

Springer Texts in Education

Springer Texts in Education delivers high-quality instructional content for graduates and advanced graduates in all areas of Education and Educational Research. The textbook series is comprised of self-contained books with a broad and comprehensive coverage that are suitable for class as well as for individual self-study. All texts are authored by established experts in their fields and offer a solid methodological background, accompanied by pedagogical materials to serve students such as practical examples, exercises, case studies etc. Textbooks published in the Springer Texts in Education series are addressed to graduate and advanced graduate students, but also to researchers as important resources for their education, knowledge and teaching. Please contact Natalie Rieborn at textbooks.education@springer.com for queries or to submit your book proposal.

More information about this series at <http://www.springer.com/series/13812>

David Andrich · Ida Marais

A Course in Rasch Measurement Theory

Measuring in the Educational, Social
and Health Sciences

 Springer

David Andrich
Graduate School of Education
The University of Western Australia
Crawley, WA, Australia

Ida Marais
Graduate School of Education
The University of Western Australia
Crawley, WA, Australia

ISSN 2366-7672

ISSN 2366-7680 (electronic)

Springer Texts in Education

ISBN 978-981-13-7495-1

ISBN 978-981-13-7496-8 (eBook)

<https://doi.org/10.1007/978-981-13-7496-8>

Library of Congress Control Number: 2019935842

© Springer Nature Singapore Pte Ltd. 2019

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd. The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

Preface

This book has arisen from two postgraduate level courses in Rasch measurement theory that have been taught both online and in intensive mode for over two decades at Murdoch University and The University of Western Australia. The theory is generally applied in the fields of education, psychology, sociology, marketing and health outcomes to create measures of social constructs. Social measurement often begins with assessments in ordered categories, with two categories being a special case. To increase their reliability and validity, instruments are composed of multiple, distinct items which assess the same variable. Rasch measurement theory is used to assess the degree to which the design and administration of the instrument are successful and to diagnose problems which need correcting. Following confirmation that an instrument is working as required, persons may be measured on a linear scale with an arbitrary unit and arbitrary origin.

The main audiences for the book are graduate students and professionals who are engaged in social measurement. Therefore, the emphasis of course is on first principles of both the theory and its applications. Because software is available to carry out analyses of real data, small hand-worked examples are presented in the book. The software used in the analysed examples, which is helpful in working through the text, is RUMM2030 (Rasch unidimensional models for measurement). Although the first principles are emphasized, much of the course is based on research by the two authors and their colleagues.

The distinctive feature of Rasch measurement theory is that the model studied in this book arises independently of any data—it is based on the requirement of invariant comparisons of objects with respect to instruments within a specified frame of reference and vice versa. This is a feature of all measurement. Deviations of the data from the model are taken as anomalies to be explained and the instrument improved. The approach taken is to provide the researcher with confidence to be in control of the analysis and interpretation of data, and to make professional rather than primarily statistical decisions. Because statistical principles are necessarily involved, reviews of the necessary statistics are provided in Appendix D.

Graduates and professionals are likely to encounter classical test theory. Therefore, introductory chapters review the elements of this theory. The perspective on the relationship between Rasch measurement theory and classical test theory is that the former is an elaboration of the ideals of the latter, not that they are entirely in conflict. However, because the centrality of invariance as a requirement for measurement had been articulated by two giants of social measurement, L. L. Thurstone and L. Guttman, reference is made to their work. In particular, Thurstone had articulated the requirements of invariance in almost identical terms as G. Rasch, but did not express it in terms of a mathematical equation, and the elementary Guttman design which is introduced in the early chapters, is shown to be a deterministic form of the Rasch model. The distinctive contribution of Rasch compared to that of Thurstone and Guttman is that the model studied in this book has built into it the principle of invariance and is immediately probabilistic. Therefore, the deviation of data from the model implies some kind of deviation from invariance and measurement. Together with the relationships shown with classical test theory, the book provides a unified theme for approaches to social measurement, rather than as a compendium of techniques.

Finally, the book stresses that the requirement of invariance, and its expression in the Rasch model, is necessary, but not sufficient to ensure sound measurement. All the principles of measurement, of experimental design and of statistical inference must be applied in the process of constructing instruments that provide invariance of comparisons and reliable and valid measurement. Indeed, the explicit requirements of invariance in the Rasch model can at times appear more demanding of the data than do other theories and approaches.

Crawley, Australia

David Andrich
Ida Marais

Acknowledgements

RUMM2030, which is a Windows, menu-driven program, has been written primarily by Barry Sheridan. He has written the program so that it permits an efficient exposition of the theory and the approach emphasized in the book for data analyses. Alan Lyne contributed to the original programming and further contributions were made by Guanzhong Luo. Irene Styles has been a colleague both in research and in improving the courses on which this book is based. Many students have also provided feedback, including Sonia Sappl who has contributed to the editing of the book. Natalie Carmody has administered the courses for more than a decade and helped prepare the book. The first author also acknowledges the deep influence of a year of study with the Danish mathematician and statistician Georg Rasch in the 1970s when Rasch had turned to the philosophy of measurement. The first author also acknowledges the support of the Australian Research Council for a range of grants over more than 30 years that have helped him conduct research into Rasch measurement theory.

Contents

Part I General Principles and the Dichotomous Rasch Model

1	The Idea of Measurement	3
	Latent Traits	3
	Assessment: A Distinction Between Latent and Manifest	4
	Scoring Assessments	4
	Dichotomous Items and Their Scoring	5
	Polytomous Items and Their Scoring	5
	Key Features of Measurement in the Natural Sciences	6
	Stevens' Levels of <i>Measurement</i>	7
	Nominal Use of Numbers	7
	Ordinal Use of Numbers	7
	Interval Use of Numbers	8
	Ratio Use of Numbers	8
	Reliability and Validity	9
	Some Definitions	9
	A Model of Measurement	10
	Exercises	10
	References	11
	Further Reading	11
2	Constructing Instruments to Achieve Measurement	13
	Constructing Tests of Proficiency to Achieve Measurements	15
	Constructing Rating Scales to Achieve Measurements	18
	Number, Order and Wording of Response Categories	19
	An Example of the Assessment of Writing by Raters	20
	An Example of the Assessment of the Early Development Indicator Instrument	22
	The Measurement of Attitudes: Two Response Mechanisms	23
	An Example: The Cumulative Mechanism	23

	An Example: The Unfolding Mechanism	24
	A Practical Approach: Likert Scales	25
	Exercises	28
	References	28
3	Classical Test Theory	29
	Elements of CTT	30
	The Total Score on an Instrument	30
	Reliability, True and Error Scores	31
	Statistics Reviews	31
	Item Analysis	33
	Facility of an Item	33
	Discrimination of an Item	34
	Person Analysis	35
	Notation and Assumptions of CTT	35
	Basic Equations of CTT	35
	Reliability of a Test in CTT r_{yy}	36
	The Standard Error of Measurement s_e	37
	Statistics Reviews	37
	Example	37
	Exercises	38
	Reference	39
4	Reliability and Validity in Classical Test Theory	41
	Validity	42
	Reliability	43
	Reliability in Terms of Items	45
	Coefficient Alpha (α): Estimating Reliability in CTT	47
	Example	48
	Factors Affecting the Reliability Index	48
	Internal Factors	49
	External Factors	50
	Common Factors Affecting Reliability and Validity	51
	Causal and Index Variables	51
	Exercises	52
	References	53
	Further Reading	53
5	The Guttman Structure and Analysis of Responses	55
	The Guttman Structure	56
	Interpretations of the Continuum in the Guttman Structure	57
	Elementary Analysis According to the Guttman Structure in the Case of a Proficiency Example	59

Item Analysis	63
Person Analysis	68
Extended Guttman Analysis: Polytomous Items	69
Exercises	73
References	74
Further Reading	74
6 The Dichotomous Rasch Model—The Simplest Modern Test Theory Model	75
Abstracting the Proportion of Successes in a Class Interval to Probabilities	75
A Two-Way Frame of Reference and Modelling a Person’s Response to an Item	78
Engagements of Persons with Items	79
Formalizing Parameters in Models	79
Effects of Spread of Item Difficulties	80
Person–Item Engagements	82
Examples	83
Item Characteristic Curve and the Location of an Item	84
The Dichotomous Rasch Model: A General Formula	85
Specific Objectivity	86
Exercises	86
References	87
Further Reading	87
7 Invariance of Comparisons—Separation of Person and Item Parameters	89
Conditional Probabilities with Two Items in the Rasch Model	90
Example	92
The Condition of Local Independence	93
The Principle of Invariant Comparisons	93
Exercises	94
Reference	94
Further Reading	95
8 Sufficiency—The Significance of Total Scores	97
The Total Score as a Sufficient Statistic	97
The Response Pattern and the Total Score	100
Exercises	103
References	103
9 Estimating Item Difficulty	105
Application of the Conditional Equation with Just Two Dichotomous Items and Many Persons	105

Estimating Relative Item Difficulties	105
Estimating Person Proficiencies	110
An Arbitrary Origin and an Arbitrary Unit	111
The Arbitrary Origin	111
The Arbitrary Unit	112
Generalizing to Many Items	113
Maximum Likelihood Estimate (MLE)	113
Item Difficulty Estimates	114
Exercises	115
Further Reading	115
10 Estimating Person Proficiency and Person Separation	117
Solution Equations in the Rasch Model	117
The Solution Equation for the Estimate of <i>Person Proficiency</i>	119
Solving the Equation by Iteration	120
Initial Estimates	121
Proficiency Estimates for Each Person	122
For Responses to the Same Items, the Same Total Score	
Leads to the Same Person Estimate	122
Estimate for a Score of 0 or Maximum Score	122
The Standard Error of Measurement of a Person	125
Proficiency Estimate for Each Total Score When All Persons	
Respond to the Same Items	125
Estimates for Every Total Score	126
Non-linear Transformation from Raw Score to Person Estimate	127
Displaying Person and Item Estimates on the Same Continuum	128
CTT Reliability Calculated from Rasch Person Parameter	
Estimates	129
Derivation of r_{β}	129
Principle of Maximum Likelihood	131
Bias in the Estimate	133
Exercises	134
References	135
Further Reading	135
11 Equating—Linking Instruments Through Common Items	137
Linking of Instruments with Common Items	137
Linking Three Items Where One Item Is Common to Two Groups	137
Estimating Differences Between Difficulties and then Adjusting	
the Origin	138
Estimating Differences Between Difficulties Simultaneously	
by Maximum Likelihood	140
Estimating Item Parameters Simultaneously by Maximum	
Likelihood in the Presence of Missing Responses	142

Equating Scores of Persons Who Have Answered Different Items from the Same Set of Items	144
Applications	146
References	147
Further Reading	148
12 Comparisons and Contrasts Between Classical and Rasch Measurement Theories	149
Motivations and Background to CTT and RMT	149
Motivation of CTT	149
Motivation of RMT	150
Relating Characteristics of CTT and RMT	151
The Total Scores of Persons	151
CTT Estimation of the True Score	153
RMT Estimation of the Person Location Estimates	155
CTT Estimation of Standard Errors of True Scores	156
RMT Estimation of Standard Errors of Person Location Estimates	157
References	158
Further Reading	158
Part II The Dichotomous Rasch Model: Fit of Responses to the Model	
13 Fit of Responses to the Model I—Item Characteristic Curve and Chi-Square Tests of Fit	161
A Graphical Test of Item Fit	161
The Item Characteristic Curve (ICC)	161
Observed Proportions in Class Intervals	162
A Formalised Test of Item Fit— χ^2	167
Interpretation of Computer Printout—Test of Fit Output	169
Exercises	171
Reference	171
Further Reading	171
14 Violations of the Assumption of Independence I—Multidimensionality and Response Dependence	173
Local Independence	173
Two Violations of Local Independence	174
Multidimensionality	175
Formalization of Multidimensionality	175
Detection of Multidimensionality	177
Other Tests of Multidimensionality	178

- Response Dependence 180
 - Formalization of Response Dependence 180
 - Detection of Response Dependence 181
 - Estimating the Magnitude of Response Dependence 182
- The Effects of Violations of Independence 184
- Exercises 184
- References 184

- 15 Fit of Responses to the Model II—Analysis of Residuals and General Principles 187**
 - The Fit-Residual 187
 - Approximations for the Degrees of Freedom 188
 - Shape of the Natural Residual Distributions 189
 - Interpreting the Sign of the Fit-Residual 190
 - Outfit as a Statistic 190
 - Infit as a Statistic 190
 - The Correlation Among Residuals 191
 - The Principal Component Analysis (PCA) of Residuals 191
 - General Principles in Assessing Fit 192
 - Interpreting Fit Statistics Relatively and in Context 192
 - Power of the Tests of Fit as a Function of the Sample Size 193
 - Sample Size in Relation to the Number of Item Thresholds 193
 - Adjusting the Sample Size 194
 - Power of Tests of Fit as a Function of the Separation Index 194
 - Test of Fit is Relative to the Group and the Set of Items 196
 - Bonferroni Correction 196
 - RUMM2030 Specifics 196
 - Exercises 197
 - References 197

- 16 Fit of Responses to the Model III—Differential Item Functioning 199**
 - Identifying DIF Graphically 200
 - Identifying DIF Statistically Using ANOVA of Residuals 201
 - Artificial DIF 205
 - Resolving Items 206
 - Exercises 207
 - References 207
 - Further Reading 207

- 17 Fit of Responses to the Model IV—Guessing 209**
 - Tailored Analysis 210
 - Identifying and Correcting for Guessing 211
 - Exercises 213

References 213

Further Reading 213

18 Other Models of Modern Test Theory for Dichotomous Responses 215

The Rasch Model 215

2PL Model 216

3PL Model 217

References 218

19 Comparisons and Contrasts Between Item Response Theory and Rasch Measurement Theory 221

Approaches to Measurement and the Data-Model Relationship in Measurement 221

 Approach 1 222

 Approach 2 222

The Function of Measurement in Quantitative Research in the Natural Sciences: Thomas Kuhn 223

 What Do Text Books Teach Is the Function of Measurement in Science? 223

 What Does Kuhn Say Is the Function of Measurement in Scientific Research? 223

 Is There a Role for Qualitative Study in Quantitative Scientific Research? 223

 What Is the Function and Role of Measurement in Science? 224

The Properties Required of Measurement in the Social Sciences: L. L. Thurstone 224

 Social Variables—What Is Distinctive About Variables of Measurement in the Social Sciences and What Are the Limits to Such Variables? 224

 Thus They Must Be Independent of Physical Variables—What Else? 224

 Why Do You Think We Have Quantification in the Social Sciences? 225

 A Requirement for Measuring Instruments 225

Georg Rasch 225

The Criterion of Invariance 226

Fit with Respect to the Model and Fit with Respect to Measurement 227

The Linear Continuum as an Idealization 228

Exercises 228

References 228

Further Reading 229

Part III Extending the Dichotomous Rasch Model: The Polytomous Rasch Model

20	The Polytomous Rasch Model I	233
	The Model for Ordered Polytomous Responses	233
	Test of Fit Between the Data and the Model	237
	Interpretation from a Computer Output	237
	Proportions in Each Category	237
	Threshold Estimates for the Items	238
	Location (Difficulty) Estimates for the Items	239
	The Test of Fit for a Dichotomous Item Scored 0 or 1	239
	The Test of Fit for a Partial Credit Item m009 Scored 0, 1, 2 or 3	240
	Threshold Order for Item m009 Scored 0, 1, 2 and 3	241
	Threshold Order for Item m010 Scored 0, 1, 2, 3 and 4	242
	Estimates of the Proficiencies of the Persons	243
	Exercises	244
	References	244
21	The Polytomous Rasch Model II	245
	Rated Data in the Social Sciences	245
	The Partial Credit and Rating Scale Specifications	246
	The Generalization to Three Ordered Categories	247
	The Expected Value Curve	248
	The Structure of the PRM	249
	The Generalization to Any Number of Categories $m + 1$	250
	The Slope of $E[X]$	252
	Latent Threshold Curves	253
	Diagnosing Problems with the Functioning of the Categories	254
	The Partial Credit and Rating Parameterizations of the PRM	255
	The Rating Scale Parameterization	256
	The Partial Credit Parameterization	257
	Exercises	258
	References	258
22	The Polytomous Rasch Model III	261
	Reparameterisation of the Thresholds	261
	Equidistant Thresholds	262
	Recovering the Thresholds	263
	Non-equidistant Thresholds	264
	Inference of an Independent Response Space	266
	Rescoring Items	266

Exercises	268
References	268
Further Reading	268
23 Fit of Responses to the Polytomous Rasch Model	269
The Fit-Residual	269
Deriving the Fit-Residual for the Persons	270
Deriving the Fit-Residual for the Items	272
References	273
Further Reading	273
24 Violations of the Assumption of Independence II—The Polytomous Rasch Model	275
A Model that Accounts for Dependent Items	275
Reparameterization of the Thresholds of the PRM—The Spread Parameter	276
Diagnosis of Multidimensionality	278
Subtest Analysis	278
Estimating the Magnitude of Multidimensionality	279
Testing the Equivalence of Person Estimates from Two Subsets of Items	281
Diagnosis of Response Dependence	282
Formalization of Response Dependence in the PRM	282
Estimating the Degree of Response Dependence Between Polytomous Items	284
Standard Errors of the Magnitude of the Estimate of d	286
Exercises	287
References	288
Further Reading	288
 Part IV Theoretical Justifications and Further Elaborations	
25 Derivation of Classical Test Theory Equations and Coefficient α	291
Formalization and Derivation of CTT Eqs. (3.1)–(3.5) in Chap. 3	291
Derivation of Covariance	292
Derivation of the Standard Error of Measurement	295
Derivation of the Equation for Predicting the True Score from the Observed Score	295
Derivation of Coefficient α	297

26 Analysis of More Than Two Facets and Repeated Measures 299
 From a Two-Facet to a Three-Facet Rasch Model Analysis 299
 Repeated Measures 301
 Repeated Measurements and Response Dependence 303
 Exercises 304
 References 305

27 Derivation of the Threshold Form of the Polytomous Rasch Model 307
 Measurement and Ordered Response Categories 307
 Minimum Proficiencies and Threshold Difficulty
 Order in the Full Space Ω 308
 Specifying the Dichotomous Rasch Model for Responses
 at the Thresholds 310
 The Response Subspace Ω^G 311
 Formalizing the Response Space Ω^G 313
 Generalizing the Notation of Grade Classification 314
 A Fundamental Identity of the PRM 314
 The Full Space Ω 315
 The Guttman Subspace Ω^G 315
 The Dichotomous Rasch Model Identity in Ω and $\Omega_{x-1,x}^G$ 316
 Exercises 317
 References 318

28 Non-Rasch Measurement Models for Ordered Response Categories 319
 The Nominal Response Model 319
 Relationship Between the PRM and the NRM 320
 The Generalized Partial Credit Model 322
 The Graded Response Model 322
 Estimation of Parameters in the Non-Rasch Models 324
 Exercises 325
 References 325
 Further Reading 326

29 Review of Principles of Test Analysis Using Rasch Measurement Theory 327
 Invariance of Comparisons and RMT 327
 Total Score as the Sufficient Statistic 329
 Dichotomous Items: The Probabilistic Guttman Structure 330
 Reasons for Multiple Items in Instruments 330
 Evidence from the Location and Thresholds of Items 331
 Construction of Items 331

- Assessing the Fit Between Responses and the Rasch Model 336
 - Meaning of Fit to the Rasch Model 336
 - Identifying Misfitting Items 337
 - Dealing with Misfitting Items 338
 - Separating the Scale Construction and Person Measurement Stages 340
 - Summary 341
 - Exercises 342
 - References 342
- Appendix A: Test Items for Chapter 3 and Solutions 343**
- Appendix B: Chapter Exercises Solutions 347**
- Appendix C: RUMM2030 Exercises and Solutions 355**
- Appendix D: Statistics Reviews, Exercises and Solutions 389**
- Index 479**