnostic tests are used to determine causes of persistent learning difficulties. Summative tests are used to determine end-of lesson/term/year/course/programme achievement for assignment of grades/certificates/prizes, and/or to certify mastery of objectives. Each of these test types has a unique time of administration in the teaching-learning process, and its own unique advantages and challenges.

## Unit 2.5: Classification of Tests based on Method of Preparation (Who Constructs?)

### 2.5.1 Introduction

Learners (pupils and students) in schools are subjected to tests of many types, and at different times, and from different test makers. The test makers could be within the school, or outside the school. When tests are classified based on method of preparation, or who constructs the tests, there are two types of tests. These are *Teacher-made tests* and *Standardised tests*. Each of these has its peculiarities and its strategic place in the instructional/school system.

### 2.5.2 Expected Learning Outcomes

At the completion of studying this Unit, you should be able to:

- compare and contrast the characteristics of each of teacher-made tests, and standardised tests.
- identify in Nigerian setting the practical testing situations that call for each of these two test types.
- identify the advantages and disadvantages of each of these two test types.
- discuss the role of the teacher and that of the learners in maximising the benefits of each of the two test types.

### 2.5.3 Test Types based on Method of Preparation (Who Constructs?)

The types of tests, on the basis of who constructs them, are *Teacher-made tests* and *Standardised tests*.

***Teacher-made Tests***

These are tests that are usually constructed by the class teacher, or a group of class teachers in one school, and are administered and scored (marked) by the teacher and interpreted by the teacher for use within his/her class or school. They are usually based on the content and objectives specific to the teacher's own class/school. A teacher-made test typically deals with specific topic(s) that may cover a week, month, term, or year; and such test is limited to the class taught by the teacher, or the school where the teacher is serving.

The norm of reference that guides the interpretation of scores is within the class/school. Teachers from different schools using the same content (syllabus) and teaching similar classes end up constructing and administrating different tests that are not comparable, thus producing different scores that are not additive and comparable in their raw forms. Except in cases where the teacher has mastered skills in test construction, items in teacher-made tests are usually characterised by low or moderate reliability and questionable validity. All Continuous Assessment (CA) tests and in-school end-of-semester examinations are examples of teacher- made tests.

***Standardised Tests***

These are tests that are constructed by experts or professionals in the subject area and/or experts in *Educational Measurement and Evaluation*. They are administered under uniform conditions and scored by the experts or trained agents and/or by machines. They are based on well selected and publicised content and objectives that are common to many schools in the same geographical area as local government area, state,

country, or region. They cover or deal with large segments of knowledge and skills that are assumed to have been taught over a time range across the entire geographical area of coverage. They contain items (questions) that have been tried out or trial-tested, pre-analysed and revised or improved before use. Thus, the items have high levels of validity and reliability. They draw their norms from the same large areas of coverage for comparison and interpretation. What makes a test standardised is the common content, common procedures and conditions of administration, and same interpretation terms and applications. Traditionally, a typical standardised test is administered to all the examinees in the geographical areas of coverage at the same time, with all the examinees engaging the same items (questions). However, with the advent of technology and new test theories (e.g., item response theory), it is now possible to construct appropriate parallel items/tests and the targeted examinees can take the test at different times and locations without compromising validity, reliability and standardisation of items and comparability of scores. A standardised test is typically a Norm-referenced test since individual's performance is usually compared or interpreted with reference to those of others in the defined norm group. The norm established makes the interpretation or scores meaningful for any examinee to whom the test is administered later. Tests constructed and administered by State Ministries/Departments of Education for the State's candidates, and those by examination bodies like JAMB, NECO, WAEC, NABTEB are examples of standardised tests. Some standardised tests are commercial in nature and can thus be “bought off the shelf.”

***Activity 2.5***

1. Compare and contrast the characteristics of teacher-made tests and standardised tests.
2. Identify the particular uses of each of teacher-made tests and standardised tests, particularly in Nigerian setting.
3. Discuss the peculiar factors that pose as challenges in the proper use of teacher-made tests and standardised tests in Nigerian school systems.
4. Discuss the necessity or otherwise of training in pedagogy and testing skills for ALL teachers at all levels who eventually test or assess learners.

### 2.5.4 Summary

In this Unit, you have learnt that:
Tests can be classified on the basis of the method of preparation (or who constructs?). On this basis, the two types of tests are Teacher-Made Tests (TMT) and Standardised Tests (ST). Teacher-made tests are constructed by the class teacher or group of teachers in one school, administered, scored, and interpreted by the same teacher, and the norm group is drawn from the class/school alone. Standardised tests are constructed by experts and examination bodies and administered and scored by paid competent agents. Content of TMT is limited to that covered in the class, but content in ST is broad and applies to all the examinees across the entire geographical area covered by the test. While STs have high validity and reliability, TMTs might have low or moderate validity and reliability because standard procedures are hardly followed in their development. Some STs are commercial in nature.

## Unit 2.6: Methods of Assessment in the Affective and Psychomotor Domains

### 2.6.1 Introduction

Learning outcomes and the instruments to assess/measure them as discussed in the preceding sections (Units 1-5) centre mainly on the cognitive aspect of teaching and learning. They fall under what

psychologists and educationalists refer to as cognitive domain, which is one of the three main domains of learning; the other two being affective domain and psychomotor domain. Learning and learning outcomes certainly manifest beyond cognitive domain. Therefore, for total or more comprehensive assessment/evaluation, all the domains have to be attended to. This implies that educators, and teachers in particular, have to think beyond use of tests, which is the basic assessment instrument in the cognitive domain. In this Unit, we are going to highlight learning outcomes in the other domains of learning and discuss the various techniques available to teachers/educators for assessment of the non-cognitive learning outcomes in the affective and psychomotor domains.

### 2.6.2 Expected Learning Outcomes

At the end of this Unit, you should be able to:

- identify learning outcomes in the affective and psychomotor domains that require non- cognitive assessment techniques.
- identify assessment instruments that are used for assessment of learning outcomes in the affective and psychomotor domains.
- explain how each of the assessment instruments for affective and psychomotor traits/skills can be used.
- explain the classification of instruments for assessment of non-cognitive learning outcomes.

### 2.6.3 Assessment of Learning Outcomes in the Affective Domains

Learning outcomes beyond knowledge that require non-cognitive assessment/evaluation techniques include:

- **Skills:** hard and soft skills, and skills of speaking, writing, listening, dancing, problem solving, collaborative, etc.
- **Habits:** work habits, study habits, eating habits, dressing habits, etc.
- **Attitudes:** social attitudes, scientific attitude, attitude toward a subject/course, etc.

- **Interests:** in subjects, careers/vocations, etc.
- **Values:** conscious and unconscious belief system, etc.
- **Appreciation:** of artwork, of nature, etc.
- **Adjustment:** relationship with others, or to situations, e.g., social adjustment, emotional adjustment, marital adjustment, and peer pressure adjustment.

These non-cognitive measures or learning outcomes cannot be assessed or evaluated using the conventional test forms. Methods (Instruments) used to assess/evaluate them include:

***Observation Technique***

Observation technique is the first and primary measurement or assessment technique a teacher uses in the classroom to see, first hand, the behaviour, and changes in behavior of his pupils/students, and this can aid assessment in any of the domains. Observation techniques are used for assessment of social behaviours, beliefs, and values. It could be direct observation (with or without gadgets like recorders) or participant observation where the observer becomes like one of those being observed.

***Rating Scale***

This takes the form of a set of words or statements (stimuli) describing traits, qualities, or characteristics to be judged in terms of degree of presence. It is used to rate behaviours, qualities of finished products, etc.

***Checklist***

This is similar to rating scale except that in checklist, the stimuli are to be judged or rated as being present or not present instead of requiring the degree of presence. Sometimes, the frequency or order of occurrence may be required by a checklist. It can also be used to assess clearly defined performance skills (in psychomotor domain).

***Anecdotal Records***

These are records of a series of observed incidents involving a particular person. The recorder records the event, the reason for the action, and or the interpretation. The teacher can use this to record factual information about classroom-observed incidents and events in the lives of the learners.

***Interest Inventory***

This is a list of questions, words or statements which assess individual's preferences for activities, topics/subjects, careers, etc. to which the person responds. Analysed results help to guide the individual's choices of activities, vocations, etc. examples of published interest inventories include Strong-Campbell Interest Inventory and Kuder General Interest Survey.

***Attitude Scale***

This instrument comprises a series of statements that show individual's likes or dislikes for an object, person, subject, career, situation, institution, or event. An individual responds by indicating the level of his agreement or disagreement by choosing one of the response categories. Various types of attitude scales abound, such as Likert scale, Thurstone scale, Guttman scale and Osgood scale, with Likert scale being the most popular.

***Personality Measures***

These are used mainly to assess the personality trait of individuals. There are many traits that come under the umbrella of personality, and each has its peculiar way of measurement/assessment. For example, traits like self-concept, locus of control, temperament, extraversion-introversion, etc are considered under personality, and are measured differently. Problem checklists, personality inventories and projective techniques are some of the popular personality measures in vogue.

***Peer Appraisal Instruments***

These are given to learners to rate their peers (fellow learners) on the same clearly defined issues or behaviours. Common types are "Guess Who Technique", and "Sociometric Technique". The "Guess Who Technique" presents a series of brief behaviour descriptions or characteristics to a learner, and he is asked to identify or guess who, among his peers, a particular description fits or represents. Sociometric technique is a method of evaluating the social acceptance of individual learners (or social distances among individuals), and the social structure of a group. Learners are asked to choose among their colleagues those they would want to have as friends or companions for some group work or assignment.

***Self-Appraisal Technique***

Another technique of assessing student learning, which fits into both cognitive and affective domains, is student self-appraisal. This is a process by which students gather information about themselves and reflect on their own learning. It is regarded as a very important component of learning because the awareness of learning progress will motivate the student for further learning (Sharma, Jain, Gupta, Garg, Battaa & Dhir, 2016). Moreover, the process of self-grading, assessment and evaluation enable the students to critically analyse their deficiencies and try to be careful when attempting further tests. Self-appraisal helps students find out what they know, what they don't, what they would like to know and how to fill the gaps discovered; and these can be done in several different ways, some of which include:

- Filling self-evaluation forms;
- Taking personal timed tests from previous examinations and tests;
- Writing revisions of work done;
- Group discussions and questioning: listening to others can reshape wrong pre-conceived perceptions;
- Asking others to rate one's performance;

- Reviewing of tests already done can shed light on areas where gaps exist;
- Reviewing set goals and achievements so far.

It is important that the teacher shows creativity in each of these steps, makes them flexible and avoid monotony and boredom. There is need to also know that multiple paths lead to knowledge, understanding and learning; and it is through constant self-appraisal that one can discover and optimise his strengths and become aware of, and minimise his weaknesses. (Joshua, 2005, 2019; Joshua & Bassey, 2011).

### 2.6.4 Classification of Instruments for Assessing Non-cognitive Learning Outcomes

In general, techniques for assessing non-cognitive learning outcomes can be classified into three main categories, namely: observational techniques, self- report techniques and peer-appraisal techniques. In most cases, the self-report techniques (rating scales, checklists, attitude scales, interest inventories, personality measures, etc) and some of the peer-appraisal techniques are designed in the form of questionnaires and or interview (Joshua, 2005, 2019; Joshua & Bassey, 2011).

### 2.6.5 Methods of Assessment/Evaluation in the Psychomotor Domain

The methods or techniques of assessing or evaluating learning outcomes in the psychomotor domain, to a great extent, are largely the same as the ones for affective domain discussed earlier. However, Vinson, in Joshua and Bassey (2011), has added that the psychomotor domain is best assessed in a face-to-face 'hands-on' situation (Practical test) where the student performs the skill with the instructor by his side observing and judging if the skill is performed to a satisfactory standard, rather than in the mediated way. Sometimes, the practical test is performed independently by the student (without the instructor by his/her side; and

the assessment of the product is undertaken subsequently by the instructor. But, even in this judgement, one or more of the techniques earlier described, e.g., checklist, rating scale, is/are used (Joshua & Bassey, 2011). **See Module 3, Unit 4 for more detailed treatment of practical tests.**

***Activity 2.6***

1. Identify learning outcomes in the affective and psychomotor domains that require non-cognitive assessment techniques.
2. Identify assessment instruments that are used for assessment of learning outcomes in the affective and psychomotor domains.
3. Explain how each of the assessment instruments for affective and psychomotor traits/skills can be used.
4. Explain the classification of instruments for assessment of non-cognitive learning outcomes.
5. Discuss in groups how each of peer-appraisal and self-appraisal assessment techniques can be used effectively to maximise their usefulness.

### 2.6.6 Summary

In this Unit, you have learnt that:

Assessment also takes, or should take, place in affective and psychomotor domains as there are certainly non-cognitive learning outcomes. These non-cognitive learning outcomes (traits/skills) include skills, habits, attitudes, interests, values, appreciations, and adjustments (social, emotional, marital, peer pressure, etc.). Instrument/techniques for assessing these non-cognitive traits and skills include observational techniques, rating scales, checklists, anecdotal records, interest inventories, attitudes scales (e.g., Likert scale), personality measures, peer appraisal techniques, and self-appraisal techniques. These assessment techniques in the two domains can be classified into three:

observational techniques, peer-appraisal techniques, and self-appraisal techniques. Practical test is one of the main assessment techniques in the psychomotor domain.

**Bibliography**

Denga, D. I. (1987). *Educational Measurement, Continuous Assessment, and Psychological Testing*. Calabar: Rapid Educational Publishers.

Umoinyang, I. E., Asim, A. E., Akwa, A. M., & Bassey, S. W. (2004). *Principles and Techniques of Educational Assessment and Evaluation*. Calabar: Helimo Associates.

**References**

Brown, F. G. (1983). *Principles of Educational and Psychological Testing* (3rd ed.). New York: Holt, Reinhart, and Winston.

Gronlund, N. E. (1985). *Measurement and Evaluation in Teaching* (5th ed.). New York: Macmillan.

Joshua, M. T. (1998). *Roles of Tests, Measurement, and Evaluation in Counselling.* Nigerian Education Journal, 2(1), 57-64.

Joshua, M. T. (2005). *Fundamentals of Tests and Measurement in Education.* Calabar: The University of Calabar Press.

Joshua, M. T. (2019). *Battles in the classroom: Evaluation of teaching and learning to the rescue* (86th Inaugural Lecture of the University of Calabar, Calabar, Nigeria). Calabar: The University of Calabar Press.

Joshua, M. T., & Bassey, S. W. (2011). *Assessment of Learning Outcomes*. In A. J. Isangedighi (Ed.), Essentials of learning (pp. 289-306). Calabar: Eti-Nwa

Sharma, R., Jain, A., Gupta, N., Garg, M., Battaa, M., & Dhir, S. (2016). *Impact of self-assessment by students on their learning-NCBI-NIH*. Retrieved online from: https://www.ncbi.nlm.nih.gov/pmc/articles/ on 23/09/2018.

## MODULE THREE

## CONSTRUCTING GOOD ACHIEVEMENT TESTS

**by**
**Mrs. C. A. Ugodulunwa**

### 3.0 Module Overview

One of the functions of the teacher is to construct and use different types of achievement tests to obtain useful data concerning learner achievement. This involves determining objectives to be measured, exposing learners to appropriate learning experience, and constructing appropriate test items for assessing the learning outcomes. The probability of developing good quality classroom tests is enhanced if the test is carefully planned, developed, and validated by the teacher. In this module, we shall discuss procedures for constructing different types of achievement tests that can be used in our school system. The module presents the concept and purpose of achievement tests, constructing essay and objective test items, designing oral assessment, designing practical tests and portfolio, as well as refining test item qualities.

### Unit 3.1: The Concept and Purpose of Achievement Tests

#### 3.1.0 Introduction

Achievement tests are developed and used by teachers for different purposes. It is therefore important for teachers to understand the concept of achievement test to enable them to use it effectively for enhancing teaching and learning in schools. This unit therefore focuses on the meaning of achievement test and its purposes in the school system.

#### 3.1.1 Expected Learning Outcomes

By the end of this Unit, you should be able to:

- Define the concept of achievement test.
- Describe the purpose of achievement test.

### 3.1.2 Concept of Achievement Test

An achievement test is a test constructed for assessing the extent to which a given content or task has been mastered after instruction. It is a test that measures relevant knowledge or skills in the school curriculum (Ugodulunwa, 2008) and involves the determination of attainment level of learners in tasks, courses, and programmes after exposure to instruction (Nworgu, 2015). It is an important tool for measuring instructional and students' progress in a course or subject area. High achievement test scores, therefore, usually indicate mastery of grade level materials and readiness for advanced instruction, while low achievement scores indicate the need for remedying learning difficulties and better teaching practices to improve performance. An achievement test is, therefore, a test that measures an individual's attainment, mastery and progression in a specific subject or topic, course, or programme after instruction over a specific period of time. Often times, you ask students questions before, during and after teaching and learning process to determine the amount of knowledge and skills they have mastered. You also give students assignments, project, quizzes, tests, and examinations in different courses at different times to assess their mastery and proficiency levels. Any test that measures the attainment and accomplishments of students after a period of instruction is an achievement test.

***Activity 3.1***

1. What is an achievement test?
2. Identify different achievement tests you use in assessing students' attainment in the courses you teach.

The goal of achievement testing is to obtain valid and reliable information (to be considered in the next module) about students'

achievement that will guide decision making in the school system. Specifically, we construct achievement tests to:

- measure students' current status with a set of knowledge and skills.
- indicate students' strengths and weaknesses.
- ensure that students are taught essential knowledge and skills.
- establish a minimum academic standard for promotion and certification.
- diagnose students' learning problems in different courses/subjects.
- place students at appropriate level for instruction.
- improve teaching and learning in school.

The different classroom tests, quizzes, assignments, projects, mid-semester, and end of semester examinations that are conducted in our institutions are examples of achievement tests.

### 3.1.3 Purpose of Achievement Tests

The purpose of an achievement test may be for placement, formative, diagnostic, or summative purpose (Gronlund, 1981).

A placement test is used at the beginning of a Unit or course to determine students' learning readiness. Most placement tests are pre-tests designed by teachers to measure the extent to which students possess the prerequisite skills required for success in an instructional unit or course. The results obtained are used to remedy entry deficiencies and to plan instruction. Formative test is used periodically during the teaching-learning process to monitor learning progress, direct learning through continuous feedback, and improve students' learning.

A test performs a diagnostic function when it is used for identifying and remedying learning difficulties, errors, and misconceptions which formative testing could not correct.

***Activity 3.2***

| Describe the purpose of achievement tests in classroom assessment. |
|---|

A Summative test is used at the end of a unit or course or programme to assign grades, evaluate teaching, and certify mastery as in the end of semester examination. Determining the purpose of a test is essential because it provides information on number and type of items to be included in an achievement test.

### 3.1.4 Summary

In this Unit, we have learnt about the meaning of achievement test and its goal. We also discussed the purpose of achievement tests, which include placement, formative, diagnostic and summative purposes.

## Unit 3.2: Constructing Essay and Objective Test Items

### 3.2.0 Introduction

Essay and objective tests are common forms of teacher-made achievement tests used for assessment of student learning outcomes. They are called teacher-made tests because they are prepared by teachers for a specific class and purposively based on what they taught. As teachers, you may write good quality items by chance, but it is not always possible. Construction of good items must adopt a laid-down procedure without which the quality of the test in terms of appropriateness and consistency cannot be guaranteed. In this Unit, you will learn the types of essay and objective tests, compare their characteristics, and describe the procedure for constructing them.

### 3.2.1 Expected Learning Outcomes

By the end of this Unit, you should be able to:

- describe the different types of essay tests.
- describe the different types of objective tests.

- describe the steps in constructing essay and objective tests.

### 3.2.2 Types of Essay Tests

Essay tests are tests which require students or testees to think and compose their responses before presenting them in writing. They are tests that allow students to select, organise, integrate, synthesise, and present answers in their own words. They are used by teachers to measure achievement based on classroom instruction.

The distinctive features of essay questions are that they require students to answer a number of questions based on the time limit of two or three hours. This does not allow teachers to cover all the topics or content taught in class. The answers to the questions are written in the students' own words. This may make spelling, handwriting and grammatical errors to influence students' test scores, if detailed scoring guide is not used for marking their answer scripts. Essay tests can be classified into two types: extended or free response items, and restricted response items.

***Extended-Response Type Essay Items***

Extended response type essay items are questions that give a wide range of freedom to students for response in their own words. They encourage creativity, originality, and divergent thinking by allowing students to respond in their own unique ways. Examples of extended essay questions are:

a. Discuss situations that illustrate the application of demand and supply in Nigeria's market economy.
b. Describe Shakespeare's view about women in his Romeo and Juliet.

***Restricted-Response Type Essay Items***

Restricted-response type essay questions are structured questions that limit the scope of the response. The answers given by the students are to some extent, controlled and limited both in content and form. They are

effective for measuring interpretation and application of ideas in specific areas but less valuable for measuring originality, integration, and organisation of ideas. Examples of restricted-response type essay items are given as follows:

a. Explain four ways in which land tenure systems affect agricultural production in Nigeria.
b. Describe five situations that illustrate the application of demand and supply in Nigeria's market economy.

### 3.2.3 Types of Objective Tests

Objective tests are highly structured test items that require the students to supply one word or few words, symbol, formulae, number or select correct answer from limited number of alternatives. It derives its name from the fact that the teachers' or scorers' personal biases do not influence the marking because a standard key or answer is provided. Objective tests enable the teachers to construct large number of items which may cover the entire course content or subject. The types of objective tests are shown in Figure 3.1

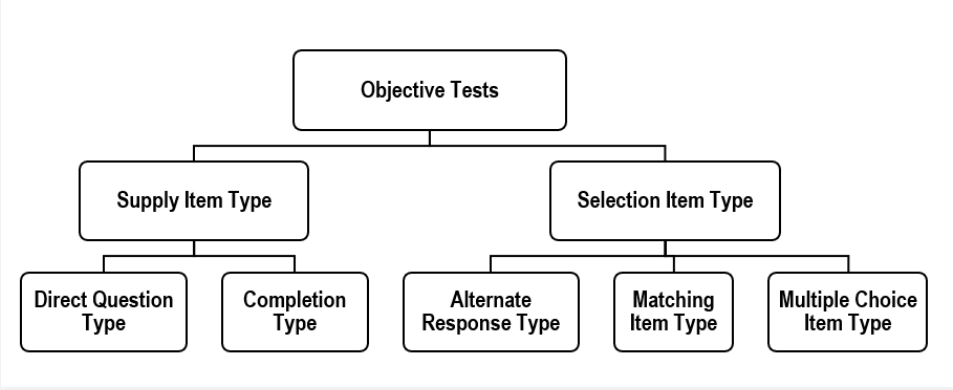



***Figure 3.1. Flow Chart Showing Types of Objective Test Item***

Let us look at each of them briefly.

***Supply Item***

Supply type items require testees to provide correct short answers to direct questions or complete sentences with correct answer or missing word(s).

Examples of short-answer items are:

a. What is the product of 2.5 and 2.05? (5.125)
b. In what year did Nigeria gain independence? (1960)

Examples of completion type objective items are:

a. Nigeria became independence in the year__
b. The green colouring matter in a plant leaf is called__

***Selection Item Type***

Selection type items require students or testees to select answer from a limited number of alternatives or options. Different types of selection type items are alternate type (true/false, or yes/no items), matching items, and multiple-choice items.

***Alternate-Response Item Type***

Alternate-Response Type items consist of a list of statements that require testees to mark true or false, right or wrong, correct or incorrect, yes or no, and agree or disagree. They measure learning outcomes that are simple and measure the students' ability to identify correctness of statements of facts and principles, definition of terms which are completely true or false.

Examples of this type of items are:

In the following statements, answer true if the statement is correct and false if the statement is wrong:

a. The movement of the earth round the sun is termed the earth's rotation. T/(F)
b. Respiration is an essential life process providing the living cells with oxygen. (T)/F

***Activity 3.3***

What do you think are the advantages and disadvantages of alternate type objective tests?

***Matching Item Type***

Matching items consist of two sets of related materials presented in two or more parallel columns with each word, number, symbol, in one column being matched to a word, sentence, or phrase in the other column. Testees are expected to match elements in one column with elements that are associated to them in the other column.

In constructing matching items, it is advisable to use homogeneous material in a single exercise, include an unequal number of responses (i.e., options) and premises (i.e., the statements or questions), keep the list of items to be matched brief and place the shorter responses on the right, indicate in the directions the basis for matching the responses and the premises, arrange the list of responses in logical order, and place all the items for one matching exercise on the same page as shown in the following example.

**Directions:** On the line to the left of each historical event in Nigeria in column A, write the letter of the date in column B when each event took place. Each date may be used once, more than once, or not at all.

| Column A(Premises) | Column B (Responses) |
|---|---|
| _____ 1. End of Nigeria's civil war | A. 1914 |
| _____ 2. Amalgamation of Nigeria | B. 1918 |
| _____ 3. Nigeria's first democracy day celebration | C. 1970 |
| | D. 1973 |
| | E. 1999 |

***Activity 3.4***

| Identify the advantages and disadvantages of Matching Items. |
|---|

***Multiple-Choice Items***

A multiple-choice item consists of two parts:

a. The stem of the item or problem and
b. A list of suggested answers referred to as the options or alternatives.

Testees are expected to read the stem and suggested alternatives and to select the one correct answer or best option. The correct option is the key or answer while the incorrect alternatives or options are the distracters. The stem is either a question or a statement.

Examples of multiple-choice items:

1. Which of the following instruments is used for measuring atmospheric pressure
   A. Anemometer
   B. Barometer
   C. Hygrometer
   D. Thermometer

2. An organised group that seeks the control of power in a State is
   A. Coalition group
   B. Pressure group
   C. Political group
   D. Social group

***Activity 3.5***

1. Identify the different parts of multiple-choice items.
2. Which of the objective test item types would you recommend for use at the tertiary level of education? Justify your answer.

### 3.2.4 Constructing Essay and Objective Test Items

The development of valid, reliable, and usable tests involves proper planning. This requires designing a blueprint or framework that can guide you in the test construction process. This is important because classroom achievement tests are vital in evaluating student learning outcomes. Let us now examine the following sequential steps you need to adopt in constructing good essay and objective achievement tests:

i. Planning the test
ii. Selecting appropriate item format
iii. Item writing
iv. Assembling the test items
v. Item vetting process or moderation

***Planning the Test***

Planning the test requires you to consider the purpose of the test, outline the instructional objectives and content it will cover, as well as develop a framework known as a table of specifications or a test blueprint for item writing.

***Determining the Purpose of the Test***

The purpose of the test may be for placement, formative, diagnostic, or summative purpose, which we have described earlier in this module. Determining the purpose of a test is essential because it provides information on number and type of items to include in the test.

***Outlining the Instructional Objectives***

You need to make an outline of the objectives that were emphasized during the teaching-learning process. You also need to analyse them according to levels of difficulty using Bloom's Taxonomy of Educational Objectives in the cognitive domain as a guide. The Bloom's taxonomy of objectives consists of the following six levels namely knowledge, understanding, application, analysis, synthesis, and evaluation in order of complexity (Bloom & Krathwohl, 1956). This Taxonomy was revised by Anderson and Krathwohl in 200l as follows: remembering, understanding, applying, analysing, evaluating, and synthesising/creating in order of complexity (Anderson & Krathwohl, 2001). Notice that in the newer taxonomy, verbs are used, and synthesis and evaluation are switched. You need to weigh each level of objectives using students' age and relative importance/emphasis during the lectures. The total unit weight should be 100 percent. For example, at the tertiary level 10% may be for remembering, 15% for understanding, 20% for applying, 25% for analysing, 20% for evaluating, and 10% for synthesizing. Action verbs you can use in statement of objectives at each level are presented in Table 3.1.

***Table 3.1: Action Verbs for Statement of Objectives***

| **Remembering** | **Understanding** | **Applying** | **Analysing** | **Evaluating** | **Creating** |
|---|---|---|---|---|---|
| Define | Describe | Apply | Analyse | Appraise | Arrange |
| List | Explain | Demonstrate | Distinguish | Assess | Assemble |
| Name | Express | Dramatize | Calculate | Choose | Build |
| Outline | Discuss | Employ | Categorize | Estimate | Compose |
| Recall | Report | Illustrate | Compare | Evaluate | Construct |
| Narrate | Locate | Interpret | Contrast | Judge | Create |
| Tell | Depict | Operate | Differentiate | Justify | Design |
| Label | Show | Practice | Debate | Rate | Formulate |
| Quote | Restate | Use | Examine | Defend | Plan |
| Recite | Review | Sketch | Experiment | Revise | Prepare |
| Reproduce | Translate | Employ | Solve | Argue | Combine |
| Recognize | Represent | Develop | Classify | Criticize | Propose |
| Memorize | Draw | Dramatize | Organise | Conclude | Generate |

Let us examine actions anticipated by some of the verbs in examination questions.

| | |
|---|---|
| ***Define:*** | Give meaning of something, idea, or concept |
| ***List:*** | Itemise or outline |
| ***State:*** | Give relevant points briefly |
| ***Describe:*** | Give details of process, properties, events, etc |
| ***Examine:*** | Take apart and describe something in great detail |
| ***Compare:*** | Put items side by side to see similarities and differences |
| ***Contrast:*** | Emphasise differences between two things |
| ***Discuss:*** | Explore various points of views which involves describing, explaining, giving examples, pros and cons, then analyse and evaluate results |
| ***Apply:*** | Apply principles and theories learnt in solving real life problem or in concrete situations |
| ***Analyse:*** | Take apart an idea, concept, or statement, examine and criticise in detail |
| ***Evaluate:*** | Judge suitability of something for a given purpose |
| ***Interpret:*** | Explain and comment on a particular issue or subject and make a judgement or evaluation |
| ***Justify:*** | Give reasons to support a statement |
| ***Illustrate:*** | Give concrete examples including use of figures or diagrams |
| ***Criticise:*** | Analysing, pointing out weaknesses as well as strengths |
| ***Assess:*** | Describe positive and negative aspects of a topic and how useful it is |
| ***Argue:*** | Defend a position, provide reasons in support or against a position |
| ***Summarise:*** | Give a brief account of something |
| ***Comment:*** | Offer statements on something, requires critiquing, discussing, evaluating, illustrating meaning of something and its implications |

***Outlining the Content to be Covered***

The topics a test will cover are the ones specified in the curriculum and expected to have been taught by you. The topics have to be weighted according to relative importance in terms of time and effort required for teaching each of them. The number of hours spent in teaching each topic should be used as a guide, while the total weight for all the topics to be covered should be 100 percent. For instance, if 2 hours was spent teaching a topic in a course that has 10 topics of 2 hours duration each week, the total number of hours spent teaching the course will be 20 hours. Each topic should therefore be allotted a weight of 10% (i.e., 2hrs/20hrs x100%).

**Developing a Test Blueprint or Table of Specifications**

A table of specifications, otherwise known as test blueprint, is a two-way matrix that relates content to objectives. It specifies proportion of questions to be written on each content and levels of objectives. It is derived from the field of architecture and it means 'detailed plan of action'. It is a good starting point for setting examination questions. Hassan (n.d) describes a blueprint of examination as an examination strategy of an institution at a glance. It is a globally accepted tool used for constructing examination or test questions for ready reference of stakeholders who engage in assessment process. Developing a test blueprint requires adoption of the following steps:

1. Identifying the purpose of test/examination.
2. Content analysis, which involves identifying and weighing topics taught, guided by relevance and time spent teaching each topic.
3. Identifying and weighing learning outcomes based on Anderson and Krathwohl's Taxonomy of cognitive objectives.
4. Determining the number of items/questions to develop.
5. Computing the number of items/questions for each topic and for each level of objectives.
6. Computing/determining the number of items for each cell in the matrix/table.

An example of a table of specifications for a 50-item multiple choice test is presented in Table 3.2.

Table 3.2: A Table of Specifications for a 50-Item Medical Biochemistry Course

| **Content** | | **Cognitive Objectives** | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | Weight | Remembering | Understanding | Applying | Analysing | Evaluating | Creating | **Total** |
| | | 10% | 15% | 20% | 25% | 10% | 20% | **100%** |
| Topic 1 | 20% | 1 | 1 | 2 | 3 | 1 | 2 | **10** |
| Topic 2 | 35% | 2 | 3 | 4 | 4 | 2 | 3 | **18** |
| Topic 3 | 45% | 2 | 4 | 4 | 5 | **2** | 5 | **22** |
| **Total** | **100%** | **5** | **8** | **10** | **12** | **5** | **10** | **50** |

Table 3.2 shows that the weight of the first topic is 20%, second topic is 35%, and the third topic is 45%, which gives a total of 100%. The weights for levels of objectives are 10% for remembering, 15% for understanding, 20% for applying, 25% for analysing, 10% for evaluating, and 20% for creating, which gave a total of 100%. To obtain the actual number of items for each of the topics, multiply the percentage weight for each of the topics by total number of items (See the marginal totals for content and for levels of objectives), that is:

Topic 1: 20% of 50 items = 10 items

Topic 2: 35% of 50 items = 18 items

Topic 3: 45% of 50 items = 22 items

To obtain the actual number of items for each level of objectives, multiply the percentage weight for each of the levels of objectives by total number of items like this:

Remembering: 10% of 50 items = 5 items
Understanding: 15% of 50 items = 7.5 rounded up to 8 items
Applying: 20% of 50 items = 10 items

| | | |
|---|---|---|
| Analysing: | 25% of 50 items | = 12.5 rounded down to 12 items |
| Evaluating: | 10% of 50 items | = 5 items |
| Creating: | 20% of 50 items | = 10 items |

To obtain the actual number of items to be assigned to each cell in the table, apply the following formula:

$$NC = WC \times WO \times NT$$

Where, NC = Number of items in a cell, WC = Weight of each content, WO =Weight of each level of objectives, NT = Total number of test items required.

You need to do some adjustments to ensure that the marginal totals are correct. A test blueprint is useful in test construction because it guides the teacher in ensuring content validity of a test or balance in testing. It also guides the teacher in item selection and arrangement, gives feedback on students' progress and teachers' delivery of curriculum. It provides sources for formative and summative assessment. The table also provides a clear framework for assessing important content and objectives in the curriculum, as well as information on how well students attain specified learning objectives.

### *Selecting Appropriate Item Format*

The format of a test can be essay or multiple-choice or a combination of the two item types. The kind of item selected should be suitable for assessing the learning objectives or outcomes. It is advisable to use multiple-choice and restricted response essay format because they can measure both lower and higher level objectives effectively. What to consider when determining whether to use or not to use essay and/or objective test items are action verbs in the intended learning outcomes that the test will assess. When you want to assess the ability to recall information, multiple-choice items should be used and when the action verbs elicit understanding of content or other higher level thought

processes, essay and multiple-choice item format can be used while action verbs that require students to construct responses only should be assessed using essay items (e.g., compose, construct, defend, justify, evaluate, explain, and propose). However, your choice of the format to select depends on your item writing skills, time, and resources available for marking the scripts.

### *Writing of the Test Items*

There are item writing rules you should bear in mind while writing or selecting test items for use in examinations. The table of specifications should guide the teacher in constructing or selecting appropriate test items. Try to generate items at appropriate levels of difficulty as specified in the table of specifications and develop a good marking scheme or scoring guide at the time the items are being constructed.

### *Guidelines for Writing Multiple-Choice Items*

Teachers should bear in mind the following guidelines suggested in the literature for writing multiple-choice items (Ugodulunwa, 2008; Nworgu, 2015):

1. Procedural guidelines such as use of correct answer format, vertical arrangement of items, avoiding tricks, avoiding complex multiple-choice options such as **'All of the above', 'None of the above', and 'A, B, and C'.**

2. Guidelines concerning content such as basing items on instructional objectives, avoiding content that has potential bias, focusing on single problem per item, testing significant materials, using items of moderate difficulty level, and keeping the vocabulary consistent with examinees' level of understanding.

3. Guidelines concerning stem of items such as stating stem of items in either, question or completion form, using clear directions, wording stem of the item positively, avoiding verbosity in the stem.

4. Guidelines concerning development of options such as placing options in logical or numerical order, avoiding overlapping options, avoiding determiners such as **'always' and 'never'**, avoiding or using the phrase '**all of the above**', and '**none of the above**', avoiding clues to the answer through use of faulty grammatical construction.

5. Guidelines concerning development of correct options such as positioning of the key so that it appears around the same number of times and ensuring that there is only one correct option.

6. Guidelines concerning development of distracters, such as using plausible distracters or options, incorporating common errors of students in the distracters, making options to be grammatically consistent with the stem.

***Guidelines for Writing Essay Items***

1. Clearly define the intended learning outcomes to be assessed.
2. Present challenging and reasonable test items.
3. Specify relative point value and time limit in the test direction.
4. Evaluate testees' responses anonymously to avoid halo effect.
5. Use large number of items that require relatively short answers.
6. Make prior decisions regarding treatment of factors such as spelling and punctuation.
7. Write the model answers while setting the questions.
8. It is preferable to give all testees same type of items to answer for comparability of standard.

***Assembling the Test Items***

There are guidelines for item review, editing, arrangement, writing direction, and reproduction.

1. Review the items (think about suitability of the items, issues of bias, clarity, alignment with learning outcomes, appropriate difficulty, suitability of vocabulary).
2. Edit the items (appropriateness of wording, spelling, and grammatical errors).
3. Arrange the items (according to item format, order of complexity of learning outcomes).
4. Prepare clear directions (provide basis for selecting or supplying the answers, specify duration of test, provide procedure for recording responses, and indicate whether or not guessing and wrong spelling will be penalised).
5. Reproduce the test (present options in vertical columns beneath stem of each item, number items serially, items should be properly spaced, ensure that graphs, maps, charts are accurate and clear, and reproduce after proofreading).

***Activity 3.6***

1. Develop a table of specifications for any course you teach.
2. Using the table of specifications, you developed in (Q1), set any one multiple-choice question and one restricted response essay question at each of the cognitive levels of objectives.

### 3.2.5 Summary

In this Unit, we have examined different classifications of essay and objective test items for use in classroom assessment. We have learnt the features of essay and objective tests items. We have also discussed the

procedure for construction of achievement tests. The general guidelines for construction of good quality essay and objective test items were also listed.

## Unit 3.3: Designing Oral Assessment

### 3.3.0 Introduction

Many sciences and humanities programmes require students to complete their programmes by taking a practical, oral or with written examinations to show how well students have understood the materials studied in the programme and what they can do. Practical and oral tests are used for assessment in areas of learning that emphasise development of speech, perceptual, manipulative, and motor skills. They are emphasised in performance aspects in science, physical and health education, creative and theatre arts, music, fine and applied arts, teaching practice, agriculture, engineering, nursing, and medicine, among others. You need to be familiar with the oral and practical tests because of the global need to produce graduates who possess requisite skills that are central to the world of work and professional practice; a world where oral communication and practical skills tend to dominate. In this Unit, you will learn how to construct tests for assessing oral and practical skills.

### 3.3.1 Expected Learning Outcomes

By the end of this Unit, you should be able to:

- describe oral assessment.
- state the importance of oral assessment.
- describe the procedure for designing oral assessment.

### 3.3.2 Meaning of Oral Assessments

An oral assessment is a direct procedure of assessing students' learning outcomes using oral questions. It is designed to measure students' attainment of learning outcomes by questioning them orally. Oral tests

do not usually have a structured list of questions, but the assessors ask questions that elicit oral responses (Assessment Resource Centre @HKU, 2014). Oral tests otherwise known as oral examination or viva voce is described as a practice in many institutions and disciplines where an examiner poses questions to a student in spoken form (Wiki, 2019).

You might have used oral test in student evaluation or have taken one in the process of your professional or career development. Can you recall your personal experience of oral examination? Oral examinations are used for assessment in different context such as in:

a. Medical, nursing, and health professions, where students are presented with series of patients and asked to discuss diagnosis and treatment plans with an examiner;
b. Graduate and postgraduate programmes before the award of Ph.D. or Master's degree;
c. Oral presentations by students in almost all disciplines to their classes either individually or in groups;
d. Assessment of language skills whether in learning foreign language or as part of the development of oral communication skills and competences;
e. Assessment of field practice of students preparing for different careers such as teaching practice by education students, interviewing of clients by psychology students, and reporting on field work by other categories of students;
f. It is used in schools as an alternative to written examination for students with a learning disability such as dysgraphia, developmental coordination disorder, or non-verbal learning disorder.

### 3.3.3 Importance of Oral Assessment

The importance of oral tests is discussed in Joughin (n.d.) and includes the fact that:

a. Globally, universities are expected to develop in their graduate's oral communication skills that are required in the world of work and professional practice.
b. Theories of learning emphasise the need for students to articulate their ideas and present same to peers and teachers through speaking, as well as develop the ability to communicate with confidence in different work environments.
c. Oral assessment helps to build confidence so that presentations by students are not plagiarised.
d. The assessment enables the use of oral processes to judge knowledge, understanding, problem solving and other abilities through oral medium.
e. It assesses attributes that require observation and interaction such as interpersonal skills, ability to think on the spot, respond to changes in situation, assume appropriate role and perform the role.
f. It assesses components of competence.
g. It assesses personal characteristics such as appearance and manner, competence, alertness, and honesty in responding, capacity for empathy and sensitivity.
h. Students are few in number who can be tested individually.
i. Examiners are properly oriented as to the testing situation and are available for sufficiently long periods of time.
j. It is used when learning outcomes demand it.
k. It helps to secure academic integrity.
l. It reflects the world of practice.
m. It allows probing of the students' knowledge and understanding.

***Activity 3.7***

What are the pros and cons of using oral examination in higher education in Nigeria?

### 3.3.4 How to Design a Good Oral Assessment

Each form of oral assessment has its unique qualities, but you need to plan and design your test to achieve the desired purpose of measuring expected student learning outcomes. The following guidelines will be useful for designing a valid and reliable oral test:

a. Clearly define areas of competence and traits to be assessed in measurable terms and ensure that students know the objectives of the assessment.
b. Provide students information on time period, location, guidelines, requirements, assessment criteria, and the assessor.
c. Provide a structured scoring guide for all the assessors.
d. Give students sufficient time to answer the questions.
e. Incorporate oral assessment into the teaching-learning process to familiarise students with the process.
f. Standardise the examination tasks by making them consistent, equivalent and comparable for all students.
g. Use more than one examiner to confirm excellent and poor performances during the assessment.
h. Use more problem-solving open-ended questions for assessing professional skills needed by the examinees.

***Activity 3.8***

1. Identify any four scenarios in which oral assessment of your students can take place.
2. Of what importance is oral assessment in evaluating student learning outcomes?

### 3.3.5 Summary

In this Unit, you have read that oral assessment deals with any assessment of student learning using oral questions. Oral assessment

deals with any assessment based on the spoken word, which includes oral presentations, viva, mock court hearings by law students, clinical examinations by medical and nursing students, among others. The importance of oral assessment was identified, and suggestions for designing good oral assessment were made.

## Unit 3.4 : Designing a Practical Test

### 3.4.0 Introduction

Many of the skills students learn in school involve some type of performance such as oral report, speech making, operating equipment, conducting laboratory experiment, performing athletic skills, music, dance drama, among others. These skills can be measured, and grades assigned as you do for skills measured by objective, essay and product development tests (Carey, 1988). In this Unit, you will learn the concept and purpose of practical test, as well as how to design a practical test in the course you teach.

### 3.4.1.Expected Learning Outcomes

By the end of this Unit, you should be able to:

- define the concept of practical test.
- describe types of practical test.
- explain how to design a simple practical test in your course.

### 3.4.2 Concept of Practical Test

A practical test is a test that assesses practical skills and techniques in laboratory, clinical or field settings, which are usually administered individually or in small groups (The Learning Centre, n.d.). They are performance tests that measure motor and manipulative skills. Practical tests are important in assessing competency-based courses such as typing, physical education, home economics, computer studies, agriculture, engineering, medicine, teaching, nursing, fine art, music, theatre arts, among others. They measure what the students can do and

require students to perform some type of action in a practical or field setting.

Race (1999) describes practical skills as a wide range of skills from deftness in assembly to speed and precision, which are measured in laboratories or workshops or real-life situations. It was further observed that the credibility of the courses we teach often depends on what students can do when in a workplace. You, therefore, need to use practical tests in courses where laboratory or studio work, use of equipment, or the development of manipulative or other psychomotor skills, are important part of a course. Practical tests are used in performance assessment. According to Oosterhof (1990), performance assessment is limited to situations where specific behaviour or outcomes of behaviour are observed, where it is possible to judge the appropriateness of student's actions, and a situation where process or outcome cannot be directly measured using paper-and-pencil tests.

Practical tests, therefore, should measure both the process and product of the skill being assessed. A process refers to the procedure a student adopts in completing a task while a product is a tangible outcome that results from completing a process (Ugodulunwa, 2008). An example of a process can be the way a fine art student draws and paints a picture, while the product is the picture produced. It is important to assess both process and product to get a complete picture of the skills possessed by a student and what he/she can do.

Basically, practical tests are used for assessment of objectives in the psychomotor domain, which deal with manipulative and motor skills. The taxonomy of educational objectives in the psychomotor domain was developed by Harrow in 1972. The domain consists of the following six levels and subcategories:

1. Reflex movement.
2. Basic fundamental movements (locomotor, non-locomotor and manipulative movements).

3. Perceptual abilities (kinaesthetic, visual, auditory, and tactile discrimination).
4. Physical abilities (endurance, strength, flexibility, and agility).
5. Skilled movement (simple, compound, and complex adaptive skills).
6. Non-discursive communication (expressive and interpretive movements).

Examples of general educational objectives you state in the psychomotor domain are: students play piano skillfully, demonstrate correct swimming style, manipulate tools properly, cut garments correctly, draw diagrams or objects, and drive automobile skillfully.

### 3.4.3 Types of Practical Tests

Practical tests can take the form of continuous assessment of practical or studio work, formal practical work, and skill tests. These tests are regarded as the most important types of practical test in higher education. Let us examine each of these practical tests.

***Continuous Assessment of Practical Work***

Ongoing or continuous assessment of practical or studio work is required in assessment of courses that have high laboratory or studio work content, such as in science, fine art, and architecture. The assessment is based on regular marking of laboratory reports and products. This type of practical test provides opportunity for monitoring students' learning progress and obtaining overall picture of their knowledge and ability. However, it is tedious and time consuming for teachers to provide continuous feedback to students, particularly in large classes.

***Practical or Studio Examinations***

Practical or studio examination is a one-shot examination of laboratory or studio work at the end of a semester or course. The use of this type of examination alone does not provide formative feedback for improving student learning. It is necessary for you to use both the continuous

assessment of practical work and the one-off end of semester or course practical examination in order to enhance student learning.

***Skill Tests***

These are tests for assessing students' ability to carry out different tasks that require manipulative and motor skills. Example of tasks that require skill tests are the ones found in engineering, home economics, music, fine art, agriculture, medicine, nursing, science and vocational courses. Skill tests are used in testing students' ability to carry out specific manipulative task such as, repairing a piece of technical equipment, identifying computer problems in an office and fixing them, changing hospital bed sheets while the bed was occupied, designing products or experiments, demonstrating physical activities, preparing different dishes, laying a table, administering injections, skills in athletics and games, driving tractors, writing, teaching skills, playing different musical instruments, drawing and painting, among others.

***Activity 3.9***

1. Describe any three types of practical tests that can be conducted in any of the courses you teach.
2. How will you design a simple practical test in any of the courses you teach?

### 3.4.4 How to Design a Simple Practical Test

It is clear to you now that you can measure and assign grades to differences in performances of students just as you do for other behaviour measured by objective and essay tests. The following procedure suggested by Carey (1988) can be used for designing and scoring practical tests:

a. Define and analyse the expected performance (that is, operationalising the skills to be measured and listing their components in observable terms and sequence).
b. Provide the relevant materials and equipment, laboratory, studio, field or simulated setting for use in demonstrating skills.
c. Develop an evaluation form for assessing process and products.
d. Observe and rate individual student's performance.
e. Formatively evaluate students' performance and identify areas requiring additional instruction.
f. Summatively evaluate products, where applicable.
g. Summarise students' scores for final grading.

### 3.4.5 Development of a Performance Rating Form

The following steps can be adopted in developing a performance rating form:

a. Identify the instructional skills and their sequence.
b. Select the skills you believe will be observable.
c. List the observable skills in order that you believe they will occur.
d. Watch different students who are proficient in the skills and compare. the skills on your list with those you are able to observe.
e. Eliminate skills you do not see and add any others you identified during the performance that are not on your own list.
f. Check the sequence of the components of the form while the students are performing the actions and reorder them if necessary.
g. Paraphrase your verbal descriptions of each component into one or two key words after sequencing the components of the form.

### 3.4.6 Guidelines for Writing Instructions for Practical Tests

Many practical tests require students to demonstrate the skills in the same sequence and with the same equipment they used during instruction and rehearsal.

The following guidelines should be borne in mind when writing instructions for a practical test:

a. The task described should match the behaviour, content, and conditions in the course objective.
b. Language should be clear, precise, and appropriate for the level of the students.
c. The skills should not be too complex for the students.
d. The conditions of performance should be the same for all the students.
e. Testing situation should require the same equipment, content, sequence and conditions in the instructions.

### 3.4.7 Summary

A performance test requires students to demonstrate skills through performance of actions. Practical tests have different varieties, which can be in the form of ongoing assessment of practical work, on-shot examination of laboratory or studio work at the end of a course, and skill tests that assess students' ability to carry out different tasks requiring manipulative and motor skills found in different professional and vocational courses, as well as suggestions for designing them.

## Unit 3.5: Designing A Portfolio

### 3.5.0 Introduction

Portfolio is an alternative to traditional assessment of student progress. The use of portfolio assessment has been recommended as the best alternative for achieving an individualised, student-centred evaluation process. Portfolios are designed to show evidence of efforts and correction of errors, which indicate progress and achievements of student learning. Thus, portfolio assessment is regarded as a comprehensive, systematic, informative, tailored and authentic instrument by Moya and O'Malley (2009). In this Unit, you will examine the concept of portfolio,

purpose and types of portfolio assessment. Finally, you will learn the innovative uses of portfolio and steps in designing portfolio assessment.

### 3.5.1 Expected Learning Outcomes

By the end of this unit, you should be able to:

1. define portfolio and state its purpose.
2. identify different types of portfolios.
3. describe the procedure and guidelines for designing portfolio.

### 3.5.2 Meaning and Purpose of Portfolio

***Concept of a Portfolio***

Portfolio is defined by Paulson, Paulson and Meyer (1991) as a purposeful collection of student work which reveals the student's efforts, progress and achievements in different areas. The collection in the portfolio should include student participation in selecting contents, the criteria for selection, the criteria for judging merit and evidence of student self-reflection. Two important terms in this definition is reflection and learning objectives. By reflection on their work, students are able to identify the strengths and weaknesses of their own work; while relation to learning objectives is assessed with reference to its specific instructional objectives. Reflections should focus on what a student has done and learnt in the past, what he/she needs to learn next, as well as elements of the instructional objectives students should address. A portfolio, is therefore, not a random collection of student products or observations, but systematic because they are related to the instructional objectives. It contains samples of student work which are related to learning objectives and shows learning growth over time.

***Purposes of Portfolio***

The purposes of portfolio are to:

a. provide students the opportunity to reflect on their growth over a period of time.

b. use it as a basis for grading students based on their efforts,
c. use it as a basis of communication with parents, peer and others.
d. monitor student learning progress and enhance student learning.
e. document and showcase students' work samples.
f. evaluate specified learning objectives or outcomes.

### 3.5.3 Types of Portfolios

Three types of portfolios identified by Danielson and Abrutyn (1997) are briefly described here. They are working portfolios, display, showcase or best works portfolios, and assessment portfolios. Working portfolios are so named because they contain work in progress as well as samples of finished work. Its major purpose is to serve as a tank, holding work for the students. Display of students' best works is the most rewarding use of portfolios to students because the exhibition and interpretation of their works make them proud. Display portfolio can be maintained from year to year, documenting growth over time. An assessment portfolio documents what a student has learned on specific curriculum outcomes. Their reflective comments focus on their judgment of the extent they believe the portfolio entries demonstrate their mastery of curriculum objectives.

Portfolio can be used to assess different content areas to indicate proficiency in various disciplines. Entries may indicate students' areas of strengths and weaknesses. An interdisciplinary unit on longitude and latitude, for example, could give rise to a portfolio containing samples of a student's accomplishment in Mathematics, geography, general studies, science, and art. Innovative uses of portfolios abound. A French language portfolio could have entries such as cultural artifacts relating to religion, art, and celebrations, as well as spoken and written proficiency in the language, while an art portfolio can contain best student performance in the arts, with drawings, slides, and examples of music composed and performed. A science portfolio could contain charts and

graphs created, projects and posters, laboratory reports, research reports, tests, and student reflections; while a Mathematics portfolio could contain samples of problem solving, written explanations of how to solve problems, charts and graphs, computer analysis reports, and student reflections that are done either weekly, monthly or bi-monthly, and criteria for evaluation.

### 3.5.4 Procedure and Guidelines for Designing Portfolio

**Procedure**

To develop a portfolio assessment, the teacher should reflect on the following guidelines identified by Martin-Kniep, (2000), Fumiss (2002) and Mueller, (2011):

a. Decide on the purpose and audience for the portfolio.
b. Identify the learning outcomes the portfolio is meant to cover.
c. Start with small outcomes and remain focused.
d. Inform parents of the use of portfolio assessment technique and the schedule or time for assigning students' work.
e. Spell out the students' engagements during the development of the portfolio. This entails that both teacher and students should work together on what to engage in and what to consider as mastery.
f. Create your own portfolio and use it as a model for portfolio. This demands that as teachers work on the portfolio, the processes they engage in and what they produce at the end should be properly documented to serve as models for future use.
g. Begin the assessment of a portfolio from the time of implementation and end at the evaluation of products. Davis and Ponnamperuma (2005) identified the following five stages in portfolio implementation and assessment: collection of evidence of learning achievement, evaluation of evidence, defence of evidence, and assessment decision. This allows the portfolio implementers to make use of clearly defined pre-validated rating scales to assess the students' best samples.

***Guidelines for Portfolio Entries***

Some guidelines for making entries in students' portfolio include:

a. Determine the purpose of the portfolio;
b. Specify time duration;
c. Describe types of work to include in the portfolio;
d. Set criteria to use for evaluating the portfolio;
e. Allow flexibility in use of different types of portfolios;
f. Ensure that students have access to resources for building their portfolios;
g. Provide ground rules for working independently or collaboratively;
h. Share the rubric with students before they work on the portfolio;
i. Allow students time to reflect on their portfolio using the rubric;
j. Rate portfolio without looking at the student's name;
k. Allow students to select the entries in the portfolio;
l. Use analytical rating scale for developmental portfolio;
m. Use holistic rating scale for a showcase portfolio;
n. Check for bias by using more than one rater.

***Activity 3.10***

1. Identify one learning objective in your course which can be assessed with portfolio assessment technique and state what should be the entries in the student portfolio.
2. What advice will you give your colleague who wants to design portfolio for assessing his/her student learning?

### 3.5.5 Summary

Portfolio assessment is a kind of authentic assessment where students work together with their teachers. Portfolio contains samples of student's work and shows development of the student over time. You can use it for diagnosing, documenting or showcasing student learning. The entries in a portfolio should, among other things, include clearly set-out criteria for

judging merits, evidence of student's involvement in selection of contents and students' self-reflection. Portfolio should reflect classroom activities, instruction and assessment in order to be part of day-to-day instruction. Since portfolio is one of the innovative forms of assessment, you need to adopt it as one of your assessment strategies following the suggested procedures and guidelines.

## Unit 3.6: Refining Achievement Test Item Qualities

### 3.6.0 Introduction

After item writing, there is need for review to ensure that the items generated are correct, fair, valid and reliable. So, item review process should include item vetting or review for quality assurance. In this Unit, you will learn different techniques you can use in reviewing your test items.

### 3.6.1 Expected Learning Outcomes

By the end of this Unit, you should be able to:

a. define item vetting and state its purpose.
b. describe the components of item vetting process.

### 3.6.2 Meaning of Item Vetting

Item vetting is defined by Gopalakrishnan and Udayshankar (2014) as the process of reviewing and evaluating items according to specified guidelines for the purpose of detecting flaws that constitute threats to quality of examinations. It is the process where test constructors examine or evaluate and edit generated items to make them free of any mistakes. This step in item writing process is considered important for assuring content validity of test items and for minimising the threats to quality of examinations.

### 3.6.3 Item Vetting Process

The areas of focus during item vetting are technical accuracy, content coverage and language aspects. The use of the prescribed criteria for item vetting ensures a consistently high standard of question presentation during examinations which will help the students to understand the questions better and to answer them correctly as observed by Gopalakrishnan and Udayshankar (2014). The following item vetting criteria and the rules of item writing presented earlier can be used by any committee assigned to validate examination questions during the item writing process:

***Technical Aspects***

The technical aspects of item vetting should focus on obtaining positive responses to the following questions:

a. Does each item comply with the prescribed format and guidelines for the item generation?
b. Is the time provided appropriate for the tasks or demands of the items?
c. Are all the keys correct answers to the item?
d. Are the total marks allotted for each question type appropriate?
e. Is there any overlap of exact content being tested in the questions within and between question types?

***Content Aspects***

The content aspects look into the relevance of the question to topics covered. The distribution of questions should be in relation to the content and objectives specified in the table of specifications or test blueprint, such as, whether or not the content being tested is within the curriculum, content is free of bias (cultural, gender, religious, etc.), difficulty level of each item is within the ability level of the students, items are arranged in order of difficulty from easy to difficult, and whether mix of skills tested by the items range from factual to understanding, application, and other higher level objectives.

***Language Aspects***

The language aspects focus on whether or not the terminology used is the current version and punctuation marks are correct and proper, language used is simple, clear, direct and appropriate for the examinees, whether or not use of tenses is consistent and subject-verb agreement is observed. They also focus on whether or not appropriate 'action' verbs are used (e.g., indicate, show, illustrate, give, and explain) for each level of objectives, items are devoid of spelling and typography errors as well as grammatical mistakes and whether spelling norms (British/American orthography) are standardised.

***Trial Testing of the Items***

The trial testing or field testing is a stage in standardised test development process, which involves the selection of a representative sample, administration of the test, scoring of the test, collection and analyses of context information from teachers. The sample should consist of subgroups such as male and female, rural and urban, low and high socioeconomic status, public and private schools. Other characteristics are cultural groups of students, state, age, ability, and geographic region. The test needs to be administered by teachers in selected classrooms following the prescribed procedure for administration. The test would also be scored following the prescribed procedure for scoring before conducting item analysis to determine the functionality of the test items. The trial testing is usually conducted by the Test Development Unit of different examination bodies for item calibration and banking.

Item analysis involves the analysis of individual items that are in the test after the items have been subjected to statistical analysis. The items that passed the analysis are selected and included in the final form of the test while those that failed are modified or discarded and tried out again in subsequent field-testing exercise. The two measurement or test theories

in use for item analysis in test development process are the Classical Test Theory (CTT) and the Item Response Theory (IRT).
The use of the two frameworks for item analysis is beyond the scope of this training workshop.

***Activity 3.11***

1. What strategies do you adopt to assure quality of tests you use in evaluating students' achievement in your course?
2. What will you do differently as a result of this workshop?

### 3.6.4 Summary

This Unit described the different techniques for refining achievement test items. Item vetting was defined and the three criteria for item vetting that were identified were described.

## References


Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching and assessment: A revision of Bloom's taxonomy of educational objectives.* Boston, MA: Allyn and Bacon.

Assessment Resource Centre @ HKU (2014). *Assessment Methods*. Retrieved from https://ar.cetl.hku.hk/am_orals.htm

Bloom, B. S. & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals by a committee of college and university examiners handbook I: Cognitive domain.* NY, NY: Longman, Green.

Danielson, C. & Abrutyn, L. (1997). *Chapter 1. The types of portfolios.*

Retrieved from https://www.itslearning.com/help/en-US/Content/Courses/portfolio assessment.htm

Davis. M. H. & Ponnamperuma, G. G. (2005). *Portfolio assessment.* Retrieved from http://www.researchgate.net/publication/7506529 portfolioassessment

De Valenzuela, J. S. (2002). *Defining Portfolio Assessment*. Retrieved from http://www.unm.edu/-devalenz/handouts/portfolio/html

Fumiss, E. (2002). *Assessing Learning Achievement* - UNICEF. Retrieved from http://www.Uniceforg/lifeskills/Assessing LearningAchievement.doc

Gopalakrishnan, S., & Udoyshankar, P. M. (2014).Question vetting: The process to ensure quality in assessment of medical students. *Journal of Clinical and Diagnostic Research, 8(9).* Doi:10.7860/JCDR/2014/9914.4793

Hassan,S.(n.d.) Retrieved January1.8.2018 from https://www.slideshare.net/DrShahidHassan/blueprint-of-exam-questions

Martin-Kniep, G. O. (2000). *Becoming A Better Teacher: Eight Innovations That Work.* U.S.A: ASCD.

Moya, S. S., & O'Malley, J. M. (2009). M *Portfolio Assessment Model For ESL.* Retrieved: http://www.ncela.gwu.edu/files/rcd/BE09739/APortfolioAssessment.pdf

Mueller, J. (2011). *Authentic Assessment Toolbox*. Retrieved from http://ifmueller.facultv. noctril./toolbox/portfolios.htm

Nworgu, B. G. (2015). *Educational Measurement and Evaluation: Theory and Practice* (2nd ed.). Enugu: University Trust Publishers.

Paulson, F. L., Paulson, P. R., & Meyer, C. A. ( 1991 ). *What makes a portfolio a portfolio?* Retrieved f r o m https://web.stanford.edu/dept/SUSE/projects/ireport/articles/eportfolio/what

Ugodulunwa, C. A. (2008). *Fundamentals of Educational Measurement and Evaluation.* Jos: Fab Educational Books.

Wikipedia (2019). *Oral Exam*. Retrieved from https://en.wikipedia.org/wiki/Oral_exam

## MODULE FOUR

## ISSUES OF QUALITY AND CONFIDENCE IN ASSESSMENT

**by**
**Christiana Agi**

### 4.0 Module Overview

Generally, in educational assessment, we base decisions about students' progress on data. Where do we get such data? The data may come from classroom and standardised tests in form of scores. Scores obtained from these tests are however influenced by errors. An obtained score may either be overestimation or underestimation of testees' true score. Hence, it is important that in using such data for decision making, quality should be ensured by minimising errors as much as possible for confidence in test results. The quality of a test is ascertained in terms of validity, reliability, objectivity, fairness and usability to achieve a result that is dependable. This module will be discussed under five units as follows: Validity, Reliability, Objectivity, Fairness and Usability.

### Unit 4.1: Validity

### 4.1.0 Introduction

As teachers and instructors, whenever validity is mentioned, what is it that comes into your mind? With the mention of validity, is your attention directed towards research or project work? If your answer is in the affirmative, you are actually thinking in the same way I used to think when I was newly-employed as a teacher. In education, however, validity cannot be tied to research work alone. As an instructor, whenever you want to make inferences and decisions on some types of data, you should be concerned about the quality of such data. This is because if inferences and decisions concerning the students' learning outcomes are based on an invalid data, it will have great consequences

on the learners' academic attainments. For instance, we might use data to make inferences from the grades the student is assigned; meanwhile, such a student has not mastered the concept or content. In the end, this could have consequences for course completion, continuation of study in the degree, and graduation. It is therefore important for instructors to be sure that those inferences are sound, and that the decisions made following these inferences are well supported by valid data. In achieving this, the teacher should ensure that his or her assessment instruments or tests are valid. What then is validity? In this Unit, we shall be discussing the concept of validity, kinds of validity estimate, factors affecting validity and ways of ensuring that a test is valid.

### 4.1.1 Expected Learning Outcomes

At the end of this Unit, you are expected to be able to:

- define test validity and explain its importance in constructing assessment instruments.
- describe the types of validity estimate.
- identify the factors affecting the validity of a test.
- state at least four (4) ways of ensuring a test is valid.

### 4.1.2 Concept of Validity

When we talk about validity, we are referring to the extent to which a given instrument measures what it claims to have measured. It is the dependability or the soundness of the interpretations and use of assessment results (Afemikhe, 2014).

Validity basically, is concerned with the specific use to be made of test result. This means that validity is concerned not with the instrument in itself but with the inferences based on the score obtained from the instrument and it is always specific to some particular use. For instance, as a teacher, if your intention is to measure spelling ability in English language, your test is valid if it successfully measures spelling ability

only, and no other abilities. This implies that a test whose validity is established in reference to a specific purpose may not be valid for other purposes. Also, a test you used to make valid predictions about a student's spelling ability may not be valid for predicting his or her mathematical ability.

Another way to look at validity of a test is that, if it is not based on what the teacher has taught, such a test is not valid. As a teacher, if you taught your class multiplication of 3-digit numbers, at the end or during your lesson, your evaluative objective would have been to ascertain whether your class would be able to multiply 3-digit numbers correctly or not. If you are evaluating your lesson, you have to construct a test whose purpose will be to measure your students' ability to multiply 3-digit numbers and nothing else. In a situation where your assessment is asking the students to solve arithmetic problems involving addition of 3-digit numbers, the test will no longer be valid. Validity requires that a test fully assesses every aspect of a domain it claims to assess.

From the forgoing therefore, because a test may be used for different purposes, there is no single validity index for a test. This means that a test that has some validity for one purpose may be invalid for another.

### 4.1.3 Kinds of Validity

There are basically, four kinds of validity. These are Face Validity, Content-related Validity, Criterion-related Validity, and Construct Validity as illustrated in figure 4.1

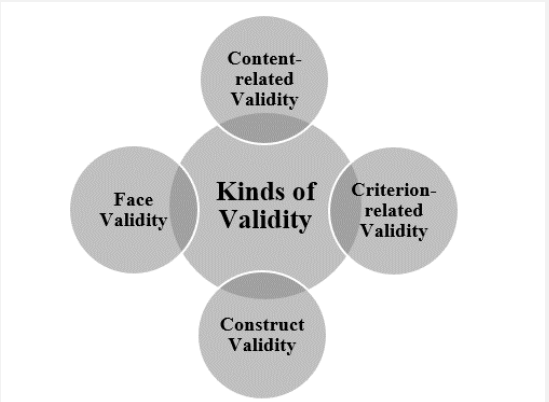


Fig. 4.1: Illustration of Kinds of Validity

***Face Validity***

This type of validity refers to the extent to which the test appears relevant, important and interesting to the testees and other people. As a teacher, whenever you give your colleagues your test instrument or to an external observer to look at it in terms of how clear the instructions are, how well structured the items are, the consistency of numbering etc, you are carrying out face validity. It is called face validity because the external observer is only concerned with the appearance of the items, language of testing, printing, binding, length and other peripheral details of the test. Technically, face validity is not really validity because this type of validity depends on the judgment of the external observers. We shall be discussing content validity in the next section.

***Content-related Validity***

Before we discuss content-related validity, let us reflect on this question. Have you ever experienced a situation where you prepared for a test, having been previously exposed to some contents and when the questions were finally presented, you found out that they were completely different from the content you were taught or as a teacher, you fail to cover the specified contents and went ahead to set questions covering all the topics in the content? Such a test has poor content

validity. As teachers, there is usually content (topics) that must be covered in a course within a stipulated period which could be a term or a session. We should remember as teachers that some topics are broader than others as such the amount of time used in teaching them varies, as a result of this, items of an instrument should be constructed in proportion to how broad and how much time was spent in teaching the topic. To ascertain that the content covered in a test is the representative of the content presented in the course, the term content validity is used. Generally, as a teacher, if items on your test instrument cover topics that were not taught in that course, or that you ignore certain important aspects and unduly give emphasis to other aspect compared with their treatment in the course, the content validity of such test will be low. Content validity of an achievement test can be achieved with the aid of a table of specification/test blueprint and this was explained and illustrated in module 3.

**Criterion-related Validity**

What is criterion-related validity? It is the degree to which test scores correlate with, predict, or inform decisions regarding another measure or outcome? How can you use test scores in predicting a student's future performance on a given task? Let us look at this example. As a teacher, you may be interested in predicting the students' scores in a West African Examinations Council (WAEC) Senior Secondary Certificate Examination (SSCE), you then subject your students to a Mock examination, the WAEC result is the criterion variable while the Mock SSCE result is the predicator variable. Please, let us note that the keyword in the definition of criterion validity is 'correlate' which is synonymous with “relate” or “correspond”. The correlation is estimated in one of two ways: concurrent-related validity and predictive related validity. Concurrent-related validity is the type of validity that both the current score and current criterion can be obtained at the same time and correlated to obtain the coefficient of concurrent validity. For instance, a teacher can use students' scores on knowledge of mathematical tables to

estimate their current use of knowledge of mathematical tables in solving multiplication of numbers. In predictive validity, scores on the criterion are collected in the future compared to collection of scores on the predictor. Predictive validity refers to the degree to which a test accurately predicts the future performance. For example, a student who does well in Mathematics aptitude test can be predicted to be able to undergo a Physics course successfully. If the student eventually does well in that physics course, we can say that the predictive validity is achieved.

***Construct Validity***

Construct validity is the degree to which a particular test can be shown to measure a hypothetical construct. It is the extent to which a test measures the construct or characteristic it claims to measure. A construct here is seen as a trait which is covert (not openly acknowledged) but can be inferred on the basis of their observable effects on the behaviours. Examples of constructs are intelligence, motivation, assertiveness, creativity, anxiety, and aptitude, among others. It is important to note that no test measures a construct directly but they measure behaviours which reflect the construct. Construct validity often pertains to tests that may measure abstract traits of a person. Consider the activity below:

**Activity 4.1**

What will you do as a teacher, if your interest is to test students' numerical aptitude to show the test possess construct validity? Pick a pen and paper and state in two sentences what to expect.

**Answer:** You need two things:

1. that the test indeed measures the desired trait, and
2. that this trait corresponds to success in Mathematics.

Construct validity of an instrument can be established using different methods as illustrated in figure 4.2 below:

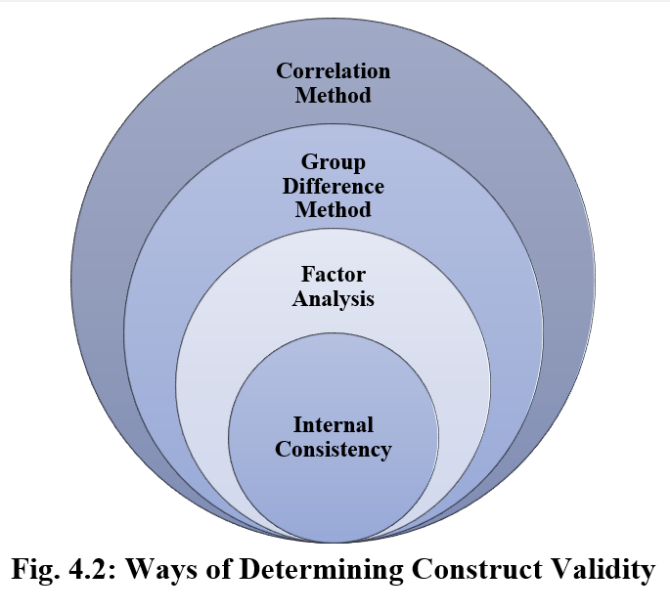



**Fig. 4.2: Ways of Determining Construct Validity**

As shown, there are different methods to establish construct validity of a test. These are explained below:

**Correlation Method:** This involves the correlation of testee's scores in a new test and scores of the testees in an earlier known test measuring the same construct.

**Group Difference Method:** This is used when the test performance of a group that exhibits the trait being measured is compared with another group that does not have the trait. For example, a primary school readiness test may be administered on a group of primary school entrants who have been classified as possessing school readiness skills and another group that has been classified as lacking school readiness skills, then the performances of the two different groups will be compared.

**Factor Analysis:** This is a sophisticated technique for establishing the construct validity of a test. It is a statistical technique used to reduce a large number of variables into a smaller number of factors or to group a test into its subtests.

**Internal Consistency:** This method involves the correlation of a subtest scores with the total test scores. It measures the test homogeneity, reason being that it compares the subtest scores or items scores with the entire test scores.

### 4.1.4 Factors Affecting Validity

What do you think is capable of affecting the validity of your instrument? A plethora of factors affect the validity of a test. These factors are classified into: factors inherent in the test itself, factors in test administration and scoring, and factors related to the testees .

***Factors inherent in the test:*** These factors include:

- Unclear directions to the students to respond to the test;
- Difficulty of the reading vocabulary and sentence structure;
- Too easy or too difficult test items;
- Ambiguous statements in the test items;
- Inappropriate test items for measuring a particular outcome;
- Inadequate time provided to take the test;
- Length of the test being too short;
- Test items not arranged in order of difficulty; and
- Identifiable pattern of answers.

***Factors in Test Administration and Scoring:***

- Unfair aid to individual students who ask for help;
- Unreliable scoring of essay type answers;
- Insufficient time to complete the test; and
- Adverse physical condition of the testing environment.

***Factors related to testees:***

- Test anxiety of the students;
- Physical and psychological state of the pupils;
- Response set - a consistent tendency to follow a certain pattern in responding to the items; and

- Cheating by the pupils during testing.

### 4.1.5 How to Ensure Validity of a Test

For a teacher to ensure the validity of a test, the following few steps should be followed while constructing the test:

1. The teacher should state the purpose of the test clearly and based same on the instructional objectives.
2. The teacher should ensure that items constituting the test should come from the topics that had been taught to the students.
3. The relative emphasis given to each topic during instruction should be clearly stated. This may be determined from the amount of time spent in teaching a topic, and the importance or coverage given to it in the syllabus. This will help to give weightings to the topics in percentages.
4. The number of questions or items to be developed from each topic should be determined based on the relative percentage weightings in order to ensure that the test faithfully represents the entire content taught.

***Activity 4.2***

1. Explain the meaning of validity.
2. Describe three kinds of validity evidence.
3. What factors affect validity of a test?

### 4.1.6 Summary

1. Validity is the extent to which the test measures what it claims to measure.
2. Validity is not of different types but a unitary concept based on various kinds of evidence.

3. The different kinds of validity evidence are content-related validity evidence, criterion related validity evidence and construct validity evidence.
4. Content validity evidence is concerned with the degree to which the sample of test items represent the content that the test is designed to measure.
5. Criterion-related validity evidence is the degree to which test scores correlate with, predict, or inform decisions regarding another measure or outcome.
6. Construct validity is the extent to which a test measures the construct or characteristic it claims to measure.
7. Face validity refers to the extent to which the test appears relevant, important and interesting to the testees' and other people (test expert).
8. Factors such as unclear directions to the students to respond to the test, difficulty of the reading vocabulary and sentence structure, too easy or too difficult test items, ambiguous statements in the test items, unfair aid to individual students who ask for help, cheating by the pupils during testing, insufficient time to complete the test, test anxiety of the students and response set, among others, affects validity of a test.

In the next Unit we shall be discussing reliability of a test

## Unit 4.2: Reliability

### 4.2.0 Introduction

Most times we hear people say that someone is not reliable or you should not rely on a particular person. What that means is that the person cannot be trusted. It also means that any information given by the person at a particular time may not be true or dependable. Hence being reliable implies that what one says or does should be depended upon. Now, let us liken the above with the reliability of a measuring instrument. If a

measuring instrument is yielding inconsistent result upon different administration on the same or similar respondents, such instrument is not reliable and it cannot be depended upon because it will be misleading. In this Unit, we shall be discussing the concept of reliability, different kinds of reliability estimate, factors affecting reliability of an instrument and how to enhance the reliability of an instrument.

### 4.2.1 Expected Learning Outcomes

After completing your study of this Unit, you should be able to do the following:

- Explain the meaning of reliability.
- Describe the different kinds of reliability estimates.
- Identify the factors that affect the reliability of a test.
- Identify ways by which the reliability of a test can be enhanced.

### 4.2.2 The Concept of Reliability

In educational measurement, all the measuring instruments that are used in gathering information about the testees' are expected to be reliable. As indicated by Mehrens and Lehmann (1991), reliability is the degree of consistency between two measures of the same thing. When an instrument is administered over and over and it is yielding approximately the same result, such an instrument is said to be consistent and dependable. But when the instrument is consistently yielding result that is not approximately the same on different administrations on the same subjects (for example students), such instrument is not trustworthy and dependable, and therefore not reliable. For instance, Christiana in the course of measuring her weight climbed a weighing scale on three consecutive times and she weighed 76kg, 49kg and 91 kg respectively. Looking at these results, is the weighing scale (measuring instrument) giving a reliable result? The answer certainly is no. One thing to note here is to ensure that no factor comes between times of measurement to either increase or reduce the weight of Christiana. As a teacher, if you

have taken a measure on students' level of achievement in a particular subject, it is hoped that the scores you will obtain on that measure will be similar under different administrations using students with similar ability, or at different intervals. Anything short of this means that such an instrument is not reliable. There are different ways of estimating reliability and these will be handled in the next section.

### 4.2.3 Types of Reliability Estimates

Let us now compare the types that we have used with the ones we shall be discussing here. Methods of estimating reliability differs in that they consider different sources of error. These methods include: Test-retest method, Alternate or Equivalent method, Split-half method, Kuder-Richardson method, Cronbach Alpha method and Inter-rater reliability methods.

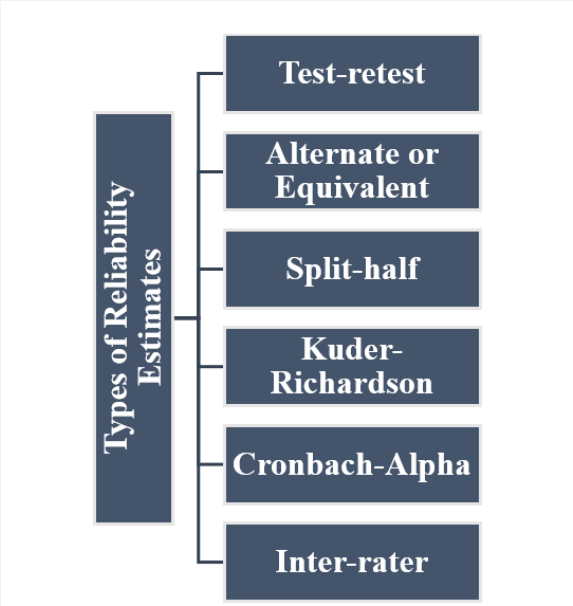



**Fig. 4.3: Types of Reliability Estimates**

***Test-Retest Method***

This method involves administering the test to the same group of testees at least two times. As teachers, if we want to estimate reliability of our test using test-retest method, we will have to administer the same test at least twice to the same group of testees with a given time interval between the administration. Then, we can go ahead to correlate the scores from the two administrations to check or measure the stability of the test. A teacher can use test re-test method to check if the result of a test is consistent over a period of time. For instance, a teacher who bought a Child Test Anxiety Scale (CTAS) to measure students' anxiety may find indices that the students have higher negative anxiety. This shows that the CTAS was able to accurately measure the anxiety level of the students. To be sure that the CTAS was a reliable measure, the teacher would re-administer the same test to the same subjects to check if the test will produce relatively the same result. The problem here will be that if we are not conscious of the time interval between the initial testing and retesting, reliability of such test will either be low or high. It is advisable that the time interval between the first and second administrations of the test should not be too long or too short. If it is too long, the result might be influenced by changes in the characteristics of the learners over the period of time. Again, if the time is too short, the students' might remember some of the answers from their first test to the second test and the reliability of the test will be affected (Emaikwu, 2005). Ugodulunwa (2008) suggested that, the time interval of the two administrations should be between two to three weeks. In estimating the reliability of the two set of scores, Product Moment Correlation Coefficient is often used. Such correlation coefficient is called coefficient of stability.

Lindstrom (2010) provided the following as a number of ways coefficients of stability can be interpreted:

- 0.9 and greater: excellent reliability
- Between 0.8 and 0.9: good reliability
- Between 0.7 and 0.8: acceptable reliability
- Between 0.6 and 0.7: questionable reliability
- Between 0.5 and 0.6: poor reliability
- Less than 0.5: unacceptable reliability

For example, let us use the table below to illustrate test-retest method of estimating reliability where X = first administration and Y = second administration.

**Table 4: Test-Retest Table**

| | | | | | |
|---|---|---|---|---|---|
| Test (X) | 40 | 25 | 18 | 38 | 27 |
| Retest (Y) | 34 | 20 | 18 | 34 | 35 |

Using Pearson Product Moment Correlation Coefficient with the following formula:

$$r = \frac{N\Sigma xy-(\Sigma x)(\Sigma y)}{\sqrt{[N\Sigma x^2-(\Sigma x)^2][N\Sigma y^2-(\Sigma y)^2]}}$$

**Table 5: Pearson's Product Moment Correlation Coefficient Table**

| Test (X) | Retest (Y) | $X^2$ | $Y^2$ | XY |
|---|---|---|---|---|
| 40 | 34 | 1600 | 1156 | 1360 |
| 25 | 20 | 625 | 400 | 500 |
| 18 | 18 | 324 | 324 | 324 |
| 38 | 34 | 1444 | 1156 | 1292 |
| 27 | 35 | 729 | 1225 | 945 |
| **148** | **141** | **4722** | **4261** | **4421** |

Now r = $\sum X = 148,\ \sum Y = 141, \sum X^2 = 4722, \sum Y^2 = 4261, \sum XY = 4421$

$$r = \frac{5 \times 4421 - (148)(141)}{\sqrt{[5 \times 4722 - (148)^2]}\sqrt{5 \times 4261 - (141)^2]}}$$

$$r = \frac{22105 - 20868}{\sqrt{[23610 - 21904]}\sqrt{[21305 - 19881]}}$$

$$r = \frac{1237}{\sqrt{1706}\sqrt{1424}}$$

$$r = \frac{1237}{\sqrt{2429344}}$$

$$r = \frac{1237}{1558.6}$$

r = 0.80

Table 5 shows that a teacher administered a test twice on the same group of respondents with a given time interval in order to measure the stability of the test. The test-retest result obtained was 0.80. This shows that the test-retest reliability result was high. Therefore, the reliability of the test can be interpreted as good.

***Equivalent or Alternate Method***

As a teacher if we do not want to administer the same test twice, we can develop two tests that are parallel. By parallel forms, we mean that the tests are equivalent so far as the content, objectives, format, difficulty and discriminating levels of items and length of the test among others are concerned. How are you going to administer parallel test? What you are expected to do is to administer the first form of the test to the students and on finishing immediately (or at least near), another form of a similar test should be administer to the same group. When the students are done writing the parallel tests, you will have to correlate the scores each student has obtained to estimate the reliability. Thus, the reliability found is called coefficient of equivalence. A high parallel form reliability coefficient indicates that the different forms of the test are very similar which means that it makes virtually no difference which version of the test a person takes; a teacher can use either version of the test. On the other hand, a teacher who obtained a low reliability coefficient of a

parallel test should know that the different forms are probably not comparable and may be measuring different things, therefore, cannot be used interchangeably. This type of reliability estimate is applied when a teacher has developed different versions of tools and decided to use them interchangeably. The calculation of the coefficient of correlation between scores from two tests is as demonstrated earlier under test-retest method.

***Split-Half Method***

In this method, the test is administered once. The test is administered to a group of testees and is then subdivided into two halves for scoring purpose. The most common method for forming the halves is to consider the odd-numbered items on the test as one form, and the even-numbered items as the other form. When the two halves have been formed, a score for each half is determined for each testee, thus giving two scores for each testee. A correlation of the two sets of scores is done using Pearson Product Moment Correlation formula and the resulting coefficient is called the coefficient of internal consistency. This correlation is based on halves of the test; the value obtained cannot be seen as the reliability of the full test. The reliability of the full test is obtained by stepping up the obtained reliability of the halves of the test with a formula referred to as Spearman Brown Prophecy which is illustrated thus:

$$\text{Reliability of the full test} = \frac{2 \times \text{Reliability of half test}}{1 + \text{Reliability of half test}}$$

For example, a teacher administered a test, and scores from the two halves of the test correlates to 0.64. The reliability estimate of the whole test would be:

$$Reliability = \frac{2r}{1 + r}$$

$$\frac{2 \times 0.64}{1 + 0.64} = \frac{1.3}{1.64} = 0.78$$

The value indicates that the reliability of the test is high and is acceptable.

***Kuder-Richardson Method***

In estimating the reliability of a test using this method, the test is administered once just as the split-half. However, this method does not require splitting the test into two halves and scoring each separately. It is very important for us to note here that, this procedure assesses the inter-item consistency or homogeneity of the items making up a test. A teacher can use either K-$R_{20}$ or K-$R_{21}$ to estimate reliability when items in the test are scored dichotomously (right or wrong answers). In this method, it is assumed that all items have same or equal difficulty value, correlation between the items are equal, all the items measure essentially the same ability and the test is homogeneous in nature.

**The formula for $P_{kr20}$ is**

$$K - R_{20} = \frac{n}{n-1}\left(1 - \frac{\sum_{j-1}^{n} P_j q_j}{\delta^2}\right)$$

*Where*

n = number of items in the test

$P_j$ = number of respondents in the sample who answered the items correctly

$q_j$ = number of respondents in the sample who answered the items incorrectly

$\delta^2$ = variance of the total scores of all respondents who took the test

**The formula for K-$R_{21}$ is**

$$K - R_{21} = \frac{n}{n-1}[1 - \frac{M(n-M)}{nSx^2}]$$

Where

$n$ = number of items in the test

$M$ = population mean of test scores

$S_{X^2}$ = variance of the total test

You may seek help in the application of these formula in determining reliability based on K-R.

***Cronbach Alpha Method***

This method provides reliability estimates for instruments or tests composed of items scored dichotomously and essay tests having items of varying point values or rating scales that provide responses on a continuum of three or more response options. The coefficient alpha can be computed using the formula:

$$\alpha = \frac{n}{n-1}\left\{\frac{Si^2}{Sx^2}\right\}$$

Where:

| | |
|---|---|
| α | coefficient Alpha reliability estimate |
| *n* | number of test items |
| $Si^2$ | variance of each test item |
| $Sx^2$ | variance of the total test |

***Inter-rater Reliability***

This method indicates how consistent test scores are likely to be if the test is scored by two or more raters. Inter-rater reliability is the consistency of the scoring process, on a test for which the scoring involves judgments by raters. It is the agreement between the scores produced by different raters scoring the same responses. It includes only the selection of raters as a source of possible inconsistency. Are you thinking that a "rater" is a machine? The answer could be yes in case of a multiple-choice test item. If the scoring of a test is purely a mechanical operation as on a multiple-choice test, the inter-rater reliability of the scores will be perfect. If it is an essay test, the raters are examiners or subject specialists or teachers. What the raters are required to do is to

independently score the same answer scripts of the examinees. For instance, a student who is dissatisfied with his grade may call for re-marking of his or her paper and either one or more examiners may be asked to score the student's answer script or rate the performance of the student independently. The scores given by the two or more of the independent examiners are correlated using a correlation technique known as Spearman Brown Prophecy formula while the Pearson Product Moment can be used to estimate the single rater's scores.

***Activity 4.3***

| |
|---|
| Take a sheet of paper and name any three types of reliability estimates that you have ever used in ensuring your test instrument is dependable? |

### 4.2.4 Factors affecting the Reliability of a Test

What we have there are: Ability range, Test length, Cheating in the test, Appropriateness of test items, Motivation of student, Homogeneity in content areas, Ensuring objectivity in scoring accuracy, Keep testing conditions constant between testing sessions, Environmental conditions should be near to ideal standard, among others.

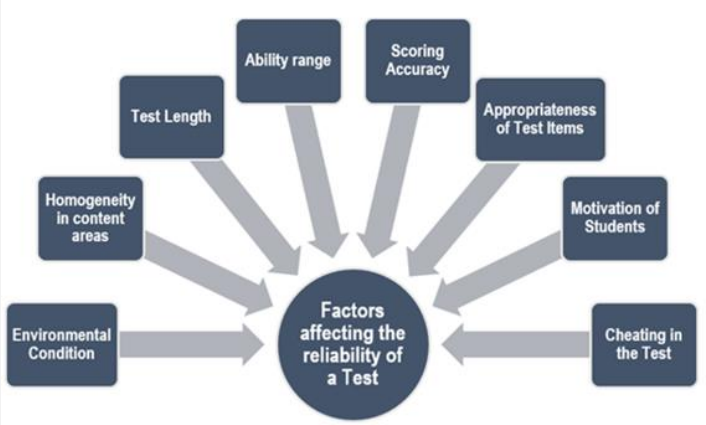



**Fig. 4.4: Factors Affecting Reliability of a Test**

***Activity 4.4***

Having known that a reliable test may be affected by factors associated with the testees, the test and its administration, how can you as a teacher ensure that reliability of your test is enhanced? List as many factors as you can.

Now, let us compare the points you have listed with what is written in this module:

- Keeping testing conditions constant between testing sessions (ensuring uniformity of conditions from one administration to another);
- Maintaining environmental conditions for testing as near to ideal standards;
- Construction of tests that has more items;
- Developing tests with moderate level of difficulty;
- Construction of tests with similarity of item content;
- Ensuring objectivity in scoring; and
- Guessing and cheating should be eliminated.

### 4.2.5 Summary

- Reliability refers to how dependable or consistent a test measures a characteristic. Reliability is a necessary but not a sufficient condition for validity. A test which is not reliable cannot be valid.
- There are different methods of estimating reliability of measuring instruments among which are test-retest method, alternate or equivalent method, split-half method, Kuder-Richardson method, Cronbach Alpha method and inter-rater reliability method.
- Test-retest method deals with the situation whereby an identical test is administered twice to the same group of students with a given interval between the two administrations of the test.

- Equivalent or alternate method involves the administration of an equivalent or parallel form of test to the same group of testees in succession and the test scores obtained are correlated.
- For the split half method, the test is administered to a group of testees once and is then subdivided into two halves for scoring purpose.
- The Kuder-Richardson method assesses the inter-item consistency or homogeneity of the items making up a test and is often used when items in the test are scored dichotomously (right or wrong answers).
- Cronbach Alpha method provides reliability estimates for instruments or tests composed of items scores dichotomously and essay tests having items of varying point values or rating scales that provide responses on a continuum of three or more response options.
- The inter-rater reliability method indicates how consistent test scores are likely to be if the test is scored by two or more raters.
- The reliability of a test can be influenced by a plethora of factors such as ability range, test length, testing conditions, cheating and motivation of students among others.
- The reliability of a test can be enhanced by keeping testing conditions constant between testing sessions (ensuring uniformity of conditions from one administration to another), maintaining environmental conditions for testing as near to ideal standards and construction of tests that have more items, among others.

***Activity 4.5***

1. Explain the meaning of reliability of a test.
2. With relevant examples and simple illustrations, discuss one reliability estimate that you were taught.
3. In what ways can the reliability of a test be enhanced?

It is not enough to stop at ensuring validity and reliability of a test. In ensuring that the process of quality assessment is complete, other factors

come to play as well, in the next unit, we shall be discussing objectivity of a test.

## Unit 4.3: Objectivity

### 4.3.0 Introduction

One of the challenges in assessing students is that sometimes we tend to score our students based on what we know they are capable of doing but not on their ability on a particular test. As soon as we come across such students' script, what we do most times is peruse through and award them some scores. Such scores are not objective but subjective. These are some of the issues or testing errors that can affect the test scores. While objectivity of a test defines information that is collected through measuring, observing, and examining facts, subjectivity describes information that is based on personal views, opinion, or value judgments. If there is no personal influence from the scorer, the test scoring is objective. Objectivity implies neutrality. In this Unit, we shall look at the meaning of objectivity, importance of test objectivity, factors that influence the objectivity of a test and how to ensure objectivity of a test.

### 4.3.1 Expected Learning Outcomes

After completing your study of this Unit, you should be able to do the following:

- Explain the meaning of objectivity.
- Enumerate the importance of test objectivity.
- Identify factors that influence objectivity of a test.
- Identify ways of ensuring objectivity of a test.

### 4.3.2 Meaning of Objectivity

Objectivity is the ability to see things without being influenced by personal feelings, emotions, or judgments. An objective assessment is based on evidence or facts. Gronlund and Linn (1995) states that, objectivity of a test refers to the degree to which equally competent scorers obtain the same results. It involves precision, specificity, absence of bias, opinion, or judgment of individual scores. So, a test is considered objective when the scorer's personal opinion and bias are eliminated. Assessments that are considered objective have a right and wrong answer that will be evaluated in the same way for every person assessed. A test that is objective measures without reference to outside influences. For example, an objective test of personality will return the same answers regardless of whether the person completing the test uses a pen or pencil. Irrelevant and unrelated factors do not influence the test results if a test is objective. It is determined by carefully studying the administration and scoring procedures to see where judgment is basic or biased. Objective type tests such as true/false and multiple choice, among others, are developed to overcome the lack of objectivity in testing. In essay type, test objectivity may be increased by careful framing of questions, and by a standard set of rules for scoring. Before reliability can be considered, a test or assessment must first be objective. Objectivity is, therefore, essential but not sufficient to ensure reliability. Objectivity of a test is an important factor as it affects both the validity and reliability of the test. There are two aspects of test objectivity which should be considered while constructing a test. These are: objectivity in test items and objectivity in scoring.

***Objectivity in Test Items***

By item objectivity, we mean that the item must call for a definite single answer. Well-constructed test items should lead themselves to one and only one interpretation by students who know the material involved. It means the test items should be free from ambiguity. A given test item

should mean the same thing to all the students. Dual meaning sentences and items having more than one correct answer should not be included in the test, as these make the test subjective. Subjective in the sense that, the scorer is at his or her discretion to either score the item right or wrong.

***Objectivity in Scoring***

Objectivity in scoring means same person or different persons, scoring the test at any time and arriving at the same result without any chance error. For test to be objective, it must necessarily be worded in a way that only correct answer can be given to it. In other words, the personal judgment of the individual who scores the answer script should not be a factor affecting the test scores so that the result of a test can be obtained in a simple and precise manner if the scoring procedure is objective. The scoring procedure should be such that there should be no doubt as to whether an item is right or wrong or partly right or partly wrong.

### 4.3.3 Importance of Test Objectivity

Objectivity is important in the following ways as illustrated in figure 4.5 below:

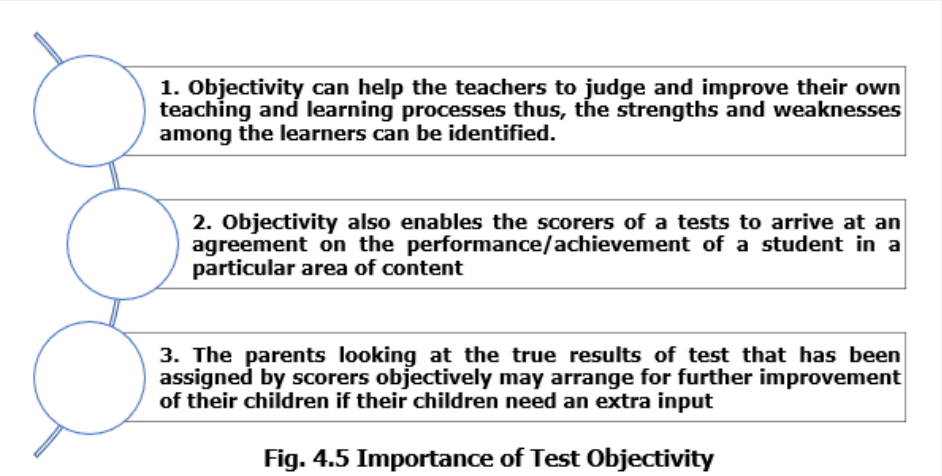



**Fig. 4.5 Importance of Test Objectivity**

### 4.3.4 Factors Affecting Objectivity of a Test

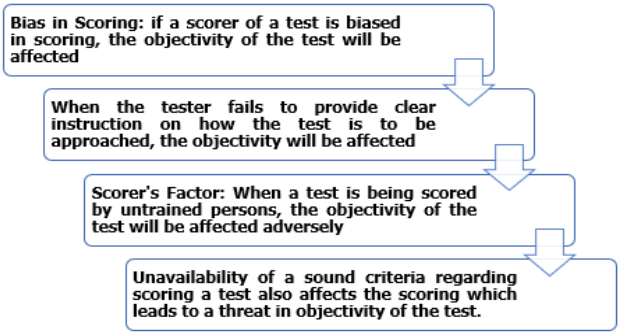



**Fig. 4.6: Factors Affecting Objectivity of a Test**

### 4.3.5 Ways of Ensuring Objectivity of a Test

The following are some of the ways to ensure objectivity of a test:

- There should be no bias in scoring the test (unbiased scoring)
- Test developers should provide a clear instruction and direction on how the test should be scored.
- Sound criteria regarding scoring of the test should be provided.
- Test should be scored by trained persons. In the case of teacher-made test, the test should be scored by the teacher who taught the subject and developed the test.

Can we briefly brainstorm on this question please? In team assessment with particular reference to scoring, which do you use in your institution?

- item writer scores his or her item,
- harmonised marking scheme for every member of the team to score irrespective of who generates the items, and
- sharing the scripts among the members of the team such that each will mark the parcel allocated to him or her.

***Activity 4.6***

1. Explain the meaning of objectivity.
2. Outline the importance of test objectivity.
3. Enumerate factors that influence the objectivity of a test.
4. List different ways of ensuring the objectivity of a test.

### 4.3.6 Summary

- Objectivity of a measuring instrument means the degree to which different persons scoring the test arrives at almost the same result.
- Objectivity is essential but not sufficient to ensure reliability.
- The two aspects of objectivity which should be considered while constructing a test are objectivity in scoring and objectivity in interpretation of test items by the testee.
- The two aspects of objectivity which should be considered while constructing a test are objectivity in scoring and objectivity in interpretation of test by the testees.
- Factors such as bias in scoring, ambiguity in test directions, lack of sound criteria regarding scoring and scorers inadequacy/inexperience affects the objectivity of a test.
- Objectivity can be enhanced through eliminating bias, provision of clear instructions, sound criteria regarding scoring of the test and scoring tests by trained persons.

In the next Unit, we shall be discussing the concept of fairness in the process of ensuring quality assessment in our schools.

## Unit 4.4: Fairness

### 4.4.0 Introduction

Let us ponder on these questions before we delve into fairness of tests:
When you were a student, did you ever feel that your teacher was not fair to you either during the process of instruction or assessment? As a

teacher/instructor, have you ever paused to reflect whether you have been fair to your students? Did you know that the idea behind testing is not to punish students? Have you given assignment to only those who paid for your material and refused to give to those who did not pay for it? When your students complain to you, how do you attend to them? Do you attend to all of them or you attend to some and push some away? Are you being fair to your students?

### 4.4.1 Expected Learning Outcomes

After completing your study of this Unit, you should be able to do the following:

- Explain the meaning of fairness in assessment
- Enumerate factors influencing fairness of assessment techniques
- List ways of ensuring fairness of assessment techniques

### 4.4.2 Meaning of Fairness

Fairness is an ethical issue and an essential quality in educational assessment. In educational setting, fairness begins right from the process of instruction through the process of evaluation. In our classroom setting, we can have students with different learning abilities. Fairness is defined as absence of bias and equitable treatment of all test takers in the testing process, and equity in opportunity to learn the material in an achievement test. In other words, fairness requires equitable treatment of all test takers in the testing process. Fairness is examined so that all participants in an assessment are provided the opportunity to demonstrate what they know and can do (Skinner, 2017). How would we tell whether a test is fair for different groups (male or female), socially advantaged or disadvantaged); ethnic grouping? The dilemma is that different groups will have different qualities and experiences, so fairness in assessment is not judged in terms of equal scores or outcomes but equitable treatment of all test takers in the testing process. A fair test is one which does not disadvantage any testee or group of testees on any

other basis other than lack of knowledge and skills the test is designed to measure. Fairness generally is marked by impartiality and honesty: free from self-interest, prejudice, or favouritism.

Fairness can also be seen as the equitable treatment of all examinees based on the conditions of testing, such as: permit to practice materials, performance feedback, retest chances, and other characteristics of test administration, among others. The notion of fairness as "equal treatment" has taken several forms: (1) fairness as a lack of bias, (2) fairness as equitable treatment in the testing process, (3) fairness as equality in outcomes of testing, and (4) fairness as opportunity to learn.

### 4.4.3 Factors that Affect Fairness

1. ***Human factor***: The day-to-day interaction of teachers with their students brings about affection. This may adversely affect the fairness of the test in terms of scoring. Sometimes, teachers may use tests as punishment to students thus the issue of fairness is eroded.
2. ***Bias***: This is also a factor that can damage the fairness in your testing. The possible sources of bias which can reduce the fairness to a considerable extent include poor test administration and inappropriate test choice, among others.

### 4.4.4 Ways of Achieving Fairness in Educational Assessments

Fairness in educational assessment is complex, and it cannot be ensured through only one practice. Some conditions for fairer educational assessment are: Opportunity to learn, Constructive Environment, Evaluative Thinking, Evaluating outcomes of assessment, Interpreting assessment results appropriately.

***Activity 4.7***

1. Explain the meaning of fairness.
2. Outline the factors that influence fairness.
3. Identify ways of ensuring fairness.
4. List different ways of ensuring the objectivity of a test.
5. Identify and state two ways in which you think lecturers are not being fair to students in your university.

### 4.4.5 Summary

1. Fairness is defined as absence of bias, equitable treatment of all test takers in the testing process, and equity in opportunity to learn the material in an achievement test.
2. A fair assessment is one in which students are given equitable opportunities to demonstrate what they know.
3. Human factor and bias affect the fairness of a test result.
4. Fairness can be enhanced by interpreting assessment results appropriately, evaluating outcomes of assessment and offering multiple opportunities to learn among others.

Finally, in this Module and in the last Unit, we shall be discussing the concept of usability in the process of quality assessment.

## Unit 4.5: Test Usability

### 4.5.0 Introduction

In this Unit, we shall be looking at another characteristic of assessment instrument which is test usability. The term 'usability' comes from the word 'use'. The main crux of this Unit, therefore, shall be on the meaning of usability of a test, considerations for test usability and how to ensure that the test is usable.

### 4.5.1 Expected Learning Outcomes

After completing your study of this Unit, you should be able to do the following:

- Explain the meaning of test usability.
- Enumerate and discuss the factors that determine the usability of a test.
- Identify ways of ensuring a test is usable.

### 4.5.2 Meaning of Usability

This refers to the extent to which the majority of the people meant to use the test can easily use it given the realities and practical conditions around its use. In other words, usability of a test is referred to as the degree to which a test can be successfully used by the classroom teachers and administrators without undue expenditure of time, money, energy and the extent to which the examinees can understand the items and finish the test without experiencing fatigue, stress, or boredom (Gronlund, 1985).The practical conditions include: the financial cost in acquiring and administering the test, the demands or conditions during test administration, the required means of scoring, the availability of equivalent forms, readability level, analysis and score interpretation. If a test is built on sound theory, and it is, therefore, valid and reliable, but makes complicated demand for its administration procedures and great expertise in scoring, such a test cannot be said to be usable or practicable or realistic. If such a valid and reliable test is too expensive, very few people or institutions will be able to afford it. Such can hardly be considered as a usable test.

### 4.5.3 Factors that Determine Usability of a Test

In judging whether a test is usable or not, the following practical considerations should be made:

- ***Ease of Administration'***. Ease of administration simply means that, a test should have a clear direction and it must be made simple and concise.
- ***Interpretation and Application***: Another important aspect of test usability is interpretation of test scores and application of test results. If the results are misinterpreted, it is harmful; on the other hand, if it is not applied, then it is useless.
- ***Availability of Equivalent Forms:*** Equivalent form tests help to verify the questionable test scores. It also helps to eliminate the factor of recall while retesting pupils on same domain of learning.
- ***Cost of Testing:*** A test should be economical from the preparation, administration and scoring points of view. An economical test is of low cost.
- ***Ease of scoring the test:*** A good test is one that is straightforward or easy to score.

### 4.5.4 Ways of Ensuring that a Test is Usable

For a test to be usable, efforts should be made to ensure the following:

- The instructions for the administration should be clear and easy to understand.
- The tester should avoid constructing a test that is bulky and will require a long time for the testees to complete.
- The guidelines for scoring the test should be easily understood by the test scorers.
- There should be guides for interpreting the scores because when test results are interpreted, it can be used for decision making in the educational sector.
- The test should be considerably cheap, so that schools can afford to buy it. However, such test must not lose its validity and reliability.

***Activity 4.8***

1. Explain the meaning of usability.
2. Identify the different considerations for test usability.
3. Enumerate the ways of ensuring a test is usable.

### 4.5.5 Summary

- Usability refers to the extent to which the majority of the people meant to use the test can easily use it given the realities and practical conditions on ground.
- The test should be financially economical.
- The test is usable, among others.

**References**


Afemikhe, O. A. (2014). *Educational Measurement and Evaluation*. Apapa-Lagos: Amfitop Books.

Barharloo, A. (2013). Test fairness in traditional and dynamic assessment. Theory and Practice in Language Studies,3( 10), 1930-1938. Retrieved from http://infonomics-societv.ie/wp-content/uploads/ licej/special-issue-

Emaikwu, S. O. (2011). *Fundamentals of test, Measurement and Evaluation with Psychometric Theories.* Makurdi: Selfers Academic Press

Gronlund, N.E. (1985). Measurement and Evaluation in Teaching. New York: MacMillan Company, New York.

Gronlund, N.E., & Linn, R. L. (1995). Measurement and Evaluation in Teaching. McMillan Company, New York.

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), *Educational and Psychological Measurement* (4th ed.). American Council on Education and Praeger.

Lindstrom, D. (2010). 'Schaums Easy Outline of Statistics. Second Edition (Schaum's Easy Outlines) 2nd Edition. McGraw-Hill Education. Retrieved from https://www.statisticshowto.com/test-retest-reliabilitv/on 09/10/2019

Mehrens, W. A. & Lehmann, I. J. (1991). Measurement and evaluation in education and psychology. USA: Ted Buchholz Objectivity: A characteristic of a good test. https://acadstuff.blogspot.com/2017/Q6/obiectivitv-characteristic-ofgood-test.html. retrieved on 05/09/2019

Shamsi, R.K.N. (2012). A fairness issue: test method facet and the validity of grammar subtests of high-stakes admissions tests. Retrieved on 5/9/2019. Retrieved from http://infonomics-societv.ie/wp-content/uploads/licej/special-issue-

Skinner, R. R. (2017). Basic concepts and technical considerations in educational assessment: a primer, congressional research service prepared for member and committees of congress. Retrieved from fas.org/sgp/crs/misc/R45048.pdf.

Ugodulunwa, C. A. (2008). Fundamentals of Educational Measurement and Evaluation. Jos: Fab Educational Books.

## MODULE FIVE

## TEST ADMINISTRATION AND SCORING

**by**
**J. Gbenga Adewale**

### 5.0 Module Overview

Generally, the term test is used to mean an instrument used in determining how much content a testee has mastered. A test is used in class work, continuous assessment, end of semester examinations which includes: objective, theory (essay) and practical. In this module, you will be exposed to four units which described tests administration and scoring. These units are:
Unit 1: Testing procedures
Unit 2: Scoring Rubrics
Unit 3: Marking scheme
Unit 4: Moderation of scores

### Unit 5.1: Testing Procedures

### 5.1.0 Introduction

In modules 4 and 5, we considered issues relating to the quality of test items and tests. If the tests are of good qualities but the administration procedures are faulty, the results obtained may leave much to be desired. In any testing process, one needs to be sure that the results obtained from the tests are consistent, free from bias, and that there is confidence in the interpretation and use of the assessment information. It is the duty of the examination body to ensure that these conditions are met. These conditions relate to safe keeping of question papers and other materials within and outside the jurisdiction of the institution.

In this unit, you will learn about testing procedure in paper pencil/biro and computer-based tests. Here, the physical environment where the test

is administered and supervised are discussed. Lastly, all forms of examination misconducts are also explained.

### 5.1.1 Expected Learning Outcomes

At the end of this unit, you should be able to:

- Moderate test items;
- Identify types of test custodians;
- Specify conditions for appropriate test custodians;
- Analyse the qualities and functions of examination coordinators;
- Identify appropriate channels of test distribution and retrieval;
- Explain the rights of the examinees; and
- Evolve strategies for combating examination malpractices.

### 5.1.2 Moderation of Tests

Moderation is a process which ensures that an assessment outcome (e.g., test, mark and/or grade) is fair, valid, and reliable and that the assessment criteria have been applied consistently, and that any differences that appear in the academic judgement of individual markers can be acknowledged and finetuned.

Moderation in this unit involves two main tasks:

a. preparing the guidelines for the preparation of draft examination papers, and
b. revising the drafts of examination papers and preparing the final versions.

***(a) Preparing the guidelines for the preparation of draft question papers***

As much as possible, a good guideline should specify:

i. the intended audience (i.e., the specific group of examinees for whom the examination is intended);
ii. the exact number of questions to be set;

iii. whether or not choices are to be allowed;
iv. the exact format of the questions (essay, short answer, etc.); and
v. the exact intellectual skills to be stressed (recall of facts and information, definition of terms, explanation, description, expression of original ideas, application of knowledge to specific situations, etc.).
vi. the intellectual skills to be tested which also determine the wordings of the questions, e.g.
    - clearly explaining the major differences between artificial and natural ventilation. (explanation); and
    - what advice will you give a village community on siting a public cemetery? (application of basic knowledge).

***(b) Revising the drafts of examination papers and preparing the final versions***

Revision of the draft questions can be done either by an individual chief examiner or by a panel of experts (e.g., subject specialists, practicing teachers, and teacher trainers). It is however preferable to have a panel vet draft questions for adequacy and usability. Among the criteria a panel of reviewers can use are the following:

a) conformity with the requirements in the syllabus and the instructional objectives,
b) level of clarity of instructions,
c) level of clarity (i.e., lack of ambiguity) of the wording of questions, and
d) the level of suitability of questions for the level of the students for whom the examination is intended.

### 5.1.3 Types of Custodians

In tertiary institutions, there are at least three types of custodians. These are, the Dean, the Head of Department, and the Examination

Coordinator. Each of the custodians has varying degrees of merits and demerits as can be seen in Table 5.1.

***Table 5.1: Strengths and Weaknesses of Different Types of Custodians***

| Custodians | Strengths | Weaknesses |
|---|---|---|
| **Dean and HOD** | High security potential | • May lack adequate storage space.<br>• May be affected by industrial unrest. |
| **Examination Officer** | Could be a dedicated examination officer | • May lack adequate storage space.<br>• If a person of questionable character is appointed, he or she may leak out the examination sheets. |

### 5.1.4 Conditions for Selecting Appropriate Custodian Points

In order to determine the appropriate custodian points, you may wish to consider the following:

- **Internal Security**
  - Is the environment secure?
  - Does the custodian have secured storage facility?

- **Adequate Space**
  - Is there a designated space?
  - Is the designated space secure?
  - Is the designated space adequate for the materials?
  - Is it free from pests such as rodents, cockroaches, and termites?

- **Commitment to Specifications of Operations**
  - Will the custodians adhere to the operational specifications?
  - Are the operational specifications likely to conflict with those of the custodians?

- Are these dedicated examination coordinators available when they are needed?
- Are there transports readily available for university courses like general courses that are offered by all the students of various faculties?

### 5.1.5 Qualities and Functions of Examination Coordinators

Examination coordinators are those officials that issue examination materials to examination supervisors and invigilators at the custodian points. In considering the qualities of examination coordinators, it is advisable to use an experienced and trustworthy staff. Examination coordinators must be of very high integrity. Such a person must be a senior staff and guaranteed by the nominee's head of department and should be capable of taking additional functions of monitoring examinations.

The examination coordinators should:

- Confirm the security of the question papers with the custodians;
- Retrieve the question papers from the custodians and hand over to the examination supervisors;
- must be with the custodians well in advance before the arrival of the supervisor; and
- retrieve the examination materials which includes: answer scripts, and other surplus materials in good conditions and at the appropriate time.

### 5.1.6 Appropriate Channels of Test Distribution and Retrieval before Examination

- Examination supervisors should take delivery following the steps below:
  a) ascertaining that examination papers are intact;

b) taking count of the number of packages involved; and
c) filling all necessary papers;

- Arrange the packages serially, irrespective of whether the examination is a one-day-all test or a span-over-time examination like the end-of-semester examinations; and
- Examination coordinators should ensure that the examination materials are stored and that the store is locked and well secured before leaving.

**On the Day of Examination - Distribution**

- Examination coordinators should arrive early at custody centre.
- Examination package is issued out to examination supervisor who signs appropriate documents which must be counter-signed or endorsed by the examination coordinators.
- Examination supervisor should return examination scripts well packaged in accordance with the institution's specifications.
- Surplus/unused examination materials (question papers, answer scripts, etc) must be retrieved from the supervisors, while strict adherence to the exact number may not be attached to the surplus question papers, the same should not be the case with surplus answer papers.
- Necessary and appropriate documents must be signed by the supervisor and the examination coordinator.

**After Examination – Retrieval**

- Examination coordinator should collect examination scripts and all other specified materials from the custodian, sign all the appropriate and necessary documents.
- Examination coordinators should move the materials to the Head of the Department's office safely.

### 5.1.7 Rights of the Examinees

The examinees must be adequately notified of the following within reasonable period:

- Examination timetable.
- Examination number.
- Examination hall/room.
- Documents to bring to the examination centre.
- Documents/materials NOT to bring to the examination centre.
- Time of commencement of examination and penalty for non-compliance e.g., "candidates shall not be allowed into the examination hall 30 minutes after the examination has started."
- Regulation on the earliest time an examinee may leave the examination hall.
- Nature/types of examination to be taken and the number of times an examination is to be written (if the nature of the examination so demands).
- That practical materials would not be shared during examination period.
- Examinees have the following rights:
    i. Be allowed out to answer the call of nature, however, within reasonable time and circumstances.
    ii. Not being disturbed by the invigilator/supervisor through undue mannerisms, unnecessary announcements, loud discussion with colleagues, and reprimand of an uncooperative assistant or examinee.
    iii. Be informed about time limits of maximum of 3 times during the examination – halfway through the examination, 10 minutes to the end of scheduled time and 5 minutes to the end of the paper. The 4th and final announcement of time is "Time Up".

iv. Availability of answer scripts that are well printed and contain all important and appropriate information for the examinee e.g., examinee's registration and examination numbers, the name of the examination (e.g., 1st semester or 2nd semester examination), etc.
v. Be attended to by the supervisor (invigilator) when he/she needs clarification on instruction regarding any question(s) on the question paper.

### 5.1.8 Strategies for Combating Examination Malpractices

As a lecturer in a tertiary institution, you must have been confronted at various times with the issue of examination malpractices. What strategies do you adopt in your institution to combat or curb examination malpractices? Draw up a list of such strategies and compare across institutions.

In addition to the strategies in your list, the following are some other strategies that institutions can employ to combat examination malpractices:

- Diversify examinations to take into consideration individual differences.
- Improve the quality of instruction and learning.
- Intensify in-service training for teachers, supervisors, and other experts.
- Improve school quality.
- Improve school–parents' relationship.
- Re-orientate the members of the society (start educating children) to appreciate the development of knowledge and skills as against the possession of mere paper qualification.

***Design prototype for each of the following:***

i. A form for the delivery of examination materials to supervisor.
ii. A form for the retrieval of examination materials from supervisor.
iii. A form for the delivery of examination materials for the course lectures.

***Activity 5.1***

1. In a group of 5, identify 3 more custodians of test in a university system; indicate the strengths and weaknesses of such test custodians.
2. Apart from the functions of an examination coordinator mentioned in this unit, identify at least four more functions of an examination coordinator.
3. What are the strategies to combat examination misconducts in the university system?

### 5.1.9 Summary

In this unit, you were exposed to testing the procedures that have to do with different types of test custodians in the universities such as the Dean/HOD or examination coordinators' offices. The weakness and strengths of each custodian were highlighted. You were also exposed to the specific conditions for appropriate test custodians like the availability of internal security, adequate space, and commitment to specifications of operations. The unit exposed you to analyse the qualities and functions of examination coordinators such as retrieving question papers from custodians and handing over to examination supervisors. You were able to identify appropriate channels of test distribution and retrieval and get familiar with the rights of the examinees. Lastly, you were exposed to the strategies for combating examination malpractices.

## Unit 5.2: Scoring Rubrics

### 5.2.0 Introduction

Rubrics are ways of describing evaluation criteria (or "grading standards") based on the expected outcomes and performances of students. Typically, rubrics are used in scoring or grading written assignments or oral presentations; however, they may be used to score any form of students' work. Each rubric consists of a set of scoring criteria and point values associated with these criteria. In most rubrics, the criteria are grouped such that the instructors and the students can differentiate among these categories by the level of performance. In classroom use, the rubric provides an "objective" external standard against which student performance may be compared.

### 5.2.1 Expected Learning Outcomes

At the end of this unit, you should be able to:

- Explain the concept of "scoring rubrics"
- Differentiate between the different types of scoring rubrics
- Discuss the uses of scoring rubrics
- Describe scoring rubrics structure
- Illustrate how to create a scoring rubric
- Display an example of scoring rubrics
- List and explain the advantages of scoring rubrics
- List and explain the disadvantages of scoring rubrics

### 5.2.2 Scoring Rubrics: What are they?

The term 'rubric' may mean different things to different professions. In education, rubric means a scoring guide used to evaluate the quality of students' constructed responses (Popham, 1997). A scoring rubric is an attempt to communicate expectations of quality around a task. In many cases, scoring rubrics are used to delineate consistent criteria for grading. Rubrics usually contain evaluative criteria, quality definitions for those

criteria at particular levels of achievement, and a scoring strategy. They are often presented in a table format and can be used by teachers when marking, and by students when planning their work (Dawson, 2015). Rubrics, when used for formative assessment purposes, have shown to have a positive impact on students' learning (Panadero & Jonsson, 2013).

### 5.2.3 Types of Scoring Rubrics

*There are two types – analytical and holistic.*

1. **Analytic Rubric:** This is the standard grid rubric that many teachers routinely use to assess students' work. This is the optimal rubric for providing clear and detailed feedback. With an analytic rubric, the criteria for the students' work are listed in the left column and the performance levels are listed across the top. The squares inside the grid will typically contain the specifics for each level. A rubric for an essay, for example, might contain criteria like "Organisation, Support, and Focus," and may contain performance levels like: "(4) Exceptional, (3) Satisfactory, (2) Developing, and (1) Unsatisfactory." The performance levels are typically given percentage points or letter grades and a final grade is typically calculated at the end.

2. **Holistic Rubric:** This is the type of rubric that is much easier to create, but much more difficult to use accurately. Typically, a teacher provides a series of letter grades or a range of numbers (1-4 or 1-6, for example) and then assigns expectations for each of those scores. When grading, the teacher matches the student’s work in its entirety to a single description on the scale. This is useful for grading multiple essays, but it does not leave room for detailed feedback on student’s work.

### 5.2.4 Functions of Scoring Rubrics

1. They provide a basis for self-evaluation, reflection, and peer review.
2. They may reduce the amount of time teachers spend evaluating student's work.
3. They are an attempt at communicating the expectations of quality around a task.
4. Scoring rubrics are used to delineate consistent criteria for grading.
5. They allow teachers and students to evaluate seemingly complex and subjective criteria.
6. They focus on measuring a stated objective (performance, behaviour, or quality).
7. A rubric can be used in individual assessment within the course, or a project or capstone project.
8. They can be used when multiple evaluators are evaluating the assessment to get focus on the contributing attributes for evaluation.
9. Rubrics are ideally suited for project assessment since each component of the project has a corresponding section on the rubric that specifies criteria for quality of work.
10. They provide timely and detailed feedback that students can use.
11. They encourage critical thinking/self-evaluation.
12. Rubrics communicate expectations to students.
13. They expose the component skills of any task.
14. They encourage a fair and consistent marking
15. Rubrics improve the reliability of scoring written assignments and oral presentations.
16. They convey the goals and performance expectations of students in an unambiguous way.

### 5.2.5 Scoring Rubric Structure

A rubric is structured like a matrix which includes two main components: criteria (listed on the left side of a matrix) and their descriptors (listed across the top of the matrix). When developing

rubrics, we should first select the most important assessment criteria which will be used to evaluate the student product. Assessment criteria listed in rubrics vary by task, discipline, and course level.

Once you identify the most important assessment criteria, you are ready to write the descriptions of what the student's performance looks like for each assessment criteria. These descriptions are called assessment standards. They can be represented through letter grades, number of points, or descriptors of quality levels, such as exemplary, competent, and developing.

### 5.2.6 How to Create a Scoring Rubric

The following six steps will help you when you decide to use a rubric for assessing an essay, a project, group work, or any other task that does not have a clear right or wrong answer.

***Step 1: Define Your Goal***

Before you can create a rubric, you need to decide the type of rubric you would like to use, and that will largely be determined by your goals for the assessment. Ask yourself the following questions: how detailed do I want my feedback to be? How will I break down my expectations for this project? Are all of the tasks equally important? How do I assess performance? Am I grading based on the work or on participation? Am I grading on both? Once you have figured out how detailed you want the rubric to be and the goals you are trying to reach, you can choose a type of rubric.

***Step 2: Choose a Rubric Type***

Although there are many variations of rubrics, it can be helpful to at least have a standard set to help you decide where to start.

***Step 3: Determine Your Criteria***

This is where the learning objectives for your unit or course come into play. In an analytic rubric, the criteria are typically listed along the left column.

***Step 4: Create Your Performance Levels***

Most rating scales are between three and five levels. Some teachers use a combination of numbers and descriptive labels like "(4) Exceptional, (3) Satisfactory, among others while other teachers simply assign numbers, percentages, letter grades or any combination of the three for each level. You can arrange them from the highest to the lowest or the lowest to the highest as long as your levels are organised and easy to understand.

***Step 5: Write Descriptors for Each Level of Your Rubric***

This is probably your most difficult step in creating a rubric. Here, you will need to write short statements of your expectations underneath each performance level for every single criterion. The descriptions should be specific and measurable. The language should be simple.

***Step 6: Revise Your Rubric***

After creating the descriptive language for all of the levels, you need to go back through and limit your rubric to a single page. Too many parameters will be difficult to assess at once and may be ineffective to assess students' mastery of a specific standard.

### 5.2.7 An Example of Scoring Rubrics for Analytic Dimension

Table 5.2: Scoring Rubrics for the quality of chapter one of a Project/Dissertation/Thesis

| | **4 Exceptional** | **3 Satisfactory** | **2 Developing** | **1 Unsatisfactory** |
|---|---|---|---|---|
| Background to the study | Justification for the study is convincing, logical, and persuasive | Justification for the study is convincing, logical but not persuasive | Justification for the study is convincing but neither logical nor persuasive | Justification for the study is neither convincing, logical nor persuasive |
| Statement of the problem | | | | |
| Purpose or objectives of the study | | | | |

| | | | | |
|---|---|---|---|---|
| Hypotheses/ Research Questions | | | | |
| Scope | | | | |
| Significance of the study | | | | |
| Operational definition of terms | | | | |

***Activity 5.2***

Complete the empty cells in Table 5.2

### 5.2.8 Good side of a Scoring Rubric

The following are the advantages of a scoring rubric:

i. Multidimensional rubrics allow students to compensate for lack of ability in one strand by improving in another one.
ii. A scoring rubric clearly shows what criteria must be met for a student to demonstrate quality on a product, process, or a performance task.
iii. Scoring rubrics can also improve scoring consistency that is, grading is more reliable while using a rubric than without one.
iv. Teachers can refer to a rubric while scoring assignments to keep grading consistent among students.
v. Teachers can also use rubrics to keep their scoring consistent with other teachers who teach the same class.

### 5.2.9 Challenges in the Use of Scoring Rubrics

The following are the difficulties encountered as one uses scoring rubrics:

i. Each level in scoring rubrics encompasses a wide range of marks. For example, if two students both receive a 'level four' mark on the project assessment, one might receive 80% and the other 100%.

ii. Distinctions between the levels do not solve the problem, because more distinctions make discriminations even more difficult. Scoring rubrics may also make marking schemes more complicated for students.
iii. Some may also find it difficult to comprehend an assignment having multiple distinct marks, and therefore makes it unsuitable for some younger students.
iv. Development of scoring rubric requires time.
v. Criteria - qualitative scales are more difficult to define than quantitative scales.
vi. Practice - both students and faculties need to practice and refine the use of rubrics for multiple types of assignments. The time to do this in class will affect "coverage."

***Activity 5.3***

1. In a group of 5, give examples for each step of a scoring rubric.
2. Develop a scoring rubric for students' knowledge of ICT.
3. Identify more strengths and limitations of scoring rubrics.

### 5.2.10 Summary

In this unit, you learnt that:

i. scoring rubrics are used in scoring or grading written assignments or oral presentations.
ii. two types of scoring rubrics are analytic and holistic scoring rubrics.
iii. there are various uses of scoring rubrics with respect to teachers and students. For students, it is a form of formative test.
iv. scoring rubrics structure is like a matrix which includes two main components: criteria (listed on the left side of a matrix) and their descriptors (listed across the top of the matrix).

v. there are six steps in creating a scoring rubric: **define your goal; choose a rubric type; determine your criteria; create your performance levels; write descriptors for each level of your rubric** and **revise your rubric.**

## Unit 5.3: Marking Scheme

### 5.3.0 Introduction

You would have experienced agitations from students with respect to the scores awarded to them in your course or other courses. Invariably, when students contest their scores, a lecturer may be embarrassed if he/she has no justification for the scores awarded to the students. This underpins the importance of a good marking scheme. Another name for a marking scheme is marking guide. In this unit, we shall consider the concept of marking scheme, how to design a good marking scheme, the steps involved in ensuring an effective scoring of students' work and how to improve your marking scheme.

### 5.3.1 Expected Learning Outcomes

At the end of this unit, you should be able to:

- Explain the concept of a marking scheme and its uses.
- Design a good marking scheme.
- Identify the steps for an effective marking of students' work.
- Identify the steps for improving a marking scheme.

### 5.3.2 Concept of Marking Scheme and its Uses

Marking guide is a plan or guidelines used in the marking of candidates' written or oral work. It could be in form of examinations (essay or objective), assignments, projects, or practical work) by the examiner. It is a guide that explains how marks are distributed to different components of the examination questions which form the basis of the overall marks awarded to a student. In practice, after the examination,

the examiner, or the course lecturer, as the case may be, reads a sample of the script. These scripts are marked in accordance with the tentative marking scheme. In the light of the candidates' interpretations of the questions, the examiner or course lecturer may amend the marking scheme. Sometimes the amendment may be substantial.

Some of the uses of a marking scheme are:

a. A good marking scheme will help you make your marking less subjective.
b. It gives you an idea of what to look out for in the candidates' scripts.
c. A non-expert can help mark candidates' scripts using the marking scheme.
d. It allows you to grade each criterion up to a maximum point value essentially by allowing you to grade within a range of points.
e. The feeling of being strict or lenient will not be necessary as the marking guide will specify the maximum scores to be obtained.
f. It allows you to justify your decision to award a particular score for candidates' responses.
g. It makes feedback objective, and such is an essential tool for responding to students' work to ensure that learning and understanding is achieved effectively. It can be a resource for teaching so that the students can identify common errors often committed.

### 5.3.3 Designing Marking Schemes

Lecturers are often faced with the challenges of marking examinations and continuous assessment scripts (e.g., assignment, presentations and/or practical work). Therefore, they need some guidelines that will help in determining the scores to be awarded. The time spent in making a good marking guide can save a lecturer some hours while marking hundreds of scripts. It becomes imperative to develop a good marking scheme as one may be required to show model answers to the departmental

reviewers, including the external examiners. It is thus important to design schemes that can withstand such scrutiny.

The following suggestions should help:

1. **As you set the questions, write a model answer for each question.** This can be a first step towards identifying the mark-bearing ingredients of a good answer. You will be able to determine the amount of task, and this will help you determine the marks to be awarded. It will also help you discover if the time allocated is adequate.
2. **Decide on what to score.** Try to allocate each mark so that it is associated with something that is either present or absent, or right or wrong, in students' answers.
3. **Make your marking scheme usable for a non-expert in the subject.** This denotes making your marking schemes useful resources for students themselves.
4. **Determine the scores to be awarded to a point in the question.** If two different people are to use your marking scheme to mark the same question, they should be able to award nearly the same or the same scores.
5. **Allow for sequential marks.** For example, when a candidate makes a mistake in the final answer, but the procedure is correct (in problem solving), allow some marks to be given for following the correct steps even when the final answer is quite wrong.
6. **Validate your marking scheme by showing it to others.** It is worth showing marking schemes to people who are not closely associated with your subject area. If they cannot see exactly what you are looking for, it may be that the scheme is not yet sufficiently self-explanatory.
7. **Look at what others have done in the past.** If it is the first time of writing a marking scheme, looking at other people's ways of doing them will help a great deal.

8. **Learn from your own mistakes.** No marking scheme is perfect. When you start applying it to a pile of scripts, you may need to make some adjustments. Keep a note of any difficulties you experience in adhering to your scheme and take account of these next time you have to make one.

### 5.3.4 How to Improve Marking of Essay Questions

The reliability of essay marks is a measure of the consistency of the examiner, not the test. Investigations have shown that essay marking has low reliabilities. The reliability of essay marks could be influenced by several factors. For example, the consistency of the marking, the variability of the examiners, that of the students, the suitability of the topic - whether the essay is measuring the same ability each time it is administered. The following suggestions have been put forward to improve the reliability of essay marks:

- The use of team marking has been suggested in order to reduce the errors arising when only one examiner marks.
- To improve the reliability of your scoring, you must do some re-scoring. If you can, score every essay question twice.
- If the essay question consists of several items, mark the answers to the first item on all papers, then go on to mark the second item, and so on.

***Activity 5.4***

1. Use the suggestions in section 5.3.3 to develop marking schemes for a question that demands the candidate to exhibit knowledge, comprehension, and higher order thinking skills.
2. Write in detail how to mark essay questions objectively.
3. Award marks to your colleagues' answer to 2 (work in peer).

### 5.3.5 Summary

In this unit, you learnt that "objectifying" essay-type tests involves:

- determining the important components of each examination question.
- determining the relative weight (or importance) of the components of each question.
- assigning marks to each component based on its relative weight.
- submitting the marking scheme to others for more profound interventions.
- joint marking of "dummy" scripts.
- team leaders' review.
- taking a second and critical review at the scripts to look for candidates' weakest points (i.e., the possible causes of failure in or avoidance of particular questions).

## Unit 5.4: Scoring Techniques

### 5.4.0 Introduction

The scores you award to your students determine the students' academic progression. Students that should ordinarily not pass have been found to pass by one mark and those that ordinarily should pass if one mark is not added, have been found to fail. Moderation of scores is a process whereby the marks are reviewed, to ensure that the individual marks awarded are appropriate in terms of consistency, fairness, and rigour in the assessment.

### 5.4.1 Expected Learning Outcomes

At the end of this unit, you should be able to:

- set criteria for scoring tests;
- explain the marking procedures for essay, objective questions, practical and projects; and
- describe the types of moderation.

### 5.4.2 Scoring Procedures

Scoring will be discussed for essay, objective, and practical examinations.

**1. Scoring for Essay Questions**

Two methods of marking discussed in this sub-section are analytic and impressionistic markings. The third which is not discussed here is the multiple marking. The analytic marking scheme is of varying degrees of rigidity while the impressionistic marking is with a minimal scheme or no scheme at all given that the examiner looks at the general impression in awarding scores. Marking by either of these two methods may be carried out by an individual working on his own or it may be done by several examiners marking each script. The later system is known as multiple marking.

a. **Analytic Marking.** This is normally carried out by having each script marked by one examiner only. The content criterion process of scoring if adopted involves assigning weights to each of the scoring categories and adding weights for each of the checked categories.
b. **Impressionistic Marking.** In impressionistic marking, the examiner reads a question rapidly, forms a general impression, and then records a rating. Individual scorers will differ in their general impressions of an essay, and these differences will be present to some extent in analytical markings.

**2. Scoring for Objective Questions**

Objective tests can be marked by another person who is not in the field of the subject or discipline. It can also be marked by machine in the case of Computer-Based Test (CBT). The paper and pencil objective tests can be answered on the question paper or on a separate answer sheet provided for the examinees (any sheet of paper or OMR). In general terms, separate answer sheets afford more accurate and reliable marking.

The two basic methods of marking objective tests are hand marking or self-marking and machine marking.

**a. Hand Marking or Self-Marking.** In a situation where separate answer sheets are used, a stencil i.e., a regular answer sheet containing holes punched at the position of the correct option to the items can be used. The examiner places the stencil on the examinees answer sheet and counts the corresponding correct responses, (i.e., the number of shaded options appearing in the holes). The use of stencil makes marking very fast, but disastrous errors could be made if the stencil is not well placed on the examinee's answer sheets. In order to ensure that the objective tests are objectively marked (i.e., error free) by using stencil, the following procedures can be taken:

i. Punch the position where the candidate's identification number and the first and last item numbers should appear.
ii. Examine the candidate's answer sheet to ensure that not more than one response has been shaded for each item.
iii. Fix the stencil to the candidate's answer sheets, that is, from edge to edge and use a clip to hold the stencil and the answer.

**b. Machine Marking**. As a result of the increase in the number of subjects as well as the number of examinees, machine scoring or marking has been introduced. This is to reduce the burden of the examiners on hand marking. Thus, the exercise of marking objective tests becomes faster, and errors are also reduced. The use of optical mark recording scanning machine facilitates this. In a recent development, CBT makes it possible to have the scores immediately a student submits his/her answers.

### 3. Scoring for Practical Examinations

In marking practical examinations, the examiner should list out the salient points and award marks for these points. Some of the areas that marks are often awarded include, setting up of apparatus, drawing, scale,

and magnification as well as labelling the steps in calculation, observations, conclusions, and inferences. In special cases or subjects such as computer, typing and operations of machines, speed is also assessed and awarded marks.

**4. Marking of Projects**

In marking projects, it is better to use the analytical approach. That is, the examiner should prepare a marking scheme, list out the points and assign weighted values. Some of the points that could be listed are introduction, statement of problem (what the project is about), method of carrying out the study, the findings and conclusion.

### 5.4.3 Criteria for Scoring Tests

You need to be aware of the various ways in which biases can be introduced into your scoring and adopt procedures designed to minimise the biases. It is essential that you focus your attention on the significant aspects of the examination paper, that your personal idiosyncrasy is minimised and that the same standards of judgement are applied to each of the papers. In order to ensure that an essay examination is effective, therefore, it is essential that you should be guided by a set of criteria. These should be predetermined and if possible pre-specified as part of the objectives of the content of the syllabus which is being tested.

**a. Content Criterion.** Although essay items are used both to elicit facts as well as to measure the thinking process, the content criterion deals only with the presence or absence of knowledge and its accuracy. All that an examiner is trying to judge is whether the content provided by the candidate can lead to conclude that the candidate is knowledgeable in the area of the essay. Even when the essay calls for reasoning, mostly, it will still be necessary to judge the content of the essay.

**b. Organisation Criterion.** Organisation and other writing skills are important components of essay performance and of many other

academic performances. If they become criteria for essay response scoring, students are likely to recognise this fact and pay more attention to them. In addition, handwriting, spelling, grammar could be evaluated since they aid in the communication process. It is, however, essential that the instructions given on the question paper should state clearly that these various aspects of essay organisation will be evaluated. Where they are not to be evaluated as it is the case in some subjects, students must not be penalised.

c. **Process Criterion.** Since the major purpose of essay type questions is to measure application, analysis, evaluation and synthesis skills, or combination of these, the most important criteria for scoring are those that reflect the adequacy with which these processes are carried out. Each of these processes results in a solution or recommendation, and justification for or reasons supporting the solution or recommendation. In other words, in scoring an essay examination paper, the adequacy of the solution and the adequacy of the reasons behind it should be evaluated. The scoring process must focus on the problem-solving behaviour that the response is intended to represent.

### 5.4.4 Moderation of Scores

Once the marking is concluded, scores are expected to be recorded on the mark sheet(s). The mark sheets are submitted to the examination coordinator who submits to the Head of the Department and a meeting of examination committee or board is arranged. The correctness of the score transfer will be ascertained, and decision will be made based on the following types of moderation procedure.

i. **Sampling Moderation**. The committee may decide to determine whether or not the scores are appropriately transferred from the students' examinations scripts into the mark sheets. In this case, a sample, say 10%, is selected randomly and the scores on the

scripts are compared to the ones on the mark sheets. Discrepancies can then be corrected.

ii. **Moderating Marking Exercise**. There is the possibility of discrepancies between the scores indicated in the marking scheme and what is finally returned to in a component or components of a question in an individual student's script. If the discrepancies are in shortage, then additional marking will be required during moderation of the scores to the actual mandatory scores. On the other hand, if the discrepancies are in surplus, then the additional marking will be required to moderate the scores by removing scores that are in excess of the actual mandatory scores.

iii. **Moderation of Transferred Scores.** Sometimes, during recording, there is the possibility of recording 53 instead of 35 or 26 instead of 62. This possibility could be moderated when an external examiner or the faculty board moderates the examination scores.

iv. **Moderation of different Markers' Scores.** Two or more examiners may be required to mark students' scripts, at the end of the marking exercise; the scores each examiner returns will be compared. If there is a significant disparity between the marks of the two or more markers in a particular unit or programme, additional marking will be required, and the scores could then be moderated.

v. **Border-Line Moderation.** Since grade points are used in the university system, there are border-line cases. For example, a student who scored 49% in an examination may actually be better than another student who scored 50% in the same test but for one reason or another, something happened, and he or she is unable to score an additional mark. Even if the ability of the student who scored 49% is lower than the ability of the student who scored 50, the difference in grade point 1 is not commensurate to 1% difference in marks. At the moderation exercise, an external

examiner or the Departmental Examination Officer or Faculty Board may decide to have an additional mark for work that has received a mark near the boundary line.

***Activity 5.5***

1. Explain marking procedures for
   a. essay,
   b. objective questions,
   c. practical and
   d. projects.
2. In a group of 5, describe how you will moderate scores of students.

### 5.3.5 Summary

In this unit, you have been exposed to the criteria for scoring tests as content criterion, organisation criterion and process criterion. You have also learnt about marking procedures for essay, objective questions, practical and projects. In essay questions, two methods of marking discussed are analytic and impressionistic markings. With respect to the objective questions, the two basic methods of marking objective tests are hand marking or self-marking and machine-marking. In marking practical examinations, it is advised that the examiner list out the salient points and award marks for these points. In marking projects, the analytical approach is recommended. Lastly, you were exposed to types of moderation which are sampling moderation, moderating marking exercise, moderation of transferred scores, moderation of different markers' scores and border line moderation.

**Bibliography**

Yoloye, T. W. and Adewale, J. G. (2006) (Eds.) A handbook on

Measurement and Evaluation for staff of examination bodies officers in Africa.

**References**

Dawson, P. (2015). "*Assessment rubrics: towards clearer and more replicable design, research, and practice Phillip".* Assessment & Evaluation in Higher Education. 42(3): 347–360. CiteSeerX10.1.1.703.8431.doi:10.1080/02602938.2015.1111294.

Fry, H., Ketteridge, S., & Marshall, S. (2003). *A Handbook for teaching and learning in higher education.* 2nd Edition. London: Kogan Page.

Gibbs, G. & Habeshaw, T. (1989). *Preparing to Teach. An introduction to effective teaching in higher education*. Bristol: Technical and Educational Services Ltd.

Habeshaw, S., Gibbs, G. & Habeshaw, T. (1995). *53 Interesting Ways to Assess Your Students.* Bristol: Technical and Educational Services Ltd.

Panadero, E and Jonsson, A. (2013*)."The use of scoring rubrics for formative assessment purposes revisited: A review".* Educational Research Review.9: 129–144. doi:10.1016/j.edurev.2013.01.002.

Popham, J. ( 1997). *"What's Wrong - and What's Right - with Rubrics".* Educational Leadership. 55 (2): 72–75.

Race, P. (2001). *The Lecturer's Toolkit: A Practical Guide to Learning, Teaching and Assessment*. 2nd Edition. London: Kogan Page:

Race, P., Brown, S. & Smith, B. (2005). *500 Tips on Assessment*. London. Routledge Falmer.

# MODULE SIX
# UNDERSTANDING TEST SCORES
**by**
**B. G. Nworgu**

## 6.0: Module Overview

As university lecturers, you administer tests or examinations to your students and in the process, you generate scores. What really are these scores? What is their nature? Are they like any other numerical quantities which we deal with in our day-to-day living such as height, length, or temperature? In this Module, we shall examine:

1. Nature of psychological tests, meaning, and types of test scores.
2. Model of test score and standard error of measurement.
3. Two frameworks for interpreting test scores.
4. Comparing scores from different tests (or examinations).

## Unit 6.1: Nature of Psychological Tests, Meaning and Types of Test Scores

### 6.1.0 Introduction

Tests and test scores are what we are all familiar with. We all started taking tests right from primary school, through secondary school to the university. In the course of our teaching, we employ tests regularly to assess our students' learning outcomes. The scores which we derive from these tests form the basis for the decisions we take about our students. These are not just important decisions, but they are sensitive in so far as they affect and, to a large extent, determine the future of our students. Being sensitive, suggests that you should exercise great caution in the way you use tests and test scores in making decisions concerning your students. This is because any wrong decision about a student can ruin the future of that student and certainly this is not the purpose of asking our students to take tests or examinations. You can imagine the number of

students in your department, faculty or university who are facing different levels of frustration owing to the wrong use or abuse of test scores.

It is obvious that before you can use test scores properly, you need to understand the nature of the instruments from where they are derived. This Unit begins by introducing you to the nature of psychological tests and then takes you through the meaning of test scores and the various types of test scores.

### 6.1.1 Expected Learning Outcomes

By the end of this Unit, you should be able to:

- explain the nature of psychological test.
- define a test score.
- identify the various types of test scores.

### 6.1.2 What is a Psychological Test?

As a lecturer, you use achievement tests to assess your students' learning. Those tests you use in assessing your students' learning provide one example of what we call psychological test. Other examples of psychological tests include intelligence tests, aptitude tests, etc. There is no doubt that given the extent of your familiarity with psychological tests, you must have formed an idea of what a psychological test is. Let us, therefore, begin this Unit by examining your own ideas about what a psychological test is.

**Activity 6.1**

> In your respective groups, share your individual views on what a psychological test is and from these views synthesize a definition of "psychological test".

You will discover that there was no consensus among the different groups in their definitions of a psychological test. Psychometrics literature is also replete with several definitions of "psychological test". We present in Box 6.1, a sample of four (4) of such definitions by notable authorities:

***Box 6.1: Sample Definitions of a Psychological Test***

| |
|---|
| **Sample 1** |
| "a device or procedure in which a sample of an examinee's behaviour in a specified domain is obtained and subsequently evaluated and scored using a standardised process" *American Educational Research Association[AERA],American Psychological Association [APA], and the National Council on Measurement in Education [NCME], 2014, p.2)* |
| **Sample 2** |
| "...a standardised procedure for sampling behaviour and describing it with categories or scores" - *Gregory (2010, p.16).* |
| **Sample 3** |
| "…a measurement device or technique used to quantify behaviour or aid in the understanding and prediction of behaviour" - *Kaplan and Saccuzzo (2013, p.7).* |
| **Sample 4** |
| ". .a tool, procedure, device, examination, investigation, assessment, or measure of an outcome (which is usually some kind of behaviour)" - *Salkind (2013, p.9).* |

***Activity 6.2***

1. Which of the five definitions did your group's definition come close to?
2. What features do these definitions share in common? Let each group identify three common features.

### 6.1.3 Characteristics of a Psychological Test

From the last activity you carried out, you are likely to find out that all psychological tests have the three (3) major characteristics or features in common. These are shown below:

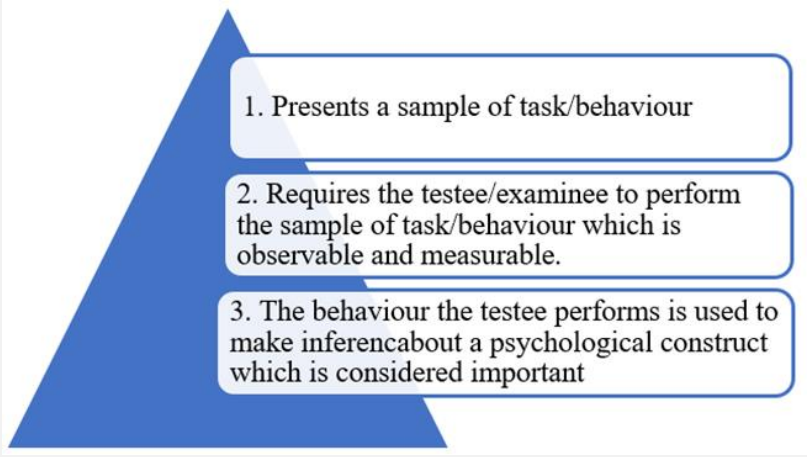



**Fig. 6.1: Characteristics of Psychological Tests**

### 6.1.4 Qualities of a Good Psychological Test

For you to develop a good psychological test, you should ensure that:

1. the behaviour/task is a representative sample of the universe of behaviour/task (i.e., domain) the test is designed to measure;
2. the behaviour is sampled under standardised conditions;
3. there are rules for scoring of the behaviours/tasks performed by the students.

***Activity 6.3***

| |
|---|
| Write down three (3) implications arising from the fact that every psychological test is a sample of the universe of the behaviour/task a test is designed to measure in terms of the scores from the tests we give to our students. |

### 6.1.5 Broad Areas of Application

There are three broad areas where we apply tests. These areas include:

1. Education services
2. Personnel services
3. Clinical Services

### 6.1.6 What is a Test Score?

When you administer a test to your students and they record their responses either mechanically on their answer scripts (for Paper-and-Pencil Test (PPT)) or electronically (for Computer-Based Tests (CBT)), what should be the next step to take? Scoring of the students' responses, isn't it? This process culminates in a quantitative or numerical value which we refer to as a 'test score'. What then is a test score? Now consider the following definition of a test score: *"a summary of the evidence contained in an examinee's responses to the items of a test that are related to the construct or constructs being measured."*

### 6.1.7 Types of Test Score

There is no doubt that you are familiar with test scores. We can broadly categorise test scores into two, namely:

1. *Raw scores, and*
2. *Transformed scores*.

***1. Raw Scores***

Raw score is that score you give to a student based on the number of questions/item he/she was able to answer correctly. We can express raw scores in two ways:

a. Number corrects or
b. Percentage score.

***Number Correct:*** This is the number of questions/items a student was able to answer correctly. For example, if a test has 50 multiple choice questions (MCQs), and a student got 30 of the questions correctly, his/her raw score will be 30 points.

***Percentage score:*** This expresses a student's number of correct score as a percentage of the total number of questions in the test or examination. To obtain a student's percentage score, we use the formula:

$$\%\, Score = \frac{Number\ correct}{Total\ number\ of\ questions/items} \times \frac{100}{1}$$

In the above example, the student's percentage score can be obtained as follows:

$$\%\, Score = \frac{30}{50} \times \frac{100}{1} = 60\%$$

### *2. Transformed Scores*

These are scores we obtain when we change or convert raw scores to other forms that will permit us to make certain decisions that cannot be supported by raw scores. We have two major categories of transformed scores, namely:

a. Standard scores, and
b. Ranks.

In this Unit, we shall consider only the standard scores.

*Standard Scores*

These are linear transformations that give rise to new distribution of scores with standard distribution characteristics: mean and standard deviation. We shall consider only two of these scores that have relevance and applicability to classroom assessment. These are the Z-score and the T-score.

*Z-score*

The Z-score is a linear transformation that results in a mean of 0 and standard deviation of 1 (Nworgu, 2019).

*T-score*

You would have observed that Z-score has two features - signs and decimal points. These features limit the usefulness of this statistic for reporting students' performance. In order to overcome these limitations, we prefer to use the T-score. The T-score, just like the Z-score, is a linear transformation that results in a distribution with a mean of 50.00 and standard deviation of 10.00. The computational procedures of Z and T scores are presented later in this module.

### 6.1.8 Summary

Psychological tests are tests that measure a sample of behaviour of examinees under standardised procedure of administration and scoring. Examples include intelligence, aptitude, and achievement tests. Test scores are numerical values obtained after scoring students' responses to a test. Two categories of test scores are raw scores and transformed scores. Examples of raw scores are the test and examination scores you obtain after testing your students; while transformed scores include standard scores (Z-score and T-score) and ranks.

## Unit 6.2: Model of Test Score and Standard Error of Measurement

### 6.2.0 Introduction

In the preceding Unit, you learned about the nature and types of test score. We can glean a deeper understanding of what a test score is if we rely on any model that explains the components of a test score and what each component represents. In this Unit, you will be introduced to one of the psychological test models (or theories) that will offer you a deeper understanding of a test score in terms of its constituent parts. This model

which we shall discuss here is known as the classical test model or theory.

### 6.2.1 Expected Learning Outcomes

By the end of this Unit, you should be able to:

- explain the basic test classical test model.
- differentiate among observed score, true score, and error score.
- explain reliability using the classical test model.
- apply the concept of standard error of measurement in interpreting test scores.

### 6.2.2 Basic Model of a Test Score

Now consider a student who scored 38% in a test/examination in your course. Assuming the student was to retake the same test/examination when there has not been any change in his or her knowledge of the course, and the conditions of the test administration and scoring remained the same across the different occasions, will this student still score 38% on each occasion? What is the implication of your answer? **This means that apart from the student's ability in the content being examined which is systematic, there are other unknown factors which are random that do influence the score obtained by a student in a test.** The systematic factors give rise to what we call the true score whereas the random or chance factors give rise to the error score. Therefore, we can say that a test score obtained by a student has two parts or components - the true score and the error score.

We can represent this relationship as follows:

$X = T + E$ where

**X** = Observed or Obtained score

**T** = *True Score*

**E=** Error of Measurement or *Error Score*.

This is the model of a test score under the Classical Test Theory (CTT).

### 6.2.3 Observed Score, True Score and Error Score

The observed score is the raw score which is the number of points a student obtained in a test. As we have seen, this score is not an accurate measure or estimate of the student's true score.

The true score is the student's score that can be attributed to his/her ability. In other words, it is that part of student's score that is not influenced by errors of measurement. Do you know that true scores can either be smaller or bigger than the observed score?

Error Score: This is part of the student's score that is due to errors of measurement. Errors of measurement are random or chance factors influencing a student's test score. Can you think of such factors? Examples include guessing, cheating, ambiguous questions/items, unclear test instructions, fatigue, and distractions. Errors of measurement or error scores have certain characteristics among which are:

1. They can be negative or positive.
2. Their algebraic sum is zero, i.e., they cancel out.

***Activity 6.4***

> In your group, discuss the implication(s) of this model of a test score in relation to your grading practices as lecturers.

### 6.2.4 Reliability of Test Scores

In Module 4, the concept of reliability was introduced as one of the key elements in ensuring the quality of measuring instruments. From that Module, you must have learnt that reliability is the degree of consistency with which the scores from a test can be reproduced and that other terms which we use to describe reliability include dependability, trustworthiness, etc. In this Unit, we shall examine the concept of

reliability from a more technical perspective that will enable us to have a better understanding and interpretation of test scores (Nworgu, 2016).

A technical meaning of reliability can be arrived at by considering the test score model, which we discussed in the preceding section.

$X = T + E$

It can be shown that the variance of the observed score will be equal to the variances of the true and error scores.

Thus:

$Var(X) = Var(T) + Var\ (E)$

If we divide both sides of the equation by $Var(X)$, we obtain

$$1 = \frac{Var(T)}{Var(X)} + \frac{Var(E)}{Var(X)}$$

$$1 = \frac{S_T^2}{S_X^2} + \frac{S_E^2}{S_X^2}$$

Where, $\frac{S_T^2}{S_X^2}$ = Observed score variance that is due to true score variance

= Reliability = $r_{xx}$

$\frac{S_E^2}{S_X^2}$ = Observed score variance that is due to error score variance

$$1 = r_{xx} + \frac{S_E^2}{S_X^2}$$

$$r_{xx} = 1 - \frac{S_E^2}{S_X^2}$$

Technically, therefore, reliability is the proportion of observed score variance that can be attributed to true score variance. It is the complement of the observed score that is due to or attributable to error score variance.

### 6.2.5 Standard Error of Measurement (SEM)

The standard error of measurement (SEM) is the standard deviation of the errors of measurement obtained from a test for a given group of testees. As Brown (1999:21) put it, "conceptually the standard error of measurement is related to test reliability in that it provides an indication of the dispersion of the measurement errors when you are trying to estimate students' true scores from their observed test scores". Mathematically, it is the square root of the error score variance.

Thus, $S_E = \sqrt{S_E^2}$

From the preceding section, we saw that $r_{xx} = 1 - \frac{S_E^2}{S_X^2}$

$\frac{S_E^2}{S_X^2} = 1 - r_{xx}$

$S_E^2 = S_X^2 (1 - r_{xx})$

Squaring both sides, we obtain

$S_E = S_X \sqrt{1 - r_{xx}}$

### 6.2.6 Interpreting Test Scores Using Score Band

One important fact we have learned from our discussion so far is that we do not know the true score of a student, but we can estimate the limits within which a student's true score will fall using the observed score and the standard error of measurement. Such limits are called score bands. An approximate method that provides reasonable limits or score band for estimating a student's true score is:

$T = X \pm SE$

*Example*

Let us assume that a student obtained a score of 38% in an examination where the standard deviation was 8.00% and the reliability of the test was 0.75. Within what reasonable limits should we expect this student's true score?

First, we calculate the SEM as follows:

$$S_E = S_X \sqrt{1 - r_{xx}} = 8.00 \sqrt{1 - 0.75} = 8.00 \sqrt{0.25}$$

$= 8.00 \times 0.50 = 4.00$

The score band within which to expect student's true score

$38 \pm 4.00$

The student's true score will lie between (38.00 – 4.00)% and (38.00 + 4.00)%, i.e., 34.00% and 42.00%

### 6.2.7 Summary

Model of test score in Classical Test Theory (CTT) is X= T+E, where X is the observed score, T is the true score, and E is the error of measurement. Sources of errors include guessing, carelessness, cheating, ambiguous questions, fatigue etc. Standard error of measurement (SEM) is the standard deviation of errors of measurement obtained from a test for a given group of examinees which is the square root of error score variance. One important lesson we have learned from this unit is that we do not know the true score of a student, but we can estimate the limits within which a student's true score will fall using the observed score and the standard error of measurement. These limits are called score bands.

## Unit 6.3 Frameworks for Interpreting Test Scores

### 6.3.0 Introduction

In assessing our students, we assign scores or grades to them. These scores or grades can only be interpreted based on a given frame of

reference. We have two such frames of reference for interpreting test scores. In our bid to understand test scores better, we shall in this Unit, discuss the two frames of reference for interpreting test scores. These are criterion-referenced and norm-referenced frames of reference.

### 6.3.1 Expected Learning Outcomes

By the end of this Unit, you should be able to:

- identify the two frames of reference used in interpreting test scores.
- explain the main features of each of the two frames of reference.
- give examples of where each frame of reference is appropriate.

### 6.3.2 Criterion-Referenced Interpretation

In criterion reference interpretation, we interpret test scores in relation to how they meet or satisfy pre-stated criteria, standards, or benchmarks. Any student whose score meets the criterion, standard, or benchmark is adjudged to have performed well or 'passed'. The criterion, standard or benchmark is usually established *a priori* and students know about it before taking the test.

Examples of such criteria, standards or benchmarks are instructional objectives, cut-off scores, 35wpm, etc. It is, therefore, possible for all the students to succeed and it is also possible for all students to fail. This type of interpretation is appropriate in proficiency tests and early stages of education. It is also used at the classroom level in assessing the extent to which the students have achieved the instructional objectives.

### 6.3.3 Norm-Referenced Interpretation

This type of interpretation involves comparing a student's score with that of a well-defined group. The performance of an average or typical member of a well-defined group is called a 'norm' (Nworgu, 2019). The basis for determining whether a student performed well or not is the performance of the other members who sat the test or examination with him/her or the performance of the average or typical member of that

group. It is not possible for all students to succeed/pass neither is it possible for all students to fail. Why is this so? Norm-referenced interpretation is appropriate in situations where the assessment is designed for selection purpose or to compare students. Examples include ranking students in a class based on their performance in end-of-term/year examination, Common Entrance Examinations, UTME, PUTME, Civil Service Promotion Examinations, etc.

***Table 6.1: Basic Approaches to Norm-referenced Interpretation and Associated Statistical Tools***

| S/N | Approach | Statistical Tool |
|---|---|---|
| 1. | Ranking the learners/testees in order of achievement or performance, either in ascending or descending order | Simple Rank |
| 2. | Indicating the proportion or percentage of the students in his/her norm group that he/she has surpassed | Percentile Rank |
| 3. | Comparing an individual's score with the mean of his group and indicating whether the person is performing below or above the mean/average (i.e., how far an individual's score is from the mean/average). | Deviation Score |
| 4. | Indicating by how many standards deviation units an individual performed above or below the average member of his group (i.e., or how many standard deviations is his/her score from the mean of the group/class). | Standard Scores<br>• Z-score,<br>• T- score |

In norm-referenced interpretation, it is not possible for all students to succeed/pass neither is it possible for all students to fail. Why is this so? Norm-referenced interpretation is appropriate in situations where the assessment is designed for selection purpose or to compare students. Examples include ranking students in a class based on their performance in end-of-term/year examination, Common Entrance Examinations, UTME, PUTME, Civil Service Promotion Examinations, etc.

***Activity 6.5***

1. Why is it possible for all students/testees to succeed/pass in criterion-referenced?
2. interpretation but in norm-referenced interpretation, that is not possible?
3. Which of these two forms is of interpretation would you prefer and why?
4. To which interpretation framework does the letter grade system adopted in tertiary institutions in Nigeria belong?

### 6.3.4 Summary

The scores or grades obtained from testing can only be interpreted based on a given frame of reference. We discussed two frames of reference for interpreting test scores. They are criterion-referenced and norm-referenced referencing. In criterion-referencing, we interpret test scores in relation to how they meet or satisfy pre-stated criteria, standards, or benchmarks; while norm-referencing involves comparing a student's score with that of a well-defined group.

## Unit 6.4: Comparing Scores from Different Tests (or Examinations)

### 6.4.0 Introduction

Consider what happens at the primary or secondary school level where pupils or students offer many subjects and sit for the examinations in these subjects. We generally tend to compare each student's performance across the different examinations or even compare the performance of two students who took different examinations based on their raw scores in those subjects. Also, the teachers obtain a composite or aggregate score for each student/pupil by adding the pupil's or student's raw scores across the different examinations or tests. Also at the University level, the raw scores obtained by each candidate in the four PUTME subjects are added to obtain a composite or aggregate score.

These assessment practices are flawed because raw scores from different examinations or subjects do not have the same unit or metric. What determines the unit or metric of any distribution of scores are the mean and standard deviation. This implies that you cannot compare raw scores from different examinations/tests directly nor can you add raw scores from different examinations/tests. According to Nworgu (2019, p.182), "in so far as any two distributions of score do not have the same mean and standard deviation, they have different units and can therefore not be added together". In fact, adding or comparing raw scores from different examinations directly is like adding different amounts of money in different currencies. Imagine someone who does something like this: 5,000 naira + 600USD + 3,000 Euro + 200 BP!

### 6.4.1 Expected Learning Outcomes

At the end of this Unit, you should be able to:

- explain the meaning of score standardisation.
- provide justification for standardising scores.
- convert raw scores to standard scores.
- interpret accurately standard scores.

### 6.4.2 What is Score Standardisation and Why Standardise Scores?

The process of transforming or converting raw scores to standard scores (i.e., Z or T score) is what is known as standardisation (Oloyede, Nworgu, Lawal and Yakasai, 2019).

This process helps us to understand test scores better and treat them more appropriately. There are two basic reasons why we standardise test scores. These are:

1. Comparability of scores obtained from different settings: Raw scores obtained from different settings cannot be directly compared because the circumstances of those settings, e.g., the difficulty/complexity of

the examinations, conditions of administration, marking, among others differ.

2. Adding or aggregating scores obtained by an individual from different subjects or courses: For the reasons mentioned above, scores from different subjects or examinations taken by a student have different units. Technically, what determines the unit of any distribution of scores is the mean and standard deviation. This implies that in-so-far-as any distribution of test scores do not have the same mean and standard deviation, they have different units. Therefore, adding the scores from those different subjects or distributions is analogous to adding amounts of money in different currencies such as BP, USD, Rupees, DM, Naira, CFA, Cedi etc. (Nworgu, 2016 and 2019). In essence, score standardisation is a mechanism for converting distributions of raw scores from different subjects/examinations to same units so as to permit the aggregation or addition of scores obtained by a student/candidate in those subjects/examinations (Oloyede, Nworgu, Lawal and Yakasai, 2019).

### 6.4.3 Standard Scores

These are linear transformations that give rise to new distribution of scores with standard distribution characteristics (mean and standard deviation). We shall consider only two of these scores that have relevance and applicability to classroom assessment. These are the Z-score and the T-score.

*Z-score*

As stated earlier in this Module, Z-score is a linear transformation that results in a mean of 0 and standard deviation of 1 (Nworgu, 2019). The formula we use for this transformation is as follows:

$$Z = \frac{X - \mu}{\sigma}$$

Where the symbols are defined as follows:

X = Raw score

$M$ = Mean

σ = standard deviation

*Example*

Suppose that in the Post-Unified Tertiary Matriculation Examination (PUTME) of a University, the mean score of the students that took the Use of English test was 58.00% with a standard deviation of 12.00%. To transform a raw score of 42.00% obtained by a candidate in the test.

Raw score = X = 42.00% ; Mean = μ = 58.00% ; standard deviation = σ = 12.00%

We then substitute these values in the formula for Z-score and simplify as follows:

$$Z = \frac{X - \mu}{\sigma} = \frac{42.00 - 58.00}{12.00} = \frac{-16.00}{12.00} = -1.33$$

***Activity 6.6***

1. Analyse the formula for the Z-score and write down three (3) things you can deduce from it.
2. Now, find out: What does a +ve, -ve or zero z-score signify?
3. How do you interpret a Z-score?

From your analysis, you can see that the Z-score:

1. Expresses a student's score (performance) in terms of how he or she is performing with reference to the mean in standard deviation units.
2. Can either be positive, negative or zero. Therefore, it has signs.

*Interpretation of Z-score*

A Z-score is interpreted as the number of standard deviations by which an individual is performing above or below the average member of his or her class (or group).

***Activity 6.7***

A 200-level student scored 58% in CHM 216 and 70% in BOT 202. If the means of the scores from the two examinations were 46.00% for CHM 216 and 78.00 % for BOT 202 with standard deviations of 6.00% and 8.00%, respectively, in which of these two subjects did the candidate perform better?

*T-score*

You would have observed the Z-score has two features: negative signs and decimal points. These features limit the usefulness of this statistics for reporting students' performance. In order to overcome these limitations, we prefer to use the T-score. The T-score, just like the Z-score, is a linear transformation that results in a distribution with a mean of 50.00 and standard deviation of 10.00.

The formula is given as:

$T\text{-score} = 10Z + 50$

Using the last example on UTME Use of English examination, we can transform the raw score of the candidate to T-score as follows:

$T\text{-score} = 10Z + 50 = 10 \times (-1.33) + 50 = -13.3 + 50 = 36.7 \simeq 37$

The T-score and Z-score serve the same purpose and are interpreted in the same manner (Nworgu, 2019).

***Activity 6.8***

1. Compare the aggregate scores reported by JAMB for candidates who sat the UTME and the aggregate scores reported by your university for candidates who sat the PUTME. What type of score does each body or institution employ?
2. A typical example of how universities combine candidates' UTME and PUTME/Screening scores is shown below:
   UTME Score = 219
   Screening Score = 233
   Aggregate Score = 226

Explain how the aggregate score was obtained and your opinion about this practice.

### 6.4.4. Summary

In this Unit, we discussed the process of transforming or converting raw scores to standard scores (i.e., Z or T score), which is known as standardisation. This process of transforming raw scores helps us to understand test scores better and treat them more appropriately. The two basic reasons why we standardise test scores are to compare scores obtained from different settings which cannot be directly compared because of issues like the difficulty/complexity of the examinations, conditions of administration, marking, among others. Transforming raw scores to Z-score and T-score is also required before adding or aggregating scores obtained by an individual from different subjects or courses.

### References

American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education, 2014 Standards for

Educational and Psychological Testing. Washington, DC: AERA.

Brown, J. D. (1999). Standard error vs standard error of measurement. Shiken: JALT Testing & Evaluation SIGNewsletter, 3(1);20-25.

Gregory, R. J. (2010). Psychological Testing: History, Principles, and Applications (6th Edition). US: Pearson Education.

Harvil, L. (2020). Standard error of Measurement. https:// www.csus. edU/indiv/b/brocks/courses/eds%20245/handouts/stan dard%20error%20of%20measurement.item.pdf

Kaplan R.M. and Saccuzzo D.P. (2013). Psychological Testing Principles, Applications, and Issues, 9th Edition

National Research Council (1999). "Front Matter." *Myths and Tradeoffs: The Role of Tests in Undergraduate Admissions.* Washington, DC: The National Academies Press. doi: 10.17226/9632.

Nworgu, B. G. (2019). *Educational Measurement and Evaluation: Theory and Practice* (Revised 2nd Edition). Nsukka: University Trust Publishers

Nworgu, B. G. (2016). *Averting Pedagogical Failure in the Sciences: Insights from Educational Measurement and Evaluation*. 103rd Inaugural Lecture of the University of Nigeria Delivered on 25th February, 2016 at the Princess Alexandra Auditorium

Oloyede, I. O, Nworgu, B. G., Lawal, R. A. &Yakasai, M. I. (2019). Social Responsiveness in Applying. Paper presented at the 45th Annual Conference of International Association for Educational Assessment [IAEA] held at Baku, Azerbaijan from 22-27 September, 2019.

Assessment Technicality: The Case of Standardisation of a Zero Score in the Unified Tertiary Matriculation Examination (UTME) Conducted By The Joint Admissions And Matriculation Board. (JAMB). Paper Presented at the 45th Annual Conference of International Association for Educational Assessment [IAEA] held at Baku, Azerbaijan from 22 – 27 September, 2019.

# MODULE SEVEN
# REPORTING AND FEEDBACK IN ASSESSMENT
**by**
**R. A. Lawal**

## 7.0 Module Overview

The ultimate goal of every meaningful assessment is feedback to be provided by all relevant stakeholders through appropriate Reporting. This Module which is centred on "Reporting and Feedback in Assessment is structured into four sequential Units. Unit 1 examines the "Rationale for Assessment Reporting"; Unit 2 discusses "Examination Reporting in Tertiary Educational Institutions"; Unit 3 highlights the "Feedback Functions of Reporting", while the last Unit explores the "Ethical Issues in Reporting".

## Unit 7.1: Rationale for Assessment Reporting

### 7.1.0 Expected Learning Outcomes

At the end of this Unit, you are expected to be able to:

- discuss the concept of "Reporting" in assessment;
- explain the rationale for "Reporting"; and
- analyse the benefits that may accrue to stakeholders from assessment reports.

### 7.1.1 Concept of Assessment

As an aspect and component of evaluation, assessment is a process of obtaining information on the twin processes of teaching and learning relative to some pre-determined objectives. In the context of this Module, assessment primarily refers to the process of ascertaining the quantity and quality of students' learning before, during and at the end of an instructional session in the lecture room, laboratory, workshop, or

field-work, or a whole educational programme. It is, thus, concerned with how much (quantity) and how well (quality) learning is either taking place (formatively) or has taken place (summatively).

Furthermore, as rightly noted by Allen (2004), assessment involves the use of empirical data on students' learning not only to improve their learning but also to refine the curriculum, including the teacher's methodology. There is, therefore, the imperative to report these data to the stakeholders, in particular students, parents and educational authorities, in a form that is easily comprehensible across board.

### 7.1.2 Concept of Reporting in Assessment

The following schematic model in Figure 7.1 illustrates the location of "Reporting" on the assessment continuum.

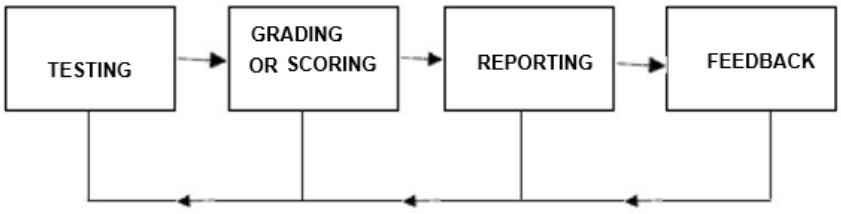



**Figure 7.1: The Location of "Reporting" in the Assessment Process**

From this figure, the meaning of Reporting in assessment can be discerned. Reporting in assessment is defined as a process of communicating the results of assessment and evaluation to various stakeholders who require it. Such stakeholders may include students, parents/guardians, other teachers, heads of institutions, administrators, among others. In short, Reporting is the process through which learning attainments of students can be made accessible to key stakeholders. That

is, learning achievements of students are reported in the form of test scores or grades.

### 7.1.3 Rationale for Reporting

Test scores and grades are quantitative and qualitative products respectively of assessment and they represent students' knowledge, skills, competencies, and ability levels. The scores for each course in the university are commonly represented in percentages covering both the Continuous Assessment (CA) and examination scores. These raw scores have limited utility value and, as such, have to undergo a form of standardisation to enable the school system to use the scores appropriately for feedback purposes. The scores are often broken roughly into three categories and presented generally as a trichotomy of ability levels - the over-achievers (Above Average), the middle achievers (Average) and the under-achievers (Below Average).

Therefore, through the simple recourse to the use of average, stakeholders are comfortable with this simple reporting system and the key rationale for assessment, which is "reforming," can be undertaken by relevant stakeholders as may be appropriate to them. The following points illustrate the rationale for assessment reporting:

i. The learner would know what was learnt, what to relearn or even unlearned.
ii. The teacher would realise the degree of success of his or her pedagogical strategies and how to bring about improvement.
iii. The educational institution would ascertain the extent to which her vision and mission are being achieved and how to keep the system on track.
iv. Parents would be able to track the progress of their wards and demand improvement.

### 7.1.4 Benefits of Reporting

Assessment reporting can have far reaching benefits to various stakeholders - the students and his/her parents/guardians, the school, other institutions of learning, community, employers, etc. It is time to brainstorm on this.

***Activity 7.1***

What are the benefits of assessment reports to each group of stakeholders - students, lecturers, parents/guardians, curriculum planners, institutional authorities? What would be your reaction if your child, at the end of the session, returns home without any form of assessment report?

***Activity 7.2***

i. What do you understand by "Reporting" in assessment?
ii. Explain the rationale for assessment reporting.
iii. Analyse the benefits that may accrue to stakeholders from assessment reports?

### 7.1.5 Summary

We have examined in this Unit the rationale for assessment reporting by taking off from the concept of assessment before discussing the concept of reporting within the process of assessment. Finally, we identified the various stakeholders who derive benefits from reporting. We will now proceed to discuss examination reporting in tertiary educational institutions in Unit 7.2.

## Unit 7.2: Examination Reporting in Tertiary Educational Institutions

### 7.2.1 Expected Learning Outcomes

At the end of this Module, you should be able to:

- identify the systems of reporting used in the university;
- demonstrate how the Grade Point Average (GPA) is calculated; and
- produce a table of degree/diploma/ classification based on cumulative GPA.

### 7.2.2 Types of Reporting

In Unit 7.1, we identified a simple type of reporting using the average to group learners into three broad performance levels. This trichotomisation is insufficient for reporting educational performance in higher institutions of learning, in particular the university.

Before going into the use GPA and its computation, it is pertinent to briefly examine some other systems of reporting used in universities:

***Absolute System:*** This method involves scaling of students' performance on numerical values (marks) from 0-100 (based on percentage). The total score of a student is a composite examination score (based on 60) and CA (based on 40).

***Relative Fixed Standard:*** This method interprets performance in terms of letter grades and grade points that have fixed ranges of marks obtained by students. An example is the five-point letter grades the university adopts as prerequisite in interpreting students' performance using CGPA system.

***Weighted Mean Average:*** This method involves the use of absolute system (numerical values in percentage) by their corresponding credit weights to determine the relative standing (Degree Class) of students in

class. Its formula depicts the quotient of the summation of the products of all the scores by the addition of the credit loads. This method takes care of the limitation of the CGPA methods where scores with wide range in values and narrow gaps (e.g., 45-49; 50-59; 60-69; 70-100) are given same grade point in determining student's degree class.

### 7.2.3 Grade Point Average

The method generally used for handling raw scores in higher education is to adopt the grading and points system for obtaining Grade Point Average (GPA) every session for each student. This eventually yields the Cumulative Grade Point Average (CGPA) used to determine the class of degree a student obtains in the degree programme. The range of scores in grading systems determines a student's class of degree. Whereas the five-point scale is used by most universities in Nigeria, there are institutions using 4, 7, 10, or other such scales in Nigeria and the rest of the world.

#### *7.2.3.1 Stages in Obtaining Students' Cumulative Grade Point Average*

1. Grades and Points Equivalent to Raw Scores: The first step towards obtaining a grade point average is to have the grade equivalents of scores obtained over hundred by students in their examinations in each course with the points as follows:

***Table 7.1: Grades and Points Equivalent to Raw Scores***

| Raw Score | Equivalent Grade | Point |
|---|---|---|
| 70-100 | **A** | 5 |
| 60-69 | B | 4 |
| 50-59 | **C** | 3 |
| 45-49 | D | 2 |
| 40-44 | **E** | 1 |
| 0-39 | **F** | 0 |

2. Obtaining the Grade Point Average (GPA): The status of a course (compulsory, required or elective), number of credits assigned it and the student's score in it are necessary for obtaining the GPA for each semester and session.

***Table 7.2: Calculating the Grade Point Average (GPA)***

| Course Code | Status | Number of Credits | Score | Grade | Point | Grade Points |
|---|---|---|---|---|---|---|
| STA 101 | C | 3 | 65 | B | 4 | 12 |
| STA 103 | C | 3 | 72 | A | 5 | 15 |
| GNS 111 | C | 2 | 55 | C | 3 | 6 |
| STA 115 | **R** | 3 | 67 | **B** | 4 | 12 |
| MAT 101 | **R** | 3 | 58 | C | 3 | 9 |
| MAT 105 | E | 2 | 68 | B | 4 | 8 |
| PHY 111 | E | 3 | 63 | **B** | 4 | 12 |
| TCO | | 21 | | | | |
| TGP | | | | | | 74 |
| GPA | | | | | | 3.52 |

KEY
C = Compulsory Course
R = Required Course
E = Elective Course

1. Cumulative Grade Point Average: The result for the first semester yielded a GPA of 3.52 and the same procedure is repeated for each session and the cumulative for the entire period of studentship is calculated to determine a student's class of degree.
2. Class of Degree: Scale for determining class of degree into which students' GPAs fall depends on the maximum and minimum points. The five-point scale is given thus:

| Cumulative GPA | Degree Classification |
|---|---|
| **4.50 - 5.00** | First Class |
| **3.50 - 4.49** | Second Class (Upper Division) |
| **2.40 - 3.49** | Second Class (Lower Division) |
| **1.50 - 2.39** | Third Class |
| **0 - 1.49** | Fail (No Degree) |

***Activity 7.3***

i. Demonstrate how the GPA is calculated.
ii. Produce a table of degree/diploma/NCE classification based on CGPA.

### 7.2.4 Summary

In this Unit we have discussed examination reporting in tertiary educational institutions and drawn parallels for the polytechnics and colleges of education by exploring different types of Report. We will now proceed to Unit 7.3 to connect Reporting with the feedback process.

## Unit 7.3: Feedback Functions of Reporting

### 7.3.1 Expected Learning Outcomes

At the end of this unit, we are expected to be able to:

- identify and explain the types of assessment report.
- identify the types of feedback.
- examine the challenges of poor feedback in assessment at the tertiary level of education in Nigeria.

### 7.3.2 Types of Report

As indicated in Figure 1 in Unit 7.1, both “Reporting” and “Feedback” are essential components of the assessment process. It is also evident that the ultimate goal of assessment is to provide adequate and timely

feedback to relevant stakeholders in the educational system. In this Unit, we would be discussing the different types of report and feedback vis-a-vis the relevant instructional processes and stakeholders.

Figure 7.2 is a schematic representation of the three main types of report used in Nigerian tertiary education.

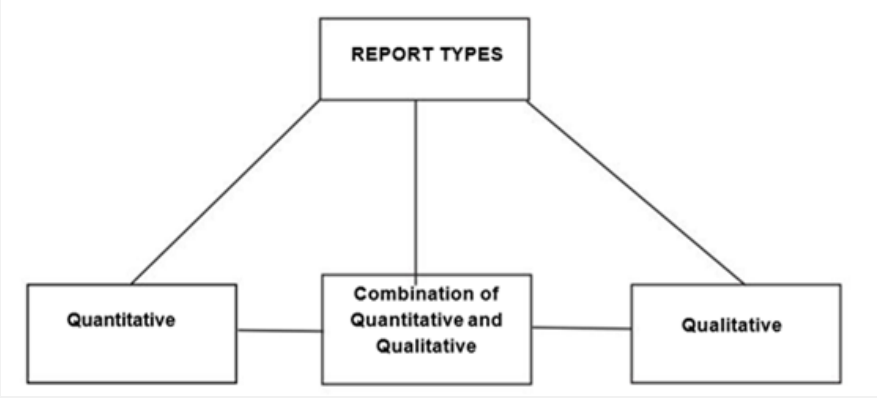



**Figure 7.2: Types of Report**

i. Quantitative Report, e.g., 70%, 7/10, CGPA= 3.67, etc.
ii. Qualitative Report: Distinction, 1st Class, Poor, Average, Merit, Credit, Lower/Upper Credit etc.
iii. Combination of Quantitative and Qualitative Report: 70% (A); 70% (Distinction); CGPA = 4.75 (1st Class); 46% (Poor); 6/10 (There is still room for improvement); 55% (Merit); etc.
iv. 1st, 2nd, 3rd, 4th, etc. (i.e., ranking)

*Activity 7.4*

**Brainstorming Session on the purpose of these Report Types:**
(a) What do you think is the purpose of each of these types of report?
(b) Relate the four types of Report to:
   i. Norm-referenced;
   ii. Criterion-referenced; and
   iii. Self-referenced testing and post-testing decisions.

### 7.3.3 Types of Feedback

To understand Feedback in the context of assessment reporting, Figure 7.3 is pertinent:

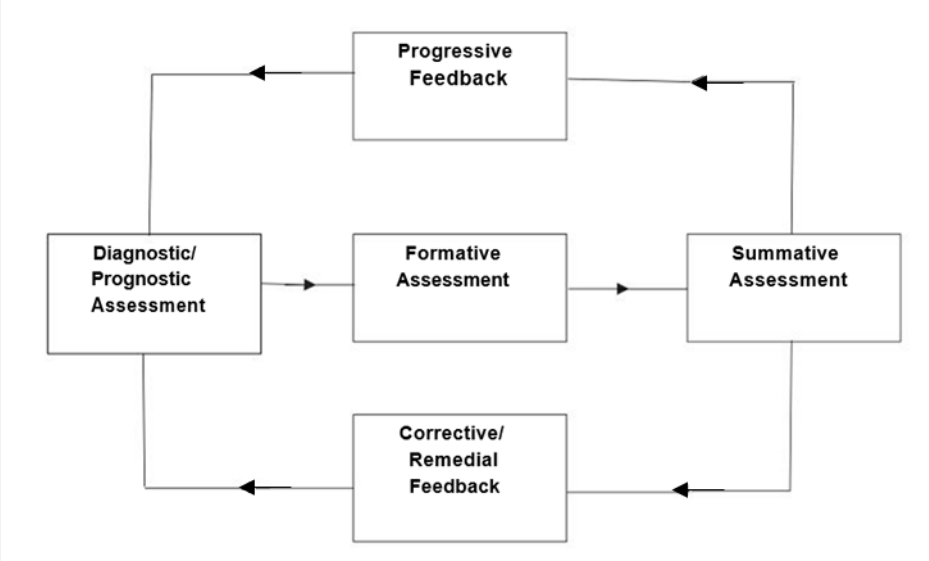



**Fig. 7.3: A Modified Version of Lawal's (2019) Model of Feedback in Education**

1. ***Corrective Feedback:*** acknowledges minor learning deficiencies based on assessment report and provides ameliorative pedagogical

intervention within the teaching session, at the end of the session, or at the beginning of the next session.

2. ***Remedial Feedback:*** identifies serious learning deficits based on assessment report and provides remedial pedagogical intervention (i.e., total re-teaching and re-learning). This may be in the form of "re-sitting" or "carrying over" of the teaching topic, course, or programme. Such stakeholders as parents, students, teachers, school authorities and curriculum experts often express divergent opinions on the sources of learning deficits and the types of pedagogical intervention learners need.
3. ***Progressive Feedback:*** is the type of feedback provided for a high-achieving group of learners whose assessment report indicates no learning deficiencies whatsoever. Although this is only possible with gifted learners, there are institutions which adopt progressive feedback in the form of "let-my-people-go" policy.

***Activity 7.5***

**Brainstorming Session:**
Let us discuss the debilitating effect of the "let-my-people-go" policy on education in Nigeria

***Activity 7.6***

1. Identify three (3) types of assessment report.
2. Critically examine the relative feedback values of the report types.
3. State the three (3) types of feedback in assessment.
4. Critically assess the significance of each of the feedback types.
5. How would you employ each of these feedback types in enhancing learning among your students?

### 7.3.4 Summary

We have located the different types of examination report within a typological framework of feedback to underscore the significance of feedback in assessment, especially its benefit in enhancing learning among students. We will now proceed to Unit 7.4 to analyse the ethical issues involved in reporting.

## Unit 7.4: Ethical Issues in Reporting

### 7.4.1 Expected Learning Outcomes

At the end of this unit, we are expected to be able to:

- identify the key ethical issues in Assessment which can compromise useful reporting.
- examine how ethical issues are handled in assessment reporting in your institution.

### 7.4.2 Ethical Issues and Educational Quality/Standards

Ethical issues in assessment reporting have to do with professional values and precepts pertaining to quality, standards, and morality in assessment practices. This Unit, therefore, highlights the ethical issues in assessment reporting.

Ethical issues pertaining to assessment reporting cover three broad areas:

a. Quality of the assessment tool including its administration,
b. Quality of scoring/grading, and
c. Quality of the assessment report.

***Quality of the assessment tool including its administration***

i. The overall consideration here is validity (the extent to which the test serves its purpose) which subsumes reliability (the consistency of an assessee's scores when re-assessed through the same assessment tool).

ii. In this regard, the assessor should not be a bully and the assessment should not be a means of threatening the assessees or a vendetta for setting personal scores.
iii. Furthermore, since the assessor is not a trickster, the assessment tool should not be used to ambush the learners and the instruction should be concise, precise, and lucid.
iv. The assessment should be based on clearly-stated objectives.
v. It should reflect expected behaviour of learners (cognitive, affective, and psychomotor competences as reflected in the instructional objectives.
vi. It should also be appropriate for students' abilities and be curriculum-based.
vii. The assessment should employ an appropriate portfolio or combination of formats - objective, subjective, performance test, oral reasoning/presentation, observation (non-test assessment), self-reports, non-verbal presentation (drawing, modelling, etc.). Emphasis should be on multiple intelligence.
viii. Test administration should be effective. An effective test administration is that in which all students are given fair chance to demonstrate their achievement of the learning outcomes being measured. It is the one that results in valid and reliable results and attains the following objectives:

***Quality of Scoring/Grading***

Objectivity, consistency, and fairness (absence of bias) within and across assessment items, assessees, and assessors (in the case of team assessment) are the essential ethical issues here.

***Quality of the Assessment Report***

The assessment report should:

i. be consistent and faithful to the assessee's performance relative to the assessment's objectives.
ii. cover multiple intelligence and behaviour domains.

iii. blend quantitative with qualitative measures.
iv. indicate the direction for learners' improvement and progress, while avoiding damaging and demoralising expressions.
v. be concise, precise, and clear.
vi. point towards appropriate feedback mechanism.
vii. be timely and accessible to key stakeholders, especially the learners themselves.

***Activity 7.7***

1. Identify and explain the core issues of ethics in reporting itself.
2. Discuss how ethical issues are handled in assessment reporting in your institution.

### *7.4.3 Summary*

In the last Unit, we x-rayed the ethical issues involved in assessment reporting by first connecting it with the tricky issues of educational quality and standards, especially as they pertain to the quality of the examination process, scoring and feedback.

**References**

Allen, M. (2004). *Assessing academic program in higher education*. Available at https//trove.nla.gov.au/work/l640957. Accessed 7/2/18

Lawal, R. A. (2018). *"Development and scoring of essay test items*". A module in a workshop on basic teaching and assessment skills for the lecturers of Al- Hikmah University, Ilorin, Nigeria, which held on the 20th and 21st February, 2018.

Lawal, R. A. (2019). A systemic framework for quality assurance in education. In R. A. Lawal et al., (eds.) *Education for Good Citizenship: Historical and 181 Contemporary Perspectives* (A festschrift in honour of Prof. Alice A. Jekayinfa), Ilorin: Haytee Press

Olasehinde-Williams, F. A. O. (2018). *"Foundation of undergraduate students' assessment"*. A module in a workshop on basic teaching and assessment skills for the lecturers of Al-Hikmah University, Ilorin, Nigeria, which held on the 20th and 21st February, 2018.

# MODULE EIGHT

# e-TESTING

**by**
**Henry O. Owolabi**
**Muhammad I. Yakassai**

## 8.0 Overview

With the introduction of the World Wide Web (www) in the 1990s, innovations are coming up in the teaching and learning processes. New developments in Information and Communication Technologies have brought with them alternatives to the traditional classroom arrangement. Transition into the electronic age has given birth to such terms as e-government, e-commerce, e-technology, and e-banking are emerging as associated practices in every facet of life. In classroom teaching and learning, the coming of e-learning, or online learning as it is sometimes referred to, has also dominated course content delivery in many institutions of learning, especially universities. Side by side with the adoption of e-learning is the use of e-testing to replace and/or supplement the traditional testing process. However, despite the rapid adoption of this new innovation in universities, particularly in Nigeria, e-testing is not fully comprehended, and its numerous advantages are not explored by examiners. This module, therefore, is intended to discuss the essential basics of e-testing. The module is divided into four units as follows:

- Understanding e-testing
- Processes in e-testing
- Online assessment models
- Computer-based testing

## Unit 8.1: Understanding e-Testing

### 8.1.0 Introduction

Any use of e-testing requires the full comprehension of it as a concept, the trends associated with it, and its application procedures. This unit is intended to explore the concept of e-testing, and its distinctions with other concepts such as e-assessment and online assessment. The unit also examines the distinctions between online assessment and traditional assessment, and online assessment of learning and assessment of online learning. The unit also discusses the components and characteristics of e-testing.

### 8.1.1 Expected Learning Outcomes

At the end of this unit, you should be able to:

- Define e-testing.
- Distinguish between e-testing and e-assessment.
- Differentiate between e-assessment and online assessment.
- Discriminate between online assessment of learning and assessment of online learning.

### 8.1.2 The Concept of e-Testing

E-testing, which is often referred to as on-screen testing, is defined as a means by which examinations and assessments are delivered on-screen through either the use of local or web-based systems (http://www.e-assessment.com). A web-based system uses internet web technologies to deliver information or services to a user or any information system while a local system depends on a network of computers that are not connected to the world wide web. Therefore, e-testing is the process of replicating paper-based tests with the use of a computer screen. It also refers to the use of internet web technology to deliver examinations for a large number of candidates.

E-testing involves the transfer of paper tests to the computer with little or no change in their appearance. Whereas it is regarded as a replication of paper-based tests, e-testing also includes variations that could not be easily replicated on paper-based tests, but which are made possible by technology such as video or audio clips and animated quizzes. To accommodate these additional features, Armenski and Gusev (2005) define e-testing as the process of using electronic devices for testing candidates. Thus, they referred to the electronic testing system as one that "accepts access from different electronic devices" (p 186).

e-Testing makes it possible to include the use of interactivity in ICT by candidates through their deployment of virtual processes to manage information before arriving at a required solution. In Nigeria, Joint Admissions and Matriculation Board (JAMB) uses e-testing for Unified Tertiary Matriculation Examinations (UTME). Likewise, many universities and other tertiary institutions employ e-tests for screening UTME candidates for admission purposes. Institutions of higher learning including universities, poly- /mono- techniques as well as colleges of education deploy e-testing techniques for their formative and summative assessment for their respective degree, diploma, and certificate examinations.

***Activity 8.1***

**Working in pairs:**
What characteristics of e-testing can you discern from this section?

### 8.1.3 Distinction between e-testing and e-assessment

Many people regard e-testing and e-assessment as synonyms, but the two terms do not have the same meaning. Authors like Purkovic and Kolumbic (2014) and Alruwias, Wills and Wald (2018) use the terms with their separate meanings. Depending on the different contexts in which they are used, however, e-assessment is broader in meaning than

e-testing. In fact, e-assessment is broader and subsumes e-testing. Whereas e-testing is delimited to those electronic procedures for presenting questions to examinees up to the termination point, e-assessment entails the end-to-end electronic processes that begin prior to the presentation of questions and go beyond, that is, up to the feedback stage. It includes all the procedures for assessing candidates' learning beginning from designing to authoring, delivery, scoring, recording, and feedback, and ending with analytics and research. The processes of e-assessment broadly cater for students' learning at different rates and they are presented with tests suited for them at an appropriate timing in a manner adaptive to their capabilities and progress made on problem-solving abilities. The e-assessment embraces the assessment of theoretical knowledge as well as practical skills with such use of an e-portfolio, simulation software, animations, design software, etc.

### 8.1.4 Components of e-Testing

Purkovic & Kolumbic, (2014) submit that e-testing consists of two main components: item bank and assessment engine. Item bank involves analysis of learning outcomes and composing e-test(s); preparation of the e-test; item analysis of the e-test; corrections and revision of the e-test based on the feedback from item analysis. The assessment engine involves a convenient user interface; a reliable subsystem for authentication; a reliable subsystem for authorisation; e-test administration and evaluation of the e-testing results.

***Activity 8.2***

**Working in pairs:**
Discuss the similarities and differences between e-testing and e-assessment. Is JAMB's UTME e-testing or e-assessment? Support your claim.

### 8.1.5 E-assessment and Online Assessment

To further understand the concept of e-testing, it is pertinent to examine the two terms, e-assessment, and online assessment. The terms e-assessment and online assessment are used to refer to procedures for testing learning outcomes through the use of technology-mediated tools such as computer software or web-based systems. Whereas online assessment is limited to assessments involving the use of the internet and therefore is web-based, e-assessment involves the use of both the internet and other ICT tools. Therefore, e-assessment encompasses online assessment.

***Activity 8.3***

> **Work in pairs to debate this contention:**
> e-assessment and online assessment are not different concepts?

### 8.1.6 Online assessment of learning and assessment of online learning

The conceptual framework on e-testing cannot be complete without examining the distinction between online assessment of learning and assessment of online learning. Battley, (2006 p. 6) distinguished between online assessment of learning and assessment of online learning. According to him, although the two terms are regarded as synonyms by some authors, there is a technical difference between them. He describes the online assessment of learning as "the use of scientifically based electronic tools (technology-mediated) to test or measure learning outcomes, in both face-to-face and distance-learning environments". Bartley (2006) refers to the assessment of online learning as "the use of scientifically based tools to test or measure learning, which occurs exclusively within the online environment.

*Activity 8.4*

**Group work:**
"The distinctions made between the two terms - *online assessment of learning* and *assessment of online learning* depend on the relative emphasis on assessment tools and learning interactions".
***Figure out the direction of emphasis of each term***.

### 8.1.7 Summary

This Unit has introduced you to the concepts of e-testing, e-assessment, and online assessment. You have also learned about online assessment of learning and assessment of online learning. The basic issues in e-testing and e-assessment have been explained to help teachers in taking decisions based on evidence obtained from their students.

## Unit 8.2: Processes in e-testing

### 8.2.0 Introduction

This section is concerned with issues and processes in conducting effective e-testing. It focuses on two approaches to e-testing. They are the cycle of e-assessment and the system of e-testing. The unit further examines support issues in e-testing.

### 8.2.1 Expected Learning Outcomes

At the end of this unit, you should be able to:

- Explain the cycle of e-assessment.
- Describe in detail the components of an e-testing system.
- Discuss the characteristics of e-testing systems.
- Explain support issues in e-testing.

### 8.2.2 Cycle of e-assessment

The cycle of e-assessment is a framework for e-testing postulated by Whitelock, Reudel, and Mackenzie (2006). It explains the process through which e-Assessment works. The cycle of assessment is replicated in the figure below:

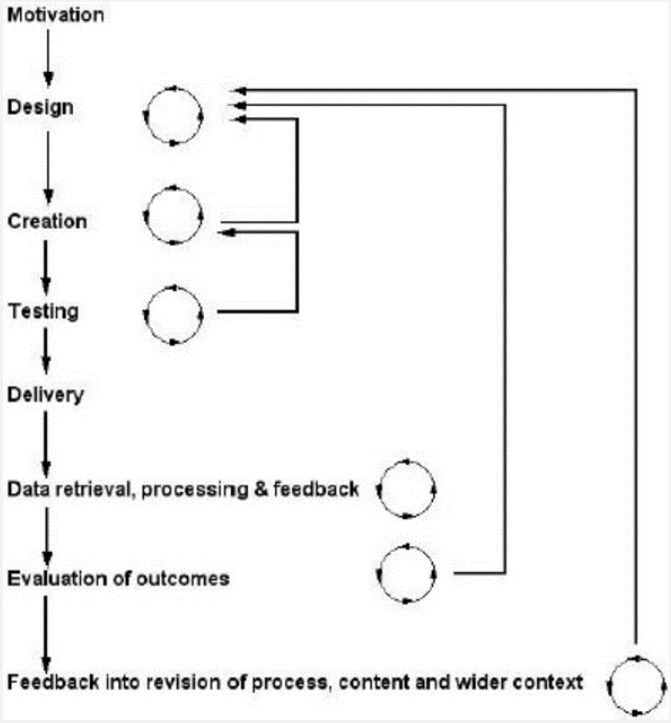



***Figure 8.1: Cycle of e-assessment***
(Source: Whitelock, D. et al, 2006)

From the figure, motivation is the first stage of e-assessment. This is the enthusiasm an examinee has for the e-testing process, and it is considered the main driver of application testing. It is regarded as a stage that is very crucial to e-assessment.

The motivation stage is followed by the design of the assessment. Design entails the process of defining the architecture, modules, interfaces, and data for the system so as to satisfy the specific requirements for the

assessment. The design is followed by the creation phase where the specific files or folders needed for the assessment are made. The next stage after creation is the testing stage. At the testing stage, the assessment is subject to and exposed to processes and steps determined in the design and creation stages to ensure they are fit for purpose. Testing is followed by the delivery of results, showing what the outcome of the assessment is. Later the data obtained are processed and feedback is obtained. Next is the evaluation stage of the outcome and decisions made as to the efficiency, utility, and relevance of the assessment. The evaluation stage is followed by the last stage whereby a review of the feedback is carried out. After the last stage, the cycle moves back to the design and creation stages.

***Activity 8.5***

**Group Work:**

a) Why do you think the motivation of the examinees is very crucial in e-assessment?
b) What is your understanding of the last step of the cycle, that is feedback into the revision process, content, and wider context?

### 8.2.3 E-testing System

Purkovic and Kolumbic (2014) postulated that the modern e-testing consists of two parts, item bank and assessment engine as illustrated in the figure below:

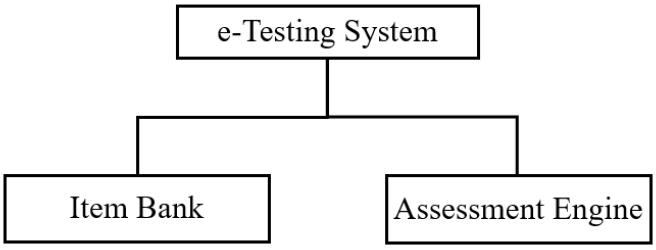



**Figure 8.2: e-Testing System**

The item bank is also referred to as the question bank. It entails item generation based on theoretical propositions on the quality of each item in a test through item analysis using the Classical Test Theory (CTT) or the modern alternative in psychometrics often called Item Response Theory (IRT). The assessment engine is the test creation algorithm which must include a convenient user interface, reliable system for authentication, and reliable system for authorisation.

According to Purkovic and Kolumbic (2014), modern Computer-Based Assessment (CBA) is driven by these two components, which can be used independently or integrated into a Learning Management System (LMS) or in a Virtual Learning Environment (VLE).

### 8.2.4 Characteristics of e-testing System

The following have been identified as the characteristics of an e-testing system.

1. Compliance of test items with the taxonomy of educational objectives on cognitive learning outcomes. This implies that questions used in e-testing reflect the different levels of thinking according to revised Bloom's taxonomy of educational objectives by Anderson and Krathwohl (2001). Test contents of instruction taken from the curriculum for each course must reflect thinking levels ranging from remembering, understanding, applying, analysing, and evaluating to creating.

2. Assessment procedures are often regulated by specific legal acts at the national, state, or institutional levels. Compliance with legislation and rules that define the procedures for assessing learners' achievement is necessary. This also constitutes the framework under which e-testing must be conducted.

3. Compliance of questions and tasks (items) with IMSQTI standards. IMS is a Global Learning Consortium while QTI

means Question and Test Interoperability. The specifications imply that items used in e-testing should comply with the standards that define the display of content items and responses to specified e-testing standards.

4. Ability to integrate with existing Learning Management Systems (LMS). Although systems for e-testing can function independently, they should be developed as modules of courses that could easily be integrated into Learning Management Systems.

5. Simplicity and reliability in use by teachers. The system for e-testing is expected to have a simple and understandable teacher's interface that has a reliable mechanism for the proper processing, interpretation, and recording of statistical indicators of the items.

6. Simplicity and availability for use by the participants. The interface of an e-testing system should be as intuitive as possible for use, authentication, and sending short and clear messages to the examinee, and the system should be designed so that it can be accessed from any computer connected to the Internet.

7. Security in the implementation of the testing procedure. Implementation of e-testing has to be safe and valid from the standpoint of:
   a. authentication and authorisation of the student, Secure Sockets Layer (SSL), monitoring network activity, safe access and storage of test solutions; and
   b. s-process control testing: monitoring of the physical activity of students, monitoring and recording of computer activities during e-testing.

8. Effectiveness and efficiency: the e-testing system is designed so that it, ultimately, facilitates the reliable evaluation of achievements to teachers and students. So, after the expiration of the time required for development, customisation and validation, and the necessary initial investment, the system should be more efficient and economical for implementation than the usual written verification of achievement.

***Activity 8.6***

**Group work:**
Work in Groups to discuss the characteristics of e-testing

### 8.2.5 Support in e-testing

Support in e-testing indicates attending to the needs of examinees before, during, and after examinations with a view to preventing each of them from being disadvantaged by the processes involved. There is no doubt that success in examinations could be hampered due to the fact that examinees are not familiar with the processes and requirements. Familiarity with the procedures and requirements should therefore not be assumed. Deliberate effort should be made to therefore do the following:

i. **Knowledge of the e-testing environment:** A good knowledge of the e-testing environment is necessary for the examinees and examiners. Location of waiting areas, points of entry to and exit from CBT halls, arrangement of workstations, location of conveniences, or requirements for online tests enable examinees to particularly enhance their motivation and confidence in the examination.

ii. **Information on the type of equipment used for e-testing:** This should be provided to enhance adequate preparation and their efficient manipulation or use. It may be necessary to adapt some equipment to meet the needs of learners. Lack of information on

them implies that examinees and examiners will be taken by surprise and may be less prepared and this may negatively affect performance.

iii. **Practice Sessions:** Opportunities should be made available for examinees and examiners to practice with their login details and thus become familiar with the environment and the equipment. This tends to affect their sense of readiness for the test.

iv. **Awareness of Centre Processes:** Such processes as checking in, movement within the halls, sitting arrangement, login, commencement, response procedures, submission, and related issues must be brought to the awareness of examinees. The role of examiners in relation to these different processes must also be brought to their awareness.

v. **Examination Regulations:** As it is with PPT, tests with the use of computers (CBT) are regulated by institutions of higher learning. Examinees and examiners must be conversant with these regulations for them to be properly guided.

***Activity 8.7***

**Group work:**
Discuss the measures you are adopting for CBT in your institution and suggest areas needing improvement.

### 8.2.6 Summary

The Unit introduced the e-testing circle from motivation, through design, creation, testing, data retrieval, to feedback components. Attempt was also made to describe the e-system and its characteristics. The support required in an e-testing system was also described.

## Unit 8.3: Online Assessment Models

### 8.3.0 Introduction

This unit describes the complex process of online assessment. It explains the different types of online assessments as they relate to various evaluation characteristics; the types of online assessment delivery methods; and online test delivery models.

### 8.3.1 Expected Learning Outcomes

At the end of reading this unit, you should be able to:

a) contrast the two broad types of online assessment delivery modes;
b) outline the eight models of online test delivery models; and
c) discuss the strengths and limitations of each model.

### 8.3.2 Types of Online Assessment Delivery Modes

Bergstrom, Fryer, and Norris (2006) identified two broad categories of online assessment delivery methods. They are Web-based and Secure sites.

***Web-based online Assessment***

According to these authorities, Web-based assessments use Web interfaces such as Internet Explorer or Netscape to deliver test content to end users via the Internet. Web-based tests can be delivered quickly and easily, any time, any place. This method of test delivery is appropriate for low-stakes tests in which the sponsor has not made a large investment in test content and wherein cheating would have little impact or no benefit to the test taker. Web-based testing requires only a common Web browser and active Internet connection for test delivery. Web-based testing is directly associated with e-learning and is very relevant for the delivery of mid- to low-stakes examinations that are diagnostic and formative in nature such as practice tests, assignments, and quizzes.

Bergstrom, Fryer, and Norris (2006) gave the following advantages of Web-based assessments:

a) For e-learning providers who cannot afford or otherwise deliver tests through a high-stakes channel of testing centres, the World Wide Web is a powerful and often less expensive way to extend quality testing services to the home, workplace, or remote testing venues.
b) Offering tests over the Web enables test developers to use a wide array of multimedia technologies that can enhance a test's authenticity or improve its accessibility.
c) As the preferred method of digital data transfer, the Web offers near-instant access to test data, which allows test results and test-taker feedback to be made available on demand.
d) Web-based test delivery can also permit organisations to take a much more active hand in the creation and administration of their testing services and offers a relatively accessible way to integrate testing with e-learning objectives.
e) As with any CBT, a Web-based test eliminates the need for test booklets and answer sheets.

***Secure site online assessment***

Secure site tests, on the other hand, take place at high-stakes test centres or proctored third-party locations such as community colleges or conference centres. The delivery of test content in this approach is highly managed and secured. Distinct from Web-based testing which requires only a common Web browser and active Internet connection for test delivery, Secure site tests use high-stakes computer-based test (CBT) delivery methods that require some form of a localised client-server network at a secure testing site for test administration.

Explaining Secure site tests delivery, Bergstrom, Fryer, and Norris (2006) asserted that as online testing technology continues to evolve, reliability and security are two key factors that continually hound the feasibility of offering high-stakes tests via the Web. As such, for high-

stakes tests that result in some form of summative assessment (certification or credentialing), Web-based delivery via e-learning is often not viewed as feasible for many programs today. A test is considered "high stakes" when there are social, legal, or professional consequences linked to the test result - for the test candidate, the testing program, or the groups involved in the testing and because of the stakes involved, administering them to the Web is not feasible. Therefore, high stakes Computer-based Tests (CBTs) are typically delivered by a variety of domestic and international vendors in dedicated test centres.

Bergstrom, Fryer, and Norris (2006) explain further that secure site tests are administered "off-line" or asynchronously. This denotes that "an electronic test file is sent to a test centre prior to administration, but the test does not rely on a real-time or synchronous connection to the Internet for delivery. "Many of these sites - even those that are established for temporary testing - rely on a local area network (LAN) to administer tests as opposed to relying on a direct connection to the Internet for assessment delivery. This ensures that the test is administered without interruption, on a secure network, under the watchful eye of a dedicated proctor (and possibly video camera), in a very secure environment. The facilities where a secure site test take place typically have been well-defined" (pp. 66-67).

***Activity 8.8***

**Group Work:**

a) Using examples, contrast the two broad categories of online delivery methods. What are the advantages of secure site tests?
b) Give some examples within Nigeria, of each of the two categories of online delivery methods.

### 8.3.3 Online Assessment Delivery Models

Various models for delivering computer-based tests have evolved over the years. In this section, the eight models reviewed by Luecht and Sireci (2011) are examined. The main focus is to understand what the models are and contrast them by their strengths and limitations. The eight models are examined below:

***Preassembled Parallel, Computerised Fixed-Test Forms***

In this model, pre-constructed (pre-assembled), intact test forms are administered through computers to large numbers of students. The items are fixed for each form that is, different examinees may see different forms of the test, but all examinees administered the same form will see exactly the same items. The model has been described (Parshall et al. 2002 as cited in Luecht & Sireci, 2011) as Computerised Fixed Tests (CFT) and is analogous to fixed-item paper-and-pencil test forms (PPT). The different forms are parallel with respect to test content and are either formally equated for difficulty (using classical test theory or item response theory) or are assumed to be randomly equivalent.

***Linear-on-the-Fly Testing***

Linear-on-the-fly testing (LOFT) is a variation of the preassembled test forms. According to Folk and Smith, (2002, cited in Luecht and Sireci 2011) the model involves the real-time assembly of a unique fixed-length test for each examinee. Like CFT, Classical Test Theory or IRT can be used to generate randomly parallel LOFT test forms. Two variations of the LOFT model are conceived: a large number of unique test forms can be developed far in advance of test administration or test forms can be generated immediately prior to testing. The primary advantage of developing the test forms in advance as claimed by Luecht and Sireci (2011) is that content and measurement experts can review each form on time.

***Computerised Adaptive Tests (CAT)***

To understand this CAT as a model of test delivery, it is pertinent to define an adaptive test. An adaptive test tailors the difficulty of the test items to the apparent ability or proficiency of each examinee. Therefore, the mechanism of CAT as a model keeps a track of an examinee's performance on each test item and then using this information selects the next item to be administered to him or her. In other words, CAT computes a score for an examinee following each item and makes a decision as to what to present to him/her next. Thus, CATs are sequentially developed item-by-item in real-time by the test-delivery software. Luecht and Sireci (2011) indicated that the goal of CAT is to ensure that reliability or score precision is maximised for each examinee. This implies that CAT caters for both high and low scorers. CAT also ensures that the difficulty of an item is tailored to the ability of examinees. Thus, tailoring the difficulty of the items to the ability of a particular examinee, according to Luecht and Sireci (2011, p.6), is proof that the reliability of the test score is maximised.

***Constrained Adaptive Testing Using Shadow Tests***

This model is an extension of the CAT which emerged to take care of the limitation of the CAT model. Many CBT researchers conceptualise the goals of CAT item selection algorithms as a multiple optimisation problem. Thus Van der Linden (2002), in Luecht and Sireci (2011) stated that whereas the optimisation problem focuses on selecting test items in such a way as to maximise the statistical information on the ability of the examinee, a large number of constraints must be met in order to ensure the same set of content specifications across examinees are obtained and thus in maximising reliability, these constraints are necessary to maintain its content validity (p. 23).

Therefore, to meet these constraints and achieve measurement efficiency Van der Linden and Reese (1998) introduced the concept of a "shadow test" as a method of achieving an optimal CAT. The shadow test method incorporates all of the required content constraints, item exposure rules,

and other constraints (e.g., cognitive levels, total word counts, test-timing requirements, clueing across items), and uses maximisation of test information at the examinee's current proficiency estimate as its objective function (Luecht and Sireci 2011).

***a-Stratified Computerised Adaptive Testing (AS)***

a-Stratified computerised adaptive testing (AS) was evolved by Chang and Ying in 1997 as a modification of the adaptive test themes. Like in CAT and "Shadow tests", AS adapts the test to the examinee's proficiency as well. However, the AS model eliminates the need for formal exposure controls and makes use of a greater proportion of the test bank than the traditional CAT. The issue of test bank use is extremely important from an economic perspective.

***Testlet-Based CATs***

This model of online test delivery was advanced by Wainer and Kiely (1987) to address the practical shortcomings of CATs as well. The authors introduced the concept of a "testlet" to describe a subset of items or a "mini-test" that could be used in an adaptive testing environment. Luecht and Sireci 2011 explained the technology of the "testlet" process by which preassembled sets of items are administered to examinees instead of single items. Examinees may be exposed to sets of items on a common reading passage, visual stimulus, a subset of specific contents of the curriculum, story, narration, picture, figure, or design. Upon completing the testlet, the computer scores the items on it and proceeds by selecting the next testlet in a series till the conclusion of the test.

***Multistage Computerised Mastery Testing***

This model is a product of many such adaptive test themes as adaptive mastery testing, computerised mastery testing, and computerised classification tests. These are explained in the literature and described by Ejim (2017) in his review as a "multi-stage form" of the CBT administration model. Ejim, (2017) explains that the multi-stage form (MST) begins presenting each examinee with the routing test which is

the first stage containing 10 or more items in its characteristic form. After completing the routing test, the score for the examinee is generated and the performance is used for taking a decision on which tests are appropriate for the second stage. In line with common adaptive test practices, a good performance on the routing test takes the examinee to a more difficult second stage test while low-performing candidates are made to attempt an easier second test. A two-stage MST is concluded with the production of a final score by aggregating performance in the examinee in both the routing and second stages. Higher branching into third, fourth, or more stages could also be designed by examiners depending on the objectives they intend to achieve.

***Computer-Adaptive Multistage Testing (ca-MST)***

This model was advanced by Luecht and Nungester in 1998 as a framework for managing real-life test construction requirements for large-scale CBT applications. The ca-MST is a pre-constructed, multistage adaptive test model which uses a "manufacturing-engineering paradigm" that incorporates multistage adaptive technologies and automated test assembly (ATA) in a way that allows test developers to maintain a greater degree of control over the quality of test forms and data. The ca-MST is adaptive in nature and is, therefore, more efficient than Computerised Fixed Tests (CFT) and Linear-on-the-fly testing (LOFT).

It is pertinent to note that there is no single CBT model that is best for all testing situations. Each model has its advantages and disadvantages. The table below from Luecht and Sireci, (2011) gives a summary of CBT delivery models and contrast them by their strength and limitations.

**Table 8.2: Summary of CBT Delivery Models**

| Model | Strength | Limitations | Selected References | Currently used by |
|---|---|---|---|---|
| Computerised | Can review test | No | Parshall et al. | Microsoft and |

| Model | Strength | Limitations | Selected References | Currently used by |
|---|---|---|---|---|
| Fixed Tests (CFT) | forms before administration. Examinees can skip and change answers to items. No item selection algorithm is needed. | improvement in measurement efficiency. Inefficient use of item pool. Poor control of item exposure (if few forms). | (2002) | other IT certification exam agencies, Physical Therapist, Physical Therapist Assistant licensure exams |
| Linear-on-the-Fly Tests (LOFT) | Better use of item pool and improved item security relative to CFT. Simple item selection algorithm. | Quality Assurance (QA) or review of operational test forms is impossible. Less efficient than any adaptive test. | Folk & Smith (2002) | Securities industry |
| Computerised Adaptive Tests (CAT) | Most efficient with respect to measurement precision and the number of items used. | Content constraints and item exposure reduce efficiency. Requires complex item selection. Test form QA is difficult to implement. Requires large item banks. Poor use of the entire item pool, even with exposure controls | Sands et al. (1997); Wainer et al. (2000); Davey & Pitoniak (2006); Segall (1996, 2010); Van der Linden & Pashley (2010). | ACCUPLACER, ASVAB, GRE, Measures of Academic Progress, Novell |
| a-Stratified Computerised Adaptive | Uses more of the entire item pool. Nearly as | QA of test forms is impossible. | Chang & Ying (1997, 1999); | None. Model-based simulations done with GRE |

| Model | Strength | Limitations | Selected References | Currently used by |
|---|---|---|---|---|
| Testing (AS) | efficient as a CAT. Simple random exposure control and content balancing are possible. | | Chang & Van der Linden (2000); Chang, Qian & Ying (2001) | data. |
| Content Constrained CAT with Shadow Tests | Maximises information while handling content and other constraints efficiently in real-time, using linear programming optimization. | Not used operationally. System-level performance issues for large-scale CBT. | Van der Linden & Reese (1998). Van der Linden (2000, 2002, 2005, 2010) | None. Model-based simulations were done with ASVAB and LSAT data. |
| Testlet-Based CAT and Multistage Computerised Mastery Tests (combined) | Adaptive selected modules or testlets. Examinees can skip and change answers to items. Adaptive component improves measurement precision relative to fixed tests | Less efficient measurement precision than pure CAT. Content balance at the test form level is not guaranteed. | Weiss & Kingsbury, (1984); Kingsbury & Zara, (1999); Lewis & Sheehan (1990); Sheehan & Lewis, (1992). | Podiatry licensing exam Gibley (1998) |
| Computer Adaptive Multistage Testing | Preconstructed content-balanced modules with targeted test information and built-in item/module exposure | Less efficient measurement precision than pure CAT. | Breithaupt & Hare (2007); Luecht (2000); Luecht & Nungester (1998, 2000); Luecht et al. (1996); | NBME (USMLE Field Tests), AICPA (Uniform CPA Examination), ETS (GRE), State of Oregon (ELPA), Massachusetts |

| Model | Strength | Limitations | Selected References | Currently used by |
|---|---|---|---|---|
| | controls. QA of tests is possible. Simplified real-time scoring and routing (score tables). The adaptive component improves measurement precision relative to fixed tests. | | Luecht et al. (2002); Sireci et al. (2008) | Adult Proficiency Test |

***Activity 8.9***

**Group work:**
Study the Summary of CBT delivery models given in the table above and try to contrast them by their strengths and limitations. Identify CBTs in Nigeria

### 8.3.4 Summary

The Unit focused on types of online assessments and discussed different types of online assessment delivery modes. This will enable participants have an understanding of alternate delivery modes to guide them in decision making on which of them to adopt.

## Unit 8.4: Computer Based Testing (CBT)

### 8.4.0 Introduction

This unit discusses the concept of CBT, the strengths and Challenges of CBT, the challenges of e-testing and test security issues in CBT.

### 8.4.1 Expected Learning Outcomes

At the end of this unit, you should be able to

a. explain the meaning of computer-based test;
b. identify the strengths of CBT;
c. highlight the challenges in e-testing;
d. define test security; and
e. distinguish different roles and responsibilities in test security.

### 8.4.2 Meaning of Computer-Based Test

As the name indicates, Computer Based Testing (CBT) is the conduct of tests with computers serving as the medium for delivery, scoring, and reporting. It refers to exposing examinees to tests and examinations with the aid of computers. In CBT, Information and Communication Technology (ICT) is used to support such testing procedures as registration and data capture, notifications, development and authoring, administration, scoring, and release of results. Though CBT covers any of these processes, a combination of some or all of them, students and candidates are more conversant with using the term to indicate the test administration component only.

The use of computers is fast replacing Paper and Pencil Tests (PPT). It is a transition that technological development has brought about that school settings, including tertiary institutions, have found necessary to adopt due to its various benefits. The new age demands all in the field of education to adopt the use of computers for all old processes, practices, and procedures. This includes learning assessment as a major component of the teaching and learning processes.

At the University of Ilorin, CBT started in 2009 and most tertiary institutions in Nigeria have also introduced its use. Although online registration of candidates for senior school certificate and entrance examinations into tertiary institutions was introduced earlier, the use of CBT in public examinations was started by Joint Admissions and

Matriculation Board in 2013 for candidates who chose to. The body went ahead in 2015 to expose all University and Tertiary Institutions Matriculation Examination candidates to CBT.

### 8.4.3 Types of CBT

CBTs could be categorised into linear and adaptive types. The linear type is similar to PPT in the sense that candidates take tests of the same length ranging from easy to difficult items. Though items on linear CBTs fall within a range of difficulty, they are not arranged in any particular order of difficulty. The Computer Adaptive Test (CAT) is the type of test in which the performance of examinees is used by the computer to select appropriate items for them. A large number of items are grouped by their contents and levels of difficulty. The computer selects items with difficulty at the ability level of examinees. Scores in CAT are generated from different items to which examinees have responded in consonance with their ability levels.

### 8.4.4 CBT Modes

There are two modes for the administration of CBT. These are:

i. ***Offline:*** This is the authoring and administration of tests on networked computers operating on local area networks only. A good number of educational institutions adopt this mode of CBT. The procedure here delivers the test to each examinee on a unit and requires them to log in with assigned details which, prior to the test administration stage, have their biometric details.

ii. ***Online:*** Usually adopted by examination bodies or test centres, this mode requires that the networked computers are also on the internet for the administration of the test. This permits tests to be administered across the district, regional, national, and international levels.

*Activity 8.10*

**Group Work:**
Identify the procedures involved in CBT as used in your institution and share that with another participant.

### 8.4.5 Strengths and Challenges of CBT

Most institutions of higher learning adopt the use of CBT to take advantage of its many strengths. These strengths include:

i. ***Ease of administration:*** The labour of examiners who invigilate, supervise, and manage tests through the deployment of examination materials and distribute them to centres and examinees has lessened with the use of computers. Technology has taken over many of these processes and thus there is a great relief from the burden of physical movement of materials here and there.

ii. ***Cost Effectiveness:*** The cost of production of papers and other writing materials is taken care of by the use of computers. The building of large spaces to serve as warehouses for storage of questions and answer booklets has become unnecessary both for educational institutions and examination bodies.

iii. ***Ease of Scoring and Recording:*** Since keys are supplied to test items at the authoring stage, computer scoring of CBTs is made possible. Machine scoring with the aid of computers applies across both the objective and essay test divides.

iv. ***Time-saving:*** Results are ready for use within a very short space of time unlike the practice with PPT whereby candidates have to endure weeks and months of sleeplessness with the attendant problems faced by examiners in the process of grading scripts and recording scores.

v. ***Customised test environment:*** Centres earmarked for the administration of CBT are deliberately designed for the purpose. Therefore, all arrangements for the comfort and safety of examinees have been planned and taken care of.

vi. ***Inclusiveness:*** Needs of examinees with special needs are addressed with the adaptive devices to help the visually or hearing impaired and people with other forms of disability.

vii. ***Possible use of Computer Adaptive Tests:*** Use of CBT has made possible the conduct of adaptive tests which has the robustness of assessing learners during the testing process.

### 8.4.6 Challenges in e-Testing

The challenges of e-testing include:

i. ***Attachment to traditional assessment practices:*** E-testing is a departure from the Paper and Pencil Test (PPT) to which most university lecturers and academic staff of tertiary institutions are used to. Most certified teachers received their training with a heavy emphasis on paper and pencil tests. The orientation and traditional assessment practices that these lecturers are familiar with have an effect on the adoption of e-testing.

ii. ***Resources for software development:*** Different institutions use alternative software for e-testing. These range from ones developed to purpose by institutions having the technical competence to those which are commercial, proprietary, or open source. The first challenge an institution faces is determining and using internal resources for the development of usable software for implementing e-testing. Going along with this challenge are the resources of alternative deployment of personnel, fund, and time to develop and validate a usable e-testing software.

iii. ***Familiarity with e-testing software:*** Various open-source e-testing software have their strengths and weaknesses. Each institution and examiner need to get familiar with them and

decide on the one that serves their purpose. Though many of these software require subscription, a good number are open source.

iv. ***Test development skills gap:*** Whereas lecturers have the content knowledge and cognate skills in their field of specialisation, that cannot be said about pedagogical skills and particularly testing. Year in year out however, these lecturers have conducted the continuous assessment, semester examinations, and orals and from these, scores used for determining classes of degree for students have been generated. Even those familiar with traditional assessment practices need some form of career development to get used to e-testing.

v. ***Psychometric Properties of Tests:*** Assessment skills gap among lecturers immediately brings the challenge of quality of tests conducted by tertiary institutions. Such qualities as validity and reliability of assessment scores aggregated for the award of degrees, diplomas, and certificates pose difficulty. If the quality of tests cannot be ascertained, then there will be Systematic error.

vi. ***Public Utilities:*** Regular electricity supply is necessary for conducting e-tests. As a result of epileptic public power supply, institutions of higher learning have sought various methods of ensuring regular and continuous supply. Alternative sources include the use of power inverters, generators, and solar. Access to toilet facilities with adequate water supply for the convenience of examinees and examiners is necessary for good assessment.

***Activity 8.11***

**Working in pairs:**
In what ways is CBT superior to PPT?

### 8.4.7 Test Security in CBT

Test security simply refers to the process which ensures that examinees have no access to contents or materials ahead of the scheduled examination, receive no aid to earn undeserved scores, and are not certified for skills they lack. Risks arise from inadvertent or vulnerable links in test development and administration while there may be deliberate steps taken to compromise the entire process. For educational institutions, maintaining the integrity of the examination system is a big issue that involves all the stakeholders. Almost the same security issues in PPT arise as they do in CBT. There are needs for security right from the item development stage to other stages as item banking, selection, test administration up to releasing the results, and record keeping. Test security in CBTs should however give attention to these essential components.

i. ***Test Authoring and Access to Item Bank:*** Subject matter experts teach their courses and develop respective tests. There is a link between them and examinees which could affect test security if not properly managed. Threats to security often arise when leakages occur at this level through the access of examinees to items on a test for which they are preparing.

   Individual academic staff and units as well as examinations office maintain items banks at their different levels. Test security is threatened with unauthorised access to such item banks. Sufficient arrangements should be made to prevent this. The risk is greater where computers have been networked and the system could be infiltrated by unauthorised users.

ii. ***Designated Environments:*** Provision is usually made for the physical structures and ICT equipment for conducting CBTs. It helps the orientation of both examiners and examinees. It further provides an opportunity to physically secure the environment and e-testing processes. It makes it easy to control movement around

and within such designated areas. This physical security is a precondition for ensuring cybersecurity in CBTs.

iii. ***Authentication of examiners and examinees:*** Most CBTs adopt the username and password authentication procedure. In addition, biometric capturing of examiners and examinees is also carried out to ensure test security. Educational institutions adopt different approaches and this requires obtaining and using such details as surname, first name, photographs, birthdate, matriculation number, level, faculty/department/college/unit, and year of entry or first registration. Details for examiners which are, most often, staff of these institutions may include their file number, birthdate, faculty/department/college/unit, date of first appointment, or other such details. Fingerprint authentication is also used to improve test security and prevent cheating tendencies.

iv. ***Invigilation and Proctoring:*** CBTs require invigilators and proctors to ensure that each examinee writes and submits their test and that cheating is prevented. Test security could be threatened as a result of the fact that invigilators and proctors have access to the examinee data, test, and test data. Role-based access control and other necessary measures should be put in place to prevent risks to test security during invigilation and proctoring.

v. ***Monitoring:*** Combinations of physical and electronic monitoring systems should be put in place to prevent examination fraud. Educational institutions adopt methods that suit their examination system by deploying personnel and technology as considered appropriate. Installing cameras in and around designated halls or centres, use of webcam and access codes may be effective for ensuring test security.

***Activity 8.12***

### 8.4.8 Summary

In this Unit, we have critically examined e-testing, e-assessment, online assessment models and computer-based testing, pointing out their similarities and differences as well as their strengths and limitations.

## References


Alruwais, N., Wills, G., & Wald, M. (2018). Advantages and Challenges of using e-assessment. *International Journal of Information and Education Technology,* 8(1), 33-37.

Armenski, G. S., & Gusev, M. (2005). Infrastructure for e-testing. Facta Universitatis – Series: Electronic Energetics 18(2), 181-204. http://doi.org/10.2298/FUEE0502181A

Bartley, J. M. (2006). Assessment is as Assessment Does: A Conceptual Framework for Understanding Online Assessment and Measurement. In M. Hricko, & S. L. Howell (Eds.). *Online Assessment and Measurement: Foundation and Measurement* (pp. 1-45). Hersley. Information Science Publishing.

Bergstrom, B., Fryer, J. & Norris, J. (2006). Defining online assessment for the adult learning market. In M. Hricko, & S. L. Howell (Eds.). Online Assessment and Measurement: Foundation and Measurement (pp. 1-45). Hersley. Information Science Publishing.

Ejim, S. (2018). Overview of computer-based test. Retrieved October 7, 2020, from http://www.reserachgate.net/publication/324208499.

Luecht, R. M., & Sireci, S. G. (2011). A review of models for computer-based testing. College Board Research and Data. Retrieved October 8, 2020, from http://www.collegeboard.org/research.

Luecht, R.M. (1998b). Computer-assisted test assembly using optimization heuristics, *Applied Psychological Measurement*, 22, 224–236.

Purkovic, D., & Kolumbic, Z. (2014). A concept of the web-based e-testing system. Retrieved September 1, 2019, from http://www.reserachgate.net/publication/26142485.

Whitelock, D., Rendel, C., & Mackenzie, D. (2006). E-assessment: Case studies of effective and innovative practice a JISC, *Jt. inf. System. comm.* p. 184.

Wainer, H., & Kiely, G. L. (1987). Item Clusters and Computerized Adaptive Testing: A case for Testlets. *Journal of Educational Measurement*, 24, 185–201.

## MODULE NINE

## CREDIBILITY ISSUES IN EXAMINATIONS

**by**
**Abubakar Hamman-Tukur**

### 9.0 Module Overview

No doubt, we have all taken many examinations in the course of our schooling. Some were bad, others were good. Some were so good or so bad that we sometimes remember them even after many years of leaving school. Do we not wonder sometimes what made such examinations good or bad? For a student, an examination he/she excels in is a good examination and the one he/she does poorly is bad. The expert is more circumspect and critical before he/she passes judgment on an examination. In this Module, we shall examine what a credible examination is and issues associated with it from preparation to reporting of results. By the end of this Module, you are expected to explain the meaning of credible examination, and discuss issues in examination preparation, administration, scoring and reporting of results.

### Unit 9.1: Meaning of Credible Examination

### 9.1.0 Introduction

This Unit explains the meaning of a credible examination as well as characteristics and importance of credible examinations.

### 9.1.1 Expected Learning Outcomes

At the end of the Unit, you are expected to be able to:

- explain the meaning of a credible examination;
- explain impact, curricular and instructional validity;
- describe the characteristics of a credible examination;
- state the importance of credible examinations.

### 9.1.2 What is a Credible Examination?

In our university system, students move from one level to another and ultimately graduate and are awarded certificates after passing series of stipulated examinations. These examinations normally come at the end of a course or semester. Before the end of each course, students normally take tests in the form of continuous assessment. In our university system examinations are tests that come at the end of a semester. Examinations can also be external, for example, Unified Tertiary Matriculation Examination (UTME) and Senior School Certificate Examination (SSCE). These examinations are often referred to as 'high-stakes' examinations because of the important decisions made based on grades or scores obtained from them and the vested interests of different groups of stakeholders. Credible examinations must have impact, curricular and instructional validity.

### 9.1.3 Meaning of Impact, Curricular and Instructional Validity

***Impact validity***

Wakeford (2003) opined that impact validity is concerned with whether the examination has made impact on the students (e.g., future learning, attitudes, etc.). An examination that has affected how students prepare for other examinations is said to have impact validity. Similarly, an examination that has resulted in students developing favourable attitudes towards learning is also said to have impact validity.

***Curricular and Instructional Validity***

Curricular validity means that "a test measures what is contained in the curriculum; while instructional validity means that the students are receiving instruction in the skills and knowledge being measured by the test" (Capper, 1996, p.39). Instructional validity is an issue of whether students are examined the way they were taught or taught the way they were examined.

### 9.1.4 Importance and Characteristics of Credible Examinations

Credible examinations in the education system will not only ensure that teaching and learning are improved, but also that reliable and valid decisions are made vis-a-vis who should be admitted or not admitted into institutions and the class of degree awarded to persons who have taken the examinations. Apart from enabling reliable decisions on students, credible examinations enable stakeholders to:

a. reliably know what students are learning,
b. understand and address inequalities and also identify factors that support learning,
c. improve teaching and learning to improve examination (Ngaware, 2017).

It is, therefore, important that such examinations should have credibility. Credibility in examinations does not come by chance. Deliberate efforts are put in place from the preparation of an examination to the reporting of scores or grades. We shall, in the next Units, examine credibility issues associated with the many phases of examinations.

***Activity 9.1***

**Class Discussion**

Evaluate the credibility of some of the examinations you have given in terms of their impact, curricular and instructional validity.

### 9.1.5 Summary

In this Unit we learnt that a credible examination is one that possesses impact, curricular and instructional validity. A credible examination makes possible reliable decisions on examinees such as degree classification, movement to next level of learning and improvement of teaching and learning, among others. Impact validity is concerned with

the extent to which an examination has affected students' learning and attitudes.

## Unit 9.2: Credibility Issues in Preparation for Examinations

### 9.2.0 Introduction

This Unit discusses credibility issues in the preparation for examinations by lecturers, institutions, and examinees.

### 9.2.1. Expected Learning Outcomes

At the end of the Unit, you should be able to:

- identify preparatory measures lecturers need to take to ensure the credibility of their examinations.
- explain preparatory measures institutions need to take to ensure the credibility of their examinations.
- discuss preparatory measures examinees need to take to ensure the credibility of the examinations they take.

### 9.2.2 Preparatory Measures for Credible Examination by Lecturers

Preparation for an examination starts the moment it is slotted into the school calendar. Lecturers have to think about what is it that they would like to examine their students on and how. As you have learnt in the previous modules, these require systematic and detailed planning, otherwise the examination will lack credibility. On the part of the lecturers, the planning involves the procedures described in Modules 3 and 4, otherwise they will set examination questions on topics that focus on narrow segments of the course. Such examinations will lack content validity. Preparation also involves preparing a scoring key for objective items or marking scheme or scoring rubrics in the case of essay questions at the time of questions preparation and not after. This step

will help remove any doubt about ambiguities in questions and responses expected from students, thus facilitating reliability.

When questions are developed before examinations, lecturers need to guard against the tendency of 'teaching to the test'. This is the tendency of consciously or unconsciously emphasising the content elements that are in pre-set examination questions.

When examinations are prepared, their security must be guaranteed. Desperate students make attempt to have knowledge of the questions they would be asked well before the examination to enable them to prepare and pass the examination. Students who are afraid of failure or are very desperate for a high grade and do not have the capacity for such are prone to engage in this form of malpractice.

Students also need to be prepared for examination, otherwise they will not give a good account of themselves. They should know what type of examination to expect and when to expect it. All information necessary for them to prepare for an examination must be conveyed to them. For example, the type of questions to expect, (i.e., whether multiple choice questions or essay or any other type of question), the duration of the examination, venue, examination regulations, and all other information necessary for the students to prepare for an examination must be conveyed to them. An examination that fails to do this will lack credibility.

### 9.2.3 Preparatory Measures for Credible Examinations by Institutions

On the part of the authorities, preparation for a credible examination requires planning to provide the resources and secure an enabling environment for the conduct of the examination. Often, failure to prepare adequately for an examination has manifested in inadequate question papers, overcrowding in examination halls, inadequate number of

invigilators, faulty and inadequate computers, electricity issues, timetabling issues, dirty and noisy environment, among others.

***Activity 9.2***

**Discuss in Groups and Report**

a. Have you ever experienced a case of examination leakage? How was it handled?
b. How do you prevent examination leakages?
c. To what extent would you say the examination environment in your institution has been affecting the credibility of its examinations?

### 9.2.4 Summary

In this unit we have learnt that the conduct of credible examinations require, invigilators to ensure a conducive environment for examination taking and to prevent examination malpractice, examinees to abide by rules and regulations of examinations and examiners to be fair and objective in their scoring.

## Unit 9.3: Credibility Issues in Administration and Scoring

### 9.3.0 Introduction

This Unit examines credibility issues associated with invigilation of examination, the roles of examinees and the scoring of objective and essay examinations.

### 9.3.1 Expected Learning Outcome

At the end of the Unit, you should be able to:

- explain the functions of the invigilator in ensuring credible examinations.

- mention the form of examination malpractice students engage in during the conduct of examination.
- explain the roles of students in ensuring the conduct of a credible examination.
- explain how to ensure credibility of examinations through scoring.

### 9.3.2 Functions of the Invigilators

The next phase in the examination process is to present the examination to the examinees. This phase requires venue, invigilators, and other personnel. Preparation in the earlier phase should ensure that the venue is conducive for examination taking and there are also adequate invigilators and other support personnel necessary for the conduct of the examination. One of the functions of the invigilators is to ensure the credibility of examinations by preventing and detecting any form of examination malpractice. The invigilator must not only be aware of the various forms of examination malpractice examinees engage in but must also be vigilant to minimise the perpetration of these and report the culprits for disciplinary measures. Failure to do this will compromise the credibility of the examination. Other functions of the invigilator are to assure that the environment for examination is conducive; as far as possible, examinees' attention should not be distracted, and they should not be subjected to any intimidation or any other experience that will prevent them from giving a good account of themselves.

### 9.3.3 Forms of Examination Malpractice by Students

Examination malpractices common among students include impersonation, spying, also called 'giraffing', exchange of answers and answer scripts and other forms of collusion, diversion of the attention of the invigilators, smuggling of answers or answer scripts, intimidation of invigilators, criminal use of ICT, among others. As days go by, and with technological advancement, new forms of malpractice are emerging.

Examinees must be aware of the rules and regulations for the conduct of the examination and must abide by them.

### 9.3.4 Ensuring Credibility of Examinations through Scoring

The issue of scoring was treated in Module 5. Objective examinations by definitions are objectively scored either by persons or computers or other machines. Ordinarily, they should not have credibility problem if the procedures for test development earlier described are complied with. However, credibility problem may arise when the students' scores are deliberately changed by persons handling the scores. These persons could be computer operators or hackers or even typists.

The scoring of essay examination is subjective. Use of an appropriate and well-developed scoring rubric, earlier described, will drastically reduce elements of subjectivity. However, even with the scoring rubric, the examiner must be conscious of the tendency to commit unintended and intended errors and bias in his/her scoring. Bias can manifest itself in many forms, vis:

1. If an examiner is impressed with empty but flowery language in students' answers.
2. Personal preference for high-sounding vocabulary.
3. Knowledge of the examinee who is active in class and therefore should do well in an examination.
4. The examinee is a child or ward of an influential person.
5. Examinees being awarded scores on the basis of their relationship with the examiner, etc.

*Activity 9.3*

**Group Discussion:**

a. You find your colleague scoring an essay test referring to a list of students with their ID numbers and names. What advice will you give him/her? What would you tell him/her about the credibility of his/her examination?

b. If you examine the distribution of grades in a course, you find that most of the examinees got As, few Bs, and Cs and no other grades. What would you hypothesise as being responsible for the distribution? What would you say about the credibility of the examination?

## Unit 9.4: Credibility Issues in Reporting of Examinations

### 9.4.0 Introduction

This Unit examines the importance of accurately reporting students' examination performance and how credibility of examinations, certificates and institutions are undermined by faking of examination results.

### 9.4.1 Expected Learning Outcomes

At the end of this Unit, you will be able to:

- explain the importance of accurately reporting of examinees' examination performance,
- describe how a strong quality assurance system will assist in enhancing credibility of examinations, certificates, and institutions.

### 9.4.2 Importance of Accurate Reporting of Examination Scores

Examination results are reported either as individual scores or cumulatively as academic status or class of graduation. It is possible to take all credible steps prior to reporting examination scores, yet if they are not accurately reported, the credibility of the whole examination

system will suffer. It is the reported results that are used in decision making about the examinee, such as, the class of degree the examinee obtained, or whether an examinee should repeat a course or proceed to the next level.

One of the biggest credibility issues is change of scores and grades, or even the award of statements of results and certificates to examinees who still have courses to pass. This issue, deriving from its legal and ethical implications, also has credibility implications. Students may, in collusion with staff, academic and non-academic, and in some cases fraudulent individuals outside the tertiary institutions, forge results and even certificates while still having outstanding courses to pass or have been expelled from the Institutions. Many students have been withdrawn from Youth Service Camps because they presented fake results. In some cases, their competence is queried even in their places of work. Tertiary institutions, therefore, need to have a strong quality assurance system to ensure that the examination results presented by students are actually theirs.

***Activity 9.4***

**Discussion Points:**

a. It is sometimes reported in the mass media that Nigerian graduates are not employable. To what extent do you think this is a consequence of the credibility of the examinations system?
b. What can your institution do to enhance the credibility of its certificates?

### 9.4.3 Summary

The Unit observes that the biggest credibility issue in reporting of examination results is the faking of scores and forgery of results/certificates by examinees. This undermines not only the credibility of the examinations but institutions themselves. Institutions,

therefore, need strong quality assurance systems to address the challenges militating against the credibility of their examinations.

**References**

Capper, J. (1996). *Testing to learn learning to test*. Washington: Academy of Educational Development and International Reading Association.

Ngaware, M. (2017). *Four reasons why we need credible examinations in Sub-Saharan Africa*. www.learningportal.iiep.unesco.org.

Wakeford, R. (2003). Principles of students' assessment. In H. Fry, S. Ketteridge and S. Marshall(Eds), *A handbook for teaching and learning in higher education* pp.42-61. London: Kogan Page.

## MODULE TEN

# RECONCILING TESTEES WITH THE TESTS: A THERAPY AGAINST EXAMINATION MALPRACTICE

**by**
**Is-haq O. Oloyede**

### 10.0 Module Overview

Examination malpractice has come to assume a prominent dimension in educational assessment so much so that no discourse on educational assessment would be complete without a focus on it. It has been described as a cankerworm that has eaten deep into the fabrics of our education system. Not only has it been the subject of several panels, commissions, or committees, several bodies/agencies, organisations, and individuals have continued to wage what appears to be an endless war against this hydra-headed monster that tends to vitiate the credibility of respective educational assessments.

Rather than diminish in the face of the multifaceted approaches and strategies adopted in this war against this monster, it has continued to grow in sophistry and complexity. The fact that these multidimensional approaches which essentially focus on a system of punishment and reward have not succeeded in curbing or eliminating examination malpractice clearly suggests alternative psychometric approaches or principles.

These alternative psychometric approaches or principles are of course embedded in the major concepts, principles and activities, and issues to which you were exposed in the preceding nine modules. In this special module, we shall tease out these psychometric principles and examine how they can effectively curb examination malpractice. These principles are predicated on the assumption that test-induced phobia is a major factor that predisposes students to indulge in examination malpractice. In effect, examination malpractice can be curbed or eliminated by

eliminating this phobia or demystifying examination/test. Taking a cue from medical science, we can utilise the symbolism of medicine for disease treatment where there are curative, preventive, and palliative methods. In the same vein we can have curative, preventive, and palliative methods to the elimination of examination malpractice. The curative (punishment and reward) approach to examination malpractice will entail the usage of all available tools for eradicating the menace of examination malpractice. The preventive method approach will be to the block every loophole for examination malpractice. Palliative approach (which in medicine refers to the provision of a holistic care for patients who are no longer responsive to curative treatment) is for teachers and examiners to reduce examination phobia for students and candidates through the involvement of the testees in test-item generation, test construction, test administration and scoring. This palliative approach may demystify tests and reconcile tests with the testees.

## 10.1 Psychometric Principles that can Demystify Tests or Examinations

What are those psychometric principles or practices that can help to demystify tests or examinations and therefore reduce or eliminate examination phobia?

### 10.1.1 Integrated Approach to Assessment

In Module One, you were introduced to three (3) types of assessment - assessment of learning, assessment for learning, and assessment as learning. The three forms of assessment serve different functions and relevance in the educational practice at all levels. However, the major challenge resides in the proper integration of these forms of assessment in order to optimise their benefits. Current assessment practices tend to over-emphasise assessment of learning to the utter neglect of assessment for learning and assessment as learning. This has given rise to a situation where the learners tend to perceive examinations or tests as a means of

'condemning' rather than helping them to fully actualise their potentials. Consequentially, they develop excessive anxiety or phobia towards examinations or tests and would go to any length to succeed in such examinations or tests.

To demystify examination and reduce the incidence of examination malpractice, assessment as learning should be given the highest priority followed by assessment for learning and the assessment of learning.

### 10.1.2 Mainstreaming tests into the learning-teaching process

In the calendar of our schools and tertiary institutions, the period mapped out for learning is different from that mapped out for examinations or tests. This practice divorces tests from learning-teaching when in actual fact, the test is an integral part of the learning process. When testing is divorced from the learning process, learners develop negative dispositions toward testing and this will fuel the desire to indulge in one form of malpractice or the other. Consider, on the other hand, a system that integrates or mainstreams tests into the learning-teaching process whereby students have to do a test at the end of every lesson or unit as part of the normal learning-teaching process. The feedback from these tests is then used by the learners and/or the teachers to improve learning as well as track their performance. In such a system, there will be no special period called the 'examination period'; learners become accustomed to tests, fear of the tests is eliminated and the test would become non-threatening to learners. Tests being conducted while the testees are unconscious that they were being tested have been found to be more reliable. For example, you may wish to interview a candidate for a sensitive position to test his personality trait and his emotional intelligence. The testee may be invited for a social outing through which he or she is assessed while oblivious to the fact that the assessment is being done.

Finland is a country which has utilised minimisation (or elimination before age of 16) of examination to improve learning.

### 10.1.3 Designing Tests that cater for different abilities and Individual Differences

It is an obvious fact that learners differ in terms of their intellectual abilities. Given a class or group of learners, some will be of high ability while others may be of moderate or low ability. Since it is an individual's intellectual ability that sets the limit to what that individual could possibly do or achieve, it would not be appropriate to subject learners of varying abilities to the same test. This is exactly what happens in tests modelled after the classical test theory. In such tests, testees of varying abilities are made to take the test. How would one expect testees endowed with moderate or low intellect to compete with testees endowed with high intellectual ability on a test that demands high intellectual ability? Such a test will be lacking in equity and fairness. It is analogous to asking 5-, 12-, and 25-year-olds to carry a load that is appropriate for a 25-year-old. In the United Kingdom, age is normally used to classify testees' expected ability for assessment. This has been found to be fair, The rate of learning is also being measured and streamed into slow and too fast categories.

The implication of 'one cap fits all' test, is that the low-ability testees will continue to face frustration and may develop a feeling of alienation from tests. This is one of the major shortcomings of the classical test model. On the other hand, this testing challenge has been resolved by the Item Response Theory (IRT).

### 10.1.4 Democratisation of Test Development Process

A major principle of modern pedagogy that is known to promote effective learning is active learner participation in the learning process. However, while we encourage active learner participation in the learning process, we tend to exclude them from participation at the stage of testing or assessment. This truncates the principle of democratisation because assessment or testing is an integral part of the learning process.

It is against this background, that we see, in the democratisation of the test development process, great potential in reconciling testees with the tests.

Shohamy (2001) clearly identified four critical requirements in practicing democratic testing:

- shared authority,
- collaboration,
- the involvement of different stakeholders (test-takers included), and
- the monitoring of undemocratic uses of tests.

Democratisation of the test development process as conceptualised in this module focuses on the first three requirements namely; shared authority, collaboration, and the involvement of test-takers. This will entail giving the testees an opportunity to participate in the test development process. In practical terms, students are asked to generate test items after every learning session. These items are edited by the teacher and pooled together with those generated by the teacher to form the test instrument. In this way, item generation becomes a collaboration which is based on shared authority between the teachers and the students. This process will inspire confidence in the students since they know that they participated in generating the test items or in setting the examination. It is this confidence that the testees need to shun examination malpractice.

### 10.1.5 Engaging Students as Markers and Drafters of the Marking Scheme

The marking scheme specifies the teacher's expectation of how students should respond to the questions as well as the distribution of marks to the critical or salient points. Whosoever will score (or mark) the students' scripts are expected to strictly follow the marking scheme in allotting marks to the critical or salient points as identified in the marking

scheme. This will ensure that different teachers or scorers will arrive at similar score for the same student or testee. Thus, the marking scheme or guide is a useful tool for ensuring scorer reliability in essay-type tests. In peer grading, students are made to exchange their scripts and each student marks or scores the script of the other student (peer assessor). Normally, such students are given the marking guide prepared by the teacher which will guide them to score the script. This practise of involvement in peer assessment using the teacher's marking guide to score each other's performance will make them become familiar with the teacher's expectations of the correct or acceptable responses. This will reduce the stress and anxiety associated with examinations or tests.

It might even be more profitable to involve the students in the process of developing the marking scheme. Engaging students in developing the marking scheme implies that the marking scheme will now reflect the students' voices and no longer that of the teacher alone. This, of course, is in tandem with the principle of democratisation of the testing process which advocates for collaboration and involvement of all stakeholders including testees (Shohamy, 2001).

### 10.1.6 Using the Marking Scheme as Instructional Resource

Marking schemes describe responses that are considered correct or acceptable. These responses are further analysed and broken down into salient or critical points to which scores (or marks) are allotted. In essence, the score a student obtains will be determined by the salient points and their relative weighting. Using the marking scheme as an instructional resource will therefore expose the students to these salient points and make them become aware of relative importance of each of the units/sections based on the scores assigned to each. This will improve their capacity to respond appropriately to questions during examinations or tests. In addition, they will become more confident in facing tests and thus reduce the possibility of examination malpractice.

### 10.1.7 Open Book Examinations (OBE)

Conventional tests or examinations require testees to close their books and not bring into the examination venue any material related to the subject matter of the test or examination. Hence, they are referred to as Closed Book Examinations (CBE). One major limitation of this type of examination is that it tends to generate anxiety in the students as it compels students to read and memorise everything in their notebooks and textbooks with most of them not understanding what they have been taught. There is usually a noticeable increase in the frequency of students who report ill during such examination due to stress. We are all familiar with the 'examination fever' syndrome that characterises such examination periods under the CBE system. These factors tend to fuel the desire on the part of both the students and their parents to indulge in examination malpractice.

Against this background, the Open Book Examination (OBE) has been suggested as an alternative to CBE which can drastically reduce the incidence of examination malpractice. Open Book Examination is a type of examination system which allows the testee to bring into the examination hall and make references to notebooks, textbooks, and any other resources. There are basically two types of OBE. One is the traditional sit-down or limited time OBE. In this type, students come into the hall with their notebooks, textbooks, and other resources and are given the question paper with a time limit within which to complete the examination. A variant to this type as stated by Jain (2020) is that students are only allowed to bring books and materials approved by the authority. The second type is the take-home OBE. In this case, the students are given the questions to take home and submit their scripts after a given period of time. This is generally known as take-home-assignment.

A study by Heijne-Penninga, Kuks, Hofman, and Cohen-Schotanus (2008) reported that 'students preferred OB examinations because they

feel less anxious and more optimistic during examinations and consequently perform better'. This appears to be in tandem with the study by Loi and Teo (1999) which reported that 'students perform better in OB examination compared to CB examination, but only on items that required simple recall. This difference was not found on items at higher taxonomic levels.'

More importantly, evidence from the study by Mokdad and Aljunaidi (2021) suggests that the use of OBE contributes positively to the reduction of academic dishonesty.

In effect, if properly implemented, the Open Book Examination can lead to a significant reduction in examination malpractice. For example, in Russia, questions are set and given to all the students before the examinations. Questions for the examinations are selected from the questions compiled for the students. No question that has not been given to the students is allowed at the lower level. The fact must be underscored that to drive maximum benefits from OBE, the focus must be on the assessment of higher-order thinking skills as opposed to lower-order thinking skills. The question should not be such that demand mere recall of information that the students can readily copy from their notebooks or textbooks rather they should be questions that call for analysing, synthesising, creating, and evaluating.

## 10.2 Summary

In this special module, an attempt was made to put together the key issues that were discussed in the nine modules of this manual. The central message is to achieve a paradigm shift away from the traditional assessment model that heaps the whole responsibility of assessing students on the teacher who only uses tests or examinations for the purpose of judging and classifying students. In that traditional model, the teacher acts as the sole authority and tends to deny the student his/her democratic rights. The assessment instruments only reflect the teacher's

voice and the real essence of assessment which is to improve learning is often lost. This heightens students' and their parents' anxiety and phobia of tests or examinations; a condition that predisposes them to indulge in examination malpractice. The new assessment paradigm must therefore be such that can reconcile testees or students with tests or examinations by eliminating anxiety or fear and inspiring confidence in the students. It must give the students the necessary assurance that tests or examinations are for and not against their interests.

The central message of this module is the need to review our assessment policy and align it to best global practices. We need to brainstorm on how our students can be engaged in collaborative teaching, collaborative testing, and in collaborative curriculum development.

**References**

Ramamurthy, S., Er, H. M., Nadaraiah. V. D. & Pook. P.C.K. (2016). Study on the impact of open and closed book formative examinations on pharmacy students' performance, perception, and learning approach. Currents in Pharmacy Teaching and Learning, 8 (3): 364-374.

Heijne-Penninga, M., Kuks, J., Hofman, W. H. A., Cohen-Schotanus, J. (2008). Influence of Open and Closed-Book Tests on Medical Students' Learning Approaches MEd Educ.

Jain, R. (2020). Open Book Exams Vs Traditional Exams. In edumpus (2022). Retrieved in https://edumpus.com/blog/open-book-exams- vs-traditional-exams.

Loi, S. L., & Teo J. C. C. (1999). The Impact of Open Book Examinations on Student Learning http://www.hkta1934.org.hk.

Mokdad, M. & Aljunaidi, S. (2021). Combating Academic Dishonesty in

Higher Education Examinations during the COVID-19 Pandemic: The Role of Open Book Examinations. 2021 Sustainable Leadership and Academic Excellence International Conference (SLAE) **DOI**:10.1109/SLAE54202.2021.9686916.

Oloyede, I. O. (2017). Curbing Examination Malpractice: The Perspectives of Joint Admissions and Matriculation Board (JAMB). Paper presented at a two-day International Summit on "Examination Malpractice: The Contemporary Realities and Antidotes" organised by the West African Examinations Council (WAEC) on Thursday, 19th October, 2017.

Oloyede, I. O., (1990). Utilisation of the Principles of Ta'limu 'l-muta'llim for Islamization of the Modern Techniques of Learning, Muslim Education Quarterly, A Journal of the Islamic Academy, Cambridge, U.K, Spring Issue, Vol. 7, No. 3, 1990, pp. 8 –18, Also published in M.O. Opeloye (Ed) Arabic and Islamic Studies in Nigerian Schools: Challenges of 6-3–3–4 Educational Systems, NATAIS Publication, Lagos, 1991, pp. 61–70.

Shohamy, E. (2001). Democratic Assessment as an Alternative. October 2001. Language Testing 18(4):373-391 D01:10.1191/026553201 682430094. Available from: https://www.researchgate. net/publication/242487327 Democratic assessment as an alternative [accessed Apr 21 2022]

# INDEX

**A**
Ability tests, 40, 42, 43
Abisoye, Mubarak Oladipo, Mr, vi
Academics, iii
Achievement tests, 40, 42, 65, 66, 67
Adamu, Adamu Mallam, vi
Adewale, J. Gbenga Professor, vi, vii, 136
Afemikhe, O.A., vi, vii, 1, 4, 29, 34
Affective, 10
Agi, Christiana Dr (Mrs)., vi, vii, 103
Alex Ekwueme Federal University, Ndufu-Alike, Ikwo, vi
All assessment types, 29
Alternative Response items tests, 38, 39, 71
Amoo, Bosede O. Dr. (Mrs), vi
Anecdotal Records, 59
Aptitude tests, 42, 43
Ariyo, Akinyele. O., vi, vii,
a-Stratified Computerised Adaptive Testing (AS), 217
Assessment, iii
- for learning, 16
- Methods, 5
- of learning, 7, 8, 9, 10, 19
- Report, 197

Attitude Scale, 59

**B**
Basic education, 31, 32
Basic Education Certificate Examination (BECE), 32
Bayero University, Kano, vi,
Board, vi, vi

**C**
CBT (computer-based test), iv, xviii, 210, 211, 2013, 214, 218, 221, 222, 224, 227
CGPA, 193

Child Test Anxiety Scale (CTAS), 115
Classical Test Theory (CTT), 208
Classification of tests, 37
Closed Book Examinations (CBE), 248
Cognitive, 10
Committee of Experts, vi
Computer-Based, 5
- assessments, 17, 208
Computer-Adaptive Multistage Testing (ca-MST), 218
Computerised Adaptive Tests (CAT), 216, 223
Computerised Fixed Tests (CFT), 215
Concepts of Validity, 104
Constrained Adaptive Testing. 216
Constructing Essay, 68
Construct Validity, 108, 109
Continuous assessment (CA), 31, 32, 33, 34, 54, 89
Corrective Feedback, 194
Counsellors, 38
Criterion-Referenced tests, 41, 47, 48
Criterion-related Validity, 107
Cronbach Alpha Method, 120
Curricular and Instructional Validity, 232
Curricular expectations, 26
Curriculum, 9, 11

**D**
Democratisation of Test Development Process, 245, 246
Demonstrations, 5
Diagnostic tests, 41, 51
Differentiating Learning, 13, 20, 27
Docudramas, 5

**E**
Earl, L.M. 29, 34
Egbewole, Qasim Mr, vi,
e-Assessment, xvii, 204, 206
e-Testing, 200, 201, 202, 203, 209, 210
- Challenges, 225

- Software, 225
- System, xvii

Examination, 5
Examination Malpractice, 237, 242
Extended-Response Type Essay Items, 69
Evaluation, iii, vi

**F**
Fairness, 129, 130
Feedback, 185
Federal Republic of Nigeria, 31
Federal University of Agriculture, Makurdi, vi
Feedback, 13, 19, 27
Finland, 244
Formative tests, 41, 50
Free-Response items tests,
Formative, 3, 30, 31

**G**
Global Learning Consortium, 208
Grade Point Average (GPA), 189, 190, 191, 192
Guess Who Technique, 60
Guttman scale, 59

**H**
Habeeb, Suleiman Mr., vi
Habits of mind, 23
Hamman-Tukur, Abubakar, vi, vii, 231
Higher order thinking, 30
Highest learning priority, 30
Homework, 5

**I**
Ihedoro, Chidinma B. Mrs, vi
Individualised learning, 19
Information and Communication Technologies (ICT), 200, 227
Information and knowledge-based economy, 29
Integration of Assessment Types, 29, 30

Integrated vision of assessment, 30
Intelligence tests, 40, 42, 44
Inter-rater Reliability, 120
Item Vetting Process, 98
Institutes, iii
Institute of Chartered Accountant of Nigeria (ICAN), 48
Invigilation, 228

**J**
Joint Admissions and Matriculation Board (JAMB), iv, v, vi, 55. 183, 222,223
Joshua, Monday, vi, vii, 63
Junior School Arabic and Islamic Certificate Examination (JAISCE), 32

**K**
Katz, S., 29, 34

**L**
Lawal, R.A., vi, vii, 185
Learners, 30
Learning, viii, ix, x, xii, xiii, xiv, xvi, xix, 5, 8, 10, 18, 25, 31
- Conversations, 5
- Interviews, 5
- logs, 5

Learning Management System (LMS), 208, 209
Learning-teaching process, 244
Likert Scale, 59
Linear-on-the-fly testing (LOFT), 215
Local Area Network (LAN), 214

**M**
Makurdi, vi
Manpower Development Institute, iv
Manual, iii, iv, vi
Marking Scheme, 152, 153, 154, 155
Marking Scheme as Instructional Resource, 247
Matching Item Type, 72
Maximal Performance tests, 40
Measurement, iii, vi,

Metacognition, 22, 23,
Metacognitive analysis, 24, 25
Method of Administration, 38
Minimisation of examination, 244
Minister, of Education, vi
Model, v
Moderation of scores, 160, 161,
Module, iii, iv, viii
Monday, Joshua Prof, 36
Multistage Computerised Mastery Testing, 217
Multiple-Choice Items, 73

**N**
NABTEB, 55
National Business Certificate (NBC), 32
National Policy on Education, 31, 32
National Technical Certificate (NTC), 32
National Vocational Certificate (NVC), 32
Ndukwe, Emmanuel Dr., vi
NECO, 55
Nigeria Institute of Management (NIM), 48
Nigerian Medical Association (NMA), 48
Norm-Referenced Interpretation, 176
Norm-Referenced Tests, 41, 46, 55
Nsukka, vi
Nworgu, B. G., vi, vii, 164
Nigeria, iii, v,
Non-Governmental Academy, iv

**O**
Objectivity, 125, 126, 127,
Objective Test Items, 68
Observation, 5
Observation technique, 58
Oloyede, Is-haq O., Professor, v, vi, vii. 242, 251
Onyekwere, Ijeoma H. Mrs, vi
Open Book Examinations (OBE), 248
Osgood scale, 59

Owolabi, O. Henry, Professor, vi, vii, 200
Oyebanji, Eniola, vi

**P**
Paper-and-pencil test forms (PPT), 215, 222
Peer Appraisal Instruments, 60
Personality Measures, 59
Placement tests, 41, 49
Planning instruction, 19
Portfolio, xii , 93, 94, 96
Post-Basic Education Certification Examinations (PBECE), 33
Post-Unified Tertiary Matriculation Examination (PUTME), 181, 183,
Power tests, 39
Presentations, 5
Proctoring, 228
Professional Assessment Academy of Nigeria, iv
Progressive Feedback, 195
Projects and Investigations, 6
Psychological test, xv, 165, 166, 167
Psychologists, 38
Psychometrics, iii, iv,
Psychometric Principles, 243
Psychometric Properties of Tests, 226
Psychomotor, 10
Psychomotor domain, 58, 61

**Q**
Questioning, 5
Quizzes, 5

**R**
Rating scale, 58
Record-Keeping, 12, 19, 27
Reference Points, 12, 19, 26
Registrar, v
Reliability, 18, 26, 113, 114, 121
Remedial Feedback, 195
Reporting, 187, 188, 196

Research skills, 6
Restricted-Response Type Essay Items, 69
Rich assessment tasks, 5

**S**
Secure site online assessment, 213
Secure Sockets Layer, 209
Scoring, 135
Scoring Rubrics, 145, 146, 148, 150
Secondary schools, iv,
- education, 32

Self-Appraisal Technique, 60
Self-management, 30
Senior Arabic and Islamic Studies Certificate (SAISC), 32
Senior School Certificate (SSC), 32
Senior Secondary School, 32
Simple recall, 249
Simulations, 5
Split-Half Method, 118
Speed tests, 39, 51
Standard Error of Measurement (SEM), 174
Standardised tests, 54, 55
Strategies for combating Examination Malpractices, 143
Supply Item, 71
Sociometric Technique, 60
Summative, 3, 41
- assessment approaches, 30, 31

**T**
Teachers, iv, 38
Teacher-made tests, 54, 56
Teachers' Registration Council of Nigeria (TRCN), 48
Test Administration, 136
Test Development Skills Gap, 226
Testlet-Based CATs, 217
Test Purpose, 42
Test Score, 168, 170, 171, 172, 175
Test-Retest Method, 115, 116

Testing, 136, 140
Testimonials, 31
Thurstone scale, 59
Typical Performance tests, 40
Typology of Educational Measurement, 36

**U**
Ugodulunwa, C.A., vi, vii, 65
Unified Tertiary Matriculation Examination (UTME), iii, 47, 183
University of
Benin, vi
Calabar, vi
Ibadan, iv, vi
Ilorin, vi, vi, 222
Maiduguri, vi
Nigeria, vi
Usability, 132

**V**
Validity, 18, 26, 110, 111, 112
Virtual Learning Environment, (VLE), 208

**W**
Web-based Online Assessment, 212
Web-based testing, 212
West African Senior School Certificate (WASSC), 32
William, D., 35
World Wide Web, 200

**Y**
Yakasai, Muhammad I., vi, vii