UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

# Comparing different assessment formats in undergraduate mathematics

by

Belinda Huntley

Submitted in partial fulfilment of the requirements for the degree

Philosophiae Doctor

in the Department of Mathematics and Applied Mathematics
in the Faculty of Natural and Agricultural Sciences

University of Pretoria
Pretoria

April 2008

© University of Pretoria

# DECLARATION

I, the undersigned, hereby declare that the thesis submitted herewith for the degree Philosophiae Doctor to the University of Pretoria contains my own, independent work and has not been submitted for any degree at any other university.

Name:……
Belinda Huntley

Date:……..

# ABSTRACT

In this study, I investigate how successful provided response questions, such as multiple choice questions, are as an assessment format compared to the conventional constructed response questions. Based on the literature on mathematics assessment, I firstly identify an assessment taxonomy, consisting of seven mathematics assessment components, ordered by cognitive levels of difficulty and cognitive skills. I then develop a theoretical framework, for determining the quality of a question, with respect to three measuring criteria: discrimination index, confidence index and expert opinion. The theoretical framework forms the foundation against which I construct the Quality Index (QI) model for measuring how good a mathematics question is. The QI model gives a quantitative value to the quality of a question. I also give a visual representation of the quality of a question in terms of a radar plot. I illustrate the use of the QI model for quantifying the quality of mathematics questions in a particular undergraduate mathematics course, in both of the two assessment formats – provided response questions (PRQs) and constructed response questions (CRQs). I then determine which of the seven assessment components can best be assessed in the PRQ format and which can best be assessed in the CRQ format. In addition I also investigate student preferences between the two assessment formats.

**Keywords:** Mathematics assessment, Quality Index, good mathematics questions, assessment components, assessment taxonomies, provided response questions, constructed response questions, multiple choice questions.

# DEDICATION

"Yea, if thou criest after knowledge, and liftest up thy voice for understanding; if thou seekest her as silver, and searchest for her as for hidden treasures; then shalt thou understand the fear of the Lord, and find the knowledge of God. For the Lord giveth wisdom; out of His mouth cometh knowledge and understanding".

PROVERBS 2: 3 - 6

# ACKNOWLEDGEMENTS

The author would hereby like to thank all people and organisations whose assistance and co-operation contributed to the completion of this thesis, and in particular:

My supervisor, Professor Johann Engelbrecht, for setting high professional standards which provided the much-needed challenge and motivation, and for his interest and moral support.

My co-supervisor, Professor Ansie Harding, for her invaluable guidance and expert assistance throughout the period of this research.

Elsie Venter, a senior lecturer from the Centre for Evaluation and Assessment, School of Education, University of Pretoria, for introducing me to the Rasch method of data analysis and for her assistance in analysing my research data.

Marie Oberholzer, for editing and type-setting the final draft of my thesis with great care and diligence.

My parents, Roland and Daisy Hill, for their prayers of upliftment and loving support.

My husband, Brian and children, Byron, Christopher and Cayla, for their total devotion and patience and on-going faith in my abilities.

# INDEX OF TABLES


| | | Page |
|---|---|---|
| Table 1.1 | Student numbers and pass rates for undergraduate mathematics courses, 2000-2004 | 8 |
| Table 1.2 | Exit level outcomes (ELOs) | 266 |
| Table 1.3 | Associated assessment criteria (AAC) | 267 |
| Table 1.4 | Critical cross-field outcomes (CCFOs) | 268 |
| Table 2.1 | MATH Taxonomy | 26 |
| Table 3.1 | MATH109 student interviewees and their academic backgrounds | 87 |
| Table 3.2 | Probabilities of correct response for persons on items of different relative difficulties | 102 |
| Table 5.1 | Mathematics assessment component taxonomy and cognitive level of difficulty | 137 |
| Table 5.2 | Mathematics assessment component taxonomy and cognitive skills | 138 |
| Table 5.3 | Decision matrix for an individual student and for a given question, based on combinations of correct or wrong answers and of low or high average CI | 154 |
| Table 5.4 | Classification of difficulty intervals | 169 |
| Table 6.1 | Characteristics of tests written | 178 |
| Table 6.2 | Misfitting and discarded test items | 269 |
| Table 6.3 | Component analysis – trends | 232 |
| Table 7.1 | A comparison of the success of PRQs and CRQs in the mathematics assessment components | 244 |

# INDEX OF FIGURES


| | | Page |
|---|---|---|
| Figure 2.1 | SOLO Taxonomy | 28 |
| Figure 2.2 | Classification according to lecturer's purpose | 29 |
| Figure 2.3 | Learning-required classification | 30 |
| Figure 2.4 | De Lange's level of understanding | 31 |
| Figure 2.5 | Cycle of formative and summative assessment | 37 |
| Figure 2.6 | Integrated assessment | 47 |
| Figure 3.1 | Number of misreadings of nine subjects in two tests | 92 |
| Figure 3.2 | How differences between person ability and item difficulty ought to affect the probability of a correct response | 98 |
| Figure 3.3 | The item characteristics curve | 99 |
| Figure 3.4 | Item characteristic curve of the dichotomous Rasch model | 103 |
| Figure 3.5 | Mathematics I Major (MATH109) assessment programme | 110 |
| Figure 5.1 | Illustration of confidence deviation from the best fit line between item difficulty and confidence | 161 |
| Figure 5.2 | Illustration of expert opinion deviation from the best fit line between item difficulty and expert opinion | 163 |
| Figure 5.3 | Visual representation of the three axes of the QI | 164 |
| Figure 5.4 | Quality index for PRQ | 165 |
| Figure 5.5 | A good quality item | 166 |
| Figure 5.6 | A poor quality item | 167 |
| Figure 5.7 | Distribution of six difficulty levels | 168 |
| Figure 7.1 | A good quality item | 238 |
| Figure 7.2 | A poor quality item | 238 |
| Figure 7.3 | A difficulty, poor quality item | 239 |
| Figure 7.4 | An easy, good quality item | 239 |

# TABLE OF CONTENTS

Page

DECLARATION i
ABSTRACT ii
DEDICATION iii
ACKNOWLEDGEMENTS iv
INDEX OF TABLES v
INDEX OF FIGURES vi

CHAPTER 1: INTRODUCTION
1.1 Purpose of study 1
1.2 Statement of problem 2
1.3 Significance of the study 4
1.4 Context of this study 7
1.5 Outline of study 11

CHAPTER 2: LITERATURE REVIEW
2.1 Terminology 15
2.2 The changing nature of university assessment in the South African context 17
2.3 Assessment models in mathematics education 21
2.4 Assessment taxonomies 24
2.5 Assessment purposes 33
2.5.1 Diagnostic assessment 33
2.5.2 Formative assessment 33
2.5.3 Summative assessment 35
2.5.4 Quality assurance 37
2.6 Shifts in assessment 38
2.7 Assessment approaches 39
2.7.1 The traditional approach 40
2.7.2 Computer-based (online) assessment 40
2.7.3 Workplace- and community-based/learnership assessment 44
2.7.4 Integrated or authentic assessment 44
2.7.5 Continuous assessment 48
2.7.6 Group-based assessment 49
2.7.7 Self-assessment 49
2.7.8 Peer-assessment 50
2.8 Question formats 51
2.9 Constructed response questions and provided response questions 52
2.10 Multiple choice questions 56
2.10.1 Advantages of MCQs 60
2.10.2 Disadvantages of MCQs 63

2.10.3 Guessing 67
2.10.4 In defense of multiple choice 69
2.11 Good mathematics assessment 70
2.12 Good mathematics questions 74
2.13 Confidence 77

CHAPTER 3: RESEARCH DESIGN AND METHODOLOGY
3.1 Research design 82
3.2 Research questions 84
3.3 Qualitative research methodology 85
3.3.1 Qualitative data collection 86
3.4 Quantitative research methodology 89
3.4.1 The Rasch model 89
3.4.1.1Historical background 91
3.4.1.2 Latent trait 96
3.4.1.3 Family of Rasch models 101
3.4.1.4 Traditional test theory versus Rasch latent trait theory 105
3.4.1.5 Reliability and validity 107
3.4.2 Quantitative data collection 109
3.5 Reliability, validity, bias and research ethics 115
3.5.1 Reliability of the study 115
3.5.2 Validity of the study 116
3.5.3 Bias of the study 118
3.5.4 Ethics 119

CHAPTER 4: QUALITATIVE INVESTIGATION
4.1 Qualitative data analysis 122
4.2 Qualitative investigation 122

CHAPTER 5: THEORETICAL FRAMEWORK
5.1 Mathematics assessment components 135
5.1.1 Question examples in assessment components 138
5.2 Defining the parameters 149
5.2.1 Discrimination index 150
5.2.2 Confidence index 153
5.2.3 Expert opinion 157
5.2.4 Level of difficulty 159
5.3 Model for measuring a good question 160
5.3.1 Measuring criteria 160
5.3.2 Defining the quality index (QI) 163
5.3.3 Visualising the difficulty level 167

CHAPTER 6: RESEARCH FINDINGS
6.1 Qualitative data analysis 172
6.1.1 Methodology 172
6.2 Data description 178
6.3 Component analysis 179
6.4 Results 232

6.4.1 Comparison of PRQs and CRQs within each assessment component 232

CHAPTER 7: DISCUSSION AND CONCLUSIONS
7.1 Good and poor quality mathematics questions 235
7.2 A comparison of PRQs and CRQs in the mathematics assessment components 239
7.3 Conclusions 242
7.4 Addressing the research questions 244
7.5 Limitations of study 247
7.6 Implications for further research 248

REFERENCES 251

APPENDIX
Appendix A1 Declaration letter 265
Appendix A2 Table 1.2: Exit level outcomes (ELOs) of the undergraduate curriculum 266
Appendix A3 Table 1.3: Associated assessment criteria (AAC) 267
Appendix A4 Table 1.4: Critical cross-field outcomes (CCFOs) 268
Appendix A5 Table 6.2: Misfitting and discarded test items 269
Appendix A6 Test items Rasch statistics 270
Appendix A7 Confidence level items Rasch statistics 274
Appendix A8 Item analysis data 279