OBAFEMI AWOLOWO UNIVERSITY

GENERALIZED PARTIAL CREDIT AND GRADED RESPONSE MODELS IN POLYTOMOUS MATHEMATICS ACHIEVEMENT TEST ITEMS IN ABILITY ESTIMATION IN LAGOS STATE

BY

IWINTOLU RUKAYAT OYEBOLA
(EDP12/13/H/0231)

A THESIS SUBMITTED TO THE DEPARTMENT OF EDUCATIONAL FOUNDATIONS AND COUNSELLING, FACULTY OF EDUCATION, OBAFEMI AWOLOWO UNIVERSITY, ILE-IFE, NIGERIA IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE AWARD OF THE DEGREE OF DOCTOR OF PHILOSOPHY IN EDUCATION (TESTS AND MEASUREMENT)

2016

OBAFEMI AWOLOWO UNIVERSITY, ILE –IFE, NIGERIA

HEZEKIAH OLUWASANMI LIBRARY

POSTGRADUATE THESIS

AUTHORIZATION TO COPY

Author:	Rukayat Oyebola IWINTOLU
Title:	Generalized Partial Credit and Graded Response Models in Polytomous Mathematics Achievement Test Items in Ability Estimation in Lagos State
Department:	Department of Educational Foundations and Counselling.
Degree:	Doctor of Philosophy in Education (Tests and Measurement)
Year:	2016

I, Rukayat Oyebola IWINTOLU hereby authorize the management of Hezekiah Oluwasanmi Library to copy my thesis in whole or in part, in response to requests from individual researchers and organisations for the purpose of private study or research.

…… ……..……

Signature Date

CERTIFICATION

This research project written by Rukayat Oyebola IWINTOLU and has been read, approved and adjudged as meeting requirements for the award of Ph.D degree in Education (Tests and Measurement) of the Obafemi Awolowo University, Ile-Ife, Nigeria

……... ……

Prof. E. R. I. Afolabi Date

Supervisor

…… ……

Dr. (Mrs.) S. A. Ehindero Date

Acting Head of Department

DEDICATION

This study is dedicated to God Almighty and Jesus Christ, for being my stronghold at all times.

ACKNOWLEDGEMENTS

I give glory to the Almighty God, my ever present help in time of need; for His guidance and protection, especially throughout the period of this programme. Indeed, He has made me to laugh, so that all that hear will laugh with me.

I sincerely wish to express my profound gratitude to my supervisor, father and mentor, Prof. E. R. I. Afolabi for his thorough supervision, constructive criticisms and willingness to render help any time the need arose. His consistent encouragement, motivation and relentless efforts at making this work a success are incomparable. I pray that the Lord bless him and continue to increase him on every side in Jesus name. My appreciation goes to the Head of Department of Educational Foundations and Counselling, Dr. (Mrs.) S. A. Ehindero for her encouragement throughout the course of my study.

I express gratitude to Professor K. A. Alao, Professor S. V. Kobiowu, Professor B. I. Popoola, Professor (Mrs.) O. O Ojo, Professor (Mrs.) B. A. Omoteso, Dr. B. A. Faleye, Dr. O. A. Ojebiyi, Dr. A. A. Adediwura, Dr. S. A. Oluwatosin, Dr. (Mrs.) A. A. Shobola, Dr. W. O. Adeniyi, Dr. T. O. Ajeigbe, Dr. (Mrs.) E. O Adeyemo, Dr. (Mrs.) A. A. Anuodo, Dr. O. F. Adebowale, Dr. A. G. Akintomide, Dr. O. O. Olatomide, Mrs. F. T. Fatoki, Mr. M. A. Omoyemiju and Miss M. Quaye for their support and encouragement during the programme. I also appreciate the administrative staff in the Department and the teachers and students who

helped in the administration of the instrument. I appreciate the useful assistance of my friends and course mates (Mrs P. O. Adetutu, Dr. J. A. Odeyemi, Dr. A. A. Adeniji, Mrs. O. E. Ogedengbe, Oyedokun Oyenike, Mr A. A. Adeniyi and Mrs T. O. Adelekun) for their academic and moral support.

My heartfelt appreciation goes to my father, HRH Oba E. O. Oyebamiji, his words of encouragement are immeasurable. He is strategically positioned in my life to be a source of motivation. I thank him so much; this is another star he has added to his cap in his life time. I pray he will eat the fruit of his labour in good health and perfect peace in Jesus name. My mothers, Mrs K. Oyebamiji and Olori R. D. Oyebamiji, I appreciate their love, prayers and care throughout the programme, may the Lord bless and keep them. I am also grateful to Prince and Mrs O. O. Oyebamiji, Prof. and Mrs O. O. Ajayi, Dr. G. O. Binuyo, Mrs. A. Oduola, Mr. B. Ogunkale, Tolani Oyebamiji, Latifat Oyebamiji, my other sisters and brothers for their prayers and moral support given to me in the course of the programme.

I appreciate my darling husband, Dr. Iwintolu Olawale, for his unquantifiable moral, spiritual and financial assistance throughout the programme and for always being my support in moments when I needed someone to say “You Can Do It”. To my princesses; Morikeoluwa and Momooreoluwa I thank them for being a constant source of love, support and strength all through the programme.

To everyone who has contributed to my fulfillment in life and this programme, but whose names were not mentioned here, I appreciate you for being part of my success story.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS AND ACRONYMS

Acronyms	Title
1PLM	One Parameter Logistic Model
2PLM	Two Parameter Logistic Model
3PLM	Three Parameter Logistic Model
BEM	Bayesian Estimation Methods
CMLE	Conditional Maximum Likelihood Estimation
CTT	Classical Test Theory
DIF	Differential Item Functioning
EAP	Expected A Posteriori
GPCM	Generalized Partial Credit Model
GRM	Graded Response Model
ICC	Item Characteristics Curve
IRF	Item response Function
IRT	Item Response Theory
JMLE	Joint Maximum Likelihood Estimation
LID	Local Item Dependence
MAP	Maximum A Posteriori
MCM	Multiple Choice Model
MLE	Maximum Likelihood Estimation
MMLE	Marginal Maximum Likelihood Estimation
NCM	Nominal Categories Model

PCM	Partial Credit Model
RSM	Rating Scale Model
SEE	Standard Error of Estimate
SSCE	Senior School Certificate Examination

ABSTRACT

The study examined the difficulty and discrimination indices of polytomous items using dichotomous scoring. It determined the effect of the systematic increase in the response levels of polytomous items on ability estimates in Generalized Partial Credit Model (GPCM) and Graded Response Model (GRM). It further compared the scores of GPCM and GRM on the accuracy of ability estimates and investigated which of the scoring formats was more effective in estimating ability. These were with a view to providing information on the ability of GPCM and GRM in improving the accuracy of scores in polytomous Mathematics achievement test items.

The study adopted the survey research design. The population for the study consisted of students who registered for the Senior School Certificate Examination (SSCE) in Lagos State in 2015. A sample of 1015 students was selected. Two Education Districts (EDs) were selected using purposive sampling technique from the six EDs in the State based on availability of federal schools for an inclusive representation of schools. Three schools were selected from each of the EDs using stratified random sampling technique with school ownership as stratum for selection. One intact SS III class from each of the schools was selected using purposive sampling technique based on students' ability groups. The instrument used in the study was an adapted version of May/June (2006-2014) SSCE General Mathematics Paper 1 titled Mathematics Achievement Test (MAT). Data collected were analysed using ANOVA and Pearson-r. Furthermore, BILOG and IRTPRO were used to generate item parameters of difficulty, discrimination and guessing.

The results showed the difficulty and discrimination indices of polytomous items using two different formats. Using dichotomous scoring in the first format difficulty and discrimination indices were high (-0.73 to 1.95 and 0.38 to 1.72 respectively) while they were higher in the second format (0.09 to 3.33 and 0.43 to 1.85 respectively). The results also showed that there was a significant effect of the systematic increase in the response levels of the items in GPCM on Mathematics ability estimates ($F= 3.98$, $p<0.05$) in the first format, while there was no significant effect in the second format ($F=3.16$, $p>0.05$). The results further showed that there was a significant effect of the systematic increase in the response levels of the items in GRM on Mathematics ability estimates ($F= 22.46$, $p<0.05$ and $F = 31.55$, $p<0.05$) in the first and second formats respectively. Furthermore, GPCM estimated ability more accurately than GRM ($r = 0.22$, $p<0.05$). Finally, the results indicated that the first format which contained items with one partially correct answer, one full correct answer and two distracters was more effective in estimating ability ($r = 0.22$, $p<0.05$).

The study concluded that Generalized Partial Credit Model had greater ability than Graded Response Model in improving the accuracy of scores in polytomous Mathematics achievement test items.

Supervisor: Prof. E. R. I. Afolabi

Number of Pages: 240

CHAPTER ONE

INTRODUCTION

Background to the Study

The inevitability of tests in fulfilling important needs in the decision-making process permeating all facets of human endeavour seems to have made it captured the attention of the public, and at the same time inspires so much anxiety from classroom to work environment. Even the least interested citizen is aware of the growing use of testing in every facet of human endeavour; it is therefore expected that tests attained prominence and usage as far back as 3000 years ago. Educators, governments and establishments' continuous search for reliable and trusted means through which students and applicants could be held accountable of a self-acclaimed ability resulted in testing. Although, several schools of thought have argued that test or examination is not the best measure of a person's ability, while on the other hand, researchers (e.g. Lumsdan, 1978; Kaplan and Saccuzzo, 2005, Afolabi, 2012) have maintained that it is everywhere evident that there is yet no viable alternative to tests in determining a person's academic ability.

A test may be defined as standard set of items which are specific stimuli to which a person overtly responds and which can be scored. There are different types of tests, with different peculiarities, strengths and weaknesses. The major and broad classification of test types is the essay and objective tests. From a practical point of view, the essay (free-response) test provides good measure of students' ability, in so far as it gives them the opportunity to respond freely to items in their own words, thereby building their reasoning and other skills in higher-thinking learning. The required response can be as simple as the writing of a single word

or as complex as the design of a laboratory experiment to test a scientific hypothesis. It helps to recognise and reward different abilities of students through the assigning of partial credits to incomplete understanding of concepts. Despite the enormous credits accrued to the essay tests however, it has been critiqued on the basis of its difficulty to use; its scoring could be subject to raters' bias and inconsistency. On the other hand, the advocates of the objective tests type among which is the multiple-choice format have adduced reasons for its dominant use which includes; limited amount of testing time, its ability to sample a broad range of content and provide a good sample of test takers' knowledge. The responses can be scored by machine, making the scoring process faster and inexpensive, with no room for differences of opinion. In other words, objective tests allow the evaluation of a greater breadth of content in a fixed testing time under limited financial budgets.

Multiple-choice test (MCT), a variant of the objective test has gained credence from classroom assessment to professional licensure examinations (Scott, 2011). It is apt to note that the use to which MCT will be put, would determine the structure of the items in terms of construction of the stem, the response options (correct answer(s)) and the distracters which are predicated on its scoring. For the vast majority of multiple-choice tests, items are scored dichotomously (i.e. correct or incorrect). According to Osterlind and Everson (2009), items with two categories or values (possibilities to respond) are called dichotomous. In this stance, it involves presenting test question and a list of alternatives ranging from three, to four or five as the case may be. The testees are to make free choice of one correct answer from the alternatives given to the item. If an examinee endorse the correct answer, in dichotomous scoring, he/she will be awarded a score of one, whereas if an incorrect answer a score of zero.

Generally, most users of the multiple-choice tests employ dichotomous scoring which are also analysed using the Classical Test Theory (CTT). It is the approach that most users have been exposed to throughout their education. In CTT the respondent's observed score on a whole instrument is the unit of focus. Therefore, a person's score is the unweighted sum of his/her responses to an instrument's items. However, educational reform efforts have led to an increased search of alternatives to the traditional dichotomously scored multiple-choice items, as there has long been a need to assess objectives that require more than a single response (Albanese, 1993). Researchers in a bid to gainfully combine the two item types (i.e essay and objective test), explored the blend of the duo item responses and established that they produced a total score that is more reliable than scores separated by item type. Despite the alternative mixed-item format, they are still subject to both item types' strength and weaknesses. Nonetheless, polytomous tests may perhaps have been observed to provide more equitable approaches to testing than the dichotomously scored multiple-choice tests. They have also been adjudged to provide more information regarding the precision of trait-level estimation than dichotomous items (Jodoin, 2003; Penfield & Bergeron, 2005). The polytomous item type could therefore be perceived as the mediator between the dichotomously scored multiple-choice item and the essay test type as it is contingent on the combined strengths of both item types in a bid to shrink their weaknesses.

In polytomous models, items in the test are not just scored right or wrong; instead, each of the categories of responses is evaluated and scored according to its degree of correctness or the amount of information provided toward the full answer. By implication, the items are constructed in such a way that it allows every examinee's efforts at items to be rewarded. Weights are assigned to options in ascending order as examinees knowledge on the item increases, in other words, polytomous items are expected to increase test validity. In a simulation

study, Lau & Wang (1998) observed that classification accuracy in computerized testing situation resulted in higher precision for polytomous items than dichotomous items. They found that lower false negative and false positive classification