

R

PERPUSTAKAAN UNIVERSITI MALAYA

ACH - 8/46  
INVC. hms. 6/2/01

**IMPLEMENTING A DIAGNOSTIC ASSESSMENT  
MODEL USING CRITERION-REFERENCED  
TESTS IN A PRIVATE COLLEGE**

**Lim Khoon Siang**

**A Dissertation Submitted to the Faculty of Education,  
University of Malaya in Partial Fulfilment  
of the Requirements for the Degree of  
Master of Education**

**1999**

Perpustakaan Universiti Malaya



A510142964

OK

## ABSTRACT

This study examines the use of criterion-referenced tests as a tool for formative assessment within a framework of a diagnostic assessment model. A total of 33 students of a pre-university physics course participated in an objectives-based diagnostic assessment programme. The study intends to assess the effectiveness of the diagnostic assessment model in improving learning outcome. According to this model, the teaching-learning processes are laid out as follows: (a) the learning objectives are specified; (b) the subject matter is organised into smaller learning units; (c) formative tests in the form of criterion-referenced tests are administered; and (d) corrective feedback is provided after student learning weaknesses are diagnosed using an item analysis. The perceptions of students toward the use of diagnostic tests as a motivating or contributing factor in their learning is also discussed. An empirical review of the effectiveness of diagnostic test items in classifying students into mastery and non-mastery states is also discussed. Three different item discrimination indices i.e., Brennan index  $B$ , point-biserial correlation coefficient  $r_{pb}$  and phi coefficient  $\phi$  that are used in mastery and non-mastery classification are also compared. The findings of this study indicate that the close relationship between teaching, diagnostic testing and remediation applied in the diagnostic assessment model results in a significant improvement in test scores for the group as well as a favourable perception of diagnostic assessment. A strong mean positive correlation of .93 between  $B$  and  $\phi$  for the diagnostic tests suggest that both these indices varies consistently irrespectively of the difficulty levels of the tests.

**PELAKSANAAN SATU MODEL PENILAIAN DIAGNOSTIK  
YANG MENGGUNAKAN UJIAN-UJIAN RUJUKAN  
KRITERIA DALAM SEBUAH KOLEJ SWASTA**

**ABSTRAK**

Kajian ini menggunakan 'ujian rujukan kriteria' yang terkandung dalam suatu kerangka model penilaian diagnostik. Sejumlah 33 orang pelajar kursus fizik pada peringkat pra-universiti telah dipilih sebagai sampel kajian program penilaian diagnostic yang berasaskan objektif-objektif pembelajaran. Kajian ini hendak menguji keberkesanan model penilaian diagnostik untuk meningkatkan prestasi pembelajaran di kalangan pelajar-pelajar. Model ini menyusun proses pengajaran dan pembelajaran seperti berikut: (a) pernyataan objektif disediakan; (b) isi kandungan matapelajaran disusun dalam bentuk unit-unit pembelajaran; (c) ujian-ujian formatif berbentuk ujian rujukan kriteria diberikan; dan (d) maklum-balas diberikan selepas analisa item dilakukan. Persepsi pelajar-pelajar terhadap penggunaan ujian diagnostik sebagai pendorong pembelajaran dibincangkan. Ulasan empirikal tentang keberkesanan ujian diagnostik untuk menyaring pelajar-pelajar sebagai 'masters' atau 'non-masters' juga dibincangkan. Tiga indeks diskriminasi item untuk klasifikasi 'mastery' dan 'non-mastery' dibincangkan iaitu indeks Brennan  $B$ , koefisien point-biserial  $r_{pb}$  dan koefisien phi  $\phi$ . Hasil kajian ini menunjukkan bahawa perkaitan erat diantara pengajaran, ujian diagnostik dan maklum-balas telah mempertingkatkan tahap pencapaian di kalangan pelajar-pelajar serta persepsi yang positif terhadap penilaian diagnostik. Korelasi purata bernilai .93 diantara  $B$  dan  $\phi$  untuk ujian-ujian diagnostik mencadangkan bahawa perkaitan diantara dua indeks ini adalah rapat samada tahap kesukaran ujian-ujian itu tinggi atau rendah.

## ACKNOWLEDGEMENTS

I express my profound gratitude to Prof. Madya Zulkifli A. Manaf for his invaluable guidance and assistance throughout all stages of this study and in whom the writer has found a great mentor and friend. The kind professor has read drafts of all the chapters, and provided helpful comments gently suggesting ways to improve the dissertation. I am likely deeply grateful to Prof. Suradi Salim, Chairman of the Vetting Committee and Prof. Chew Sing Buan, Reader of the Research Proposal, for their helpful comments and recommendations.

Sincere appreciation is given to the management of Sepang Institute of Technology for granting the writer time off to pursue the degree in Master of Education. Students of the March 1997 Intake Degree Foundation Studies (Science & Engineering) course deserves a great deal of credit for their cooperation and participation in the study.

I owe a very special acknowledgement to my wife, Ng Peng Lan. When I was faced with the decision of pursuing this post-graduate course, she encouraged me to proceed, and in fact took the course herself! For her love, understanding and unwavering support, I dedicate this work to her.

Finally, the writer would like to share the following philosophy which has guided him during the course of writing this dissertation.

‘When things seem difficult, remember the easy and its passing.

When things seem easy, remember the difficult and its passing. Thus patience and perspective are kept and complacency is avoided’.

## TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iii
TABLE OF CONTENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	ix
CHAPTER 1. INTRODUCTION	1
Significance of Study	2
Research Questions	4
Limitations of the Study	4
CHAPTER 2. LITERATURE REVIEW	7
Instructional Objectives and Intended Learning Outcomes	10
Taxonomies of Educational Objectives	10
Instructional Objectives and Assessment	14
Validity	16
Content-related Evidence	16
Construct-related Evidence	18
Criterion-related Evidence	22
Criterion-referenced Tests	23
Setting Performance Standards	26
Standards Based on Theories	28
Standards Based on Expert Consensus	28
Standards Based on Practical Necessities	31
Discrimination Indices for Item Analysis	33
Brennan Discrimination Index, $B$	34

Point-biserial Correlation Coefficient, $r_{pb}$	35
Phi Coefficient, $\phi$	36
Interpretation of Discrimination Indices with Criterion-referenced Tests	36
Mastery Learning	37
Feedback	42
A Diagnostic Assessment Model	45
CHAPTER III. METHODOLOGY	48
Sample	48
Procedure	48
Construction Phase	49
Application Phase	49
Instrumentation	60
The Student Questionnaire	60
Semantic Differential Scale	61
Diagnostic Instrument	61
Test Item Review	66
Logical Review Procedure	66
Empirical Review Procedure	66
Item Difficulty	67
Item Discrimination	68
Test Review	70
Logical Procedure for Construct Validation	71
Empirical Procedure for Construct Validation	71
Test Reliability	73
Standard Error of Measurement	75
CHAPTER IV. RESULTS	76
Performance Data	75
Gain Score Analysis	76
Analysis of Variance	78
Repeated Measures Data	79

Judges' Ratings of Difficulty Level of Tests	83
Perceptions Data	84
Perceptions Data Using Semantic Differential Scale	85
Data on Essay and Multiple-choice Format Tests	91
Test Item Analysis Data	95
Item Data for Diagnostic Decisions	101
 CHAPTER V. DISCUSSIONS AND RECOMMENDATIONS	 128
Introduction	128
Academic Achievement	129
Performance Standards	134
Attitudes and Perceptions	135
Perceptions of Multiple-choice and Essay Formats	141
Interpretation of Observed Test Scores for Diagnostic Decisions	143
Agreement among Discrimination Indices	148
Conclusions and Recommendations	150
Implications of Diagnostic Assessment for Educational Reform	157
 REFERENCES	
 APPENDICES	
Appendix A: Survey of Perceptions towards the Use of Diagnostic Tests	
Appendix B: Survey to Assess Group Differences in the Perceptions of the Diagnostic Test	
Appendix C: Survey to Assess Perceptions towards Multiple-choice Versus Essay Type Diagnostic Tests	
Appendix D: Item Rating Form for Determining Content-related Validity	
Appendix E: List of Formulae	
Appendix F: Organisation of Syllabus into Modules, Units and Behavioural Objectives	
Appendix G: Semester Tests	
Appendix H: Test Specifications and Item Ratings of Judges for Diagnostic Tests 1-7	
Appendix I: Test Difficulty Ratings for First Semester Tests and Second Semester Tests	

## LIST OF FIGURES

Figure		Page
1.	A Model of Diagnostic Assessment	9
2.	Content-Validity Triangle	16
3.	Flowchart Showing the Steps in Monitoring Student Progress	40
4.	A Model of Diagnostic Assessment	47
5.	Pretest and Posttest Scores for Three Achievement Groups	77
6.	Pattern of Test Scores for Three Achievement Groups Using a Time-Series Design	82
7.	Comparison of the Pattern of Test Scores for Three Achievement Groups Before and After Intervention of Diagnostic Assessment	82
8.	Bar-chart of Mean and Composite Ratings on the Likert-scale for the Three Achievement Groups	84
9.	Bar-chart Showing Mean Ratings of the Whole Group on Individual Scales of the Semantic Differential Instrument	87
10.	Bar-chart of Composite and Mean Ratings of the Whole Group on the 5 Dimensions of the Semantic Differential Instrument	87
11.	Bar-chart Showing the Mean Ratings of the Low and high Achievement Groups on the 5 scales in the 5 Dimensions of the Semantic Differential Instrument	89
12.	Bar-chart Showing the Mean Ratings of the Low and High Achievement Groups on Individual Scales of the Semantic Differential Instrument	89
13.	Means for Likert-scale Ratings of essay versus Multiple-choice Format Tests	93
14.	Scatterplots of $B$ , $r_{pb}$ and $\phi$ against $p$ for Test 4	107
15.	Scatterplots of $B$ , $r_{pb}$ and $\phi$ against $p$ for Test 5	110

16.	Scatterplots of $B$ , $r_{pb}$ and $\phi$ for Test 6	113
17.	Scatterplots of $B$ , $r_{pb}$ and $\phi$ for Test 7	116
18.	Scatterplots of $B$ versus $\phi$ for Tests 4-7	122
19.	Scatterplots of $B$ versus $r_{pb}$ for Tests 4-7	124
20.	Scatterplots of $\phi$ versus $r_{pb}$ for tests 4-7	126

# LIST OF TABLES

Table		Page
1.	Means, Standard Deviations, SEM and Reliabilities for the Pretest and Posttest	76
2.	Gain Score Analysis for Each Achievement Group	77
3.	Posttest Mean Scores and Standard Deviations for the Three Achievement Groups	78
4.	One-way Analysis of Variance (ANOVA) for Posttest Mean Scores	78
5.	Mean Pre-treatment and Post-treatment Scores for the Achievement Groups	79
6.	Means and Standard Deviations for the Ratings	83
7.	Analysis of Variance to Compare Mean Scores on Likert Scale for Three Groups	85
8.	Means and Standard Deviations for Semantic Differential Ratings on Classroom Diagnostic Assessment	86
9.	Means and Standard Deviations of Semantic Differential Ratings for the Low and High Achievement Groups	88
10.	Correlation Matrix for Items in the Evaluation Dimension of the Semantic Differential Scale	90
11.	Means and Standard Deviations for Likert-scale Ratings of Essay versus Multiple-choice Format Tests	93
12.	Mean Ratings and Standard Deviations for the Low and High Achievement Groups on the Likert-scale	94
13.	Test Format, Test Length, Test Statistics and Test Reliabilities	96
14.	Correlations of Physics Test Scores with Further Mathematics and Malaysian Studies	96

15.	Item-total Correlations for Essay Format Tests	97
16.	Item Difficulty and Discrimination Indices for Tests 4-7	98
17.	Item Data for Tests 4-7	102
18.	2 X 2 Contingency Tables Showing the Number of Good and Poor Items	119
19.	Inter-correlations between Discrimination Indices Using the Phi Coefficient	120
20.	Inter-correlations for Discrimination Indices for Tests 4-7	121
21.	Matrix to Classify Item and Test Mastery Distinctions	143