

**MAINTAINING A COMMON ARBITRARY  
UNIT IN SOCIAL MEASUREMENT**

**STEPHEN HUMPHRY**

**2005**

**Submitted in fulfillment of the requirements of the degree of**

**Doctor of Philosophy**

**School of Education,**

**Murdoch University, Perth, Western Australia**

I declare that this dissertation is my own account of my research and contains as its main content work which has not previously been submitted for a degree at any tertiary institution.

STEPHEN HUMPHRY.

## ABSTRACT

In educational assessment, it is common to equate test forms in order to draw comparisons between different populations of students. The process of test equating presents a number of challenges, many of which relate inherently to the problem of maintaining a common unit and origin.

In order to develop a general theoretical approach to maintaining a common unit and origin in the measurement of quantitative attributes, the role of the unit is carefully examined. Classical physics is explicitly adopted as the guiding paradigm during the investigations throughout the dissertation. Accordingly, the central objective is to develop a theoretical foundation for maintaining a common unit and origin which meets two criteria: (i) it must be congruent with the definition of measurement in physics captured in the *classical theory of measurement* (Michell, 1999); and (ii) it must meet a key requirement of measurement in the physical sciences identified by Rasch (1960/1980). Rasch identified the relevant requirement, that of *invariant comparison*, based on analysis of Newton's second law and showed that the Principle of Invariant Comparison is formally embodied in his measuring function for dichotomous data (Rasch, 1960/1980). This model provides the basis for the development and exposition of general concepts and principles in the dissertation.

In order to achieve the central objective, the unit is made formally explicit and specified in relation to the experimental frame of reference. Rasch (1977) defined a *Specified Frame of Reference* (SFR) in terms of a collection of objects, a collection of agents, and outcomes of the interaction between these. Drawing on a fundamental

distinction introduced by Andrich (2003), the unit of a SFR is referred to as a *natural unit* and is distinguished from an *arbitrary unit*, the magnitude of which is theoretically independent of any particular SFR and instrument contained within. From this distinction, a definition of *discrimination* arises naturally; a definition that is also congruent with classical physics. The distinction and related definitions provide the basis for derivation of a general form of Rasch's measuring function for dichotomous data, referred to as the Extended Frame of Reference Model (EFRM). It is shown that the EFRM provides a rational basis for maintaining a common unit and origin in assessment contexts involving two or more Specified Frames of Reference.

Simulation and empirical studies are employed to illustrate application of the EFRM. These studies also serve to illustrate that quantitative hypotheses entailed by the EFRM are open to empirical tests by providing a context for the use of graphical methods and statistical tests of fit. Empirical investigations are used to illustrate consequences of differences between natural units in the context of applied educational assessment. The studies also provide a context in which to characterise the model, and the structure of data that it entails. Although the simulation studies demonstrate the basic efficacy of the model, they also indicate scope for improvement in terms of the precision of estimates. To explore possible approaches to refining the estimation process, Maximum Likelihood (ML) equations are derived and examined. Firstly, Joint Maximum Likelihood (JML) equations are presented. Following this, Conditional Maximum Likelihood (CML) equations are derived. It is shown that while the CML equations permit separation of the person and item parameters, item locations are expressed in terms of natural, rather than arbitrary, units. A particular approach is proposed, emphasising links to the classical

theory and the Principle of Invariant Comparison. In considering the proposed approach, a distinguishing feature of the definition of discrimination is highlighted: specifically, the nature of its definition represents the importance of *relationships between* quantitative attributes, and the specific *structure* of these relationships, to the measurement of any *particular* attribute. Although it is not possible to fully study this feature given the scope of the work, it is a key to the implications of the general theoretical framework embodied in the EFRM. Accordingly, these implications are touched on before concluding the dissertation.

## PREFACE

This work is about the importance of the context of the process of measurement to the unit. The *Merriam-Webster* dictionary defines *context* as “the interrelated conditions in which something exists or occurs” and lists, as a definition of *condition*, “something essential to the appearance or occurrence of something else”. This work is, accordingly, about the interrelated conditions essential to the occurrence of a unit of a quantitative attribute.

An important feature of the work is the explicit incorporation of symbols within functions to denote units of continuous quantity. In physics, symbols for quantities are frequently omitted from equations and understood to be implicit. However, as stated in a classic textbook: “The beginning student will do well to include units of all physical quantities, as well as their magnitudes, in all his calculations” (Sears et al., 1981, p. 4). One might object that there are no standard units outside of the physical sciences. The question should be asked, though: how are we to seek to obtain units with any level of universality if their existence is not explicitly hypothesised? The purpose of this work is, accordingly, to develop a theoretical framework within which to formally specify the unit with respect to the empirical context as a basis for stating and testing quantitative hypotheses.

It is noteworthy that the definition of each of the *base* units in the *Système International d'Unités* (SI) makes reference to specific empirical conditions. The metre is defined in terms of the path of light traveling through a vacuum in a specific time interval, the kelvin in terms of a fraction of the triple point of water, the kilogram in terms of an empirical prototype, and so on. Specification of an empirical context has been

instrumental to obtaining precise universal units. In addition, all but one of the definitions of the SI units of quantity, base or derived, makes reference to one or more *other* quantitative properties. Even in the case of the one exception, the kilogram, the original definition referred to volume. It seems, therefore, that precise definition or specification of a unit of any *particular* quantitative property must be framed in terms of other quantitative properties. Indeed, derived SI units are defined purely in terms of relationships between quantities: specifically, any given derived unit is defined in terms of multiplicative relationship involving itself and other units.

A feature of the theoretical framework developed in this dissertation is that specification of the unit with respect to one or more characteristics of an experimental frame of reference entails the quantitative hypothesis of a relationship *between* attributes. Although this feature was not deliberately sought, it potentially carries implications of considerable importance, as highlighted at key points in the body of the work. Indeed, I am hopeful that the congruence of this feature of the work with physics signifies the potential of the framework as a general foundation for generating and investigating specific quantitative hypotheses.

## ACKNOWLEDGEMENTS

I would like to express my appreciation to David Andrich for his guidance, encouragement, generosity, and especially for the balance between rigor and open-mindedness he brings to scientific endeavour. On a specific feature of the work, I would particularly like to thank David for persuading me that Rasch's concept of a Specified Frame of Reference should assume a focal point for the development and exposition of the central thesis. As should become clear, the concept became an indispensable element of the theoretical framework developed within the dissertation.

In addition, I would like to acknowledge the critical and constructive input of my co-supervisor and friend, Guanzhong Luo, particularly in relation to mathematical features of the work. I would also like to acknowledge the support of my family and friends; and especially my father, Peter Humphry, for the many discussions about elements of this work related to physics, and my wife, Jia-Yu, for her endless support and understanding. Lastly, I am grateful for the support I have received within the Department of Education and Training, WA, and wish to acknowledge permission for the use of empirical data sets featured within the dissertation.

## TABLE OF CONTENTS

ABSTRACT	i
PREFACE	iv
ACKNOWLEDGMENTS	vi
LIST OF FIGURES	ix
LIST OF TABLES	xi
LIST OF ACRONYMS	xii
CHAPTER 1: FOUNDATIONS	1
1.1 Introduction	1
1.2 Contextual overview	10
1.3 The RM and its features	13
1.4 Specification of experimental context	19
1.5 Measurement in the physical sciences	22
1.6 The RM and the classical theory of measurement	41
CHAPTER 2: THE EXTENDED FRAME OF REFERENCE	46
2.1 The unit in the context of a SFR	47
2.2 The unit and origin in the context of an EFR	50
2.3 The Extended Frame of Reference Model	56
2.4 Extending the domain of the Principle of Invariant Comparison	63
2.5 Maintaining a common unit and origin	68
CHAPTER 3: SIMULATION STUDIES	74
3.1 Introduction and overview	74
3.2 Methods for interrogating the data	76
3.3 PGD simulation study	84
3.4 ISD simulation study	101

3.5	ISD-PGD interaction simulation study	114
3.6	Discussion and summary of the simulation studies	127
CHAPTER 4: EMPIRICAL INVESTIGATIONS		129
4.1	Introduction and overview	129
4.2	Empirical investigations regarding PGD	132
4.3	Empirical investigations regarding ISD	142
CHAPTER 5: MAXIMUM LIKELIHOOD EQUATIONS		161
5.1	Joint Maximum Likelihood	163
5.2	Conditional Maximum Likelihood	166
5.3	A proposed approach to ML estimation	171
5.4	Standard errors of the parameters of the EFRM	175
CHAPTER 6: DISCUSSION AND CONCLUSION		178
6.1	Discussion	178
6.2	Conclusion	181
APPENDICES		183
BIBLIOGRAPHY		208

## LIST OF FIGURES

Figure	Title	Page
1.1	Specified Frame of Reference	19
1.2	Multiple frames of reference	21
1.3	Balance scale	28
1.4	Distribution of $m_{ji}$ for the prototype example	35
1.5	Distribution of $m_{ji}$ for Case 2 of the prototype example	37
2.1	An example of a common scale with common origin and arbitrary unit, and its relation to scales obtained from two Specified Frames of Reference	54
2.2	ICCs for three items within two Specified Frames of Reference, $F_{11}$ and $F_{12}$ , as shown in legend	58
3.1	Scatterplot of Group 2 and Group 4 item estimates	89
3.2	Scatterplot of Group 2 and Group 4 item estimates against the identity line	90
3.3	Scatterplot of Group 3 and Group 4 item estimates	91
3.4(a)	RM ICCs for Item 14	94
3.4(b)	EFRM ICCs for Item 14	95
3.5(a)	The RM ICC for Group 4 on Item 14	95
3.5(b)	The EFRM ICC for Group 4 on Item 14	96
3.6	Scatterplot of mean abilities derived from item Sets 2 and 3	105
3.7	Scatterplot of mean abilities derived from item Sets 2 and 4	106
3.8	Scatterplot of mean abilities derived from item Sets 1 and 3	106
3.9	Comparisons of transformed group means with simulated values for each of the four item sets	107
3.10(a)	Simulated and estimated item locations for the RM, ISD study	108
3.10(b)	Simulated and estimated item locations for the EFRM, ISD study	109
3.11(a)	RM ICC for Item 16	112
3.11(b)	EFRM ICC for Item 16	112
3.12(a)	RM ICC for Item 17	112
3.12(b)	EFRM ICC for Item 17	113
3.13	Item locations derived from Groups 2 and 3 according to item set	120
3.14(a)	Simulated and estimated item locations for the EFR analysis	121

3.14(b)	Simulated and estimated item locations for the RM analysis	122
3.15(a)	Simulated and estimated person locations based on the RM for the interaction simulation study	123
3.15(b)	Simulated and estimated person locations based on EFRM for the interaction simulation study	123
4.1	Schematic representation of the vertical equating design	133
4.2	ICC for link Item R03519/R03706	134
4.3	ICC for link Item R03524/R03711	134
4.4	ICC for link Item R03523/R03710	135
4.5	ICC for link Item R03526/R03713	135
4.6	Scatterplot of common item locations derived from separate analyses of year group data after equating transformation	138
4.7	Schematic representation of the WALNA 2003 mathematics equating design	142
4.8(a)	RM ICC for Item N00327	152
4.8(b)	EFRM ICC for Item N00327	152
4.9(a)	RM ICC for Item N03321	153
4.9(b)	EFRM ICC for Item N03321	153
4.10(a)	RM ICC for Item N00316	153
4.10(b)	EFRM ICC for Item N00316	154
4.11(a)	RM ICC for Item N03324	154
4.11(b)	EFRM ICC for Item N03324	154
4.12(a)	RM ICC for Item N00313	155
4.12(b)	EFRM ICC for Item N00313	155
4.13(a)	RM ICC for Item N00320	156
4.13(b)	EFRM ICC for Item N00320	156

---

## LIST OF TABLES

Table	Title	Page
3.1	Summary of the PGD simulation study	84
3.2	Standard deviations of item locations according to group	85
3.3	Generated and recovered means and standard deviations for person groups based on standard analysis of combined data for the four groups	87
3.4	Log ratios of the standard deviations of item estimates	87
3.5	Simulated and estimated PGDs	88
3.6	Simulated versus estimated item locations for the PGD study using the EFRM	93
3.7	Global Chi-square test of fit for the RM and EFRM, PGD study	97
3.8	Group fit residuals for the PGD simulation study	99
3.9(a)	Design of the ISD study	101
3.9(b)	Simulated item locations within each of the four sets	102
3.10	Summary results of separate analyses of data for the item sets	103
3.11	Simulated and estimated ISDs	103
3.12	Global Chi-square test of fit for the RM and EFRM, ISD study	111
3.13	Comparison of mean item fit residuals, ISD study	113
3.14(a)	Design of the ISD-PGD interaction simulation study	115
3.14(b)	Two-way table of simulated ISDs, PGDs, and frame discriminations	116
3.15	Ratios of standard deviations of ability estimates in natural units	117
3.16	Simulated and estimated ISDs for the interaction simulation study	118
3.17	Ratios of standard deviations of item estimates in natural units	118
3.18	Simulated and estimated PGDs for the interaction simulation study	119
3.19	Estimated frame discriminations for the interaction simulation study	119
3.20	Standard deviations of ability estimates for person groups derived from standard analysis of each SFR	124
3.21	Global Chi-square test of fit for the RM and EFRM, interaction simulation study	125
3.22	Summary of fit residuals for the EFRM and RM, interaction simulation study	126
4.1	WALNA 2003 Reading common Year 5-Year 7 items	136

4.2	Fit residuals by group for the PGD empirical study	140
4.3	ISD estimates for the Year 3 Mathematics 2003 Assessment	145
4.4	ISD estimates for the Year 3 Mathematics 2000 Assessment	145
4.5	Item and ISD estimates for the 2003 and 2000 Assessments	147
4.6	Comparison of abilities derived from the RM and EFRM for the ISD empirical investigation	148
4.7	Comparison of Benchmark results based on each of the analyses	149
4.8	2003 Assessment, comparison of fit residuals from the analyses	150
4.9	2000 Assessment, comparison of fit residuals from the analyses	151

---

## LIST OF ACRONYMS

2PLM	Two Parameter Logistic Model
CI	Class Interval
CML	Conditional Maximum Likelihood
DIF	Differential Item Functioning
EFR	Extended Frame of Reference
EFRM	Extended Frame of Reference Model
ICC	Item Characteristic Curve
ISD	Item Set Discrimination
JML	Joint Maximum Likelihood
LCJ	Law of Comparative Judgment
ML	Maximum Likelihood
MP	Measurement Poisson
OPLM	One Parameter Logistic Model
PGD	Person Group Discrimination
RM	Rasch Model for dichotomous data
RMSD	Root Mean Squared Difference
RMSE	Root Mean Squared Error
SFR	Specified Frame of Reference
SI	Système International d'Unités
WLE	Weighted Likelihood Estimate