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

LIST OF FIGURES

LIST OF TABLES

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