

# *Issues in Contemporary Evaluation*

UNIVERSITY OF IBADAN LIBRARY

Edited by

Olajide Olorunnisola  
Adams Otuoze U. Onuka  
Oyebamiji Babalola  
Aderemi I. Alarape  
S. A. Babarinde

ISSUES IN CONTEMPORARY EVALUATION

UNIVERSITY OF IBADAN LIBRARY

Edited by  
Olatide Olorunsola  
Adams Oluwole O. Ousika  
Oyebanjo P. Salola  
Aderemi I. Alarape  
S. A. Babalola

# ISSUES IN CONTEMPORARY EVALUATION

UNIVERSITY OF IBADAN LIBRARY

Edited By  
**Olajide Olorunnisola**  
**Adams Otuoze U. Onuka**  
**Oyebamiji Babalola**  
**Aderemi I. Alarape**  
**S. A. Babarinde**

ISSUES IN CONTEMPORARY EVALUATION

Published by  
The Postgraduate School,  
University of Ibadan,  
Ibadan, Nigeria.

E-mail: [postgrad@mail.ui.edu.ng](mailto:postgrad@mail.ui.edu.ng)

Website: <http://www.postgraduateschool.ui.edu.ng>

First Published 2013

No part of this publication may be reproduced, stored in a retrieval system or transmitted by any form or by any means, electronic, mechanical, photocopying, recording, or otherwise without the expressed written consent of the copyright owner.

All rights reserved

ISBN 978- 978-50948-7-9

Printed by Sapphire Prints  
08034892535. 08070709894

Olajide Olorunnisola  
Adams Oluwole U. Onyika  
Aderemi I. Alarape  
S. A. Babarinde

## Contents

<b>Foreword</b>	<b>vii</b>
<b>Preface</b>	<b>viii</b>
<b>Acknowledgements</b>	<b>x</b>
<b>Notes on Contributors</b>	<b>xii</b>
<b>Chapter 1</b>	
Issues in Contemporary Evaluation: An Overview <i>Mac Ade Araromi</i>	<b>1</b>
<b>Chapter 2</b>	
Concepts in Evaluation including Evaluation For and of Learning <i>Charles V. Abe</i>	<b>5</b>
<b>Chapter 3</b>	
Determining Learning and Evaluation Objectives ( 1) <i>Monica N. Odinko</i>	<b>15</b>
<b>Chapter 4</b>	
Determining Learning and Evaluation Objectives (2) <i>Ikmat O. Junaid &amp; Ifeoma M. Isiugo-Abanihe</i>	<b>27</b>
<b>Chapter 5</b>	
Test Construction and Administration <i>Joseph A. Adegbile</i>	<b>36</b>
<b>Chapter 6</b>	
Continuous Assessment: Concept, Practices and Tools <i>Monica N. Odinko</i>	<b>45</b>

## Contents

<b>Chapter 7</b> Non-Cognitive Evaluation Tools <i>Benson A. Adegoke</i>	63
<b>Chapter 8</b> Construction of Non-cognitive Instruments <i>Eugenia A. Okwilagwe</i>	73
<b>Chapter 9</b> Test Security <i>Adams O.U. Onuka</i>	88
<b>Chapter 10</b> Evaluation in Science <i>Modupe M. Osokoya &amp; Tolulope.W. Yoloye</i>	97
<b>Chapter 11</b> Evaluation in the Humanities and Social Sciences <i>Folajogun V. Falaye &amp; Serifat F. Akorede</i>	111
<b>Chapter 12</b> Test Development <i>Joshua O. Adeleke</i>	128
<b>Chapter 13</b> Evaluation as Feedback Mechanism for Quality Assurance and System Improvement <i>Folajogun V. Falaye</i>	151
<b>Chapter 14</b> Item Banking <i>Adams O. U. Onuka</i>	167
<b>Chapter 15</b> Computer Based Testing: Principles and Practice <i>Jeremiah G. Adewale</i>	179

## Foreword

*Issues in Contemporary Evaluation* is a product of collaborative efforts between the University of Ibadan Postgraduate School and the Institute of Education. It explores the current issues of best practices in evaluation in assuring the provision of quality education that meets up with contemporary global requirements of higher education. The original idea was to conduct a workshop on modern evaluation practices in higher institutions. However, when a training workshop was not feasible, it was agreed that a book should be developed at least in the interim for the use of postgraduate lecturers so that they could personally undertake some basic individual and corporate evaluation which consists of tests, measurement, assessment and evaluation for improved teaching and learning.

Topics discussed in the book include concepts like evaluation *for* learning and evaluation *of* learning, the principles and practice of continuous assessment, determining learning objectives and concomitantly evaluation objectives. Other topics covered are: item banking, test security and feedback mechanism for quality assurance in the higher education system, evaluation practices in the sciences and in the humanities and social sciences. Test construction and development and computer-based testing were also reviewed.

The book has made it very clear that evaluation is more than examining or testing, but a holistic approach to quality assurance and quality control while a number of often misconstrued concepts in evaluation were clarified. This compendium is recommended to Postgraduate teachers as a guide to evaluation practices.

Olajide Olorunnisola  
Dean, Postgraduate School,  
University of Ibadan, Nigeria  
March, 2013

## Preface

*Issues in Contemporary Evaluation* highlights the issues of best practices in modern evaluation. Basically, the book is designed to be a manual for lecturers and other teachers in tertiary institutions in Nigeria in particular and Africa in general. It is the modest contribution by the staff of the Institute of Education, University of Ibadan to enhance Postgraduate teachers' capacity in current trends in evaluation. It must be stated that this is just the first of such contributions as evaluation covers many facets of human endeavour and not just students' achievement. A well-designed and implemented comprehensive evaluation affects a whole lot of human endeavours and continues to engender holistic educational programme development. Thus, several important evaluation items are covered. Theories [principles] and practices were succinctly discussed by experts in various aspects of evaluation. A number of usually misconstrued concepts in evaluation such as measurement, assessment and evaluation were clarified. Various ways of carrying out tests as well as numerous types of tests were enunciated. The basic ways of determining learning objectives which invariably gives birth to evaluation objectives were brought to the fore.

The distinction between evaluation *for* and *of* learning was clearly stated. The fact that these concepts are however, congruent to the concepts such as 'formative' and 'summative' evaluation was stated. The principles and practice of continuous assessment [CA], which is a form of assessment or evaluation *for* learning, were clearly discussed in detail and the various components of CA were highlighted. The book obviously showed that for CA to be comprehensive, it has to take cognisance of the three domains of learning, if it were to achieve its objective of developing the total learner as well as improve learning outcomes. Other topics treated in this book include: Item banking, test security, feedback mechanism for quality assurance in the higher education system, evaluation practices in the sciences and in the humanities and social sciences.



## *Preface*

Test construction and development and computer-based testing were reviewed in every material detail. These topics were covered by renowned evaluators which include Professor T.W. Yoloye, Dr. Ifeoma M. Isiugo-Abanihe, Drs. J. A. Adegbile, Folajogun V. Falaye, J. G. Adewale, Modupe M. Osokoya, Eugenia A. Okwilagwe, Adams O. U. Onuka, Monica N. Odinko, B. A. Adegoke, J. O. Adeleke, Serifat F. Akorede and Ikmat O. Junaid. The book is a must read for teachers who desire to make meaningful impact on their students through unbiased assessment of their educational worth and for improving their educational worth. It is, therefore, commended to colleagues as an evaluation companion.

**Adams O. U. Onuka**

UNIVERSITY OF IBADAN LIBRARY

## Acknowledgement

The following persons are hereby acknowledged for their invaluable contributions to making the debut of the book a huge success: Professors Mac Araromi, Director, Institute of Education, Prof. Abel Olajide Olorunnisola, Dean, Postgraduate School, University of Ibadan for the encouragement through sponsorship of the book and being the lead editor, Prof. T. W. Yoloye, a former Director, Institute of Education, Prof. Adenike Emeke, a senior colleague at the Institute of Education, Drs. Onuka, Alarape and Babalola for being co-editors. Immense gratitude is hereby duly extended to contributors of chapters: Professor T. W. Yoloye, Drs. Ifeoma M. Isiugo-Abanihe, J.A Adegbile, Folajogun V. Falaye, J. G. Adewale, Modupe M. Osokoya, Eugenia A. Okwilagwe, A.O.U. Onuka, Monica N. Odinko, B.A. Adegoke, J. O.Adeleke, Serifat F. Akorede and Ikmat O. Junaid.

## Notes on Contributors

**Mac Ade Araromi** is a Professor and Director of the Institute of Education, University of Ibadan. He specializes in Language Education.

**Charles Vincent Abe** is an Adjunct Senior Research Fellow at the Institute of Education, University of Ibadan, Nigeria. He is a specialist in Monitoring and Evaluation and Social Science Education. He was a Sub-Dean (Postgraduate) at the Institute.

**Monica Ngozi Odinko** is a Senior Research Fellow at the Institute of Education, University of Ibadan and Head of the ICT unit of the Institute. She is also the Coordinator of the M.Ed-in-Service Programme. She specializes in Early Childhood Education and Evaluation.

**Ikamat Olanrewaju Junaid** is a Research Fellow II and a specialist in Distance Learning and Special Education and Evaluation at the Institute of Education, University of Ibadan, Nigeria.

**Ifeoma Mercy Isiugo-Abanihe** is a Reader and Head of the International Centre for Educational Evaluation, Institute of Education, University of Ibadan. She was a two-time Sub-Dean (PG) at the Institute of Education. She is a specialist on Gender Issues and Language Education and Evaluation.

**Joseph Abiodun Adegbile** is a Reader at the Institute of Education, University of Ibadan. His specialty is in Language Education and Evaluation. He is the immediate past Head of the outreach services unit of the Institute.

**Benson Adesina Adegoke** is a Research Fellow I at the Institute of Education, University of Ibadan and currently the examinations Coordinator at the Institute. He specializes in Physics Education and Evaluation.

**Eugenia A. Okwilagwe** is a Senior Research Fellow and presently Head of School Services unit of the Institute of Education, University of Ibadan. She is a specialist in Social Science Education and Evaluation.

### *Notes on Contributors*

**Adams Otuoze Umoru Onuka** is a Senior Research Fellow at the Institute of Education. He is the Director of the Centre for Social Orientation and the current Head of the Outreach Services Unit of the Institute of Education, University of Ibadan. He is a specialist in Business and Management Education and Evaluation. He is the coordinating editor of this book.

**Modupe M. Osokoya** is a Senior Research Fellow and currently the Sub-Dean (PG) at the Institute of Education, University of Ibadan. She is a specialist in Science Education and Evaluation.

**Tolulope Wale Yoloye** is a Professor and former two-time Director, Institute of Education, University of Ibadan. He is a specialist in Learning and Science Education.

**Folajogun Veronica Falaye** is a Senior Research Fellow and immediate past Sub-Dean (PG) at the Institute of Education, University of Ibadan, Nigeria. She is a town Planner and Educational Evaluator in Social Science Education.

**Joseph Oluwatoyin Adeleke** is a Research Fellow I and Admissions Coordinator at the Institute of Education as well as a specialist in Mathematics Education and Evaluation at the Institute of Education, University of Ibadan.

**Jeremiah Gbenga Adewale** is a Senior Research Fellow in Physics Education and Evaluation at the Institute of Education, University of Ibadan. He is the current Head of Research, Consultancy and Publications at the Institute.

**Serifat F. Akorede**, a Research Fellow I is a specialist in Economic Education and Evaluation at the Institute of Education, University of Ibadan, Nigeria.

## Chapter 12

### Test Development

Joshua O. Adeleke

#### Introduction

Test is an important tool in any educational system, because it is the major means by which the level of attainment or accomplishment of educational outcomes (whether cognitive or psychomotor) is determined. Test can be theoretical or practical or both. In other words, test is usually utilized to determine the level to which a particular knowledge area and/or skill has been acquired by a learner or participant in a course. Therefore, test can be seen as a regulated procedure to sample behaviour as well as to describe the behaviour in terms of scores or categories. In a nutshell, the essence of test does not lie in just determining achievement, but also to predict ability and other types of behavior. There are many forms of test, such as achievement, aptitude, affective etc. This module focuses on achievement test.

#### Achievement Test

An achievement test should fully measure the status of the individual in all the hierarchical levels of understanding as proposed in Bloom's taxonomy of educational objectives. The test should measure:

- Recall of information (knowledge),
- Understand the meaning, translation, interpolation, and interpretation of instructions and problems. Stating a problem in one's own words (comprehension),
- Using a concept in a new situation or unprompted use of an abstraction. Applying what was learned in the classroom to novel situations in the workplace. (application),
- Separating material or concepts into component parts so that its organizational structure may be understood. Distinguishing between facts and inferences (analysis),
- Building a structure or pattern from diverse elements. Putting parts together to form a whole, with emphasis on creating a new meaning or structure. (synthesis), and;

- Making judgments about the value of ideas or materials (evaluation).

Test experts have classified achievement tests using different parameters. Whereas, some classify tests on the basis of the behaviour that is being measured, others classify considering the types of items contained in the test, the purpose of tests, etc. However, achievement tests may be classified on the basis of the **essay-type** and the **objective-type**. The two major types are presented in the table below:

**Table 12.1: Types of Achievement Test**

Essay	Objectives	Others
1. Extended Response	1. Fill in a. Short Answer	1. Oral
2. Restricted Response	2. Selective Type a. True or False b. Matching Items c. Multiple Choice Question	2. Student Portfolios 3. Performance

### Essay Tests

The essay test has been a very popular type of achievement test. It is a test that allows the testees to apply their ideas on the items in a personal way. The two forms of essay tests are Extended Response and Restricted or Short-Answer Tests.

### The Extended Response

In this type of essay test, the testee answers a small number of items. The examiner is expected to develop a valid marking scheme to award marks on each item. The testees may be instructed to answer all the questions or to choose out of the number of items given. Instructions that guide the testees on how to write the test are very important.

Two examples of the extended response type are:

- Differentiate between Bar chart and Histogram.
- Discuss methods of establishing the Reliability Coefficient of an Achievement Test.

*Extended Response test is good for:*

- Application, synthesis and evaluation levels

*Advantages:*

- Students less likely to guess
- Easy to construct
- Stimulates more study
- Allows students to demonstrate ability to organize knowledge, express opinions, show originality.

*Disadvantages:*

- Can limit the amount of material tested, therefore has decreased validity.
- Subjective, potentially unreliable scoring.
- Time consuming to score.

*Tips for Writing Good Extended Response Items:*

- State the instructional objectives in specific terms
- Outline the course content
- Prepare the table of specification (test blue print)
- Provide reasonable time limits for thinking and writing.
- Avoid letting them answer a choice of questions (Instruct all the testees to answer the same set of items.)
- Give definitive task to student-compare, analyze, evaluate, etc.
- Use checklist point system to score with a model answer: write outline, determine how many points to assign to each part
- Score one question at a time-all at the same time.

### **The Restricted or Short-Answer Essay**

In this kind of test, the candidate is given a number of questions to respond briefly to. It limits both the content and the type of learner's response. The following are examples of such a question:

1.
  - Mention any four functions of an Achievement test.
  - Write about six sentences on any 3 of them.
2. List the two major types of Achievement test and discuss only one of them.

### Short Answer items are good for

- Application, synthesis, analysis, and evaluation levels

#### Advantages:

- It is easy to construct
- It minimizes guessing
- It encourages more intensive study that students must know the answer rather than just recognizing the answer.

#### Disadvantages:

- It may overemphasize memorization of facts
- Testees may give different types of answers on an item.
- Scoring is laborious

#### Tips for Writing Good Short Answer Items:

- State the instructional objectives in specific terms;
- Outline the course content;
- Prepare the table of specification (test blue print);
- Provide reasonable time limits for thinking; and
- For numbers, indicate the degree of precision/units expected.

### Objective Test

It is a test consisting of factual questions requiring extremely short answers that can be quickly and unambiguously scored by anyone with an answer key, thus minimizing subjective judgments by both the person taking the test and the person scoring it.

#### Types of Objective Test Items

*Objective tests are of various types:* The commonly used among the types of Objective test are:

- the true-false type,
- the fill-in type, i.e. short answer or completion,
- the matching type, and
- the multiple choice type,

#### The True-False Type

In this type of test, the testee is given some statements to which s/he should respond. The statements have to be marked as either "true" or



false". Let us consider the following examples:

- The triangle is an example of polygon. (True or False).
- The state capital of Rivers is Port Harcourt. (True or False).

**True/False items are good for:**

- Knowledge level content;
- Evaluating students' understanding of popular misconceptions; and
- Concepts with two logical responses

**Advantages:**

- They can test large amounts of content; and
- Students can answer 3-4 questions per minute

**Disadvantages:**

- They are easy to construct
- It is difficult to discriminate between students that know the material and students who do not
- Students have a 50-50 chance of getting the right answer by guessing
- Need a large number of items for a high level reliability

**Tips for Writing Good True/False items:**

- Avoid double negatives.
- Avoid long/complex sentences.
- Use specific determinants with caution: never, only, all, none, always, could, might, can, may, sometimes, generally, some, few.
- Use only one central idea in each item.
- Don't emphasize the trivial.
- Use exact quantitative language
- Don't lift items straight from the book.
- Make more false than true (60/40). (Students are more likely to answer true.)

**The Fill-in Type (Short Answer or Completion)**

This is another simple type of objective test where a testee is expected to provide short answers or complete some statements.

Examples:

- What is the name of a side of a triangle that is opposite a right angle?
- The name of a side of a triangle that is opposite a right angle is .....

**Fill-in type of Objective test items is good for**

- Knowledge level content
- Evaluating student understanding of popular misconceptions

*Advantages:*

- Easy to construct
- Good for recalling date, idea, fact etc.
- Minimizes guessing
- Encourages more intensive study. Students must know the answers
- Scoring is easy.

*Disadvantages:*

- May overemphasize memorization of facts
- Take care - questions may have more than one correct answer

*Tips for Writing Good Fill-in type Items:*

- When using with definitions: supply term, not the definition-for a better judge of student knowledge.
- For numbers, indicate the degree of precision/units expected.
- Use direct questions, not an incomplete statement.
- If you do use incomplete statements, don't use more than 2 blanks within an item.
- Arrange blanks to make scoring easy.
- Try to phrase the questions so there is only one answer possible.

### **The matching type**

As the name denotes, matching type presents two columns containing domains of elements that can be independently assigned.

**Matching Items are good for:**

- Knowledge level
- Some comprehension level, if appropriately constructed

## Example

Use straight lines to match the items below

Question Column		Answer Column
Circle	goes to	Diagonal
Rectangle		Edges
Cube		Average
		Mode
		HCF
		LCM
		Radius

*Types:*

- Terms with definitions
- Phrases with other phrases
- Causes with effects
- Parts with larger units
- Problems with solutions

*Advantages:*

- Maximum coverage at knowledge level in a minimum amount of space/prepare time
- Valuable in content areas that have a lot of facts

*Disadvantages:*

- Time consuming for students
- Not good for higher levels of learning

*Tips for Writing Good Matching items:*

- Need 15 items or less.
- Give good directions on basis for matching.
- Use items in response column more than once (reduces the effects of guessing).
- Use homogenous material in each exercise.
- Make all responses plausible.
- Put all items on a single page.
- Put response in some logical order (chronological, alphabetical, etc.).
- Responses should be short.

### The multiple choice type

The multiple-choice (MC) item is one of the most popular item formats used in educational assessment. A typical MC item has three parts: a stem that presents a problem; the correct or best answer; and several distractions (i.e., the wrong or less appropriate options).

MC items can be constructed to assess a variety of learning outcomes, from simple recall of facts to Bloom's highest taxonomic level of cognitive skills – evaluation (Osterlind, 1998). It is common knowledge that the correct answers should be distributed evenly among the alternative positions of MC items, but there are many other important guidelines for writing good items. Eight steps though not exhaustive are provided as guidelines to construction of multiple choice questions.

#### Multiple Choice items are good for:

- Application, synthesis, analysis, and evaluation levels

#### Types:

- Question/Right answer
- Incomplete statement
- Best answer

#### Advantages:

- Very effective
- Versatile at all levels
- Minimum of writing for student
- Guessing reduced
- Can cover broad range of content

#### Disadvantages:

- Difficult to construct good test items.
- Difficult to come up with plausible distractors/alternative responses.

#### Tips for Writing Good Multiple Choice items:

- Stem should present single, clearly formulated problem.
- Stem should be in simple, understood language; delete extraneous words.

- Avoid "all of the above"--can answer based on partial knowledge (if one is incorrect or two are correct, but unsure of the third...).
- Avoid "none of the above."
- Make all distractors plausible/homogenous.
- Don't overlap response alternatives (decreases discrimination between students who know the material and those who don't).
- Don't use double negatives.
- Present alternatives in logical or numerical order.
- Place correct answer at random (A answer is most often).
- Make each item independent of others on test.
- Way to judge a good stem: students who know the content should be able to answer before reading the alternatives
- List alternatives on separate lines, indent, separate by blank line, use letters vs. numbers for alternative answers.
- Need more than three alternatives, four is best.

### **Specific Guidelines for Constructing MC Items**

#### ***State the objectives of the domain of knowledge to be assessed***

The test developers can list major topics covered or expected to be covered in a term, semester or session if the focus of the test is summative assessment. The components of a unit of instructional content can form the objective of the domain if formative assessment is the focus.

#### ***Prepare Table of Specification***

The next task is to prepare a table of specification (Test Blueprint) to cover appropriate levels of cognition using Bloom's taxonomy of learning outcomes: Knowledge, Comprehension, Application, Analysis, Synthesis and Evaluation. The cognition level of the testees should be considered while preparing the test blueprint. It is a table that has rows as the list of topics and columns as cognition levels. An example of test blueprint adaptable for developing multiple choice items on mathematics is presented below. The proportion of time spent on each topic can be a guide while distributing items on topics. Appendix I provides details on Bloom's taxonomy of learning outcomes.

**Test Blueprint**

Content or Topic	Know 25%	Comp 20%	Appl 20%	Analy 15 %	Synt 15%	Eval 5%	Total 100%
Number and Numeration (24%)							24
Mensuration (10%)							10
Statistics/ Prob (16%)							16
Algebraic Proc (30%)							30
Plane Geo (8%)							8
Trigonometry (12%)							12
<b>Total</b>	25	20			15	5	100

Note: The percentages assigned to the cognition levels are suggested.

In addition to Table of Specification, filling system should also be developed to ensure effective scoring and analysis of the scores.

**Filling System**

S/N	Topic / Content	Key	Level of Cognition
1.	Number and Numeration	B	Knowledge
2.			
3.			
4.			
.			
.			
.			
.			
.			
.			
100			

Test blueprint should include item number, level of cognition, topic from which each item was drawn and the key. The table above presents a sample of filling system.

**The stem should be meaningful by itself and should present a definite problem**

A common fault in MC item writing is to have a brief, meaningless stem with problem definition revealed in the options. In such cases, it can be difficult to see the intent of the item after reading the stem. To write a focused item, we should include the central idea in the stem instead of the options. In Item 1, the stem does not present a definite problem.

ITEM 1

A triangle

- A. Can be constructed by a ruler and a pencil only.
- B. possesses four interior angles and three exterior angles.
- \*C. has three sides.
- D. is a three dimensional shape.

The correct answer is indicated with an asterisk. Students are faced with four true-false options; each is about a triangle, but only option C is correct. Furthermore, the four options cover a set of widely dissimilar ideas about a triangle so that evaluation by comparison is not possible. The stem can be judged to be clearly presenting a problem if it forces the options to be parallel in type of content. Item 2 demonstrates one way to make the stem become a definite problem. Students can think about the correct answer rather than figuring out what the problem is. Also, the clearly stated problem in the stem has forced the four options to be parallel in content.

ITEM 2

How many sides has a triangle?

- A. 1; B. 2; \*C. 3; D. 4

**The use of internal or beginning blanks in completion-type MC items should be avoided**

The stem may be written as an incomplete statement that needs to be completed by insertion of the correct option. Measurement specialists have advised not to use the completion format because a student has to retain the stem in short-term memory while completing the stem with each option. Test anxiety is even higher if the student is not a native English speaker. If the completion format is unavoidable, the omission

should occur toward the end of the stem rather than in the middle or at the beginning of the stem as shown in Item 3. Item 4 shows an improved version.

## ITEM 3

---- has four lines of symmetry.

- A. Kite; \*B. Square; C. Rectangle; D. Circle

## ITEM 4

Which of the following has four lines of Symmetry?

- A. Kite; \*B. Square; C. Rectangle; D. Circle

**Use a negatively stated stem only when significant learning outcomes require it**

Most students have difficulty understanding the meaning of negatively phrased items. They often read through the negative terms such as *not*, *no*, and *least*, and forget to reverse the logic of the relation being tested. For example, Items 5 and 6 assess the same concept of chemistry, but some students may answer Item 5 incorrectly merely because of the word *least*. Since *least* and *concentrated* are opposites, the phrase *least concentrated* is more difficult to understand than the phrase *most concentrated*. Research by Cassels and Johnstone (1984) has confirmed that the change from *least concentrated* to *most concentrated* will increase the percentage of correct responses.

## ITEM 5

Which of the following solutions is the least concentrated?

- A. 50 g of calcium carbonate in 100 cm<sup>3</sup> of water  
 B. 60 g of sodium chloride in 200 cm<sup>3</sup> of water  
 C. 65 g of potassium nitrate in 100 cm<sup>3</sup> of water  
 \*D. 120 g of potassium sulphate in 200cm<sup>3</sup> of water

## ITEM 6

Which of the following solutions is the most concentrated?

- A. 50 g of calcium carbonate in 100 cm<sup>3</sup> of water  
 B. 60 g of sodium chloride in 200 cm<sup>3</sup> of water  
 \*C. 65 g of potassium nitrate in 100 cm<sup>3</sup> of water  
 D. 120 g of potassium sulphate in 200 cm<sup>3</sup> of water



Although negatively phrased stems should generally be avoided, they are useful if we want to assess whether students can identify dangerous laboratory practices that may damage expensive equipment or result in bodily injury, and which should not be carried out. Item 7 is an example of such an item. However, when a negative term is used, it should be emphasized by being underlined or capitalized. Replacing the negative term with the word *except* can sometimes improve clarity, as illustrated in Item 8. Few students would overlook the negative element in the stem because the word *except* is deliberately placed at the end of the stem and is capitalized.

ITEM 7

Water-type extinguisher is not suitable for putting out fire caused by burning

\*A. alcohol. B. cotton. C. paper D. wood.

ITEM 8

Water-type extinguisher is suitable for putting out fire caused by burning all of the following

EXCEPT

\*A. alcohol. B. cotton. C. paper D. wood

**Irrelevant difficulty should be avoided**

The difficulty of an item should not be increased by incorporating more complicated information in the stem than is necessary. For example, if we want to assess whether students can solve dilution problems using the concept of molarity, Item 9 contains confounding detail. The values used in Item 10 will assess the same learning outcome and will avoid irrelevant sources of difficulty and error.

ITEM 9

A pentagon like any other polygon has how many sides?

\*A. 5 sides; B. 6 sides; C. 7 sides; D. 8 sides

The words "like any other polygon" are irrelevant. Item 9 can be reworded as

## ITEM 10

A pentagon has how many sides?

\*A. 5; B. 6; C. 7; D. 8

**All distractors should be plausible**

Designing plausible distractors is the most difficult part of MC item writing. A good distractor should be selected by low achievers and not by high achievers. To construct plausible distractors, teachers are encouraged to use common misconceptions. For example, the correct answer in both Items 11 and 12 is 7. Many students are familiar with the number of sides of a triangle and a square; they can easily eliminate the distractors to pick the right answer unlike in question 12.

## ITEM 12

A Heptagon has how many sides?

A. 2; B. 3; C. 4; D. 7

## ITEM 12

A Heptagon has how many sides?

A. 5; B. 6; \*C. 7; D. 8

**Avoid the use of complex MC format**

Sometimes, teachers like to design complex MC items to make them harder. A complex MC item consists of a list of potentially correct answers called primary responses and a list of combinations of the primary responses called secondary options. Students have to select one of the secondary options in answering the item, as shown in Item 13. This item is equivalent to a set of four true-false items, but knowing that a particular primary response is correct or incorrect would help the examinee identify the correct secondary option by eliminating distractors (Ebel & Frisbie, 1991; Haladyna, 1999). For example, if students know that the primary response "rectangle" is untrue, they tend to pick option D because sulphur dioxide does not appear in options A and D and more than one primary response are usually included in the correct answer to a complex MC item. Although the complex MC format may make the items more difficult, research reports (Albanese, 1993) reveal that it is less discriminating and reliable than the single-answer format.

## ITEM 13

Which of the following shapes is/are 3 dimensional?

(1) Cuboid; (2) Rectangle; (3) Frustum; (4) Prism

A. (3) only; B. (1) and (2) only; C. (2) and (4) only; \*D. (1), (3) and (4)

**The relative length of the options should not provide a clue to the answer**

Teachers are mostly unaware of this item-writing principle (Rodriguez, 1997). It is common to express the correct response more carefully and at greater length than the distractors. However, research (Chase, 1964) has indicated that longer options tend to result in higher response rates. In Item 14, testwise students will notice that option B is much longer than the other options. Even without a good understanding of the concepts of Energy and Matter, they will guess that the correct answer is B because it stands out from the others. Note that the longest options may not be correct.

## ITEM 14

Energy like matter can:

- be destroyed and changed to matter.
- never be destroyed but converted from one form to another.
- be changed from solid to liquid. D. sublime

**Avoid using "none of the above" or "all of the above" as an option**

The use of *none of the above* and *all of the above* as options in MC items is tempting to many teachers because they appear to fit easily into many items. However, many measurement specialists do not recommend the use of the option *none of the above*. For example, the correct answer for Item 15 is option D. A student may explain this way: "The correct answer is none of the above because, as everyone knows, hydrogen relights a glowing splint." Another student may be surprised to hear that explanation: "What! The correct answer is not hydrogen, but sulphur dioxide." It does not matter; neither gas is listed. Thus, the correct answer could be selected Using wrong ideas. This item may be modified to form Item 16.

ITEM 16

Which of the following substances would relight a glowing splint?

- A. carbon dioxide; B. chlorine; C. nitrogen; \*D. none of the above

ITEM 16

Which of the following substances would relight a glowing splint?

- A. carbon dioxide; B. chlorine; C. nitrogen; \*D. oxygen

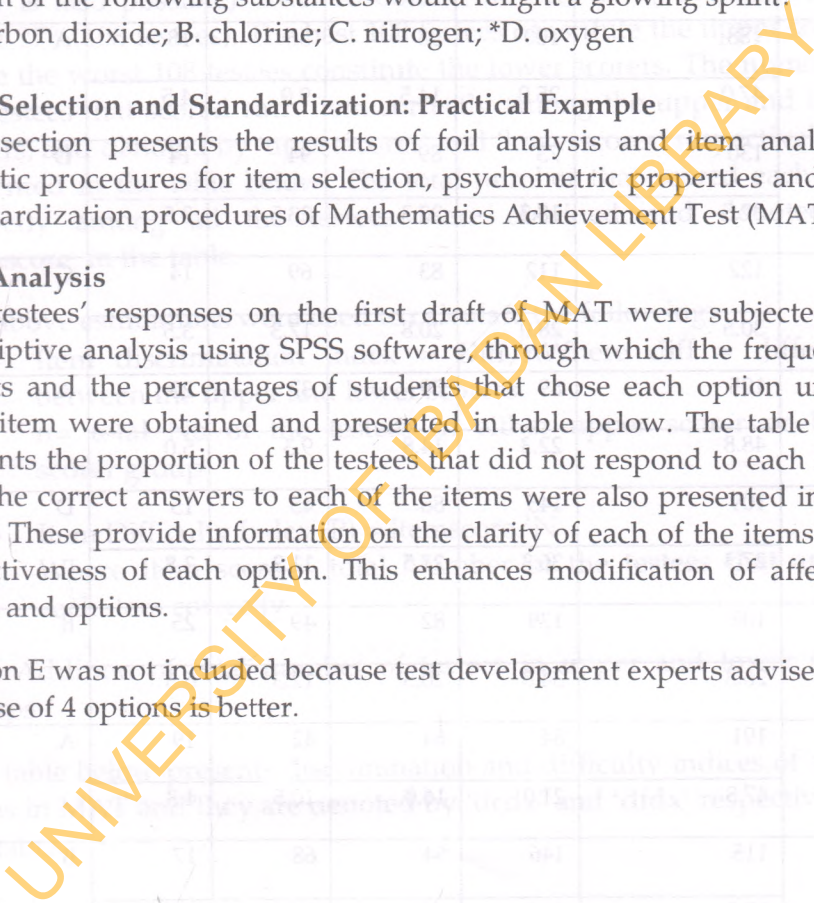
**Item Selection and Standardization: Practical Example**

This section presents the results of foil analysis and item analysis, analytic procedures for item selection, psychometric properties and the standardization procedures of Mathematics Achievement Test (MAT).

**Foil Analysis**

The testees' responses on the first draft of MAT were subjected to descriptive analysis using SPSS software, through which the frequency counts and the percentages of students that chose each option under each item were obtained and presented in table below. The table also presents the proportion of the testees that did not respond to each item and the correct answers to each of the items were also presented in the table. These provide information on the clarity of each of the items and attractiveness of each option. This enhances modification of affected items and options.

Option E was not included because test development experts advise that the use of 4 options is better.



Item	Option A	Option B	Option C	Option D	Option E	Correct Answer
1	10	15	20	25	30	A
2	12	18	22	28	33	B
3	14	20	24	30	36	C
4	16	22	26	32	38	D
5	18	24	28	34	40	E

Note: NR means No Response while A/E means Answer

## Foil Analysis

S/N	A	B	C	D	NR	ANS.
1	156	59	121	44	20	C
	39.0	14.8	30.3	11.0	5.0	
2	188	100	58	36	18	A
	47.0	25.0	14.5	9.0	4.5	
3	130	73	89	94	14	D
	32.5	18.3	22.3	23.5	3.5	
4	122	112	83	69	14	C
	30.5	28.0	20.8	17.3	3.5	
5	195	89	59	37	20	A
	48.8	22.3	14.8	9.3	5.0	
6	109	145	86	45	15	D
	27.3	36.3	21.5	11.3	3.8	
7	105	139	82	49	25	B
	26.3	34.8	20.5	12.3	6.3	
8	191	84	64	42	19	A
	47.8	21.0	16.0	10.5	4.8	
9	115	146	54	68	17	B
	28.8	36.5	13.5	17.0	4.3	
10	129	89	71	90	21	C
	32.3	22.3	17.8	22.5	5.3	

**Note:** NR means No Response while ANS means Answer.

**Item Analysis**

SPSS software was used to mark the responses of the testees obtained from the first administration of the drafted version of MAT. The overall scores obtained by the testees were obtained by using computers to add up scores item by item. The data generated from the responses of 400 testees was sorted by their overall scores in descending (ascending order is also possible) order to obtain the best 27%(108 testees) and lower 27%(108 testees). The best 108 testees constitute the upper scorers while the worst 108 testees constitute the lower scorers. The number of the testees that scored each item correctly among the upper and lower scorers, and denoted by 'upperscorer' and 'lowerscorer' respectively are presented in the table below. The total number that scored each item correctly among all the testees were obtained and denoted by 'itemscore' in the table.

The above estimations were used to calculate the following:

- Item discrimination index =  $\text{Diff}/n$  Where diff = difference between the upper and lower scorers  
 $n$  = total No of the testees in either upper scorer or lower scorer group
- Item Difficulty Index (P) =  $\text{Itemscore}/N$   
 Where Item score = total Number of the testees that answer each item correctly

$N$  = Addition of total number of testees in upper and lower scorer groups.

The table below presents discrimination and difficulty indices of all the items in MAT and they are denoted by 'dcdx' and 'dfdx' respectively in the table.

Item	Upperscorer	Lowerscorer	Itemscore	dcdx	dfdx
1	10	12	22	0.10	0.05
2	15	18	33	0.15	0.08
3	20	25	45	0.20	0.12
4	25	30	55	0.25	0.15
5	30	35	65	0.30	0.18
6	35	40	75	0.35	0.20
7	40	45	85	0.40	0.22
8	45	50	95	0.45	0.25
9	50	55	105	0.50	0.28
10	55	60	115	0.55	0.30
11	60	65	125	0.60	0.32
12	65	70	135	0.65	0.35
13	70	75	145	0.70	0.38
14	75	80	155	0.75	0.40
15	80	85	165	0.80	0.42
16	85	90	175	0.85	0.45
17	90	95	185	0.90	0.48
18	95	100	195	0.95	0.50
19	100	105	205	1.00	0.52
20	105	110	215	1.00	0.55
21	110	115	225	1.00	0.58
22	115	120	235	1.00	0.60
23	120	125	245	1.00	0.62
24	125	130	255	1.00	0.65
25	130	135	265	1.00	0.68
26	135	140	275	1.00	0.70
27	140	145	285	1.00	0.72
28	145	150	295	1.00	0.75
29	150	155	305	1.00	0.78
30	155	160	315	1.00	0.80
31	160	165	325	1.00	0.82
32	165	170	335	1.00	0.85
33	170	175	345	1.00	0.88
34	175	180	355	1.00	0.90
35	180	185	365	1.00	0.92
36	185	190	375	1.00	0.95
37	190	195	385	1.00	0.98
38	195	200	395	1.00	1.00
39	200	205	405	1.00	1.00
40	205	210	415	1.00	1.00
41	210	215	425	1.00	1.00
42	215	220	435	1.00	1.00
43	220	225	445	1.00	1.00
44	225	230	455	1.00	1.00
45	230	235	465	1.00	1.00
46	235	240	475	1.00	1.00
47	240	245	485	1.00	1.00
48	245	250	495	1.00	1.00
49	250	255	505	1.00	1.00
50	255	260	515	1.00	1.00
51	260	265	525	1.00	1.00
52	265	270	535	1.00	1.00
53	270	275	545	1.00	1.00
54	275	280	555	1.00	1.00
55	280	285	565	1.00	1.00
56	285	290	575	1.00	1.00
57	290	295	585	1.00	1.00
58	295	300	595	1.00	1.00
59	300	305	605	1.00	1.00
60	305	310	615	1.00	1.00
61	310	315	625	1.00	1.00
62	315	320	635	1.00	1.00
63	320	325	645	1.00	1.00
64	325	330	655	1.00	1.00
65	330	335	665	1.00	1.00
66	335	340	675	1.00	1.00
67	340	345	685	1.00	1.00
68	345	350	695	1.00	1.00
69	350	355	705	1.00	1.00
70	355	360	715	1.00	1.00
71	360	365	725	1.00	1.00
72	365	370	735	1.00	1.00
73	370	375	745	1.00	1.00
74	375	380	755	1.00	1.00
75	380	385	765	1.00	1.00
76	385	390	775	1.00	1.00
77	390	395	785	1.00	1.00
78	395	400	795	1.00	1.00
79	400	405	805	1.00	1.00
80	405	410	815	1.00	1.00
81	410	415	825	1.00	1.00
82	415	420	835	1.00	1.00
83	420	425	845	1.00	1.00
84	425	430	855	1.00	1.00
85	430	435	865	1.00	1.00
86	435	440	875	1.00	1.00
87	440	445	885	1.00	1.00
88	445	450	895	1.00	1.00
89	450	455	905	1.00	1.00
90	455	460	915	1.00	1.00
91	460	465	925	1.00	1.00
92	465	470	935	1.00	1.00
93	470	475	945	1.00	1.00
94	475	480	955	1.00	1.00
95	480	485	965	1.00	1.00
96	485	490	975	1.00	1.00
97	490	495	985	1.00	1.00
98	495	500	995	1.00	1.00
99	500	505	1005	1.00	1.00
100	505	510	1015	1.00	1.00



### Item Selection

In order to select good items that will constitute the final version on MAT, the following criteria were used:

- Difficulty indices ranged between 0.4 and 0.6
- Discrimination indices ranged between 0.3 and above
- Results of foil analysis were also used for the modification of items and options noticed to be ambiguous.

Based on the above stated criteria, the following items were discarded.

2	3	4	5	6	8	9	10	13	14
15	16	17	19	20					

The items that satisfied the criteria and are retained to constitute the final version of MAT are:

1	7	11	12	18
---	---	----	----	----

**Observation:** Only five of twenty items were retained. Item developers are advised to develop five times the number of items they finally needed, if the final items are going to be good. For example a test developer who intends to have 100 good items at the end of item analysis should develop 500.

### Validity of MAT

To ensure that the instrument (MAT) measures what it is purported to measure, content validity was established using test blue print. The nature of the test (objective achievement test) informed the choice of test blue print. The test blue print ensures that:

- All the topics taught are covered
- Expected cognition levels are equally covered out of Bloom's hypothesized six levels of cognition (knowledge, comprehension, application, analysis, synthesis and evaluation).

Hence, the Sample of test blue print for the final version of MAT is presented below.



Topic	Level of Cognition						Total
	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation	
Number and Numeration	Q1						1
Basic Operation		Q7					1
Measurement							
Algebraic Process		Q11					1
Geometry and Mensuration	Q12						1
Everyday Statistics			Q18				1
Total	2	2	1				5

The limited number of items does not allow for adequate representation of topics in a test. It is suggested that the number of items in a test should be large enough to ensure content validity.

#### Reliability Coefficient of MAT

The reliability coefficient of MAT was established using Kuder Richardson 20 (KR 20) formula.

∴ Reliability coefficient

$$r_{11} = \left( \frac{n}{n-1} \right) \left( \frac{SD_t^2 - \sum p_i q_i}{SD_t^2} \right)$$

Where  $SD_t^2$  = variance of the testees' scores

$P_i$  = proportion of the testees' that answered each item correctly

$q_i$  = proportion of the testees that answered each item wrongly.

$n$  = the sum of testees in Upper and Lower groups.

$$R = \frac{21683 - 4.01}{215 \quad 83}$$

$$\therefore R = 0.9561$$

$$R^2 = 0.9141$$

The estimated reliability coefficient on MAT was 0.9561. This shows that the instrument is highly reliable, and possesses high internal consistency for measuring cognitive achievement of Primary and Junior Secondary One Students' Mathematics Achievement. R square ( $R^2$ ) estimated was 0.9141 which implies that 91.41% variation in JSS 1 cognitive achievement in Mathematics was measured by MAT while the remaining 8.59 percent is traceable to other factors that can cause variation in the cognitive achievement of JS 1 Students.

#### Note:

Test development experts advise that only data generated through testees in upper and lower scorer groups should be used for item analysis.

#### References

- Adeleke J.O. (2010) *The Basics of Research and Evaluation Tools*. Published by SomerestVentures (213pgs).
- Albanese, M. A. (1993). Type K and other complex multiple-choice items: An analysis of research and item properties. *Educational Measurement: Issues and Practice*, 12, 28-33.
- Bloom, B. S., Hastings, J. T. & Madaus G. F. (1971) *Handbook on Formative and Summative Evaluation of Student Learning*. McGraw-Hill Inc.
- Cassels, J. R. T., & Johnstone, A. H. (1984). The effect of language on student performance on multiple choice tests in chemistry. *Journal of Chemical Education*, 61(7), 613-615.
- Chase, C. I. (1964). Relative length of option and response set in multiple choice items. *Educational and Psychological Measurement*, 24(4), 861-866.

- Cohen, J. (1977). *Statistical power analysis for the behavioural sciences*. NY: Academic Press.
- Ebel, R.L. and Frisbie, D.A. (5th ed). (1991) *Essentials of Educational Measurement*. Englewood Cliffs, NJ: Prentice Hall.
- Haladyna, T. M. & Downing, S. M. (1989). Validity of a taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, 2(1), 51-78.
- Haladyna, T. M. (1999). *Developing and validating multiple-choice test items*. Mahwah: Lawrence Erlbaum.
- Kerlinger F.N. & Lee H.B. (2000) *Foundations of Behavioural Research*. 4<sup>th</sup> Edition, Wadsworth Publisher.
- Osterlind, S. J. (1998). *Constructing test items: Multiple-choice, constructed-response, performance, and other formats* (2nd ed.). Boston: Kluwer.
- Thondike, R.M (1997) *Measurement and Evaluation in Psychology and Education*, 6<sup>th</sup> Edition, Prentice-Hall, Inc.

UNIVERSITY OF IBADAN LIBRARY