UNIVERSITI TEKNOLOGI MARA

# DETERMINING THE PERFORMANCE OF FIVE MULTIPLE CHOICE SCORING METHODS IN ESTIMATING EXAMINEE'S ABILITY

**PREPARED BY**

**LAU SIE HOE**
**DR. PAUL LAU NGEE KIONG**
**LING SIEW ENG**
**HWA TEE YONG**

**DECEMBER 2006**

PERPUSTAKAAN UiTM SARAWAK

# TABLE OF CONTENTS

CHAPTER ONE

INTRODUCTION

CHAPTER TWO

LITERATURE REVIEW

# ABSTRACT

Despite the current popularity of performance-based assessment and the emergence of new assessment methods, multiple choices (MC) item remain a major form of assessment. Conventional Number Right (NR) scoring method where one point for correct response and zero for other response has been consistently criticized for failure to credit partial knowledge and encourage guessing. Various alternative-scoring methods such as Number Right with Correction for Guessing (NRC), Elimination Testing (ET), Confidence Weighting (CW) and Probability Measurement (PM) had been proposed to overcome these two weaknesses. However to date, none has been widely accepted although the theoretical rationale behind various scoring methods under Classical Test Theory (CTT) is sound. A major cause of concern is the possibility that complicated scoring instruction might introduce other factors, which may affect the reliability and validity of the test scores. Studies on whether examinees can be trained to follow the new test instructions realistically have been inconclusive. Whether they can consistently follow the test instruction throughout the whole test remain an open question. There have been intense comparisons studies on scores obtain through various CTT scoring methods with NR scores. What yet to be explore is the comparison of these scores with Item Response Theory (IRT) ability estimates. This study attempt to close the three knowledge gaps identified above.

Firstly, it attempts to determine the extent to which the examinees can be trained to follow a new MC test instruction realistically. Under the new test instruction, an examinee must first eliminate the option(s) which is/are sure incorrect, and based on the remaining option(s), choose one as the answer. It also determines whether there

# CHAPTER ONE

# INTRODUCTION

## 1.0 Introduction

From the day children learn how to read and write tests play an important role in their lives. Brown (2005) is of the notion that tests are political by virtue that they decide to a great extent people's lives, in the sense of their future choice and directions. The purpose of testing is to assign a score to an examinee that reflects the examinee's ability as measured by the test (Linn, 1990).

Multiple-choice (MC) tests format are the most common, and perhaps the best tool for objective measurement of knowledge, ability, or achievement (Chevaliaer, 1998). This format is favored by both testing organizations and classroom teachers because it provides broad content sampling, high score reliability, ease of administration and scoring, usefulness in testing varied content, and objective scoring (Kurz, 1999). Under the conventional number right (NR) scoring method, all items are weighted equally and the examinees are required to pick one alternative as the answer. An examinee is awarded one point for the correct responses and zero for incorrect responses. The test score is the sum of item scores.

However, this method, while simple to use, has been constantly criticized due to several weaknesses. These weaknesses include decrease in validity due to guessing and failure to credit partial knowledge (Kurz, 1999). According to Bar-Hillel M. Budescu and Attali, (2003), in NR scoring method tests one cannot distinguish lucky guesses from answers

# CHAPTER TWO

# LITERATURE REVIEW

## 2.0 Introduction

This literature review chapter is divided into eight main sections. The first section briefly describes how the literature review was conducted. The second section gives a brief comparison between two popular frameworks in addressing assessment issues; Classical Test Theory (CTT) and Modern Test Theory. The third section gives a review on the historical background of multiple-choice testing and the development of various CTT scoring methods. The fourth section touches on the historical background of IRT models, its characteristics, the assumptions and the features of IRT. It concludes with a theoretical framework of IRT models. The fifth section focuses on the two most commonly used IRT-based scoring methods. This is followed by a review of the different methods in IRT ability estimation and the factors influencing it. The seventh section summarizes the comparison studies of various scoring methods and the issues related to the ability of examinees following test instruction. The concluding section discusses how the literature review impacted on the present study; the formulation of the research questions, design and the methodology.

## 2.1 Conducting the Literature Review

To obtain an overview of the research topic, a few key texts were referred. Two doctorate dissertations by Holmes (2002) and Ndalichako (1977) were referred to give an overview on the comparison studies on scores of various CTT scoring methods with NR scores. In order to obtain an overview on the development of IRT and issues related to IRT ability