

**Malaysian Journal of Social Sciences and Humanities (MJSSH)**

Volume 4, Issue 6, October 2019

e-ISSN : 2504-8562

Journal home page:
www.msosocialsciences.com**Book Review on Language Testing, Written by Tim McNamara****Sisay Ayalew Tsegaw¹**¹Department of English Language and Literature, Bahir Dar Institute of Technology, Bahir Dar University

Correspondence: Sisay Ayalew Tsegaw (sisaya@bdu.edu.et / lesisay12@gmail.com)

Abstract

The book in Language Testing was written by Tim McNamara, which is published in 2000 by Oxford University Press. The book provides access to specialist knowledge and stimulates an awareness of its significance. This series as a whole has been designed to provide this access and promote awareness in respect of different areas of language study or testing, which is concerned on its scope and principles of enquiry, basic concerns and key concepts. This book offers a succinct (short and precise) theoretical introduction to the basic concepts in language testing in a way that is easy to understand in some extent. In educational context, this book is highly recommended for university faculty members involved in obtaining assessment literacy, teachers who want to validate their exploratory teaching and testing, applied linguistics students new to the language testing field or for those people who pick interest in language without being academically engaged. Therefore, the book might accommodate those interests as meant to be introductions to language more generally.

Keywords: language testing, test types, test purpose, test construct, test cycle, test content**Introduction**

The Tim McNamara book is divided into eight main sections or chapters. The first one provides an overview of the principles of language testing. The survey in Section 1 is thorough, clear, and easy to follow - though more theoretical than practical. It focuses on language testing principles rather than practical advice on how to design and implement good tests. This background knowledge, nevertheless, can help teachers critically evaluate their testing practices.

Testing

Under section 1, there are five sub sections that explore the concept of testing. The first one is exploration about testing. Testing is a universal feature of social life. It is a matter of using data to establish evidence of learning and Tests to see how a person performs particularly in relation to a threshold of performance. And also Tim McNamara says that what is true of testing is general is true also of language testing, that may conjure up an image of an examination room, a test paper with question, desperate scribbling against the clock. The second sub section elaborates about the understanding of language testing, which is relevant to involving in creating language tests and using information or tests, which is provided in practical and research contexts.

It also involves recognizing the distinction between the criterion (relevant communicative behavior in the target situation or a series of performances subsequent to the test) and the test that is a performance or series of performances stimulating / representing from the criterion. There are also reasons for developing a critical understanding of the principles and practice of language assessment: First, language tests play a powerful role in many people's lives. Second, relying on information from tests to make decisions on the placement of students on particular courses and finally to conduct research concerning language study, we have to measure of the language proficiency. The third sub section is about types of tests which is elucidated that all language tests differ from how they are designed (test method) and what they are for (test purpose). In terms of method, paper –and-pencil language tests (used for the assessment micro and the macro receptive skills) whereas, performance tests assessed is an act of communication language skills (productive skills).

Test purpose, design and Cycle

The fourth is concerning to test purpose. In this section, there is distinction between achievement and proficiency tests. The first one is associated with the process of instruction to reflect the progressive aspects of the goals of teaching and learning by relating the past in that they measure what language the students have learned as a result of teaching. But the later look the future situation of real life language use. In addition to this, the opponent idea on direct testing has several reasons, for instance: The materials and tasks in language teaching can never be real but relatively real (limits of authentic); the period in which behavior observed is quite extended; and the testing equivalent of observer' paradox. The final sub-section is about the test –criterion relationship.

The book in section two is through communication and the design of language tests in the first sub section of this main section, Test construct is taken in to the aspects of knowledge or skill possessed by the candidate which are being measured, which involves what knowledge of language consists of and how that knowledge is deployed in actual performance or language use. In second sub-section, discrete point test is the practice of testing aspects of knowledge in isolation or practice of testing with separate or minimal context, individual points of knowledge is what it stood for. In third, integrative tests are the development of tests which integrated knowledge of relevant systematic features of language (pronunciation, grammar, vocabulary.) in integrative manner though the problem to such integrative tests tended to be expensive, time consuming, and difficult to score but Pragmatic tests are the ability to integrate grammatical, lexical, contextual and pragmatic knowledge in test performance that the author of the book gave an example of close tests for all skills .

Communicative language testing /CLT/ is the fourth subsection that has a scope of what was covered by an understanding of language and the ability to use it in the language context, and it is also to mean that knowing the language. And also, Performance tests and paid attention to the social roles are the main features of it. In the last sub section, models of communicative ability is the challenge of CLT was matched by a continuing theoretical engagement with the idea of communicative competence and its implications for the performance requirement of CLT. This model also specified the components of knowledge language without dealing in detail with their role in performance. For instance, grammatical or formal, strategic, socio linguistic and discourse competences are the main parts of it.

The extracts in Section 3 introduce topics such as the testing cycle: understanding the constraints, test content, method and specification, authenticity of response, fixed and constructed response formats and test trials. This section provides that constraints for testing are under which the test will be administered, what resources, physical and financial, are available for test development and test operation, test security...etc should be take in to account or establish. The test content is concerned on the decision on what will go in to the test, and it involves careful sampling from the domain of the test. Depending on the construct, the test domain typically defined as a set of practical, real world tasks in operational way, and also in terms of a more abstract construct (for example in terms of theory of components of knowledge and ability) but the test method is related with the way in which candidates will be required to interact with the test materials, particularly the response format, aspects of design together with the issues of how candidate responses will be rated or scored and then sees method as an

aspect of content, and raise issues of authenticity and treats method independently of contents are the two major approaches in order to understand the relation of test method to test content.

During the book discussion it included that different response formats are sometimes conventionally associated with different types of test content. For instance, fixed response format is ones in which the candidates possible responses have been anticipated and their task is to chose the appropriate response from those offered. And also constructed response formats may also be used, although these are more complex and usually more expensive to score; It has the advantage of not constraining the candidate to the same degree, and reducing the effect of guessing, that enables the candidates to feel responsible, more demanding and authentic even though expensive to score is their backwash effect of this format.

According to Tim points of view, test specifications are the result of the design process in terms of test content and test method, and they are recipe or blueprint for test construction that include information on such matters as the length and structure of each part of the test, the types and sources of the material where as test trail involves careful design of data collection to see how well the test is working.

In section four, it has discussed about rating procedures used in language assessment. Establishing a rating procedure and scales, problem and frameworks of raters, is introduced in this section. Three main aspects or set methods for yielding the judgment in question. First, agreement about the condition; Second, certain features of the performance are agreed to be critical (considering various components of competence) and finally raters understanding the characteristics' of a performance by allocating grade or rating. The problem with the raters will be subjectivity and significance degree of chance. In order to avoid this problem, establishing a frame work for making judgments such as considering the criteria by which performances at a given level will be recognized, and decide how many different levels of performance we wish to distinguish are the major points what we should be take in to account. The writer also extensively elaborated about rating scales.

He stated that an order series of such description is called rating scales. The preparation of such scales involves developing level descriptors. And also They are used to guide and constrain the behavior of rates and used to report the outcome of a rating process to score users, and allow the greatest flexibility to the users, who may want to use the multiple distinctions available from a scale or who may choose to focus on only one cut-point or region of the scale. According to Tim discussion, ratings are holistic and analytic. Holistic rating is getting raters to record a single impression of the impact of performance as a whole but analytic rating is getting raters to provide separate assessments for each of a number of aspects of performance. Finally he illustrated that an important way to improve the quality of rater mediated assessment scheme is to provide initial and ongoing training to raters that usually takes the form of a moderation meeting.

The extracts in Section five introduce topics such as testing the test: threats to test validity, test content, method, and construct and test impacts, and then have questions which encourage readers to consider matters in more depth. This is particularly useful for those discussing their ideas with their colleagues and / or supervisors that are the follow up questions and more practical examples are designed to help readers consider possible solutions. In this section, test validation is elaborated as a purpose of investigating the defensibility of the inferences about candidates that have been made on the basis of test performance.

It involves understanding how, in principle; performance on the test can be used to infer performance in the criterion and empirical data from test performance to investigate the defensibility of that understanding, close analysis of data, and thinking about the logic of the test. And also it elaborated in some extent about face validity, content and consequential validity. Threats to test validity is deals about the necessary procedures of test validity, the meaningfulness, interpretability, fairness of assessment (sores, rating), and about test content, method, construct and etc. Test content, method and construct, are introduced in unit five are taken mainly threats to test validity.

The key discussion was that test content makes satisfactory basis for the inferences to be made from test performance and the emphasis issue is content related validity. Also one way of approaching this

issue is to ask to what extent the method is properly part of the test construct (the underlying ability or trait being measured by the test). In general, the more general the context of performance, the more there is to jeopardize the validity of the ratings though the preparation of test and the fairness of inferences about the candidates have their own impact on tests.

Assessment and Procedures

Section 6 is discussed about measurement, quality controlling for raters, grading systems. This section gave brief elucidation to a small selection of measurement concepts and procedures commonly used in language assessment, and in particular to make the reader for that they are accessible and worthy understanding and also about grading system. Measurement investigates the quality of the process of assessment by looking at scores and also involved two main steps. These are: quantification, which is the assigning of numbers or scores to various outcomes of assessment (data matrix) and checking for various kinds of mathematical and statistical patterning within the matrix. According to Hughes (2003) what we need are techniques that will extract behavior which is reliable and valid pointer of the ability in which we are interested and which can be reliably scored.

Those procedures help us to achieve quality control (to improve the meaningfulness and fairness of the conclusions reached about individual candidates) and for predication of one set of measures from one class to the other. The main concern of this section is relating with the quality control of the raters. It means that understanding of language tests and ability to develop fair, careful data collection, understand scales for the degree of rater agreement, meaningful tests and statistical methods for predictive relationship such as correlation coefficient, reliability coefficient that express inter-rater reliability, and also some way of classification analysis that should be recognized in a qualified and effective test raters. According to Tim discussion, Quality control procedures for others such as paper and pencil is most of the time for objectively scored test items we carry out item analysis. The perception of reliability is defined as 'the consistency of measurement' (Bachman and Palmer 1996). In other words, a test is reliable to the degree that whatever it measures, it measures it every time.

This analysis tells us how well each item is working (making to the overall picture of candidates' ability emerging from the test). It is also possess through the stages of pilot version, trailing or try out, and operational version. Item analysis provides two kinds of information. The first one is, Item facility, which helps us decide if test items are at the right level for the target group, and also it express the proportion of the people taking the test who got a given item right. It is also expressed from 0 to 1. And the second one is item discrimination, which allows us to see if individual items are providing information on candidates' abilities consistent with that provided by the other items on the test.

When we overlook about the grade system, norm referenced measurement is the way of using as one's standard some kind of norm, which an expectation is derived from a typical range of performance and it is a system by which students are placed in rank order, where grades are assigned by comparison to other students' performance rather than upon the absolute quality of their performance. It is very easy for teachers to use and prevent grade inflation by limiting the number of high grades awarded; and it identifies those students whose work truly stands out as exceptional, which is useful when you write letters of recommendation are the main advantages of this grade system.

Whereas, Criterion- referencing relates the pass or fail of a candidate to some explicitly stated criteria and Individual performances are evaluated against a verbal description of a satisfactory performance at a given level. The last but not least and that actuated to me when I read this book is that about the new approaches to measurement. The new approach is known by the general name of IRT. It greatly facilitates formerly very difficult business of test equating (producing tests of equivalent difficulty). IRT also permits test linking. And IRT also makes possible ways for development of computer adaptive tests.

The extracts in section seven are introduced sub topics on the main topic of the social character of language tests. The perspectives on assessment which focus on the institutional and social meaning of

assessment including their responsibility in addition to about the social and educational policy are the main detail ideas during the book discussion. When we see the institutional character of an assessment, it concerned on when assessment is made, it is not done by someone acting in a private capacity, motivated by personal curiosity about the other individual, but in an institutional role, and serving institutional purposes. Regarding to about the social responsibility of the language tester, those who advocate the position of socially responsible language testing reject the view that language testing is merely a scientific and technical activity.

Under this issue, ethical and critical language testing are discussed. Ethical language testing is language testing practice can be made ethical, and stresses the individual responsibility of testers to ensure that they are. Generally, ethical testing practice as involving test developers in taking responsibility for the effects of tests. The three area main concerns are: accountability, wash back, and impact. But critical language testing sees tests as essentially sociopolitical constructs, which since they are designed an instruments of power and control, and It is best understood as an intellectual project to expose the role of tests in this exercise of power. Finally, regarding to the codes of professional ethics for language testers, Tim recommended that professional bodies of language testers should formulate codes of practice which will guide language testers in their work; the emphasis is on good professional practice: that is, language testers should in general take responsibility for the development of quality language tests.

The last section or chapter of the book is concerned in this eon educational reforms of new directions' –and dilemmas concerning to language testing, which is described as it is a field in crisis, one which is masked by the impressive appearance of technological advance. Especially, Computers based testing is the ability of computers to carry out various kinds of automatic processes on spoken or written texts is having an impact on testing. The proponents of computer based testing can point to a number of advantages; first, scoring of fixed response items can done automatically, and the candidate can be given a score immediately; second, the computer can deliver tests that are tailored the particular abilities of the candidate, and finally the existence of large item banks. Apart from this, the speed of technological advances affecting language testing; Language testing as the fields of an assessment is crucially dependent on definitions of the test construct; and Language testing remains a complex and perplexing activity are the main dilemmas that mentioned relating to the performance of the learners or the candidates.

Conclusion

I found this book to be very well thought out and well written. It is easy to use this book as a reference book. The nature of this survey, with short chapters and subheadings, makes it easy to find specific topics readers are interested in. Generally the cross-referencing is very helpful. Throughout the text readers are referred back and forth to places where the material is discussed in more detail. For instance, in all sections I have understood or examined that a language test is a procedure for gathering evidence of general and specific language abilities from performance on tasks designed to provide a basis for predications; a number of influence school of language testing; the process through which testing procedure is conceptualized, developed and put in operation.

We have considered test content as an expression of test construct and looked at how that content may be examined; rating procedures used in language assessment; the need for questioning the basis for inferences about candidates abilities residing in test procedures, and the way in which these inferences may be at risk from aspects of test design and test method, or lack of clarity in our thinking about what we are measuring; a number of ways in which concepts and practices from the field of educational measurement or psychometrics have had an impact on the area of language assessment. Test developers, should go one more steps to determine the characteristics of the total test. In other words, simply putting good individual items together would not be sufficient for a test to function satisfactorily. (Farhady and Jafarpur and Birandi, 2010).

We distinguish different approaches to measurement, with different sets of assumptions and some of the most techniques' associated with each for investigating the quality of language testing; and the institutional character of tests and the implications of this for understanding the nature of language testing as a social practice, and the responsibility of language testers. To my mind, all sections of the book do encourage readers to both seek further information and relate this theoretical knowledge to their testing practice. The main purpose of this book is to enable readers to understand testing theory more deeply and relate this knowledge to their testing practice. In Ethiopia this book could be a much needed catalyst for change to change the change for a better change about language testing in classroom or national levels in various ways according to the context.

There are only four references for the entire topic of validation, arguably the most important aspect of the entire testing process. Most of the references appear to be academic in nature, which is appropriate for applied linguistics students, but practicing teachers may appreciate the inclusion of more practical texts. It is easy to use this book as a reference book. The nature of this survey, with short chapters and sub-headings, makes it easy to find specific topics readers are interested in. Generally the cross-referencing is very helpful.

In contrast, the list of references in Section 3, though mentioning many key language testing texts, is sometimes disappointingly brief. For example, there are only two references for the problematic area of test rating. Throughout the text readers are referred back and forth to places where the material is discussed in more detail. The glossary in Section 4 also lists page numbers where terms can be found in the main survey, but it is disappointing that only one occurrence, and not a list of all occurrences of the term are included. For example, critical language testing is referenced to page 72, but the main discussion of this movement is found on pages 76 and 77. It would therefore have been helpful to include these pages in the glossary. In spite of the few criticisms made here, this book remains an excellent overview of the main principles of language testing. Although not as thoroughly developed as possible, Sections 2 and 3 do encourage readers to both seek further information and relate this theoretical knowledge to their testing practice.

References

- Alderson J. Clapham, C. and Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge. Cambridge University Press.
- Bachman, L. and Palmer, A. (1996). *Language testing in practice: Designing and Developing Useful Language*. Oxford: Oxford University Press.
- Bachman, L.F., and Palmer A.S (1996). *Language testing in practice*. Oxford. Oxford University Press.
- Banerjee, S. (2000). Integrated Testing. In Byram, M. (2000) Routledge Encyclopaedia of *Language Teaching and Learning*. London. Routledge.
- Farhady, H., Jafarpur, A., and Birjandi, P. (2010). *Testing language skills from theory to Practice*. Iran: SAMT
- Hughes, A. (2003). *Testing for language teachers*. Cambridge university press.
- Tim, M. (2000). *Language Testing*. Oxford: Oxford University