

**School of Psychology and Speech Pathology**

**Rasch Analysis of the Parental Reflective Functioning Questionnaire:  
A Critical Examination of Data from a Non-Clinical Sample of Mothers  
and Fathers with a One-Year-Old Child**

**Dawson Campbell Cooke**

**This thesis is presented for the Degree of  
Doctor of Philosophy  
of  
Curtin University**

**July 2015**



Declaration

To the best of my knowledge and belief this thesis contains no material previously published by any other person except where due acknowledgment has been made.

This thesis contains no material which has been accepted for the award of any other degree or diploma in any university.

Signature: .....

Date: .....

### Acknowledgements

This project was only possible with the support and involvement of a number of people who I would like to acknowledge and thank.

Firstly, my thanks to my three supervisors, Dr Lynn Priddis, Dr Garth Kendall and Professor Robert Cavanagh, who have each contributed in unique ways and been wonderfully supportive throughout this journey. I'm particularly grateful for their gracious way of challenging me to learn new skills and explore new topics. In addition, my thanks to Dr Ida Marais who was supportive in my learning and research with the Rasch model.

My involvement with the Peel Child Health Study made this project so much more meaningful and interesting than I could have predicted. I am thankful for the efforts of the Peel Study families who gave of their time to participate in interviews and complete questionnaires.

Finally and most importantly, I extend my sincere thanks and deep appreciation for the support of my friends and family – in particular my wife Sharon, our three children, Amy, Harry and Matilda, and my parents.

## Table of Contents

<b>Acronyms .....</b>	<b>1</b>
<b>Abstract .....</b>	<b>3</b>
<b>Chapter 1 Introduction .....</b>	<b>3</b>
<b>Chapter 2 Part A: Background Literature Review .....</b>	<b>7</b>
Early Child Development.....	8
The Father’s Influence on Child Development .....	10
Models and approaches to researching father involvement .....	11
From fathering to parenting .....	13
The quality of the father-child relationship.....	14
Attachment Theory.....	15
Child-father attachment.....	16
Comparisons of assessments of attachment with mothers and fathers .....	18
Alternative assessments of father-child relationships .....	22
Reflective Functioning, Mentalizing and Related Terms .....	25
Application of mentalizing.....	27
The development of mentalizing capacity .....	31
Parental reflective functioning .....	33
Measurement of Reflective Functioning .....	37
Parental Reflective Functioning Questionnaire (PRFQ).....	39
Summary .....	42
<b>Chapter 2 Part B: Measurement in Social Science Literature Review .....</b>	<b>44</b>
Overview of Measurement in Social Sciences.....	44
Assumptions and requirements for measurement .....	47
Likert’s assumptions .....	48
Thurstone’s requirements.....	50
Guttman scaling .....	51
Rasch model as a probabilistic variant of Guttman scaling .....	55
Two approaches to modern measurement.....	58
Classical test theory.....	59
Modern test theory .....	61
Distinctive Qualities of Rasch Measurement Theory .....	64

iv Rasch Analysis of the PRFQ

Error of measurement .....	64
Reliability .....	65
Missing data .....	66
Test equating .....	67
Sample dependence and scale dependence .....	69
Meaningfulness of measurement.....	70
Interval scale measurement .....	73
Primacy of the model .....	74
Threshold order .....	75
Invariance .....	77
Summed score sufficiency .....	78
Rasch Measurement Literature.....	79
Analysis of Data using the Rasch Model .....	79
<b>Chapter 3 Research Methodology.....</b>	<b>81</b>
Participants .....	81
Apparatus and Materials.....	82
Demographics, parental mental health, family functioning and child development questionnaires .....	82
Parental self-report scales.....	84
Parent Development Interview Revised Short Version (PDI) .....	88
Procedure.....	89
Data Analysis .....	89
Research Question 1: Do data from mothers and fathers PRFQ conform to the requirements of the Rasch Measurement Model? .....	90
i. Parameterisation and choice of Rasch model.....	91
ii Sample size and category frequencies .....	92
a. Ordering of response category thresholds .....	92
b. Targeting and distribution of items and persons .....	94
c. Overall fit to the Rasch model.....	95
d. Reliability .....	96
e. Item and person fit to the Rasch model .....	96
f. Differential item functioning .....	97
g. Local item dependence.....	98

h. Dimensionality .....	99
i. Differences between groups of persons .....	100
Iterative analysis process and removal of items.....	101
Research Question 2: Is there an association between self-report of parental depression or anxiety and PRFQ scores? .....	103
Research Question 3: Do one or more specific sets of PRFQ items show temporal (test–retest) stability? .....	103
Research Question 4: Is there a relation between PRFQ scores and PDI-RF scores that demonstrates convergent validity? .....	104
<b>Chapter 4 Results .....</b>	<b>105</b>
Response Rates and Participant Characteristics.....	105
Research Question 1: Do data from mothers and fathers PRFQ conform to the requirements of the Rasch measurement model?.....	108
HL subscale analysis .....	108
i. Parameterisation and choice of Rasch model .....	109
ii. Sample size and category frequencies .....	109
a) Order of response category thresholds.....	110
b) Targeting and distribution of items and persons.....	117
c) Overall fit to the Rasch model .....	119
d) Reliability.....	119
e) Item and person fit to the Rasch model.....	119
f) Differential item functioning .....	121
g) Local dependence .....	122
h) Dimensionality .....	123
Iterative analysis process and removal of items.....	125
HL child-focused subscale analysis .....	128
b) Targeting of the PRFQ-CF .....	129
c) Overall fit of the PRFQ-CF.....	130
d) Reliability of the PRFQ-CF .....	130
e) Individual item and person fit of the PRFQ-CF.....	130
f) DIF with the PRFQ-CF subscale.....	132
g) Local dependence and h) dimensionality.....	134
i) Variance in PRFQ-CF scores accounted for by categorical parental	

characteristics .....	134
HL self-focused subscale analysis.....	137
LH subscale analysis .....	140
M subscale analysis .....	143
Summary of Rasch analyses .....	146
Research Question 2: Is there an association between self-report of parental depression or anxiety and PRFQ scores? .....	146
Research Question 3: Do one or more specific sets of PRFQ items show temporal (test–retest) stability? .....	148
Research Question 4: Is there a relation between PRFQ scores and PDI-RF scores that demonstrates convergent validity? .....	149
Results Summary .....	149
<b>Chapter 5 Discussion.....</b>	<b>152</b>
Rasch Analysis of the PRFQ .....	153
HL subscale analysis .....	153
HL threshold disorder.....	154
Self-other dimensions of parental interest in mental states.....	156
PRFQ-CF analysis .....	158
Targeting .....	158
Fit to the Rasch model.....	159
Reliability .....	159
DIF .....	160
Local item dependence and dimensionality .....	161
Variance accounted for by categorical parental characteristics .....	161
PRFQ-SF analysis .....	163
LH subscale analysis .....	164
M subscale analysis .....	164
PRFQ-CF Associations with Parental Depression or Anxiety .....	165
PRFQ-CF Test-Retest Stability .....	166
Convergent Validity with the PDI-RF .....	168
Limitations .....	170
<b>Chapter 6 Conclusions .....</b>	<b>173</b>
<b>References .....</b>	<b>178</b>
<b>Appendices .....</b>	<b>210</b>

Appendix A – PRFQ .....	210
Appendix B - Peel Child Health Information Sheet and Consent Forms .....	213
Appendix C – BDI-II .....	225
Appendix D – STAI .....	227
Appendix E – PDI .....	229
Appendix F – Participation Information Sheet .....	236
Appendix G – Consent Form .....	237
Appendix H – GFAD Rasch Journal Article.....	238

### Table of Figures and Tables

<i>Figure 2.1.</i> Heuristic model of the dynamic of father influence on children over time (from Cabrera et al., 2007, p. 186) .....	12
<i>Figure 2.2.</i> A data-based model of the family attachment network, Figure from Bretherton (2010, p.15) – based on meta-analyses by van IJzendoorn (1995), De Wolff and van IJzendoorn (1997) and van IJzendoorn & de Wolff (1997).....	19
<i>Figure 2.3.</i> Proposed diagram showing dual attachment systems of security and exploration (from Richard Bowlby's personal lecture notes, Newland & Coyl, 2010, p. 27) .....	23
<i>Figure 2.4.</i> Model of the relationship between failures of mentalizing and possible consequences (from Luyten, 2014, p. 2) .....	29
<i>Figure 2.5.</i> Biobehavioural switch model of the relation between stress and controlled or automatic mentalization (from Fonagy & Luyten, 2009, p. 1367). .....	32
<i>Figure 2.6.</i> Typical Likert Scale categories as presented to the respondent.....	50
<i>Figure 2.7.</i> Typical Likert scale categories from a measurement perspective.....	50
<i>Figure 2.8.</i> Guttman response probability for a single item with a dichotomous response.....	54
<i>Figure 2.9.</i> Guttman response probability for four items. ....	54
<i>Figure 2.10.</i> Item Characteristic Curve for affirming a single item with a dichotomous response. ....	55
<i>Figure 2.11.</i> Item Characteristic Curve for a correct and incorrect response to a single item. ....	56
<i>Figure 2.12.</i> Item Characteristic Curve for four items.....	56

<i>Figure 2.13.</i> Category Probability Curve for a five-category rating scale item.....	58
<i>Figure 2.14.</i> Linear equating of tests simulation (from Embretson, 1996, p. 344)....	68
<i>Figure 2.15.</i> Common Person Linking Plot (from Bond & Fox, 2007, p. 87).....	69
<i>Figure 2.16.</i> Example of an item map showing persons and items calibrated onto a common scale.....	70
<i>Figure 2.17.</i> Embretson’s universal system for validity (from Embretson, 2007, p. 453).....	72
<i>Figure 2.18.</i> Relationship between raw scores and measures .....	74
<i>Figure 2.19.</i> Example of a category probability curve of a five-category item.....	75
<i>Figure 3.1.</i> Category probability curve showing disordered threshold of a four- category item. ....	94
<i>Figure 3.2.</i> Category probability curve showing ordered thresholds after rescoring first two categories as 0. ....	94
<i>Figure 3.3.</i> Example of an Item Characteristic Curve showing an under discriminating item.....	97
<i>Figure 3.4.</i> Example of an Item Characteristic Curve displaying DIF. ....	98
<i>Figure 3.5.</i> The iterative process of Rasch analysis.....	102
<i>Figure 3.6.</i> Example of a Common Person Linking Plot.....	104
<i>Figure 4.1.</i> Category Probability Curves for HL Item 11 .....	111
<i>Figure 4.2.</i> Category Probability Curves for HL Items (except item 11) .....	112
<i>Figure 4.3.</i> Threshold Map for 17-item HL scale .....	116
<i>Figure 4.4.</i> Person-Item Threshold Distribution of 17-item HL scale.....	118
<i>Figure 4.5.</i> Item Map of 17-item HL scale .....	118
<i>Figure 4.6.</i> Item Characteristic Curve for HL item 13.....	120
<i>Figure 4.7.</i> Item Characteristic Curve for HL item 24.....	121
<i>Figure 4.8.</i> ICC showing differential item functioning for Gender of Parent with HL Item 1.....	122
<i>Figure 4.9.</i> ICC for item 6 .....	123
<i>Figure 4.10.</i> ICC for HL item 13 showing differential item functioning between mothers and fathers .....	126
<i>Figure 4.11.</i> ICC for HL item 30 .....	127
<i>Figure 4.12.</i> Person-Item Threshold Distribution of the PRFQ-CF subscale .....	129
<i>Figure 4.13.</i> Item Map of the PRFQ-CF subscale .....	130

<i>Figure 4.14.</i> Item Characteristic Curve for the PRFQ-CF subscale item 6 .....	132
<i>Figure 4.15.</i> Item Characteristic Curve for the PRFQ-CF subscale item 20 .....	132
<i>Figure 4.16.</i> ICC for item 31 of the PRFQ-CF subscale showing DIF with parent's first child or not .....	133
<i>Figure 4.17.</i> ICC for item 7 of the PRFQ-CF subscale showing DIF with parent's education .....	134
<i>Figure 4.18.</i> Distribution of mother and father locations of the PRFQ-CF Subscale .....	135
<i>Figure 4.19.</i> Plot of mother and father PRFQ-CF scores .....	135
<i>Figure 4.20.</i> Distribution of locations showing levels of parent's education (without extreme scores) of the PRFQ-CF Subscale .....	136
<i>Figure 4.21.</i> Distribution of locations showing levels of parent's education (with extreme scores) of the PRFQ-CF Subscale .....	137
<i>Figure 4.22.</i> Person-Item Threshold Distribution of the PRFQ-SF subscale .....	138
<i>Figure 4.23.</i> Distribution of mother and father locations of the PRFQ-SF Subscale .....	139
<i>Figure 4.24.</i> Distribution of locations showing levels of parent's education of the PRFQ-SF Subscale .....	139
<i>Figure 4.25.</i> Threshold Map for 14-item LH Subscale .....	141
<i>Figure 4.26.</i> Distribution of mother and father locations, and item thresholds for 9-item LH Subscale .....	143
<i>Figure 4.27.</i> Distribution of locations by level of education for 9-item LH Subscale .....	143
<i>Figure 4.28.</i> Distribution of mother and father locations, and item thresholds for M Subscale .....	145
<i>Figure 4.29.</i> Invariance Plot of PRFQ-CF Test and Retest .....	149
Table 4.1. <i>Comparison of Participants and Non-participants Characteristics</i> .....	106
Table 4.2. <i>Australian Bureau of Statistics Comparative Population Demographics</i> .....	108
Table 4.3. <i>Item Category Frequencies for 17-Item HL Scale</i> .....	110
Table 4.4. <i>Item Thresholds for 17-Item HL Scale</i> .....	111
Table 4.5. <i>Category Scoring for 17 HL Items</i> .....	117

Table 4.6. <i>Item-Person Interaction Statistics for 17-item HL Scale</i> .....	119
Table 4.7. <i>Individual Item-Fit Statistics for 17-item HL Scale</i> .....	120
Table 4.8. <i>Person-item Residual Correlation Matrix for 17-item HL Scale (from RUMM203)</i> .....	122
Table 4.9. <i>HL Items PC loadings (from RUMM203)</i> .....	123
Table 4.10. <i>HL Items Principle Components Summary</i> .....	124
Table 4.11. <i>Two Groups of Items from 15-item HL Scale Based on PCA</i> .....	125
Table 4.12. <i>HL Item Residual Correlation Matrix</i> .....	127
Table 4.13. <i>PC Loadings of 15-item HL Scale</i> .....	128
Table 4.14. <i>Principal Components Summary for 15-item HL Scale</i> .....	128
Table 4.15. <i>Item-Person Interaction Statistics for PRFQ-CF Subscale</i> .....	130
Table 4.16. <i>Individual Item-Fit Statistics for the PRFQ-CF Subscale</i> .....	131
Table 4.17. <i>Item-Person Interaction Statistics for PRFQ-SF Subscale</i> .....	137
Table 4.18. <i>Individual Item-Fit Statistics for the PRFQ-SF Subscale</i> .....	138
Table 4.19. <i>Category Frequencies for LH Subscale</i> .....	140
Table 4.20. <i>Category Scoring for 14-item LH Subscale</i> .....	141
Table 4.21. <i>Individual Item-Fit Statistics for 14-item LH Subscale</i> .....	142
Table 4.22. <i>Item-person Interaction Statistics for 14-item LH Subscale</i> .....	142
Table 4.23. <i>Category Frequencies for Centre Scored M Subscale</i> .....	144
Table 4.24. <i>Item-person Interaction Statistics for M Subscale</i> .....	144
Table 4.25. <i>Individual Item-Fit Statistics for M Subscale</i> .....	145
Table 4.26. <i>Matrix of Correlations Between Father PRFQ-CF and Father Mental Health</i> .....	147
Table 4.27. <i>Matrix of Correlations Between Mother PRFQ-CF and Mother Mental Health</i> .....	148
Table 4.28. <i>Summary Statistics for Rasch Analysis of PRFQ Subscales</i> .....	151
Table 5.1. <i>Seven Items of the PRFQ-CF</i> .....	156

### Acronyms

AAI	Adult Attachment Interview
ASQ	Attachment Behaviour Q-Set
BDI	Beck Depression Inventory
CTT	Traditional or Classical Test Theory
DIF	Differential Item Familiarity or Functioning
FAD	McMaster Family Assessment Device
GFAD	General Functioning Subscale of the McMaster Family Assessment Device
GRM	Graded Response Model
ICC	Item Characteristic Curve
IRT	Item Response Theory
MMM	Maternal Mind-Mindedness
PCHS	Peel Child Health Study
PCA	Principal Components Analysis
PDI	Parent Development Interview Revised Short Version
PDI-RF	Reflective Functioning scored from the PDI
PRF	Parental Reflective Functioning
PRFQ	Parental Reflective Functioning Questionnaire
PRFQ	
Subscales:	<p>HL (PRFQ high-low scored subscale) Items worded such that higher scores reflect higher levels of parental mentalizing.</p> <p>LH (PRFQ low-high scored subscale) Items were worded in a way that lower scores reflect higher levels of parental mentalizing.</p> <p>M (PRFQ middle scored subscale) Items were worded such that a response in the middle of the rating scale reflects a high level of parental mentalizing, and lower levels by responses towards either extreme of the scale.</p>

## 2 Rasch Analysis of the PRFQ

PRFQ-CF (Child focused mentalizing subscale)

PRFQ-SF (Self focused mentalizing subscale)

PSI Person Separation Index

RF Reflective Functioning

RM Rasch Model

RMT Rasch measurement theory

SCIP Sensitive and Challenging Interactive Play

SE Standard Error of Measurement

SSP Strange Situation Procedure

STAI State Trait Anxiety Inventory

ToM Theory of Mind

### Abstract

Research on the care of infants and their development has been predominantly with mothers and more research is needed on the nature and quality of the father-infant relationship. The focus of this cross-sectional research is the capacity of paternal mentalizing or fathers' Parental Reflective Functioning (PRF): the parental capacity to reason about one's own and one's child's behaviours by taking into consideration intentional mental states. Rasch measurement analysis and theory was used to critically examine the validity of the recently developed Parental Reflective Functioning Questionnaire (PRFQ) for both mothers and fathers. Mothers' and Fathers' levels of PRF were assessed using scores based on the Parent Development Interview (PDI-RF) and self-report ratings from the 39-item PRFQ, in the context of the longitudinal Peel Child Health Study ( $N = 120$  families). Data in this study from seven items of the PRFQ conformed to the requirements of Rasch measurement theory and this set of items is proposed to reflect a specific aspect of mentalizing: child-focused parental mentalizing (PRFQ-CF). Properties of the PRFQ-CF are described in detail and the shortcomings of other PRFQ subscales are identified. The PRFQ-CF showed marginal test-retest stability and limited convergent validity with the PDI-RF; these limitations are discussed. Parental ratings of mental health were not associated with scores on the PRFQ-CF and variance in the PRFQ-CF scores was not accounted for by demographic variables including parent age, gender of child, parent birthplace, birth order, parent education, and parent occupation. Mother and father PRFQ-CF scores were found to be unrelated ( $r = -.001$ ), and mothers scored on average higher PRFQ-CF than fathers. The measurement of fathers' PRF has the potential to advance the understanding of father-child relationships in early childhood. This aspect of the father-child relationship is fundamental for a better appreciation of the father's role in the family and for the progress of 'family-friendly' government policies and interventions that target the specific needs of fathers and children.

## Chapter 1 Introduction

This thesis provides a critical examination of the qualities of the Parental Reflective Functioning Questionnaire (PRFQ) using the Rasch model. The PRFQ was recently developed (Luyten et al., 2009) to be a self-report of *parental mentalizing*, also referred to as *Parental Reflective Functioning*. Mentalizing is the process of considering mental states and how they influence behaviour both for one's self and for others. The assessment of parental mentalizing has to date only been undertaken by the analysis of a transcribed clinical interview (Slade, Bernbach, Grienenberger, Levy, & Locker, 2005). Furthermore, parental mentalizing has mostly been studied in the context of the mother-infant relationship and rarely with the fathers and their infants or with consideration of gender differences.

Supportive parent-child relationships have been recognised as having a critical influence on the child's emotional and social regulatory capacity, and in developing resilience against the harmful effects of stress (Shonkoff, 2010). Although most parenting research has been with mothers, the role of fathers in children's development has been found to be uniquely influential and independent of the mother's role (Lamb, 2010a). One aspect of the father's role that has been identified as requiring further investigation is the quality of the father-infant relationship. Attachment theory has been a dominant framework within which to study the quality of mother-child relationship and the foundations of children's social development (Cassidy, 2008). More recently, broader conceptualisations of a child's psychological security with parents have been developed to be inclusive of behaviours more typical in the father-child relationship (K. Grossmann, Grossmann, Kindler, & Zimmermann, 2008). Parental mentalizing may be a common factor in mothers' and fathers' capacity to offer their child a secure parent-child relationship. The PRFQ could provide a valuable source of evidence in the investigation of the foundational capacities of healthy parent-infant relationships.

Mentalizing has emerged as a construct of interest from the study of attachment relationships and more broadly as an aspect of social cognition (Fonagy, Bateman, & Luyten, 2012). It overlaps with a number of other related concepts, such as theory of mind, empathy and alexithymia, which assess different facets of mentalizing. Research into mentalizing and parental mentalizing in particular has been mostly limited to small clinical samples of mothers and identified with

observational methods or ratings based on interview transcripts. A self-report measure of mentalizing would enable collection of data from large samples of parents, which in turn could provide norms for levels of mentalizing and the opportunity to identify causal relationships with other parent and child characteristics.

Rasch measurement theory provides an approach to evaluating rating scales against a set of requirements for achieving measurement (Andrich, 1985). The methods of Rasch analysis have fundamental differences to the methods of classical test theory, which is the dominant paradigm for social science measurement. This thesis and the associated analyses demonstrate the utility of using Rasch measurement theory in the critique and development of rating scales. Establishing a scale's suitability as an instrument of measurement is necessary before inferences can be made from data. The objective of this thesis is to apply Rasch measurement theory to critically examine the validity of the recently constructed PRFQ as a measure of parental reflective functioning for both mothers and fathers.

The Literature Review chapter provides a rationale for the examination of parental reflective functioning and its assessment. In particular, the argument is developed that assessment of a father's capacity to appreciate his own and his infant's mental states is important in gaining a better understanding of father-child relationships and children's social-emotional development. A central aspect of this thesis is the use of the Rasch analyses to examine the qualities of the PRFQ. Since the use of Rasch analysis is underutilized in the field of psychometrics, a relatively thorough review of Rasch measurement theory is given in Part B of the Literature Review chapter. This review of the Rasch model describes how it fits within the broader field of social science measurement, and the distinctive qualities of the model.

The Methods chapter provides a description of the participants in the study, the instruments used in the collection of data. Particular strengths of this study are that the sample is generally representative of parents in Australia and that comparable data were collected from both mothers and fathers. Furthermore, resource intensive interviews (Parent Development Interview) and specialised coding of transcripts for reflective functioning (PDI-RF) were undertaken with 40 couples, which makes this project one of the largest studies of PDI-RF and possibly the only

## 6 Rasch Analysis of the PRFQ

PDI-RF study to have matched mother and father data. The analyses described in the Methods section address the following research questions:

1. Do data from mothers and fathers PRFQ conform to the requirements of the Rasch measurement model?

In particular, do data from one or more specific sets of PRFQ items show evidence of:

- a. ordered response category thresholds,
  - b. targeted item/person distributions,
  - c. overall scale fit to the Rasch model,
  - d. reliability as indicated by the Person Separation Index,
  - e. individual item and person fit to the Rasch model,
  - f. invariance in individual item functioning between persons with membership of different groups,
  - g. local item independence,
  - h. unidimensionality, and
  - i. variance in PRFQ scores accounted for by categorical parental characteristics.
2. Is there an association between self-report parental depression or anxiety and PRFQ scores?
  3. Do one or more specific sets of PRFQ items show temporal (test-retest) stability?
  4. Is there a relation between PRFQ scores and PDI-RF scores that demonstrates convergent validity?

## Chapter 2 Part A: Background Literature Review

This section of the literature review will provide an overview of early child development and parental influence on psychosocial functioning, fatherhood and father-infant relationships, attachment theory, parental mentalizing, and assessment of parental reflective functioning. This material is provided to place the thesis in its psychological context of evidence based developmental theory.

Firstly a summary is provided of current literature that outlines how the early months and years of human life provide the experiences that lay a foundation for future capacity and development. This section highlights recent research in developmental epidemiology and developmental psychology that underscores the complexity of the interacting factors which provide an optimal environment for the growth of a healthy individual. Conclusions are drawn regarding the importance of early social-emotional development and the influence of parent-child relationships.

The second area of review introduces the topic of fatherhood. Attention is given to recent evidence-based literature that identifies the particular impact the father-infant relationship has on child development and how this compares with the mother-infant relationship.

An overview of how the quality of a parent-infant attachment relationship influences a child's social-emotional development and future psychosocial outcomes is provided. Emphasis will be given to recent research and theory describing the father-infant attachment relationship and its implications for children.

Fourthly, relevant research and theories are presented showing how parental mentalizing is helpful in the understanding of parent-infant relationships. The specific area of interest to this thesis is then addressed in detail, namely, the mentalizing of fathers and the relevance of this for child development.

Finally, the fifth area of review is the assessment of parental reflective functioning. Literature on existing and potential instruments for assessing reflective functioning is critiqued and the challenges of measuring reflective functioning are highlighted. In particular, the literature regarding the Parent Development Interview and the Parental Reflective Functioning Questionnaire is examined with regards to their application with non-clinical parents and fathers in particular.

### **Early Child Development**

The factors that contribute to a child's healthy development are a complex interaction of parental and child characteristics, the nature of the relationship that develops between them, and the many contextual influences in which a family lives (Bronfenbrenner & Morris, 1998; Collins, Maccoby, Steinberg, Hetherington, & Bornstein, 2000; Shonkoff, Boyce, & McEwen, 2009). A consensus has now been reached that developmental systems theory is most appropriate for guiding research, practice and policy in this field (Belsky, 2010), and those in research and practice consider a wide range of mediating and moderating variables in order to better understand the environment that impacts upon the developing child, such as psychological, social, cultural, educational, physical and economic factors (Davies, 2004). A number of scientific advances in the brain sciences are improving our understanding of these complex developmental processes. This section of the literature review will briefly introduce some of the most recent and influential advances in the understanding of early child development, which integrate the biological, psychological, and social sciences.

The National Scientific Council on the Developing Child (2013) has played a prominent role in synthesizing and raising the awareness of child development science and evidence. Jack Shonkoff, Megan Gunnar and the other council members, who are leaders in a variety of scientific fields, have produced persuasive reports on the cumulating evidence linking early childhood experiences to later health and well-being. Progress in the biological science areas of brain development, the endocrine stress response system and the emerging field of epigenetics have been central in revealing the mechanisms by which early experiences influence developmental outcomes. Other leading researchers have also made informative investigations into these processes, such as parental care mediating the effects of adversity on neural development (Meaney, 2001) and gene expression (Meaney, 2010), changes in the developing brain through the hypothalamo-pituitary-adrenal (HPA) axis activity (McEwen, 2003; Repetti, Taylor, & Seeman, 2002), exposure to stress in childhood associated with DNA methylation in adolescence (Essex et al., 2013), DNA methylation and changes in human development (van IJzendoorn, Bakermans-

Kranenburg, & Ebstein, 2011), and the embedding of early life experiences in the genome (Szyf & Bick, 2013).

The social determinants of health have been identified as being of critical importance during a child's vulnerable early years of development (Maggi, Irwin, Siddiqi, & Hertzman, 2010). These social influences range from individual interactions with parents and other caregivers to broader social environments of the neighbourhood or community, which are all set within a socio-political context (Irwin, Siddiqi, & Hertzman, 2007). These first years of life are a time during which vulnerability to health risks can be biologically embedded via sensitive periods of neurological, endocrine, immune, and metabolic system development. Biological embedding occurs when early experiences impact health and developmental outcomes over the whole life course. This process includes the influence of experiences on neural sculpting, gene expression, epigenetics, and the establishment of the stress response, "allostasis" (Cynader & Frost, 1999; S. E. Fox, Levitt, & Nelson, 2010; McCain, Mustard, & Shanker, 2007; Shonkoff et al., 2009; Ting-Fang, 2007). Cynader and Frost (1999) highlight the importance of social experiences and their biological pathways to future social competence; they concluded that "... inadequate and inappropriate social and emotional experiences in the early environment could... result in compromised higher level neural systems whose task is to provide information necessary to bond, imitate, and generally respond in socially appropriate ways" (p. 183).

Although there are many aspects of early child development that are important for life long health and well-being, the development of emotional and social regulatory capacity is crucial because of its profound influence on the development of various other new skills and abilities throughout childhood (Shonkoff et al., 2009; Shonkoff & Phillips, 2000). An obvious and influential factor that impacts on child development is the nature of the parent-child relationship that they jointly create with social-emotional interactions. This influential role of parents is inherent in two core concepts of development noted in the highly regarded report on early child development by the National Academy of Sciences (From Neurons to Neighborhoods: Shonkoff & Phillips, 2000), and based on a wide range of unequivocal research findings. One of these concepts focuses on the powerful influence of culture on every aspect of human development, which is most clearly expressed in childrearing beliefs and practices (Kendall & Li, 2005). Another core

## 10 Rasch Analysis of the PRFQ

concept highlights human relationships as the building blocks of healthy development and how relationships affect other relationships. The most influential of relationships in a child's first years of life are within the immediate family, and when this environment is not sensitive, warm and responsive, the associated risks for poor mental and physical health outcomes are immediate, long term and life-long (Repetti et al., 2002).

The influence of supportive adult care has been identified as an important aspect of understanding children's resilience and vulnerability through their different experiences of stress, which can be categorized as positive, tolerable or toxic (National Scientific Council on the Developing Child, 2007; Shonkoff, 2010). A defining aspect of both positive stress and tolerable stress is that both are typically experienced within the context of a stable and supportive relationship with an adult, who assists the child to recover from elevated levels of physiological stress response. Toxic stress occurs when stress is experienced in the absence of a supportive relationship and with strong and sustained or repeated stimulation of the physiological stress response system. Understood from this perspective, the role of the supportive parental relationship is fundamental to mitigating toxic stress for a child, and thereby protecting a child from the associated long term psychological and physical health problems. Based on the science of human development from multidisciplinary sources, Shonkoff et al. (2012) make the weighty statement that "many adult diseases should be viewed as developmental disorders that begin early in life and that persistent health disparities associated with poverty, discrimination, or maltreatment could be reduced by the alleviation of toxic stress in childhood" (p. e232).

In summary, increasing empirical evidence from the domains of biology, neuropsychology and developmental psychology supports the central role of parental influence on child development, and an important aspect of this role is to provide a nurturing and protective relationship. The following sections will review literature on father-child relationships and examine the characteristics of the parental relationship that are most beneficial for healthy child development.

### **The Father's Influence on Child Development**

The study of fatherhood is a growing area of parent-child research that has changed markedly over the past 40 years (Lamb, 2000; Lamb, Pleck, Charnov, &

Levine, 1985; Pleck, 2010), particularly the influence of fathers on child development (Lamb, 2010a). Unlike the relatively stable and mainly biologically defined role of a mother, the definition of fatherhood can fluctuate widely and is largely dependent on cultural and societal influences (Blankenhorn, 1995; Lupton & Barclay, 1997; Park & Brott, 1999). Reflecting this variation, the demographic and socio-economic circumstances, and the structural position of fathers within families are remarkably and increasingly diverse (Hofferth, 2002). In addition to changing over time and being culturally specific, the role of fatherhood changes for individual fathers as their circumstances alter and their child develops.

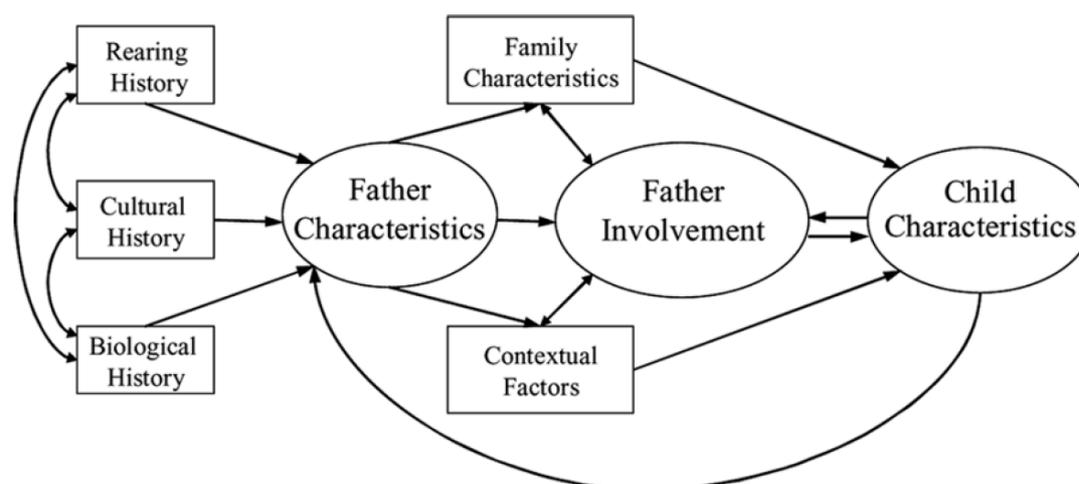
A number of comprehensive reviews of evidence regarding father involvement have been produced over recent years, including evidence of the effects of father involvement by the Canadian *Father Involvement Research Alliance* (S. M. Allen & Daly, 2007), report on father's involvement as a determinant of child health (Ball, Moselle, & Pedersen, 2007) for the Public Health Agency of Canada, evidence of costs and benefits of active fatherhood to inform the development of policy and practice produced by *Fathers Direct* in the UK (Burgess, 2007), a *World Health Organisation* report of Fatherhood and Health Outcomes in Europe (World Health Organisation, 2007), a *US Department of Health and Human Services* report on the importance of fathers in the healthy development of children (Rosenberg & Wilcox, 2006), and the comprehensive *The Role of the Father in Child Development* now in its 5<sup>th</sup> edition (Lamb, 2010b). These reports draw attention to the growing evidence that demonstrates the important role of the father in families and in children's development; however, they also reveal that the effects of father involvement are varied and complex, and that fatherhood research often has inherent limitations and methodological challenges. Consequently, research on the influence of father involvement requires a multidimensional approach. This section will consider the various approaches to studying complexity of father involvement and current directions in fatherhood research with a focus on the fathers' influence on child development.

### **Models and approaches to researching father involvement**

Until recent times, the primary attributes of fatherhood were generally accepted to be economic provisioning and the psychosocial support of the mother (Lamb, 2000). In more current investigations of father involvement and its

## 12 Rasch Analysis of the PRFQ

influences, a number of different approaches have been utilized. The integrative model (**Error! Reference source not found.**) developed by Cabrera, Fitzgerald, Bradley and Roggman (2007) shows many potential pathways of a father's influence on their child. This model is impressively comprehensive and helpful in a number of ways: it incorporates a developmental perspective, it acknowledges bidirectional influences, it places the influence of the father within a broad ecological framework, and it recognizes the direct and indirect influences which would include both mediating and moderating variables.



*Figure 2.1.* Heuristic model of the dynamic of father influence on children over time (from Cabrera et al., 2007, p. 186)

Importantly, the Cabrera et al. (2007) model shows the indirect influence fathers have on their children through family characteristics and contextual factors. For example, the 'breadwinning' role, being a positive male role model, support provided to the mother's parenting and making responsible fertility decisions, are all valuable contributions that benefit children (Pleck, 2010). Also, the model shows that numerous factors influence child outcomes via facilitating (or impeding) the involvement of the father – these would include variables such as father work hours, paternal mental health, the couple relationship and co-parenting (Baxter & Smart, 2010).

A fundamental problem with any practical application of the Cabrera et al. (2007) model is that the central variable of father involvement is not easily defined

or agreed upon. A popular distinction has been made between three aspects of father involvement: accessibility (availability), engagement (interaction) and responsibility (Lamb, 2000; Lamb et al., 1985). This broad conceptualization of father involvement has been helpful in contextualizing the many specific assessments of father involvement used in research projects. More recently, leading fatherhood researchers have found limitations with these three dimensions and have suggested alternative approaches to studying fatherhood (Palkovitz, 2007; Pleck, 2010).

### *From fathering to parenting*

A major limitation of the three fathering dimensions is that researchers have operationalized them in different ways (Pleck, 2010). The first dimension of engagement has been commonly interchanged with general father involvement and increasingly interpreted to refer to positive engagement that is understood to foster positive child outcomes. A systematic review of 24 longitudinal studies examining the effects of father involvement on children's developmental outcomes (Sarkadi, Kristiansson, Oberklaid, & Bremberg, 2008) indicated that active and regular engagement predicts positive child outcomes. Pleck (2010) recently revised the three-dimensional model of father involvement in a number of ways based on how the original model had subsequently been conceptualized in fathering research. The revised model replaces accessibility and responsibility with the two dimensions of 'warmth and responsiveness' and 'control'. In addition to three core dimensions are two auxiliary domains of 'indirect care' and 'process responsibility'. Pleck argues for the inclusion of traditionally generic parenting dimensions by asserting that a father's parenting characteristics are more important than his maleness or anything specific to being a father. This modified conceptualization of fathering places more emphasis on the quality of the relationship between the father and child, and on integration with the family and wider social systems.

A move away from unique father specific assessment to more generic parenting assessments of fathering could be understood as neglecting all the studies showing the uniqueness of the father's contribution. A justification for this generic approach is that each parent, regardless of gender or role, makes a unique contribution to a child's development. Pleck (2007) draws attention to the ecological model of Bronfenbrenner (1979) and highlights the value of proximal caregivers with different characteristics and roles. Rather than simply being an additional

## 14 Rasch Analysis of the PRFQ

replicated parent or ‘microsystem partner’ in the model, a father has a unique contribution due to the simple fact that his attributes are different to the mother. Even though their roles may potentially be similar, the difference between a mother and father introduces complexities that stimulate the child’s social and emotional development.

Therefore, the inclusion of generic parenting assessment of fathering does not diminish the value of father specific research, or lessen the importance of examining the role of father. Considering the parenting attributes of the father continues to help broaden appreciation of the diverse characteristics of parents - both mothers and fathers (Spicer, 2007). Although mothers and fathers can perform very similar parenting roles, some differences in roles are influenced by cultural norms and socio-economic contexts. For this reason caution needs to be taken with assessment instruments designed for mothers that are later applied comparatively with fathers (Lamb & Lewis, 2004).

### *The quality of the father-child relationship*

Evaluation of the quality of the father-child relationship has received much less attention than other aspects of fathering and has therefore been less clearly understood. In particular, more research has been called for that investigates the nature and quality of father-child interaction (Palkovitz, 2002), which is not necessarily reflected in the quantity of engagement. For example, analysis of data from the Longitudinal Study of Australian Children (Baxter & Smart, 2010) found that the amount of time fathers spent in their children’s company was a poor predictor of child outcomes compared to other assessments of fathering. The assessment of the parent-child relationship has the benefit of results being comparable for both mothers and fathers, whereas the assessment of time spent engaged in child caring activities is confounded by differing parent roles and opportunities due to social norms.

An interesting challenge in the study of parent-child relationships is the relative dependence of a dyadic relationship on the family system. The quality of the father-child relationship has been suggested to be more vulnerable to the influence of family functioning quality than that of the mother-child relationship. Cummings, Goeke-Morey and Raymond (2004) present a number of arguments and sets of evidence that support the *fathering vulnerability hypothesis* (for an exception see;

Carlson, Pilkauskas, McLanahan, & Brooks-Gunn, 2011). Possible reasons for this vulnerability include primarily the more salient and defined role of motherhood for women compared to the role of fatherhood for men, and the difference in time alone with the child such that men develop the father-child relationship mostly with the mother present.

Palkovitz (2007) has suggested that the style or quality of father-child relationship is a central and overarching factor that is mirrored in the various indicators of father involvement. Based on “meta-analytic thematic analysis” and “professional and relational intuitions”, Palkovitz proposed what he considers to be the most important long-term qualities of the father-child relationship: affective climate (e.g. attachment, warmth, love), behavioural style (e.g. developmentally facilitative interactions, modelling, relational style), and relational synchrony (e.g., developmentally appropriate and sensitive, tuned in to signals, capitalising on emerging interests and abilities). Conceptualising and assessing these qualities was argued to be a more valuable and productive focus of future research into father involvement.

Attachment theory provides a widely appreciated approach to conceptualising and researching the qualities of parent-child relationships described by Palkovitz (2007). This field of study has a rich history of theoretical and clinical exploration, and it continues to develop and hold promise for elucidating the antecedents of optimum social-emotional health.

### **Attachment Theory**

The theory of attachment (Bowlby, 1969) is a prominent and influential framework for understanding social and emotional development, which has its foundations in children’s relationships with their caregivers (Cassidy, 2008). The attachment relationship has been characterised primarily by the nature of a child’s strategies in seeking proximity and therefore protection or comfort from a primary caregiver. A caregiver’s responses to a child’s appeals for comfort have been found to shape these attachment strategies, which over time can become reinforced and resistant to change. The child-parent attachment pattern developed during infancy has been shown to provide a foundation for subsequent psychosocial development over time, particularly if this pattern of care is maintained (Thompson, 2008; van IJzendoorn, 1995b).

## 16 Rasch Analysis of the PRFQ

Two assessments of attachment have dominated the psychological literature examining attachment relationships: one from observations of infant-parent attachment behaviour and the other from an interview assessing adult representations of early attachment relationships. The Strange Situation Procedure (SSP; Ainsworth, Blehar, Waters, & Wall, 1978) is generally considered the gold standard assessment of infant-parent attachment. This procedure involves a laboratory-based observation of an infant and parent, which includes two brief separations and two 3-minute reunions. Ainsworth's student, Main and her colleagues (Main, Kaplan, & Cassidy, 1985) later developed an assessment of adult attachment representations named the Adult Attachment Interview (AAI). The AAI is a semi-structured interview containing questions that elicit reflection upon memories of attachment relationships (Hesse, 2008). In contrast to the observation of nonverbal interactions in the Strange Situation (attachment styles), the transcript of the AAI is analysed with attention to coherency of the discourse and nature of collaboration with the interviewer.

Based on research undertaken using data from the longitudinal Minnesota study spanning 30 years, Sroufe (2005) concluded that early attachment security provides an “organising core in human development that is always integrated with later experience and never lost” (p. 365). This statement is supported by evidence from numerous studies indicating that children with secure attachments are more likely to have greater social-emotional competence, more developed cognitive abilities, and positive physical and mental health outcomes (Ranson & Urichuk, 2008). Also, longitudinal studies indicate early attachment patterns influence later attachment representations in childhood and adulthood, particularly when environmental and relational circumstances remain relatively stable. The processes involved in attachment changes across the life course, and the intergenerational transmission of attachment are controversial amongst attachment researchers and theorists (Fraley, 2002; van IJzendoorn, 1995b). Recent developments in understanding of parent-infant relationship are showing promise in conceptualizing these processes, and are discussed later in this chapter.

### **Child-father attachment**

Two of the earliest published studies on attachment acknowledged the significance of the infant-father attachment relationship (Ainsworth, 1967; Schaffer & Emerson, 1964). Schaffer and Emerson (1964) studied 60 infants and their

behaviour when separated from caregivers/family members (see Table 2.1). They found that even at one month of age, 27% of the infants identified the father as a joint attachment object/figure (person with whom the infant wants proximity). This percentage increased to 59% by 6 months and to 71% when the child was 18 months old. Only very few infants identified the father as the sole object at any age, and understandably the majority of infants (65%) identified the mother as the sole object at one month of age. However, this sole preference for the mother rapidly declined such that by the sixth month only 17% of the infants identified the mother as the sole object and by 18 months only 5%. Also notable was that at 18 months 22% of the principal objects had very little role in the child's physical care – indicating the role of primary child carer did not necessarily indicate a primary attachment relationship. These attachment relationships tended to fluctuate, although mostly due to changes in the infant's contact with the object.

Table 2.1

*Identity of Attachment Objects*

**Identity of Attachment Objects: Percentage of Subjects Forming Specific Attachments According to Identity of Object at Successive Age Periods**

<i>Identity of Object</i>	<i>Lunar Month Following Age at Onset in First Year</i>						<i>18-Months (CA)</i>
	<i>1st</i>	<i>2nd</i>	<i>3rd</i>	<i>4th</i>	<i>5th</i>	<i>6th</i>	
Mother (sole object) .....	65%	53%	32%	50%	47%	17%	5%
Mother (joint object) .....	30	35	54	43	50	77	76
Father (sole object) .....	3	9	7	2	0	5	4
Father (joint object) .....	27	23	42	29	44	59	71
Grandparent (sole object) .....	2	0	0	0	0	0	0
Grandparent (joint object) .....	9	12	14	12	10	29	45
Other relative (sole object) .....	0	0	0	0	0	0	2
Other relative (joint object) .....	5	5	5	14	10	18	16
Friend or neighbor (sole object) .....	0	0	2	0	0	0	0
Friend or neighbor (joint object) .....	3	7	7	9	3	12	26
Sibling (sole object) .....	0	0	0	0	0	0	2
Sibling (joint object) .....	2	5	7	7	7	12	22
Other child (sole object) .....	0	0	0	0	0	0	0
Other child (joint object) .....	3	5	14	7	3	12	14

Note: Table from Schaffer and Emerson (1964, p.31)

## 18 Rasch Analysis of the PRFQ

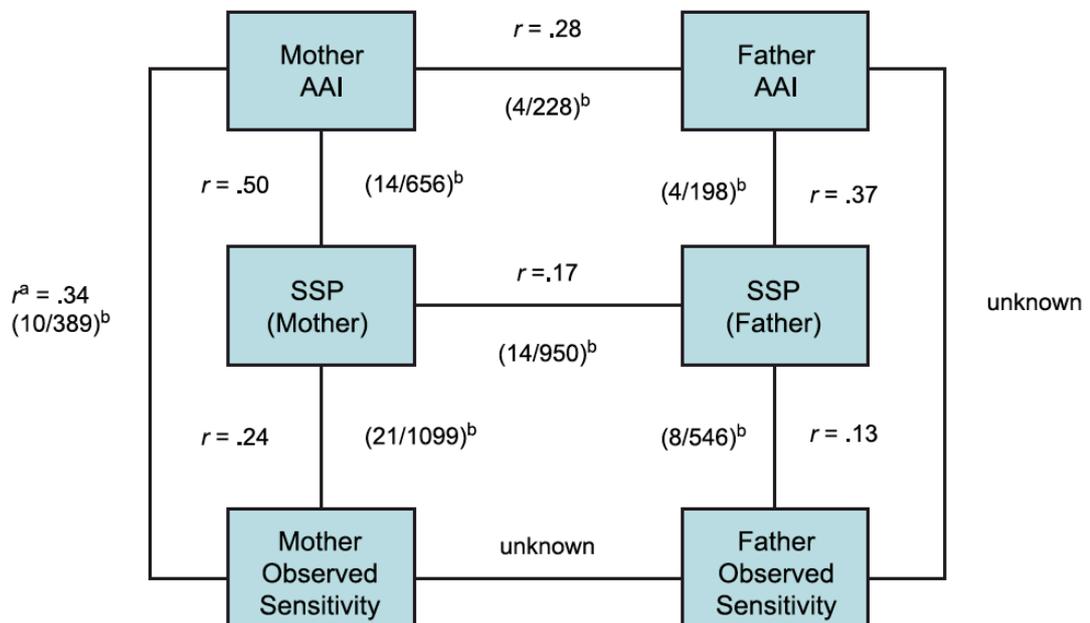
Bowlby (1969) supported the finding of Shaffer and Emerson that showed infants could develop a number of attachment relationships; however, his subsequent focus of study and theory stressed the infant's preference for a one principal attachment figure – in most cases the mother. Ainsworth (1967) also made note of the father-infant attachment, which she considered to be “disproportionate to the frequency of his interaction with the baby” (p.352). Bowlby, however, discounted this conclusion and suggested that the methods of assessment used in both these studies did not distinguish between playmates and attachment figures and that the subsidiary attachment figures may have been mostly fulfilling a playmate role rather than that of an attachment figure (Bowlby, 1969). Later in his career, Bowlby referred to both mothers and fathers as attachment figures, with fathers as a protector and mothers as a secure base (Newland & Coyl, 2010).

Most attachment research since these first studies has neglected assessment of the father-child relationship and father specific variables. Only more recently have studies begun to explore in detail and with alternative instruments the differences and similarities of infant security with mothers and fathers. The following section firstly reviews attachment research that has included both mothers and fathers using widely accepted attachment instruments such as the AAI and SSP. Following this, some examples of research will be considered that have used assessments specifically designed with the infant-father relationship in mind.

### *Comparisons of assessments of attachment with mothers and fathers*

The associations between mother and father attachment representations (AAI), infant-parent attachments (SSP) and observations of parent-infant sensitivity have been examined in three meta-analyses (De Wolff & van IJzendoorn, 1997; van IJzendoorn, 1995a; van IJzendoorn & De Wolff, 1997). The strength of the associations between scores from these instruments are presented in Figure 2.2, with the number of studies included in each analysis and the number of participants shown in parentheses. The modest association of  $r(N = 950) = .17$  between infant-mother and infant-father classifications of secure or insecure attachment indicates that although there is some commonality between each parent's attachment with their child, infant-parent attachment tends to be generally independent and relationship specific (K. E. Grossmann, Grossmann, Huber, & Wartner, 1981; Lamb, 1978; Main & Weston, 1981). Possible explanations for the associations that have been found

between child attachments with mothers and fathers include the influence of the child's temperament and similarities between the parents, such as parenting values or caregiving behaviours. (N. A. Fox, Kimmerly, & Schafer, 1991).



<sup>a</sup>86% mothers, <sup>b</sup>first number in parentheses = number of samples, second = total participants.

Figure 2.2. A data-based model of the family attachment network, Figure from Bretherton (2010, p.15) – based on meta-analyses by van IJzendoorn (1995), De Wolff and van IJzendoorn (1997) and van IJzendoorn & de Wolff (1997)

Interestingly, the results shown in Figure 2.2 indicate the associations between attachment classifications for mothers and fathers is stronger for their AAI ( $r = .28$ ) compared to their SSP ( $r = .17$ ). Therefore, mothers and fathers seem to have a greater similarity between their adult representations of attachments than with their infant-parent attachments. A suggested explanation for these similarities is an effect of *assortative mating* (van IJzendoorn & De Wolff, 1997; Vandenberg, 1972), whereby men and women are attracted as partners partly due to similarities in their attachment representations or perhaps some related characteristic. Also, van IJzendoorn (1995b) suggested that the similarity of infant-parent attachment between

## 20 Rasch Analysis of the PRFQ

parents may be explained by the similarity of mother and father adult attachment representations rather than by characteristics of the child.

The relative independence of infant attachments with mothers and fathers is made clearer when it is considered that from the total sample of 950 families, a large proportion (38%) was classified as secure with only one parent and insecure with the other (van IJzendoorn & De Wolff, 1997). Furthermore, couples with only one secure child-parent classification were fairly equally represented by mothers ( $n = 174$ ) and fathers ( $n = 188$ ), and more generally, there is no evidence of a gender difference in the proportions of attachment classifications between mothers and fathers with their infants (van IJzendoorn & De Wolff, 1997) or between male and female adult representations (van IJzendoorn & Bakermans-Kranenburg, 1996).

It can be seen in Figure 2.2 that fathers' AAI attachment representations were predictive of SSP infant-father classifications ( $r = .37$ ), and this association was stronger for mothers' ( $r = .50$ ). Similarly, but to a lesser degree, the parental sensitivity was associated with infant-parent attachment for fathers ( $r = .13$ ), and for mothers ( $r = .24$ ). It appears that representations of attachment relationship in adulthood are more strongly predictive of infant-parent attachment behaviour than observations of parental sensitivity. These associations indicate that although there is evidence of relationships between these variables, there are clearly other substantial factors influencing the nature of infant-parent attachment interactions.

Bretherton (2010) noted that these meta-analyses of studies undertaken during infancy tend to have similar findings regarding mother and father comparisons; however, studies of how early attachments predict outcomes for children at older ages provide a more complex picture. Studies that explore early attachments and child outcomes rarely include fathers, usually have small sample sizes due to the labour intensive assessments and employ a wide variety of outcome assessment methods. In many cases, the Strange Situation assessments with the infant-father appear to be a poor predictor of significant child outcomes in comparison to the assessments with infant-mother (K. Grossmann et al., 2002). Results from a selection of studies that have found results of interest with fathers will be briefly summarised as examples of the diversity of results between mothers and fathers.

The Bielefeld longitudinal study (K. Grossmann, Grossmann, & Kindler, 2005) has followed 44 children and their parents from infancy to early adulthood. The results showed that both mothers' and fathers' infant-parent attachments

independently predicted their children's attachment representations at age 6 years. At age 10 year the children's attachment representations were only predicted by the infant-mother attachment. By the age of 16 and 22 years neither infant-mother nor infant-father SSP classifications predicted their children's AAI attachment representations. These studies also included an alternative assessment of infant security that proved to have greater utility in predicting later attachment. This will be discussed in the next section on father-child relationships.

A recent study of mothers and fathers with their 15-month old children (Kochanska & Kim, 2013) found that insecure SSP attachment with both parents predicted higher levels of behavioural problems at age 8 years compared to children with secure attachments with both parents. Interestingly, a secure attachment with either parent moderated this effect such that a secure attachment with one parent (either mother or father) had a protective effect against behavioural problems almost seven years later. Also, there appeared to be no additional protective benefit for a child to have secure attachments with both parents.

Verschueren and Marcoen (1999) found a somewhat similar result using a story completion task with kindergarten age children. One secure attachment with either parent was more beneficial than an insecure attachment with both parents. In this case, however, secure attachments with both parents predicted better child outcomes than having only one secure attachment with either parent. Interestingly, the specific child outcomes that were predicted by attachment security differed for mothers and fathers. Attachment with mother predicted child's "positiveness of self" more powerfully than attachment with the father, whereas the attachment with the father was more powerful at predicting the degree of anxious and withdrawn social behaviour. It was argued that this finding supported a common generalisation that a mother's more nurturing role facilitates development of the inner security whereas fathers have more of a playmate role that provides support for exploration and confidence with the outer world (Lamb & Lewis, 2010; Steele, Steele, & Fonagy, 1996).

It has been argued that observations of attachment behaviours in the home environment, assessed with the Attachment Behaviour Q-Set (AQS; Waters, 1995; Waters & Deane, 1985), have the benefit of being a more ecologically valid assessment of attachment relationships. Using the AQS, Verissimo et al. (2011)

## 22 Rasch Analysis of the PRFQ

found that father's and not mother's secure base behaviour with their child of about 32 months predicted the child's number of reciprocated friendships at between the ages of 4 and 5 years. No significant differences were found between AQS scores for mothers and fathers.

These studies indicate that the father-child attachments can be predictive of child outcomes, and in some cases appear to predict different outcomes to that of the mother. The following section examines assessments of the father-child relationship that have taken into account these differences.

### *Alternative assessments of father-child relationships*

Although mothers and fathers can potentially play a very similar role in relation to their children, the dominant means of assessing the early father-child relationships has been based on the traditional attachment framework. This approach has certainly provided insights into parent-child relationships, and yet at the same time limited consideration of relationship characteristics more likely to be typical of the infant-father relationship (Lamb, 2005). The following research has explored characteristics of the father-child relationship that have progressed the understanding of parental influence on child development.

Although infant-father attachment has been found to predict some later child attachment representations, fathers' sensitive and challenging interactions during play with toddlers has been found to be a better predictor of child attachment representations at a range of ages through childhood and into early adulthood. Grossmann and colleagues (2005; K. Grossmann et al., 2008) argue for a broader view of attachment that incorporates *security of exploration* in addition to attachment security, a concept they've called *psychological security*. Secure exploration is characterised by "confident, attentive, eager and resourceful exploration of materials or tasks, especially when a child is facing disappointment" (K. Grossmann et al., 2008, p. 857).

The distinction between the security of attachment and security of exploration has been presented graphically (Figure 2.3) in a way that shows the complexity of multiple attachment figures and the commonality of the two aspects of security. Bowlby's (1969) theory of attachment included analysis of exploratory behaviour and described an attachment figure as a secure base from which to explore. Subsequent attachment research, however, has focused mostly on proximity

behaviours rather than a child’s need for stimulation in the context of exploration (Paquette & Bigras, 2010).

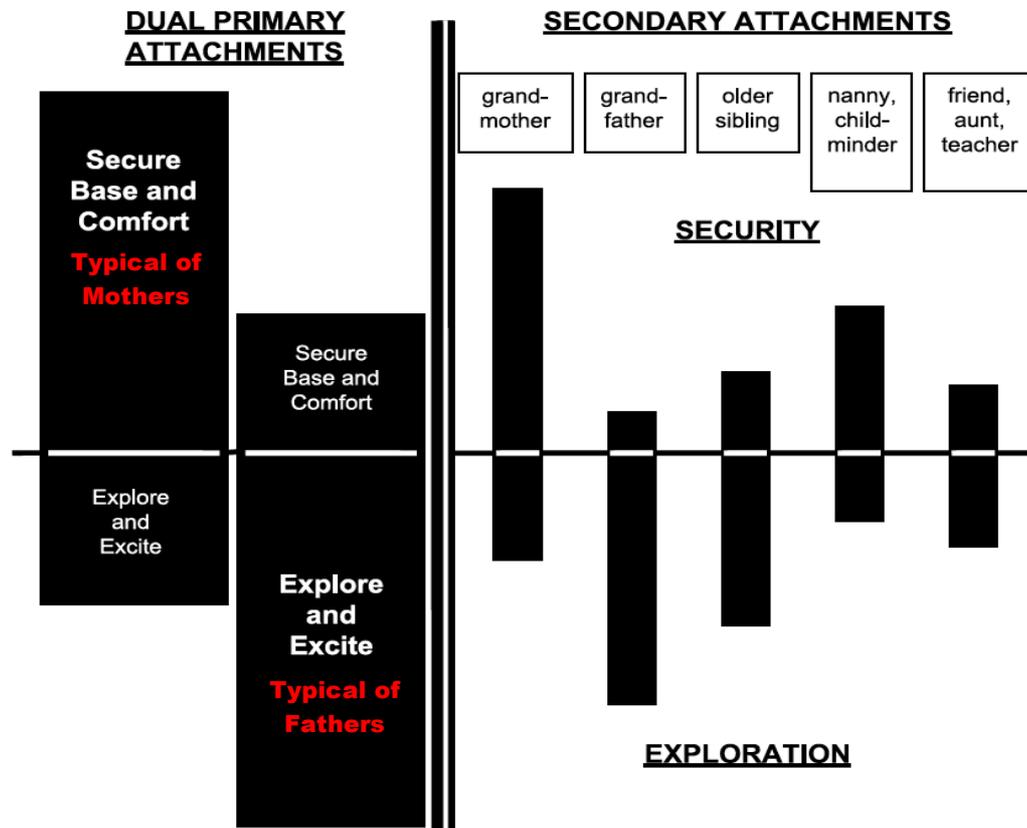


Figure 2.3. Proposed diagram showing dual attachment systems of security and exploration (from Richard Bowlby's personal lecture notes, Newland & Coyl, 2010, p. 27)

The Bielefeld longitudinal study (K. Grossmann et al., 2005), previously discussed with reference to the AAI and SSP, also included an observational assessment of security of exploration with the assessment of *sensitive and challenging interactive play* (SCIP). Results showed that fathers’ interactive play with their toddlers predicted the children’s attachment representations at ages 10, 16 and 22 years (K. Grossmann & Grossmann, 2009; K. Grossmann et al., 2005). At the ages of 16 and 22 years, these children’s attachment representations based on the AAI could not be predicted by their infant SSP classifications from either parent. These findings confirm long-standing theories that suggest play behaviour is a salient aspect of early father-child relationships that promotes children’s later social-

## 24 Rasch Analysis of the PRFQ

emotional development (Lamb & Lewis, 2004). A longitudinal study that spans development from infancy to adulthood is a valuable source of evidence, although it has the shortcomings of representing only a specific birth cohort, in this case the context was German families of the mid 1970s. Nevertheless this assessment of father-infant interaction is particularly useful in understanding the father-child relationship security, the unique role of father sensitivity, and how to conceptualise the complexity of wider family systems and attachment processes (Hill, Fonagy, Safier, & Sargent, 2003).

A study by Ramchandani et al. (2012) illustrated how a specific assessment method of early parent-infant interactions appeared to restrict or frustrate father-infant interactions. The procedure was originally developed such that the infant is seated directly in front of the parent (Murray, Fiori-Cowley, Hooper, & Cooper, 1996), this aspect of the procedure was modified to include an additional setting with the parent and infant interacting on a floor mat, allowing for a wider range of interactions that were inclusive of more active play (widely considered more typical of father-infant interactions). Observations of intrusiveness and remoteness of father-infant interactions with three-month old infants (in this additional setting) predicted mothers' reports of children's externalising behavioural problems at 12 months. These associations were not found using the conventional seated observations, indicating the greater predictive validity of the instrument in the more interactive context.

An observational assessment called the Risky Situation (Paquette, 2004a, 2004b) is based on similar principles to the SSP and was developed to place a more central focus on exploration in the parent-child *activation* relationship. Activation refers to the stimulation of a child's exploration of their social and physical environment, while ensuring they are safe and protected. Initial studies examining this instrument indicate the construct of activation is sufficiently different from that of the SSP security with only 31% of children having both an activated and secure relationship with the same parent (Paquette & Bigras, 2010). Further study of this potentially informative assessment is required to confirm its validity and determine its potential for prediction of later child development.

Assessment instruments such as those developed by Paquette (Paquette & Bigras, 2010) and Grossman et al. (2008) are addressing the need to better understand a father's contribution to child development. In particular, the findings

from this area of research are contributing to the issue of relationship quality that has been identified as lacking in the study of fatherhood. The growing evidence demonstrating the importance of the father-child relationship accentuates the need to investigate the social and personal factors that can impede fathers connecting well with their children. This endeavour will require larger studies with alternatives to interview or observation assessments that are time consuming for research participants and costly to administer.

Attachment theory continues to develop as it is integrated with or contrasted with other approaches to examine the broad spectrum of human development and relationships (Giudice & Belsky, 2010; Thompson, 2010). Two areas of growth in attachment related research are of interest in this study: firstly, the father-infant relationship which has already been discussed, and secondly the relatively new field of *reflective functioning*, which will be reviewed in the next section.

### **Reflective Functioning, Mentalizing and Related Terms**

Fonagy and colleagues (Fonagy, Steele, Steele, Moran, & Higgitt, 1991) developed a scale of *reflective functioning* (RF) in order to research attachment-related metacognition and to operationalise an individual's general level of *mentalizing*, a term that has been used interchangeably with RF (J. G. Allen, Fonagy, & Bateman, 2008). In its broadest sense, mentalizing can be understood as “attending to states of mind in oneself and others – in Fonagy’s apt phrase, “holding mind in mind” (p. 3; J. G. Allen, 2006). While this assessment of RF is based primarily in the field of attachment theory, studies of mentalizing have parallel lines of investigation within the domains of neurophysiology and social cognition (U. Frith & Frith, 2003). The importance of individual competence in the area of understanding and managing emotions has also become widely accepted beyond the academic and therapeutic literature. Daniel Goleman’s (1995) internationally best-selling book, *Emotional Intelligence* and the development of tests to assess emotional intelligence (analogous to recognised IQ tests; Mayer, Roberts, & Barsade, 2008) highlight the popular interest in emotional competence.

The importance of and interest in the mentalizing construct is clearly not new. A variety of other concepts related to mentalizing have a long history of exploration within various fields of psychology including theory of mind (ToM; Baron-Cohen, Wheelwright, Hill, Raste, & Plumb, 2001), intersubjectivity (Lyons-Ruth, 2006),

## 26 Rasch Analysis of the PRFQ

mindblindness (Baron-Cohen, 1997), mindfulness (Shapiro, 2009), mind mindedness (Meins, Fernyhough, Arnott, Leekam, & de Rosnay, 2013), meta-emotion (Katz, Maliken, & Stettler, 2012) and folk or naïve psychology (Poulin-Dubois, Brooker, & Chow, 2009). The distinctions between many related concepts and mentalizing have been detailed by Allen and Fonagy (2006) and a summary table was produced by Allen et al. (2008), which is reproduced below (Table 2.2).

Table 2.2

### *Differentiating Mentalizing from Overlapping Terms*

*(from J. G. Allen et al., 2008, p. 41)*

Term	Distinction
<i>Mentalizing</i>	Attending to mental states in self and others, and interpreting behaviour accordingly
<i>Mindblindness</i>	Antithesis of mentalizing; employed originally to characterize autism
<i>Mindreading</i>	Applies to others and focuses on cognition
<i>Theory of Mind</i>	Conceptual framework for mentalizing, focuses on cognitive development
<i>Metacognition</i>	Focuses primarily on cognition in the self
<i>Reflective Functioning</i>	Measurement of mentalizing in attachment context
<i>Mindfulness</i>	Focuses on present and not limited to mental states
<i>Empathy</i>	Focuses on others and emphasizes emotional states
<i>Emotional Intelligence</i>	Pertains to mentalizing emotion in self and others
<i>Psychological Mindedness</i>	Broadly defined, the disposition to mentalize
<i>Insight</i>	Mental content that is the product of the mentalizing process

As can be noticed in the distinctions listed in Table 2.2, there are a number of aspects of mentalizing that are given more or less focus in different concepts.

Recently these aspects have been described as four polarities of mentalizing (Table

2.3; Fonagy & Luyten, 2009) and are considered in depth by Fonagy, Bateman and Luyten (2012). Firstly, mentalizing can be with regards to the self or the other. This distinction is clear in the concepts of mindfulness, which is typically focused on the self and empathy, which is focused on the other. The interplay of reflecting on the mental states of both self *and* others is important for healthy social functioning. Secondly, automatic or implicit mentalizing can be contrasted with that which is controlled or explicit. Implicit mentalizing does not require effort or awareness and functions instinctively. Explicit mentalizing typically takes longer to process, requires reflection and can be verbalised or expressed. Mentalizing can be cognitive and/or affective. These two aspects of mental state awareness have parallels with Baron-Cohen's (Baron-Cohen, 2002) more recent formulation of theory of mind which distinguishes an empathizing system (primarily affective) from a systemizing system (primarily cognitive). Finally, the focus of mentalizing can be directly on internal mental states (e.g., thoughts, feelings, desires, intentions), or rely mostly on the external and visible indications of mental states. These external cues include facial expressions, body posture and nonverbal behaviour. In addition to these polarities, mentalizing has been recognised as being context and relationship specific (Luyten, Fonagy, Lowyck, & Vermonte, 2012).

Table 2.3.

*Four Polarities or Dimensions of Mentalizing*

	Self	Other
Automatic / Implicit		Controlled / Explicit
	Cognitive	Affective
	Internal	External

**Application of mentalizing**

An imbalance or deficit in any of the four polarities of mentalizing is suggested to contribute to various types of relational difficulties and psychopathology, such as the problematic traits associated with disorders of the personality (Fonagy & Luyten, 2009). For example, antisocial personality disorder is characterised by heightened understanding of the mind of others, yet often a failure

## 28 Rasch Analysis of the PRFQ

to have insight into one's own mental states. Whereas for people with borderline personality traits the implicit and affective aspects of mentalizing often overwhelm them; they often struggle to intentionally engage in controlled or reflective consideration of mental states.

The automatic and controlled polarities of mentalizing could be thought of as being unconscious and conscious respectively. Behaviour can be understood to result from an interaction between conscious and unconscious processes (Baumeister, Masicampo, & Vohs, 2011). Conscious thought processes that are typical of healthy mentalizing include reflection, perspective taking, and anticipation as well as potential to override automatic responses. Such processes can be indirect and delayed in comparison to unconscious processes; however, they can make possible behaviours that are influenced by information or factors that are not present in the current situation. In this way, mentalizing is associated with self-agency and self-regulation – a sense of competence and degree of influence over one's experience or behaviour. Also, appropriate mentalizing enables realistic understanding and prediction of the experience and behaviour of others, a basic requirement of interpersonal communication and close relationships.

Deficits in mentalizing have been described in terms of *pseudomentalizing* and the pre-mentalizing modes of *pretend*, *psychic equivalence* and *teleological thinking* (Bateman & Fonagy, 2008b; Skarderund & Fonagy, 2012).

Pseudomentalizing is a term used to describe what appears as to be mentalizing; however, when examined it is found to be self-serving or has features such as being overly implausible, destructively inaccurate or intrusive (e.g., incorrectly attributing mental states to someone's behaviour based on personal bias). Psychic equivalence and pretend mode are two opposite extremes of distortions of reality that indicate a lack of mentalizing - they are to be expected in the early childhood, but can be problematic in later stages of development. Psychic equivalence is when the mental world is experienced as if it was real and is equated with concrete thinking or understanding (such as with vivid dreams and delusions). In this mode, there is no consideration of alternative perspectives and rather than thoughts and feelings representing reality, they are taken to be equated with real experience. The pretend mode reflects a disconnection between mental states and reality. Internal experiences can be dissociated and have little meaning with regards to genuine experience, behaviour or reality. The teleological mode is noticeable when mental

states are required to be proven by behaviour or expressed by action. These failures of mentalizing, their inter-relationships and possible consequences are presented in Figure 2.4. How these separate aspects of mentalizing deficits can be assessed in a way that distinguishes each is yet to be resolved.

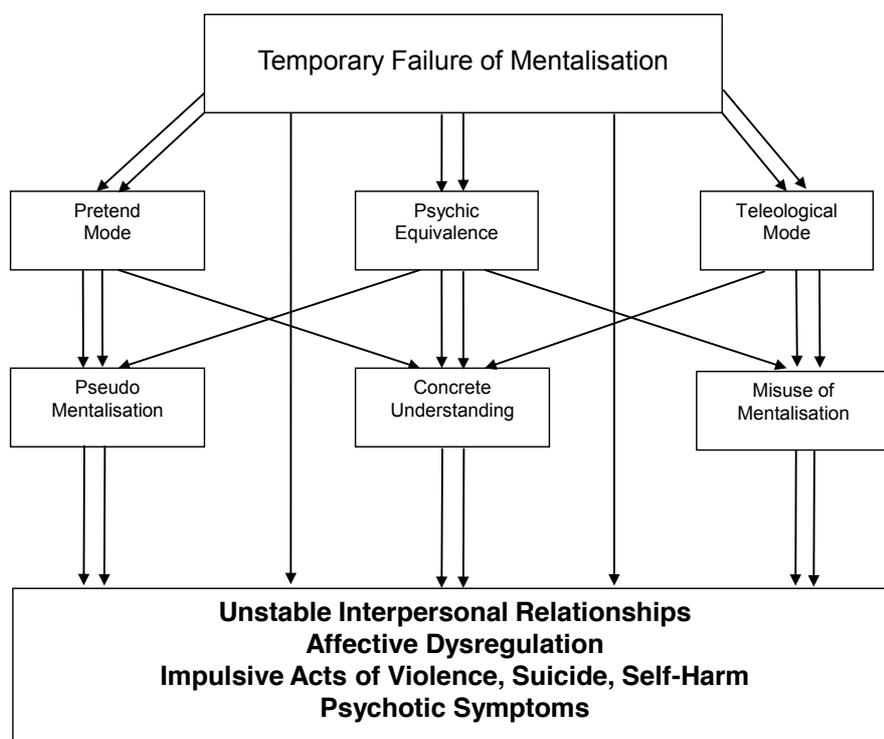


Figure 2.4. Model of the relationship between failures of mentalizing and possible consequences (from Luyten, 2014, p. 2)

In addition to a general application to relationships and psychotherapy, there are specific mentalizing focused treatments that have shown positive results, particularly for the treatment of borderline personality disorder (Bateman & Fonagy, 2001, 2008a, 2012; Fonagy & Luyten, 2009). Mentalization-based treatment (MBT) or principles have also been applied to a wide range of contexts and psychopathology, including family therapy (Asen & Fonagy, 2012; Safier, 2003), violence and antisocial behaviour (Fonagy, 1999; Leichsenring, Kunst, & Hoyer, 2003), posttraumatic psychopathology (Stein & Allen, 2007), eating disorders (Skarderund & Fonagy, 2012), drug addiction (Philips, Kahn, & Bateman, 2012), schizophrenia (Lysaker et al., 2011), obsessive-compulsive personality disorder

### 30 Rasch Analysis of the PRFQ

(Dimaggio et al., 2011), anxiety (Nolte, Guiney, Fonagy, Mayes, & Luyten, 2011), depression (Luyten, Fonagy, Lemma, & Target, 2012), alexithymia (Vanheule, Verhaeghe, & Desmet, 2011), mediation (N. D. Howieson & Priddis, 2012), difficult patients (Simonsen, Nørgaard, Larsen, & Bjørnholm, 2011), and adolescents with complex difficulties (Bevington, Fuggle, Fonagy, Target, & Asen, 2012; Rossouw & Fonagy, 2012). These investigations add to the widely published literature of related concepts such as theory of mind, emotional intelligence, empathy and mindfulness.

Allen et al. (J. G. Allen, 2012; J. G. Allen, Bleiberg, & Haslam-Hopwood, 2003) have proposed that mentalizing serves as a “conceptual compass” for psychotherapy treatment. For this reason, the principles of mentalizing have been applied to the understanding of many adult mental health difficulties and integrated into a wide range of psychotherapy treatments of differing theoretical approaches (J. G. Allen & Fonagy, 2006; J. G. Allen et al., 2008). These mentalizing approaches have been described as “a useful way of thinking about the psychotherapeutic treatment... as it conceptually harmonizes cognitive and psychodynamic interventions” (Stein & Allen, 2007, p. 287). Even though the support of mentalizing as a concept appears to be strong and it appears to be clinically useful, the evidence for the quality of the assessment instruments still requires more thorough investigation and development. Challenges with the measurement of reflective functioning are addressed in a subsequent section.

In the context of parenting, the concept of reflective functioning has only recently been applied in treatment or intervention (Baradon, Fonagy, Bland, Lénárd, & Sled, 2008; J. A. Howieson & Priddis, 2011; Reynolds, 2003; Slade, 2006; Sled, Baradon, & Fonagy, 2013; Suchman, Pajulo, Kalland, DeCoste, & Mayes, 2012). Although these studies are promising with respect to the application of mentalizing to parenting, the studies of parental mentalizing that do include assessments have similar limitations: small samples that are usually specific to clinical or special groups, fathers are rarely considered and not included, and the instruments are newly developed with minimal examination of their quality. Further study of PRF with larger normative samples including fathers is needed to better understand the construct of mentalizing in parent-child relationships.

### **The development of mentalizing capacity**

The concept of mentalizing has utility for better understanding of both normative development and dysfunctional aspects of social-emotional development. In addition, mentalizing could help uncover the factors associated with challenging problems such as cycles of abusive or neglectful relationships (Drielsma, 2001) and the intergenerational transmission of attachment insecurity (van IJzendoorn, 1995a).

It has been argued that the parental capacity for mentalizing contributes to a child's development of affect regulation, providing the child with a sense of coherence and predictability to his or her inner experiences and experience of the social environment (Fonagy, Gergely, Jurist, & Target, 2002; Fonagy, Gergely, & Target, 2007). In contrast, low parental mentalizing is implicated as a likely risk factor for the development of childhood psychopathology (Sharp & Fonagy, 2008). A newly emerging source of evidence for the role of mentalizing is from neurological research. Studies of neurological functioning have shown that parent-child mentalizing is a possible contributor to the hardwiring of a child's specific brain system, which is associated with certain mentalizing tasks (C. D. Frith & Frith, 2006; U. Frith & Frith, 2003). Other neurological evidence includes the important function of mirror neurons in the experience of empathy (Rizzolatti & Craighero, 2004), the study of infants' early preferences for social stimuli, and examining the ability of infants to distinguish between mechanical and biological movement (U. Frith & Frith, 2003). Such research on brain functioning is rapidly accumulating and providing insights that contribute further to the understanding of mentalizing capacity and its development.

There is a complex association between the experience of stress or arousal and both the capacity for mentalizing and its development (Fonagy et al., 2012). As already discussed in this review, early childhood stress experienced in the absence of a supportive relationship is detrimental to many aspects of a child's development. A supportive caregiving relationship implies a degree of attachment security and mentalizing capacity. Although there may be some general tendencies for some people to have high or low levels of mentalizing capacity, the ability to mentalize is understood to be context and relationship specific. This specificity has been conceptualised as being determined largely by an individual's level of stress and attachment strategy.

### 32 Rasch Analysis of the PRFQ

Fonagy and Luyten (2009) have proposed a model (Figure 2.5) that locates a switching point at which mentalizing changes from controlled to automatic at a certain level of stress. Based on an understanding of neurocognitive systems, affect regulation and the activation of the attachment system (Mayes, 2006), this model presents a process of complex interactions of some important factors that influence mentalizing quality. In brief, at a low levels of stress or arousal controlled mentalizing is dominant; however, as arousal increases the capacity for controlled mentalizing decreases and there comes a point where automatic mentalizing dominates. For individuals with mentalizing related deficits (e.g., resulting from trauma or mental health difficulties), automatic mentalizing in high stress is likely to feature pseudomentalizing or the pre-mentalizing modes of psychic equivalence, pretend or teleological thinking. Depending on a person’s attachment strategies (e.g., secure, anxious, avoidant) the switch may be at a high or low level of stress, and the recovery back to controlled mentalizing may be fast or slow. For those individuals with secure attachment and without significant mental health difficulties, high stress would likely still result in biases and distortions of their account of mental states and behaviour – although these deficits in mentalizing are not as likely to be problematic for them. Their switch to automatic mentalizing would most probably be at a high level of stress and would return to controlled mentalizing relatively quickly.

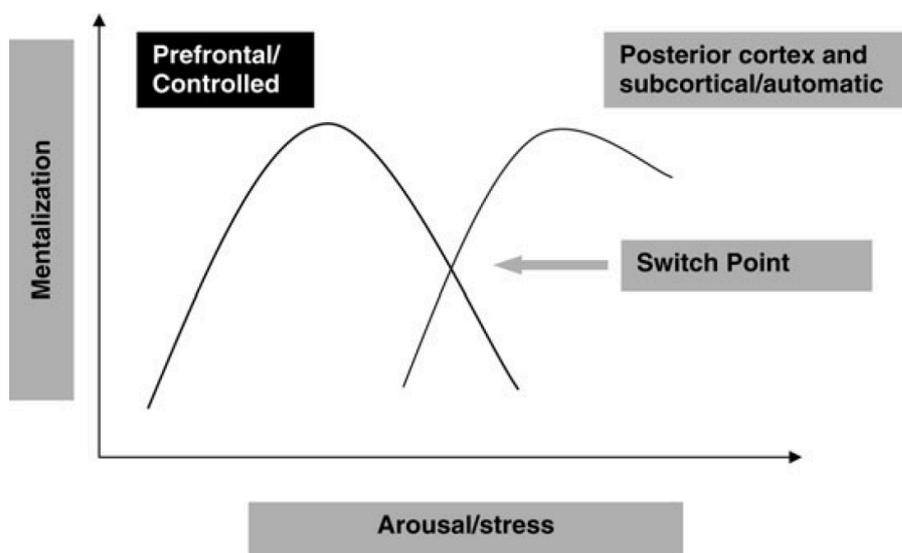


Figure 2.5. Biobehavioural switch model of the relation between stress and controlled or automatic mentalization (from Fonagy & Luyten, 2009, p. 1367).

These patterns of mentalizing have implications for the capacity of a stressed parent to provide a supportive relationship to his/her child. A parent with a switch point that is at very low levels of stress may experience personal distress when the child is upset, and therefore find it difficult to appropriately mentalize the child's behaviour. In such a scenario, the parent's ability to comfort the child or help the child self-regulate would be limited.

The next section examines literature that has addressed parent-child mentalizing, also referred to as parental reflective functioning (PRF).

### **Parental reflective functioning**

The first study of RF was in the context of parenting, based on AAI interviews with 100 couples from the London Parent-Child Project (Fonagy et al., 1991). Parents' RF assessed before the birth of their child predicted the child's attachment security with each parent independently with mothers at 12 months ( $r = .51$ ) and fathers at 18 months ( $r = .36$ ). Arnott and Meins (2007) were the next to replicate this early study (finding similar results), although only 18 mother-infant and 15 father-infant assessments were undertaken. These two studies were based on mentalizing of past childhood experiences from the Adult Attachment Interview (AAI).

The children of the Parent-Child Project were followed up at age 5 years ( $N = 72$ ) and 11 years ( $N = 46$ ), and child behaviours were compared with pre-birth assessments of mothers' and fathers' RF (Steele & Steele, 2008). At age 5 years, mothers' RF was not associated with any child behaviour problems reported by mothers or fathers. Higher fathers' RF however, was associated with lower father report of their child's withdrawn delinquent and aggressive behaviour (particularly with sons), as well as mother report of withdrawn behaviour problems. This association was confirmed by children of fathers' with high RF reporting lower emotional, behavioural and peer problems at age 11 years. Also at age 11 years, mothers' and fathers' RF was positively associated with their child's self-report of higher self-esteem. Although less than half the families were available for the 11 year follow-up, the distribution of parental RF from the reduced sample was not statistically different from the original sample of 100 families. Clearly the small

### 34 Rasch Analysis of the PRFQ

sample size and large dropout rate of this study mean the results must be interpreted with caution and further study is required to confirm the findings.

Parental mentalizing has been more directly assessed by observation of appropriate mind-related comments – an assessment referred to as Maternal Mind-Mindedness (MMM; Meins, Fernyhough, Fradley, & Tuckey, 2001). Studies of MMM with mothers and infants have provided interesting predictions of their children's ToM. Mothers' appropriate mental state language independently predicted children's ToM assessed in a range of tasks at 45 and 48 months, whereas inappropriate language and observed security of attachment were not significantly correlated with ToM (Meins et al., 2002). These findings were confirmed in a larger study with 51-month-old children ( $n = 161$ ), which also indicated that nonattuned mind-related comments may be indirectly linked with ToM (Meins et al., 2013).

It is interesting to note that the association between parental attachment classifications from the AAI and mind-minded comments have been found to be stronger for fathers than for mothers (Arnott & Meins, 2007), a similar finding to that of van IJzendoorn (1995a) with regards to attachment classification and parental sensitivity. Also, fathers' proportions of appropriate and inappropriate mind-related comments were found to be positively correlated (Arnott & Meins, 2007) – indicating that fathers who made proportionally more appropriate comments also made more inappropriate comments. Although interesting, this result was from a very small sample and there was no reference to examination of the instrument's suitability with fathers, so the finding needs to be considered with caution. Nonetheless, it appears that fathers' sensitive and appropriate interactions with his child may be associated with their representations of childhood attachment more strongly than mothers. Arnott and Meins (2007) suggested that mothers' interactions with their children are less influenced by their own attachment experiences compared to fathers, possibly because mothers have more support and opportunity to learn about their infant's moods and interests. This conclusion implies that sensitive or mind-minded behaviour may require opportunity, time and learning for it to be assessable, and therefore, fathers' (who are typically engaged in less childcare than mothers) capacity for mentalizing may not be accurately assessed with these instruments. This hypothesis requires further study with the appropriate modelling and controlling of variables for it to be justified with empirical evidence.

The distinctions between mother and father capacity for mentalizing have to date been examined in relatively small samples, predominately with clinical samples. Development of a clearer understanding of parental mentalizing of mothers and fathers requires further investigation with larger samples. This is particularly true if differences are to be identified between specific groups and if findings are to be generalised to broader populations. Of relevance to this thesis, are the distinctions between mothers' and fathers' mentalizing with their infants. There is no specific theory of gender difference proposed in the mentalizing or RF literature. Baron-Cohen (2002), however, concluded from a range of evidence and experience from the study of Theory of Mind - a concept close to mentalizing - that there are two psychological dimensions of "empathizing" and "systemizing" that tend to characterise females and males respectively. Females are suggested to have a preference to analyse mental states and respond emotionally, whereas males tend to analyse systems or construct systems based on 'underlying lawful regularities'. This theory implies females would have higher levels RF than males. Systematizing may be a genuine attempt at making sense of behaviour and can reflect a curiosity in mental states; however, as a strategy for understanding social behaviour it is unlikely to adequately appraise irrational, ambiguous or emotionally motivated behaviour. Also, mental states that are not evidenced by behaviour such as intentions, desires and fears, would be difficult to systematize, and may be understood in a concrete manner characteristic of the pre-mentalizing teleological mode.

Similarly, it has been argued that lower levels of the capacity to be aware of and describe emotions is normative in men (the so-called "normative male alexithymia hypothesis"; Levant, 1992; Levant, Hall, Williams, & Hasan, 2009) as it is assumed that gender role socialization may normatively lead to less emphasis on emotions for men compared to women, at least in men who have been reared to endorse more "traditional" male values. The measurement of mentalizing would provide a means of assessing whether such gender differences exist in parent-child relationships.

Fonagy and Target (2005) proposed that the mother-infant capacity for mentalizing is an important mechanism in the intergenerational transmission of adult attachment security to infant-parent attachment. Slade and colleagues (Slade, Grienenberger, Bernbach, Levy, & Locker, 2005) investigated this possibility by assessing parental RF (PRF) from the Parent Development Interview (PDI) with 40

### 36 Rasch Analysis of the PRFQ

mothers. This preliminary study found mother PRF was correlated with infant security more strongly than the correlation found between mother sensitivity and infant security (Bretherton, 2010). Furthermore, when mothers' PRF was controlled, the association between mothers' attachment representations and infant-mother attachments was no longer significant. These results are yet to be replicated with a larger sample or with inclusion of the fathers' attachment and PRF assessments.

A number of therapeutic interventions have been developed specifically to improve PRF for mother-baby dyads, so as to improve outcomes for children. Most notable is the *Minding the Baby* reflective parenting program (Slade et al., 2006), which has shown to improve the capacity of mothers to reflect on their children's experience in a randomised control trial (Slade et al., 2006). Other examples include an intervention for mothers in hostels for the homeless (Sadler et al., 2013) and mothers in prison (Baradon et al., 2008), which have shown a significant improvement in child development and PRF respectively, following the therapeutic program. Results have recently been published from the first longitudinal study of a mentalizing based intervention with 24 mothers and infants (Ordway et al., 2013). An increase was found in mothers' PRF at follow-up for the intervention group; however, the change was not statistically significant compared to control group mothers. An attachment-based intervention for substance-using mothers with toddlers has shown improvements in PRF compared to those mothers in standard parent education programs (Suchman et al., 2010; Suchman et al., 2012).

From investigations of mentalizing over more than 20 years, and more recent study of PRF, many questions still remain regarding the contribution of parents' mentalizing to children's developmental outcomes. One area yet to be examined is the assessment of both mother and father PRF with infants, and more research is required with larger non-clinical community samples to establish norms and increase the power of analysis. Based on RF findings from the AAI, it is possible that father PRF could be a key characteristic of positive fathering that could contribute to improved likelihood of healthy child development. It may be that mentalizing is a central characteristic of "good-enough" parenting (Winnicott, 1971) for both mothers and fathers. And this may be especially so for fathers, who compared to mothers tend to be less available, have less opportunity for involvement, and sometimes have to go against social norms to be actively involved with their infant. For these possibilities to be thoroughly examined, data from instruments assessing PRF needs to be verified

as invariant across mothers and fathers for comparisons to be made, and ideally these instruments would be suitable for larger population studies (i.e., a self-report questionnaire).

### **Assessment of Reflective Functioning**

The most commonly used assessment of RF is derived from a system of coding references to mental states from the Adult Attachment Interview (AAI, Fonagy, Target, Steele, & Steele, 1998) and was first reported as a rating of Reflective Self-Function (Fonagy et al., 1991). The assessment from the AAI considers representations of attachment relationships within the context of attachment narratives and is considered a general assessment of a person's ability for RF. The single score indicates global mentalizing capacity – the assessment is not designed to give any indication of the variability of the responses or the different dimensions or capacities in different contexts (Katznelson, 2014); however, the capacity for RF is thought to differ in regards to specific relationships (Sharp & Fonagy, 2008), with the parent-infant relationship being of particular interest given its importance in the child's social-emotional development.

Scores of RF from the AAI (Fonagy et al., 1996) have shown evidence of discriminant validity, with no significant relationships found with personality scores of extraversion, neuroticism or psychoticism. This analysis also found no significant relationship with any of the scales of the Langer 22 (screening for psychiatric disorders), the Sources of Self Esteem Inventory, or Epstein's Mother-Father-Peer Scale (assessments of independence vs. overprotection, and acceptance vs. rejection). Mood state was assessed immediately prior and following administration of the AAI and was not found to have any relationship with RF scores. No significant association was found between RF and parent demographic variables such as social class, socioeconomic status, ethnic background or education (Fonagy et al., 1991).

A more direct assessment of parent-child mentalizing or PRF has been developed using the PDI (PDI-RF, Slade, 2005). The clear difference between these two assessments is that the participant in an AAI reflects on incidences and memories relating to his/her primary caregivers that are usually formed many years ago in infancy, whereas the PDI draws on a current and developing relationship with the participant's own child. A reflective parent of an infant is in a dynamic relationship with his/her child in times of rapidly changing developmental growth. In

### 38 Rasch Analysis of the PRFQ

contrast to a reflective adult who would provide an elaborate picture of their experience of being parented, a high PDI-RF manuscript is likely to include examples of a parent acknowledging a struggle to understand his or her child (*opacity* of mental states). Reflective parents actively imagine how their children feel and think; however, they also recognise their own limited insight and the likelihood of alternative explanations for behaviours.

The capacity for PRF is assessed with detailed scoring of a transcript from the PDI using a procedure based on the manual for scoring RF from the AAI (see the Methods section of this proposal; Fonagy et al., 1998; Slade, Bernbach, et al., 2005). The questions in the PDI prompt the parent to reflect on his or her child, the relationship with the child, the experience of being a parent and his or her own childhood experiences. These questions are grouped under the headings: View of the Child, View of the Relationship, Affective Experience of Parenting, Parent's Family History, Separation/Loss (relating to feelings towards the child), and Looking Behind, Looking Ahead. In the same manner as the scoring of the AAI, the PDI-RF rating is a single score that reflects a global or gestalt assessment of parental mentalizing capacity.

The PDI-RF is arguably less subject to situational biases than observational procedures such as the SSP and MMM, since the questions reflect more generalised ways of thinking and being, compared to a single short period of observation. Also, parental reflections may be able to show the capacity or potential for parental mentalizing, when circumstances or experiences in some way impede parent-infant interaction, such as a father who has had limited contact with their child. As suggested by Arnott and Meins (2007), some observational procedures may be biased against parents with limited experience with their child. Validation studies have found significant relationships between the PDI-RF and key criterion variables (Slade, Bernbach, et al., 2005), including adult and child attachment security (Slade, Grienberger, et al., 2005) and disruptive affective communication (Grienberger, Kelly, & Slade, 2005).

Although the AAI and PDI, and the scoring for RF provide a rich and detailed source of information, they are time-consuming processes, requiring specialist training and costly human resources. Also, the final score from these assessments does not reflect the variability of capacity in the different dimensions or polarities of mentalizing. Interview procedures may be more easily undertaken than some

observational techniques; however, they remain impractical for large-scale research and routine clinical applications. There is a need to measure this important aspect of parenting with a simpler and easier instrument that is suitable for use with mothers and fathers. A measurement instrument of this kind would be a valuable screening tool in the context of time and resource limited child health services, and it would be valuable for population-based research where the complex pathways to many developmental outcomes remain poorly understood. As one of the primary authors of mentalizing research has stated: “An easily applied measure of mentalizing capacity is urgently needed” (p. 107; Fonagy, Bateman, & Bateman, 2011). Yes, there is need to test this theory and its concepts in large population-based samples - to refine it and establish its utility in promotion, prevention and practice

An easily applied measure of mentalizing would ideally be a short self-report questionnaire. Attempts to develop such measures have been initiated, such as the Reflective Functioning Questionnaire by Fonagy and Ghinai (as cited in Luyten, Fonagy, Lowyck, et al., 2012). To date, however, there has been no published examination of a questionnaire specifically assessing parental mentalizing (Schiborr, Lotzin, Romer, Schulte-Markwort, & Ramsauer, 2013). Luyten and colleagues (2012) have provided an overview of a collection of questionnaires that assess some of the dimensions of mentalizing. These scales were not developed for the mentalizing concept or to distinguish between the mentalizing polarities, although the examples offered suggest many aspects of mentalizing are assessable with self-report questionnaires. A self-report questionnaire that attempts to assess dimensions of parental mentalizing is discussed in the next section.

### **Parental Reflective Functioning Questionnaire (PRFQ)**

A self-report questionnaire of parent-child mentalizing is under development by Luyten et al. (2009). The scale is called the Parental Reflective Functioning Questionnaire (PRFQ, Appendix A) and research is currently being undertaken in various international samples to examine its psychometric properties. The PRFQ consists of 39 items that are rated on a 7-point scale indicating the respondent's level of agreement with the statement. The authors of the PRFQ are leading international researchers of mentalizing, and the items have been developed on the basis of the authors' considerable depth of understanding of the construct.

## 40 Rasch Analysis of the PRFQ

The assessments of PRF from the PRFQ and from scoring the PDI are conceptually similar. However, as highlighted by Luyten et al. (2009), the PRFQ cannot take into account the clinically rich, detailed and idiosyncratic information that is acquired from administering and interpreting a PDI. Examples of idiosyncratic information include the assessment of the interviewer-participant interaction and information gained from specific probing or clarifying by the interviewer. The scoring of PRF from the PDI transcript is a thorough process of evaluation that takes into account the subtleties and complexity of the participant's narrative. In addition, the PDI requires the recollection of concrete examples of recent interactions, which is likely to evoke an emotional reaction. As previously discussed, automatic or implicit mentalizing is more likely in conditions of arousal, therefore the PDI-RF has potential to evoke automatic aspects of pseudomentalizing or pre-mentalizing.

The PRFQ has undergone an initial unpublished process of validation during its development (Luyten et al., 2009). Following a consideration of relevant literature on mentalization and social cognition, items were generated based on content from the RF manuals for the AAI and PDI, and from items of comparable questionnaires. The items of the PRFQ were created to assess three broad dimensions of PRF: (a) curiosity in, and explicit efforts to tease out mental states underlying behaviour, which includes an awareness of the opaqueness of mental states; (b) repudiation or defence against mentalization as reflected in various pre-mentalizing modes (i.e., teleological model, pretend mode, psychic equivalence mode); and (c) recognizing the developmental aspects of states of mind. Experts were asked to rate the initial pool of items for statements considered prototypical of high mentalizing mothers or low mentalizing mothers, and to rate how well the items captured that level of capacity.

The PRFQ consists of three subscales, which are distinguished by their scoring procedure and the aspect of PRF the items assess:

1. HL (high-low scored subscale) - 17 items, which include items that refer to the curiosity in mental states and explicit efforts to tease out mental states underlying behaviour (e.g., I am often curious to find out how my child feels). Some of these items reflect an awareness of the opaqueness of mental states (e.g., My child can react to a situation very differently than I think he or she will). Other items assess an aspect of PRF that is particularly relevant to the parental context - the recognition of the developmental aspects of states

of mind (e.g., I believe that how I think about my child will change over time.). Some of the items primarily address mental states of the child, others are mostly about the self and some are about both the self and the child fairly equally. These HL items were worded such that higher scores reflect higher levels of parental mentalizing.

2. LH (low-high scored subscale) - 14 items, which were created to assess repudiation or defence against mentalizing as reflected in various pre-mentalizing modes (i.e., teleological model, pretend mode, psychic equivalence mode). For example, “When my child is fussy he or she does that just to annoy me”, and “I believe there is no point in trying to guess what my child feels.” These items were worded in a way that lower scores reflect higher levels of parental mentalizing.
3. M (middle scored subscale) - 8 items, which were created to indicate pseudo-mentalizing or socially desirable response patterns, and an awareness of the opaqueness of mental states or the limits of mentalizing. For example, “I always know what my child wants”, and “No matter how sick my child is, I can always tolerate him or her”. These items were worded such that a response in the middle of the rating scale reflects high levels of parental mentalizing, and lower levels as responses are towards either extreme of the scale.

The results of the PRFQ validation studies will hopefully shed some light on the multidimensional and polarity aspects of PRF, which until now have only been theorised. Luyten et al. (2012) have indicated that the PRFQ has potential to assess most of the mentalizing dimensions, including both polarities of the self/other and the cognitive/affective aspects of mentalizing. The controlled and internally focused aspects of mentalizing are also assessable by the PRFQ; however, implicit mentalizing is beyond the scope of self-report, as are the automatic interpretation of many external cues. Another limitation of self-report assessment of mentalizing is that it is off-line, in that it is a retrospective report of mentalizing whereas a parent-infant observation can provide an on-line, in the moment assessment. On-line assessments of mentalizing are more likely to be able to assess implicit and external aspects of mentalizing, as well as various pre-mentalizing modes that may not be accurately perceived in one’s self. Both positive mentalizing and the various types of

## 42 Rasch Analysis of the PRFQ

pre-mentalizing or pseudomentalizing are of interest in an assessment of parental mentalizing capacity. These two extremes may not be assessable on the one dimension as is the assumption in the single score rating of RF from the AAI and PDI. It is more likely that the pre-mentalizing modes and evidence of pseudomentalizing require separate assessments, and may be evidenced in different degrees in each of the self/other, cognitive/affective and internal/external polarities.

Detailed analyses of the PRFQ, with comparisons with the PDI-RF will reveal the utility of the PRFQ and potentially inform the understanding of parental mentalizing. This endeavour is the aim of the current study.

### **Summary**

This literature review supports an argument for the value of measuring paternal mentalizing – the capacity of fathers to reflect on the mental states of their children. Two main sources of evidence suggest father’s mentalizing to be an important aspect of the father-child relationship. Firstly, studies of early relationship security have shown that the quality of the father-infant relationship has implications for child, adolescent and adult social-emotional outcomes. Secondly, assessments of father mentalizing in relation to his own childhood experiences have been found to be predictive of psychosocial outcomes later in childhood. The assessment of father mentalizing in relation to his infant has potential to provide further clarity on the intriguing influence fathers have on their child’s development.

A self-report of PRF would make it possible to obtain population trends in father’s PRF in relation to various demographics, and to screen for men at risk of low PRF. If the PRFQ is proven to have sound psychometric properties, its predictive validity for family and child developmental outcomes could be examined and the developmental origins of mentalizing investigated. Importantly, the instrument could help identify needs for targeted early interventions and preventative measures to improve child developmental outcomes (Brooks-Gunn, 2003).

The following quote from Shonkoff (2010) highlights the responsibility of all aspects of society to support healthy childhood development:

Science tells us that the early childhood period is a time of both great opportunity and considerable risk, and its influence can extend over a lifetime. The foundational importance of the early years is increasingly appreciated across the political spectrum, and there is

growing recognition that families, communities, the workplace, and government each has a shared interest and distinctive, noninterchangeable role to play in assuring the healthy development of all young children. Stated simply, the science of early childhood and brain development is strong and growing, the moral imperative for preventive action is compelling, and the potential social and economic returns on investment are substantial. (p. 365)

This literature review has presented evidence for the importance of the father-infant relationship for child development, and therefore the need to better understand, assess and support that relationship. Adequate reflective functioning is argued to be a fundamental aspect of a father's capacity for a healthy father-infant relationship. Ensuring optimum father-infant relationships has potential of contributing to the substantial returns referred to by Shonkoff in the quote above. The next chapter addresses the challenge of measurement in the social sciences and provides an overview of the Rasch model.

## **Chapter 2 Part B: Measurement in Social Science Literature Review**

This section of the review chapter will provide background information that will help the reader understand the importance of and rationale for use of the Rasch model in the evaluation of rating scale questionnaires. Firstly, the Rasch model is situated within the context of measurement in the social sciences. Secondly, a critique of classical test theory is provided and comparisons made with modern test theory and the Rasch model. Finally, the distinctive qualities of the Rasch model are elaborated, and this provides the theoretical basis for the analysis techniques described in the Methods section of this thesis. The objective of this section is to present a rationale for the use of the Rasch model in the evaluation of scales such as the Parental Reflective Functioning Questionnaire – the focus of the analysis in this thesis. Such an extensive rationale is deemed necessary because of the dominance and entrenchment of the current measurement paradigm, which has a number of fundamental differences to Rasch measurement theory.

### **Overview of Measurement in Social Sciences**

Cattell (1893) noted the need for measures of mental phenomenon and made comparisons with measurement within the physical sciences, stating “the history of science is the history of measurement”, and “we may affirm without hesitation that quantity is the beginning and end of material science” (p. 316). It seems that at the end of the nineteenth century, quantitative study of psychological phenomenon was only beginning the course that the physical sciences had begun centuries earlier. The scientific discoveries and advancements gained in many fields of science over that long period were not easily won and were progressed slowly over generations. Cattell acknowledged that the science of psychology was understandably the last of all the sciences to develop advanced measurement – because of the obscure and complex nature of mental phenomenon. Like the other sciences in early stages of development, the science of psychology was often associated with speculations, anecdotes, basic observations and artificial classifications.

Kuhn (1961/1977) traced the early developments of quantitative measurement in fields such as astronomy, optics and mechanics, to changing attitudes in the seventeenth century. Other fields such as temperature, electricity, magnetism and chemistry have an even longer history reaching back to ancient times. For all these examples, Kuhn emphasizes that breakthroughs in productive measurement are

preceded by a long tradition of theory development and qualitative study of phenomenon. Based on the history of science, Kuhn concludes:

The road from scientific law to scientific measurement can rarely be traveled in the reverse direction. To discover quantitative regularity one must normally know what regularity one is seeking and one's instruments must be designed accordingly; even then nature may not yield consistent or generalizable results without a struggle. (p. 219)

Regarding the advancement of measurement in the social sciences, Kuhn observed that at his time of writing in the mid twentieth century:

...the fundamental agreement which physicists, say, can normally take for granted has only recently begun to emerge in a few areas of social science research. Most other areas are still characterized by fundamental disagreements about the definition of the field, its paradigm achievements, and its problems. (p. 223)

A committee of the British Association for the Advancement of Science (representing Mathematical and Physical Sciences, and Psychology) was appointed in 1932, and debated over the problems of measurement for a period of eight years (Stevens, 1946). According to Stevens, the members were divided, with some arguing that measures of subjective experience were meaningless unless the concept of addition could be applied to the variable being considered. This principle, the requirement of additivity, was considered necessary for *fundamental measurement*, a widely accepted theory made popular by Campbell in 1928 (van der Linden, 1994).

Stevens responded to the committee's report with an argument for a broader definition of measurement, and for a variety of forms of measurement. He suggested that the "most liberal and useful definition of measurement [is]...the assignment of numerals to things so as to represent facts and conventions about them." (p. 680). He proposed a structure of classification of four types of measurement scales and the corresponding statistics that are admissible for each. In Table 2.4, each classification is listed, with relevant empirical operations and permissible statistics. The statistics column is cumulative, such that all listed statistics are permissible for data that are of a ratio scale.

## 46 Rasch Analysis of the PRFQ

Table 2.4.

*Scales of Measurement, Based on Stevens Classifications of Scales*

Scale	Basic Empirical Operations	Permissible Statistics
NOMINAL	Determination of equality	Number of cases Mode Contingency correlation
ORDINAL	Determination of greater or less	Median Percentiles
INTERVAL	Determination of equality of intervals or differences	Mean Standard deviation Rank-order correlation Product-moment correlation
RATIO	Determination of equality of ratios	Coefficient of variation

Note: Table from Stevens (1946, p. 678)

Stevens makes clear that most psychological data have the property of an ordinal scale, and therefore the statistics that are permissible with such data should be limited to those appropriate for nominal and ordinal scale data. He stated:

A classic example of an ordinal scale is the scale of hardness of minerals. Other instances are found among scales of intelligence, personality traits, grade or quality of leather, etc.

As a matter of fact, most of the scales used widely and effectively by psychologists are ordinal scales. In the strictest propriety the ordinary statistics involving means and standard deviations ought not to be used with these scales, for these statistics imply knowledge of something more than the relative rank-order of data. On the other hand, for this 'illegal' statisticizing there can be invoked a kind of pragmatic sanction: In numerous instances it leads to fruitful results. While the outlawing of this procedure would probably serve no good purpose, it is proper to point out that means and standard deviations computed on an ordinal scale are in error to the extent that the

successive intervals on the scale are unequal in size. When only the rank order of data is known, we should proceed cautiously with our statistics, and especially with the conclusions we draw from them. (p. 679)

Criticisms of the Stevens classification include an inadequate reference to the substantive issues of the property being measured, and its relationship to the numerical assignment (Borsboom & Scholten, 2008) – a problem that can lead to a conclusion such as Lord's (1953) playful saying that "...the numbers don't remember where they came from, they always behave just the same way, regardless." (p. 751). Also, Michell (1997a, 2002) identifies a lack of detail in the specified conditions for scale classification, and that this has resulted in Stevens's theory being used to support superficial psychological measurement – rather than being used to test the theory or investigate if a variable can be measured. The general acceptance of Stevens's broad and vague definition of measurement, opened the way for theories and arguments by authors such as Likert (1932) and Lord to provided rationales for what is now the common use of parametric statistics with ordinal scale data (Hobart, Cano, Zajicek, & Thompson, 2007).

The Rasch model is based on measurement models developed by Georg Rasch (1960) and offers a very different approach to the use of data compared to the approach by Stevens. In the following section, the Rasch model is shown to offer a theory and means to address some of the primary challenges facing social science measurement, in particular measurement using rating scales.

### **Assumptions and requirements for measurement**

Technology and methods available for analysing data have improved exponentially over the past few decades, with increasingly sophisticated computer software making complex analysis relatively easy. However, there seems to be a corresponding decline in the interest in quantitative training and critique of analysis methods within psychology education curriculum (Embretson, 1996, 2006; Osborne, 2010). With such an array of powerful and easily accessible tools for analyzing data, now more than ever, researchers need to be wary of the quality of their data, and the rationale for using the chosen model or method of analysis. Andrich (1989) makes the point that from early in the history of standard psychometric methods, the

requirements for measurement were reframed as assumptions, which were subsequently not considered necessary assumptions.

### *Likert's assumptions*

Quantitative measurement in the social sciences is often undertaken using rating scales. Likert (1932; Likert, Roslow, & Murphy, 1934) first made popular the summing of scores from a number of rated items, and the use of this sum to represent the measurement of a *latent* trait – referred to as latent because the variable of interest is not directly observable and its measurement is inferred from the observable ratings of items. For example, the 12-item General Functioning Subscale of the McMaster Family Assessment Device (FAD; Epstein, Baldwin, & Bishop, 1983; Kabacoff, Miller, Bishop, Epstein, & Keitner, 1990) is a commonly used Likert type scale. The GFAD offers a range of statements regarding family functioning that each have four response categories labelled strongly agree (SA), agree (A), disagree (D) and strongly disagree (SD). A statement of positive family functioning would typically be scored in the following manner, SA = 4, A = 3, D = 2, and SD = 1. A statement reflecting family dysfunction or poor functioning would be reverse scored: SA = 1, A = 2, D = 3, and SD = 4. Each person's scores from the 12 items are then added to provide their total GFAD score of family functioning. These scores are assumed to reflect levels of the latent trait 'family functioning'. Once a number of scores such as these have been collected from a sample people, they are ready to be analysed using a range of statistics. Table 2.5 presents hypothetical scores for two items of the GFAD and a summed total score.

Table 2.5.

### *Hypothetical Responses to Two GFAD Items*

Item statement	Strongly agree	Agree	Disagree	Strongly Disagree	Item Score
We avoid discussing our fears and concerns	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	3
There are lots of bad feelings in our family	<input type="radio"/>	<input type="radio"/>	<input checked="" type="radio"/>	<input type="radio"/>	3
<b>Total summed score for two items</b>					<b>6</b>

Although Likert’s method of summed ratings seems straightforward and perhaps reasonable, it is as one might say, ‘too good to be true’. This method makes a number of assumptions about the data and provides no means to examine the truth of these assumptions. The ratings on items clearly represent an order, however the ‘distance’ or interval between each response category on the continuum of the trait is assumed to be equal. Also when the item scores are summed, there is an assumption that each item score of the same number represents the same ‘amount’ of the trait. To illustrate how these assumptions are likely to be inaccurate, consider the following hypothetical placement (Table 2.6) of how family members may actually respond to the different categories – aligned under the theorized continuum of family functioning.

Table 2.6.  
*Hypothetical Placement of Categories Along a Continuum of Family Functioning*

	Low Family Functioning	----- trait continuum-----			High Family Functioning
We avoid discussing our fears and concerns		SA	A	D	SD
There are lots of bad feelings in our family	SA	A	D		SD

If the categories were functioning as per the table above, a response of D to both items is clearly not representing the same level of the trait. Also, it is clear that the difference between A and D, is not the same as D and SD for either item. The Likert approach to summated rating assumes that these issues are inconsequential.

Linacre (2002) also recognised this problem and noted that Likert scales are typically presented as being equal categories that presume equal importance and reflect an equal division of the underlying variable (see Figure 2.6); however, from a measurement perspective, Linacre pictured the range of responses quite differently, and suggests they could look more like Figure 2.7. The two distinctive differences between these two visual representations of scale categories are that each category can vary in size across the continuum of the latent variable, and that the extreme end categories are infinitely wide, since no matter how much someone agrees, there is the possibility of someone who agrees more strongly.

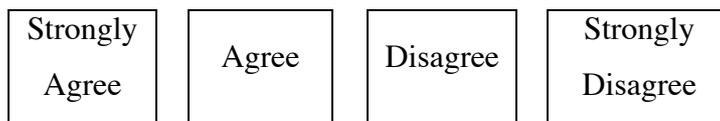


Figure 2.6. Typical Likert Scale categories as presented to the respondent.

Strongly Agree	Agree	Disagree	Strongly Disagree
----------------	-------	----------	-------------------

Figure 2.7. Typical Likert scale categories from a measurement perspective

While Likert’s methods were appealing for reasons of convenience, others such as Thurstone and later Guttman proposed requirements and methods of measuring psychological variables, which raised the bar and more closely resembled the science of physical measurement.

***Thurstone’s requirements***

At around the same time as Likert, Thurstone (1928, 1929) was making early attempts to measure attitudes, although with much more rigorous methods. Thurstone identified a number of requirements for social measurement. Based on Thurstone’s work, Andrich (1989) has drawn attention to four essential aspects of constructing a scale: the definition of a continuum, the use of a statistical model to test the consistency of the results, additivity and invariance. Each of these is briefly expanded upon below:

1. Measurement implies a linear continuum along which one can say there is ‘more’ or ‘less’ of a trait or construct. This rules out measurement of some attributes, and Thurstone provided the example of opinions on prohibition, which he conceded were multidimensional and that they cannot all be represented along a single continuum. Alternatively, a more specific attitude, such as the degree of restriction that should be imposed on individual liberty in the consumption of alcohol, could be placed on a linear continuum.

2. Prior to analysis and independent of any data set, a statistical measurement model can be developed that can then be used to check if the data met the necessary requirements for measurement.
3. Thurstone (1928) recognised the requirement of additivity, as indicated in the following statement: “The scale is so constructed that two opinions separated by a unit distance on the base line seem to differ as much in the attitude variable involved as any other two opinions on the scale which are also separated by a unit distance.” (p. 529-530)
4. Thurstone (1928) also recognised the requirement of invariance for measurement, insisting that: “A measuring instrument must not be seriously affected in its measuring function by the object of measurement. To the extent that its measuring function is so affected, the validity of the instrument is impaired or limited.” (p. 547)

Likert referred to Thurstone’s methods of scale construction as “exceedingly laborious” that made “unnecessary statistical assumptions” and that his own methods were a “radical departure” from those of Thurstone’s (Likert, 1932, p. 6). The difference in the approach and methods by Likert and Thurstone were contentious at the time, and continued to be debated in the field of social measurement over a decade later (McNemar, 1946).

Andrich (1989, p. 11) criticised the a-theoretical approach of Likert to variable construction and has contended that Likert totally misunderstood Thurstone’s methods. Thurstone’s principles were presented as *requirements* for measurement not *assumptions* to be considered. His statistical models formalized these requirements independently of data and therefore did not require verification with data. The importance of Thurstone’s work is not so much the procedures he used, but the primacy he placed on the role of measurement and scale development theory, and the use of a statistical model to test the theory requirements (Hobart & Cano, 2009). In this respect, Thurstone anticipated some of the most important properties of the Rasch model.

### **Guttman scaling**

An approach to scaling and measurement was introduced independently of Rasch by Louis Guttman in the 1950s (Andrich, 1985, 2002a; Engelhard, 2008;

## 52 Rasch Analysis of the PRFQ

Fisher, 2000). This section summarises the fundamental principles of Guttman scaling, which provides a useful foundation for understanding the Rasch model.

A number of terms will be used throughout this thesis that are common to Guttman scaling and the Rasch model. The perfect hypothetical Guttman scale requires items to range in *difficulty* in a hierarchical ordering, and respondents also to be hierarchically ordered by their level of *ability*. The terms ‘difficulty’ and ‘ability’ are used to convey the item and person levels of the construct that are the focus of the scale. Alternatively items could be said to have a level of *agreeability* on the trait continuum, and persons could vary in levels of a trait or attitude, without the items being difficult or easy, and without the responses necessarily being correct or incorrect. Both items and persons are understood to have a *location* on the same continuum of the latent trait.

As an example of Guttman scaling, consider a dichotomously scored four-item scale created to measure a trait, with a correct response scored 1 and incorrect response scored 0. All possible responses are displayed in Table 2.7. Only 5 out of 16 sets of response data fit the Guttman pattern. For example, a person who scores a 2 out of 4 (their ability score) would have to respond correctly to the two least difficult items and incorrectly to the two most difficult items. This pattern of responses is what would be expected. As can be seen from the Table 2.7, any of the other five unexpected combinations of responses for a total score of 2 would not meet the expected order, and one would rightly question if the total score could justifiably be used to represent performance on the trait. When all responses are ordered as expected, the total score is the *sufficient statistic* that fully and accurately conveys the responses that contribute to that score.

Table 2.7.

*Dichotomous Response Matrix Conforming to the Guttman Scale*

Items (in order of difficulty or location on trait)				Total Score
1	2	3	4	(person "ability")
Expected Responses				
0	0	0	0	0
1	0	0	0	1
1	1	0	0	2
1	1	1	0	3
1	1	1	1	4
Unexpected Responses				
0	0	0	1	1
0	0	1	0	1
0	1	0	0	1
0	0	1	1	2
0	1	0	0	2
0	1	1	0	2
1	0	1	0	2
1	0	0	1	2
0	1	1	1	3
1	0	1	1	3
1	1	0	1	3

Guttman's requirements are a deterministic approach to scaling. The probability of a dichotomous response to a single item is graphically represented in Figure 2.8 (Diagram based on Andrich & Styles, 2004), such that the probability is on the vertical axis and the location on the continuum of trait (both item difficulty and person ability) is on the horizontal axis. A person located below the item location has no chance of affirming the item (0 probability), whereas a person located higher than the item location is certain to affirm the item (probability of 1).

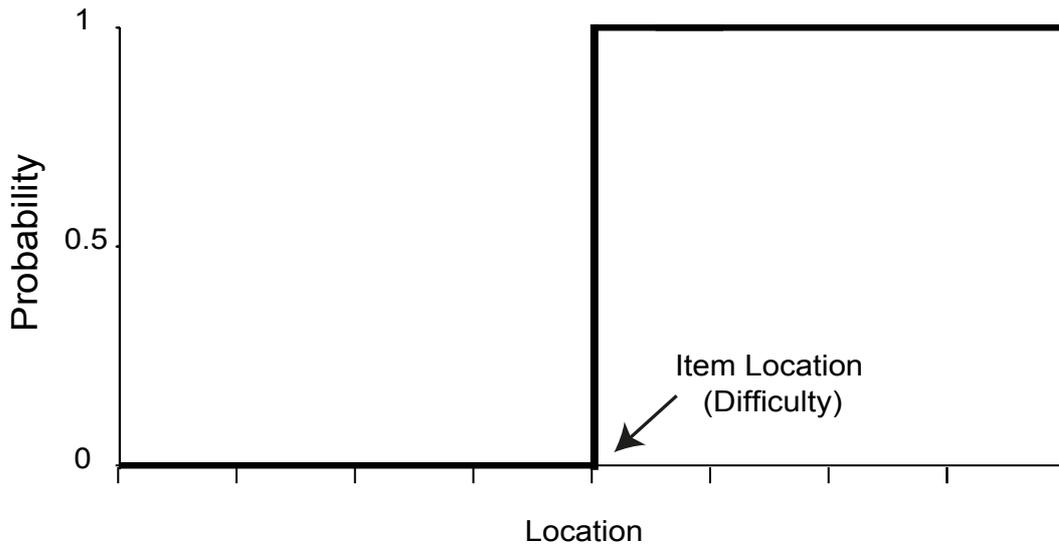


Figure 2.8. Guttman response probability for a single item with a dichotomous response.

When a number of items are mapped out along the continuum such as the four in Figure 2.9, an instrument with greater precision is created. With responses to these four items, a person can now be located in one of the five areas along the same continuum as the item locations. For example, a correct response to Item 1 and an incorrect response to the other three items, fits the model and locates the person's trait level in the location of the "1" along the trait continuum.

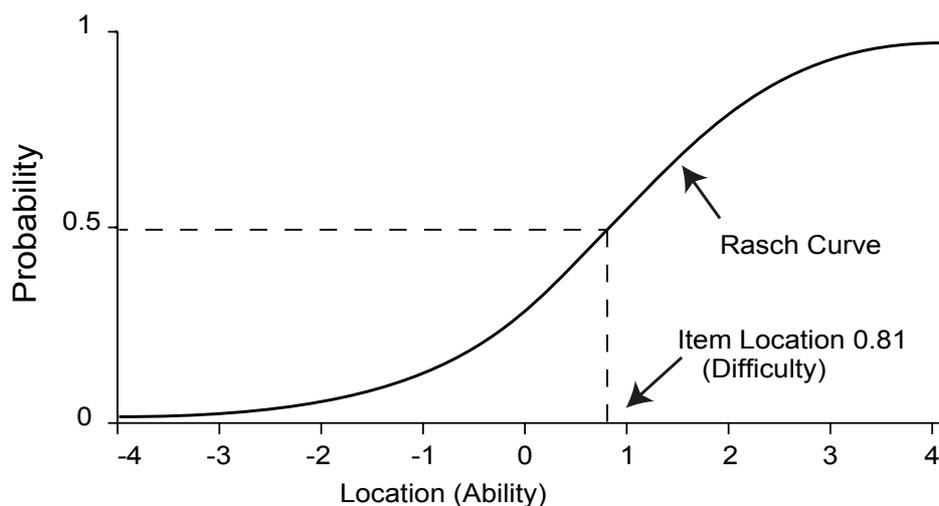


Figure 2.9. Guttman response probability for four items.

### Rasch model as a probabilistic variant of Guttman scaling

The Rasch model is a probabilistic form of Guttman scaling (Andrich, 1985). Although the mathematics of the Rasch model will not be detailed in this thesis, the following section provides an overview of the properties of the model in way that requires minimal background knowledge. The mathematical expressions of the Rasch model can be found in numerous publications: the original dichotomous Rasch model was first published by Georg Rasch (1960, 1961) with thorough rationale of the mathematical equations provided, and subsequent developments by others have been well documented (Andrich, 1988a; Bond & Fox, 2007; Wright & Stone, 1979).

Compared to Guttman's deterministic model, the probabilistic Rasch model is more representative of data from social science rating scales. The expected response probabilities to an item are non-linear and produce the Rasch curve, referred to as an Item Characteristic Curve (ICC), as seen in Figure 2.10. This curve shows the probability of affirming an item as a function of the ability of the person. As the ability estimate of the person (location) increases the probability of affirming the item increases. The difficulty estimate of the item located at the point where there is equal likelihood (probability of .5) that the person will affirm the item (at location 0.81 in this case). As the person location increases higher than the item location, the probability rises. Because there is never certainty that a person would always or never affirm an item, the curve never reaches probabilities of 1 or 0.



*Figure 2.10.* Item Characteristic Curve for affirming a single item with a dichotomous response.

Figure 2.11 shows the probabilities of both an incorrect and correct response (agreeing and disagreeing) to an item. In this case the location of the item is the threshold of where the probability of an incorrect response is equal to the probability of a correct response.

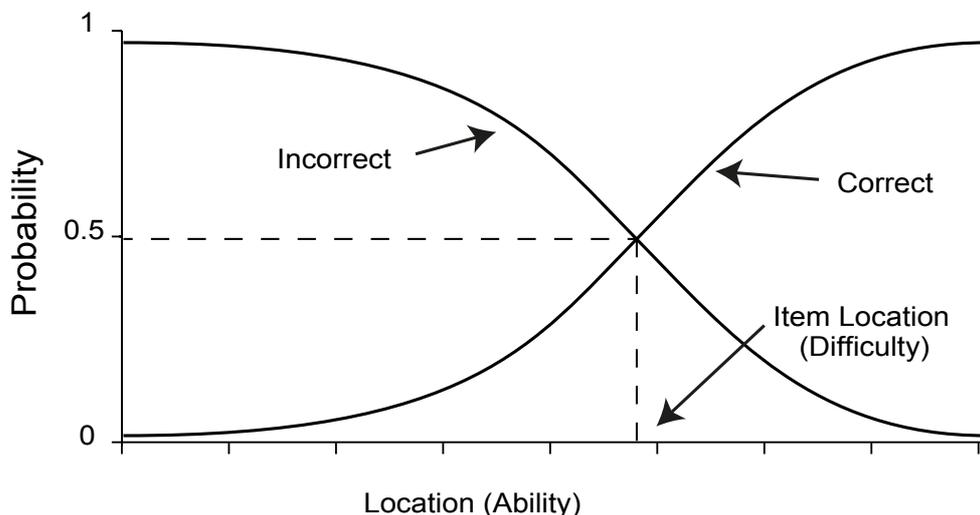


Figure 2.11. Item Characteristic Curve for a correct and incorrect response to a single item.

When the probabilities of four items are plotted according to the Rasch model the ICC appears as seen in Figure 2.12. Each item difficulty level is located on the trait continuum where the probability of affirming the item is at 0.5.

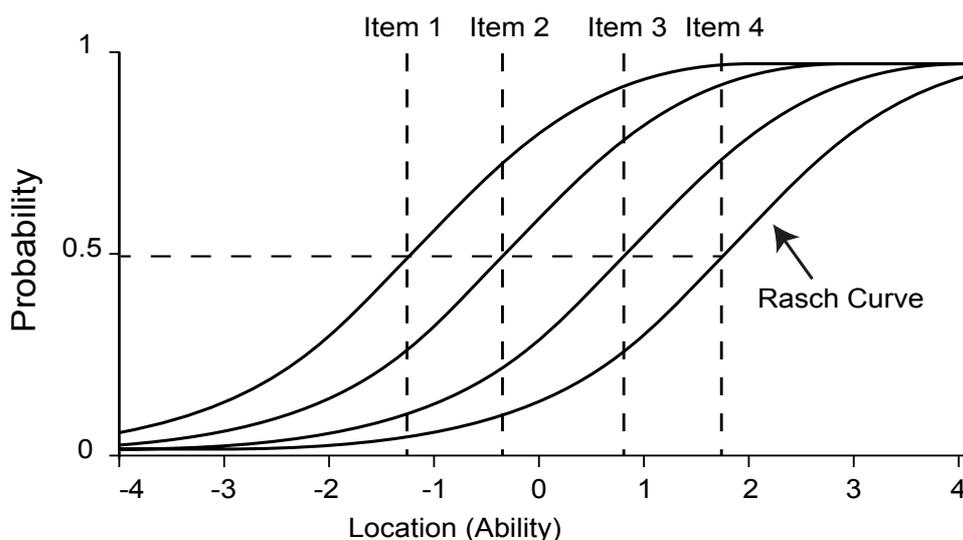


Figure 2.12. Item Characteristic Curve for four items

Now consider the case of a Likert-type scale – a rating scale that has items with three or more categories. A perfect Guttman distribution of responses to four items each with four categories would appear as in Table 2.8.

Table 2.8.

*Polytomous Response Matrix of Four Items with Four Categories, Conforming to the Guttman Scale.*

	Items (in order of difficulty or location)				Total Score (person “ability”)
1	2	3	4		
Expected Responses					
0	0	0	0	0	
1	0	0	0	1	
2	0	0	0	2	
3 (2)	0 (1)	0 (0)	0 (0)	3	
3	1	0	0	4	
3	2	0	0	5	
3 (3)	3 (2)	0 (1)	0 (0)	6	
3	3	1	0	7	
3	3	2	0	8	
3 (3)	3 (3)	3 (2)	0 (1)	9	
3	3	3	1	10	
3	3	3	2	11	
3	3	3	3	12	

Note: Patterns in parentheses also conform to the Guttman pattern with the same probability and total score.

Figure 2.13 is a plot referred to as the Category Probability Curve, which shows the probabilities of responses to a single five-category item. Notice how five categories provide four locations (shown with a vertical dotted line) at the thresholds of where the probability of responding to one category crosses and falls below the probability of the next category.

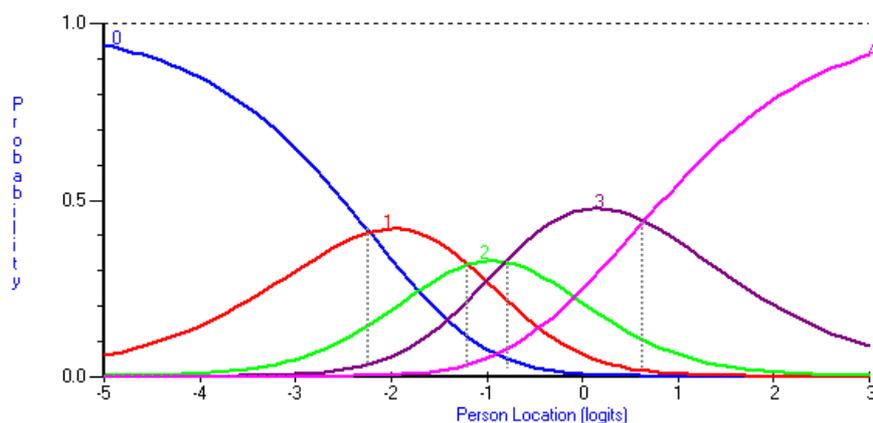


Figure 2.13. Category Probability Curve for a five-category rating scale item.

When analysing data to examine fit to the Rasch model, the Item Characteristic Curve and the Category Probability Curve provide important sources of evidence of data fit/misfit and anomalies in the data that require explanation or further investigation. Essentially, Rasch analysis provides a means to determine how well a set of data conforms to the requirements of measurement that are inherent in the probabilistic form of the Guttman scale, which is the Rasch model. This approach to measurement and to data is unique and in contrast to the traditional or commonly used psychometric methods of scale development and evaluation.

### Two approaches to modern measurement

Andrich (2002b, 2004, 2011) proposes that there are two approaches to the data-model relationship in the development or evaluation of measurement instruments.

1. *Model fit to the data – Statistical or Psychometric Approach:* There is no a priori restriction on the model choice, and consequently models with more parameters are likely to account for the data better than models with fewer parameters.
2. *Data fit to the model – Experimental Approach:* A model is chosen a priori, to best account for the data and meet specific criteria to provide measurement. These criteria and the model specify the parameters required for measurement. The case for the model is not dependent on any data set.

The essential difference between these two approaches is in the relationship between the data and the model used to help clarify the instrument and understanding

of the construct and participants. The statistical approach has no predetermined class of models that will be used, but rather the model is determined by which one best fits the data with the least number of parameters (Andrich, 2004). The experimental approach applied with Rasch analysis has specific a priori restrictions on the class of models and the parameters in these models that can be used, and these are independent of the set of data to be examined. Therefore the assumptions and conditions regarding the data that apply to the statistical approach are not necessarily relevant for Rasch analysis approach.

The statistical and experimental approaches represent two paradigms of test theory. Kuhn (cited in Andrich, 2004, 2011) used the term paradigm to describe a set of mutually reinforcing theories or approaches that underpin a school of science. Experiments and programs that take a paradigm for granted and are therefore limited to a specific view are within the bounds of *normal science*. An alternative paradigm would by definition be incompatible with the existing paradigm and likely to be controversial. The controversy is not only because of the break from the tradition, but because a new scientific school operating from a different paradigm is incommensurable with the paradigm of the original school. Those debating the benefits of their own paradigm may not acknowledge the irreconcilable differences of their approaches to research, and consequently the opposing parties talk through each other and misunderstandings persist. The following sections of this paper present the rationale for considering Rasch Measurement Theory as a paradigm of test theory with a number of key benefits over other Classical Test Theory, and that is unique in its experimental approach to the data-model relationship.

### ***Classical test theory***

The dominant paradigm for the development or evaluation of measurement instruments in social sciences uses the statistical approach of selecting models to explain the data, and is referred to as *traditional* or *classical test theory* (CTT). A foundational assumption of CTT is that the observed score a person achieves on a scale is comprised of a hypothetical true score plus associated measurement error (Spearman, 1904). Although many advances have been achieved using CTT methods, this approach has a number of fundamental problems. Hobart and Cano and colleagues (Cano & Hobart, 2011; Hobart & Cano, 2009; Hobart et al., 2007) have

## 60 Rasch Analysis of the PRFQ

highlighted a number of the limitations common to CTT methods, which are summarised in the following points:

- Both the hypothetical true score and the error score are unobservable variables that are poorly defined. Therefore the theory and equations based on these variables are difficult (or impossible) to test or challenge, and the assumptions are met by most datasets. For these reasons CTT has also been referred to as Weak True Score Theory.
- Raw rating scale scores are only ordered counts (ordinal scale data) and this limitation is not addressed or overcome with CTT. Rather, as previously discussed, the limitations on statistical methods with ordinal data are often waived without adequate rationale.
- The scale performance is completely dependent on the sample, making the measurement unstable since it is influenced by the sample being measured.
- Likewise, a person's level of performance is dependent on the particular choice of scale or set of items being used. Because of this, a person's measurement cannot be confidently compared to others from another sample, or to their own if they are within two different samples, or to their own on two scales assessing the same trait (refer to Invariance section).
- Complete data are required for common CTT methods, and techniques for substituting values for missing data are dubious (see following Missing Data section).
- The standard error of measurement (SE) in CTT methods is estimated as a constant value, rather than being specific to different scores across the continuum of the scale or level of trait. The SE is also sample dependent and typically large, therefore making individual assessment impractical (see following Error section).
- Methods from CTT do not provide a unidimensional scale on which to locate items or the scores from persons (see following Summed Score Sufficiency section).

These and other issues regarding the failings of CTT and mainstream psychometrics have been argued by numerous authors (Borsboom, 2006; Cliff, 1992; Embretson, 1996, 2006; Fisher, 1998; Kline, 2000; Michell, 1997a, 2002, 2009; Woods, 2011). A number of alternative approaches to measurement have developed in parallel to CTT, and to varying degrees these approaches address the limitations of CTT. How the Rasch Measurement Theory compares to these other approaches and how it is useful in addressing the shortcomings of CTT is discussed in the following section.

### ***Modern test theory***

*Rasch measurement theory* (RMT) is often classified within a broad group of approaches to measurement that are distinct from CTT – referred to as *modern test theory* (Andrich, 2011) and dominated by *Item Response Theory* (IRT) (Bock, 1997). The Rasch model (i.e. the mathematical model) can be considered one of the models within IRT, although RMT has distinct differences to IRT, which are detailed later in this section. The primary advantages of IRT (inclusive of RMT) over CTT are summarized well by Embretson (1996; Embretson & Reise, 2000), in her comparison between what she called the old and new rules of measurement, referring to CTT and IRT respectively. These comparisons are included in Table 2.9, which compares aspects of the two approaches to measurement. Also in the table is the corresponding numbered section of this thesis, which provides further details of each issue.

Table 2.9.

*Advantages of Item Response Theory Over Classical Test Theory*

Issue addressed in following section	Old Rule (CTT)	New Rule (IRT)
1. Error of Measurement	The standard error of measurement applies to all scores in a particular population.	The standard error of measurement differs across scores, but generalizes across populations.
2. Reliability	Longer tests are more reliable than shorter tests.	Shorter tests can be more reliable than longer tests.
3. Missing Data	Missing data is imputed using person specific mean scores (based on an untestable assumptions).	A person's score is calculated only from available data (completed items).
4. Test Equating	Comparing test scores across multiple forms depends on test parallelism or adequate equating.	Comparing scores from multiple forms is optimal when test difficulty levels vary across persons.
5. Sample Dependence and Scale Dependence	Unbiased assessment of item properties depends on representative samples from the population.	Unbiased estimates of item properties may be obtained from unrepresentative samples.
6. Meaningfulness of Measurement	Meaningful scale scores are obtained by comparisons of position in a score distribution.	Meaningful scale scores are obtained by comparisons of distances from various items.
7. Interval Scale Properties	(Implicit) Interval scale properties are achieved by selecting items that yield normal raw score distributions.	Interval scale properties are achieved by justifiable measurement models, not score distributions.

Table adapted from Embretson (1996)

In addition to these advantages of IRT over CTT, RMT is distinct from other IRT approaches in a number of ways, which are summarized in Table 2.10 and expanded upon in the thesis sections indicated within this table. Firstly, RMT gives primacy to the model (compared to primacy of the data), and this is a fundamental

difference between RMT and all other approaches. This distinction has been explained in the description of the statistical and experimental approaches outlined earlier in this thesis: RMT is an experimental approach that has a priori specifications for model choice that meet the requirements for providing measurement. Secondly, rating categories are expected to function (be responded to) in a particular way, so as to reflect increasing levels of the latent variable with higher category ratings, and this hypothesis is tested by inspecting the category probability curves for disordered thresholds. Thirdly, maintaining the property of invariance is central to the requirement for measurement. And finally, the sufficiency of the total score is applicable when the data fit the Rasch model and therefore meet the requirement of invariance.

Table 2.10.

*Unique Properties of Rasch Measurement Theory*

Issue addressed in following section (continued from previous table)	Rasch Measurement Theory
8. Primacy of the Model	The inherent properties of the Rasch model satisfy the requirement of additivity for fundamental measurement. An experimental approach is taken, with data required to show fit to the model that is chosen <i>a priori</i> .
9. Threshold Order	Category thresholds are examined for expected ordering to test the hypothesis that item categories are responded to as would be expected.
10. Invariance	The person parameters and item parameters can be estimated separately, on a common metric.
11. Summed Score Sufficiency	If the data fit the model, the totalled scale score is the sufficient statistic, and therefore no information from the response pattern is needed.

These distinctions between RMT and other approaches to psychometric data analysis have been noted by other authors in the context of social sciences (Bond, 2004) health (Fisher, 1998; Hobart & Cano, 2009), health economics (Cano et al., 2010), marketing (Salzberger & Koller, In Press) and education (Andrich, 2010a).

### **Distinctive Qualities of Rasch Measurement Theory**

The following numbered sections correspond to the issues summarized in Tables 2.9 and 2.10. Each section calls to attention a specific quality of RMT, which arguably makes it a valuable alternative to other data analysis methods, and in some cases distinct from or advantageous to CTT and IRT methods.

#### **Error of measurement**

Calculation of the standard error of measurement (SE) is critical for judging the precision of a measurement instrument and interpreting results from measurements. With CTT methods of calculating the SE, the error is assumed to be the same for all scores and is calculated using the standard deviation of the sample (Howell, 2002). However, it is generally accepted that scores towards the extremes (most items correct or incorrect) are less precise than scores closer to the mean (Andrich, 2010a). In contrast to CTT methods, calculation of SEs with the Rasch model is performed for both individual person location estimates and item location estimates, the SE estimates vary across the trait continuum, and the calculation does not rely on sample dependent statistics such as the standard deviation of the raw scores. Therefore the following characteristics are found in the SE estimates from Rasch analysis:

- Error estimates are calculated for individual item and person estimates.
- Since item and person estimates are located along the trait continuum, estimates at either extreme of the continuum have larger SE.
- An item estimate that is distant to person estimates in location will typically have a relatively small SE, and likewise, a person estimate that is distant to item estimates will have a small SE. Conversely, SE estimates are larger when the location difference between person and item is relatively large.

- The SE can be reduced by improving targeting of items to persons, so person estimates have closely matching item estimates.
- When items are well targeted, SE is low and therefore more precise measures are obtained for individuals.
- Item and person specific SEs are particularly useful in calculating confidence intervals for use in testing hypothesized relationships between linked persons or linked items (see In variance section).

### **Reliability**

Scale reliability is commonly indicated by an index of internal consistency, calculated from the proportion of variance of the person distribution to the error variance (Fisher, 2010). Cronbach's alpha ( $\alpha$ ) is an index of reliability often used in CTT methods (Cronbach, 1951; Sijtsma, 2009). A similar index of reliability is calculated using the Rasch model and is known as the Person Separation Index (PSI; Andrich, 1982). The PSI statistic can also be thought of as an indication of how well a measure can distinguish between scores for groups of subjects on the continuum of the trait in question (Fisher, 2010). A scale showing a relatively low psi of .67 will be only sufficient to distinguish between two groups with 95% confidence, whereas a relatively high PSI of .94 provides the possibility of distinguishing between five groups. Other authors have considered high and low PSI as indicating the suitability of a scale for use with groups of scores (minimum of .70) compared with individual use (minimum of .85–.90; Nunnally, Bernstein, & Berge, 1967; Parkitny et al., 2012; Tennant & Conaghan, 2007).

Although the PSI is calculated using similar reasoning and theory as  $\alpha$ , there are a number of differences that are evident between the two because of the properties of the Rasch Model. The following points from the RUMM Laboratory Website (RUMM Laboratory, 2012) highlight the differences between the PSI and Cronbach's  $\alpha$ :

- Cronbach's  $\alpha$  can be calculated with complete data only, while the PSI can be calculated with random missing data. With substantial missing data, the values for the two indices might be different.
- The PSI is based on the estimated locations of the persons, which are non-linear transformations of the raw scores.

- Cronbach's  $\alpha$  is calculated with all scores, including the maximum and minimum scores, while the estimate of the PSI requires extrapolated values for extreme scores. This is because there is no finite estimate for extreme scores.
- When the item difficulties relative to the person proficiencies are misaligned, so that there is a skewed distribution with extreme raw scores, a difference emerges between Cronbach's  $\alpha$  and the PSI: Cronbach's  $\alpha$  remains more constant than the PSI. The reason for this difference is in their respective constructions: Cronbach's  $\alpha$  is based on raw scores while the PSI involves a non-linear transformation of these raw scores. The error variance for persons increases as the scores become more extreme, so with scores close to the extreme, the error variance increases in the PSI while there is no such effect in the construction of Cronbach's  $\alpha$ .

An important implication of calculating SE and location estimates for each person is that the PSI is a more sensitive index than Cronbach's  $\alpha$ . Analysis with the Rasch model enables development of carefully targeted scales and selected items (e.g. adaptive testing; Linacre, 2000), which can lead to creation of relatively short (few items) scales with high reliability (Embretson, 1996). This is in contrast to the common expectation that longer tests are more reliable than shorter tests.

Although an index of reliability such as Cronbach's  $\alpha$  or PSI is a helpful indicator of internal consistency, it does not provide any evidence of where there are departures from modelled expectations (Fisher, 2010). Therefore, when using the Rasch model, the PSI is only one of a number of indicators of fit to the model.

### **Missing data**

The person and item locations, standard errors and fit statistics of the Rasch model are all based on observed data only, and "the measurement structure specified by the model needs only enough data to identify a finite estimate for each person and each item." (p. 140; Wright & Stone, 1999). Data can also be analysed with structurally missing data when resolving Differential Item Functioning (DIF) or comparing subtests (Andrich & Hagquist, 2012). Unexpected missing data should be considered further for possible influence of issues such as fatigue due to testing,

confusion or bias regarding the item, or possible effects of the testing conditions or format of the questionnaire. Of particular interest could be missing data that have some commonality between persons (e.g. mostly one gender, mostly at the end of the questionnaire, mostly a particular item). Bond and Fox (2007) suggest that when missing data are assumed to be random and have no particular systemic cause, no interpolation based on average or typical scores is required, only sufficient density of data to provide links throughout the data set. The substitution of a value in the place of a missing response makes assumptions about person responses that cannot be tested. For example, using the mean of the person's completed responses assumes (questionably) the difficulty level of the item, or that all the items are of the same level of difficulty (Hobart & Cano, 2009).

The ability to use data with missing values makes other analysis possible, such as creating an item bank and then selecting subsets of items that are targeted for groups of persons thought to have differing levels of the trait. For example, two year groups of students could be given different sets of items from the same bank of test items, with the higher group not required to complete the easiest items and the lower group not required to complete the hardest items - thus reducing burden on the test takers. Similarly, a clinical and non-clinical group could provide self-report on an aspect of mental health. Each group's data could have values missing for items that are not applicable; however, all the data could be analysed as one dataset.

### **Test equating**

Embretson (2006) emphasized the difficulty of comparing scores from two tests that were developed to measure the same latent trait. For example an 'easy' and 'hard' test were compared in a simulation (Figure 2.14), and as would be expected the hard test shows a floor effect (many scores of 0) and the easy test a ceiling effect (many scores in the high 20s but few above 30). The relationship is not linear and the amount of equating error is large.

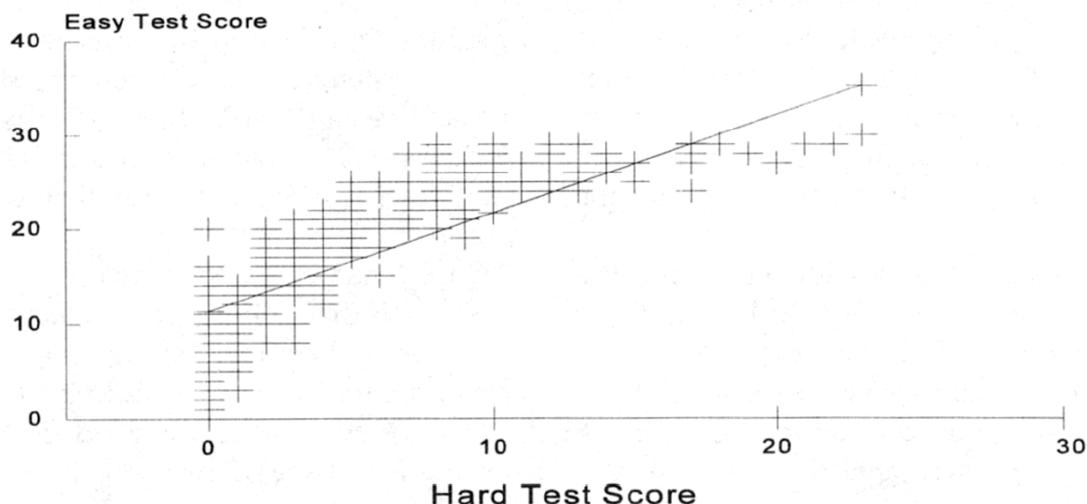


Figure 2.14. Linear equating of tests simulation (from Embretson, 1996, p. 344)

In contrast to the correlation method of CCT shown in Figure 2.14, item estimates from a Rasch analysis of two scales can be plotted to assess how well they can be used interchangeably. An example is provided by Bond and Fox (2007) in a comparison of the Bond's Logical Operations Test (BLOT) and the Piagetian Reasoning Task (PRTIII). The plotted results from testing 150 children can be seen in Figure 2.15. The standard errors of the individual person scores (not available with traditional/classical statistical methods) are used to create the 95% confidence lines. These lines make clear the lower precision at the extremes of low and high scores (larger error terms). If all the plotted locations fall within the SE confidence lines, this would indicate that the person estimates are invariant when parents are assessed for their level of development on the two different scales - indicating that their scores are measuring the same latent trait (allowing for error). In this example several cases fall outside the SE confidence lines and therefore examination of the possible reasons for this need to be examined. The property of invariance in the Rasch model is discussed in greater detail later in this section.

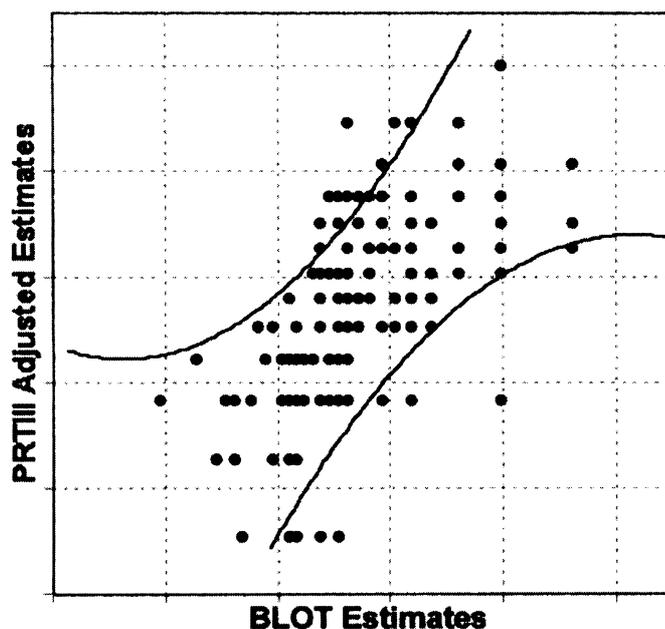


Figure 2.15. Common Person Linking Plot (from Bond & Fox, 2007, p. 87)

### Sample dependence and scale dependence

A foundational assumption of CTT methods is that the sample is representative of the target population (Embretson, 1996), and therefore the performance of the scale (reliability and validity) is dependent on the sample used for its evaluation. The converse is also true. The set of items used in a scale is assumed to be representative of all items that could measure the particular trait, making a person's measurement dependent on the set of items in the scale.

The Rasch model, however, is not dependent on assumptions about the population (Rasch, 1966) and the item and person parameters are estimated separately (Hobart & Cano, 2009). This separation of the parameters is the basis of objective measurement or invariance (see Invariance section). As clearly stated by Thurstone (1928) "A measuring instrument must not be seriously affected in its measuring function by the object of measurement. To the extent that its measuring function is so affected, the validity of the instrument is impaired or limited" (p. 547). Parameter separation is a mathematical property of the Rasch model, and only if data from a measure are found to adequately fit the model can the person locations be considered independent of the item set used, and the item locations considered independent of the sample used. Once measures with this quality are developed, subsets of items can be used and these scores compared, as with computerized adaptive testing (Linacre, 2000).

**Meaningfulness of measurement**

Embretson (1996) notes that in CTT the meaningfulness of a score is found in its relative position in a distribution of scores from a norm referenced group. One major limitation of this requirement for measurement is the lack of connection between a person’s score and their performance on the individual items. As shown earlier in the explanation of the Rasch model as a probabilistic form of Guttman scaling, the person’s trait estimates are based on the probability of their responses to individual items, and this is a more meaningful measure than the simple summing of their item counts. With the Rasch model, both persons and items are calibrated onto a common scale (Andrich, 1990), which can be portrayed easily in an item map, shown in Figure 2.16. The item map gives a visual display of both item and person distributions, such that the individual item or threshold locations and order can be observed, and matched to the sample distribution. The example in Figure 2.16 is from a three-category scale, and therefore two threshold locations have been estimated for each item.

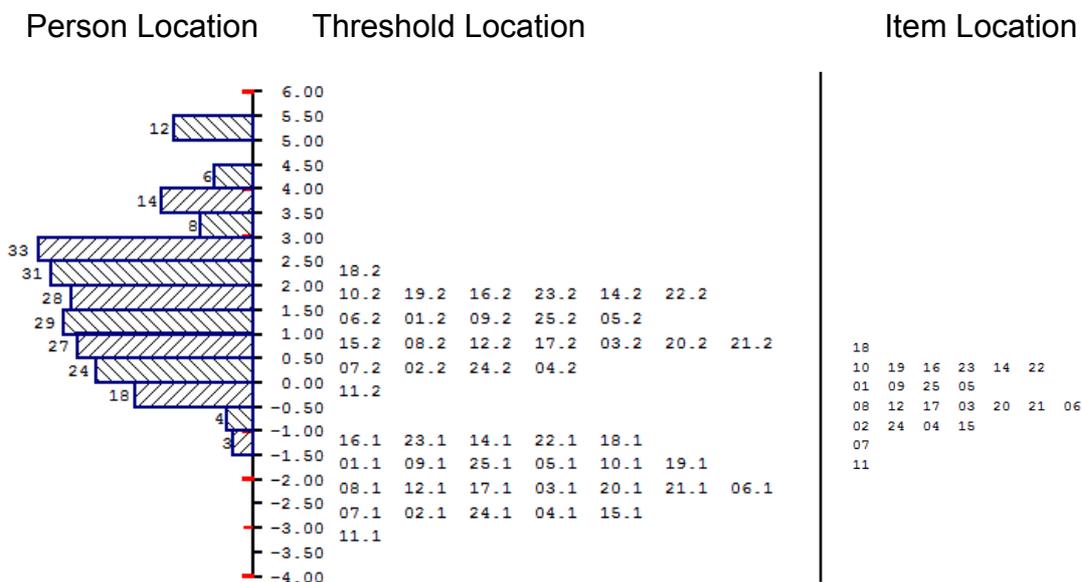


Figure 2.16. Example of an item map showing persons and items calibrated onto a common scale.

The threshold locations provide the basis for calibrating the scale, like units along a ruler or thermometer, to provide the metric against which measurement is possible (Wright & Stenner, 1999). The unit of measurement in the Rasch Model is

the logit – explained in the next section ‘Interval Scale Measurement’. Figure 2.16 provides an example of how thresholds and person are located on the same scale. The threshold 15.2 is seen to be located at approximately 0.8 on the continuum, and this has a direct meaning for the likely responses of persons to item 15. Those persons located below 0.8 are more likely to rate item 15 in the lower two categories, and persons located above 0.8 are more likely to rate with category 3. The same detailed analysis and information are available for each item and each person, including location estimate, a number of statistics indicating fit to the model, and the standard error of measurement.

Rasch (1961, 1966) asserted that his model addressed a concern that psychometric methods were entirely group-centred and inadequate for meaningful investigations of individuals – a criticism that remains true for CTT methods. Analysis using the Rasch model is truly individual centred, providing clear and meaningful comparison between an individual’s responses and the individual items (assuming data fit to the model and with consideration of measurement error).

Detailed and persuasive arguments for adopting the Rasch model for meaningful measurement in the social sciences have been published by Fisher (1998, 2000, 2004, 2005, 2010). Fisher makes strong calls for improved practice of measurement within the social sciences with particular reference to developing more meaningful, reliable, accountable and adaptive measures through application of the Rasch model.

In a similar vein to the search for more meaningful measurement, there is need for greater integration between psychometrics and substantive psychology. Cano and Hobart (2011) argue that the current methods of ensuring the validity of rating scales are mostly inadequate, and therefore we cannot be sure of what scales are measuring. Many researchers concerned with social measurement have called for more qualitative assessment and consideration of substantive issues associated with measurement scale construction or evaluation, application and interpretation of results (Borsboom, 2006; Michell, 2009).

Although many aspects of validity have been proposed in the past, more recently there has been a general consensus that construct validity is the overarching and unified type of validity, which includes many different aspects of validity (Messick, 1989). The varied aspects of validity provide the multiple sources from

with one can accumulate evidence of construct validity (Wolfe & Smith, 2007a, 2007b).

Embretson (2007) proposed a model for a universal system of construct validity of educational and psychological tests. From the model below (Figure 2.17) it can be seen that the test specifications and psychometric properties of a test play a critical and central role in the process and these are the result of a number of processes that together provide internal/meaning evidence of validity. Also important is the feedback loop of the external/significance categories of evidence back into all aspects of the internal/meaning categories – informing continual improvement and re-evaluation of construct theory and testing procedures.

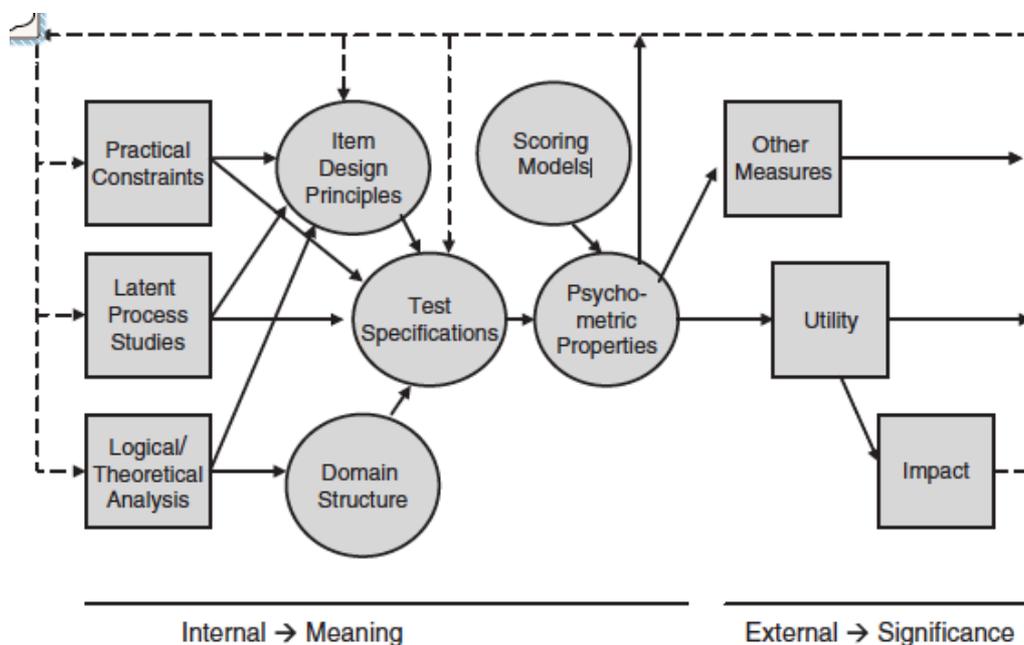


Figure 2.17. Embretson's universal system for validity (from Embretson, 2007, p. 453)

Scale development should ideally be a bottom-up approach beginning with sound development of construct theory and operationalization, rather than using top-down methods of item grouping and looking for correlations with other measures. These top-down approaches are a poor means of establishing a rationale for the meaning and utility of a score – the essence of construct validity (Cano & Hobart, 2011; Embretson, 2007). When psychometric statistics are used without careful ongoing consideration of the qualitative meaning of the scale and its scores, they are

in danger of being misleading and unscientific, and likely to lead to unethical or adverse social consequences (Cano & Hobart, 2011; Messick, 1989).

Embretson (2006) argued for greater effort to develop nonarbitrary metrics in psychology – such that scores from a scale have a meaningful location on the continuum of the construct, and that there is meaning associated with a unit change on that continuum. Similarly, Cano and Hobart (Cano & Hobart, 2011) suggest greater use of inductive and deductive approaches to the development of construct theory, operationalization and evaluation. The experimental approach to measurement incorporating the Rasch model meets both of these appeals for what can simply be called meaningful measurement.

### **Interval scale measurement**

As previously discussed, psychological rating scale data are simply ordered counts and summed scores that provide an ordinal scale of measurement. However, an interval scale is required for summing items to provide a total score and for common statistical analysis such as calculating means, variances and Pearson correlations (Stevens, 1946). With the Rasch model, the scale of location estimates meets the requirements of an interval scale – ratios of difference can be calculated, such as one difference being twice that of another. The units of measurement in the probabilistic Rasch model are called *logits*, which are the logarithmic odds of affirming an item – calculated from the difference between the person location and the item location.

In the Rasch model, the ordinal raw scores relate to the logit measurement units in an S-shaped (ogival) fashion as can be seen in Figure 2.18. Notice in this figure how the difference between scores 10 and 11 compares to the difference between scores 15 and 16. The score differences at the lower and higher ends of the spectrum represent much larger intervals than those around the centre. The scores between 5 and 15 approximate a linear relationship with the logit intervals and therefore with well-targeted samples raw scores can have a high correlation with Rasch estimates (Wright & Stone, 1999) – as was argued by Likert (Likert et al., 1934) in his rationale for using summed ratings. However, this similarity is superficial and ignores the requirements for measurement. These requirements are discussed further in the Invariance section.

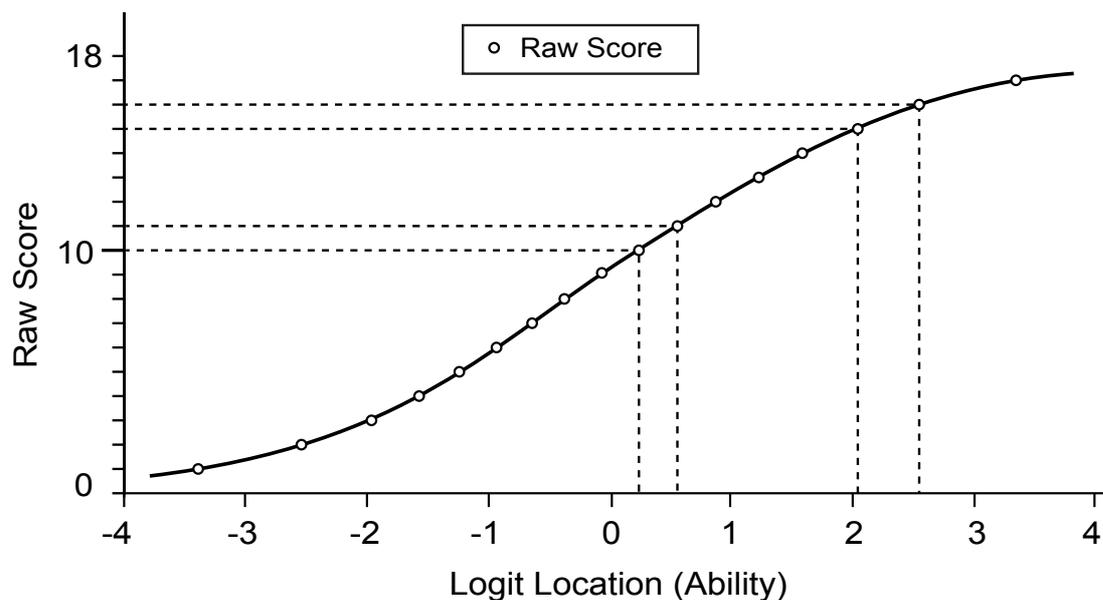


Figure 2.18. Relationship between raw scores and measures

### Primacy of the model

As already described in earlier sections, Thurstone defined a number of requirements for measurement. Although IRT has its origins in the work of Thurstone, the focus for IRT is Thurstone's application of mathematical models to social measurement. In addition to the application of similar mathematics, RMT remains true to Thurstone's principles of giving primacy to the theory and the model. With RMT, the data from a scale are evaluated against the requirements of measurement, which are theoretically sound and are found in the Rasch model. The practical benefit of this approach is that detailed analyses with the Rasch model reveals the specific occurrences of misfit or anomalies in the data.

It may be thought that a scientist uses measurement to produce data in the search for theories, however Khun argued the opposite. The function of measurement is to reveal anomalies in data, given that the data was produced by theories in which the scientist is already confident (Khun, 1970, as cited in Andrich, 2004). In a similar way, the Rasch Model represents a sound theory of measurement that is independent of the data. When data are based on an instrument (such as a rating scale) that has a sound substantive theory there is reason to be curious regarding any data that do not

fit the Rasch model, and it is imperative to investigate the specific circumstances of the misfit so as to propose some meaningful explanation. This diagnostic approach to analysis provides ways for researchers to examine the performance of individual or subsets of items and persons, gain insight into the nature of the latent variable being investigated, and opportunity to refine the definition and measurement of such variables (Wright, 2000). The diagnostic features of the Rasch model provide unique opportunities to investigate anomalies and departures from what we expect, and thus provide an opportunity to make new discoveries (Fisher, 2004).

### Threshold order

Rating scale measurement implies ordered rating categories, which correspond with increasing or decreasing levels of the variable being measured (Andrich, 1998). In measurement with the polytomous RM, the probability of each item category being affirmed is plotted against the sample's estimated person location of the attribute as a category probability curve (see Figure 2.19). As the person estimates increase, each category in turn should be the most probable one to be affirmed. The probability curves at each extreme of the scale extend indefinitely, approaching a probability of 1, since the possibility of a higher or lower ability/capacity is always possible. The thresholds of each category curve provide the measurement locations, where the probabilities of affirming one category or the other are estimated to be 50:50. When an item is functioning as would be expected in the RM, the thresholds estimates are sequentially ordered. Disordered threshold estimates (Andrich, de Jong, & Sheridan, 1997) indicate that the response categories have not been responded to as would be expected according to the RM.

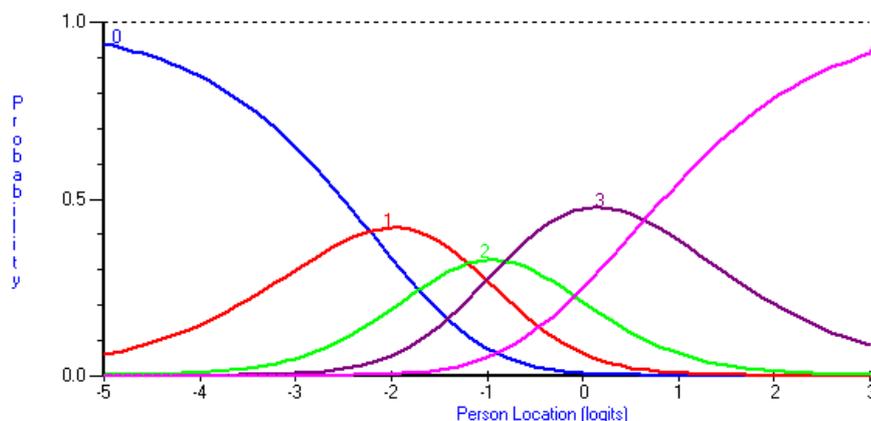


Figure 2.19. Example of a category probability curve of a five-category item.

The assumption that rating scale categories function in an orderly way that reflects the attribute being measured is often taken for granted and not able to be tested using statistical analysis other than examination with the RM (Andrich, 2011). Some researchers using the Rasch model have argued that ordered thresholds are not necessary (Adams, Wu, & Wilson, 2012; Linacre, 1999), although these arguments are based on a paradigm that prioritizes the statistical modeling approach and the fit of the model to the data, rather than the primacy of substantive issues and the mathematical model (Andrich, 2012, 2013). The requirement of ordered thresholds is also a source of difference between the Polytomous RM and the polytomous IRT model, the Graded Response Model (GRM). The polytomous RM partitions the underlying continuum of responses to generate measurement locations, whereas the GRM partitions the frequency distribution of responses (Andrich, 2010b). The GRM does not address reversed thresholds and assumes categories can be joined without affecting the model.

The ordering of thresholds is an unknown property of the data prior to analysis, and therefore the expectation that thresholds are ordered is a hypothesis that needs to be investigated with the RM - the only model that has the property of being able to test this hypothesis (Andrich et al., 1997). Identifying disordered thresholds in itself does not suggest why the responses do not conform to the expected model and Andrich (2011) argues “the solution to the problem needs to be substantive, empirical and experimental. The analysis cannot reveal the source of the problem, only the location of the problem.” (p.581). When thresholds are found to be disordered, there is potential to identify where the operationalisation of the variable may be inadequate, and the anomaly in the data can be further investigated qualitatively (Andrich, 2002b).

What are some possible reasons why thresholds might be disordered? Although the items in a questionnaire may be carefully developed to assess a particular variable, there is the possibility that the respondents have a poorly defined understanding of the variable (Linacre, 2002), or even a different definition due to cultural or developmental reasons. The use of too many categories is a common problem that can result in respondents showing inability to distinguish between similar categories. A number of problems associated with using a middle category with labels such as “unsure”, “undecided”, “do not know”, or “not applicable” (Enos,

2001), with the obvious issue being this category often cannot be conceptualized as representing a place on a continuum between the two adjoining categories. The meaning respondents associate with a central category can vary widely and qualitative research (Kulas & Stachowski, 2013) has shown that its use was more strongly related to the item than to the person's level of the trait – and that the most common meaning attributed to the middle category was “it depends”.

### **Invariance**

The principle of invariance is central to the Rasch model and a requirement of achieving objective measurement. A number of authors have detailed the uniqueness of the Rasch model in its relation to objective measurement (Engelhard, 1994, 2008; Wright & Stone, 1999). Although it is not a commonly addressed requirement for psychological measurement, the properties of invariance are generally assumed in everyday physical measurement. Consider a ruler and a tape measure - two instruments from a class of instruments that measure length - both instruments are expected to have comparable measures of the same one object. Also, any two objects (from a class of objects) are expected to have measures of length that can be compared. The nature of invariance in rating scale measurement is such that the comparative location of any two people on a trait continuum does not depend on the set of items used to assess them, and likewise, the comparative location of any two items on the same trait continuum does not depend on the sample of people who respond to those items.

Rasch (Rasch, 1960) referred to the property of invariance as *parameter separation* or *specific objectivity*, and explained this aspect of his model in the following way:

The comparison between two stimuli should be independent of which particular individuals were instrumental for the comparison; and it should also be independent of which other stimuli within the considered class were or might also have been compared.

Symmetrically, a comparison between two individuals should be independent of which particular stimuli within the class considered were instrumental for comparison; and it should also be independent of which other individuals were also compared, on the same or on some other occasion. (Rasch, 1961)(p.332).

The requirement of invariance that is inherent in the Rasch model provides the probabilistic ideal against which anomalies in the data can be highlighted. Once identified, these anomalies in the performance of either persons or items can be investigated further with analysis of the deviation from the model or error. Evidence from such analysis can provide the basis of reasons for the misfit to the model.

The seemingly unattainable standard of direct fundamental measurement of psychological properties can appear to prove measurement to be largely impossible in the social sciences. The primary problem is that most psychological variables of interest are conceptualized constructs – latent variables that are inferred by a person's behaviour. The additivity that is possible with physical measurement such as length (e.g. one metre rods can be placed end to end and demonstrate the additivity of the unit) is not applicable with a latent variable. However, Luce & Tukey (1964) were able to identify the requirements for *conjoint measurement*, which enable an indirect way to confirm if there is additive structure in data from latent variables. This approach to fundamental measurement is also referred to as derived measurement, because the dependent variable is modelled by its relationship with independent variables (van der Linden, 1994). The Rasch model has been shown to meet the requirements of conjoint measurement (Brogden, 1977; Newby, Conner, Grant, & Bunderson, 2009; Wright, 1985) although this conclusion is not without controversy and critique (Heene, 2013; Karabatsos, 2001; Kyngdon, 2008; Michell, 1997b)

### **Summed score sufficiency**

The total raw score (sum of all item responses) is the key statistic for both CCT and the RM. However, the summed score in CCT is assumed to adequately represent a unidimensional latent trait, and this assumption is untestable. Whereas, with the RM the total score is considered the *sufficient statistic*, and is testable by the fit of the data to the unidimensional Rasch model (Andrich, 2010a, 2011). The sufficiency of the total score is comparable to the requirement of invariance, as one implies the other. The attribute of sufficiency can be observed when the response matrix conforms to the Guttman scale. When data fits a model with the requirement of invariance, the total score provides a statistic that completely describes the data. As previously discussed, the Rasch model is a probabilistic variant of the Guttman scale and therefore a perfect fit of the data to the model is not expected.

### **Rasch Measurement Literature**

A number of papers by Rasch (1960, 1961, 1966) provide a detailed account of the development of his model and personal overviews of the life and background of Georg Rasch have been provided by two leading advocates of the Rasch model, Andrich (2005) and Wright (1998). Other authors have provided reviews of the theory and development of the Rasch model, including: “Applying the Rasch Model: Fundamental Measurement in the Human Sciences”, by Bond and Fox (2007); “Objective Measurement: Theory into Practice (Vol. 2)”, edited by Wilson (1994) (Vol. 1-5 also available); and “Best Test Design: Rasch Measurement” (1979) and “Measurement Essentials” (1999), by Wright and Stone. Recent journal articles that specifically address the rationale and application of the Rasch model have been valuable resources in the writing of this literature review (Andrich, 2011; Cano & Hobart, 2011; da Rocha, Chachamovich, de Almeida Fleck, & Tennant, 2013; Fisher, 2000, 2004; C. M. Fox & Jones, 1998; Hagquist, Bruce, & Gustavsson, 2009; Hobart & Cano, 2009; Hobart et al., 2007; Royal, 2010; Tatum, 2000; Wright, 1997, 2000). In addition, the website “Rasch Measurement Transactions” ([www.rasch.org](http://www.rasch.org)) provides an archive of articles addressing a broad range of issues associated with the Rasch model. The mathematics of the Rasch model is explained in detail in a number of publications (Andrich, 1978; Hobart & Cano, 2009; Rasch, 1960), and has intentionally not been addressed in this thesis.

### **Analysis of Data using the Rasch Model**

This Part B of the literature review has provided a rationale for the use of Rasch measurement theory as a valuable and effective means to evaluate important aspects of validity of the Parental Reflective Functioning Questionnaire. The twelve distinctive qualities of the Rasch measurement theory that have been addressed in this section can be examined in a dataset using a range of different techniques with computer software such as RUMM2030 (Andrich, Sheridan, & Luo, 2012). These techniques are outlined in the Research Methodology section of this thesis under the following headings.

## 80 Rasch Analysis of the PRFQ

Firstly:

- i. Parameterisation and Choice of Rasch Model
- ii. Sample size and category frequencies

Then:

- a. ordered response category thresholds,
- b. targeted item/person distributions,
- c. overall scale fit to the Rasch model,
- d. reliability,
- e. individual item and person fit to the Rasch model,
- f. invariance in individual item functioning between persons with membership of different groups (differential item functioning),
- g. local item independence,
- h. unidimensionality, and
- i. variance in PRFQ scores accounted for by categorical parental characteristics.

### Chapter 3 Research Methodology

The design of this research was a cross-sectional study of Parental Reflective Functioning (PRF) using two instruments: the self-report Parental Reflective Functioning Questionnaire (PRFQ) and the semi-structured Parent Development Interview (PDI) – scored for reflective functioning (PDI-RF). The participants were parents recruited into the Peel Child Health Study (PCHS), and assessments were completed when their child was approximately 12 months of age. The PRFQ was examined for conformity to the requirements of the Rasch measurement model. The PCHS presented a unique opportunity to examine this measure of PRF with a relatively diverse sample of non-clinical mothers and fathers.

This chapter begins with a description of the participants and the questions or measures that were used to obtain data for this study. The data analysis is presented in four sections that correspond to the four research questions:

1. Do data from mothers and fathers PRFQ conform to the requirements of the Rasch measurement model?
2. Is there an association between self-report parental depression or anxiety and PRFQ scores?
3. Do one or more specific sets of PRFQ items show temporal (test–retest) stability?
4. Is there a relation between PRFQ scores and PDI-RF scores that demonstrates convergent validity?

#### Participants

All parents participating in the PCHS with a 12-month-old child were invited to participate in this study of PRF. The PCHS is a longitudinal population study, designed to help identify the conditions that provide children with the maximum opportunity for achieving their developmental potential. The project included collection of psychosocial, environmental, biological and genetic data, with a focus on the complexity of individual behaviours in context. A wide range of community, research and government agencies collaborated to apply multidisciplinary and multilevel research approaches to this project. The Peel Study Principal Investigators have given their approval for this PRF research to be undertaken and the research is consistent with the ethics approval obtained by the PCHS (Curtin Human Research

Ethics Committee: HR 02/2008, and the Murdoch University Human Research Ethics Sub-Committee: Permit Number: 2007/238).

The PCHS recruited 451 families over a period from September 2009 to January 2012. Medical practitioners identified families with a pregnancy, and following consent (see Peel Study consent forms: Appendix B) each family was recruited just prior to the 18th week of pregnancy. Sufficient data were provided by 120 couples (240 participants) from the one-year follow-up questionnaires. Parents who completed these questionnaires were contacted regarding their willingness to be interviewed. Forty of the couples consented to participating in interviews and a second administration of the PRFQ. Analysis is provided to indicate statistical difference between those participants who were recruited into the PCHS and those who participated in this study of PRF (see Results Chapter).

The population of the Peel Region includes a wide range of social and economic stratifications, in addition to unique small communities reflecting diverse industries (Peel Development Commission, 2012). The participants were all English-speaking mothers and fathers living together with the study child at the time of the study.

## **Apparatus and Materials**

### **Demographics, parental mental health, family functioning and child development questionnaires**

Data were used from self-report questionnaires completed at a number of stages by families in the PCHS. Parent's demographics were received from the 18-week expectant parent questionnaires (18-week) or the 12-month old follow-up (one-year). These questionnaires were administered via self-report surveys delivered and collected by the PCHS staff. Demographics of both parents included: age (18-week), country of birth (18-week), education (18-week), occupation (18-week), and if the study child was their first child (mother: 18-week, Father: one-year). The following family characteristics were from the mother questionnaires: child age and gender (one-year), family income (18-week), and English as main language spoken at home (18-week). Some of these demographics were coded so as to simplify the data for analysis and these are described below.

*Country of birth*

Mothers and fathers indicated their country of birth and results were dichotomized into either “Australia” or “Not Australia”.

*Education*

Participants who had completed or were currently undertaking education indicated the course and level of study. Data from the 18-week questionnaire were coded and categorized into four levels of education: Less than Year 12, Year 12, Certificate or Diploma, and Degree or higher.

*Occupation*

Each participant’s description of occupation at 18-weeks was coded according to ANZSCO criteria (ABS, 2013) to create four categories of occupation. ANZSCO occupation categories are as follows: 1 Managers, 2 Professionals, 3 Technicians and Trades Workers, 4 Community and Personal Service Workers, 5 Clerical and Administrative Workers, 6 Sales Workers, 7 Machinery Operators and Drivers, 8 Labourers. The four categories in this study were created using ANZSCO categories 1 and 2, 3 and 4, 5 and 6, and 7 and 8.

*Income*

Family income was reported by the mother in the 18-week questionnaire and data coded into four categories: \$60,000 or less, \$60,001 to \$78,000, \$78,001 to \$104,000, and Over \$104,000.

*First child*

In order to assess prior parenting experience, the following questions were coded as either “First Child” or “Parent has an older child”. Mothers indicated if there were other older children living in their home or if they had given birth to a child that lived elsewhere. Fathers indicated if they were previously a father to another child in addition to the study child, even if the additional child was not biologically related to them.

*Parental self-report scales*

The PCHS questionnaires included two self-report scales that were used as an indication of parental mental health: the Beck Depression Inventory II (BDI) and the State Trait Anxiety Inventory (STAI). The trait anxiety subscale of the STAI was completed with the 18-week questionnaires. The STAI-state subscale and BDI scale were completed at the one-year follow-up. These instruments have demonstrated validity and are standardized measures commonly used in population-based studies, such as the Household, Income and Labour Dynamics in Australia Survey (HILDA) and the Longitudinal Study of Australian Children (LASC). An overview of each measure is provided below, including report of scale validity and reliability.

*Beck Depression Inventory II*

The Beck Depression Inventory II (BDI-II; Beck, Steer, & Brown, 1996; Appendix C), a revision of the original BDI, is a 21-item self-report scale for the assessment of a wide range of symptoms and attitudes associated with depression. The items of the BDI-II address themes such as pessimism, guilt, agitation, loss of energy, worthlessness, sleep patterns and appetite. Each item provides the choice of a number of statements that are scored from 0 to 3, with higher scores indicating a higher level of depression. One statement is chosen to best describe how the person has been feeling during the past two weeks. The depressive symptoms addressed in the BDI-II closely match the diagnostic criteria for depression detailed in the DSM-IV (American Psychiatric Association, 2000; Sprinkle et al., 2002). Although originally validated with a psychiatric population the BDI has been commonly used with non-clinical samples (Sprinkle et al., 2002) and with mothers and fathers in the postpartum period (Gaynes et al., 2005; Magalhães, Pinheiro, Horta, Pinheiro, & Da Silva, 2008; Milgrom, Negri, Gemmill, McNeil, & Martin, 2005).

Validation studies of the BDI-II have mostly been undertaken with the approach of Classical Test Theory. These studies consistently report high internal consistency (Storch, Roberti, & Roth, 2004) and test-retest reliability (Sprinkle et al., 2002). Factor analysis has found the BDI-II items tend to load onto two highly correlated factors labelled Somatic and Cognitive-Affective, although the results have been dependent on the sample studied (Storch et al., 2004; Whisman, Perez, & Ramel, 2000).

Two Rasch analyses of the BDI-II have indicated some items showed misfit and their removal provided a scale that met the requirements of the Rasch measurement model. These studies with neurorehabilitation inpatients (Siegert, Tennant, & Turner-Stokes, 2010) and with stroke survivors (Lerdal, Kottorp, Gay, Grov, & Lee, 2014) both recommended that the following items be removed from the scale: Changes in Sleeping Pattern (16; Sleep), Changes in Appetite (18; Appetite), and Loss of Interest in Sex (21; Sex). Although these analyses were with specific clinical samples, the misfit of the three items is consistent with the previously reported findings of somatic items representing a distinct construct.

The current study's BDI-II scores were examined to determine the utility of using the 18-item BDI-II recommended by (Siegert et al., 2010). Specifically relevant to this current study's sample are the known difficulties with sleep disturbance and changes in sexual intimacy in early parenting (von Sydow, 1999). The 'Sleep' item in this study was on average the highest scored BDI item with the largest variation in scores for both mothers ( $M = 1.03$ ,  $SD = 1.00$ ) and fathers ( $M = 0.66$ ,  $SD = 0.92$ ). A study with patients suffering from insomnia (Carney, Ulmer, Edinger, Krystal, & Knauss, 2009) found the cut-off for mild depression on the BDI has poor specificity and can incorrectly assess depression for patients with insomnia. Also notable was that for mothers, the 'Appetite' item was the second highest in variability ( $SD = .83$ ). A Rasch analysis of the BDI-II data was beyond the scope of this study, so further more meaningful examination of the item responses was not undertaken. In light of these considerations and observations, the recommendations of item removal by (Siegert et al., 2010) were adopted, in an attempt to provide the best possible analysis for the current sample.

Given the non-clinical sample of this study, the distribution of depression scores was understandably skewed towards the lower levels. An examination of missing data revealed 30 mothers and 15 fathers with missing values in their BDI-II responses. Ten mothers and four fathers did not complete any BDI-II items, therefore, no replacement of missing values was attempted for these participants. The remaining participants with missing data were mostly missing only one item, with nine mothers and six fathers missing the 'Changes in Sleep Patterns' item, and five mothers missing the 'Changes in Appetite' item. Chronbach's alpha calculated for the 18-item BDI-II with complete data sets from the 120 couples was slightly higher

than that of the full BDI-II (mothers: .91/.89; fathers = .90/.88), indicating an improvement in internal consistency.

### *State-Trait Anxiety Inventory*

Postpartum mood disorders are commonly assessed with measures of depression, while anxiety symptoms are rarely assessed even though they tend to both co-exist with depressive symptoms as well as present independently of depression (Matthey, Barnett, Howie, & Kavanagh, 2003). The Spielberger State-Trait Anxiety Inventory (STAI; Spielberger, 1983; Appendix D) is a commonly used assessment of anxiety that consists of two 20-item self-report scales, one assessing how the person currently feels (state), the other requires the person to reflect on how they generally feel (trait). Each item on the scales is rated on a 4-point scale from 1 to 4. The state scale ratings are labelled from “Not at all” to “Very much so”, and the trait ratings from “Almost never” to “Almost always”. Examples of state scale item wording are: “I feel tense”, “I am jittery”, and “I feel content”. Examples of trait scale item wording are: “I feel pleasant”, “I have disturbing thoughts”, and “I lack self-confidence”. When scoring, items that indicate positive states, such as “I am a steady person” are reverse scored such that higher scores indicate less experience of positive states and higher anxiety. A score for each scale is obtained by calculating a mean from valid responses.

A review of STAI studies (Barnes, Harp, & Jung, 2002) found that internal consistency reliability estimates were satisfactory for both the Trait scale ( $n = 51$ ;  $M = .89$ ;  $SD = .05$ ) and the State scale ( $n = 52$ ;  $M = .91$ ;  $SD = .05$ ). The test-retest reliabilities, assessing temporal stability, were also satisfactory for the Trait scale ( $n = 7$ ;  $M = .88$ ;  $SD = .05$ ) and the State scale was understandably less stable than the trait scale. Factor analysis of the STAI has shown support for the two dimensions represented by the state and trait items (Kennedy, Schwab, Morris, & Beldia, 2001).

The STAI has been successfully applied in the study of early parenting (Grant, McMahon, & Austin, 2008) with a sample of mothers during pregnancy. Postnatal anxiety and mood disorders were found to be equally well predicted by the self-report STAI and by diagnostic interviews, with scores over 40 on the trait scale associated with a six-fold increase in the prevalence of postnatal anxiety or

depression. Interestingly, antenatal assessments of depression were not predictive of postnatal mood outcomes.

An early Rasch analysis of the STAI (Tenenbaum, Furst, & Weingarten, 1985) was undertaken to measure athlete performance anxiety, and more recently a Rasch analysis was published on a short 6-item form of the scale (Court, Greenland, & Margrain, 2010). Therefore, since no Rasch analysis or other validity studies have indicated alterations that would improve the STAI, the scale was used in this study in its current form.

Missing data for the STAI responses were examined and it was found that some parents did not complete any items from the STAI scales (STAI-Trait: 11 women and 18 men; STAI-State: ten women and three men). In addition, a number of parents did not respond to one item from each of the 20-item scales (STAI-Trait: eight women and seven men; STAI-State: five women and five men), however no item failed to elicit responses from more than two men or women. All missing data were coded and did not contribute to individual final scores. In this sample, the STAI scale demonstrated high internal consistency with Chronbach's alpha for both mothers and fathers (Trait = .93 for both parents, State = .95 for mothers and .94 for fathers).

#### *Parental Reflective Functioning Questionnaire (PRFQ)*

Development of the PRFQ (Luyten et al., 2009; Appendix A) is currently in progress and undergoing reliability and validity evaluation in various samples in Europe and the United States. There were no published validation results at the time of writing this thesis. The current study was initiated with the support of and in consultation with Luyten and colleagues, and will contribute to the validation process. At present, the PRFQ is a 39-item, self-administered questionnaire rated on a 7-point Likert-type scale with categories ranging from *strongly disagree* to *strongly agree*. The questions were designed to relate directly to the mother-child relationship and aspects of maternal RF.

The PRFQ consists of three subscales as described in Chapter 2A. The HL subscale is scored such that higher scores indicate higher levels of PRF and the LH subscale is the opposite with higher scores indicating lower levels of PRF. The M subscale is scored such that the middle category has the highest score and represents

highest PRF, and categories at either extreme of the rating categories have the lowest scores.

Unlike most statistical methods the Rasch model does not require a complete data set and can calculate person locations with some missing data (see discussion of missing data in Chapter 2B). Of the 120 couples that completed the PRFQ, one mother and one father from different couples did not attempt any items; therefore, results on the PRFQ analysis are based on responses from 119 mothers and 119 fathers. The PRFQ responses for the two parents without matching partner data were included in the analysis because they participated in the PDI interviews that provided PRF scores for the test of convergent validity. Fourteen mothers and eight fathers were found to have partially incomplete data, which in most cases was from only one or two items, and was in no case more than half the items of the PRFQ. There were no items or groups of items were missed by most of these participants. Because Rasch analysis is able to accommodate missing data, all of the available data from other questionnaires of the 120 couples were used in the analyses.

#### ***Parent Development Interview Revised Short Version (PDI)***

The PDI (Slade, Aber, Berger, Bresgi, & Kaplan, 2003; Appendix E) is a semi-structured clinical interview designed to explore parents' representations of their children, themselves as parents, and their relationships with their children. The revised version of the PDI is less age specific than the original and was developed to allow for assessment of PRF. This version has 33 questions, with scoring for PRF being identical in structure and organisation to the process developed by Fonagy et al. (1998) for the scoring of RF from the Adult Attachment Interview (AAI). The overall score of RF is rated on an 11-point scale, from negative 1 to positive 9. Negative RF represents actively hostile resistance to mentalizing, bizarre or paranoid attributions, with a total lack of any reflection on mental states. High RF is rare and characterised by exceptional and highly sophisticated reflection, which integrates a number of different aspects of RF into a unified perspective.

The two interviewers (DC and DG) had qualifications and experience relevant for the task of interviewing the parents. Detailed instructions were followed that ensured a common procedure for conducting the PDI (Appendix E). The interviews were approximately 60 minutes in duration and audio recorded at the parent's home.

The recordings were transcribed verbatim and subsequently scored for PRF according to the Addendum to the Reflective Functioning Scoring Manual Version 2.0 (Slade, Bernbach, Grienenberger, Levy, & Locker, 2005).

Scoring of the de-identified transcripts was undertaken by an independent research associate (JL) who completed specific training in scoring the PDI for PRF, and achieved an acceptable level of reliability based on scoring a sample set of transcripts (intraclass correlation  $>0.70$ ). The principal author (DC) and supervisor (LP) have both also received this training and demonstrated acceptable reliability. The research associate (JL) and the author (DC) demonstrated high inter-rater agreement (96.4%) with the overall PRF scores in a concurrent study with a random selection of 30 PDIs from a sample of 100.

### **Procedure**

All parents who participated in the PCHS completed the PRFQ soon after their child was 12 months old. Data were collected from questionnaires completed at 18 weeks of pregnancy and when the study child was approximately one year old. From all the PCHS participants, 40 couples agreed to participate in the PDI. A participant information sheet (Appendix F) was provided to clarify their participation in this PRF study. Following completion of the consent form (Appendix G), the interview took place at the family's home at a time arranged in consultation with the participants. Immediately following the interview the participants completed a retest of the PRFQ. This interview and retest were administered within two months of the initial administration of the PRFQ.

### **Data Analysis**

In the following sections, each research question is presented with details of the analytic techniques used to answer the question. Part B of the Literature Review complements the description of the Rasch analyses and provides more detailed explanation of many concepts specific to Rasch measurement theory.

**Research Question 1: Do data from mothers and fathers PRFQ conform to the requirements of the Rasch Measurement Model?**

Analysis was performed using RUMM2030 computer software (Andrich, Sheridan, & Luo, 2012), which assisted in examining the data from the PRFQ according to Rasch Measurement Theory and provided evidence of data fit to the Rasch model. The development of the Australian Early Development Index (AEDI) is a useful example of applying the Rasch model for the purpose of evaluating rating scale instruments. Rasch analysis was used to examine the psychometric properties of the Early Development Index (EDI) with Australian data (Andrich & Styles, 2004a, 2004b), resulting in a modified version of the five scales (Brinkman, Silburn, & Lawrence, 2006). Rasch analysis was used to determine the optimal number of response categories for items, fit of the data to a unidimensional model, and identification of items that were not performing as expected (including potential implications/benefits of their removal). Pallant, Tennant and colleagues (Pallant & Tennant, 2007; Shea, Tennant, & Pallant, 2009; Tennant & Conaghan, 2007) have similarly applied and reported the primary and relevant aspects of Rasch analyses in the evaluation of psychological and health rating scales.

The author undertook advanced training in Rasch analysis and the use of RUMM2030 software. In preparation for the analysis of the PRFQ, Rasch analyses were performed on the mother and father responses to the 12-item General Functioning subscale of the Family Assessment Device (Epstein, Baldwin, & Bishop, 1983). A journal article based on this analysis has been published and is included as Appendix H (Cooke, Marais, Cavanagh, Kendall, & Priddis, 2015).

Based on these examples, and the author's training in Rasch measurement theory, the psychometric properties (validity and reliability) of the PRFQ were analysed by addressing the following issues:

Firstly:

- i. Parameterisation and choice of Rasch model
- ii. Sample size and category frequencies

Then:

- a. Ordering of response category thresholds
- b. Targeting and distribution of items and persons

- c. Overall fit to the Rasch model
- d. Reliability
- e. Item and person fit to the Rasch model
- f. Differential item familiarity
- g. Local item dependence
- h. Dimensionality
- i. Differences between groups of persons

The theoretical background for these procedures is provided in Part B of the Literature Review chapter of this thesis. The process used to evaluate the PRFQ was iterative using multiple sources of evidence. The analyses were undertaken in the order indicated, although each analysis was not considered in isolation and in some cases data were re-analysed on the basis of subsequent results. The process of iterative analysis is presented following a description of each individual procedure.

#### *i. Parameterisation and choice of Rasch model*

The *polytomous* Rasch model (Andrich, 1978) is applicable to items with Likert-type response scales using three or more response categories, such as the seven category PRFQ items. RUMM2030 offers two parameterisations of the Polytomous model: the *partial credit* (also called the unrestricted model; Masters, 1982) and the *rating scale model* (Andrich, 1978). These models differ in that data suitable for the rating scale model parameterisation must have items with the same number of categories (therefore the same number of thresholds) and the threshold distance is assumed to be uniform across all items in the scale. In contrast, the partial credit parameterisation is suitable for use with items of varying category sizes or varying numbers of categories, and is therefore advantageous in scales where ‘partial credit’ is given to items partially correct such as some proficiency tests. The rating scale parameterisation of the model is advantageous when all the items have the same number of categories (e.g., in attitude questionnaires such as Likert-style questionnaires).

In order to access the most appropriate model of parameterisation, Fisher’s Likelihood-ratio Test was performed in RUMM2030. A significant statistic from this test indicates the unrestricted (partial credit) model contains more information than

the rating scale model, and further analysis should be continued with the unrestricted model (Andrich, 2010). A non-significant statistic indicates data can be analysed using the more efficient rating scale model. It is important to note that both these models conform to all the requirements and assumptions of the Rasch model; as stated by Andrich (2002), “at the level of the response of one person to one item, there is no difference between these models, either in their structure or the response processes they can characterize” (p. 355).

### *ii Sample size and category frequencies*

Linacre (1994, 2002) calculated that a sample of size of 50 (99% confidence of  $\pm 1$  logit) would be sufficient to produce statistically stable measures of data to model fit, and 30 (95% confidence of  $\pm 1$  logit) for well-designed pilot studies. Linacre points out that in the case of polytomous scales there is more information provided in each item compared to dichotomous items, so potentially the sample size could be smaller; however, he also cautions that adequate observations are required for each category within each item (at least 10 is suggested), otherwise estimations of the thresholds could be imprecise or unstable (Linacre, 1999).

The ideal of having a certain minimum frequency in all categories may not be possible for substantive reasons and needs to be investigated if not achieved (e.g., a scale measuring symptoms of a severe psychological disorder with a random community sample is likely to have very few extreme ratings affirming these symptoms). Of greater need for consideration, is whether the persons are responding in the categories expected, based on their location estimates (see section a. Ordering of response categories thresholds). Category frequencies are reported and discussed, with consideration of any impact on measurement precision and stability.

### *a. Ordering of response category thresholds*

A common assumption when analysing data from rating scales is that the item categories are responded to in an ordered fashion, such that sequential categories correspond to increasing levels of the trait (however, there are cases of alternatives to sequential scoring of categories which are referred to as unfolding models; Andrich, 1988). The ordering of thresholds is examined prior to any examination of data or analysis of fit to the Rasch model (as explained in detail in the Literature Review).

Disordered thresholds indicate the responses are not consistent with the expected ordering of the item categories. Collapsing the data into fewer categories by rescoreing the categories is a post hoc procedure that can facilitate further examination and analysis. However, consideration should be given to the possible cause of disordered thresholds and the substantive issues associated with collapsing categories (such as the meaningfulness of the category labels).

After considering response category frequencies, distributions, and substantive reasons for category disordering, examination of the Category Probability Curves provides evidence of where the problem of disordered thresholds lies and which categories might be collapsed to remedy the problem. However, collapsing categories unnecessarily or further than is necessary will reduce the scale fit to the Rasch model (Andrich, de Jong, & Sheridan, 1997). Further research with the modified scale would be necessary to confirm the validity of an altered scale. An example will now be provided to illustrate the process of creating ordered thresholds through collapsing categories.

In Figure 3.1, the category probability curve provides an example of disordered thresholds. The highlighted section between thresholds numbered 2 and 1 indicates that there is no range of person locations in which a score of 1 is the most probable response, and this is also observed as disordered thresholds. The data in Figure 3.1 were rescored such that categories scored 0 and 1 were both scored 0, and the next two categories scored 1 and 2, resulting in three possible scores and two thresholds. Figure 3.2 shows the Category Probability Curve after rescoreing with categories 1 and 2 collapsed, and the result is that the thresholds are ordered such that for increasing levels of person ability (location) there are regions of higher probability of responding to each item category.

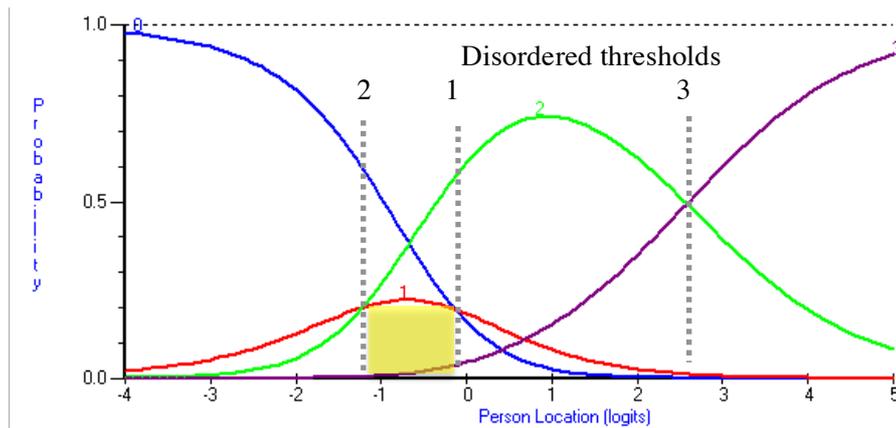


Figure 3.1. Category probability curve showing disordered threshold of a four-category item.

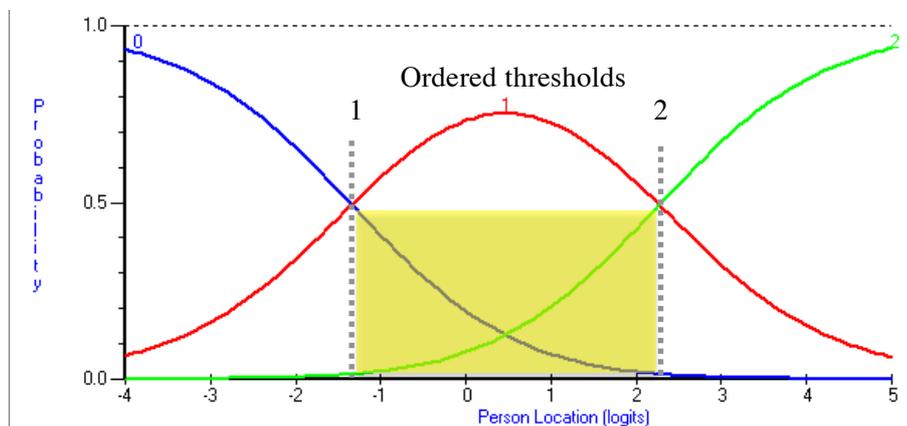


Figure 3.2. Category probability curve showing ordered thresholds after rescaling first two categories as 0.

### ***b. Targeting and distribution of items and persons***

Extreme scores occur when a person receives a “perfect” high score or a zero score for all attempted items. Rasch locations are based on probabilities in relation to the other scores, and an extreme score has no outer reference point and therefore provides no indication of how high the person’s level of the underlying trait could be. The estimate of location parameter for persons with extreme scores is “derived [from the] geometric mean algorithm which uses, respectively, the three highest person location estimates [for the items attempted] and the three lowest person location estimates [for the zero score]” (RUMM Laboratory, 2012). The RUMM2030 analysis identifies extreme scores and provides the choice to either

include or not include these scores when viewing statistics and graphical displays. With many extreme person locations, the targeting of the scale and variability across the scale may be influenced and this should be considered prior to further Rasch analysis. When extreme scores are found to be influencing the results of the analysis, both statistics and graphics with and without extreme persons included will be considered.

The Rasch model scale is centred on a mean of zero logits for the average difficulty estimates of the items. A well-targeted scale would ideally also have a mean of zero for the person location estimates. If the mean of person locations was positive, this would indicate that the sample generally found the scale items easy to affirm, signifying (if the data did not misfit the model) that on average they have high levels of the construct being measured.

The power to detect misfit of data to the model decreases as the distance (in logits) between person and item estimates increases. Likewise the standard error of measurement for person estimates typically increases where there are few or no items targeted at a similar logit estimate – making it more difficult to distinguish between persons of similar ability when there are no items of corresponding difficulty level. Graphical displays and statistics of estimate means and standard deviations are reported to provide an indication of targeting and distribution, in addition to standard errors.

### *c. Overall fit to the Rasch model*

RUMM2030 provides an overall statistic of data fit to the Rasch model – the item-trait interaction chi-square statistic (chi-square). This statistic reflects the agreement among persons with similar levels of a trait (their total score) regarding the level of difficulty of items and the ordering of the difficulties. A significant chi-square (probability of less than .05) indicates that the data does not fit the model as expected and the hierarchical ordering of the items is inconsistent for persons across levels of the trait.

If chi-square shows evidence of data fit to the model, this evidence alone is not necessarily enough to confirm the construct being measured is unidimensional. The Rasch model is a unidimensional model independent of the data set, so if the data fits the model according to the chi-square statistic, it is conforming to the

unidimensional requirements of the model. However, more detailed analysis of dimensionality is better established using principle components factor analysis of item residuals (detailed in section h).

#### *d. Reliability*

RUMM2030 provides a person separation index (PSI) as an indication of internal consistency reliability which is comparable to Chronbach's  $\alpha$  (see Reliability section of the Measurement in Social Science Literature Review chapter). A comparison of  $\alpha$  and PSI can be made if persons with missing data are removed from the data set. This comparison is particularly relevant if floor or ceiling effects are observed in the distribution of scores, or similarly if there is a high proportion of extreme scores (RUMM Laboratory, 2012). Chronbach's  $\alpha$  is also checked when the degree of dimensionality is being determined (see section h).

#### *e. Item and person fit to the Rasch model*

Graphical and statistical evidence of data fit to the Rasch model are available in RUMM2030 for individual items and persons. The Item Characteristic Curve (ICC) is a plot of class intervals (averages based on score estimates, and the proportion of correct responses) against the expected Rasch curve. Items that under-discriminate are characterised by the observed proportions being flatter than the theoretical Rasch curve, whereas over-discriminating items have observed proportions steeper than the theoretical curve. Figure 3.3 is an example of an ICC showing an under discriminating item. The proportion of observed scores in the lower class interval is greater than expected by the model and the higher class interval is lower than expected.

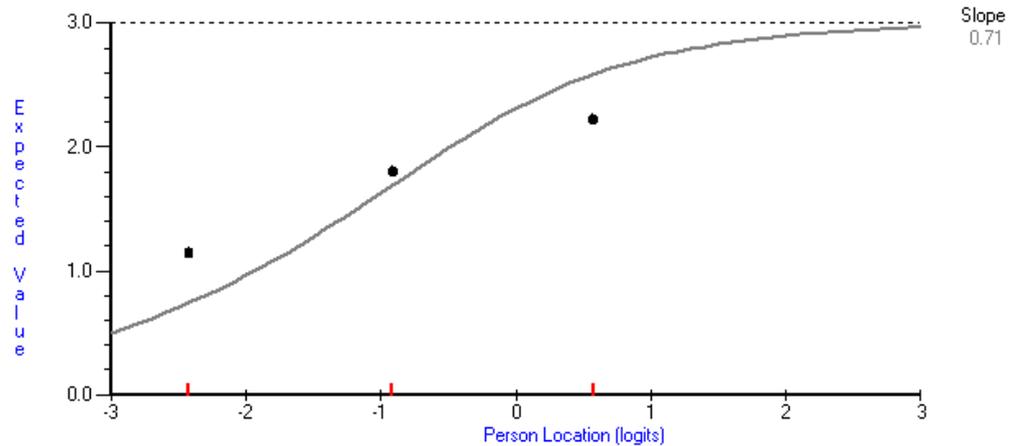


Figure 3.3. Example of an Item Characteristic Curve showing an under discriminating item.

There are two statistics that indicate data misfit to the model. Firstly a fit-residual provides an index of deviation of the observed estimate from the expected Rasch model estimate, as a standardized statistic, which ideally has a mean of 0 and a standard deviation of 1. When items or persons are ordered according to their fit residual it is possible to identify items or persons at the extremes; these cases may be relatively extreme compared to others, and a cut off level of  $\pm 2.5$  is commonly used (Default setting in RUMM2030 software, Andrich, Sheridan, & Luo, 2010) beyond which further investigation is considered.

The other statistic indicating fit to the model is a chi-square statistic that is considered with Bonferroni correction. When the chi square  $p$  value is significant this indicates a lack of fit to the model, and further investigation of other more specific evidence is warranted.

#### *f. Differential item functioning*

Using the RUMM2030 software, the PRFQ items were inspected for Differential Item Familiarity or Functioning (DIF). This analysis identifies whether any particular item or items are biased in the way they operate across comparable groups, and therefore not invariant as required by the Rasch model. Significant DIF indicates that persons of one group have a higher probability of affirming an item than persons of another group (despite being of a similar level of the trait). DIF can be seen in the ICC patterns for different groups, and an ANOVA  $F$  statistic indicates significant differences. Figure 3.4 shows an example of an item that displays DIF, where the average scores from one group's class intervals are consistently higher

than the other group. DIF was checked against parent gender, age, education level, income, family size, work hours, and child's gender. Ideally, no items of the PRFQ should show DIF with any of these variables. If DIF is found, the item showing DIF (or the item showing the largest DIF) can be resolved into new items for each category of the variable and the scale analysed again with the original item removed (e.g., an item with DIF for parent gender would become two items, one for mothers and an additional item for fathers). The properties of the scale can be assessed for any improvements and the new items can be examined for differences in their location estimates, which shows the magnitude of the DIF. If more than the one item shows DIF, resolving the item with largest DIF will reveal whether the remaining items have real or artificial DIF (Andrich & Hagquist, 2012).

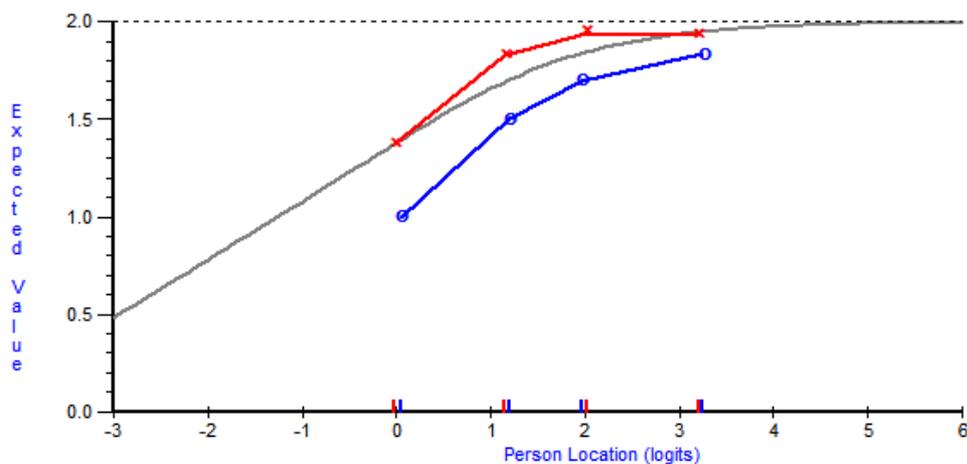


Figure 3.4. Example of an Item Characteristic Curve displaying DIF.

### *g. Local item dependence*

The Rasch model assumes local item independence, which means a person's response to any one item is not dependent (correlated) with any other item and each person's response to any particular item is not dependent on any other person's response (Andrich & Kreiner, 2010). If dependency exists in a data set, the reliability and standard errors are compromised, giving a false impression of the test's precision and quality (Baghaei, 2008; Marais & Andrich, 2008). This dependency is also called a violation of the assumption of local independence.

RUMM2030 produces a residual correlation matrix from which high positive correlations between item residuals can be easily identified (see 'Data Fit to the

Rasch Model' section of the Background Literature Review for an explanation of item residuals). A high correlation between a pair of item residuals indicates that the items may violate the assumption of independence. However, the judgment of how high a correlation needs to be to indicate dependence, is relative to the other correlations in the matrix (RUMM Laboratory, 2012). Often an examination of the wording and meaning of each item can reveal possible substantive reasons why the items are showing dependence.

If two items have relatively high residual correlations, Andrich and colleagues (Andrich, Humphry, & Marais, 2012) detail a procedure to estimate the amount of response dependence. In summary, after investigating the items' relative difficulties and fit to the item characteristic curve, the dependent item is split and resolved based on the responses to the other item. The original items are then removed from the analysis and the data reanalysed.

#### ***h. Dimensionality***

Multidimensionality is another source of violation of local independence (Andrich & Kreiner, 2010) and like the procedure for identifying local item dependence dimensionality can be examined by analysis of item residuals. Although the item-trait interaction chi-square statistic provides a summary statistic for dimensionality (as explained in section c), a Principal Components Analysis (PCA) of the item residuals (the part of the data unexplained by the Rasch Model) and *t*-tests of dimension subsets provides a more detailed analysis with greater sensitivity (Smith, 2002; Tennant & Pallant, 2006). A PCA of residuals is an examination of the variability in the items after the unidimensional Rasch 'factor' has been accounted for in all the items.

The RUMM2030 manual provides the following procedure to perform a PCA of residuals and *t*-test comparisons of subsets of items (Andrich, Sheridan, & Luo, 2009). The PCA of residuals is reported by firstly identifying the percentage of total variance attributable to the first factor and its relative size compared to subsequent factors. A relatively high percentage of variance in the residuals accounted for by the first component indicates multidimensionality (ie. has items that are measuring more than one construct). In addition, an examination of the loadings of items on the first factor can identify whether groups of items that are showing some commonality and

perhaps together represent a construct that is subsidiary to the main construct being measured by the complete set of items. A threshold of approximately + or -.3 is used to select two groups of items that have high correlations with the first component. These items should be examined for substantive differences. If there appears to be two groups of items that are distinct according to their residuals, results from a series of independent paired *t*-tests will provide the percentage of the persons which have significantly different scores on these groups of items.

A subtest analysis is a further test of dimensionality developed by Andrich (2009) that can be performed in RUMM2030. This procedure performs an analysis using the two groups of items (suspected subscales) identified from the PCA of residuals and *t*-tests. The analysis is based on the standard deviation of the person distribution and the standard error of measurements, and uses Cronbach's  $\alpha$  as an indication of how these vary in the analysis (RUMM Laboratory, 2009) This analysis provides a more accurate estimate of the reliability of the scale that allows for local dependence between the items of the same group. The subtest values for the Cronbach's  $\alpha$  are considered a more accurate estimates of unidimensional reliability and a decrease in  $\alpha$  when subtests are formed is evidence of multidimensionality. An overall test-of-fit Chi Square is provided which can indicate any improvement when the two subscales are analysed as two items.

If two subscales are clearly identified by the PCA of residuals and subtest analysis, a plot of the person scores can clearly show the relationship between the different dimensions. Scores on each subscale plotted together with boundaries indicating the area of 95% confidence (based on measurement error) shows if many persons significantly differ in their scores (Bond & Fox, 2007; Wright & Stone, 1979, 1999).

#### ***i. Differences between groups of persons***

Differences in means for groups of persons can be tested for significance with the ANOVA *F*-statistic, although this is best done once other aspects of fit to the model and reliability have been addressed. Mean estimates were tested for significant between-group differences when grouped according to sex of parent, parent age, sex of child, parent birthplace, birth order, parent education, and parent occupation. Of primary interest was to determine whether mothers and fathers differed in their level

of PRF and whether variance in PRFQ scores could be accounted for by these categorical parental characteristics. Generally, RF has been found not to be associated with these demographic variables, although few studies have included fathers and rarely have mother and father associations been compared.

*Iterative analysis process and removal of items*

The procedures detailed above provide evidence that a scale, including the individual items in a scale, is functioning according to the Rasch model,. The justification for removal of an item from a scale is based on consideration of these multiple forms of evidence within the context of the iterative process graphically portrayed in Figure 3.5.

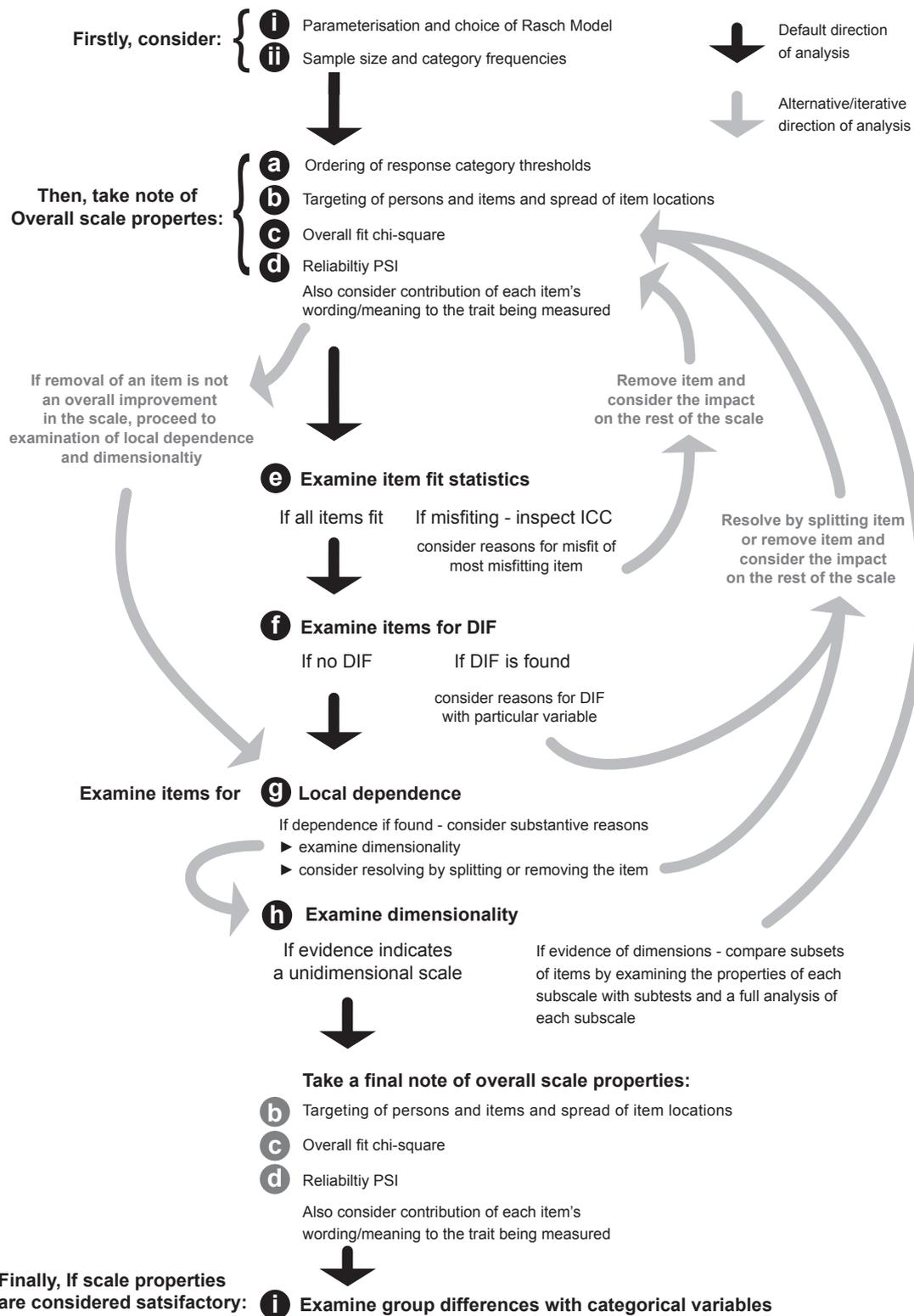


Figure 3.5. The iterative process of Rasch analysis

**Research Question 2: Is there an association between self-report of parental depression or anxiety and PRFQ scores?**

Spearman's correlations were computed between the PRFQ scores on one hand and ordinal scores from the subscales of the STAI (state  $N = 116$ ; trait  $N = 102$ ) and the BDI ( $N = 105$ ) on the other. Separate correlations were computed for mothers and fathers. According to Cohen's power tables ([Cohen, 1992](#)), 102 participants are sufficient for an 80% chance of capturing a 'small to medium' association (correlation of .24+) between two variables at a one-tailed alpha-level of .05.

**Research Question 3: Do one or more specific sets of PRFQ items show temporal (test–retest) stability?**

The couples that participated in the PDI interviews completed a retest of the PRFQ for the purpose of testing the temporal stability of the scale. The PRFQ is expected to be subject to measurement error and therefore it can be assumed the stability of scores would be less than perfect. A realistic expectation is that the PRFQ has sufficient reliability such that individuals obtain *similar* scores on each testing occasion. The more similar test scores are across two testing occasions, the greater the Pearson correlation between the two sets of scores. The Pearson correlation can therefore be used as an index of test reliability and is referred to as a *reliability coefficient*.

Data were used from the couples that completed a retest of the PRFQ (32 mothers and 37 fathers) to provide an analysis of test–retest validity or stability. Firstly, the item and threshold values from the first PRFQ analysis were used to anchor the item locations for the analysis of the test and retest data. With the item locations anchored, a related-samples *t*-test was conducted in RUMM2030 and a Pearson's correlation performed with the locations. Separate correlations were computed for mothers and fathers. A relatively strong correlation (.7+) would indicate satisfactory reliability of test-retest stability.

In addition to a correlation analysis, Rasch person estimates from the two PRFQ measures were plotted with the standard errors (SE) of the individual person scores used to create the 95% confidence lines (Bond & Fox, 2007). Figure 3.6 shows an example of such a plot. Ideally all plotted locations fall within the SE

confidence lines, and this indicates the degree with which the person estimates are invariant when persons are assessed at the two different times (allowing for error). Less precision can be identified at the extremes of low and high scores (larger error terms).

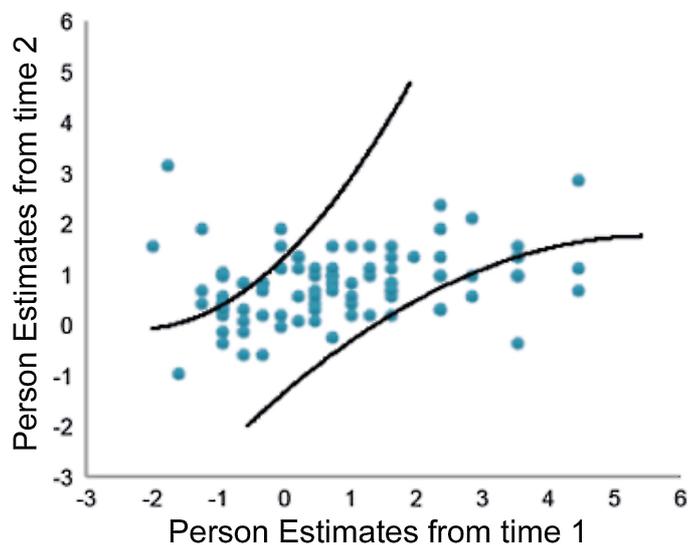


Figure 3.6. Example of a Common Person Linking Plot

A comparison of the two tests is made with a *t*-test, which provides details of the proportion of parents with significantly different scores. Means and standard deviations of both tests will also indicate differences or similarities with how parents responded to the two tests.

**Research Question 4: Is there a relation between PRFQ scores and PDI-RF scores that demonstrates convergent validity?**

A psychological test is said to have construct validity if it is ‘driven’ by the psychological construct that it purports to measure (i.e., in this case parental reflective functioning). If a measure has construct validity, it should correlate highly with other valid measures of that construct demonstrating convergent validity. Spearman’s correlations were computed separately for mothers ( $N = 39$ ) and fathers ( $N = 39$ ) to examine how PRFQ scores were associated with ordinal ratings of PRF from the PDI. A relatively strong correlation ( $.7+$ ) would indicate satisfactory reliability of test-retest stability.

## Chapter 4 Results

### Response Rates and Participant Characteristics

The one-year follow-up questionnaires from the PCHS families provided data for this study from 120 couples/families (240 participant parents; 62 male and 58 female children). At the 18-weeks gestation point of recruitment into the study, 439 mothers and 310 fathers completed the initial questionnaires. When a parent's child was approximately 12 months of age, 211 (48%) mothers and 137 (44%) fathers completed the one-year follow-up questionnaires, and of these parents, 120 couples had the required data to be included in the final analyses. The non-participant families either had: incomplete questionnaires, only one parent complete the one-year follow-up questionnaires, or had withdrawn from the study.

A comparison of the characteristics between final participants and non-participants is presented in Table 4.1. From this table it can be seen that there was no statistically significant difference between the two groups in terms of age, father education, occupation, country of birth, and whether or not the study child was their first child. The participant group differed from the non-participant group on two of demographic characteristics, with the participant group having a higher proportion of mothers with a degree or higher education, and higher family incomes. Parents were administered the BDI and STAI questionnaires at the one-year follow-up (See Table 4.1 for means and standard deviations)

Comparative statistics from the Australian Bureau of Statistics (Table 4.2) appeared to be generally similar, although education levels were slightly higher and the proportion of parents born outside Australia slightly lower than the research sample. Family income was not directly comparable with other statistics due to the categorical response format of the Peel questionnaire.

Table 4.1.

*Comparison of Participants and Non-participants Characteristics*

	Mother			Father			p-value
	Participants n = 120	Non-Participants n = 319	Total n = 439	Participants n = 120	Non-Participants n = 190	Total n = 310	
Parent Age at 18 Week*							
Under 28	26% (31)	35% (106)	33% (137)	18% (20)	24% (44)	22% (64)	
28-38	70% (82)	61% (184)	63% (266)	66% (73)	61% (112)	63% (185)	
Over 38	4% (5)	4% (13)	4% (18)	16% (18)	15% (27)	15% (45)	
Total	100% (118)	100% (303)	100% (421)	100% (111)	100% (183)	100% (294)	.478
Education Level							
< Year 12	8.7% (10)	16.4% (52)	14.4% (62)	6.7% (7)	16.7% (29)	12.9% (36)	
Year 12	13.9% (16)	8.8% (28)	10.2% (44)	10.6% (11)	13.8% (24)	12.6% (35)	
Cert / Dip	33.9% (39)	49.5% (157)	45.4% (196)	62.5% (65)	54.6% (95)	57.6% (160)	
≥ Degree	43.5% (50)	25.2% (80)	30.1% (130)	20.2% (21)	14.9% (26)	16.9% (47)	
Total	100% (115)	100% (317)	100% (432)	100% (104)	100% (174)	100% (278)	.064
ANZSCO Occupation Category							
1 & 2	46.2% (36)	32.3% (61)	36.3% (97)	27.6% (29)	20.1% (35)	22.9% (64)	
3 & 4	17.9% (14)	27.0% (51)	24.3% (65)	43.8% (46)	44.3% (77)	44.1% (123)	
5 & 6	28.2% (22)	34.4% (65)	32.6% (87)	7.6% (8)	8.6% (15)	8.2% (23)	
7 & 8	7.7% (6)	6.3% (12)	6.7% (18)	21.0% (22)	27.0% (47)	24.7% (69)	
Total	100% (78)	100% (189)	100% (267)	100% (105)	100% (174)	100% (279)	.447
Country of Birth**							
Australia	72.9% (86)	78.1% (246)	76.7% (332)	75.9% (85)	78.9% (150)	77.8% (235)	
Not Australia	27.1% (32)	21.9% (69)	23.3% (101)	24.1% (27)	21.1% (40)	22.2% (67)	
Total	100% (118)	100% (315)	100% (433)	100% (112)	100% (190)	100% (302)	.537

\* Mean age of participant mothers = 31.48 years (SD = 4.98), fathers = 33.62 (SD = 5.97)

\*\* No participants in this sample described themselves as an Aboriginal or Torres Islander.

Note: p values are derived from the chi-square test of association



Table 4.2.

*Australian Bureau of Statistics Comparative Population Demographics*

Australian median gross household income (2011-12)	\$74,984 per annum
Australian average gross household income for couple families with dependent children (2011-12)	\$134,160 per annum
Western Australian population born outside of Australia (2011)	33%
Median age of mothers for all births in Australia (2009)	30.6 years
Median age of fathers for all births in Australia (2009)	33.0 years
No post school qualification <sup>a</sup>	30.6%
Cert/Dip <sup>a</sup>	31.3%
Bachelor degree or higher <sup>a</sup>	38.1%

<sup>a</sup> Australian residents aged 25-34

Analysis was also performed to compare the 40 couples who were interviewed with the PDI and the remaining 80 who only completed questionnaires. No differences were found for the demographic characteristics listed in Table 4.1.

### **Research Question 1: Do data from mothers and fathers PRFQ conform to the requirements of the Rasch measurement model?**

The PRFQ consists of three subscales, which are each scored differently and represent different aspects of parental reflective functioning. Each subscale was examined separately using RUMM2030 software in conjunction with the methods detailed in the Methods Section. Following are the full details of the HL subscale analysis and a further analysis of two dimensions that were revealed within this subscale; the Child- Focused and Self-Focused subscales. A 7-item scale from within the HL subscale showed the best psychometric properties from all the items of the PRFQ and is the only subscale that has all results reported in full. An abbreviated summary is provided for the results of the LH and the M subscales of the PRFQ.

#### **HL subscale analysis**

The following results from the Rasch analysis of the HL subscale are presented under headings that correspond with the headings of the Methods section.

*i. Parameterisation and choice of Rasch model*

The HL data best fit the partial credit parameterisation of the polytomous Rasch model, as indicated by a significant  $\chi^2$  statistic ( $p < .001$ ) from the likelihood ratio test. This result indicated that parents responded to the seven rating categories of items in a manner that varied across the 17 items, i.e., the difference between two thresholds in one item was not necessarily the same as the distance between those thresholds in another item. Therefore, all further analysis was undertaken using the Partial Credit Model, which assumes the threshold distances vary across the items. Using this model also allowed for rescoring of individual items if the category thresholds were disordered (described in next section), and therefore using data from items with differing numbers of response categories.

*ii. Sample size and category frequencies*

The category frequencies of the HL subscale items are displayed in Table 4.3, which shows that there are many items that have low response rates across the three levels of *disagree* categories: 0, 1 and 2. Some items have responses fairly evenly distributed and a few are skewed in the opposite direction to what would be expected, with lower frequencies in the *agree* categories (in particular Item 10).

Table 4.3.

*Item Category Frequencies for 17-Item HL Scale*

Item	Abbreviated Item Wording	Response Category						
		0	1	2	3	4	5	6
1	My Child and I can feel differ	6	21	18	80	32	39	37
3	I am often curious	4	5	5	26	33	99	64
4	How I am feeling can affect	7	12	10	28	60	76	44
6	I like to think about the	2	6	10	30	53	72	64
7	I try to see situations	2	3	12	33	47	89	51
10	I believe that how I think	82	39	15	56	18	17	8
11	My child can react to a	15	38	37	62	40	29	14
13	At times, it takes several	32	66	33	38	54	8	5
15	Now that I am a parent, I	19	21	18	52	41	56	28
17	How I see my child changes	36	27	17	71	47	20	12
20	I wonder a lot about what my	4	6	13	42	63	72	34
22	I can sometimes misunderstand	13	36	22	62	71	28	4
24	I believe that how my parents	24	11	9	28	59	64	40
26	I pay attention to what my	0	4	1	12	34	93	91
28	Understanding why my child	5	3	2	33	44	108	38
30	I often think about how I felt	20	42	27	57	53	25	11
31	I try to understand the reason	6	9	9	33	55	86	35

*a) Order of response category thresholds*

The threshold locations for the 17-item HL scale are displayed in Table 4.4. All but three of the 17 HL items had disordered thresholds, indicating these item categories were not responded to in the manner required for measurement.

Table 4.4.

*Item Thresholds for 17-Item HL Scale*

Item	Location	Thr1	Thr2	Thr3	Thr4	Thr5	Thr6	
1	-0.15	-0.71	-0.76	-0.24	0.42	0.80	0.48	*
3	-0.42	-0.19	-0.01	-0.29	-0.53	-0.20	1.21	*
4	-0.19	-0.14	-0.27	-0.47	-0.46	0.04	1.30	*
6	-0.50	-0.40	-0.54	-0.43	-0.11	0.40	1.07	*
7	-0.52	-0.87	-0.43	-0.31	-0.18	0.31	1.48	
10	0.63	0.50	-0.54	-0.60	-0.15	0.35	0.43	*
11	0.24	-1.01	-0.53	-0.21	0.08	0.49	1.16	
13	0.88	-1.28	-0.46	-0.61	-0.74	0.12	2.96	*
15	0.09	-0.16	-0.31	-0.38	-0.27	0.15	0.98	*
17	0.43	0.35	-0.91	-0.85	-0.12	0.67	0.86	*
20	-0.26	-0.39	-0.66	-0.62	-0.25	0.44	1.47	*
22	0.59	-1.24	-0.82	-0.85	-0.64	0.45	3.11	*
24	0.07	1.04	-0.33	-0.81	-0.62	-0.02	0.74	*
26	-0.83	-0.50	-0.08	-0.20	-0.36	-0.06	1.19	*
28	-0.27	0.33	-0.35	-0.81	-0.80	-0.05	1.69	*
30	0.39	-0.81	-0.54	-0.43	-0.22	0.38	1.62	
31	-0.19	-0.26	-0.33	-0.56	-0.56	0.05	1.66	*

\* Disordered threshold

The item threshold order can be seen visually in the Category Probability Curves (CPCs). For example Figure 4.1 shows the ordered thresholds of Item 11. Figure 4.2 provides the other 16 CPCs for the HL scale.

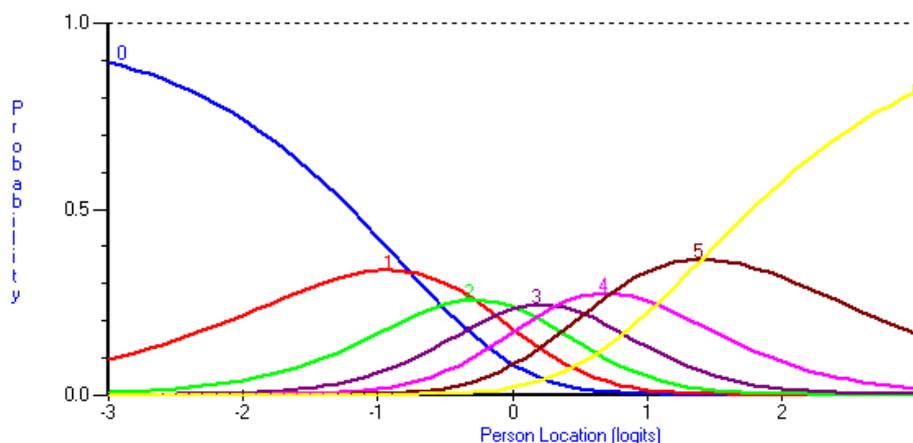


Figure 4.1.

Category Probability Curves for HL Item 11

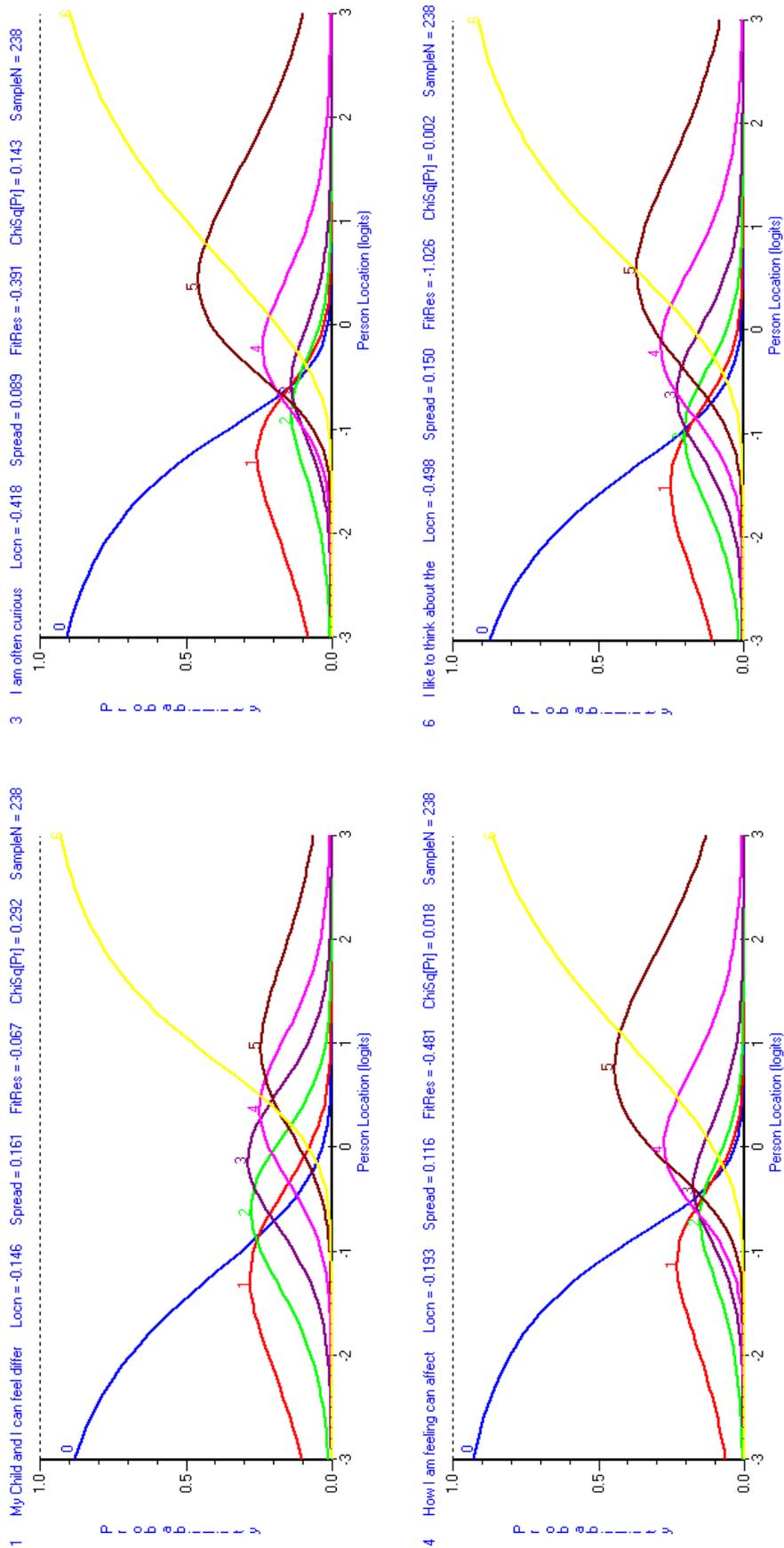


Figure 4.2.

Category Probability Curves for HL Items (except item 11)

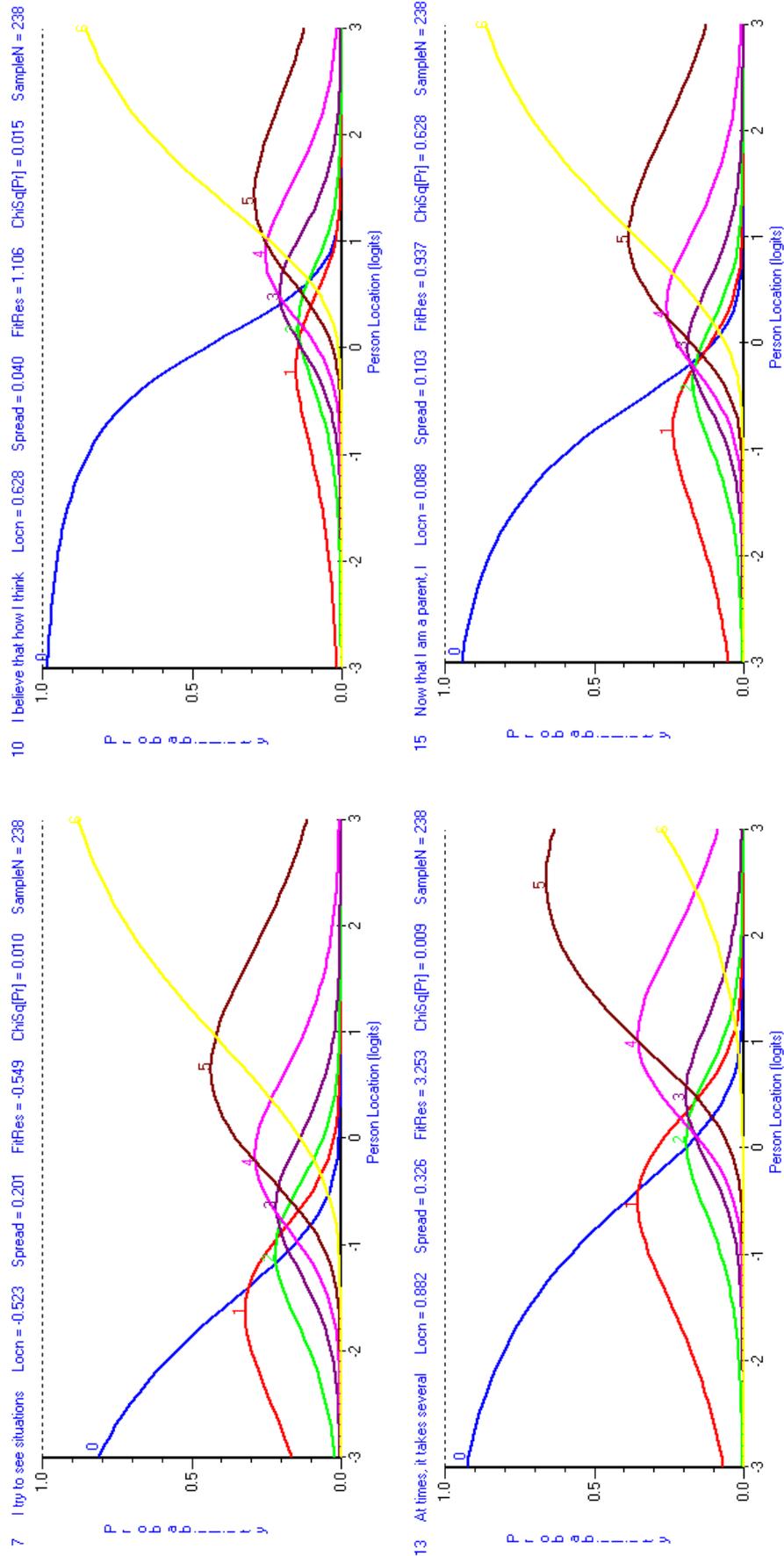


Figure 4.2. Continued

Category Probability Curves for HL Items (except item 11)

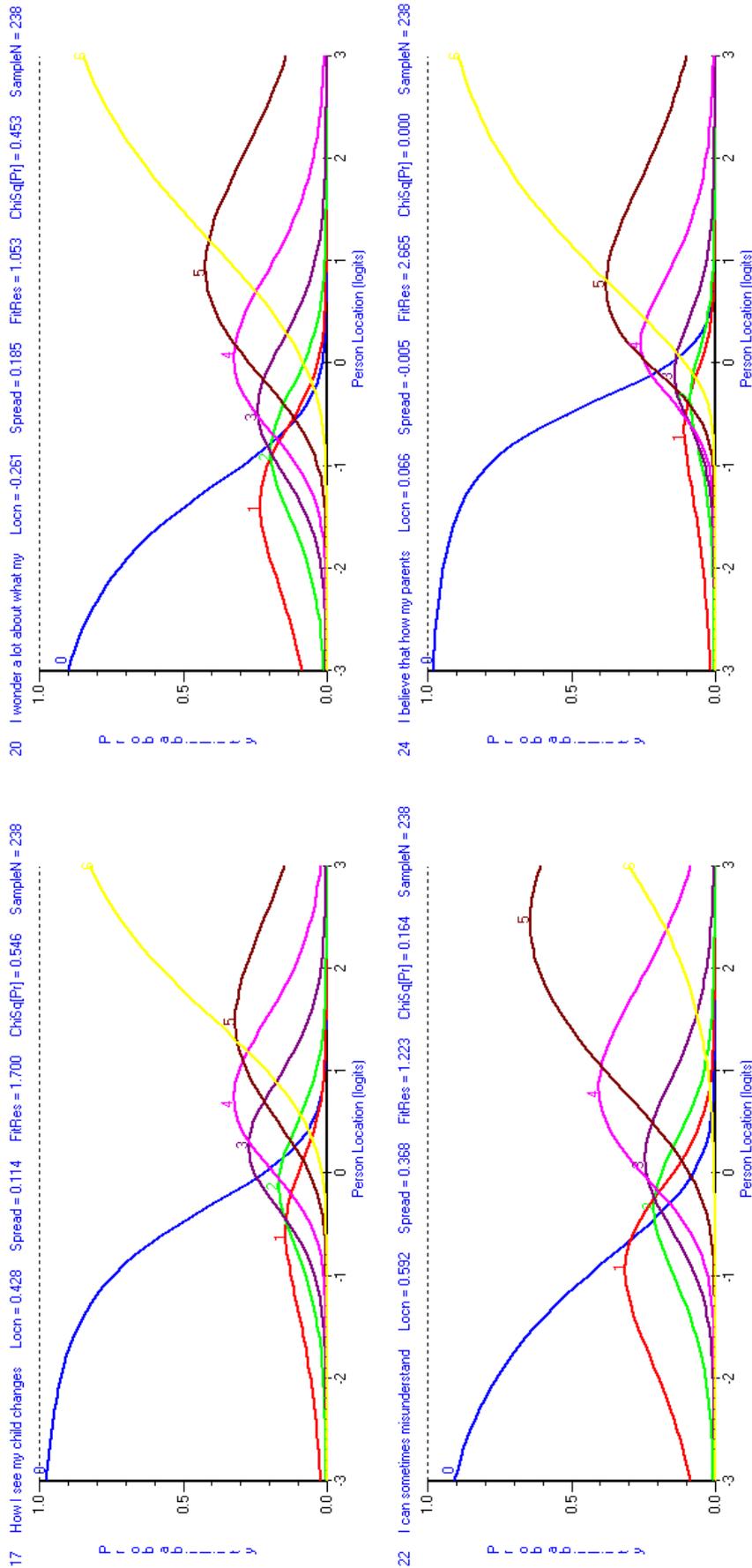


Figure 4.2. Continued

Category Probability Curves for HL Items (except item 11)

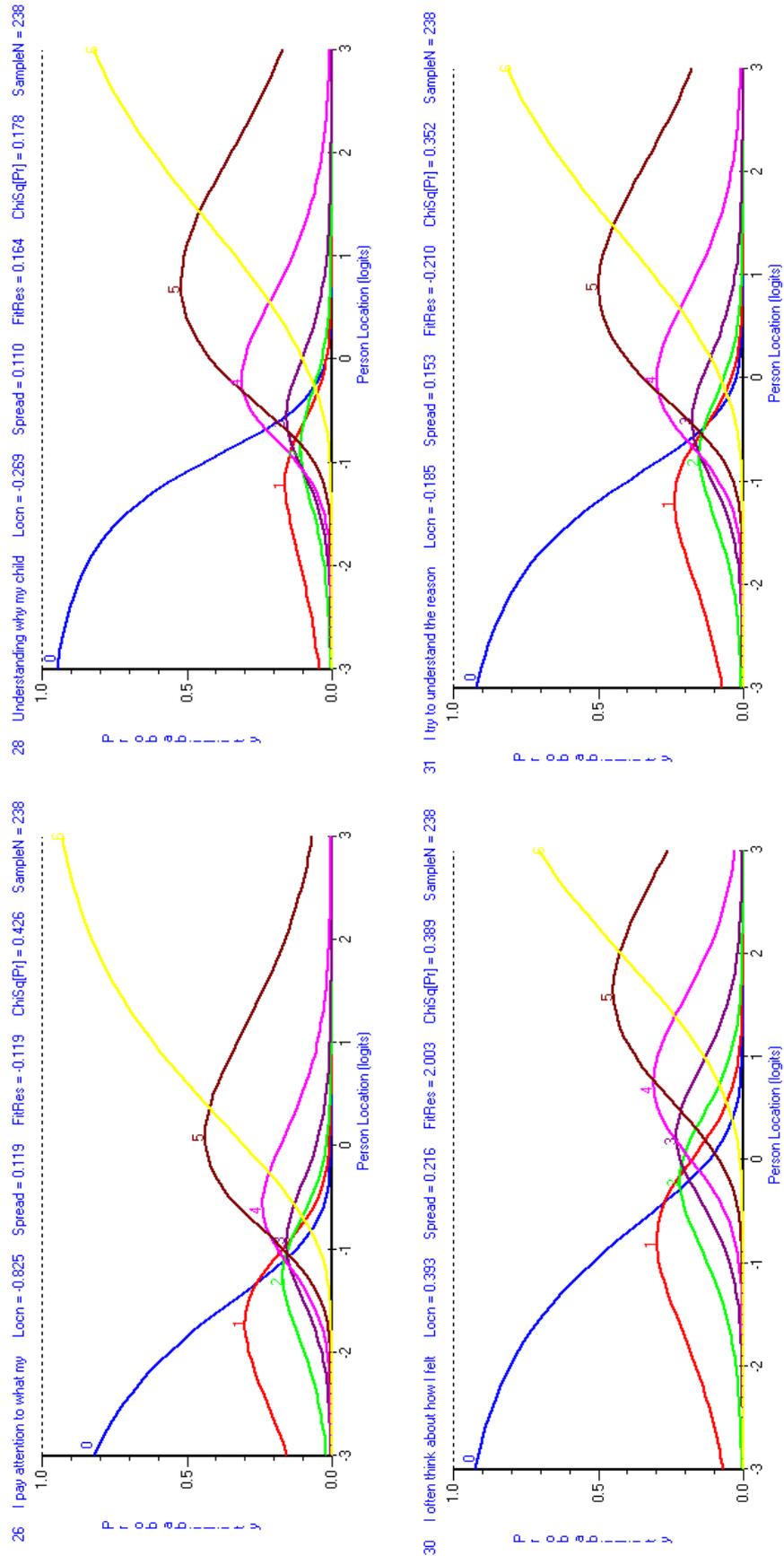


Figure 4.2. Continued

Category Probability Curves for HL Items (except item 11)

From observation of the threshold locations and the CPC of Item 4, it can be seen that there is an order to the thresholds 4, 5 and 6; the *agree* end of the ratings to the statement “How I am feeling can affect how I understand my child’s behaviour”. The three categories at the *agree* end of the scale (labelled 4, 5 and 6 in Figure 4.2) each have a range of person locations for which the probability of selecting that category is higher than the probability of selecting any other category. The categories labelled 1, 2 and 3 have no person locations at which they are the most probable categories to select. The CPCs were examined for all the PRFQ items in the HL scale and the areas of threshold disorder identified.

The HL items were rescored so as to provide rating scale data with ordered thresholds, as displayed in the Threshold Map (Figure 4.3), which shows the range of person locations for which each response category has the highest probability of being selected. The scoring that achieved ordered thresholds is displayed in Table 4.3.

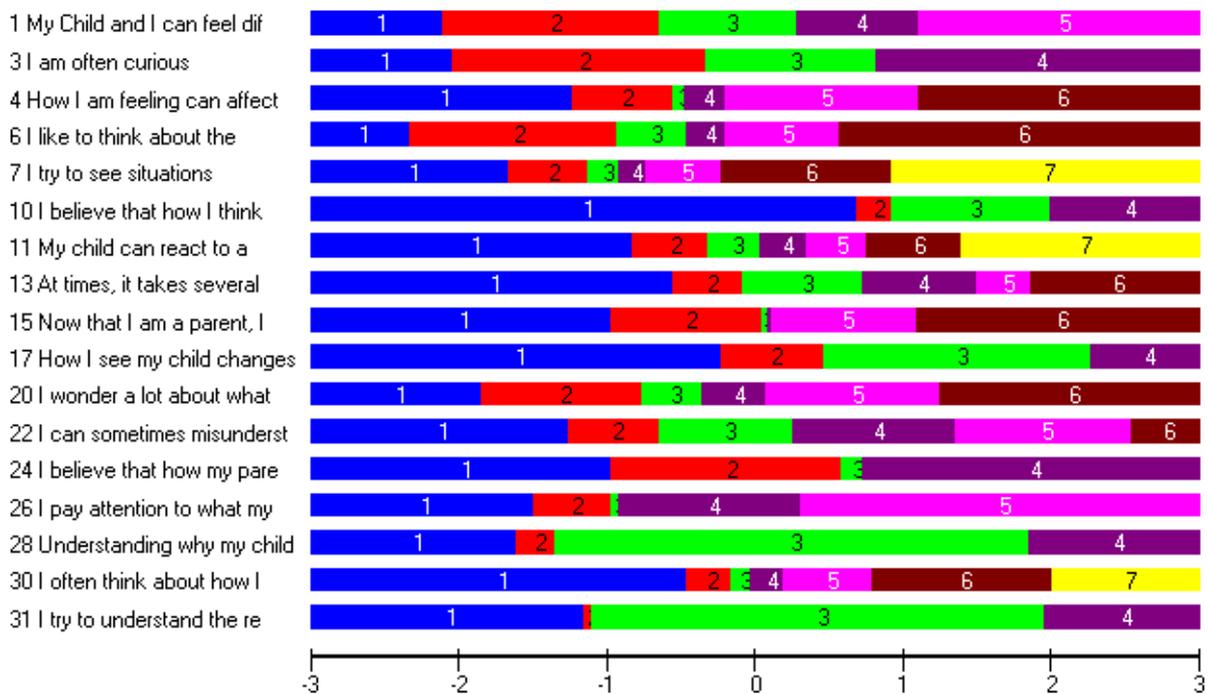


Figure 4.3. Threshold Map for 17-item HL scale

Table 4.5.

*Category Scoring for 17 HL Items*

Item	Max Score	Score for each category						
		1	2	3	4	5	6	7
1	4	0	1	1	2	3	3	4
3	3	0	0	1	1	1	2	3
4	5	0	1	1	2	3	4	5
6	5	0	1	1	2	3	4	5
7	6	0	1	2	3	4	5	6
10	3	0	0	1	1	2	2	3
11	6	0	1	2	3	4	5	6
13	5	0	1	2	2	3	4	5
15	5	0	1	1	2	3	4	5
17	3	0	0	1	1	2	2	3
20	5	0	1	1	2	3	4	5
22	5	0	1	2	2	3	4	5
24	3	0	0	1	1	1	2	3
26	4	0	0	1	1	2	3	4
28	3	0	0	1	1	2	2	3
30	6	0	1	2	3	4	5	6
31	3	0	0	1	1	2	2	3

***b) Targeting and distribution of items and persons***

Examination of the individual person location scores and fit residuals showed that when locations were calculated using the 17 rescored HL items there were no parents with extreme scores.

The distribution of person locations and item locations can be seen in the upper and lower sections of Figure 4.4 respectively. An alternative display of distributions is presented in the Item Map (Figure 4.5), which shows the same distribution of persons on the left, against the item locations on the right. From these distribution figures it can be seen that the item threshold locations are well matched and spread across the range of person locations. Table 4.4 displays means and standard deviations of the item and person locations and fit residuals. These statistics are examined and referred to when the same statistics are produced from a refined set of items.

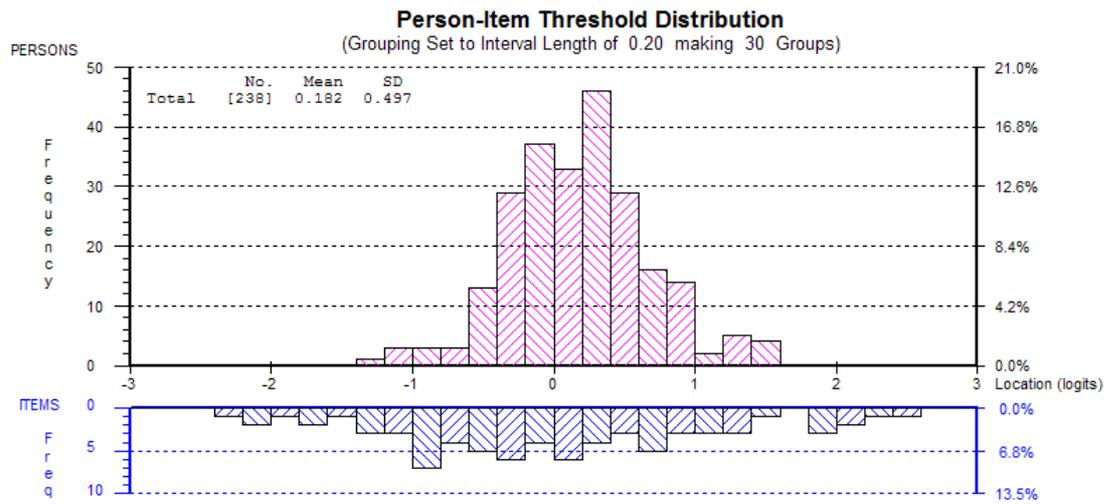


Figure 4.4. Person-Item Threshold Distribution of 17-item HL scale

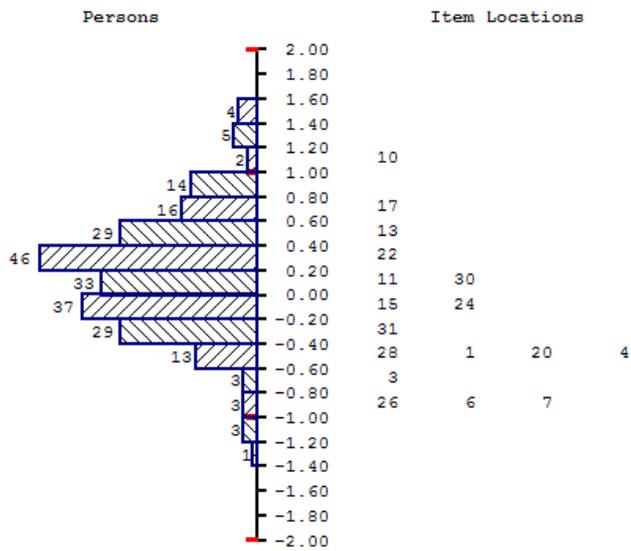


Figure 4.5. Item Map of 17-item HL scale

Table 4.6.

*Item-Person Interaction Statistics for 17-item HL Scale*

	Items		Persons <sup>a</sup>	
	Location	Fit Residual	Location	Fit Residual
Mean	0	0.44	0.18	-0.29
SD	0.57	1.18	0.50	1.48

<sup>a</sup>*n* = 238***c) Overall fit to the Rasch model***

The 17-item HL scale was found to misfit the Rasch model according to the item-trait interaction chi-square statistic (118.85; *p* = .000).

***d) Reliability***

The Person Separation Index (PSI) for the 17-item HL scale was .76.

***e) Item and person fit to the Rasch model***

The fit residual and chi square statistics (Table 4.5) were examined for both items and persons. Item 13 and 24 had significant chi-square values (highlighted), which indicated misfit. The average locations for each of the four class intervals are plotted against the expected curve in Figures 4.6 and 4.7. For both these items it can be seen that they under-discriminate, with parents in the high group scoring lower than expected and those in the lowest group scoring higher than expected.

Table 4.7.

*Individual Item-Fit Statistics for 17-item HL Scale*

Item	Location	SE	FitResid	DF	ChiSq	DF	Prob
26	-0.77	0.07	-0.24	217.40	4.23	3	0.237967
6	-0.67	0.06	-1.73	219.25	8.93	3	0.03025
7	-0.63	0.06	-1.01	219.25	6.80	3	0.078448
3	-0.52	0.08	-0.45	218.32	11.49	3	0.009374
28	-0.37	0.10	-0.08	215.55	10.44	3	0.015153
1	-0.34	0.07	-0.19	215.55	3.09	3	0.377835
20	-0.33	0.06	0.85	216.47	3.72	3	0.292841
4	-0.27	0.06	-0.43	219.25	6.10	3	0.106702
31	-0.10	0.10	-0.47	215.55	10.57	3	0.01429
15	0.07	0.05	0.94	217.40	5.16	3	0.16017
24	0.12	0.07	2.29	217.40	17.48	3	**0.000565
11	0.23	0.05	0.69	217.40	3.33	3	0.34301
30	0.39	0.05	2.41	217.40	4.67	3	0.197411
22	0.45	0.07	1.03	218.32	2.11	3	0.549127
13	0.69	0.06	2.07	218.32	17.78	3	**0.000489
17	0.83	0.08	1.50	212.77	0.34	3	0.951554
10	1.21	0.08	0.31	217.40	2.59	3	0.458984

\*\*Chi-square is significant with Bonferroni-adjustment alpha-level of 0.000588

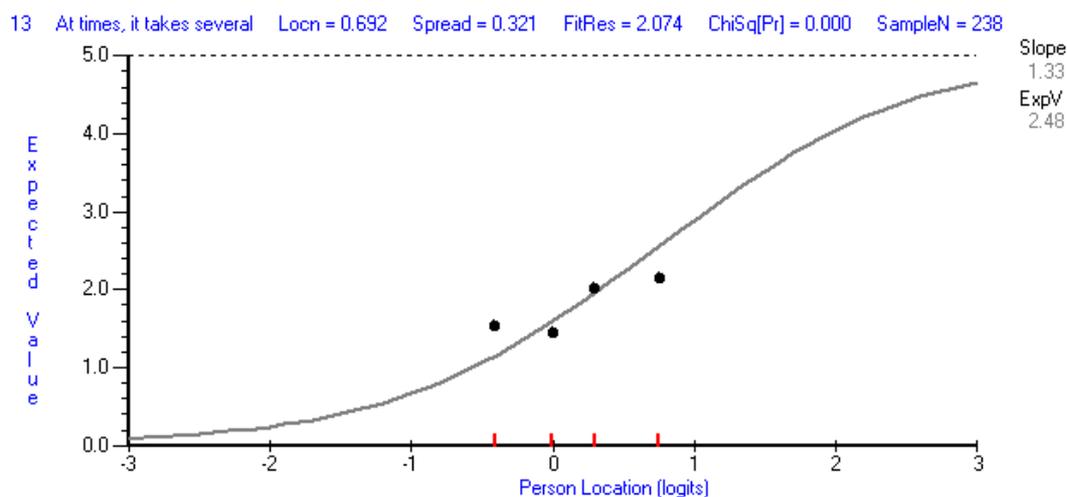


Figure 4.6. Item Characteristic Curve for HL item 13

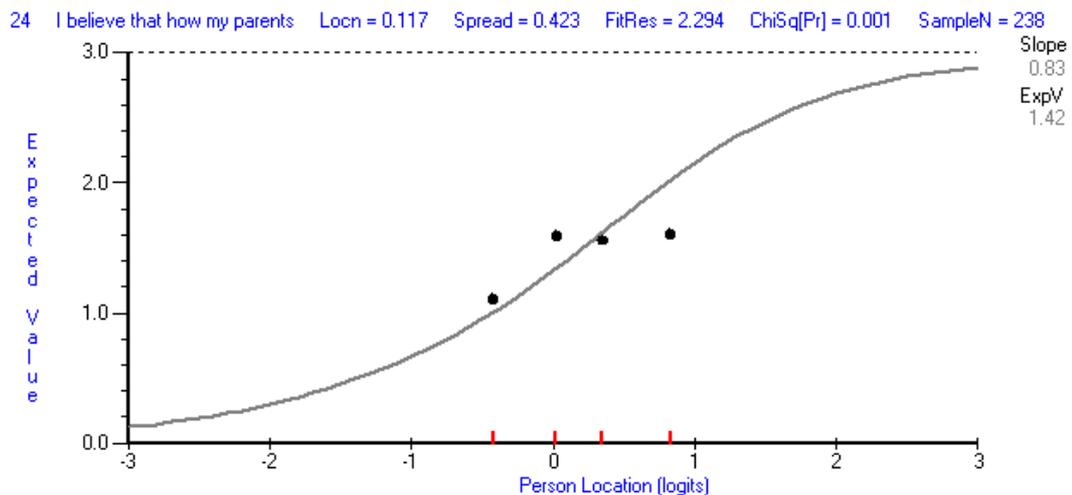


Figure 4.7. Item Characteristic Curve for HL item 24

The individual person fit residuals were examined and it was found that data for five parents (2%) had fit residuals over 2.5 and sixteen parents (7%) had residuals below -2.5.

#### *f) Differential item functioning*

The set of 17 HL items was examined for differential item functioning (DIF) to identify any individual item bias for particular groups of parents. Seven variables were considered for DIF: gender of parent, parent age, gender of child, parent birthplace, birth order, parent education, and parent occupation. Only parent gender showed statistically significant DIF with Items 1 and 13 at the Bonferroni-adjusted alpha-level of .000980. This DIF is shown graphically in the item characteristic curves. For example, in Figure 4.8 it can be seen that mothers were more likely to agree to Item 1 (My child and I can feel differently about the same thing) compared to fathers for all four of the class intervals.

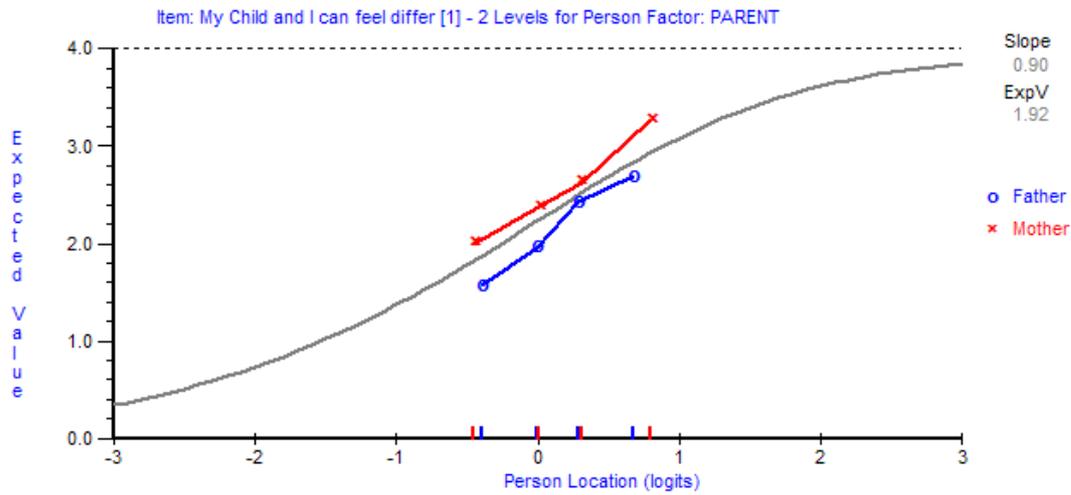


Figure 4.8. ICC showing differential item functioning for Gender of Parent with HL Item 1

**g) Local dependence**

Item residual correlations were examined for indications of item dependence. From the correlation matrix (Table 4.6) it can be seen that four pairs of items had residual correlations above 0.3.

Table 4.8.

*Person-item Residual Correlation Matrix for 17-item HL Scale (from RUMM203)*

Item	1	3	4	6	7	10	11	13	15	17	20	22	24	26	28	30	31
1	1.000																
3	-0.005	1.000															
4	0.086	0.022	1.000														
6	-0.073	0.410	0.061	1.000													
7	-0.022	0.281	0.012	0.503	1.000												
10	-0.071	-0.286	0.000	-0.242	-0.262	1.000											
11	-0.101	-0.239	-0.151	-0.347	-0.361	0.188	1.000										
13	-0.161	-0.208	-0.169	-0.331	-0.223	0.163	0.290	1.000									
15	-0.113	-0.194	-0.168	-0.244	-0.233	-0.020	-0.063	-0.008	1.000								
17	-0.001	-0.137	0.083	-0.211	-0.243	0.368	0.023	-0.023	-0.019	1.000							
20	-0.107	0.250	-0.159	0.072	0.098	-0.224	-0.153	-0.153	-0.098	-0.211	1.000						
22	0.035	-0.243	0.028	-0.376	-0.419	0.002	0.192	0.196	0.040	0.100	-0.153	1.000					
24	-0.058	-0.226	-0.029	-0.175	-0.144	-0.018	-0.061	0.004	-0.116	-0.097	-0.172	0.036	1.000				
26	0.040	0.242	-0.127	0.276	0.293	-0.320	-0.259	-0.364	-0.096	-0.229	0.163	-0.260	0.000	1.000			
28	0.033	0.052	-0.062	0.186	0.176	-0.171	-0.250	-0.238	0.031	-0.178	0.008	-0.230	-0.142	0.177	1.000		
30	-0.205	-0.203	-0.201	-0.152	-0.119	-0.025	-0.110	-0.089	-0.076	-0.036	-0.023	-0.099	0.131	-0.143	-0.044	1.000	
31	-0.045	0.072	0.027	0.278	0.152	-0.197	-0.253	-0.289	0.075	-0.181	0.001	-0.208	-0.077	0.130	0.346	-0.075	1.000

Correlations greater than an absolute value of .3 are highlighted

Local item dependence was most noticeable for Item 6, with residual correlations of .50 with Item 7 and .41 with Item 3. An examination of the statistics in Table 4.5 shows that Item 6 has the largest fit residual of these three items, and tended to slightly over discriminate (see Figure 4.9).

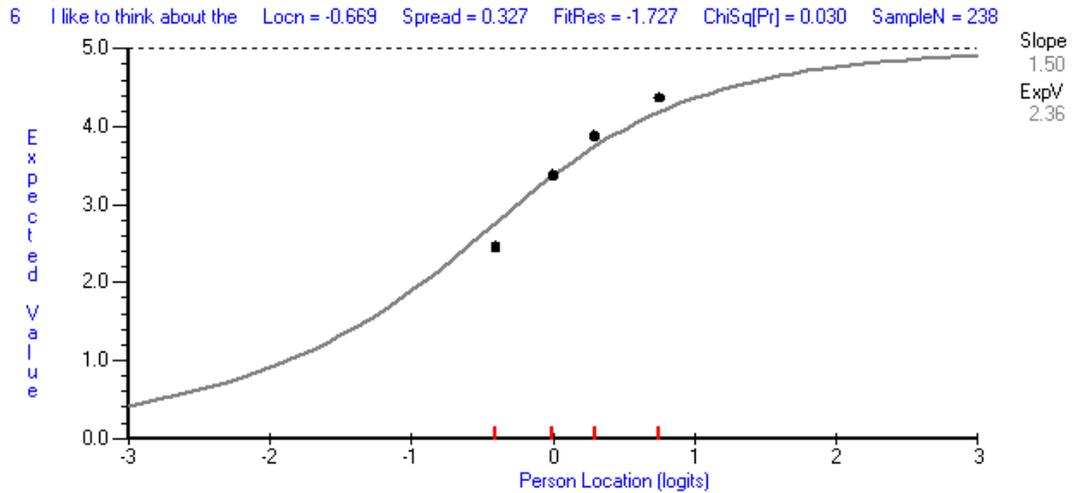


Figure 4.9. ICC for item 6

**h) Dimensionality**

Item residuals were examined with a Principal Components Analysis (PCA) in order to examine the dimensionality within the scale items. Items loaded onto two dimensions, as can be seen in the PC loadings in Table 4.7. The Principal Components Summary (Table 4.8) shows the first component explained 20.76% of the total variance among residuals with an eigenvalue of 3.53. Components 2 and 3 explained comparatively smaller proportions of variance, 9.09% (Eigen = 1.54) and 8.47% (Eigen = 1.44) respectively.

Table 4.9.

*HL Items PC loadings (from RUMM203)*

Item	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15
6	0.725	0.187	-0.099	0.125	-0.188	-0.226	0.077	0.060	0.004	-0.205	0.035	0.121	-0.133	0.257	-0.083
7	0.681	0.095	-0.155	0.175	-0.130	-0.247	-0.108	0.050	-0.311	-0.012	0.034	-0.222	-0.279	-0.083	-0.217
26	0.587	-0.121	-0.063	-0.023	0.239	0.156	-0.337	0.379	0.234	-0.127	0.010	-0.292	-0.001	-0.081	0.365
3	0.569	0.192	-0.426	-0.100	-0.074	0.142	0.153	0.070	-0.013	-0.165	-0.069	0.346	0.424	0.072	0.054
31	0.457	-0.091	0.497	-0.192	-0.090	-0.241	0.178	-0.106	0.308	0.025	-0.053	0.366	-0.264	-0.218	0.165
28	0.441	-0.134	0.450	-0.229	-0.105	-0.153	-0.204	-0.350	0.098	0.019	-0.330	-0.268	0.340	0.083	-0.140
20	0.332	-0.285	-0.446	-0.140	-0.049	0.478	0.166	-0.157	0.161	0.463	-0.149	-0.031	-0.166	-0.023	-0.101
1	0.051	0.422	0.131	-0.176	0.497	0.202	-0.486	-0.291	-0.269	0.099	0.106	0.247	-0.065	-0.007	0.016
4	0.044	0.651	0.173	0.136	0.185	-0.104	0.500	-0.058	0.044	0.268	0.205	-0.261	0.158	-0.064	0.115
30	-0.117	-0.514	0.168	0.572	-0.103	0.255	0.106	-0.379	-0.146	-0.219	0.202	-0.002	0.050	0.005	0.152
15	-0.172	-0.319	0.421	-0.537	-0.146	0.147	0.137	0.416	-0.298	0.068	0.235	0.017	0.068	0.076	-0.089
24	-0.177	-0.243	0.157	0.548	0.501	-0.177	-0.028	0.343	0.119	0.189	-0.192	0.191	0.113	0.015	-0.217
17	-0.385	0.461	0.177	0.146	-0.361	0.404	-0.058	0.151	0.066	-0.234	-0.233	-0.001	-0.009	-0.346	-0.166
10	-0.503	0.274	0.112	0.193	-0.490	0.005	-0.267	0.084	0.119	0.265	-0.088	0.049	-0.086	0.378	0.193
22	-0.532	0.073	-0.046	-0.268	0.417	0.064	0.301	-0.074	0.070	-0.362	-0.282	-0.120	-0.240	0.249	-0.001
13	-0.536	-0.138	-0.371	-0.131	-0.103	-0.452	-0.004	-0.030	-0.330	0.086	-0.282	0.025	0.059	-0.218	0.242
11	-0.554	-0.051	-0.334	-0.201	-0.054	-0.264	-0.211	-0.119	0.448	-0.092	0.389	-0.016	0.069	-0.074	-0.187

Loadings greater than an absolute value of .3 are highlighted

Table 4.10.

*HL Items Principle Components Summary*

PC	Eigen	Percent	CPercent	StdErr
PC001	3.529	20.76%	20.76%	0.489
PC002	1.545	9.09%	29.85%	0.204
PC003	1.440	8.47%	38.32%	0.194
PC004	1.320	7.76%	46.08%	0.172
PC005	1.258	7.40%	53.48%	0.166
PC006	1.078	6.34%	59.82%	0.145
PC007	0.983	5.78%	65.60%	0.129
PC008	0.885	5.21%	70.81%	0.117
PC009	0.813	4.78%	75.59%	0.111
PC010	0.743	4.37%	79.96%	0.100
PC011	0.701	4.12%	84.08%	0.096
PC012	0.659	3.87%	87.95%	0.091
PC013	0.611	3.60%	91.55%	0.086
PC014	0.528	3.10%	94.65%	0.080
PC015	0.497	2.92%	97.58%	0.076
PC016	0.405	2.38%	99.96%	0.079
PC017	0.007	0.04%	100.00%	0.045

A number of related-samples *t*-tests were performed using person estimates from different groupings of the items, based on the loadings in the first PC. Firstly, loadings that were greater than an absolute value of .3 were grouped and compared (i.e., positively loaded Items 6, 7, 26, 3, 31, 28 and 20; negatively loaded Items 11, 13, 22, 10 and 17), then items that were loaded between the  $\pm 0.3$  thresholds were grouped and analysed with either the positive or negative loaded items. Grouping the seven positively loaded items and comparing them with the other items provided the most distinct differences shown by the *t*-tests. A comparison of the seven positively loaded items to the other ten items showed 21.01% ( $n = 50$ ) of the parents' locations differed significantly ( $p < .05$ ). A comparison of the same seven items with the five most negatively loaded items showed a similar result of 21.43% ( $n = 51$ ).

Consideration of the item wording of the two sets of items revealed a tendency for the positively loaded items to be focused on the child's mental states as much or more so than the parent's mental states (Child-Focused; PRFQ-CF); and the converse was true for the negatively or less positively loaded items, which mostly focused on the parent's own mental states and the opacity of mental states (Self-Focused; PRFQ-SF). Table 4.9 presents these two subsets of HL items.

Table 4.11.

*Two Groups of Items from 15-item HL Scale Based on PCA*

Positively loaded items - Other / Child Focused	
6	I like to think about the reasons behind the way my child behaves and feels.
7	I try to see situations through the eyes of my child.
26	I pay attention to what my child is feeling.
3	I am often curious to find out how my child feels.
31	I try to understand the reasons why my child misbehaves.
28	Understanding why my child behaves in a certain way helps me not to be upset with him or her.
20	I wonder a lot about what my child is thinking and feeling.
Negatively (and less positively) loaded items – Self Focused	
1	My child and I can feel differently about the same thing.
4	How I am feeling can affect how I understand my child's behaviour.
30	I often think about how I felt when I was a child.
15	Now that I am a parent, I realize how my parents could have misunderstood my reactions when I was a child.
24	I believe that how my parents raised me affects how I raise my child.
17	How I see my child changes as I change.
10	I believe that how I think about my child will change over time.
22	I can sometimes misunderstand the reactions of my child.
13	At times, it takes several tries before I understand what my child needs or wants.
11	My child can react to a situation very differently than I think he or she will.

A subtest analysis was performed in RUMM2030 to determine the Cronbach's alpha when these two sets of items were grouped as subtests. The result was a reduction in alpha from .75 to .44, and the correlation of these two sets of items was weak ( $r = .43$ ).

***Iterative analysis process and removal of items***

An iterative process of examining the fit and suitability of individual items was undertaken. Consideration for removal of an item was informed initially by individual item fit statistics and examination of the ICCs. However, multiple sources of evidence were considered prior to any item being removed, and relevant evidence has been noted in the following report of this process. Possible substantive reasons for the removal of items are addressed in greater detail in the Discussion section.

As reported, items 13 and 24 both showed misfit according to the chi-square statistic. Figures 4.6 and 4.7 show how responses to both these items under

discriminated compared to the Rasch curve. The analysis of DIF (reported in previous section) showed that Item 13 was biased with respect to mothers and fathers. Inspection of the ICC with mothers and fathers plotted separately (Figure 4.10) shows mothers scoring noticeably lower than expected in the second and third class intervals, whereas the observed scores for father class intervals is in the expected direction. With item 13 data removed, the PSI improved marginally from .76 to .77, and the interaction chi-square statistic remained similar (120.90;  $p < .001$ ).

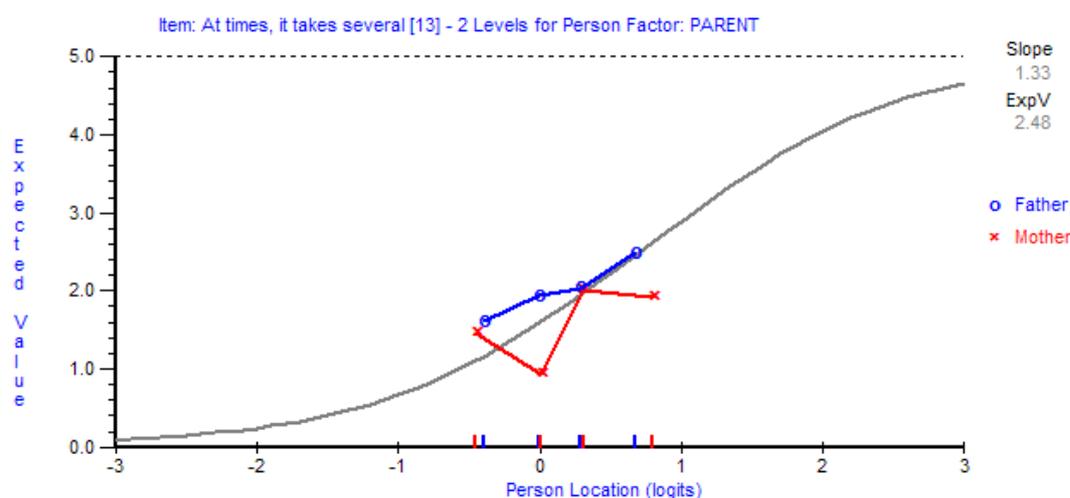


Figure 4.10. ICC for HL item 13 showing differential item functioning between mothers and fathers

Following the removal of Item 13, Item 24 was found to remain misfitting, with a significant chi-square misfit and a fit residual of 2.53. With the removal of Item 24, the PSI improved marginally (.78), and the interaction chi-square statistic remained similar (121.77;  $p < .001$ ).

With both Item 13 and Item 24 data removed from the HL scale, individual item-fit statistics showed all items fitting according the chi-square statistic and only item 30 with an elevated fit residual (3.10). Inspection of the ICC for Item 30 (Figure 4.11) shows evidence of slight under discrimination, although the observed probabilities of the class intervals were in the expected direction.

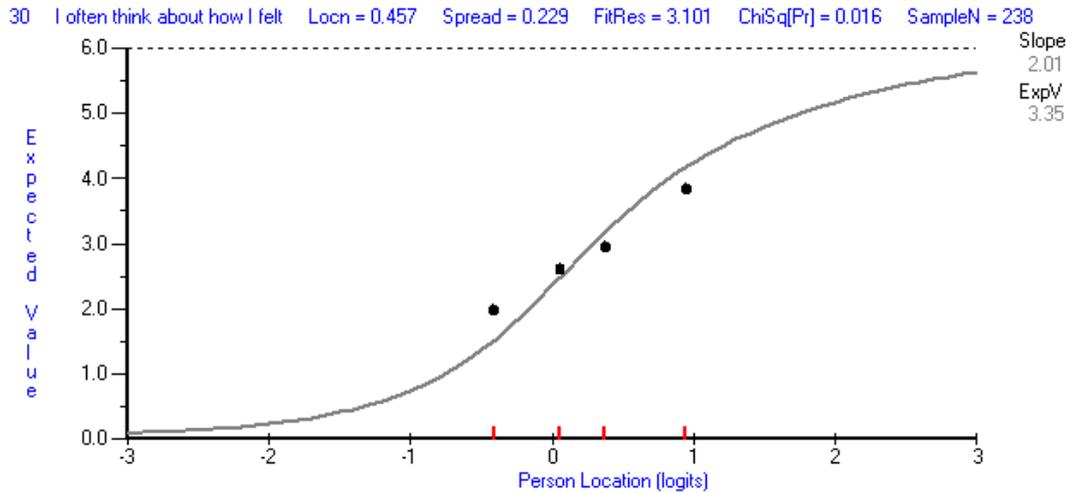


Figure 4.11. ICC for HL item 30

The removal of Item 30 from the analysis provided no change in the PSI and a minimal reduction in the interaction chi-square statistic (107.07;  $p = .000$ ). Therefore item 30 was retained in further analysis.

Analysis of item dependence and scale dimensionality with the item residuals was performed again following the removal of Items 13 and 24. Items residuals correlated similarly to the previous analysis (Table 4.10) and loaded similarly to the previous PCA (Table 4.11) - the first component explaining 21.96% of the total variance with an eigenvalue of 3.30 and the subsequent components comparatively smaller (Table 4.12). *T*-tests were performed and the results were very similar to the first analysis with Items 13 and 24 included.

Table 4.12.

HL Item Residual Correlation Matrix

Item	1	3	4	6	7	10	11	15	17	20	22	26	28	30	31
1	1.000														
3	-0.033	1.000													
4	0.064	-0.008	1.000												
6	-0.112	0.385	0.028	1.000											
7	-0.051	0.247	-0.015	0.472	1.000										
10	-0.068	-0.298	0.000	-0.253	-0.268	1.000									
11	-0.092	-0.254	-0.147	-0.366	-0.363	0.222	1.000								
15	-0.139	-0.231	-0.186	-0.291	-0.266	-0.004	-0.036	1.000							
17	-0.022	-0.162	0.071	-0.241	-0.269	0.373	0.026	-0.021	1.000						
20	-0.134	0.220	-0.188	0.023	0.061	-0.226	-0.152	-0.117	-0.231	1.000					
22	0.042	-0.255	0.031	-0.385	-0.418	0.038	0.246	0.057	0.103	-0.151	1.000				
26	0.010	0.210	-0.160	0.237	0.260	-0.330	-0.271	-0.125	-0.257	0.133	-0.265	1.000			
28	0.009	0.020	-0.097	0.152	0.153	-0.183	-0.266	0.003	-0.201	-0.029	-0.240	0.153	1.000		
30	-0.210	-0.234	-0.202	-0.185	-0.129	-0.012	-0.067	-0.081	-0.046	-0.040	-0.062	-0.169	-0.067	1.000	
31	-0.075	0.040	-0.005	0.248	0.123	-0.207	-0.267	0.051	-0.206	-0.034	-0.218	0.106	0.328	-0.098	1.000

Correlations greater than an absolute value of .3 are highlighted

Table 4.13.

*PC Loadings of 15-item HL Scale*

Item	PC1	PC2	PC3	PC4	PC5	PC6	PC7	PC8	PC9	PC10	PC11	PC12	PC13	PC14	PC15
6	0.710	0.245	-0.025	-0.228	-0.110	0.087	0.219	-0.113	0.085	-0.190	-0.032	0.145	0.195	0.445	0.020
7	0.681	0.170	0.033	-0.221	0.073	-0.096	0.149	-0.259	-0.256	0.015	0.222	0.333	-0.233	-0.270	0.021
26	0.560	-0.104	0.128	0.220	-0.031	-0.347	-0.033	-0.305	0.269	0.508	-0.182	-0.140	0.108	-0.032	0.018
3	0.558	0.254	0.349	0.071	-0.284	0.093	-0.103	0.072	0.196	-0.402	0.044	-0.337	0.088	-0.264	0.015
31	0.418	-0.228	-0.553	0.014	-0.085	0.218	0.149	0.274	0.146	-0.044	-0.481	0.155	-0.098	-0.164	0.010
28	0.408	-0.275	-0.508	0.044	0.109	-0.188	0.083	0.381	0.143	0.074	0.496	-0.156	0.019	0.036	0.012
20	0.297	-0.262	0.585	0.099	-0.147	0.098	-0.324	0.480	-0.202	0.185	-0.015	0.196	-0.035	0.099	0.021
1	-0.021	0.363	-0.151	0.539	0.421	-0.445	-0.182	0.092	-0.192	-0.262	-0.168	0.032	0.006	0.054	0.020
4	-0.030	0.634	-0.261	0.041	0.117	0.540	-0.098	0.024	-0.208	0.331	0.019	-0.243	-0.001	0.008	0.021
30	-0.171	-0.441	0.146	-0.524	0.618	0.122	-0.158	-0.092	0.078	-0.112	-0.087	-0.144	0.045	-0.022	0.031
15	-0.251	-0.524	-0.351	0.219	-0.439	0.078	-0.249	-0.351	-0.277	-0.139	0.054	-0.072	0.018	0.050	0.029
17	-0.476	0.356	-0.136	-0.322	-0.245	-0.208	-0.383	0.026	0.367	0.008	0.012	0.069	-0.353	0.091	0.018
10	-0.551	0.208	-0.120	-0.430	-0.239	-0.318	0.075	0.187	-0.162	0.090	-0.055	0.084	0.427	-0.168	0.017
22	-0.566	0.028	0.086	0.442	0.107	0.306	0.068	-0.074	0.384	-0.010	0.190	0.365	0.184	-0.097	0.022
11	-0.572	-0.056	0.284	0.114	-0.103	-0.087	0.664	0.093	-0.054	0.007	-0.054	-0.197	-0.242	0.062	0.031

Loadings  $\pm 0.3$  highlighted

Table 4.14.

*Principal Components Summary for 15-item HL Scale*

PC	Eigen	Percent	CPercent	StdErr
PC001	3.294	21.96%	21.96%	0.455
PC002	1.536	10.24%	32.20%	0.204
PC003	1.402	9.35%	41.55%	0.184
PC004	1.278	8.52%	50.07%	0.168
PC005	1.044	6.96%	57.03%	0.133
PC006	0.995	6.63%	63.66%	0.130
PC007	0.943	6.29%	69.95%	0.116
PC008	0.819	5.46%	75.41%	0.108
PC009	0.740	4.93%	80.34%	0.102
PC010	0.715	4.77%	85.11%	0.094
PC011	0.645	4.30%	89.41%	0.086
PC012	0.622	4.14%	93.55%	0.090
PC013	0.524	3.50%	97.05%	0.081
PC014	0.436	2.90%	99.95%	0.078
PC015	0.007	0.05%	100.00%	0.049

**HL child-focused subscale analysis**

Based on the results of the tests for dimensionality of the HL subscale, the seven child-focused items were analysed together as a PRFQ-CF subscale. The analysis of the PRFQ-CF was undertaken using the Partial Credit Model and with the seven items rescored as indicated in Table 4.5.

**b) Targeting of the PRFQ-CF**

The distribution of the person and item locations can be seen in Figure 4.12. The parents tended to score high levels of the trait and although the item thresholds were reasonably well distributed across the range of locations, they were not well targeted for those high scoring parents. The Item Map (Figure 4.13) shows that even though the thresholds mostly covered the range of locations the individual item locations were limited in their distribution and targeted only the lower portion of person locations. The comparative distributions are also reflected in the mean and standard deviation (*SD*) of the item and person locations (Table 4.13). The item *SD* of 0.36 is very narrow in comparison to the person *SD* of 1.36, and the person mean of 1.01 reflects the parents' relatively high scores.

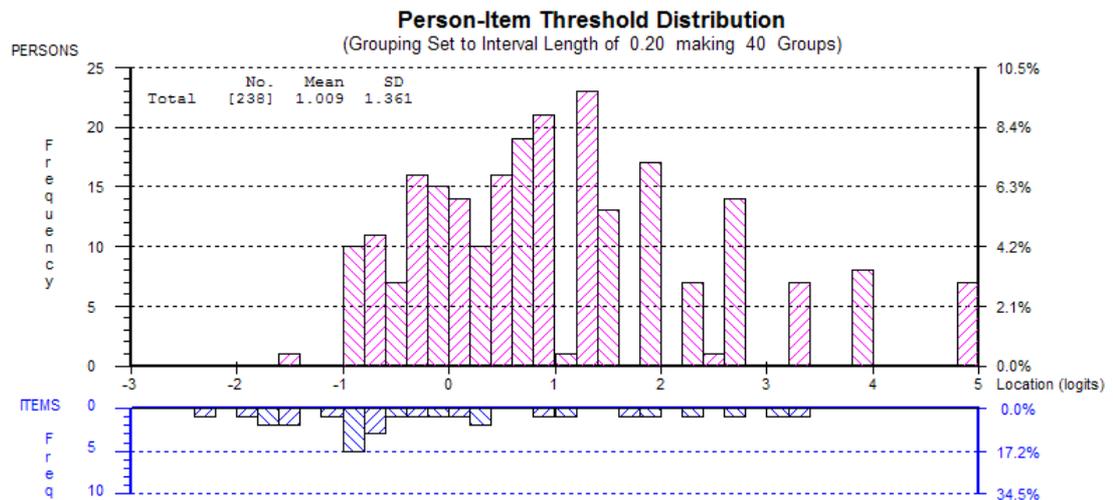


Figure 4.12. Person-Item Threshold Distribution of the PRFQ-CF subscale

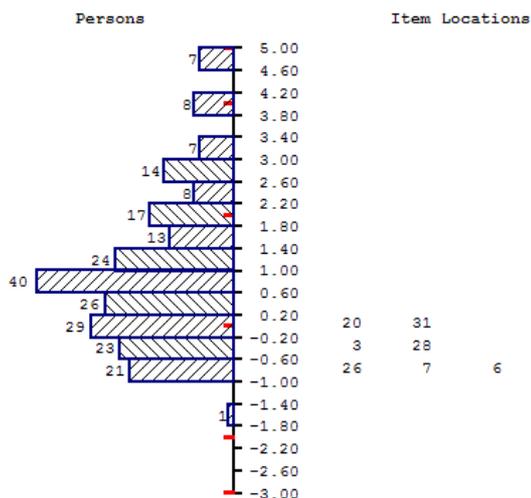


Figure 4.13. Item Map of the PRFQ-CF subscale

Table 4.15.

Item-Person Interaction Statistics for PRFQ-CF Subscale

	Items		Persons <sup>a</sup>	
	Location	Fit Residual	Location	Fit Residual
Mean	0	0.08	1.01	-0.45
SD	0.36	1.97	1.36	1.26

<sup>a</sup>n = 238

**c) Overall fit of the PRFQ-CF**

The seven items of the PRFQ-CF subscale were analysed together and were found to show overall fit to the Rasch model with an interaction chi-square of 28.93 ( $p = .116$ ).

**d) Reliability of the PRFQ-CF**

The PSI of .82 indicated good internal consistency reliability, which was only slightly lower when extreme person locations ( $n = 7$ ) were excluded (0.80).

**e) Individual item and person fit of the PRFQ-CF**

Examination of the individual person location scores and fit residuals with the PRFQ-CF subscale showed that seven parents received the highest possible score for

all items. These parent's locations were therefore considered to be extreme scores and a number of the analyses in RUMM2030 provided the option to report statistics either including or excluding these parent's scores.

All seven items fit the Rasch model according to chi-square statistics; however, two items (6 and 20) exceeded the  $\pm 2.5$  fit residual criteria (see Table 4.14 for fit statistics and the wording for these two items). Inspection of the ICCs for these two items showed only slight over discrimination for Item 6 (Figure 4.14) and slight over discrimination for Item 20 (Figure 4.15). Removal of Item 20 from the analysis resulted in similar satisfactory overall fit (chi-square 22.15;  $p = .225$ ) and slight decrease in the PSI (.81). After removal of Item 20, Item 6 showed similar misfit. With Item 6 removed the overall fit remained satisfactory (chi-square 14.32;  $p = .501$ ) and PSI was reduced to .75 (.68 with extreme scores removed).

Table 4.16.

*Individual Item-Fit Statistics for the PRFQ-CF Subscale*

Item	Location	SE	FitResid	DF	ChiSq	DF	Prob
26	-0.47	0.09	-0.21	191.57	1.13	3	0.769996
7	-0.28	0.07	-1.53	193.25	3.32	3	0.345614
6	-0.26	0.07	*-2.63	193.25	7.91	3	0.047839
3	0.01	0.10	-0.46	192.41	3.50	3	0.320696
28	0.20	0.12	0.68	189.89	2.57	3	0.4625
20	0.22	0.07	*3.43	190.73	8.74	3	0.033006
31	0.58	0.11	1.31	189.89	1.77	3	0.621791

Note: Probability evaluated against a Bonferroni-adjusted alpha-level of .001429

\*Items with fit residual statistics exceeding  $\pm 2.5$

Item 6: I like to think about the reasons behind the way my child behaves and feels.

Item 20: I wonder a lot about what my child is thinking and feeling.

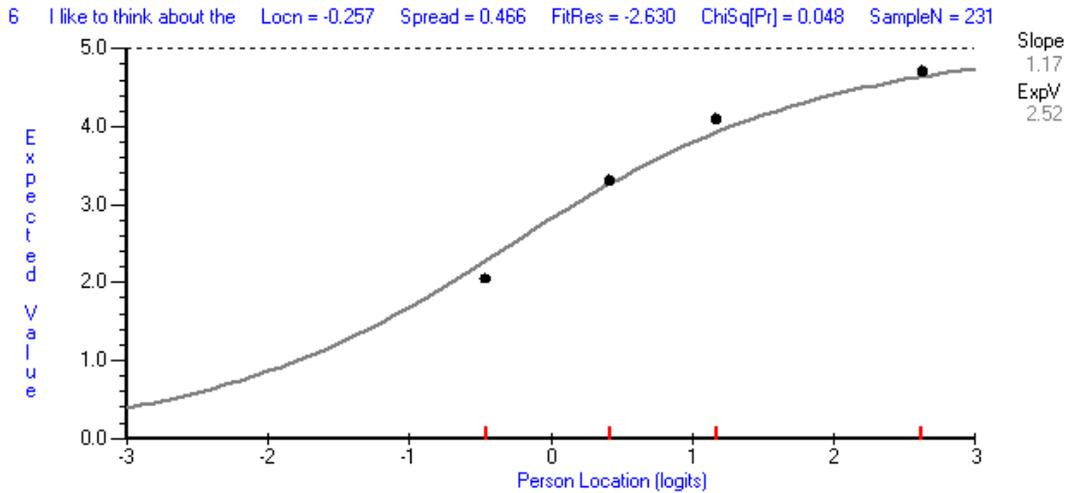


Figure 4.14. Item Characteristic Curve for the PRFQ-CF subscale item 6

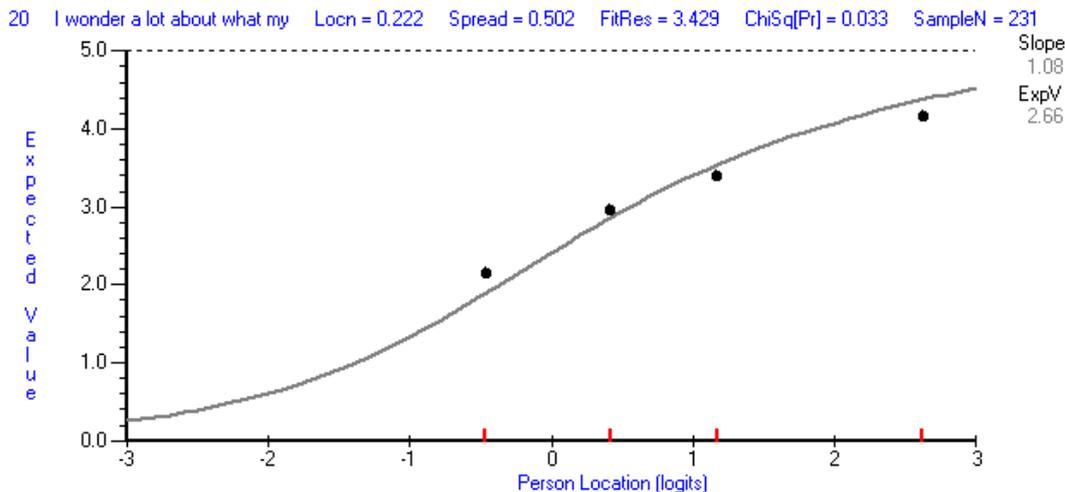


Figure 4.15. Item Characteristic Curve for the PRFQ-CF subscale item 20

Examination of the item locations showed two pairs of items that are relatively closely located: Items 6 and 7, and Items 20 and 28. The independence of these items is examined in the analysis of residual correlations (Section g).

#### *f) DIF with the PRFQ-CF subscale*

The PRFQ-CF items were examined for DIF as an indication of individual item bias with the following categorical variables: gender of parent, parent age, gender of child, parent birthplace, birth order, parent education, and parent occupation. The only items to show DIF with any of the variables at the Bonferroni-adjusted alpha-level of .002381 were Item 31 (I try to understand the reasons why my

child misbehaves) with the parent's first child or not, and Item 7 (I try to see situations through the eyes of my child) for education.

The ICC for Item 31 (Figure 4.16) shows that parents whose study child was their first child tended to score lower than other parents for Item 31, and this was consistent across parents in all class intervals. When this item was split such that it was considered a different item for the two groups of parents, the item locations showed a difference of over one logit, with the item for parent of a first child showing the highest location (0.122; 1.135). With this item split there was no improvement on PSI, the overall fit reduced but remained satisfactory (chi-square 33.87;  $p = .087$ ), and there was no change in the item location order since both of the items resulting from the split were at the highest end of the scale.

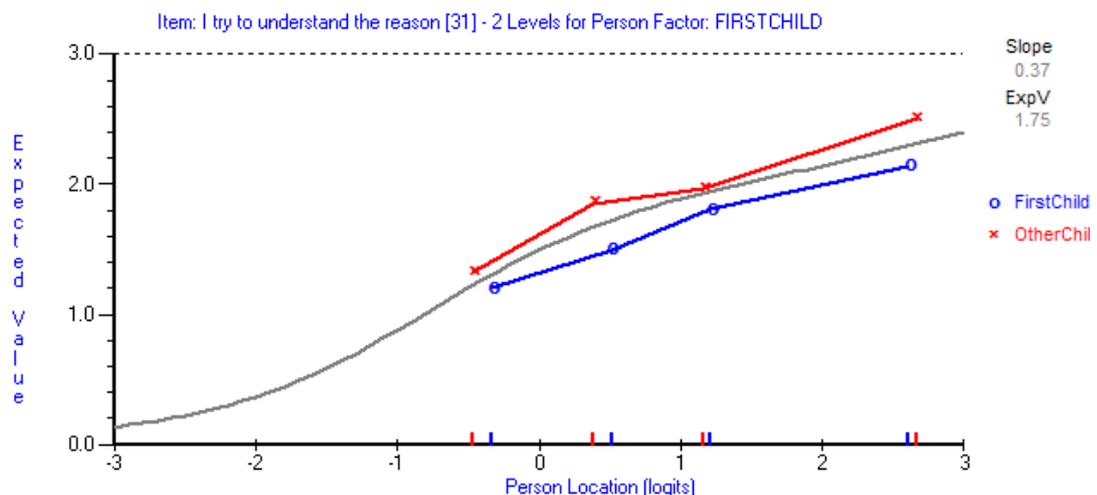


Figure 4.16. ICC for item 31 of the PRFQ-CF subscale showing DIF with parent's first child or not

Item 7 DIF is shown graphically in the ICC display (Figure 4.17). There is evidence that parents with a Year-12 level of education scored lower than expected for this item for all but the highest class-interval. Also notable is the variation of expected values for the less than Year-12 category, with one class interval being less than expected and two others being higher than expected.

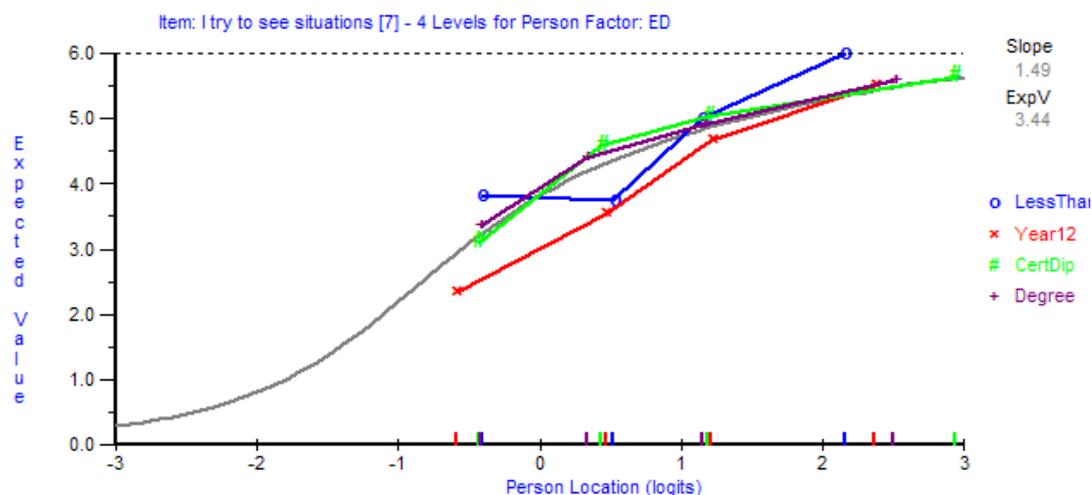


Figure 4.17. ICC for item 7 of the PRFQ-CF subscale showing DIF with parent's education

***g) Local dependence and h) dimensionality***

Analysis of fit residual correlations and PCA showed no clear evidence of local item dependence or dimensionality. Therefore items 6 and 20 (which showed slight misfit to the model and similar location estimates to other items) were retained and further analysis was undertaken with the 7-item PRFQ-CF.

***i) Variance in PRFQ-CF scores accounted for by categorical parental characteristics***

A one-way ANOVA was performed to compare groups of persons using the same seven categorical variables that were examined in the DIF analyses: gender of parent, parent age, gender of child, parent birthplace, birth order, parent education, and parent occupation. From all the analyses, only parent gender and education variables showed any differences between groups.

The mean of mother and father scores were found to be statistically different ( $F[1,236] = 11.41; p < .001$ ), with mothers ( $M = 1.30; SD = 1.47$ ) scoring on average higher than fathers ( $M = 0.72; SD = 1.17$ ), and this difference remained significant with extreme person locations excluded ( $F [1,229] = 7.58; p = .006$ ). A comparison of the distributions including extreme person locations can be seen in Figure 4.18.

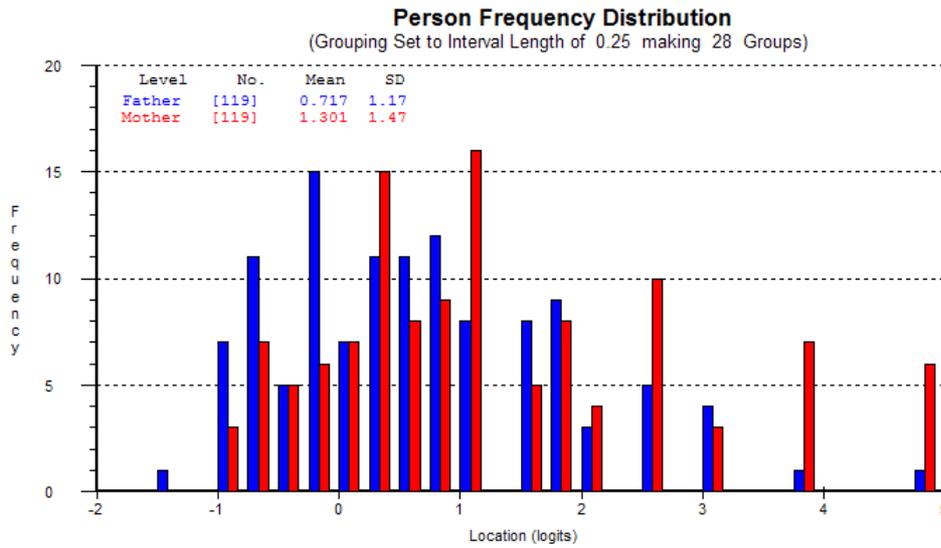


Figure 4.18. Distribution of mother and father locations of the PRFQ-CF Subscale

A Pearson correlation was computed to further examine the relationship between mother and father (couple) PRFQ-CF scores. The analysis showed there was no association ( $r = -.001$ ), as seen in Figure 4.19.

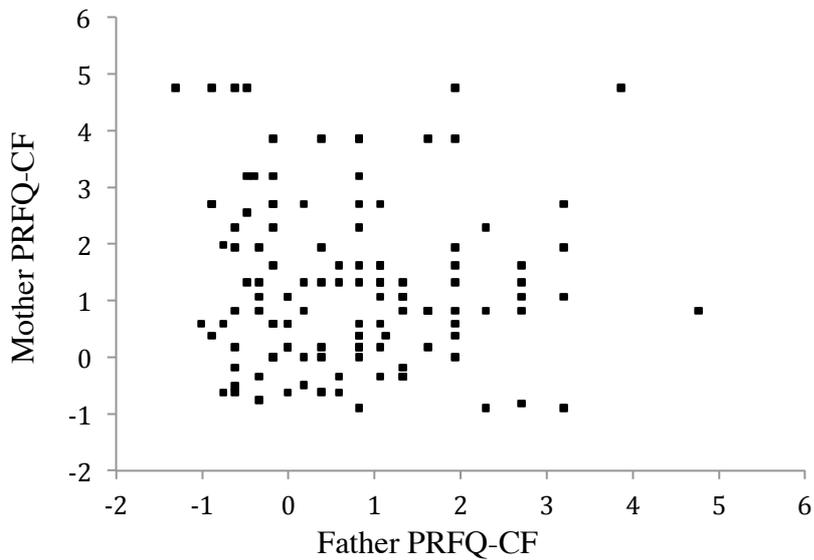


Figure 4.19. Plot of mother and father PRFQ-CF scores

When parents were grouped into four categories based on their level of education, the differences between the mean scores was not quite at a level of significant difference with extreme scores included ( $F[3,215] = 2.48; p = .062$ ), although the difference was significant with extreme scores excluded ( $F[3,208] = 3.29; p = .021$ ). Examination of the means shows that with higher levels of education, parents on average have higher person locations. The order of location means for levels of education is sequential with the extreme scores removed (Figure 4.20), however with extreme score included (Figure 4.21) the *Less than Year 12* category had a mean that was higher than the *Year 12* category and similar to the *CertDip* category.

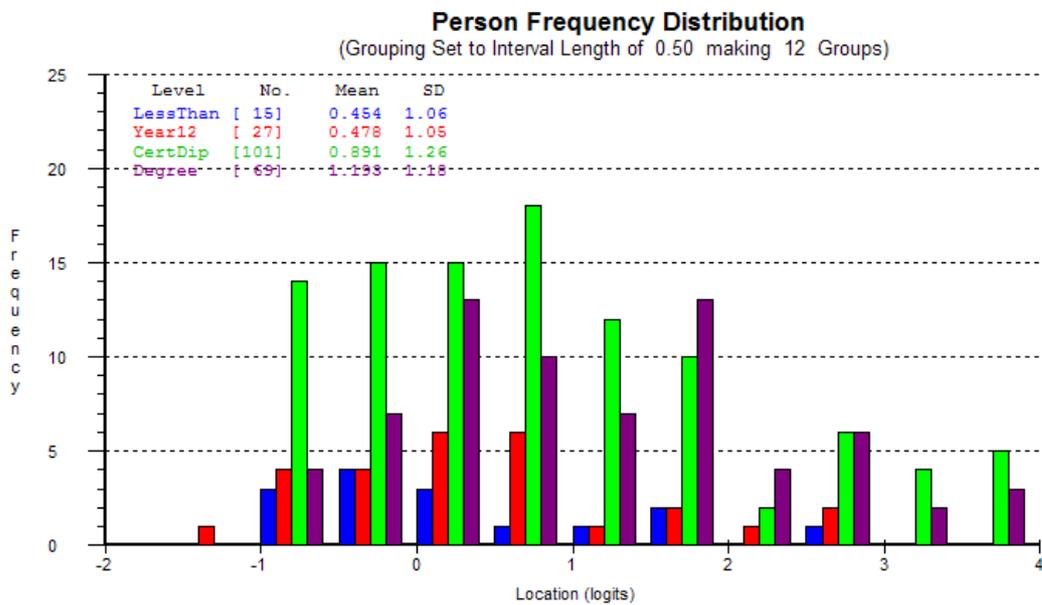


Figure 4.20. Distribution of locations showing levels of parent’s education (without extreme scores) of the PRFQ-CF Subscale

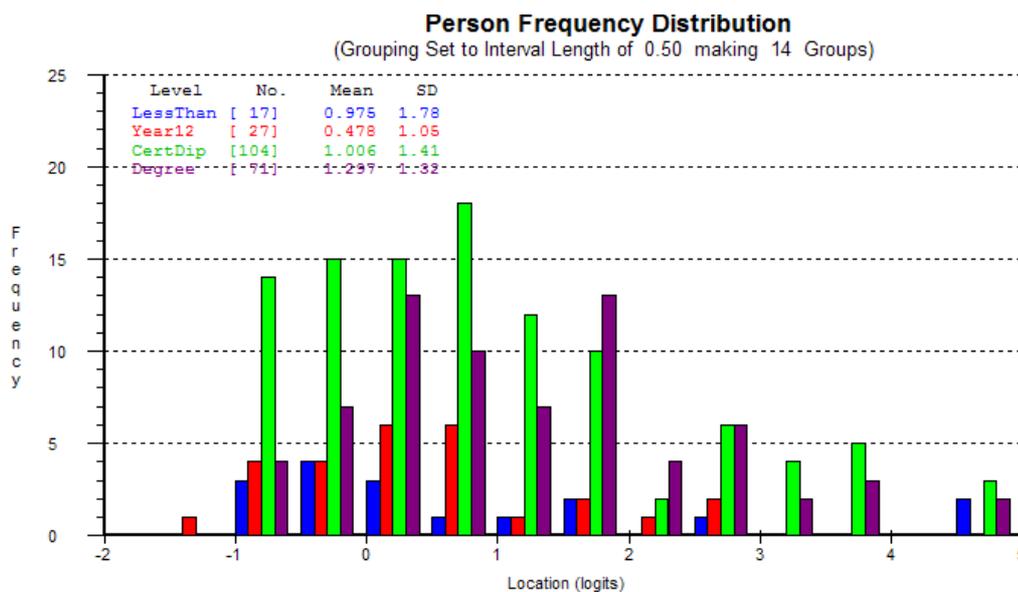


Figure 4.21. Distribution of locations showing levels of parent’s education (with extreme scores) of the PRFQ-CF Subscale

**HL self-focused subscale analysis**

The 10 items from the self-focused subscale (PRFQ-SF) were analysed together and the following results are presented in a condensed form. The results showed overall fit to the Rasch model with an interaction chi-square of 46.73 ( $p = .026$ ), and PSI of .66 (see Table 4.15 for detailed location and fit residual statistics). The individual item-fit statistics showed all items fitting according the chi-square statistic and only item 30 with a slightly elevated fit residual of 2.61 (Table 4.16). Item thresholds were well distributed over the levels of the trait and well targeted to the person locations that were normally distributed (Figure 4.22).

Table 4.17.

*Item-Person Interaction Statistics for PRFQ-SF Subscale*

	Items		Persons <sup>a</sup>	
	Location	Fit Residual	Location	Fit Residual
Mean	0	0.59	-0.16	-0.26
SD	0.50	1.16	0.52	1.27

<sup>a</sup> $n = 238$

Table 4.18.

*Individual Item-Fit Statistics for the PRFQ-SF Subscale*

Item	Location	SE	FitResid	DF	ChiSq	DF	Prob
1	-0.68	0.07	0.75	205.80	1.44	3	0.696637
4	-0.63	0.06	-0.08	209.33	2.55	3	0.466799
15	-0.27	0.05	0.70	207.57	1.86	3	0.60272
24	-0.22	0.07	2.21	207.57	7.15	3	0.067423
11	-0.12	0.05	-1.05	207.57	11.19	3	0.010741
30	0.02	0.05	2.61	207.57	3.45	3	0.327002
22	0.08	0.07	-0.11	208.45	5.84	3	0.119894
13	0.36	0.06	0.93	208.45	2.20	3	0.532567
17	0.52	0.08	0.71	203.15	4.60	3	0.203203
10	0.93	0.08	-0.72	207.57	6.47	3	0.091018

Note: Probability evaluated against the Bonferroni-adjusted alpha-level of .001000

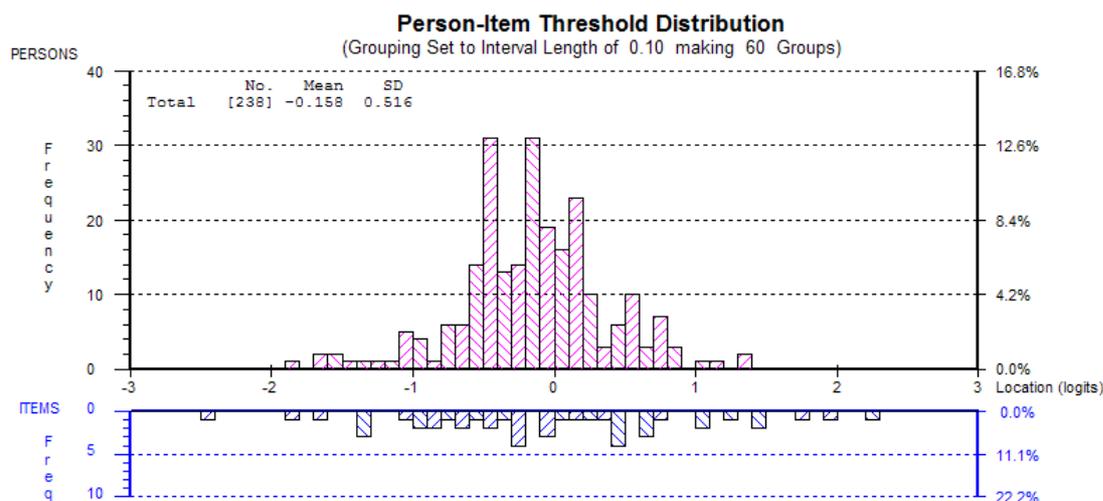


Figure 4.22. Person-Item Threshold Distribution of the PRFQ-SF subscale

No difference was found between means of mothers and fathers (Figure 4.23), with mothers ( $M = -0.131$ ;  $SD = 0.56$ ) only slightly higher than fathers ( $M = -0.185$ ;  $SD = 0.47$ ). The only difference between means for the other variables was for education, with level of education increasing with level of the trait (Means and SDs detailed in Figure 4.24).

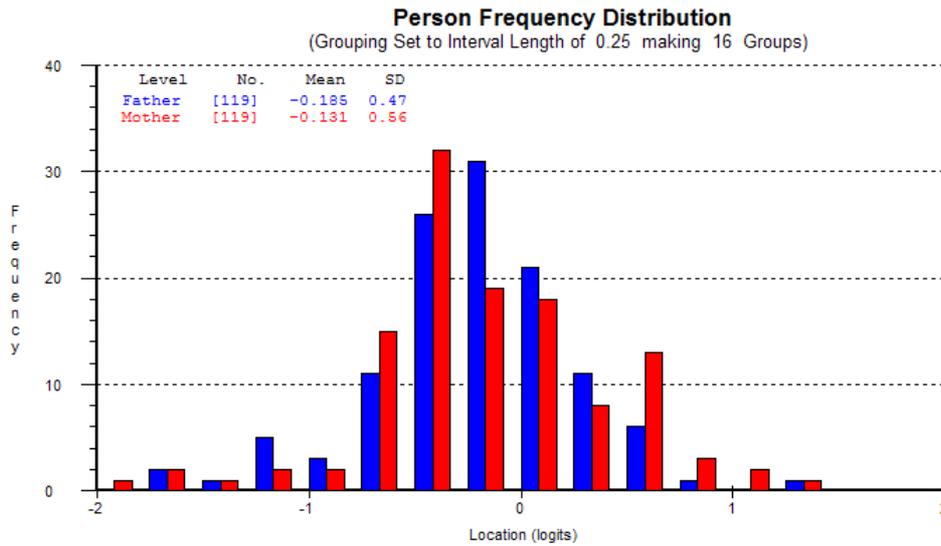


Figure 4.23. Distribution of mother and father locations of the PRFQ-SF Subscale

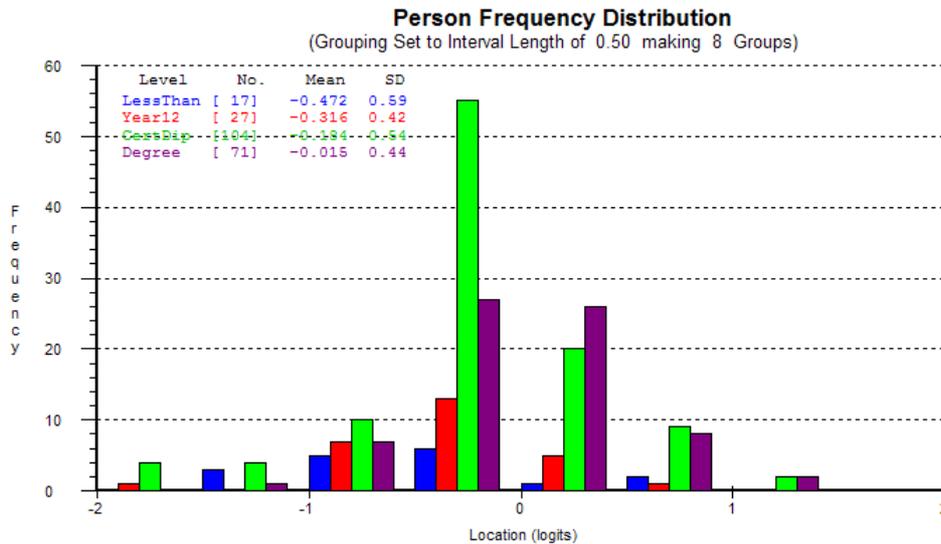


Figure 4.24. Distribution of locations showing levels of parent’s education of the PRFQ-SF Subscale

Items 1, 4, and 13 showed DIF for gender of parent, with Items 1 and 4 showing consistent higher scores than expected for mothers, and Item 13 showing higher scores for fathers. No other DIF was found for any items with any other variables. There was no evidence of local item dependence or dimensionality with the PRFQ-SF items from the examination of item residuals.

### LH subscale analysis

The same processes of analysis that were undertaken with the HL Subscale were applied to the LH Subscale and the most relevant results are reported in brief. The subscale was best analysed with the Partial Credit Model. The category frequencies (Table 4.17) showed very low frequencies for item categories representing low levels of PRF (high scores). Seven of the items were particularly skewed in their distribution of category frequencies (Items 5, 9, 14, 23, 25, 34 and 36), with 50% or more of the responses in the first *strongly disagree* category and less than 9% of responses in any the three *agree* categories.

Table 4.19.

#### *Category Frequencies for LH Subscale*

Item	Abbreviated Item Wording	Response Category						
		0	1	2	3	4	5	6
5	My child knows when I am	119	51	17	29	14	3	3
9	My child sometimes gets sick	194	24	5	9	2	0	2
12	I find it hard to actively	107	65	20	26	14	3	1
14	When my child is fussy	163	45	11	13	0	0	4
18	My behaviour towards my child	31	47	34	59	22	28	11
21	Often, my child's behaviour is	90	89	28	22	4	1	1
23	When my child is misbehaving	206	20	5	3	0	0	1
25	My child cries around stranger	200	23	5	5	1	0	1
27	I can completely read my child	50	59	34	42	40	8	0
29	I believe there is no point in	72	84	29	23	8	13	5
33	I hate it when my child cries	59	54	17	50	35	12	5
34	The only time I'm certain my	159	50	7	12	3	1	1
36	The best way to know your	124	61	16	22	7	5	0
37	My child's temperament is what	34	61	47	40	22	15	16

Threshold disorder for all but two items was addressed by rescoring (Table 4.18) and the resolved category thresholds are displayed in Figure 4.25. The LH Subscale showed overall misfit to the Rasch model according to the item-trait interaction chi-square statistic (227.43;  $p < .001$ ), and had a low reliability PSI of .66. Five items showed misfit according to the chi-square statistics and an additional three items showed fit residuals above 2.5 (Table 4.19). Item-person interaction statistics are presented in Table 4.20. Through a process of examining evidence of misfit and removal of items, five items were removed from the subscale, which

improved the chi-square statistic (59.85;  $p < .001$ ) but with minimal change of the PSI (.65; with extreme person scores removed PSI = .62).

Table 4.20.

Category Scoring for 14-item LH Subscale

Item	Max Score	Score for each category						
		1	2	3	4	5	6	7
5	4	0	1	1	2	3	4	4
9	2	0	1	1	1	2	2	2
12	3	0	1	1	2	2	3	3
14	2	0	1	1	2	2	2	2
18	8	0	1	2	3	4	5	6
21	5	0	1	2	3	4	5	5
23	2	0	1	1	2	2	2	2
25	2	0	1	1	2	2	2	2
27	3	0	1	1	2	2	3	3
29	3	0	1	2	2	2	3	3
33	6	0	1	2	3	4	5	6
34	3	0	1	1	2	2	3	3
36	3	0	1	1	2	2	3	3
37	4	0	1	2	2	3	3	4

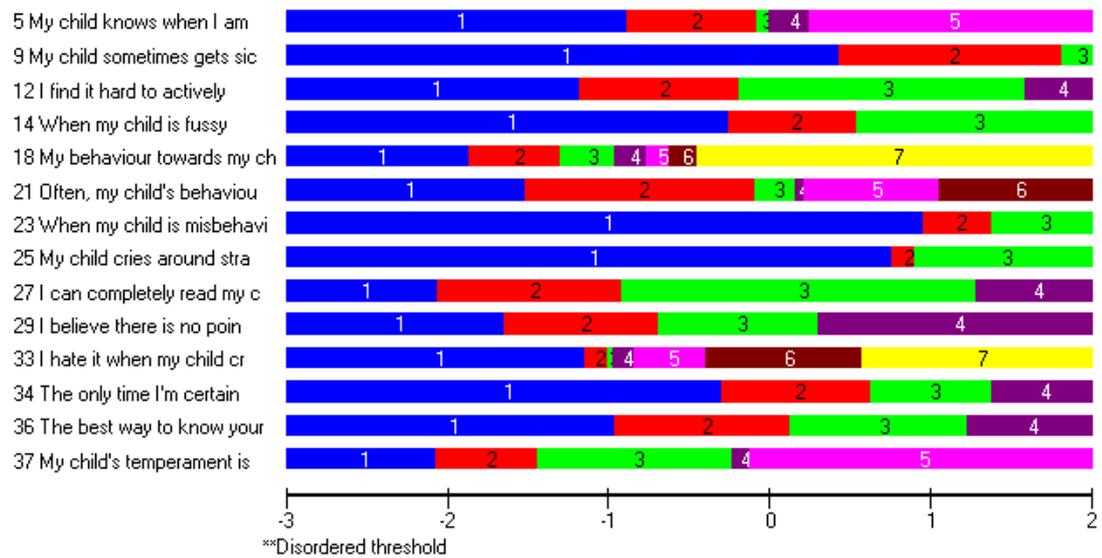


Figure 4.25. Threshold Map for 14-item LH Subscale

Table 4.21.

*Individual Item-Fit Statistics for 14-item LH Subscale*

Item	Location	SE	FitResid	DF	ChiSq	DF	Prob
18	-0.99	0.05	*5.02	212.28	30.21	3	**0.000001
37	-0.97	0.07	*3.37	215.03	15.29	3	0.001586
29	-0.68	0.08	1.27	214.11	4.56	3	0.206983
33	-0.63	0.05	*3.05	212.28	8.14	3	0.043194
27	-0.56	0.09	*3.30	213.20	13.06	3	0.004503
5	-0.18	0.07	-0.48	215.94	7.36	3	0.06128
21	-0.04	0.07	-0.33	215.03	15.35	3	0.001539
12	0.08	0.09	-0.46	215.94	10.66	3	0.013699
36	0.13	0.09	-1.27	215.03	16.46	3	0.000912
14	0.14	0.11	-2.37	215.94	27.47	3	**0.000005
34	0.57	0.11	-1.50	213.20	15.37	3	0.001531
25	0.83	0.15	-2.00	215.03	20.37	3	**0.000143
9	1.13	0.16	-1.70	215.94	23.83	3	**0.000029
23	1.17	0.17	-2.03	215.03	19.30	3	**0.000238

\*Items with fit residual statistics exceeding  $\pm 2.5$ 

\*\* Probabilities are evaluated against the Bonferroni-adjusted alpha-level of 0.000714

Table 4.22.

*Item-person Interaction Statistics for 14-item LH Subscale*

	Items		Persons <sup>a</sup>	
	Location	Fit Residual	Location	Fit Residual
Mean	0	-0.25	-2.13	-0.29
SD	0.60	1.67	1.27	0.87

<sup>a</sup> $n = 238$ 

The person-item threshold distribution is displayed in Figure 4.26 shows no item thresholds at the location levels of the lowest scoring parents (assumed to reflect highest levels of PRF) and no persons corresponding to the threshold locations at the high scoring end of the scale. Examination of mean scores showed fathers ( $M = -1.52$ ;  $SD = 0.98$ ) tended to score higher ( $p < .001$ ) than mothers ( $M = -2.05$ ;  $SD = 0.96$ ), and parents with a degree level of education tended to score lower ( $p < .005$ ) than other levels of education on this scale (Figure 4.27). Examination of

the residual correlations showed no clear evidence of local item response dependence or dimensionality.

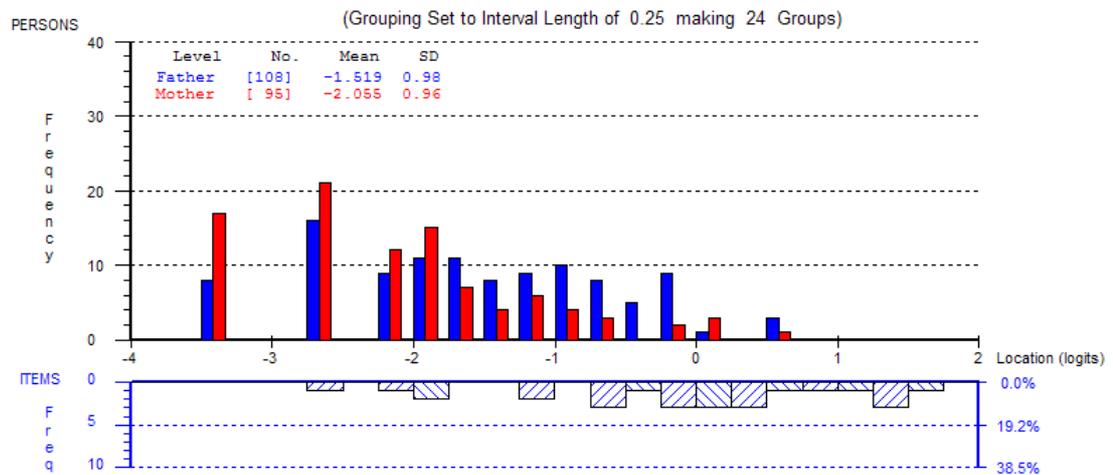


Figure 4.26. Distribution of mother and father locations, and item thresholds for 9-item LH Subscale

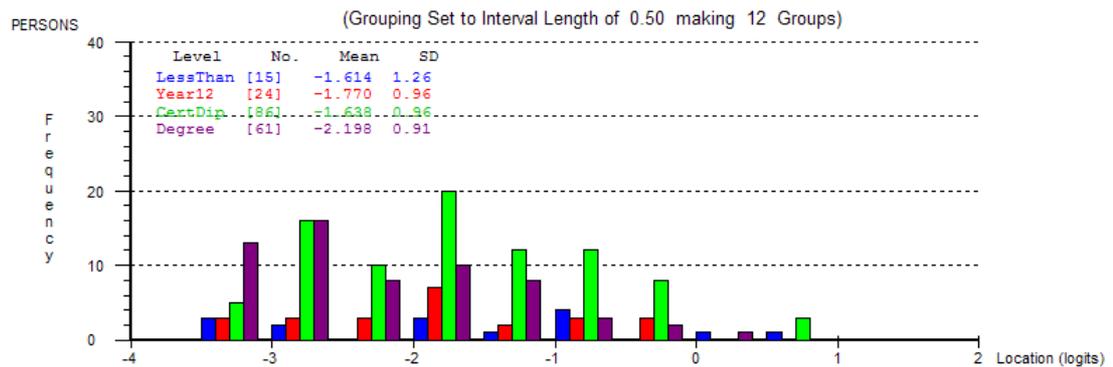


Figure 4.27. Distribution of locations by level of education for 9-item LH Subscale

### M subscale analysis

Table 4.21 presents the category frequencies of the M subscale, showing the frequencies were generally well distributed when centre scored, although items 16 and 35 showed higher frequencies towards the extremes. Frequencies are also displayed for the full seven categories of the M subscale, which show items 16 and 35 have a different pattern of responses to the other items with most responses in the “Strongly Agree” category scored as a 6. This pattern of response is not consistent with the expectation of the middle response category reflecting highest PRF.

Table 4.23.

*Category Frequencies for Centre Scored M Subscale*

Item	Abbreviated Item Wording	Response Categories			
		0 & 7	1 & 6	2 & 5	4
2	When I get angry with my child	50	88	52	43
8	I always know why my child	13	65	105	53
16	No matter how sick my child is	132	71	27	5
19	I can always predict what my	14	55	110	54
32	I always know what my child	17	63	102	51
35	I'm certain that my child know	172	51	11	1
38	I always know why I do what I	36	83	66	49
39	At times I get confused about	20	62	90	63

		0	1	2	3	4	5	6
2	When I get angry with my child	4	13	9	43	43	75	47
8	I always know why my child	7	37	41	53	64	28	7
16	No matter how sick my child is	4	5	6	5	21	66	129
19	I can always predict what my	15	34	64	54	46	21	0
32	I always know what my child	16	37	52	51	50	26	2
35	I'm certain that my child know	1	1	0	1	11	50	172
38	I always know why I do what I	2	8	17	49	49	75	35
39	At times I get confused about	16	46	29	63	61	16	5

Thresholds ordering was checked for the eight M subscale items and only Item 2 showed slight disorder and required minor rescoring in the following manner: 0122210. Interestingly, the thresholds for items 16 and 35 were ordered even though the frequencies showed a pattern that was contrary to the expected pattern. The rescored M subscale showed overall fit according to the item-trait interaction chi-square statistic (25.79;  $p = .36$ ); however, the reliability index was low (PSI = .63). Item-person interaction statistics are shown in Table 4.24. All individual items showed fit according to the chi-square and residual statistics (Table 4.25).

Table 4.24.

*Item-person Interaction Statistics for M Subscale*

	Items			Persons <sup>a</sup>	
	Location	Fit	Residual	Location	Fit Residual
Mean	0	0.42		-0.11	-0.24
SD	1.02	0.53		0.81	1.00

<sup>a</sup> $n = 238$

Table 4.25.

*Individual Item-Fit Statistics for M Subscale*

Item	Location	SE	FitResid	DF	ChiSq	DF	Prob
8	-0.69	0.09	1.32	203.38	3.05	3	0.38406
19	-0.66	0.09	1.06	200.80	2.55	3	0.46655
39	-0.60	0.08	0.09	202.52	2.77	3	0.42843
32	-0.55	0.09	0.31	200.80	0.48	3	0.92321
2	-0.48	0.10	0.41	200.80	0.60	3	0.89541
38	-0.22	0.08	-0.29	201.66	7.12	3	0.06829
16	1.21	0.09	0.37	202.52	4.73	3	0.19293
35	1.98	0.12	0.12	202.52	4.50	3	0.21271

Note: Probability are evaluated against the Bonferroni-adjusted alpha-level of .001250

The Person-Item threshold distribution (Figure 4.28) shows relatively good targeting of items to persons. Fathers ( $M = -0.01, SD = 0.74$ ) tended to score higher ( $p < .05$ ) than mothers ( $M = -0.22, SD = 0.87$ ). No other variables showed differences between means or showed DIF for any items. The M subscale showed no evidence of local item dependence or dimensionality in the analysis of residual correlations.

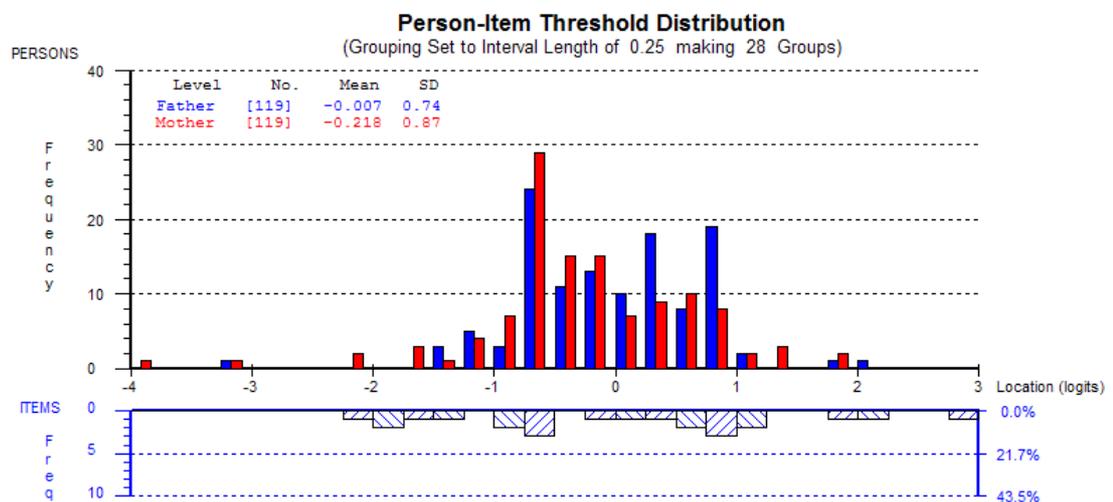


Figure 4.28. Distribution of mother and father locations, and item thresholds for M Subscale

### **Summary of Rasch analyses**

Analysis of data using the HL subscale of the PRFQ revealed disordered response category thresholds and rescoring (collapsing categories) was required to achieve ordered thresholds. Seven HL items showed conformity to the Rasch measurement model and potential for measurement of a child-focused dimension of parental reflective functioning: the PRFQ-CF. This subscale had good internal consistency reliability as indicated by a PSI of .82. The items were best targeted for low to mid levels of PRF. Some evidence of DIF was found for two items. No evidence was found of local item dependence or dimensionality. When the mean PRFQ-CF score from groups of parents were compared, mothers tended to score higher than fathers and parents with higher levels of education tended to have higher scores than those with lower levels.

The PRFQ-CF subscale showed evidence of being distinct from the ten other HL items that had mostly a self-focus (PRFQ-SF), although the internal consistency of these items as a subscale was poor. The LH and M subscales of the PRFQ were analyzed and both had indications of poor internal consistency reliability. In addition the LH subscale showed overall misfit to the Rasch model, even after removal of all individual misfitting items.

The PRFQ-CF subscale was used for further analysis of test–retest stability, convergent validity with the PDI-RF and for associations with other measures completed by the mothers and fathers.

### **Research Question 2: Is there an association between self-report of parental depression or anxiety and PRFQ scores?**

Spearman's correlations were computed between the father and mother PRFQ-CF and ordinal measures of parent depression and anxiety (Table 4.24 and 4.25). Father state anxiety scores were the only mother or father scores to significantly correlate with PRFQ-CF scores ( $n = 116$ ;  $\rho = -.166$ ;  $p = .038$ ), such that lower levels of father state anxiety corresponded with higher child-focused PRFQ scores.

Table 4.26.

*Matrix of Correlations Between Father PRFQ-CF and Father Mental Health.*

Spearman's rho		FPRFQ- CF	Father BDI	Father STAI State	Father STAI Trait
FPRFQ-CF	Correlation Coefficient	1.000			
	Sig. (1-tailed)	.			
	<i>N</i>	119			
Father BDI	Correlation Coefficient	-.106	1.000		
	Sig. (1-tailed)	.130	.		
	<i>N</i>	116	116		
Father STAI State	Correlation Coefficient	-.166*	.590**	1.000	
	Sig. (1-tailed)	.038	.000	.	
	<i>N</i>	116	116	116	
Father STAI Trait	Correlation Coefficient	-.027	.424**	.498**	1.000
	Sig. (1-tailed)	.395	.000	.000	.
	<i>N</i>	102	99	99	102

Table 4.27.

*Matrix of Correlations Between Mother PRFQ-CF and Mother Mental Health.*

• Spearman's rho		MPRFQ- CF	Mother BDI	Mother STAI State	Mother STAI Trait
MPRFQ-CF	Correlation Coefficient	1.000			
	Sig. (1-tailed)	.			
	<i>N</i>	119			
Mother BDI	Correlation Coefficient	.044	1.000		
	Sig. (1-tailed)	.324	.		
	<i>N</i>	109	90		
Mother STAI State	Correlation Coefficient	-.079	.677**	1.000	
	Sig. (1-tailed)	.208	.000	.	
	<i>N</i>	109	110	110	
Mother STAI Trait	Correlation Coefficient	-.099	.438**	.546**	1.000
	Sig. (1-tailed)	.152	.000	.000	.
	<i>N</i>	109	102	102	109

**Research Question 3: Do one or more specific sets of PRFQ items show temporal (test–retest) stability?**

The PRFQ was completed a second time by parents who participated in the PDI interviews, which was on average 45 days ( $SD = 32$ ) after the date of the first administration of the questionnaire. A Pearson's correlation was computed with the PRFQ-CF Rasch locations from the two administrations of the PFRQ. Mothers test-retest scores correlated at  $r = .64$  ( $n = 36$ ) and fathers at  $r = .29$  ( $n = 32$ ).

An invariance plot was created (Figure 4.29), which shows the 12 locations that fall outside of the 95% confidence lines created with the individual item error estimates. One parent (a father) had the lowest score for the first test and an extreme high score for the retest, which can be easily seen as an outlier in the top left of the plot. With this father's scores removed from the analysis, the reliability coefficient for the fathers' test-retest improved to  $.52$  ( $n = 31$ ).

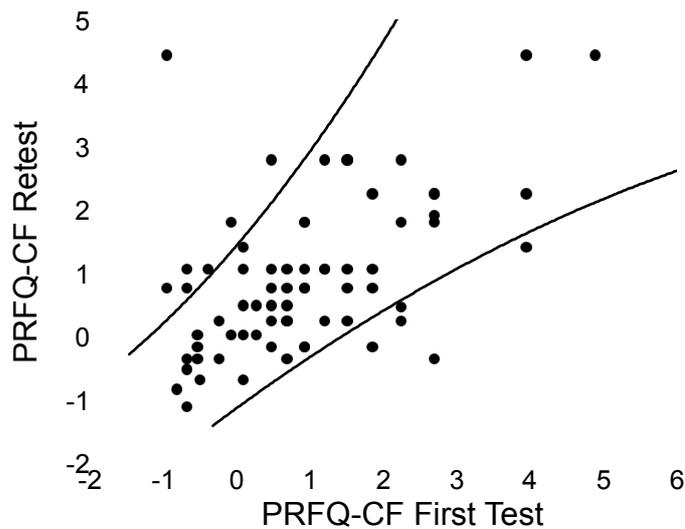


Figure 4.29. Invariance Plot of PRFQ-CF Test and Retest

The mean of the retest scores ( $M = 1.38$ ;  $SD = 1.23$ ) was higher than the mean of the first test scores ( $M = 0.93$ ;  $SD = 1.30$ ) by a difference of 0.45. A  $t$ -test showed the scores for the two tests differed significantly ( $p < .05$ ) for 17.39% ( $n = 12$ ; 6 mothers and 6 fathers) of the parents.

**Research Question 4: Is there a relation between PRFQ scores and PDI-RF scores that demonstrates convergent validity?**

Spearman's correlations were computed to determine the association between PRFQ-CF scores and the ordinal PDI-RF scores. They were found to have a weak positive (non-significant) association, which was similar for fathers ( $n = 39$ ;  $\rho = .13$ ;  $p = .21$ ) and mothers ( $n = 39$ ;  $\rho = .14$ ;  $p = .19$ ).

**Results Summary**

Rasch analysis was performed separately with each of the three subscales of the PRFQ. The HL subscale was found to have two dimensions, with some items tending to have a self-focus (PRFQ-SF) and others a child-focus (PRFQ-CF). Seven items of the PRFQ-CF were found to conform to the requirements of the Rasch model, with good reliability and no clear evidence of dimensionality. The PRFQ-CF

analysis was reported in detail and this scale was used for the analysis in addressing Research Questions 2, 3 and 4.

The key findings from the detailed analysis of the PRFQ-CF data are as follows: response categories were mostly not responded to as would be expected and required rescoring to achieve ordered thresholds (this was also the case for the other HL and LH items); targeting of items to the sample was acceptable for screening purposes; the range of item difficulties was limited although compensated for by the range of category thresholds; the data generally fit the Rasch model with only minor evidence of misfit for two items; reliability was adequate for basic research purposes; items were generally invariant for all demographic variables considered; the data showed no evidence of local item dependence or dimensionality; and both mothers and parents with higher levels of education were found to have higher PRFQ-CF scores (no difference in scores were found for parents grouped by child gender, parent age, parent birthplace, parent occupation or child's birth-order).

The analysis of the PRFQ-SF and the remaining two subscales of the PRFQ (LH and M) found these scales to demonstrate low reliability. In addition the LH subscale showed overall misfit to the Rasch model with and without individual misfitting items included in the analysis. The analysis of the PRFQ-SF, LH and M subscales was reported in brief and the evidence of specific shortcomings was highlighted. Further analysis of these subscales was beyond the scope of this study's research questions and focus. Summary statistics and results for the two HL subscales, the LH and the M subscale are presented in Table 4.26.

Table 4.28.

*Summary Statistics for Rasch Analysis of PRFQ Subscales*

Subscale	Threshold Disorder	Model Fit <sup>b</sup>	Individual Item Fit	PSI	DIF	Dependence or Dimensionality
PRFQ -CF	6/7 items disordered	28.93 ( $p = .116$ )	Slight indications of misfit for 2 items.	.82	None	None
PRFQ -SF	8/10 items disordered	46.73 ( $p = .026$ )	Slight indications of misfit for 1 item.	.66	3 items for Parent Gender	None
LH	12/14 items disordered	59.85 <sup>c</sup> ( $p = .0003$ )	8 items with evidence of misfit	.62 <sup>c</sup>		
M	2/8 items disordered	25.79 ( $p = .36$ )	All fit	.63	None	None

<sup>a</sup>Items with disordered thresholds were rescored prior to further analysis

<sup>b</sup>Overall item-trait interaction chi-square. Significant result ( $p < .05$ ) indicates misfit to the model

<sup>c</sup>Results with 5 misfitting items removed

## Chapter 5 Discussion

This study was an investigation of parental reflective functioning (PRF) with mothers and fathers, with a focus on the validity of the new rating scale instrument, the Parental Reflective Functioning Questionnaire (PRFQ). The literature review highlighted two central themes of this thesis: the utility of the Rasch measurement model in the analysis of rating scales and the need for greater understanding of father-infant relationship quality. A key goal of this thesis is to advance the development of an instrument that can assess the PRF of fathers with infants, and therefore help identify one of many important characteristics of healthy father-child relationships and optimal child development.

The PRFQ was administered to 120 couples with a one-year old child and 40 of these couples were also interviewed with the Parent Development Interview (PDI), which was scored for Parental Reflective Functioning (PDI-RF). Firstly, the PRFQ was analysed to examine the validity of the scale with mothers and fathers. This analysis identified a subset of PRFQ items that showed the best conformity to the requirements of the Rasch measurement model. This subscale was then used for further analyses of temporal (test-retest) stability and convergent validity with PDI-RF scores. Finally, scores from mental health, parenting self-efficacy and family functioning scales were examined for associations with the PRFQ subscale scores. An important aspect of this study was the inclusion of fathers and the comparisons with the mothers.

The detailed analysis of the PRFQ with the Rasch Model has been presented in the Results section of this thesis. The first part of this discussion considers the results of this analysis, including consideration of item and subscale characteristics that were problematic. The analysis of the proposed 7-item PRFQ Child- Focused subscale (PRFQ-CF) is addressed in detail since this set of items showed better conformity to the Rasch measurement model compared to other potential subscales of the PRFQ. The other PRFQ subscales are discussed briefly to identify issues for further investigation. Subsequent sections of the discussion consider the results of the PRFQ-CF test-retest stability, convergent validity with the PDI-RF, and associations with parent characteristics as reported by self-report questionnaires. Finally, the limitations of this study and the implications for assessing PRF are discussed.

## **Rasch Analysis of the PRFQ**

A rationale has been presented in this thesis for using the Rasch measurement model for the analysis of the PRFQ. It was argued that there are benefits for using an experimental approach of fitting data to a measurement-based model – a model with clear requirements for measurement of a latent trait (Andrich, 1989). The 39-item PRFQ was developed by (Luyten et al., 2009) as three subscales: HL, LH and M subscale. Each subscale is scored in a particular manner and represents different aspects of parental reflective functioning. Therefore, the three subscales were analysed separately and the results are discussed in the following three sections. A subscale within the HL subscale, the 7-item PRFQ-CF, was analysed in detail and presented as the most viable set of items that measure a specific aspect of PRF, namely, child-focused interest and attention to mental states.

### **HL subscale analysis**

The likelihood ratio test indicated that an analysis of the HL subscale with threshold distances varying across the items (Partial Credit Model) provided greater information from the data than an analysis with uniform thresholds (Rating Scale Model). As previously discussed, non-uniform thresholds are not problematic for the Rasch model analysis and do not alter the ability to examine the data for conformity to the requirements of the model. It is not surprising to find that parents responded to the seven rating categories of HL items in different ways for different items, as this is common for other rating scales in the social sciences, e.g., (Hagquist, Bruce, & Gustavsson, 2009; Pallant & Tennant, 2007; Parkitny et al., 2012); however, the variation does highlight the incorrect assumption with classical approaches to rating scales that all item categories have equal weighting and each item contributes equally towards the total scale score. Further development of the PRFQ subscales may result in subscales with rating categories that do suit the more efficient or parsimonious meaning achieved with consistent rating categories that can be analysed with the Rating Scale Model.

The Rasch analysis of the complete 17-item HL subscale is reported in full in the Results section, however only two particular analyses will be addressed in this discussion, namely, the examination of response category threshold order and scale dimensionality. Threshold disorder was resolved with rescoring of items, which applied to all subsequent analysis of those items. The finding of two dimensions of

the HL subscale led to two separate analyses of each subscale. The other analyses are addressed in the detailed analysis of the PRFQ-CF subscale.

### ***HL threshold disorder***

Examination of the thresholds showed that most of the HL items had disordered thresholds (Table 4.4 in the Results chapter) and required rescoring prior to further analysis. A thorough investigation of the reasons for threshold disorder would require a qualitative analysis such as a parent interview regarding their responses – which was beyond the scope of this study’s research questions and focus. The three items with ordered thresholds do not appear to have any common feature that distinguishes them from the other items. What is clear from the data is that these parents have not responded to the categories in a manner that would be expected (as described in the literature review). It may be that seven categories created too fine a distinction for these parents to clearly identify one particular category that represented their degree of agreement. The lack of labels for each of the seven categories could further hinder the parent’s ability to clearly identify a specific category that represents their response. This fine distinction between categories could be particularly difficult for parents who have low reflective functioning and are less likely to have considered the issues implied in these questions. Also, the terms *strongly agree* and *strongly disagree* may not be the most appropriate labels for the extreme categories. It could be imagined that to *agree* to an item such as Item 1 (My child and I can feel differently about the same thing) would indicate sound PRF, whereas to *strongly agree* may not necessarily indicate higher PRF but could possibly reflect an overemphasis of the difference of feelings between a parent and child. Perhaps the extremes of the scale could be labelled *agree* and *disagree* to avoid this overemphasis.

In most cases the threshold disorder was around the middle and the disagreement (low scoring) end of the scale. As mentioned, the parents with low PRF were the most likely to select these categories and therefore their low capacity for reflection may have contributed to the unexpected response patterns. Although many of the items had low frequencies of responses (which has been argued to contribute to threshold disorder) in the low-end categories, the disordered items did not correspond only to items with low frequencies at the low-end (see Table 4.3 in the Results chapter). For example, Item 7 had relatively low frequencies at the low end

of the scale and showed no disorder of thresholds, and Item 15 had fairly evenly distributed frequencies and yet showed disorder among the low end of the scale.

It is interesting to note the unexpectedly high frequencies at the low-end of the scale for Item 10. Once this item was rescored to four categories, the thresholds were ordered and the showed acceptable fit statistics. The location of this item indicates it is was by far the most difficult item to agree with, which suggests why the response pattern was contrary to the other more agreed to items. Further qualitative analysis of responses to this item are required to confirm this conclusion.

The use of an odd number of categories (with a middle category) is known to be problematic for rating scale measurement (Andrich, 1998; Enos, 2001). In this study, it is likely that the middle category was used in ways that were not related to PRF and therefore would have contributed to threshold disorder and measurement error. Parents could have used the middle category in response to confusion about the question, or when their response depended on contextual factors, or if they thought the question was not relevant or applicable to them. The use of a middle category is particularly problematic in this case where the instructions to the questionnaire refer to the midpoint as *neutral or undecided*. Use of the middle category as “undecided” is clearly not a point situated on the continuum of the latent variable between *strongly agree* and *strongly disagree*. Future research with this scale could remove this option and have only four or six categories. Alternatively, allow this response as a separate category, which would allow for a clear analysis of how frequently items were not relevant or were confusing. Remaining data could be analysed as usual with these responses entered as missing data, which is not problematic for Rasch analysis.

The HL items were rescored to achieve ordered thresholds with six items requiring categories to be ‘collapsed’ to four categories. The post hoc collapsing categories is a useful experimental strategy although not ideal because the meaning of new categories is different from what the respondents were presented. If respondents had been presented items with the revised response format, the items may have elicited a different response. Also, the categories for each item are not interpreted in isolation – they are associated with the set of items and categories that were administered together. Ideally, all items in a set would have the same number and labelling of categories unless there was a substantive reason for difference. If the

HL scale were to be developed further, therefore, this evidence points towards using four-categories for all items, which would have the benefits of not requiring an extreme response of *strongly* agreeing or disagreeing and not requiring an ambiguous response for the middle category.

### *Self-other dimensions of parental interest in mental states*

The process of examining the dimensionality of the HL subscale resulted in the identification of two distinct HL subscales (Table 4.7 in the Results chapter): seven items that tended to focus on the child's behaviour and mental states (PRFQ-CF), and 10 items that focused on the parent's behaviour and mental states (PRFQ-SF; repeated in Table 5.1). Although the capacities for reflection on mental states of the self and other are closely related with regards to development (Allen, Fonagy, & Bateman, 2008), neurological activity (Lombardo et al., 2010) and pathology (Fonagy, Bateman, & Luyten, 2012), the dimensionality found in the HL items confirms the self-other polarity of mentalizing described by Fonagy and colleagues (Fonagy et al., 2012; Fonagy & Luyten, 2009). This self and child focused distinction has also been identified in a factor analysis of the individual question PRF scores from the PDI with a small sample of mothers (Suchman, DeCoste, Leigh, & Borelli, 2010). These two facets of PRF could be understood to have commonality with the concepts of empathy (child-focused PRF) and mindfulness (self-focused PRF) within the context of the parent-child relationship.

Table 5.1.

#### *Seven Items of the PRFQ-CF*

---

3	I am often curious to find out how my child feels.
6	I like to think about the reasons behind the way my child behaves and feels.
7	I try to see situations through the eyes of my child.
20	I wonder a lot about what my child is thinking and feeling.
26	I pay attention to what my child is feeling.
28	Understanding why my child behaves in a certain way helps me not to be upset with him or her.
31	I try to understand the reasons why my child misbehaves.

---

In relation to the four polarities of mentalizing identified by Fonagy and Luyten (Fonagy & Luyten, 2009) the PRFQ-CF is predominantly child (other) focused, however the items also include aspects of self-mental states as well, such as curiosity (Item 3), wonder (Item 20) paying attention (Item 26), and understanding (Items 28 and 31). The items could be understood to reflect only the explicit or controlled aspect of mentalizing, since the assessment is a self-report and the items are understood and rated at face value without any implied meanings. Implicit or automatic aspects of mentalizing are more likely to be assessed with observational or interview assessments. The scale includes both cognitive and affective aspects of mental states, both of the child and the parent. Items refer to feelings (Items 3, 6, 20 and 26), as well as to thinking (Items 6 and 20), curiosity (Item 3), understanding (Items 28 and 31), wonder (Item 20), and attention (Item 26). The fourth polarity of mentalizing identifies a focus on internal mental states and external bodily cues or behaviour. The PRFQ-CF items focus on the internal mental states of thoughts and feelings; however, the questions do not address the means by which the respondent infers these states. A parent may respond positively to the items yet have a tendency to rely on external cues to make assessments of mental states of self and others rather than directly considering internal mental states, or vice versa. It may be that the PRFQ-CF is applicable for responses based on internal or external mentalizing, although the scale does not distinguish between these two polarities.

Finally, the focus of mentalizing can be directly on internal mental states (e.g., thoughts, feelings, desires, intentions), or the external and visible indications of mental states. These external cues include facial expressions, body posture and nonverbal behaviour. In addition to these polarities, mentalizing can be context and relationship specific

In addition to the PRFQ-CF and PRFQ-SF items reflecting the distinction in self-other focus of mentalizing, the items have other characteristics that are common to one or the other subscale. The child-focused items appear to mostly refer to the fundamental aspect of mentalizing which is a curiosity or interest in mental states. Two PRFQ-CF items stand out as being slightly different: Item 26 (I pay attention to what my child is feeling) is specifically about paying attention to the child's mental states, and Item 28 (Understanding why my child behaves in a certain way helps me not to be upset with him or her) refers to the interaction between the child's and the parent's mental states. Item 28 is similar in this way to Item 4 in the self-focused

subscale. Perhaps the reason for these two items fitting into the different scales lies in the ordering of the subject in the wording and where one places the emphasis. The subject order is consistent with the item placement the different scales, such that Item 28 begins with the child as the focus whereas Item 4 begins with the parent's feelings (i.e., self-focussed).

The PRFQ-SF items appear to have a common theme of self-focus in the context of parenting; however, they also represent a much broader mix of mentalizing characteristics in comparison to the PRFQ-CF items. With reference to the PDI-RF scoring manual (Slade, Bernbach, Grienenberger, Levy, & Locker, 2005), the PRFQ-SF items reflect the following PRF characteristics (with item numbers in parentheses): opaqueness or uncertainty of accurately knowing mental states (13, 15 and 22), the limitations of insight into mental states (11), an intergenerational perspective (15, 24 and 30), envisioning changes over time (10 and 17), diverse perspectives (1), and the interaction between parent and child mental states (4). In keeping with the self-focus, several Items (30, 15, 24, and 10) have no reference to the child's behaviour or mental states.

### ***PRFQ-CF analysis***

The seven items of the child-focused PRFQ-CF subscale showed evidence of conforming to the requirements of the Rasch model, with good internal consistency reliability and minimal DIF. All the results of the analyses of the PRFQ-CF will be discussed in this section.

### ***Targeting***

The generally high scores of the PRFQ-CF in this study are not surprising given the sample of non-clinical families with relatively higher incomes and higher levels of education compared to the population. The items were well targeted for the mid to low levels of PRF for this sample although they were located over only a very narrow range of PRF. This narrow range of item locations is somewhat compensated for by the distribution of item thresholds, which varied sufficiently so as to provide a fairly good distribution across all but the very highest scoring parents. Further examination of the PRFQ-CF with a sample of parents who have difficulties in their relationship with their children, or experience known stressors of the parent-child relationship (e.g. poor attachment, mental health difficulties, poor family

functioning), would be required to more reliably examine the threshold locations for the lower levels of PRF.

#### *Fit to the Rasch model*

The PRFQ-CF was found to show overall fit to the Rasch model indicating that hierarchical ordering of items was consistent for parents across all levels of PRF. The individual items mostly showed fit to the model, with the exception of some evidence of slight misfit for Items 6 and 20. On examination of these two items the misfit was considered to be minimal and removal of the items from subsequent analysis reduced the index of reliability. The location of items is important to note when considering possible redundancies of similarly located items.

Interestingly, Items 6 and 20 were very close in location to Items 7 and 28 respectively. These two pairs of items are worded very differently and are relatively different in their meanings, so it could be argued they contribute in different ways to the trait even though they have similar locations. Further examination of these pairs is discussed in the dimensionality section.

Individual person fit to the Rasch model revealed seven parents with extreme scores, which is not surprising given the generally high scoring sample. These parents strongly agreed to all seven PRFQ-CF items. Such strong agreement may be simply an accurate indication of high parental mentalizing. Alternatively, such consistent extreme scores may indicate the influence of a response biases such as faking good, a tendency to agree, or a tendency to use extreme ratings (Paulhus, 1991). These response biases were not assessed in this study. Regardless of the reason for these extreme responses, the analysis the PRFQ-CF was considered both with and without extreme scores, and no differences were observed.

#### *Reliability*

The PSI indication of internal consistency reliability of the PRFQ-CF was .82 (slightly less with extreme scores removed, PSI = .80), which is considered adequate for basic research purposes (Nunnally, Bernstein, & Berge, 1967). Fisher (2010) reported that a PSI of between .80 and .90 is sufficient for distinguishing between three groups enabling low, medium and high levels of the trait to be separated with 95% confidence. Therefore, the PRFQ-CF has potential for further research and investigation of PRF through comparisons of groups of parents; however, the level of

reliability indicates that caution is required if attempting to accurately distinguish between individual scores such as in clinical settings.

### *DIF*

The analysis of Differential Item Functioning (DIF) found no item bias with the PRFQ-CF for most of the variables considered. The result indicated that, for this sample, the items showed invariance across groups of parents categorised by parent gender, child gender, parent age, parent birthplace, and parent occupation.

The DIF of Item 31 (I try to understand the reasons why my child misbehaves) for child's birth-order was examined further by splitting the item into two items (one only answered by first time parents for the other answered by parents with other children) and reanalysing the scale; however, the scale properties were not improved. Also, the order of the remaining item locations was not influenced by the splitting of Item 31, because both item locations remained at the highest end of the scale. Although there appears to be a bias in the response to Item 31 based on child birth-order, this bias does not seem to influence the functioning of the scale. Even so, the question remains why first time parents would be less likely than other parents (of the same level of PRF) to agree to this item. One distinctive feature of this item is the use of the word "misbehaves", which could be considered an inappropriate descriptor for a 12-month-old's behaviour. Item 6 uses the terms "reasons" and "behaves", and Item 28 likewise refers to "understanding why my child behaves in a certain way", yet neither of these items revealed DIF for child birth-order. Perhaps a first time parent is less likely to agree to the use of the word "misbehaves" as a description of their 12-month-old's behaviour. Whereas a parent with older children might be more likely to respond in a manner that is influenced by their attitude to their older children.

DIF was also found for parent's level of education (Item 7); however, there was no clear distinction between any of the four levels of education that was consistent for all levels of PRF (i.e. class interval levels). Therefore, neither instance of DIF was considered sufficiently problematic to justify removal of items from the scale and the PRFQ-CF items appear to not bias the total score with regards to the seven variables considered with this sample. The DIF analysis has however highlighted potential problems with these items that could be examined further. Both these cases of DIF could be examined qualitatively by asking parents of different

groups for their interpretation of the question's wording and the reasons for their responses.

*Local item dependence and dimensionality*

The PRFQ-CF scale showed no evidence of local item dependence or dimensionality when the residual correlations were examined. As previously mentioned, some items had similar location estimates (Items 6 and 7, and Items 20 and 28); however, no items were found to have residuals that were correlated, indicating the items satisfied the assumption of independence. Therefore, all seven items were considered beneficial for the optimum functioning of the PRFQ-CF scale as a unidimensional rating of child focused PRF.

*Variance explained by categorical parental characteristics*

No significant differences in PRFQ-CF scores were found when parents were grouped by child gender, parent age, parent birthplace, parent occupation and child's birth-order. While child focused PRF did not appear to differ for child's gender in this study, a study of parent child sensitivity and attachment (Schoppe-Sullivan et al., 2006) found interaction effects with parent and child gender. Further analysis is required to test for such interaction effects of child gender with the PRFQ-CF scale. Interestingly older parents and parents with prior experience with other children did not have significantly higher child focused PRF, indicating this aspect of PRF may not necessarily increase with age or experience. The 27% of parents born outside of Australia were found to have generally similar scores to Australian born parents. This result does not provide any evidence of the PRFQ-CF's validity for other cultures, but rather the similarity of scores and the finding of no DIF for birthplace support the validity of the scale with an English speaking Australian population (not including Aboriginal people or Torres Island Islanders).

The two variables that did show significant differences in PRFQ-CF scores were parent gender and education. Mothers were found to score higher than fathers in ratings of child-focused PRF and parents with higher levels of education showed a tendency to have higher levels of PRF. These two findings are now discussed.

Mother and father PRF comparisons have not been previously reported in published studies. Research with the Adult Attachment Interview has shown mother and father adult attachment classifications to have very similar distributions (van IJzendoorn & Bakermans-Kranenburg, 1996), indicating fathers do not tend to have

a different pattern of attachment security compared to mothers. Likewise the proportions of mother-infant and father-infant secure attachments based on the Strange Situation Procedure have been found to be very similar (van IJzendoorn & Bakermans-Kranenburg, 1996; van IJzendoorn & De Wolff, 1997). Only one comparable indication of a gender difference in parental mentalizing has been found in a study of parental mind-mindedness with a very small sample of mothers and fathers (Arnott & Meins, 2007). Arnott and Meins (2007) found a tendency for fathers to produce a higher proportion of inappropriate mind-minded comments compared to mothers. A possible explanatory factor in the current finding of gender differences is the gender associated parenting role and average amount of time spent caring for the infant. Further research is required with a sample of parents with sufficiently varying parenting roles across genders in order to examine the contribution of average time of infant care to the variance in PRF. Such studies will enable more accurate gender differences to be identified and examined.

The analysis of correlations between mother and father PRFQ-CF scores showed no relationship and this confirms previous studies indicating the mother-infant and father-infant relationship is independent and relationship specific (Arnott & Meins, 2007). In this respect, the current finding of no parent gender association is contrary to hypotheses that would suggest that men and women self-select partners with similar characteristics such as PRF (Vandenberg, 1972) or that the characteristics of the child may influence the parent's PRF.

The gender difference found for levels of child focused PRF could be understood as confirmation of similar differences found in theory of mind and alexithymia research. Men have generally been found to have lower awareness of emotions and less able to describe emotions compared to women. This difference is known as the normative male alexithymia hypothesis (Levant, 1992; Levant, Hall, Williams, & Hasan, 2009). Also, Baron-Cohen's (2002) proposed ToM dimensions of "empathizing" and "systemizing" imply that females generally have greater capacity to think in terms of mental states, whereas males tend to have greater abilities with systems type thinking. These findings are consistent with the current finding of higher mother PRFQ-CF scores.

Comparisons of PRFQ-CF means showed parents with higher education tended to score higher in child-focused PRF. This trend is comparable with

associations between RF (scored from the AAI) and education found in the first studies of RF (Fonagy, Target, Steele, & Steele, 1998). For example, Fonagy et al. (1998) also found similar correlations between RF and verbal IQ, and no correlations with social class or socio-economic groups. Likewise, the current study did not find any differences in PRF based on occupation. These results suggest that the association found between PRF and education may be a reflection of comprehension rather than socio-economic status. Since the PRFQ-CF relies on written items, the ability of the parent to comprehend the items is likely to bias or add measurement error to the responses. Ideally, observational measures would be used to reliably assess the parent-infant relationship with a greater degree of independence from parent IQ, such as the proposed concept of Parental Embodied Mentalizing (Shai & Belsky, 2011). Interestingly, in a study of maternal mind-mindedness (Meins, Fernyhough, Fradley, & Tuckey, 2001), education was found to be related to maternal responsiveness to change in infant direction of gaze, but not associated with a range of other infant-mother interaction variables. Therefore, parent IQ or a proxy measure, such as education, should routinely be examined and accounted for in the study of PRF as a predictive variable.

### ***PRFQ-SF analysis***

The self-focused PRFQ-SF subscale was found to have low internal consistency reliability (PSI = .66) and three items showed DIF for parent gender (Items 1, 4, and 13). Low reliability for self-focused parental mentalizing compared to child-focused mentalizing was also found in the study of PDI-RF scores by Suchman et al. (2010). Other aspects of the analyses of the PRFQ-SF subscale satisfied the requirements of the Rasch measurement model and indicate there is potential for a scale that assesses the self-focused dimension of PRF.

The analyses of item fit, local item dependence or scale dimensionality did not provide any indications of ways to modify the scale so that the problem of low reliability might be improved; therefore, these results from these analyses are not discussed further in this thesis. As previously discussed, the PRFQ-SF items represent a broad range of mentalizing characteristics, and it may be that a multitude of dimensions in the subscale contribute to a difficulty with finding clear results. Perhaps analysis of a larger selection of similar items might reveal more defined dimensions of self-focused PRF, which might provide subscales that show higher

internal reliability. Further investigation and development of this subscale was deemed beyond the scope of this thesis, primarily because it would involve recruitment of an additional sample of parents for the collection of additional qualitative and quantitative data. The main focus of analysis and discussion in this thesis was limited to the child-focused PRFQ-CF scale, which is arguably more directly indicative of the context specific parent-child relationship.

### **LH subscale analysis**

The examination of category response frequencies revealed that the item response distribution was severely skewed for most of the LH items. Half of the items were endorsed *strongly disagree* by 50% or more of the parents, and parents rarely (less than 9%) responded to any of the three *agree* categories for these items. Although the non-clinical sample of parents in this study would be expected to generally disagree with these statements (i.e., have generally higher PRF), the degree of extreme responses or poor targeting of the LH scale is likely to have been problematic for the analysis. Following rescoring of items in order to resolve disordered thresholds (mostly at the *agree* end of the rating scale), a majority of the items showed misfit to the Rasch model and the scale overall did not show fit to the model or satisfactory reliability (PSI = .66). Attempts to remove items did not resolve these problems and no analysis provided clear indications on ways to improve the scale (i.e., no local item dependence or dimensionality).

The problems found with the LH subscale could be in part due to the implicit and automatic nature of the aspects of mentalizing that are targeted by these items. The pre-mentalizing modes (deficits in mentalizing) that the LH items represent are more likely to be observed as an automatic response in the context of stress and when the attachment system is triggered in intimate relationships (Fonagy & Luyten, 2009). Therefore parents may honestly disagree with these statements in the safe environment of completing an anonymous questionnaire in their home and yet behave, think and feel quite differently in a stressful exchange with their child. Also, the proposed modes of pre-mentalizing may represent separate dimensions that, if analysed as targeted scales, may show more promise of validity and reliability.

### **M subscale analysis**

The M subscale was scored in a novel way with the centre category representing the highest score for PRF and the extremes of *strongly agree* and

*strongly disagree* representing lowest PRF. This scoring pattern was generally supported by the relatively even frequency distribution of responses to categories for all but two items (Item 16 and 35) and almost completely ordered thresholds for all but one item (slight disorder for Item 2). Therefore, parents appear to have responded to the M questions in the expected manner, such that responses towards either extreme of the rating scale were more likely for low scoring parents and central responses more likely for high scoring parents.

Further support for the M subscale was found with overall fit to the Rasch model and all items individually showing satisfactory fit to the model; however, the PSI index of reliability was very low at .63. The M subscale showed no signs of local item dependence or dimensionality, and therefore no obvious ways were available to improve the low reliability. In brief, the M subscale may suffer the same challenge as the LH subscale in that pseudo-mentalizing is understood to be mostly implicit and may not be accurately reflected in a self-report questionnaire. No further analysis was performed with the M or LH subscales since additional development of these subscales would involve qualitative analysis, item development and another administration to a new sample of parents. Moreover, there is no directly relevant literature with which these preliminary results can be compared.

Further qualitative investigation could provide insight into the unexpected response patterns for Items 16 and 35, which had much higher frequency of responses at the “Strongly Agree” end of the rating scale compared to other items. The evidence from the item locations indicates the frequency pattern for these items was a result of them being the most ‘difficult’ to elicit a middle response. It is possible that these two items are worded such that only the parents with the highest levels of reflective functioning tend to rate with a less extreme response.

### **PRFQ-CF Associations with Parental Depression or Anxiety**

The correlations between mothers’ and fathers’ PRFQ-CF scores and their depression and anxiety scores were all similarly weak and mostly in a negative direction. This result is confirmation of the discriminative validity of RF reported in the Fonagy et al. (1996) study of parents’ RF scored from the AAI, which was not significantly associated with a self-report of mental health; however, mental health difficulties are understood to commonly feature deficits in mentalizing, including conceptualizations of anxiety (Nolte, Guiney, Fonagy, Mayes, & Luyten, 2011) and

depression (Luyten, Fonagy, Lemma, & Target, 2012). Therefore, a weak association between low PRF and reports of mental health difficulties is not unexpected.

Although mental health difficulties would be expected to limit a parent's capacity for mentalizing, this limitation may only be substantial in times of distress or for those parents with high levels of dysfunction. Few parents in this sample reported high levels of depression or anxiety and the questionnaires were unlikely to have been completed at times when the parents were experiencing high levels of distress. The state subscale of the STAI is specifically designed to assess current anxiety and these scores were no more related to the PRFQ-CF than the trait subscale scores.

Mental health difficulties are likely to be more closely related to particular aspects of mentalizing, such as implicit or automatic mentalizing and with distorted modes of mentalizing (Luyten, Fonagy, Lemma, et al., 2012). Therefore, since the PRFQ-CF is expected to only measure explicit mentalizing, it may be that this particular aspect of the measure limits its association with mental health. The child-focused characteristic of the PRFQ-CF may also be a factor influencing the association with mental health. Further development of valid measures for the different polarities of mentalizing (Luyten, Fonagy, Lowyck, & Vermonte, 2012) will enable examination of the differential relationships between mentalizing and important parent and child characteristics, such as mental health.

### **PRFQ-CF Test-Retest Stability**

The reliability coefficients from mother and father test-retest correlations indicated poor test-retest stability. With one outlier removed, the fathers' test-retest reliability improved but was still lower than the mothers', and was below the expected minimum of .70. The invariance plot showed matching parent test and retest scores showed a more optimistic result with parent scores generally within or close to the 95% confidence lines, and a t-test indicated that mothers and fathers had an equal frequency of retest scores that were significantly different from the first test. Therefore, it can be concluded that the test-retest analysis of the PRFQ-CF with this sample showed indications of test-retest stability although they were lower than expected. Furthermore, mothers and fathers were equally likely to have significantly differing scores between the two tests. Possible reasons for the low test-retest stability are now discussed.

Firstly, the accuracy of the test-retest coefficient as a measure of stability is attenuated by the proportion of error of measurement in the scale (Fisher, 2010), in this case with the PRFQ-CF reflected in the PSI of .82. Due to this effect of attenuation, self-report measures tend to have lower test-retest reliability coefficients than internal reliability coefficients; see, for example, reports of test-retest stability for measures of mindfulness (Baer, Smith, Hopkins, Krietemeyer, & Toney, 2006) and empathy (Davis, 1980). The test-retest result for the PRFQ-CF would likely be slightly higher if the internal reliability of the scale were higher.

A parent's general capacity for parental reflective functioning with a specific child is expected to be relatively stable over a short time, much like the traits of mindfulness and empathy, and is likely to have similarities with the moderate stability of attachment representations over early childhood assessed with the PDI (Aber, Belsky, Slade, & Crnic, 1999). Studies of infant-parent attachment with the Strange Situation Procedure (SSP) have shown generally low test-retest stability. A meta-analysis by Fraley (2002) summarised results from 15 studies ( $N = 896$ ) that applied the SSP at 12 months and repeated 6-8 months later. The weighted test-retest coefficient from these studies was  $r = .32$  ( $SD = .28$ ), indicating a high portion of variability in the way an infant behaves towards his or her parent over time. This variability may have some degree of influence on a parent's reflective functioning and contribute to test-retest variability. Factors found to influence infant-parent attachment may also contribute to PRF instability and therefore reduce test-retest reliability. A review of attachment stability across the lifespan (McConnell & Moss, 2011) identified a number of factors that reduced stability in early childhood attachment, including changes in maternal employment, negative life events, parental mental ill-health, and poor parental sensitivity

The PRFQ-CF scores were found to generally increase over the test-retest period. This increase could possibly be inflated by systematic differences between the contexts of the two tests. The first administration of the PRFQ was via a posted package to the parents and consisted of a large number of questionnaires, which they completed in their own time, compared to the retest, which was completed as a single questionnaire. The reduced burden of completing a single questionnaire may contribute to higher scores from the retest. Furthermore, the retest was completed during the visit by the interviewer when the PDI was conducted. The presence of the interviewer and the process of the interview may have in some way influenced the

parents' responses to the questionnaire, perhaps the interviewers presence prompted more reflective responses or maybe prompted a social desirability bias (Paulhus, 1991). In addition to the change of context, it may be that these parents had developed slightly greater PRF over the period between tests, which spans a time of rapid child development, including increased communication. Studies of PDI-RF scores with mothers have found levels of PRF generally increase over time and it has been suggested that this is due to increased familiarity with the infants (Sadler et al., 2013).

Mentalizing capacity is likely to vary with the parent's level of arousal and this would contribute to variation in test-retest stability. High levels of arousal or distress are known to impair mentalizing capacity, and at these times pseudomentalizing or pre-mentalizing modes of pretend, psychic equivalence and teleological thinking (Bateman & Fonagy, 2008) are more likely to be activated. The influence of arousal or distress on the measurement of PRF is deserving of further research. In this study there was no measure of arousal or distress taken at the times of the test and retest, so this likely contributor to variability in the scores could not be accounted for.

### **Convergent Validity with the PDI-RF**

The scores from the PRFQ-CF were not associated significantly with PDI-RF scores even though both assessments have been designed to measure parental mentalizing. Nevertheless, the weak non-significant correlation between the PRFQ-CF and PDI-RF scores was in the expected positive direction, and was of similar strength for both mothers ( $r = .14$ ) and fathers ( $r = .13$ ). Assuming the PDI-RF assesses the construct of parental mentalizing in the broadest sense, the specific child-focus of the PRFQ-CF clearly limits the potential for the two measures to have a strong correlation. The PDI-RF is administered with an emphasis on the parent relationship with a specific child; however, many of the questions are solely about the experience of being a parent and some are about the participant's own parents. Further research could examine the correlation between the PRFQ-CF and sub-scores of the PDI-RF differentiating child-focused and self-focused questions, such as the analysis conducted by Suchman et al. (2010).

A further explanation for the lack of concordance between the PDI-RF and the PRFQ-CF is the difference in the method or type of measurement instrument,

with one being a self-report questionnaire and the other a score from analysis of an interview transcript. In an analysis of method-specific differences between two instruments (with measures of family functioning, Sigafos, Reiss, Rich, & Douglas, 1985), three factors have been suggested to effect participant responses: “(a) the stimulus provided by the instrument itself; (b) the ‘language’ the procedure requires the subject to use in responding; and (c) the interpersonal context of the research setting in which these responses are made” (p. 199). Each of these factors will be briefly described in terms of how they operate differently between the PRFQ-CF and the PDI-RF.

The act of undertaking the interview and being asked direct questions about intimate thoughts and feelings is very different to completing a questionnaire in one’s own time and choice of space. In the language of the PDI-RF scoring manual (Slade et al., 2005), some questions are referred to as “demand” questions, in reference to the way these questions explicitly ask the participant to reflect on mental states. The questions are demanding in another sense in that they require the participant to respond in the moment, and in their own words. In contrast, the PRFQ-CF questions are presented on paper and participants are required to respond with a level of agreement or disagreement. There is no limit on time taken to respond or any way of knowing if the participant thought deeply about the question or responded impulsively. Since the questionnaire is confidential and completed in private, it is possible that the participant is psychosocially removed from the experience of a particular person wanting to know the response to each item. There is also no opportunity for unique expression or responses outside the dimension of agreement. The stimulus of the instrument in this case certainly differs between the two measures and this distinction substantiates the proposition by Luyten et al. (Luyten, Fonagy, Lowyck, et al., 2012) that the PDI-RF is at least a partial assessment of the automatic or implicit dimension of mentalizing, whereas the PRFQ only assesses controlled or explicit mentalizing.

A further distinction between the two measures of PRF is the “language” of the instrument. The PRFQ is purely a written questionnaire that requires skills such as reading comprehension and vocabulary. There is no way of knowing how much of the measurement error in scores is due to participants misunderstanding of the questions. The PDI-RF is a face-to-face interview with no written component; as such the participant requires verbal comprehension and expression with an active

listener, with the benefit of the interviewer's availability to rephrase or further explain any questions and seek clarification if responses are not clear. Further examination of the relationship of verbal and written comprehension with the two measures of PRF would help determine the extent to which these skills explain the variance between the scores.

The interpersonal differences between the PRF measures are mostly accounted for in the explanation of the language and stimulus factors. An additional consideration is the potential for the interviewer's presence, engagement and questioning to increase the arousal of the participant. An increase in arousal in the context of discussing the parent-infant relationship has the possibility of reducing explicit mentalizing and increasing pre-mentalizing modes of thinking. The impersonal nature of the questionnaire method is less likely to be arousing and have this effect.

The current comparison of PRFQ-CF and PDI-RF scores has similarities to the comparison of self-report questionnaire measures of adult attachment and the attachment representations rated from the AAI. Therefore, it is interesting to consider the explanations for a similarly weak correlation ( $r = .09$ ) generally found between self-report and AAI measures of attachment (Roisman et al., 2007). Unlike the measures of PRF, the two methods of measuring attachment evolved from two relatively distinct lines of research and distinct groups of researchers, even though both had a common focus of attachment theory (Shaver & Mikulincer, 2002). Some authors (van IJzendoorn & Bakermans-Kranenburg, 2010) have argued that some of the self-report measures such as the Experiences in Close Relationships scale are not measures of attachment, but assess a distinctly different construct. The PRFQ-CF and PDI-RF have the same substantive and theoretical bases, so the differences can only be a result of the characteristics and limitations of each instrument.

### **Limitations**

The sample used for this analysis had a number of strengths and weaknesses. An important strength of this sample is the inclusion of the mother and the father with comprehensive data for both, allowing for comparisons and examination of data generalizable to parents, not only mothers. A problem with using data from couples is the possibility of intra-dyad dependence, such that the scores from each couple are not independent. The lack of any association between mother and father score is

evidence suggesting intra-dyad dependency was not problematic in this data set in analyses where mother and father data were analysed together.

The sample was in most respects representative of the population, although a large percentage of the initial study's cohort discontinued their involvement by the one-year follow-up. This dropout resulted in a higher representation of mothers with higher levels of education and families with higher income. Within these limitations, the sample generally represents non-clinical English speaking couples. The findings are not generalizable beyond this sample to groups such as clinical populations, parents with older or younger aged children, single or separated parents, and other language groups or cultures.

The Rasch analysis using RUMM2030 software provided a comprehensive examination of the data available from the Peel Child Health Study. The recommended PRFQ-CF subscale for child-focused reflective functioning was identified using post hoc analysis procedures, which can only indicate the potential of these items for valid measurement. These procedures include the rescoring or collapsing of rating categories to resolve disordered thresholds and the selection of specific items to group together as a scale based on item fit and residual correlations. These methods alter the items such that if they were presented in the form of the suggested PRFQ-CF, the change of rating categories and the particular grouping of the items in the child-focused reflective functioning subscale may elicit a different response from parents. Prior to further research with this scale, the items should be administered again in this recommended format of the PRFQ-CF with a community sample of parents to confirm the evidence of validity reported in this study.

The objective of this study was to critically examine the validity of the PRFQ with the application of Rasch measurement theory. Not all aspects of scale validity could be tested with the methods used in this analysis. For example, predictive validity of the PRFQ would require a longitudinal study design in conjunction with causal modelling techniques in order to test causal hypotheses such as, higher levels of mothers' and fathers' PRF uniquely contribute to their children's higher performance in theory of mind tasks or the strange situation procedure. The testing of convergent validity could have been tested more broadly (it was only tested against the PDI-RF in this study), including observational measures of parent- infant mind-mindedness (Meins et al., 2001) and self-report of closely related constructs such as theory of mind and empathy. Observational measures are of particular interest

because they have the potential to assess on-line mentalizing, whereas the PRFQ and the PDI-RF are limited to assessing off-line mentalizing.

This examination of the PRFQ is limited to being primarily a description of data fit to the Rasch measurement model. Although this method is very informative, the construct validity of the PRFQ is dependent to a large degree on the theoretical foundations of the PRF construct and the development of the original pool of PRFQ items. The results of this analysis have highlighted both strengths and weaknesses with the PRFQ, nevertheless the analysis is not sufficient to test the validity of the underlying construct. There is the possibility that the some subscales of the PRFQ would show better fit to the Rasch model with alternative approaches to the conceptualization of PRF and the development or construction of additional items.

## Chapter 6 Conclusions

The objective of this thesis was to apply Rasch measurement theory in a critical examination of the validity of the *Parental Reflective Functioning Questionnaire* (PRFQ) as a measure of *parental reflective functioning* (PRF) for both mothers and fathers. In fulfilling this objective, three unique aspects of the current research project are noteworthy. Firstly, PRF has yet to be measured with a validated self-report scale and no published studies have compared mother and father PRF rated from the *Parent Development Interview* (PDI-RF). Therefore, this examination of the PRFQ makes an important contribution to the study of parental mentalizing. This study also adds to the understanding of father-infant relationships, which is limited in comparison to research with mothers. Finally, this project uses Rasch analysis to determine the potential of the PRFQ for measurement, which provides advantages over the more commonly used classical test theory methods.

The analyses in this thesis provided a means to answer the following four research questions:

1. Do data from mothers and fathers PRFQ conform to the requirements of the Rasch measurement model?
2. Do ordinal variables of parental depression or anxiety predict variance in PRFQ scores?
3. Do one or more specific sets of PRFQ items show temporal (test-retest) stability?
4. Is there a relation between PRFQ scores and PDI-RF scores that demonstrates convergent validity?

Numerous sources of evidence were obtained from this Rasch analysis of the PRFQ, which inform the ongoing development of parental mentalizing measurement. Earlier sections of the Discussion chapter summarised these findings and their implications. A synthesis of these findings is now provided to answer each of the above research questions.

The results from separate analyses of the three PRFQ subscales indicate that the scale requires substantial further development. Only the child-focused dimension of the HL subscale has shown promise as a measure of parental mentalizing. A primary shortcoming of the self-focused dimension of the LH subscale, the LH and the M subscale was low internal consistency reliability, as indicated by the person

separation index. In addition to low reliability, the LH subscale data showed poor targeting of items to persons and misfit the Rasch model. The analysis did not show obvious signs of the cause of these problems or show means to improve reliability or model fit. Therefore, these subscales are not recommended for use in their current form. Strategies for further development of these three subscales would include qualitative analysis with careful consideration of the most recent developments in mentalizing theory (Fonagy et al., 2012).

Qualitative analysis of the PRFQ items could provide alternative ways to group existing items into more targeted subscales as well as indicate additional items that could be developed, which might improve construct validity. For example, separate scales could be developed that target the pre-mentalizing modes of teleological thinking, pretend and psychic equivalence. Further consideration could also be given to how individual items or subscales assess the mentalizing polarities of self/other, implicit/explicit, cognitive/affective and internal/external.

The polarities and modes of mentalizing would ideally be measured separately to be able to examine the interrelationships between them and to test the theories of how different profiles are associated with particular factors or psychopathology (e.g., borderline personality disorder; Fonagy & Luyten, 2009). If these dimensions were found to show a reasonable degree of independence, then separate subscales for each dimension would be the ideal goal for a holistic assessment of a parent's mentalizing profile. Furthermore, evidence may suggest some specific dimensions of mentalizing lack validity when assessed with self-report measures, as could possibly be the case with some aspects of pre-mentalizing modes that are mostly implicit or automatic in nature. In such cases, and when deemed necessary, specific observational or physiological measures could be developed for those dimensions. For example, the self-report of explicit, child focused mentalizing could prove to be useful as a screening instrument, such that a low score indicates a risk of mentalizing deficits and suggests the need for further investigation of the parent's use of pre-mentalizing modes of thinking. It may be necessary to assess implicit forms of pre-mentalizing under stressful conditions that prompt the failure of automatic mentalizing and during which behavioural and perhaps neurological observations could be informative.

The 7-item child-focused PRFQ-CF scale generally met the Rasch measurement theory requirements for measurement. The scale's level of reliability

and less than optimal targeting of items to persons indicates that the PRFQ-CF is recommended only for screening purposes and use with research groups rather than individual use in a clinical setting. Although the potential of the scale is promising within these limits, replication of the results is required using the recommended four-category rating scale with the Rating Scale Model. The original seven-category rating scale appears to be problematic, most likely for reasons previously discussed such as confusion regarding a middle response option, issues with the labelling of the categories, and possible difficulty distinguishing between categories if there are too many.

The analysis of the PRFQ-CF data did not show problematic item bias (DIF) or differences in mean scores when parents were grouped by child gender, parent age, parent birthplace, parent occupation or child's birth-order. The parents did, however, differ in PRFQ-CF scores for groupings by parent gender and education, with mothers and fathers with higher education levels scoring on average higher on the PRFQ-CF. The difference in mentalizing with regards to education is particularly interesting given that the sample was more highly educated than the general population, which could indicate the difference is underestimated. Future analysis with the PRFQ would be best undertaken with education as a covariate.

The finding that mothers on average scored higher than fathers for child-focused mentalizing raises a number of questions. The normative male alexithymia hypothesis (Levant et al., 2009) and Baron-Cohen's (2002) ToM theory of gender differences both indicate that this gender difference would be expected; however, in this specific context of the parent-child relationship, it could be that the gender differences be confounded by the time each parent spends caring for the infant. Also, gender biased attitudes regarding mother and father roles could be a source of some of the variability of scores. It is possible this gender difference could be reversed if the PRFQ-CF was administered to a sample of parents with stay-at-home fathers and full-time working mothers. The PRFQ-CF showed no DIF for parent gender, which indicates no items were responded to in a gender-biased manner. Therefore, in this respect, the PRFQ-CF scale appears to be well suited to answering these questions in future studies with the appropriate measures and analyses.

The mother and father PRFQ-CF scores were found to be uncorrelated. This independence of scores could be understood to show that the scores do not appear to be confounded by characteristics common to the couple or strongly influenced by

characteristics of the child, such as temperament. This is in contrast to the commonly found association (albeit weak) between mothers and fathers with infant attachment. The PRFQ-CF scores were also not associated with measures of anxiety or depression. Further studies of these associations are needed to confirm this independence and the conditions under which it is maintained.

A serious potential weakness of the proposed PRFQ-CF is the evidence of very low test-retest stability. Possible reasons for this result have been discussed in an earlier section of the thesis. Empirical investigation of test-retest stability of the PRFQ-CF is required with more rigorous control of the timing and conditions of the testing. In particular, further studies may consider a shorter and consistent timespan between testing and statistical control of factors that may influence the variance of scores.

The examination of convergent validity of the PRFQ-CF in this study was inconclusive. The correlations between the PRFQ-CF and the PDI-RF were not significant. This result is understandable given the stark differences in the instrument methods of assessment and the global RF score of the PDI-RF (including pre-mentalizing modes) compared to the PRFQ-CF's very specific assessment of child-focused mentalizing. If the three subscales of the PRFQ had proven to be valid for measurement, the three PRFQ-CF scores would have together provided an indication of convergent validity by their degree of prediction of the PDI-RF score in a regression analysis. The PRFQ offers the potential to examine a wider range of hypotheses with more powerful research methods than is possible with the PDI-RF. In particular, the PRFQ presents a relatively easy to administer self-report questionnaire for the collection of normative data, for the comparison of groups of interest, and for use in modelling that could identify causal relationships.

In addition to further validation of the PRFQ-CF scale, the other potential subscales of the PRFQ are in need of development and validation. Each facet of parental mentalizing that could possibly be measured by self-report would ideally be examined for associations with other established measures of closely related constructs so as to identify the specific traits measured by each subscale. For example, in addition to comparisons with the PDI-RF, the following measures of parent-infant relationships would offer informative comparisons: the Strange Situation Procedure (SSP; Ainsworth, Blehar, Waters, & Wall, 1978), mind mindedness (Meins et al., 2001), RF rated from the Adult Development Interview

(Fonagy et al., 1998), sensitive and challenging interactive play (Grossmann, Grossmann, & Kindler, 2005), and the Risky Situation (Paquette & Bigras, 2010). There are a number of considerations to bear in mind when making comparisons between these measures. Firstly, the specificity of the construct is likely to influence associations, as this study has shown in the contrast between the specific PRFQ-CF and the more broadly defined PDI-RF. The context of the assessment can differ greatly, such as the very specific observations of separation and reunion with a child in the SSP, compared to the more generalised scenarios presented in the PRFQ. A parent measure needs to be empirically validated with both mothers and fathers, as has been attempted in this study with the PRFQ-CF. The measures can be assessing on-line immediate responses and reactions (e.g., mind-mindedness) compared to off-line attitudes and opinions like those in the PRFQ.

Further development and analysis of other dimensions of PRF apart from the child-focused mentalizing is an obvious next step from this current project. Specific problems identified with other subscales of the PRFQ include: the low reliability of the self-focused HL subscale and three items that showed bias for parent gender; the pre-mentalizing LH subscale showed poor targeting, misfit to the Rasch model and low reliability; and the pseudo-mentalizing M subscale showed very low reliability. This analysis of the PRFQ has demonstrated the utility of the Rasch measurement model in identifying the ways in which scale meets critical requirements for measurement.

The measurement of the parent-child relationship and the quality of father-child relationship in particular is a challenging task. Using a rigorous measurement model, this analysis has shown potential for a short self-report measure of child-focused mentalizing with both mothers and fathers. Although other subscales of the PRFQ were found to have shortcomings, the analysis provides an informative starting point for further development of these scales. Measurement instruments such as the PRFQ could play a key role in furthering our understanding of the parent-infant relationship beyond what has already been achieved with resource intensive instruments and small samples.

### References

- Aarons, G., McDonald, E., Connelly, C., & Newton, R. (2007). Assessment of family functioning in Caucasian and Hispanic Americans: Reliability, validity, and factor structure of the Family Assessment Device. *Family Process, 46*(4), 557-569.
- Aber, J. L., Belsky, J., Slade, A., & Crnic, K. (1999). Stability and change in mothers' representations of their relationship with their toddlers. *Developmental Psychology, 35*(4), 1038.
- ABS. (2013). ANZSCO - Australian and New Zealand Standard Classification of Occupations, Version 1.2 from <http://www.abs.gov.au/ausstats/abs@.nsf/mf/1220.0>
- Adams, R. J., Wu, M. L., & Wilson, M. (2012). The Rasch rating model and the disordered threshold controversy. *Educational and Psychological Measurement, 72*(4), 547-573. doi: 10.1177/0013164411432166
- Ainsworth, M. D. S. (1967). *Infancy in Uganda: Infant care and the growth of love*. Baltimore: John Hopkins Press.
- Ainsworth, M. D. S., Blehar, M. C., Waters, E., & Wall, S. (1978). *Patterns of attachment: A psychological study of the Strange Situation*. Hillsdale, NJ: Erlbaum.
- Akister, J., & Stevenson-Hinde, J. (1991). Identifying families at risk: exploring the potential of the McMaster Family Assessment Device. *Journal of Family Therapy, 13*(4), 411-421. doi: 10.1046/j..1991.00437.x
- Allen, J. G. (2006). Mentalizing in practice. In J. Allen, G & P. Fonagy (Eds.), *Handbook of Mentalization-Based Treatment* (pp. 4-30). West Sussex: John Wiley & Sons.
- Allen, J. G. (2012). *Restoring mentalizing in attachment relationships: Treating trauma with plain old therapy*. American Psychiatric Publishing.
- Allen, J. G., Bleiberg, E., & Haslam-Hopwood, T. (2003). *Mentalizing as a compass for treatment*. The Menninger Clinic. Retrieved from <http://www.menningerclinic.com/resources/Mentalizing-compass.htm>
- Allen, J. G., & Fonagy, P. (Eds.). (2006). *Handbook of mentalization-based treatment*. West Sussex: John Wiley & Sons.

- Allen, J. G., Fonagy, P., & Bateman, A., W. (2008). *Mentalizing in clinical practice*. London: American Psychiatric Publishing.
- Allen, S., & Daly, K. (2007). The Effects of Father Involvement: An Updated Research Summary of the Evidence Inventory. .  
[http://fira.ca/cms/documents/29/Effects\\_of\\_Father\\_Involvement.pdf](http://fira.ca/cms/documents/29/Effects_of_Father_Involvement.pdf)
- Allison, C., Baron-Cohen, S., Wheelwright, S. J., Stone, M. H., & Muncer, S. J. (2011). Psychometric analysis of the Empathy Quotient (EQ). *Personality and Individual Differences, 51*(7), 829-835. doi: 10.1016/j.paid.2011.07.005
- American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders (DSM-IV-TR)* (4th ed.). Washington, DC: Author.
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561-574.
- Andrich, D. (1982). An index of person separation in latent trait theory, the traditional KR-20 index, and the Guttman scale response pattern. *Education Research and Perspectives, 9*(1), 95-104.
- Andrich, D. (1985). An elaboration of Guttman scaling with Rasch models for measurement. In N. Brandon-Tuma (Ed.), *Sociological Methodology* (pp. 33-80). San Francisco: Jossey-Bass.
- Andrich, D. (1988a). Rasch models for measurement. In J. L. Sullivan & R. G. Niemi (Eds.), *Sage University Paper series on Quantitative Applications in the Social Sciences, series no. 07-068*. Beverley Hills: Sage.
- Andrich, D. (1988b). The application of an unfolding model of the PIRT type to the measurement of attitude. *Applied Psychological Measurement, 12*(1), 33-51.
- Andrich, D. (1989). Distinction between assumptions and requirements in measurement in the social sciences. In J. A. Keats, R. Taft, R. A. Heath & S. H. Lovibond (Eds.), *Proceedings of the XXIVth International Congress of Psychology, 4, 7-16. Mathematical and Theoretical Systems*. B.V. North Holland: Elsevier Science Publications.
- Andrich, D. (1990). The ability of an item. *Rasch Measurement Transactions, 4*(2), 101.
- Andrich, D. (1998). Understanding ordered category response data from the perspective of the Rasch model. *Education Research and Perspectives, 25*(2), 25-35.

- Andrich, D. (2002a). Implications and applications of modern test theory in the context of outcomes based education. *Studies In Educational Evaluation*, 28(2), 103-121. doi: 10.1016/s0191-491x(02)00015-9
- Andrich, D. (2002b). Understanding resistance to the data-model relationship in Rasch's paradigm: A reflection for the next generation. *Journal of Applied Measurement*, 3(3), 325-359.
- Andrich, D. (2004). Controversy and the Rasch model: a characteristic of incompatible paradigms? *Med Care*, 42(1 Suppl), I7-16. doi: 10.1097/01.mlr.0000103528.48582.7c
- Andrich, D. (2005). Rasch, Georg. In K. Kempf-Leonard (Ed.), *Encyclopedia of Social Measurement* (Vol. 3, pp. 299-306). New York: Elsevier Academic Press.
- Andrich, D. (2009). Interpreting RUMM2030 Part IV. Multidimensionality and subtests in RUMM. Perth, Australia: RUMM Laboratory.
- Andrich, D. (2010a). Educational measurement: Rasch models. In E. Baker, P. Peterson & B. McGaw (Eds.), *International Encyclopedia of Education (3rd Edition)* (pp. 111-122). Amsterdam: Elsevier.
- Andrich, D. (2010b). Understanding the response structure and process in the polytomous Rasch model. In M. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models: Developments and applications* (pp. 123-145). New York: Taylor and Francis.
- Andrich, D. (2011). Rating scales and Rasch measurement. *Expert review of pharmacoeconomics & outcomes research*, 11(5), 571-585. doi: 10.1586/erp.11.59
- Andrich, D. (2012). An expanded derivation of the threshold structure of the polytomous Rasch model that dispels any 'threshold disorder controversy'. *Educational and Psychological Measurement*, 73(1), 78-124. doi: 10.1177/0013164412450877
- Andrich, D. (2013). The legacies of R. A. Fisher and K. Pearson in the application of the polytomous Rasch model for assessing the empirical ordering of categories. *Educational and Psychological Measurement*, 73(4), 553-580. doi: 10.1177/0013164413477107
- Andrich, D., de Jong, J. H. A. L., & Sheridan, B. E. (1997). Diagnostic opportunities with the Rasch model for ordered response categories. In J. Rost & R.

- Langeheine (Eds.), *Applications of Latent Trait and Latent Class Models in the Social Sciences* (pp. 59-72). New York: Waxmann Verlag GMBH.
- Andrich, D., & Hagquist, C. (2012). Real and artificial differential item functioning. *Journal of Educational and Behavioural Statistics*, 37(3), 387-416.
- Andrich, D., Humphry, S. M., & Marais, I. (2012). Quantifying local, response dependence between two polytomous items using the Rasch model. *Applied Psychological Measurement*, 34(4), 309-324
- Andrich, D., & Kreiner, S. (2010). Quantifying response dependence between two dichotomous items using the Rasch model. *Applied Psychological Measurement*, 34(3), 181-192.
- Andrich, D., Sheridan, B., & Luo, G. (2010). RUMM2030 (Computer software and manual). *RUMM Laboratory, Perth, Australia*.
- Andrich, D., Sheridan, B. E., & Luo, G. (2009). *RUMM2030 manual: Part IV: Multidimensionality and subtests in RUMM*. Perth, Western Australia: RUMM Laboratory.
- Andrich, D., Sheridan, B. E., & Luo, G. (2012). *RUMM2030: Rasch Unidimensional Models for Measurement*. Perth, Western Australia: RUMMLaboratory.
- Andrich, D., & Styles, I. (2004a). Final report on the psychometric analysis of the Early Development Instrument (EDI) using the Rasch Model: A technical paper commissioned for the development of the Australian Early Development Instrument (AEDI). Perth, Western Australia: Murdoch University.
- Andrich, D., & Styles, I. (2004b). Report on the Rasch analysis of the Australian Early Development Instrument (AEDI) using 2004 data from 6 states. Perth, Western Australia: Murdoch University.
- Andrich, D., & Wright, B. D. (1994). Rasch sensitivity and Thurstone insensitivity to graded responses. *Rasch Measurement Transactions*, 8(3), 382.
- Arnott, B., & Meins, E. (2007). Links among antenatal attachment representations, postnatal mind-mindedness, and infant attachment security: A preliminary study of mothers and fathers. *Bulletin of the Menninger Clinic*, 71(2), 132-149.
- Asen, E., & Fonagy, P. (2012). Mentalization-based therapeutic interventions for families. *Journal of Family Therapy*, 34(4), 347-370. doi: 10.1111/j.1467-6427.2011.00552.x

- Baer, R. A., Smith, G. T., Hopkins, J., Krietemeyer, J., & Toney, L. (2006). Using self-report assessment methods to explore facets of mindfulness. *Assessment, 13*(1), 27-45.
- Baghaei, P. (2008). Local dependency and Rasch measures. *Rasch Measurement Transactions, 21*(3), 1105-1106.
- Ball, J., Moselle, K., & Pedersen, S. (2007). Father's involvement as a determinant of child health *Father involvement for healthy child outcomes: Partners supporting knowledge development and transfer*: Public Health Agency of Canada.
- Baradon, T., Fonagy, P., Bland, K., Lénárd, K., & Sled, M. (2008). New Beginnings—an experience-based programme addressing the attachment relationship between mothers and their babies in prisons. *Journal of Child Psychotherapy, 34*(2), 240-258. doi: 10.1080/00754170802208065
- Barnes, L. L. B., Harp, D., & Jung, W. S. (2002). Reliability generalization of scores on the Spielberger state-trait anxiety inventory. *Educational and Psychological Measurement, 62*(4), 603-618. doi: 10.1177/0013164402062004005
- Baron-Cohen, S. (1997). *Mindblindness: An essay on autism and theory of mind*: MIT press.
- Baron-Cohen, S. (2002). The extreme male brain theory of autism. *Trends in Cognitive Sciences, 6*(6), 248-254.
- Baron-Cohen, S., Wheelwright, S. J., Hill, J., Raste, Y., & Plumb, I. (2001). The "reading the mind in the eyes" test revised version: A study with normal adults, and adults with asperger syndrome or high-functioning autism. *Journal of Child Psychology and Psychiatry, 42*(2), 30.
- Bateman, A., W., & Fonagy, P. (2001). Treatment of borderline personality disorder with psychoanalytically oriented partial hospitalization: An 18-month follow-up. *American Journal of Psychiatry, 158*(1), 36-42.
- Bateman, A., W., & Fonagy, P. (2008a). 8-Year follow-up of patients treated for borderline personality disorder: Mentalization-based treatment versus treatment as usual. *The American Journal of Psychiatry, 165*(5), 631-638.
- Bateman, A., W., & Fonagy, P. (2008b). Comorbid antisocial and borderline personality disorders: Mentalization-based treatment. *Journal of Clinical Psychology, 64*(2), 181-194.

- Bateman, A., W., & Fonagy, P. (2012). Borderline personality disorder. In A. Bateman & P. Fonagy (Eds.), *Handbook of mentalizing in mental health practice* (pp. 273-288). Arlington, VA: American Psychiatric Publishing.
- Baumeister, R. F., Masicampo, E. J., & Vohs, K. D. (2011). Do conscious thoughts cause behavior? *Annual Review of Psychology*, *62*, 331-361.
- Baxter, J., & Smart, D. (2010). *Occasional Paper no. 37, Fathering in Australia among couple families with young children*. Canberra: Commonwealth of Australia.
- Beck, A., Steer, R., & Brown, G. (1996). Manual for the BDI-II: San Antonio, TX: Psychological Corporation.
- Belsky, J. (2010). *Experiencing the lifespan* (Second ed.). New York: Worth.
- Bevington, D., Fuggle, P., Fonagy, P., Target, M., & Asen, E. (2012). Innovations in Practice: Adolescent Mentalization-Based Integrative Therapy (AMBIT) – a new integrated approach to working with the most hard to reach adolescents with severe complex mental health needs. *Child and Adolescent Mental Health*, n/a-n/a. doi: 10.1111/j.1475-3588.2012.00666.x
- Blankenhorn, D. (1995). *Fatherless America: confronting our most urgent social problem*. New York: Harper Perennial.
- Bock, R. D. (1997). A brief history of item theory response. *Educational Measurement: Issues and Practice*, *16*(4), 21-33. doi: 10.1111/j.1745-3992.1997.tb00605.x
- Bond, T. G. (2004). Validity and assessment: a Rasch measurement perspective. *Metodologia de las Ciencias del Comportamiento*, *5*(2), 179-194.
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: fundamental measurement in the human sciences* (2nd ed.). Mahwah, N.J.: Lawrence Erlbaum.
- Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, *71*(3), 425-440. doi: 10.1007/s11336-006-1447-6
- Borsboom, D., & Scholten, A. Z. (2008). The Rasch model and conjoint measurement theory from the perspective of psychometrics. *Theory & Psychology*, *18*(1), 111-117. doi: 10.1177/0959354307086925
- Bowlby, J. (1969). *Attachment and loss: Attachment* (Vol. 1). London: Hogarth Press.

- Bretherton, I. (2010). Fathers in attachment theory and research: a review. *Early Child Development and Care, 180*(1-2), 9-23. doi: 10.1080/03004430903414661
- Brinkman, S., Silburn, S., & Lawrence, D. (2006). Construct and concurrent validity of the Australian Early Development Index. Perth, Western Australia: Centre for Community Child Health and Telethon Institute for Child Health Research.
- Brogden, H. E. (1977). The Rasch model, the law of comparative judgment and additive conjoint measurement. *Psychometrika, 42*(4), 631-634.
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, Massachusetts: Harvard University Press.
- Bronfenbrenner, U., & Morris, P. A. (1998). The ecology of developmental process. In R. Lerner (Ed.), *Handbook of child psychology* (4 ed., Vol. 1). New York: Wiley.
- Brooks-Gunn, J. (2003). Do you believe in magic? What can we expect from early childhood intervention programs. *Social Policy Report, 17*(1), 11.
- Burgess, A. (2007). The costs and benefits of active fatherhood: Evidence and insights to inform the development of policy and practice. London: Fathers Direct.
- Byles, J., Byrne, C., Boyle, M. H., & Offord, D. R. (1988). Ontario Child Health Study: Reliability and validity of the general functioning subscale of the McMaster Family Assessment Device. *Family Process, 27*(1), 97-104.
- Cabrera, N. J., Fitzgerald, H. E., Bradley, R. H., & Roggman, L. (2007). Modeling the dynamics of paternal influences on children over the life course. *Applied Developmental Science, 11*(4), 185-189. doi: 10.1080/10888690701762027
- Cano, S. J., & Hobart, J. (2011). The problem with health measurement. *Patient Preference and Adherence, 2011*(5), 279-290.
- Cano, S. J., Klassen, A. F., Scott, A., Thoma, A., Feeny, D., & Pusic, A. (2010). Health outcome and economic measurement in breast cancer surgery: challenges and opportunities. *Expert review of pharmacoeconomics & outcomes research, 10*(5), 583-594. doi: 10.1586/erp.10.61
- Carlson, M. J., Pilkauskas, N. V., McLanahan, S. S., & Brooks-Gunn, J. (2011). Couples as partners and parents over children's early years. *Journal of*

*Marriage and Family*, 73(2), 317-334. doi: 10.1111/j.1741-3737.2010.00809.x

- Carney, C. E., Ulmer, C., Edinger, J. D., Krystal, A. D., & Knauss, F. (2009). Assessing depression symptoms in those with insomnia: An examination of the beck depression inventory second edition (BDI-II). *Journal of Psychiatric Research*, 43(5), 576-582. doi: <http://dx.doi.org/10.1016/j.jpsychires.2008.09.002>
- Cassidy, J. (2008). The nature of the child's ties. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research and clinical applications* (2nd ed., pp. 3-22). New York: Guildford Press.
- Cattell, J. M. (1893). Mental measurement. *The Philosophical Review*, 2(3), 316-332.
- Cliff, N. (1992). Abstract measurement theory and the revolution that never happened. *Psychological Science*, 3(3), 186-190.
- Cohen, J. (1992). A power primer. *Psychological Bulletin*, 112(1), 155. doi: 10.1037/0033-2909.112.1.155
- Collins, W. A., Maccoby, E. E., Steinberg, L., Hetherington, E. M., & Bornstein, M. H. (2000). Contemporary research on parenting: The case for nature and nurture. *American Psychologist*, 55(2), 15.
- Cook, W. L., & Kenny, D. A. (2006). Examining the validity of self-report assessments of Family Functioning: A question of the level of analysis. *Journal of Family Psychology*, 20(2), 209-216. doi: 10.1037/0893-3200.20.2.209
- Cooke, D., Marais, I., Cavanagh, R., Kendall, G., & Priddis, L. (2015). Differences Between Mothers' and Fathers' Ratings of Family Functioning With the Family Assessment Device The Validity of Combined Parent Scores. *Measurement and evaluation in counseling and development*, 48(3), 226-237.
- Court, H., Greenland, K., & Margrain, T. H. (2010). Measuring patient anxiety in primary care: Rasch analysis of the 6-item Spielberger State Anxiety Scale. *Value in Health*, 13(6), 813-819.
- Cox, M. J., & Paley, B. (1997). Families as systems. *Annual Review of Psychology*, 48(1), 243-267. doi: 10.1146/annurev.psych.48.1.243
- Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297-334. doi: 10.1007/BF02310555

- Cummings, E. M., Goeke-Morey, M. C., & Raymond, J. (2004). Fathers in family context: Effects of marital quality and marital conflict. In M. E. Lamb (Ed.), *The role of the father in child development* (4 ed.). Hoboken, New Jersey: John Wiley.
- Cynader, M. S., & Frost, B. J. (1999). Mechanisms of brain development: Neuronal sculpting by the physical and social environment. In D. Keating, P & C. Hertzman (Eds.), *Developmental health and the wealth of nations: Social, biological, and educational dynamics* (pp. 185-200). New York: The Guildford Press.
- da Rocha, N. S., Chachamovich, E., de Almeida Fleck, M. P., & Tennant, A. (2013). An introduction to Rasch analysis for psychiatric practice and research. *Journal of Psychiatric Research*, *47*(2), 141-148. doi: 10.1016/j.jpsychires.2012.09.014
- Davies, D. (2004). *Child Development: a Practitioners Guide* (2 ed.). New York: Guildford Press.
- Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *Catalog of Selected Documents in Psychology*, *10*, 85.
- De Wolff, M., S, & van IJzendoorn, M., H. (1997). Sensitivity and attachment: A meta-analysis on parental antecedents of infant attachment. *Child Development*, *68*(4), 571-591.
- Dimaggio, G., Carcione, A., Salvatore, G., Nicolò, G., Sisto, A., & Semerari, A. (2011). Progressively promoting metacognition in a case of obsessive-compulsive personality disorder treated with metacognitive interpersonal therapy. *Psychology and Psychotherapy: Theory, Research and Practice*, *84*(1), 70-83. doi: 10.1348/147608310x527240
- Dozier, M., Stovall-McClough, K. C., & Albus, K. E. (2008). Attachment and psychopathology in adulthood. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (2nd ed., pp. 718-744). New York: Guildford Press.
- Drielsma, P. (2001). Hard wiring young brains for intimacy: The essential first step to breaking generational cycles of abuse and neglect. *Journal of the HEIA*, *8*(2), 19-23.
- Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment*, *8*(4), 341-349.

- Embretson, S. E. (2006). The continued search for nonarbitrary metrics in psychology. *The American psychologist, 61*(1), 50-55. doi: 10.1037/0003-066X.61.1.50
- Embretson, S. E. (2007). Construct validity: A universal validity system or just another test evaluation procedure? *Educational Researcher, 36*(8), 449-455.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Engelhard, G. (1994). Historical views of the concept of invariance in measurement theory. In M. Wilson (Ed.), *Objective measurement: Theory into practice* (Vol. 2, pp. 73-100). Norwood, New Jersey: Ablex.
- Engelhard, G. (2008). Historical perspectives on invariant measurement: Guttman, Rasch, and Mokken. *Measurement: Interdisciplinary Research & Perspective, 6*(3), 155-189. doi: 10.1080/15366360802197792
- Enos, M. (2001). The language of response. In B. D. Wright, M. Enos, M. Enos & J. M. Linacre (Eds.), *Adventures in questionnaire design: Poetics, posters, and provocations* (pp. 62-69). Chicago: MESA Press.
- Epstein, N. B., Baldwin, L., & Bishop, D. S. (1983). The McMaster Family Assessment Device. *Journal of Marital and Family Therapy, 9*(2), 171-180.
- Essex, M. J., Thomas Boyce, W., Hertzman, C., Lam, L. L., Armstrong, J. M., Neumann, S. M. A., & Kobor, M. S. (2013). Epigenetic vestiges of early developmental adversity: Childhood stress exposure and DNA methylation in adolescence. *Child Development, 84*(1), 58-75. doi: 10.1111/j.1467-8624.2011.01641.x
- Fisher, W. P. (1998). A research program for accountable and patient-centered health outcome measures. *Journal of Outcome Measurement, 2*(3), 222-239.
- Fisher, W. P. (2000). Objectivity in psychosocial measurement: What, why, how. *Journal of Outcome Measurement, 4*(2), 527-563.
- Fisher, W. P. (2004). Meaning and method in the social sciences. *Human Studies, 27*(4), 429-454. doi: 10.1007/s10746-004-3339-z
- Fisher, W. P. (2005). Meaningfulness, measurement and item response theory (IRT). *Rasch Measurement Transactions, 19*(2), 1018-1020.
- Fisher, W. P. (2010). Reliability, precision, and measurement in the context of data from ability tests, surveys, and assessments. *Journal of Physics: Conference Series*(238), 1-6.

- Fonagy, P. (1999). Male perpetrators of violence against women: An attachment theory perspective. *Journal of Applied Psychoanalytic Studies, 1*, 7-27.
- Fonagy, P., Bateman, A., W., & Bateman, A. (2011). The widening scope of mentalizing: A discussion. *Psychology and Psychotherapy: Theory, Research and Practice, 84*(1), 98-110. doi: 10.1111/j.2044-8341.2010.02005.x
- Fonagy, P., Bateman, A., W., & Luyten, P. (2012). Introduction and overview. In A. Bateman & P. Fonagy (Eds.), *Handbook of mentalizing in mental health practice* (pp. 3-42). Arlington, VA: American Psychiatric Publishing.
- Fonagy, P., Gergely, G., Jurist, E., & Target, M. (2002). *Affect regulation, mentalization, and the development of the self*. New York: Other Press.
- Fonagy, P., Gergely, G., & Target, M. (2007). The parent-infant dyad and the construction of the subjective self. *Journal of Child Psychology and Psychiatry, 48*(3-4), 288-328. doi: doi:10.1111/j.1469-7610.2007.01727.x
- Fonagy, P., Leigh, T., Steele, M., Steele, H., Kennedy, R., Mattoon, G., . . . Gerber, A. (1996). The relation of attachment status, psychiatric classification, and response to psychotherapy. *Journal of Consulting and Clinical Psychology, 64*(1), 22-31.
- Fonagy, P., & Luyten, P. (2009). A developmental, mentalization-based approach to the understanding and treatment of borderline personality disorder. *Development and psychopathology, 21*(4), 1355-1381. doi: 10.1017/S0954579409990198
- Fonagy, P., Steele, M., Steele, H., Moran, G., S., & Higgitt, A., C. (1991). The capacity for understanding mental states: The reflective self in parent and child and its significance for security of attachment. *Infant Mental Health Journal, 12*(3), 201-218.
- Fonagy, P., & Target, M. (2005). Bridging the transmission gap: An end to an important mystery of attachment research? *Attachment & Human Development, 7*(3), 333 - 343.
- Fonagy, P., Target, M., Steele, H., & Steele, M. (1998). *Reflective-functioning manual: Verson 5.0 for application to the adult attachment interview*. Unpublished manual. University College. London. Retrieved from <http://mentalizacion.com.ar/images/notas/Reflective Functioning Manual.pdf>

- Fox, C. M., & Jones, J. A. (1998). Uses of Rasch modeling in counseling psychology research. *Journal of Counseling Psychology, 45*(1), 30-45.
- Fox, N. A., Kimmerly, N. L., & Schafer, W. D. (1991). Attachment to mother/attachment to father: A meta-analysis. *Child Development, 62*(1), 210-225.
- Fox, S. E., Levitt, P., & Nelson, C. A. (2010). How the timing and quality of early experiences influence the development of brain architecture. *Child Development, 81*(1), 28-40.
- Fraley, R. C. (2002). Attachment stability from infancy to adulthood: Meta-analysis and dynamic modeling of developmental mechanisms. *Personality and Social Psychology Review, 6*(2), 123-151. doi: 10.1207/s15327957pspr0602\_03
- Frith, C. D., & Frith, U. (2006). The neural basis of mentalizing. *Neuron, 50*(4), 531-534. doi: 10.1016/j.neuron.2006.05.001
- Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions: Biological Sciences, 358*(1431), 459-473.
- Gaynes, B. N., Gavin, N., Meltzer-Brody, S., Lohr, K. N., Swinson, T., Gartlehner, G., . . . Miller, W. C. (2005). Perinatal depression: Prevalence, screening accuracy, and screening outcomes: summary *Evidence Report/Technology Assessment: Number 119*. Rockville, MD: Agency for Healthcare Research and Quality.
- Georgiades, K., Boyle, M. H., Jenkins, J. M., Sanford, M., & Lipman, E. (2008). A multilevel analysis of whole family functioning using the McMaster Family Assessment Device. [References]. *Journal of Family Psychology, 22*(3), 344-354.
- Giudice, M. D., & Belsky, J. (2010). Evolving attachment theory: Beyond bowlby and back to darwin. *Child Development Perspectives, 4*(2), 112-113.
- Goleman, D. (1995). *Emotional Intelligence*: Bantam Books.
- Grant, K. A., McMahon, C., & Austin, M. P. (2008). Maternal anxiety during the transition to parenthood: a prospective study. *J Affect Disord, 108*(1-2), 101-111. doi: 10.1016/j.jad.2007.10.002
- Green, R. G., & Vosler, N. R. (1992). Issues in the assessment of family practice. *Journal of Social Service Research, 15*(3-4), 1-19. doi: 10.1300/J079v15n03\_01

- Grienenberger, J., Kelly, K., & Slade, A. (2005). Maternal reflective functioning, mother–infant affective communication, and infant attachment: Exploring the link between mental states and observed caregiving behavior in the intergenerational transmission of attachment. *Attachment and Human Development, 7*(3), 12.
- Grossmann, K., & Grossmann, K. E. (2009). The impact of attachment to mother and father and sensitive support of exploration at an early age on children’s psychosocial development through young adulthood. *Encyclopedia on Early Childhood Development, 1-8*. <http://www.child-encyclopedia.com/Pages/PDF/GrossmannANGxp3.pdf>
- Grossmann, K., Grossmann, K. E., Fremmer-Bombik, E., Kindler, H., Scheuerer-Engelisch, H., & Zimmermann, P. (2002). The uniqueness of the child-father attachment relationship: Fathers' sensitive and challenging play as a pivotal variable in a 16-year longitudinal study. *Social Development, 11*(3), 301-337. doi: [doi:10.1111/1467-9507.00202](https://doi.org/10.1111/1467-9507.00202)
- Grossmann, K., Grossmann, K. E., & Kindler, H. (2005). Early care and the roots of attachment and partnership representations: The Bielefeld and Regensburg longitudinal studies. In K. Grossmann, K. E. Grossmann & E. Waters (Eds.), *Attachment from infancy to adulthood: The major longitudinal studies* (pp. 98-136). New York: Guilford Press.
- Grossmann, K., Grossmann, K. E., Kindler, H., & Zimmermann, P. (2008). A wider view of attachment and exploration: the influence of mothers and fathers on the development of psychological security from infancy to young adulthood. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (2nd ed., pp. 857-879). New York: Guilford Press.
- Grossmann, K. E., Grossmann, K., Huber, F., & Wartner, U. (1981). German children's behavior towards their mothers at 12 months and their fathers at 18 months in insworth's strange situation. *International Journal of Behavioral Development, 4*(2), 157-181. doi: [10.1177/016502548100400202](https://doi.org/10.1177/016502548100400202)
- Hagquist, C., Bruce, M., & Gustavsson, J. P. (2009). Using the Rasch model in nursing research: An introduction and illustrative example. *International journal of nursing studies, 46*(3), 380-393.

- Hayden, L. C., Schiller, M., Dickstein, S., Seifer, R., Sameroff, S., Miller, I., . . . Rasmussen, S. (1998). Levels of family assessment: I. Family, marital, and parent-child interaction. *Journal of Family Psychology, 12*(1), 7.
- Heene, M. (2013). Additive conjoint measurement and the resistance toward falsifiability in psychology. *Frontiers in Psychology, 4*. doi: 10.3389/fpsyg.2013.00246
- Hesse, E. (2008). The Adult Attachment Interview: Protocol, method of analysis, and empirical studies. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment: Theory, research, and clinical applications* (2nd ed., pp. 552-598). New York: Guilford Press.
- HILDA. (2014). The Household, Income and Labour Dynamics in Australia Survey. from <http://www.aifs.gov.au/growingup/>
- Hill, J., Fonagy, P., Safier, E., & Sargent, J. (2003). The ecology of attachment in the family. *Family Process, 42*(2), 205-221.
- Hobart, J. C., & Cano, S. J. (2009). Improving the evaluation of therapeutic interventions in multiple sclerosis: The role of new psychometric methods. *Health Technology Assessment, 13*(12).
- Hobart, J. C., Cano, S. J., Zajicek, J. P., & Thompson, A. J. (2007). Rating scales as outcome measures for clinical trials in neurology: problems, solutions, and recommendations. *The Lancet Neurology, 6*(12), 1094-1105. doi: 10.1016/s1474-4422(07)70290-9
- Hofferth, S. L. (2002). The demography of fathers. In C. S. Tamis-LeMonda & N. J. Cabrera (Eds.), *Handbook of father involvement: Multidisciplinary perspectives*. New Jersey: Lawrence Erlbaum Associates.
- Howell, D. C. (2002). *Statistical methods for psychology* (5th ed.). Pacific Grove, CA: Duxbury.
- Howieson, J., & Priddis, L. (2011). Building resilience for separating parents through mentalizing and constructive lawyering techniques. *Psychiatry, Psychology and Law, 18*(2), 202-211. doi: 10.1080/13218711003739532
- Howieson, N. D., & Priddis, L. (2012). Mentalising in mediation: Towards an understanding of the "mediation shift". *Australasian Dispute Resolution Journal, 23*(1), 52-60.

- Irwin, L., Siddiqi, A., & Hertzman, C. (2007). Early child development: A powerful equalizer *World Health Organisation: Commission on the Social Determinants of Health*. Vancouver BC: University of British Columbia.
- Kabacoff, R. I., Miller, I. W., Bishop, D. S., Epstein, N. B., & Keitner, G. I. (1990). A psychometric study of the McMaster Family Assessment Device in psychiatric, medical, and nonclinical samples. [References]. *Journal of Family Psychology*, 3(4), 431-439.
- Karabatsos, G. (2001). The Rasch model, additive conjoint measurement, and new models of probabilistic measurement theory. *Journal of Applied Measurement*, 2(4), 389-423.
- Katz, L. F., Maliken, A. C., & Stettler, N. M. (2012). Parental meta-emotion philosophy: A review of research and theoretical framework. *Child Development Perspectives*, n/a-n/a. doi: 10.1111/j.1750-8606.2012.00244.x
- Katznelson, H. (2014). Reflective functioning: A review. *Clinical Psychology Review*, 34(2), 107-117. doi: <http://dx.doi.org/10.1016/j.cpr.2013.12.003>
- Kendall, G. E., & Li, J. (2005). Early childhood socialization and social gradients in adult health: A commentary on Singh-Manoux and Marmot's "Role of socialization in explaining social inequalities in health"(60: 9, 2005, 2129–2133). *Social Science & Medicine*, 61(11), 2272-2276.
- Kennedy, B. L., Schwab, J. J., Morris, R. L., & Beldia, G. (2001). Assessment of state and trait anxiety in subjects with anxiety and depressive disorders. *Psychiatric Quarterly*, 72(3), 263-276.
- Kline, P. (2000). *The handbook of psychological testing* (Second ed.). New York: Routledge.
- Kochanska, G., & Kim, S. (2013). Early attachment organization with both parents and future behavior problems: From infancy to middle childhood. *Child Development*, 84(1), 283-296. doi: 10.1111/j.1467-8624.2012.01852.x
- Kuhn, T. S. (1961/1977). The function of measurement in modern physical science. In T. S. Kuhn (Ed.), *The essential tension* (pp. 161-190). Chicago: The University of Chicago Press.
- Kulas, J. T., & Stachowski, A. A. (2013). Respondent rationale for neither agreeing nor disagreeing: Person and item contributors to middle category endorsement intent on Likert personality indicators. *Journal of Research in*

*Personality*, 47(4), 254-262. doi:

<http://dx.doi.org/10.1016/j.jrp.2013.01.014>

- Kyngdon, A. (2008). Conjoint measurement, error and the Rasch model: A reply to Michell, and Borsboom and Zand Scholten. *Theory & Psychology*, 18(1), 125-131. doi: 10.1177/0959354307086927
- Lamb, M. E. (1978). Qualitative aspects of mother- and father-infant attachments. *Infant Behavior and Development*, 1, 265-275.
- Lamb, M. E. (2000). The history of research on father involvement. *Marriage & Family Review*, 29(2), 23 - 42.
- Lamb, M. E. (2005). Attachments, social networks, and developmental contexts. *Human Development*, 48(1-2), 108-112.
- Lamb, M. E. (2010a). How do fathers influence children's development? Let me count the ways. In M. Lamb (Ed.), *The role of the father in child development* (5 ed., pp. 1-26). Chichester: Wiley.
- Lamb, M. E. (2010b). *The role of the father in child development* (5 ed.). Chichester: Wiley.
- Lamb, M. E., & Lewis, C. (2004). The development and significance of father-child relationships in two-parent families. In M. E. Lamb (Ed.), *The Role of the Father in Child Development* (4 ed., pp. 272-306). New Jersey: John Wiley & Sons.
- Lamb, M. E., & Lewis, C. (2010). The development and significance of father-child relationships in two-parent families. In M. Lamb (Ed.), *The role of the father in child development* (5 ed., pp. 94-153). Chichester: Wiley.
- Lamb, M. E., Pleck, J. H., Charnov, E. L., & Levine, J. A. (1985). Paternal behavior in humans. *American Zoologist*, 25(3), 883-894.
- LASC. (2014). Growing Up In Australia: The Longitudinal Study of Australian Children. from <http://www.growingupinaustralia.gov.au>
- Leichsenring, F., Kunst, H., & Hoyer, J. (2003). Borderline personality organization in violent offenders: Correlations of identity diffusion and primitive defense mechanisms with antisocial features, neuroticism, and interpersonal problems. *Bulletin of the Menninger Clinic*, 67(4), 314-327.
- Lerdal, A., Kottorp, A., Gay, C. L., Grov, E. K., & Lee, K. A. (2014). Rasch analysis of the Beck Depression Inventory-II in stroke survivors: A cross-sectional

- study. *Journal of Affective Disorders*, 158, 48-52. doi:  
<http://dx.doi.org/10.1016/j.jad.2014.01.013>
- Levant, R., F. (1992). Toward the reconstruction of masculinity. *Journal of Family Psychology*, 5(3-4), 379.
- Levant, R., F., Hall, R., J., Williams, C., M., & Hasan, N., T. (2009). Gender differences in alexithymia. *Psychology of Men and Masculinity*, 10(3), 190-203. doi: 10.1037/a0015652
- Likert, R. A. (1932). A technique for the measurement of attitudes. *Archives of Psychology*, 140, 5-55.
- Likert, R. A., Roslow, S., & Murphy, G. (1934). A simple and reliable method of scoring the Thurstone attitude scales. *Journal of Social Psychology*, (5), 228-238. [http://www.brocku.ca/MeadProject/Likert/Likert\\_1934.html](http://www.brocku.ca/MeadProject/Likert/Likert_1934.html)
- Linacre, J. M. (1994). Sample size and item calibration stability. *Rasch Measurement Transactions*, 7(4), 328.
- Linacre, J. M. (1999a). Category disordering (disordered categories) vs. threshold disordering (disordered thresholds). *Rasch Measurement Transactions*, 13(1), 675.
- Linacre, J. M. (1999b). Investigating rating scale category utility. *Journal of Outcome Measurement*, 3(2), 103-122.
- Linacre, J. M. (2000). Computer-adaptive testing: A methodology whose time has come. In S. Chae, U. Kang, E. Jeon & J. M. Linacre (Eds.), *Development of computerized middle school achievement test*. Seoul, South Korea: Komesa Press.
- Linacre, J. M. (2002). Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85-106.
- Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Wheelwright, S. J., Sadek, S. A., Suckling, J., . . . Consortium, M. A. (2010). Shared neural circuits for mentalizing about the self and others. *Journal of Cognitive Neuroscience*, 22(7), 1623-1635.
- Lord, F. M. (1953). On the statistical treatment on 'football numbers'. *American Psychologist*, 9, 750-751.
- Luce, R. D., & Tukey, J. W. (1964). Simultaneous conjoint measurement: A new type of fundamental measurement. *Journal of Mathematical Psychology*, 1, 1-27.

- Lupton, D., & Barclay, L. (1997). *Constructing fatherhood: Discourses and experiences*. London: Sage.
- Luyten, P. (2014). *The Assessment of Mentalization*. Presentation Slides. Retrieved from [http://www.deviersprong.nl/files/Viersprong\\_Academy/The\\_Assessment\\_of\\_Mentalization.pdf](http://www.deviersprong.nl/files/Viersprong_Academy/The_Assessment_of_Mentalization.pdf)
- Luyten, P., Fonagy, P., Lemma, A., & Target, M. (2012). Depression. In A. Bateman & P. Fonagy (Eds.), *Handbook of mentalizing in mental health practice* (pp. 385-417). Arlington, VA: American Psychiatric Publishing.
- Luyten, P., Fonagy, P., Lowyck, B., & Vermonte, R. (2012). Assessment of mentalizing. In A. Bateman, W & P. Fonagy (Eds.), *Handbook of mentalizing in mental health practice* (pp. 43-65). Arlington, VA: American Psychiatric Publishing.
- Luyten, P., Mayes, L., C, Sadler, L. S., Fonagy, P., Nicholls, S., Crowley, M. J., . . . Slade, A. (2009). *Manual for the Parental Reflective Functioning Questionnaire-1 (PRFQ-1)*. Unpublished Manual. Yale Child Study Center. New Haven, CT.
- Lyons-Ruth, K. (2006). The interface between attachment and intersubjectivity: Perspective from the Longitudinal Study of Disorganized Attachment. *Psychoanalytic Inquiry*, 26(4), 595-616.
- Lysaker, P. H., Buck, K. D., Carcione, A., Procacci, M., Salvatore, G., Nicolò, G., & Dimaggio, G. (2011). Addressing metacognitive capacity for self reflection in the psychotherapy for schizophrenia: A conceptual model of the key tasks and processes. *Psychology and Psychotherapy: Theory, Research and Practice*, 84(1), 58-69. doi: 10.1348/147608310x520436
- Magalhães, P., V, Pinheiro, R., T, Horta, B., L, Pinheiro, K., A, T, & Da Silva, R., A. (2008). Validity of the Beck Depression Inventory in the postpartum period. *International Journal of Psychiatry in Clinical Practice*, 12(1), 81-84.
- Maggi, S., Irwin, L., Siddiqi, A., & Hertzman, C. (2010). The social determinants of early child development: an overview. *Journal of Paediatrics and Child Health*, 46(11), 627-635. doi: 10.1111/j.1440-1754.2010.01817.x
- Main, M., Kaplan, N., & Cassidy, J. (1985). Security in infancy, childhood, and adulthood: A move to the level of representation. *Monographs of the Society for Research in Child Development*, 50(1/2), 66-104.

- Main, M., & Weston, D. R. (1981). The quality of the toddler's relationship to mother and to father: Related to conflict behavior and the feadiness to establish new relationships. *Child Development, 52*(3), 932-940.
- Marais, I., & Andrich, D. (2008). Formalizing dimension and response violations of local independence in the unidimensional Rasch model. *Journal of Applied Measurement, 9*(3), 200-215.
- Masters, G. A. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174.
- Matthey, S., Barnett, B., Howie, P., & Kavanagh, D., J. (2003). Diagnosing postpartum depression in mothers and fathers: whatever happened to anxiety? *Journal of Affective Disorders, 74*, 139-147.
- Mayer, J. D., Roberts, R. D., & Barsade, S. G. (2008). Human abilities: Emotional intelligence. *Annual Review of Psychology, 59*, 507-536. doi: 10.1146/annurev.psych.59.103006.093646
- Mayes, L. C. (2006). Arousal regulation, emotional flexibility, medial amygdala function, and the impact of early experience. *Ann N Y Acad Sci, 1094*(1), 178-192.
- McCain, M. N., Mustard, J. F., & Shanker, S. (2007). *Early years study 2: Putting science into action*. Roronto Ontario, Canada: Council for Early Child Development.
- McConnell, M., & Moss, E. (2011). Attachment across the life span: Factors that contribute to stability and change. *Australian Journal of Educational & Developmental Psychology, 11*, 60-77.
- McEwen, B., S. (2003). Early life influences on life-long patterns of behavior and health. *Mental Retardation and Developmental Disabilities Research Reviews, 9*(3), 149-154.
- McNemar, Q. (1946). Opinion-attitude methodology. *Psychological Bulletin, 43*, 289-374.  
[http://www.brocku.ca/MeadProject/McNemar/McNemar\\_1946.html](http://www.brocku.ca/MeadProject/McNemar/McNemar_1946.html)
- Meaney, M. J. (2001). Maternal care, gene expression, and the transmission of individual differences in stress reactivity across generations. *Annual Review of Neuroscience, 24*(1), 1161-1192. doi: doi:10.1146/annurev.neuro.24.1.1161

- Meaney, M. J. (2010). Epigenetics and the biological definition of gene x environment interactions. *Child Development, 81*(1), 41-79.
- Meins, E., Fernyhough, C., Arnott, B., Leekam, S. R., & de Rosnay, M. (2013). Mind-mindedness and theory of mind: Mediating roles of language and perspectival symbolic play. *Child Development, n/a-n/a*. doi: 10.1111/cdev.12061
- Meins, E., Fernyhough, C., Fradley, E., & Tuckey, M. (2001). Rethinking maternal sensitivity: Mothers' comments on infants' mental processes predict security of attachment at 12 months. *Journal of Child Psychology and Psychiatry, 42*(5), 637-648.
- Meins, E., Fernyhough, C., Wainwright, R., Gupta, M. D., Fradley, E., & Tuckey, M. (2002). Maternal mind-mindedness and attachment security as predictors of theory of mind understanding. *Child Development, 73*(6), 1715-1726.
- Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*(2), 5-11.
- Michell, J. (1997a). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology, 88*(3), 355-383. doi: 10.1111/j.2044-8295.1997.tb02641.x
- Michell, J. (1997b). Reply to Kline, Laming, Lovie, Luce and Morgan. *British Journal of Psychology, 88*(3), 401-406. doi: 10.1111/j.2044-8295.1997.tb02647.x
- Michell, J. (2002). Stevens's theory of scales of measurement and its place in modern psychology. *Australian Journal of Psychology, 54*(2), 99-104.
- Michell, J. (2009). The psychometricians' fallacy: Too clever by half? *British Journal of Mathematical and Statistical Psychology, 62*(1), 41-55.
- Milgrom, J., Negri, L. M., Gemmill, A. W., McNeil, M., & Martin, P. R. (2005). A randomized controlled trial of psychological interventions for postnatal depression. *British Journal of Clinical Psychology, 44*(4), 529-542.
- Miller, I. W., Kabacoff, R. I., Epstein, N. B., Bishop, D. S., Keitner, G. I., Baldwin, L. M., & van der Spuy, H. I. J. (1994). The development of a clinical rating scale the McMaster model of family functioning. *Family Process, 33*(1), 53-69. doi: 10.1111/j.1545-5300.1994.00053.x
- Miller, I. W., Ryan, C. E., Keitner, G. I., Bishop, D. S., & Epstein, N. B. (2000a). "Factor analyses of the Family Assessment Device," by Ridenour, Daley, &

- Reich. *Family Process*, 39(1), 141-144. doi: 10.1111/j.1545-5300.2000.39112.x
- Miller, I. W., Ryan, C. E., Keitner, G. I., Bishop, D. S., & Epstein, N. B. (2000b). The McMaster approach to families: Theory, assessment, treatment and research. *Journal of Family Therapy*, 22(2), 168-189. doi: 10.1111/1467-6427.00145
- Miller, I. W., Ryan, C. E., Keitner, G. I., Bishop, D. S., & Epstein, N. B. (2000c). Why fix what isn't broken? A rejoinder to Ridenour, Daley, & Reich. *Family Process*, 39(3), 381-384.
- Murray, L., Fiori-Cowley, A., Hooper, R., & Cooper, P. (1996). The impact of postnatal depression and associated adversity on early mother-infant interactions and later infant outcome. *Child Development*, 67(5), 2512-2526. doi: 10.1111/j.1467-8624.1996.tb01871.x
- National Scientific Council on the Developing Child. (2007). The science of early childhood development. <http://www.developingchild.net>
- Newby, V. A., Conner, G. R., Grant, C. P., & Bunderson, C. V. (2009). The Rasch model and additive conjoint measurement. *Journal of Applied Measurement*, 10(4), 348-354.
- Newland, L., & Coyl, D. (2010). Fathers' role as attachment figures: an interview with Sir Richard Bowlby. *Early Child Development and Care*, 180(1), 25-32. doi: 10.1080/03004430903414679
- Nolte, T., Guiney, J., Fonagy, P., Mayes, L., & Luyten, P. (2011). Interpersonal stress regulation and the development of anxiety disorders: an attachment-based developmental framework. *Front Behav Neurosci*, 5, 55. doi: 10.3389/fnbeh.2011.00055
- Nunnally, J. C., Bernstein, I. H., & Berge, J. M. t. (1967). *Psychometric theory* (Vol. 226): McGraw-Hill New York.
- Ordway, M. R., Sadler, L. S., Dixon, J., Close, N., Mayes, L., & Slade, A. (2013). Lasting effects of an interdisciplinary home visiting program on child behavior: Preliminary follow-up results of a randomized trial. *Journal of pediatric nursing*.
- Osborne, J. W. (2010). Challenges for quantitative psychology and measurement in the 21st century. *Frontiers in Psychology*, 1, 1-3. doi: 10.3389/fpsyg.2010.00001

- Palkovitz, R. (2002). Involved fathering and child development: Advancing our understanding of good fathering. In C. S. Tamis-LeMonda & N. J. Cabrera (Eds.), *Handbook of father involvement: Multidisciplinary perspectives*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Palkovitz, R. (2007). Challenges to modeling dynamics in developing a developmental understanding of father-child relationships. *Applied Developmental Science, 11*(4), 190-195. doi: 10.1080/10888690701762050
- Pallant, J., & Tennant, A. (2007). An introduction to the Rasch measurement model: An example using the Hospital Anxiety and Depression Scale (HADS). *British Journal of Clinical Psychology, 46*, 1-18.
- Paquette, D. (2004a). Dichotomizing paternal and maternal functions as a means to better understand their primary contributions. *Human Development, 47*(4), 237.
- Paquette, D. (2004b). Theorizing the father-child relationship: Mechanisms and developmental outcomes. *Human Development, 47*(4), 193.
- Paquette, D., & Bigras, M. (2010). The risky situation: a procedure for assessing the father-child activation relationship. *Early Child Development and Care, 180*(1-2), 33-50. doi: 10.1080/03004430903414687
- Park, R., & Brott, A. A. (1999). *Throwaway dads: the myths and barriers that keep men from being the fathers they want to be*. New York: Houghton Mifflin.
- Parkitny, L., McAuley, J. H., Walton, D., Pena Costa, L. O., Refshauge, K. M., Wand, B. M., . . . Moseley, G. L. (2012). Rasch analysis supports the use of the Depression, Anxiety, and Stress Scales to measure mood in groups but not in individuals with chronic low back pain. *Journal of Clinical Epidemiology, 65*(2), 189-198. doi: 10.1016/j.jclinepi.2011.05.010
- Paulhus, D. L. (1991). Measurement and control of response bias. In J. P. Robinson, P. R. Shaver & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 17-59). San Diage: Academic Press.
- Peel Development Commission. (2012). Peel Profile. Retrieved April 3, 2014, from <http://peel.wa.gov.au/wp-content/uploads/PeelProfileMarch2012.pdf>
- Philips, B., Kahn, U., & Bateman, A., W. (2012). Drug addiction. In A. Bateman & P. Fonagy (Eds.), *Handbook of mentalizing in mental health practice* (pp. 445-461). Arlingron, VA: American Psychiatric Publishing.

- Pleck, J. H. (2007). Why could father involvement benefit children? Theoretical perspectives. *Applied Developmental Science, 11*(4), 196-202. doi: 10.1080/10888690701762068
- Pleck, J. H. (2010). Paternal involvement: Revised conceptualization and theoretical linkages with child outcomes. In M. Lamb (Ed.), *The role of the father in child development* (5 ed., pp. 58-93). Chichester: Wiley.
- Poulin-Dubois, D., Brooker, I., & Chow, V. (2009). The developmental origins of naïve psychology in infancy. *Advances in Child Development and Behavior, 37i*, 55-104.
- Ramchandani, P. G., Domoney, J., Sethna, V., Psychogiou, L., Vlachos, H., & Murray, L. (2012). Do early father–infant interactions predict the onset of externalising behaviours in young children? Findings from a longitudinal cohort study. *Journal of Child Psychology and Psychiatry, 54*(1), 56-64. doi: 10.1111/j.1469-7610.2012.02583.x
- Ranson, K., E., & Urichuk, L., J. (2008). The effect of parent–child attachment relationships on child biopsychosocial outcomes: a review. *Early Child Development and Care, 178*(2), 129-152.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests* (reprinted 1980, expanded ed. with forward and afterward by B. D. Wright). Chicago: The University of Chicago Press.
- Rasch, G. (1961). *On general laws and the meaning of measurement in psychology*. Paper presented at the Proceedings of the fourth Berkeley symposium on mathematical statistics and probability, University of California, Berkeley.
- Rasch, G. (1966). An item analysis which takes individual differences into account. *The British Journal of Mathematical and Statistical Psychology, 19*(1), 49-57.
- Repetti, R. L., Taylor, S. E., & Seeman, T. E. (2002). Risky families: Family social environments and the mental and physical health of offspring. *Psychological Bulletin, 128*(2), 330-366. doi: 10.1037//0033-2909.128.2.330
- Reynolds, D. (2003). Mindful Parenting: a group approach to enhancing reflective capacity in parents and infants. *Journal of Child Psychotherapy, 29*(3), 357-374. doi: 10.1080/00754170310001625413
- Ridenour, T. A., Daley, J., & Reich, W. (1999). Factor analyses of the family assessment device. *Fam Process, 38*(4), 497-510.

- Ridenour, T. A., Daley, J., & Reich, W. (2000). Further evidence that the family assessment device should be reorganized: Response to Miller and colleagues. *Family Process, 39*(3), 375-380.
- Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience, 27*(1), 169-192. doi:  
doi:10.1146/annurev.neuro.27.070203.144230
- Roisman, G. I., Holland, A., Fortuna, K., Fraley, R. C., Clausell, E., & Clarke, A. (2007). The Adult Attachment Interview and self-reports of attachment style: an empirical rapprochement. *Journal of Personality and Social Psychology, 92*(4), 678.
- Rosenberg, J., & Wilcox, W. B. (2006). *The importance of fathers in the healthy development of children*: US Department Health and Human Services, Administration for Children and Families, Administration on Children, Youth and Families, Children's Bureau, Office of Child Abuse and Neglect.
- Rossouw, T. I., & Fonagy, P. (2012). Mentalization-based treatment for self-harm in adolescents: a randomized controlled trial. *Journal of the American Academy of Child & Adolescent Psychiatry, 51*(12), 1304-1313. e1303.
- Royal, K. D. (2010). Making meaningful measurement in survey research: A demonstration of the utility of the Rasch model. *IR Applications, 28*(Sept 24), 1-15.
- RUMM Laboratory. (2009). Interpreting RUMM2030 Part IV. Multidimensionality and subtests in RUMM. Perth, Australia: RUMM Laboratory.
- RUMM Laboratory. (2012). Frequently asked questions. from  
<http://www.rummlab.com.au/>
- Sadler, L. S., Slade, A., Close, N., Webb, D. L., Simpson, T., Fennie, K., & Mayes, L. C. (2013). Minding the Baby: Enhancing reflectiveness to improve early health and relationship outcomes in an interdisciplinary home visiting program. *Infant Ment Health J, 34*(5), 391-405. doi: 10.1002/imhj.21406
- Safier, E. (2003). Seven ways that the concepts of attachment, mentalization and theory of mind transform family treatment. *Bulletin of the Menninger Clinic, 67*(3), 260-270.
- Salzberger, T., & Koller, M. (In Press). Towards a new paradigm of measurement in marketing. *Journal of Business Research*(0). doi:  
10.1016/j.jbusres.2012.02.030

- Sarkadi, A., Kristiansson, R., Oberklaid, F., & Bremberg, S. (2008). Fathers' involvement and children's developmental outcomes: a systematic review of longitudinal studies. *Acta Pædiatrica*, *97*(2), 153-158.
- Sawyer, M. G., Sarris, A., Baghurst, P. A., Cross, D. G., & Kalucy, R. S. (1988). Family Assessment Device: Reports from mothers, fathers, and adolescents in community and clinic families. *Journal of Marital and Family Therapy*, *14*(3), 287-296. doi: 10.1111/j.1752-0606.1988.tb00748.x
- Schaffer, H. R., & Emerson, P. E. (1964). The development of social attachments in infancy. *Monographs of the Society for Research in Child Development*, *29*(3), 1-77. doi: 10.2307/1165727
- Schiborr, J., Lotzin, A., Romer, G., Schulte-Markwort, M., & Ramsauer, B. (2013). Child-focused maternal mentalization: A systematic review of measurement tools from birth to three. *Measurement*, *46*(8), 2492–2509.
- Schoppe-Sullivan, S. J., Diener, M. L., Mangelsdorf, S. C., Brown, G. L., McHale, J. L., & Frosch, C. A. (2006). Attachment and sensitivity in family context: the roles of parent and infant gender. *Infant and Child Development*, *15*(4), 367-385. doi: 10.1002/icd.449
- Shai, D., & Belsky, J. (2011). When words just won't do: Introducing parental embodied mentalizing. *Child Development Perspectives*, no-no.
- Shapiro, S. L. (2009). The integration of mindfulness and psychology. *Journal of Clinical Psychology*, *65*(6), 555-560.
- Sharp, C., & Fonagy, P. (2008). The parent's capacity to treat the child as a psychological agent: Constructs, measures and implications for developmental psychopathology. *Social Development*, *17*(3), 737-754.
- Shaver, P. R., & Mikulincer, M. (2002). Attachment-related psychodynamics. *Attachment & Human Development*, *4*(2), 133-161.
- Shea, T., Tennant, A., & Pallant, J. (2009). Rasch model analysis of the Depression, Anxiety and Stress Scales (DASS). *BMC Psychiatry*, *9*(1), 21.
- Shonkoff, J. P. (2010). Building a new biodevelopmental framework to guide the future of early childhood policy. *Child Development*, *81*(1), 357-367.
- Shonkoff, J. P., Boyce, W. T., & McEwen, B. S. (2009). Neuroscience, molecular biology, and the childhood roots of health disparities: Building a new framework for health promotion and disease prevention. *JAMA*, *301*(21), 2252-2259. doi: 10.1001/jama.2009.754

- Shonkoff, J. P., Garner, A. S., Committee on Psychosocial Aspects of Child and Family Health, Committee on Early Childhood Adoption and Dependent Care, Section on Developmental Behavioral Pediatrics, Siegel, B. S., . . . Wood, D. L. (2012). The lifelong effects of early childhood adversity and toxic stress. *Pediatrics*, *129*(1), e232-e246. doi: 10.1542/peds.2011-2663
- Shonkoff, J. P., & Phillips, D. A. (2000). *From neurons to neighbourhoods: the science of early childhood development*. Washington, D.C.: National Academy Press.
- Siegert, R. J., Tennant, A., & Turner-Stokes, L. (2010). Rasch analysis of the Beck Depression Inventory-II in a neurological rehabilitation sample. *Disability and Rehabilitation*, *32*(1), 8-17. doi: doi:10.3109/09638280902971398
- Sigafoos, A. N. N., Reiss, D., Rich, J., & Douglas, E. (1985). Pragmatics in the measurement of family functioning: An interpretive framework for methodology. *Family Process*, *24*(2), 189-203. doi: 10.1111/j.1545-5300.1985.00189.x
- Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107-120.
- Simonsen, S., Nørgaard, N. L., Larsen, K. A., & Bjørnholm, K. I. (2011). Mentalization-inspired case formulation: Minding the difficult patient. *Personality and Mental Health*, *5*(1), 73-79. doi: 10.1002/pmh.154
- Skarderund, F., & Fonagy, P. (2012). Eating disorders. In A. Bateman & P. Fonagy (Eds.), *Handbook of mentalizing in mental health practice* (pp. 347-383). Arlington, VA: American Psychiatric Publishing.
- Slade, A. (2005). Parental reflective functioning: An introduction. *Attachment and Human Development*, *7*, 269-281.
- Slade, A. (2006). Reflective parenting programs: Theory and development. *Psychoanalytic Inquiry*, *26*(4), 640-647.
- Slade, A., Aber, J., Berger, B., Bresgi, I., & Kaplan, M. (2003). *Parent Development Interview Revised Short Version (PDI-R2-S)*. New York.
- Slade, A., Bernbach, E., Grienberger, J., Levy, D., & Locker, A. (2005). *Addendum to the reflective functioning scoring manual for use with the parent development interview*. Unpublished manual. City University of New York. New York.

- Slade, A., Grienenberger, J., Bernbach, E., Levy, D., & Locker, A. (2005). Maternal reflective functioning, attachment, and the transmission gap: A preliminary study. *Attachment and Human Development, 7*(3), 283-298.
- Slade, A., Sadler, L. S., de Dios-Kenn, C., Webb, D., Ezepchick, J., & Mayes, L. C. (2006). Minding the Baby: a reflective parenting program. *Psychoanalytic study of the child, 60*, 26.
- Sleed, M., Baradon, T., & Fonagy, P. (2013). New Beginnings for mothers and babies in prison: A cluster randomized controlled trial. *Attachment & Human Development, 1*-19. doi: 10.1080/14616734.2013.782651
- Smith, E. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement, 3*(2), 205-231.
- Spearman, C. (1904). The proof and measurement of association between two things. *The American Journal of Psychology, 15*(1), 72-101.
- Spicer, P. (2007). Commentary: From fathering to parenting and back again. *Applied Developmental Science, 11*(4), 203-204. doi: 10.1080/10888690701762084
- Spielberger, C. D. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y)*. Palo Alto, CA: Mind Garden.
- Sprinkle, S. D., Lurie, D., Insko, S. L., Atkinson, G., Jones, G. L., Logan, A. R., & Bissada, N. N. (2002). Criterion validity, severity cut scores, and test-retest reliability of the Beck Depression Inventory-II in a university counseling center sample. *Journal of Counseling Psychology, 49*(3), 381.
- Sroufe, L. A. (2005). Attachment and development: A prospective, longitudinal study from birth to adulthood. *Attachment & Human Development, 7*(4), 349-367. doi: 10.1080/14616730500365928
- Steele, H., & Steele, M. (2008). On the origins of reflective functioning. In F. Busch, N (Ed.), *Mentalization: Theoretical considerations, research findings, and clinical implications* (pp. 133-158). New York: The Analytic Press.
- Steele, H., Steele, M., & Fonagy, P. (1996). Associations among attachment classifications of mothers, fathers, and their infants. *Child Development, 67*(2), 541-555.
- Stein, H., & Allen, J. G. (2007). Mentalizing as a framework for integrating therapeutic exposure and relationship repair in the treatment of a patient with

- complex posttraumatic psychopathology. *Bulletin of the Menninger Clinic*, 71(4), 273-290.
- Stevens, S. S. (1946). On the theory of scales of measurement. *Science*, 103(2684), 677-680.
- Stevenson-Hinde, J., & Akister, J. (1995). The McMaster model of family functioning: Observer and parental ratings in a nonclinical sample. *Family Process*, 34(3), 337-347. doi: 10.1111/j.1545-5300.1995.00337.x
- Stevenson-Hinde, J., Curley, J. P., Chicot, R., & Jóhannsson, C. (2007). Anxiety within families: Interrelations, consistency, and change. *Family Process*, 46(4), 543-556. doi: 10.1111/j.1545-5300.2007.00231.x
- Storch, E. A., Roberti, J. W., & Roth, D. A. (2004). Factor structure, concurrent validity, and internal consistency of the Beck Depression Inventory—Second Edition in a sample of college students. *Depression and Anxiety*, 19(3), 187-189.
- Suchman, N. E., DeCoste, C., Castiglioni, N., McMahon, T. J., Rounsaville, B., & Mayes, L. (2010). The Mothers and Toddlers Program, an attachment-based parenting intervention for substance using women: Post-treatment results from a randomized clinical pilot. *Attachment & Human Development*, 12(5), 483-504. doi: 10.1080/14616734.2010.501983
- Suchman, N. E., DeCoste, C., Leigh, D., & Borelli, J. (2010). Reflective functioning in mothers with drug use disorders: Implications for dyadic interactions with infants and toddlers. *Attachment & Human Development*, 12(6), 567-585. doi: 10.1080/14616734.2010.501988
- Suchman, N. E., Pajulo, M., Kalland, M., DeCoste, C., & Mayes, L. (2012). At-risk mothers of infants and toddlers. In A. Bateman & P. Fonagy (Eds.), *Handbook of mentalizing in mental health practice* (pp. 309-346). Arlington, VA: American Psychiatric Publishing.
- Szyf, M., & Bick, J. (2013). DNA methylation: A mechanism for embedding early life experiences in the genome. *Child Development*, 84(1), 49-57. doi: 10.1111/j.1467-8624.2012.01793.x
- Tatum, D. S. (2000). Rasch analysis: An introduction to objective measurement. *Laboratory Medicine*, 31(5), 272-274.

- Tenenbaum, G., Furst, D., & Weingarten, G. (1985). A statistical reevaluation of the STAI anxiety questionnaire. *Journal of Clinical Psychology, 41*(2), 239-244. doi: 10.1002/1097-4679(198503)41:2<239::aid-jclp2270410218>3.0.co;2-5
- Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it? Why use it? When should it be applied and what should one look for in a Rasch paper? . *Arthritis & Rheumatism, 57*(8), 1358-1362.
- Tennant, A., & Pallant, J. (2006). Unidimensionality matters! (A tale of two Smiths?). *Rasch Measurement Transactions, 20*(1), 1048-1051.
- The National Scientific Council on the Developing Child. (2013). Reports & working papers. from [http://developingchild.harvard.edu/index.php/resources/reports\\_and\\_working\\_papers/](http://developingchild.harvard.edu/index.php/resources/reports_and_working_papers/)
- Thompson, R. A. (2008). Early attachment and later development: Familiar questions, new answers. In J. Cassidy & P. R. Shaver (Eds.), *Handbook of attachment* (2nd ed., pp. 348-365). New York: Guilford.
- Thompson, R. A. (2010). Attachment and life history theory: A rejoinder. *Child Development Perspectives, 4*(2), 106-108.
- Thurstone, L. L. (1928). Attitudes can be measured. *American Journal of Sociology, 33*(4), 529-554.
- Thurstone, L. L. (1929). The measurement of psychological value. In T. V. Smith & W. Kelley (Eds.), *Essays in philosophy by seventeen doctors of philosophy of the University of Chicago* (pp. 157-174). Chicago: Open Court. Retrieved from [http://www.brocku.ca/MeadProject/Thurstone/Thurstone\\_1929a.html](http://www.brocku.ca/MeadProject/Thurstone/Thurstone_1929a.html).
- Ting-Fang, W. (2007). *Executive Summary*. Understanding the brain: The birth of a learning science. OECD. Retrieved from <http://www.oecd.org/dataoecd/42/33/38811529.pdf>
- van der Linden, W. J. (1994). Fundamental measurement and the fundamentals of Rasch measurement. In M. Wilson (Ed.), *Objective measurement: Theory into practice* (Vol. 2, pp. 3-24). Norwood, New Jersey: Ablex.
- van IJzendoorn, M., H. (1995a). Adult attachment representations, parental responsiveness, and infant attachment: A meta-analysis on the predictive

- validity of the Adult Attachment Interview. *Psychological Bulletin*, 117(3), 387-403.
- van IJzendoorn, M., H. (1995b). Of the way we are: On temperament, attachment, and the transmission gap: A rejoinder to Fox (1995). *Psychological Bulletin*, 117(3), 411-415.
- van IJzendoorn, M., H., & Bakermans-Kranenburg, M. J. (1996). Attachment representations in mothers, fathers, adolescents, and clinical groups: A meta-analytic search for normative data. *Journal of Consulting and Clinical Psychology*, 64(1), 8.
- van IJzendoorn, M., H., & Bakermans-Kranenburg, M. J. (2010). Stretched until it snaps: Attachment and close relationships. *Child Development Perspectives*, 4(2), 109-111.
- van IJzendoorn, M., H., Bakermans-Kranenburg, M. J., & Ebstein, R. P. (2011). Methylation matters in child development: Toward developmental behavioral epigenetics. *Child Development Perspectives*, 5(4), 305-310. doi: 10.1111/j.1750-8606.2011.00202.x
- van IJzendoorn, M., H., & De Wolff, M., S. (1997). In Search of the Absent Father - Meta-analyses of infant-father attachment: A rejoinder to our discussants. *Child Development*, 68(4), 604-609.
- Vandenberg, S. (1972). Assortative mating, or who marries whom? *Behavior Genetics*, 2(2-3), 127-157. doi: 10.1007/BF01065686
- Vanheule, S., Verhaeghe, P., & Desmet, M. (2011). In search of a framework for the treatment of alexithymia. *Psychology and Psychotherapy: Theory, Research and Practice*, 84(1), 84-97. doi: 10.1348/147608310x520139
- Verissimo, M., Santos, A. J., Vaughn, B. E., Torres, N., Monteiro, L., & Santos, O. (2011). Quality of attachment to father and mother and number of reciprocal friends. *Early Child Development and Care*, 181(1), 27 - 38.
- Verschueren, K., & Marcoen, A. (1999). Representation of self and socioemotional competence in kindergartners: Differential and combined effects of attachment to mother and to father. *Child Development*, 70(1), 183-201.
- von Sydow, K. (1999). Sexuality during pregnancy and after childbirth: a metacontent analysis of 59 studies. *Journal of Psychosomatic Research*, 47(1), 27-49.

- Waters, E. (1995). Appendix A: The Attachment Q-Set (Version 3.0). *Monographs of the Society for Research in Child Development*, 60(2/3), 234-246. doi: 10.2307/1166181
- Waters, E., & Deane, K. E. (1985). Defining and assessing individual differences in attachment relationships: Q-methodology and the organization of behavior in infancy and early childhood. *Monographs of the Society for Research in Child Development*, 50(1/2), 41-65. doi: 10.2307/3333826
- Whisman, M. A., Perez, J. E., & Ramel, W. (2000). Factor structure of the Beck Depression Inventory—Second Edition (BDI-II) in a student sample. *Journal of Clinical Psychology*.
- Wilson, M. (Ed.). (1994). *Objective measurement: Theory into practice* (Vol. 2). Norwood, New Jersey: Ablex.
- Winnicott, D. W. (1971). *Playing and reality*. London: Tavistock Publications.
- Wolfe, E. W., & Smith, E. (2007a). Instrument development tools and activities for measure validation using Rasch models: Part 1 - Instrument development tools. *Journal of Applied Measurement*, 8(1), 97-123.
- Wolfe, E. W., & Smith, E. (2007b). Instrument development tools and activities for measure validation using Rasch models: Part 2 - Validation activities. *Journal of Applied Measurement*, 8(2), 204-234.
- Woods, B. (2011). *What's still wrong with psychology, anyway? Twenty slow years, three old issues, and one new methodology for improving psychological research*. (Master of Arts in Psychology Thesis), University of Canterbury.
- World Health Organisation. (2007). *Fatherhood and health outcomes in Europe*  
Retrieved from  
<http://apps.who.int/iris/bitstream/10665/108571/1/E91129.pdf>
- Wright, B. D. (1985). Additivity in psychological measurement. *Measurement and Personality Assessment*, 8, 101-111.
- Wright, B. D. (1997). A history of social science measurement. *Educational Measurement: Issues and Practice*, 16(4), 33-45. doi: 10.1111/j.1745-3992.1997.tb00606.x
- Wright, B. D. (1998). Georg Rasch: The man behind the model - The mathematician. *Popular Measurement, Spring*, 15-22.
- Wright, B. D. (2000). Rasch analysis for surveys. *Popular Measurement*, 61(Spring), 36.

Wright, B. D., & Stenner, J., A. (1999). One fish, two fish: Rasch measures reading best. *Popular Measurement*, 2(1), 34-38.

Wright, B. D., & Stone, M. H. (1979). *Best test design: Rasch measurement*. Chicago: Mesa Press.

Wright, B. D., & Stone, M. H. (1999). *Measurement essentials* (2 ed.). Wilmington, Delaware: Wide Range Inc.

Every reasonable effort has been made to acknowledge the owners of copyright material. I would be pleased to hear from any copyright owner who has been omitted or incorrectly acknowledged.

## Appendices

### Appendix A – PRFQ

#### Parental Reflective Functioning Questionnaire

Rgcu'eqpcev'vj g'c'wvj qt u'ht 'c'eqr { 'qh'vj k's wgm'ppct g

Luyten, P., Mayes, L., C, Sadler, L. S., Fonagy, P., Nicholls, S., Crowley, M. J., . . .  
Slade, A. (2009). Manual for the Parental Reflective Functioning Questionnaire-1  
(PRFQ-1). Unpublished Manual. Yale Child Study Center. New Haven, CT.

This document has not been included in the published version of this thesis

## **Appendix B - Peel Child Health Information Sheet and Consent Forms**

### **Participant Information Letter**

#### **Invitation**

You are being asked to be part of an important research study to see how newborns born to families in the Peel Region respond to their environment and grow into healthy children. We are speaking with hundreds of families in the Peel Region to help us understand how health, education and other community services can do the best job possible in supporting children's health and development. All the information collected will be kept strictly confidential. Your information will only be seen by the research team, and published findings will not identify you.

#### **Benefits of the Study**

While we cannot guarantee that you will personally benefit, the knowledge gained from your participation will inform future policies and services to improve the health and development of your child/ren and it may help others in the future.

#### **Who is conducting the study?**

The research team members are from Murdoch University, the Telethon Institute for Child Health Research (TICHR) the University of Western Australia, and Curtin and Edith Cowan Universities. Each of us also has a network of national and international researchers who support this work and will provide additional expertise if this is required as we analyse this information. The research is led by Professor Anne McMurray of Murdoch University, who will be available via phone or email to provide any information about the study as required. The other research leader is Professor Fiona Stanley, from TICHR. Our study manager, Ms Martinique Sandy, will also be available locally to ensure that any queries are answered promptly and sensitively. The manager will monitor an email address to respond to any queries as they arise by telephone (95825559), or electronically ([childhealthstudy@murdoch.edu.au](mailto:childhealthstudy@murdoch.edu.au)).

#### **What will I be asked to do?**

We would like your permission to contact the medical practitioner who is assisting your pregnancy (GP or obstetrician), so that we can have access to your pregnancy ultrasound scans and any pregnancy related tests (such as blood tests), and to request additional ultrasound scans at 26 and 34 weeks. This will help us assess your child's growth prior to birth. We would also like to visit your home during the pregnancy to examine a sample of your blood and saliva, and that of your partner. During that visit we'll ask you a few questions about your home and the products you use, and sample your household dust and air to try to link any toxins in the home with your child's health and growth.

The blood and saliva will be used for hormone analysis to give us an indication of the level of stress experienced by you, your partner and your baby in the womb. Blood taken early in the pregnancy from you and your partner and umbilical cord blood taken at birth will also be used to map the DNA of you, your partner and your child, which will help us understand the role played by genetics in children's development. In no case will the DNA tests be used for any other purposes. The blood sample from the cord will help us study your child's immune function. If you agree, we would also like to collect a sample of the placenta to analyse its function during your pregnancy. After birth, when your child is aged 1, 2, and 3 we would like to check your child's blood pressure, heart rate and skinfold measurements, repeat the blood and saliva tests, and check urine and blood samples from your child to see how your child is growing and developing. Once your child has been immunised we will also analyse the blood

for vaccine responses. With your permission we would like to link the information we gather on your child with Australian Bureau of Statistics and Medicare data and the unique W.A. database which includes information from medical and hospital admissions, education, child protection and the criminal justice system.

***Please note that you may choose to participate in all or only some of these tests***

Before the home visit we will drop off a questionnaire for you to complete when you have time. Questions are about you and your family. They ask about your pregnancy, general health and well-being, work-life balance, housing, cultural influences and social support. We would also like you to complete a questionnaire about your new baby and one for each of your other children aged up to five years. These questions ask how your child is feeding, sleeping and communicating, and their physical health and vaccination status. In addition, we are interested in your views on parenting and use of child care (see flow chart). Our research assistant will pick up your questionnaires once they're completed and answer any questions you may have.

We will also be studying your community to identify the supports or risks that influence your ability to make various parenting decisions. Questions will ask about your access to health services, schools and child care facilities, the links between home and school, any issues of transportation or safety, and other features of community life that affect child rearing. If you wish, we will invite you to participate in either an individual interview or a group discussion to voice your opinions on community issues, on a strictly voluntary basis.

---

## Flow Chart

1. We will enrol you in the study either in person at Murdoch University (Peel Campus) or in your home. The researcher will explain the study, and gain your written consent to participate. We will then contact your medical practitioner (via phone) to access your record of ultrasounds and/or any other pregnancy test. Then we will schedule further ultrasounds if necessary, and leave the questionnaires to be completed when you have time. After signing the consent form, we will ask you to complete the following tasks:
    2. Complete the printed questionnaire on your health, family background, parenting experiences and community life at five different times: during pregnancy, at birth, and when your child is age 1, 2, and 3. A separate questionnaire will also ask you for information on any of your other children up to age 5 and the father of the newborn.
    3. Provide the following samples at 18, 26 and 34 weeks of pregnancy:
 

a. Blood	mother/father
b. Saliva (3 time points during the day)	mother
c. Urine	mother
c. House dust, air (18 weeks of pregnancy)	in family home
d. Pregnancy Health Record & Ultrasound (reports from medical doctor: dating scan, 10 week, 18 weeks, new ultrasounds at 26, 34 weeks)	
    4. Provide the following samples at birth:
 

a. Umbilical Cord blood	child
b. Placenta tissue	mother
    5. Provide the following samples when the child is aged 1, 2, and 3:
 

a. Blood	child
b. Saliva (5 time points during the day)	child
c. Urine	child
d. House dust, air	family home
-

### Explanation of sample collection

#### At the birth

**Ultrasound:** A professional radiographer will take the ultrasound scans. They will take around one hour and be done at your local radiography clinic at no cost to you.

**Placenta:** The placenta will be weighed and measured following the birth, and a sample of placental tissue will be collected.

**Umbilical Cord Blood:** Immediately following the birth, a midwife or a phlebotomist (registered blood collector) will collect a blood sample from the cord.

#### Home Visit: Once in Pregnancy and after the birth and when your child is 1, 2, and 3

**Fasting Blood:** You, your partner and your child will be asked to provide a blood sample to a registered phlebotomist, after abstaining from food for one hour prior to the blood collection. The phlebotomist will come to your home and draw 20 ml of blood intravenously from the inside of the elbow (parents and child), staying with you long enough to make sure there is no blood leakage from the vein.

**Saliva:** One week before we collect saliva samples, collection tubes and cotton swabs will be dropped off for you to take samples from yourself and your child. A special information sheet will help you with this. Collection of saliva will involve chewing on a cotton swab for 1-3 minutes, then placing the swabs in a special collection tube and keeping them in the fridge to be picked up by the research assistant.

**Urine:** A urine container will also be provided one week before the home visit. We'd like you to collect the sample, keeping it in the refrigerator until it is collected by the research assistant that day.

**Dust:** When the research assistant comes to your home, (s)he will vacuum up a small amount of house dust from the floor or carpet and, where possible, collect an air sample.

### What will we do with this information?

Enrolment in the study will require both you and your partner (if he is available) to sign the consent forms attached. All information you provide will be kept strictly confidential by giving you and your child a coded number so that you are not identifiable. No other researchers will have access to your information without your written consent, and no material or information related to you or your child will be released without your consent unless required by law. If the research reveals information of potential importance to the health of your child or your family we will contact you to help you receive medical advice.

During the study all information will be kept in a locked cabinet at Murdoch University. The blood, saliva, urine, dust and air samples will be kept in an appropriate storage facility at Murdoch until they are analysed. Instead of destroying the remaining blood samples we will keep these in the storage facility in case we are able to add a further study to follow your child's development into adolescence.

At various times throughout the study, we will provide you with plain language summaries of what we are finding, maintaining your privacy by reporting only group information. We will also have group feedback sessions for members of your community on the views of parents and the issues that should be discussed at community or government levels. We will also be conducting an annual seminar in the Peel Region by eminent child health researchers to provide the most current research knowledge on child health and development back to the local community. If at any time you wish to speak to a person not involved in the study who can provide further information you can contact Murdoch University's Human Research Ethics Committee on 9360 6677 or email [ethics@murdoch.edu.au](mailto:ethics@murdoch.edu.au).

### **Voluntary Participation and Withdrawal from the Study**

Your participation in this study is entirely voluntary. You may withdraw at any time without discrimination or prejudice and without providing a reason. All information will be confidential and no names or other details that might identify you will be used in any publication arising from the research. If you choose to withdraw from the study, you may also choose to have all information about you related to previous follow-ups and the current follow-up destroyed.

If you consent to take part in this research study, it is important that you understand the purpose of the study and the procedures you will be asked to undergo. Please make sure that you ask any questions you may have, and that all your questions have been answered to your satisfaction before you agree to participate in each aspect of the study.

### **Risks of the Study**

Every effort will be made to eliminate any risks to you, your child and other family members. However, we feel it is important to explain that there is always some degree of risk in collecting blood and tissue samples. If you have a bleeding disorder or if your medical practitioner believes it is inappropriate for you to participate, please accept that advice and we thank you for considering our study. Regardless of your decision you are welcome to keep this information sheet. In making a decision to participate in this study you are welcome to involve any other person(s) for whom the research is relevant. We will make every effort to ensure your child's best interest and to provide for his/her safety, emotional and psychological security and wellbeing. If you consent to involve your children other than the birth child we will discuss with you any implications for your child based on his/her developmental level and capacity.

Although unlikely, it is possible at some time during this study that you may experience some anxiety or stress as a result of some of the tasks and we will make every effort to minimise this risk by monitoring your condition. You will be able to withdraw at any time, and if feelings of anxiety or stress persist after the completion of the session, arrangements will be made for you to access support from the counselling support services the Peel Community Mental Health Service, Lakes Rd., Mandurah, 95318080 or the local branch of Relationships Australia, 7 Anzac Parade, Mandurah, 95355711. There will be no cost to you for these services. If you would like private counselling you will be guided to the Child Health Service Directory developed as a preliminary step to this project, which lists all counselling, psychiatric and child health services on the website [www.peelhealth.com.au](http://www.peelhealth.com.au).

If you are willing to consent to participation in this study, please **complete the Consent Form**. If you have any questions about this project please feel free to contact either myself, Anne McMurray on ph. 0409587850, or the study manager, Martinique Sandy on 9582 5559 or [childhealthstudy@murdoch.edu.au](mailto:childhealthstudy@murdoch.edu.au)

I will be pleased to discuss with you any concerns you may have on how this study has been conducted. Alternatively you can contact Murdoch University's Human Research Ethics Committee on 9360 6677 or email [ethics@murdoch.edu.au](mailto:ethics@murdoch.edu.au).

This study has been approved by the Murdoch University Human Research Ethics Committee (Approval No. 2007/238)

## Our Children our Families our Place

### Participant Consent Form – Primary Care Giver (Mother)

I have read the participant information sheet, which explains the purpose and nature of the research and the possible risks. Any questions asked have been answered to my satisfaction and I have been given a copy of the information sheet to keep.

I agree that the research findings from the study may be published provided my name or any identifying data is not used. I have also been informed that I may not receive any direct benefits from participating in this study.

I understand that all information provided by me is treated as confidential and will not be released by the researcher unless required to do so by law.

I am happy to be interviewed and complete the questionnaire as part of this research. I understand that I do not have to answer particular questions or participate in some of the medical tests if I do not want to and that I can withdraw at any time without consequences to myself.

I agree to participate in the following parts of the study (please tick as appropriate)

	Yes	No
<b>At Birth:</b> Sample of cord blood and placental tissue for DNA and biochemical analysis	<input type="radio"/>	<input type="radio"/>
Blood sample at 18, 26 and 34 weeks pregnant for DNA and biochemical analysis	<input type="radio"/>	<input type="radio"/>
Urine sample at 18, 26 and 34 weeks pregnant for biochemical analysis	<input type="radio"/>	<input type="radio"/>
Abdominal ultrasound at 26 and 34 weeks pregnant	<input type="radio"/>	<input type="radio"/>
Saliva sample at 18, 26 and 34 weeks pregnant for biochemical analysis	<input type="radio"/>	<input type="radio"/>
<b>At Home:</b> Home inventory, dust, air sample – 18 wks pregnancy, when child is age 1, 2, and 3	<input type="radio"/>	<input type="radio"/>
<b>Linked information:</b>		
W.A. Births, medical, hospital records, ABS, Medicare data	<input type="radio"/>	<input type="radio"/>
Educational records	<input type="radio"/>	<input type="radio"/>
Child Protection records	<input type="radio"/>	<input type="radio"/>
Criminal justice records	<input type="radio"/>	<input type="radio"/>
Health record on pregnancy from GP	<input type="radio"/>	<input type="radio"/>

Dated \_\_\_\_\_ day of \_\_\_\_\_ 20 \_\_\_\_\_

Signed \_\_\_\_\_ (Parent)

**Chief Investigator**

I have fully explained to \_\_\_\_\_ the nature and purpose of the research, the procedures to be employed, and the possible risks involved. I have provided the participant with a copy of the Information Sheet.

\_\_\_\_\_  
Signature of Investigator

\_\_\_\_\_  
Date

\_\_\_\_\_  
Print Name

\_\_\_\_\_  
Position

## Our Children our Families our Place

### Participant Consent Form - Father

I have read the participant information sheet, which explains the purpose and nature of the research and the possible risks. Any questions asked have been answered to my satisfaction and I have been given a copy of the information sheet to keep.

I agree that the research findings from the study may be published provided my name or any identifying data is not used. I have also been informed that I may not receive any direct benefits from participating in this study.

I understand that all information provided by me is treated as confidential and will not be released by the researcher unless required to do so by law.

I am happy to be interviewed and complete the questionnaire as part of this research. I understand that I do not have to answer particular questions or participate in some of the medical tests if I do not want to and that I can withdraw at any time without consequences to myself.

I agree to participate in the following parts of the study (please tick as appropriate)

	Yes	No
<b>At 18 wks of pregnancy</b>		
Blood sample for DNA and biochemical analysis	<input type="radio"/>	<input type="radio"/>
<b>At birth:</b> I have no objection to my partner providing a sample of cord blood and placental tissue for DNA and biochemical analysis	<input type="radio"/>	<input type="radio"/>
<b>Linked information:</b>		
W.A. Births, medical, hospital records, ABS, Medicare data	<input type="radio"/>	<input type="radio"/>
Educational records	<input type="radio"/>	<input type="radio"/>
Child Protection records	<input type="radio"/>	<input type="radio"/>
Criminal justice records	<input type="radio"/>	<input type="radio"/>
Health record on pregnancy from GP	<input type="radio"/>	<input type="radio"/>

Dated \_\_\_\_\_ day of \_\_\_\_\_ 20 \_\_\_\_\_

Signed \_\_\_\_\_ (Parent)

**Chief Investigator**

I have fully explained to \_\_\_\_\_ the nature and purpose of the research, the procedures to be employed, and the possible risks involved. I have provided the participant with a copy of the Information Sheet.

\_\_\_\_\_  
Signature of Investigator

\_\_\_\_\_  
Date

\_\_\_\_\_  
Print Name

\_\_\_\_\_  
Position

**Our Children our Families our Place**

**Participant Consent Form – Birth Child**

I have read the participant information sheet, which explains the purpose and nature of the research and the possible risks. Any questions asked have been answered to my satisfaction and I have been given a copy of the information sheet to keep.

I agree that the research findings from the study may be published provided my name or any identifying data is not used. I have also been informed that I may not receive any direct benefits from participating in this study.

I understand that all information provided by me is treated as confidential and will not be released by the researcher unless required to do so by law.

I am happy to be interviewed and complete the questionnaire as part of this research. I understand that I do not have to answer particular questions or participate in some of the medical tests if I do not want to and that I can withdraw at any time without consequences to myself.

I agree to my child \_\_\_\_\_ participating in the following parts of the study (please tick as appropriate)

	Yes	No
<b><u>Birth Child at 1, 2, 3 years of age:</u></b>		
physical examination	<input type="radio"/>	<input type="radio"/>
measurement of cardiovascular function and body composition	<input type="radio"/>	<input type="radio"/>
collection of blood, saliva and urine for biochemical analysis at age 1, 2, and 3	<input type="radio"/>	<input type="radio"/>
developmental screening tests	<input type="radio"/>	<input type="radio"/>
<b>Linked information:</b>		
W.A. Births, medical, hospital records, ABS, Medicare data		
Educational records	<input type="radio"/>	<input type="radio"/>
Child Protection records	<input type="radio"/>	<input type="radio"/>
Criminal justice records	<input type="radio"/>	<input type="radio"/>

Dated \_\_\_\_\_ day of \_\_\_\_\_ 20 \_\_\_\_\_

Signed \_\_\_\_\_ (Parent/Guardian)

Signed \_\_\_\_\_ (Parent/Guardian)

**Chief Investigator**

I have fully explained to \_\_\_\_\_ the nature and purpose of the research, the procedures to be employed, and the possible risks involved. I have provided the participant with a copy of the Information Sheet.

\_\_\_\_\_  
Signature of Investigator

\_\_\_\_\_  
Date

## Our Children our Families our Place

### Participant Consent Form – Sibling

I have read the participant information sheet, which explains the purpose and nature of the research and the possible risks. Any questions asked have been answered to my satisfaction and I have been given a copy of the information sheet to keep.

I agree that the research findings from the study may be published provided my name or any identifying data is not used. I have also been informed that I may not receive any direct benefits from participating in this study.

I understand that all information provided by me is treated as confidential and will not be released by the researcher unless required to do so by law.

I am happy to be interviewed and complete the questionnaire as part of this research. I understand that I do not have to answer particular questions or participate in some of the medical tests if I do not want to and that I can withdraw at any time without consequences to myself.

I agree to my child \_\_\_\_\_ participating in the following parts of the study (please tick as appropriate)

	Yes	No
<b>Sibling</b>	<input type="radio"/>	<input type="radio"/>
Questionnaire data		
<b>Linked information:</b>		
W.A. Births, medical, hospital records, ABS, Medicare data	<input type="radio"/>	<input type="radio"/>
Educational records	<input type="radio"/>	<input type="radio"/>
Child Protection records	<input type="radio"/>	<input type="radio"/>
Criminal justice records	<input type="radio"/>	<input type="radio"/>

Dated \_\_\_\_\_ day of \_\_\_\_\_ 20 \_\_\_\_\_

Signed \_\_\_\_\_ (Parent/Guardian)

Signed \_\_\_\_\_ (Parent/Guardian)

**Chief Investigator**

I have fully explained to \_\_\_\_\_ the nature and purpose of the research, the procedures to be employed, and the possible risks involved. I have provided the participant with a copy of the Information Sheet.

\_\_\_\_\_  
Signature of Investigator

\_\_\_\_\_  
Date

## Appendix C – BDI-II

## Beck Depression Inventory II (from PCHS questionnaire)

**C39. This questionnaire consists of 21 groups of statements. Please read each group of statements carefully, and then pick out one statement in each group that best describes the way you have been feeling during the past two weeks, including today. Circle the number for the statement you have picked. If several statements in the group seem to apply equally well, circle the highest number for that group. Be sure that you do not choose more than one statement for any group, including items 16 and 18.**

**1. Sadness**

- 0 I do not feel sad.
- 1 I feel sad much of the time.
- 2 I am sad all the time
- 3 I am so sad or unhappy that I can't stand it.

**8. Self-Criticalness**

- 0 I don't criticise or blame myself more than usual.
- 1 I am more critical of myself than I used to be.
- 2 I criticise myself for all my faults.
- 3 I blame myself for everything bad that happens.

**2. Pessimism**

- 0 I am not discouraged about my future.
- 1 I feel more discouraged about my future than I used to be.
- 2 I do not expect things to work out for me.
- 3 I feel my future is hopeless and will only get worse.

**9. Suicidal Thoughts or Wishes**

- 0 I don't have any thoughts of killing myself.
- 1 I have thoughts of killing myself, but I would not carry them out.
- 2 I would like to kill myself.
- 3 I would kill myself if I had the chance.

**3. Past Failure**

- 0 I do not feel like a failure.
- 1 I have failed more than I should have.
- 2 As I look back, I see a lot of failures.
- 3 I feel I am a total failure as a person.

**10. Crying**

- 0 I don't cry anymore than I used to.
- 1 I cry more than I used to.
- 2 I cry over every little thing.
- 3 I feel like crying, but I can't.

**4. Loss of Pleasure**

- 0 I get as much satisfaction out of things as I used to.
- 1 I don't enjoy things the way I used to.
- 2 I don't get any real satisfaction out of anything anymore.
- 3 I am dissatisfied or bored with everything.

**11. Agitation**

- 0 I am no more restless or wound up than usual.
- 1 I am more restless or wound up than usual.
- 2 I am so restless or agitated that it's hard to stay still.
- 3 I am so restless or agitated that I have to keep moving or doing something.

**5. Guilty Feelings**

- 0 I don't feel particularly guilty.
- 1 I feel guilty over many things I have done or should have done.
- 2 I feel quite guilty most of the time.
- 3 I feel guilty all of the time.

**12. Loss of Interest**

- 0 I have not lost interest in other people or activities.
- 1 I am less interested in other people or things than before.
- 2 I have lost most of my interest in other people or things.
- 3 It's hard to get interested in anything.

**6. Punishment Feelings**

- 0 I don't feel I am being punished.
- 1 I feel I may be punished.
- 2 I expect to be punished.
- 3 I feel I am being punished.

**13. Indecisiveness**

- 0 I make decisions about as well as ever.
- 1 I find it more difficult to make decisions than usual.
- 2 I have much greater difficulty in making decisions than I used to.
- 3 I have trouble making any decisions.

**7. Self-Dislike**

- 0 I feel the same about myself as ever.
- 1 I have lost confidence in myself.
- 2 I am disappointed in myself.
- 3 I dislike myself.

**14. Worthlessness**

- 0 I don't feel I am worthless.
- 1 I don't consider myself as worthwhile and useful as I used to.
- 2 I feel more worthless as compared to other people.
- 3 I feel utterly worthless.

## New Page

**15. Loss of Energy**

- 0 I have as much energy as ever..
- 1 I have less energy than I used to have.
- 2 I don't have enough energy to do very much.
- 3 I don't have enough energy to do anything.

**19. Concentration Difficulty**

- 0 I can concentrate as well as ever.
- 1 I can't concentrate as well as usual.
- 2 It's hard to keep my mind on anything for very long.
- 3 I find I can't concentrate on anything.

**16. Changing in Sleeping Pattern**

- 0 I have not experienced any change in my sleeping pattern.
- 1a I sleep somewhat more than usual.
- 1b I sleep somewhat less than usual.
- 2a I sleep a lot more than usual.
- 2b I sleep a lot less than usual.
- 3a I sleep most of the day.
- 3b I wake up 1-2 hours early and can't get back to sleep.

**20. Tiredness or Fatigue**

- 0 I am no tired or fatigued than usual.
- 1 I get more tired or fatigued more easily than usual.
- 2 I am too tired or fatigued to do a lot of things I used to do.
- 3 I am too tired or fatigued to do most of the things I used to do.

**17. Irritability**

- 0 I am not more irritable than usual.
- 1 I am more irritable than usual.
- 2 I am much more irritable than usual.
- 3 I am irritable all the time.

**21. Loss of Interest in Sex**

- 0 I have not noticed any recent change in my interest in sex.
- 1 I am less interested in sex than I used to be.
- 2 I am much less interested in sex now.
- 3 I have lost interested in sex completely.

**18. Changes in Appetite**

- 0 I have not experienced any change in my appetite.
- 1a My appetite is somewhat less than usual.
- 1b My appetite is somewhat greater than usual.
- 2a My appetite is much less than before.
- 2b My appetite is much greater than usual.
- 3a I have no appetite at all.
- 3b I crave food all the time.

**Appendix D – STAI**

**State-Trait Anxiety Inventory (from PCHS questionnaire)**

**State scale**

**C40. A number of statements which people have used to describe themselves are given below. Read each statement and then mark the appropriate circle to the right of the statement to indicate how you feel *right now*, that is, *at this moment*. There are no right or wrong answers. Do not spend too much time on any one statement but give the answer which seems to describe your present feelings best.**

Not at all	1			
Somewhat		2		
Moderately so			3	
Very much so				4
	1	2	3	4
I feel calm .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel secure .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel tense .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel strained .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel at ease .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel upset .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am presently worrying over possible misfortunes .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel satisfied .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel frightened .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel comfortable .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel self-confident .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel nervous .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am jittery.....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel indecisive .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am relaxed .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel content .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am worried .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel confused .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel steady.....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel pleasant .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Trait scale**

<b>C41. A number of statements which people have used to describe themselves are given below. Read each statement and then mark the appropriate circle to the right of the statement to indicate how you <i>generally feel</i>.</b>				
<b>Almost never</b>	<b>1</b>			
<b>Sometimes</b>		<b>2</b>		
<b>Often</b>			<b>3</b>	
<b>Almost always</b>	<b>4</b>			
	<b>1</b>	<b>2</b>	<b>3</b>	<b>4</b>
I feel pleasant .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel nervous and restless .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel satisfied with myself .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I wish I could be as happy as others seem to be .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel like a failure .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel rested .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am "calm, cool, and collected" .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel that difficulties are piling up so that I cannot overcome them	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I worry too much over something that really doesn't matter .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am happy .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I have disturbing thoughts .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I lack self-confidence .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel secure .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I make decisions easily .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I feel inadequate .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am content .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Some important thought runs through my mind and bothers me .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I take disappointments so keenly that I can't put them out of my mind	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I am a steady person .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I get in a state of tension or turmoil as I think over my recent concerns and interests .....	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

**Appendix E – PDI**

**Parent Development Interview**

PDI-R2-S

PARENT DEVELOPMENT INTERVIEW  
REVISED

*SHORT VERSION*

Arietta Slade, J. Lawrence Aber, Brenda Berger, Ivan Bresgi, Merryle Kaplan

Adapted with the help of:  
Linda Mayes, Mary Target, Sidney Blatt

October, 2003

*PRIVILEGED COMMUNICATION*

This interview is an adaptation of the Parent Development Interview (Aber, Slade, Berger, Bresgi, & Kaplan, 1985). This protocol may not be used or adapted without written permission from Arietta Slade, Ph.D., The Psychological Center, R8/130, The City College of New York, 138<sup>th</sup> Street & Convent Avenue, New York, NY 10031, [asladephd@earthlink.net](mailto:asladephd@earthlink.net) .

"

"

Vj ku'f qewo gpv'j cu'pqv'dggp'kpenw gf 'lp'yj g'r wdrkuj gf 'xgtukp'qh'yj ku'yj guku0

## Appendix F – Participation Information Sheet

Project Title: **Parental reflective functioning of fathers and its relevance for children's health and development: a population-based study**

**You are being contacted because you previously agreed to participate in interviews as part of the Peel Child Health Study. You are being asked to agree to a one-hour, audio-recorded, interview, which will be completed by 80, other mothers and fathers. Your involvement in this interview is completely voluntary and you can withdraw at any time without it affecting your rights or further involvement in other parts of the Study.**

**What is this interview for?**

**Dawson Cooke is undertaking this investigation as part of his PhD in Psychology at Curtin University of Technology. The term “parental reflective functioning” refers to the ways a parent understands the thoughts and feelings of their child. This study will help with understanding more about the influence parents have on the health and development of their children, and specifically the relevance of the father's reflective functioning. In addition to this interview, you will also be asked to complete a two page questionnaire about reflective functioning (The Parental Reflective Functioning Questionnaire) that is included as part of the Peel Study One Year Follow-up.**

**You can meet the interviewer at a time and location that suits you.**

**Alternatively, you can arrange to have the interview via video conferencing over the internet. There are no financial costs associated with participating in this interview. Although we cannot guarantee you will personally benefit from the interview, the knowledge gained from the study will inform future policies and services to improve the health of your child. You can choose to not answer any question and can stop the interview at any time.**

**The questions asked in the interview will be about your feelings and family relationships.** Although unlikely, it is possible at some time during this study that you may experience some anxiety or stress as a result of some of the questions asked. If feelings of anxiety or stress persist after the completion of the session, arrangements will be made for you to access support from the counselling support services of the Peel Community Mental Health Service, Lakes Rd., Mandurah, 95318080 or the local branch of Relationships Australia, 7 Anzac Parade, Mandurah, 95355711. There will be no cost to you for these services.

**Please note:**

- This study is being carried out in accordance with the National Statement on Ethical Conduct in Research Involving Humans (NHMRC).
- **Any information you provide will remain confidential. Your name or any identifying information will not be reported or published.**
- **The recording will be destroyed once a written transcript has been made with any identifying information excluded.** This transcript will only be used for this research. In adherence to university policy, the study's collected data will be kept in a secure location for five years and then destroyed.
- The only way the information could be shared with others is if police or a law court requires it.

This study has been approved by the Curtin University Human Research Ethics Committee (Approval Number HR 133/2009). If needed, verification of approval can be obtained either by writing to the Curtin University Human Research Ethics Committee c/- Office of Research and Development, Curtin University of Technology, GPO Box U1987, Perth, 6845 or by telephoning 9266 2784 or emailing hrec@curtin.edu.au.

Further questions? If you would like further information about the study, please contact:  
Dawson Cooke Email: dawson.cooke@postgrad.curtin.edu.au Phone: (08) 9266 3086  
Or you may wish to contact the project supervisor, Dr. Garth Kendall  
Email: g.kendall@curtin.edu.au Phone: (08) 9266 2191

***Thank you for taking time to read this information sheet. If you consent to participate please sign the enclosed form and keep this letter for your information.***

**Appendix G – Consent Form**

Project Title: **Parental Reflecting Functioning**

**Participant Statement**

I.....(Print full name), have read the information sheet on this study of parental reflective functioning.

- I understand the nature and purpose of the study, and what my participation involves.
- I agree to participate in a one-hour interview.
- I agree for the interview to be audio recorded.
- I understand that my participation in this study is voluntary and I can withdraw at any time without problem.
- I have had the opportunity to ask questions.
- I understand that all my information will remain confidential.

Signed: \_\_\_\_\_  
Participant

Date: \_\_\_\_/\_\_\_\_/\_\_\_\_

Signed: \_\_\_\_\_  
Researcher

Date: \_\_\_\_/\_\_\_\_/\_\_\_\_

**Appendix H - Rasch Journal Article**

Published in *Measurement and Evaluation in Counseling and Development*.

Differences between mothers' and fathers' ratings of family functioning with the

Family Assessment Device: the validity of combined parent scores

## Running head: MOTHER AND FATHER FAMILY FUNCTIONING

Differences between mothers' and fathers' ratings of family functioning with the Family

Assessment Device: the validity of combined parent scores

### **Short Abstract**

The psychometric properties of the General Functioning subscale of the McMaster Family Assessment Device (GFAD) were examined using the Rasch Model ( $n = 237$  couples). Mothers' and fathers' ratings of the GFAD are recommended provided these are analysed separately. More than a quarter of couples differed significantly in their ratings.

### **Abstract**

Inconsistent results have been found regarding the differences between mothers' and fathers' ratings of family functioning. The present study evaluates the psychometric properties of the General Functioning subscale of the McMaster Family Assessment Device (GFAD) using the Rasch measurement model. In particular, the performance of a combined parent GFAD rating scale is evaluated. The GFAD was completed by 237 couples as part of an Australian population-based study. The analysis revealed that, even though the construct of family functioning was the same for mothers and fathers, more than a quarter of couples differed significantly in their ratings. Fathers were generally more severe in their rating of family functioning than mothers. Because of these differences, combined parent scores were not considered valid. The GFAD is recommended for use with both mothers' and fathers' ratings, provided these are analysed separately.

## 2 MOTHER AND FATHER FAMILY FUNCTIONING

Self-report measures of family functioning are commonly used in population studies and as part of clinical assessments as a means of gaining understanding of an individual within the wider family system ([Cox & Paley, 1997](#)). The McMaster model of family functioning is a well established approach to conceptualising the problems and health of families and is the basis of a number of assessment instruments ([Miller, Ryan, Keitner, Bishop, & Epstein, 2000b](#)). This study focuses on the validity of scores from the general functioning subscale (GFAD) of the McMaster Family Assessment Device (FAD), with particular attention to its use with mothers and fathers.

The GFAD consists of 12 items (see Table A) with four response categories labelled strongly agree, agree, disagree and strongly disagree. Item scores are summed into an overall general score of family functioning. The items of the GFAD represent the subscales of the FAD, which were developed to assess the six dimensions of the McMaster model of family functioning ([Epstein, Baldwin, & Bishop, 1983](#); [Kabacoff, Miller, Bishop, Epstein, & Keitner, 1990](#); [Miller et al., 2000b](#)). The GFAD is well suited to large-scale population studies such as the Ontario Child Health Study ([Byles, Byrne, Boyle, & Offord, 1988](#)), in which the scale was required to be short, and for general screening of family problems rather than detailed assessment of specific areas of functioning ([Georgiades, Boyle, Jenkins, Sanford, & Lipman, 2008](#); [Kabacoff et al., 1990](#); [Ridenour, Daley, & Reich, 1999](#)).

### **Limitations of Family Assessment with the GFAD**

A number of studies have used the GFAD with mothers and fathers from non-clinical samples (e.g., [Kabacoff et al., 1990](#); [Stevenson-Hinde, Curley, Chicot, & Jóhannsson, 2007](#)). A benefit of the GFAD is that the item statements are targeted at family level functioning rather than characteristics of individual or dyad functioning. Therefore, multiple family members can evaluate the family's functioning and differences between ratings of the same family can be examined. However, findings are inconsistent regarding the level of agreement between mother and father ratings. One study of parents with children aged seven years (N = 55) found the GFAD to have

### 3 MOTHER AND FATHER FAMILY FUNCTIONING

the lowest level of agreement of all the seven FAD subscales ( $r = .24$ ), although there was no difference between means of mother and father GFAD scores ([Akister & Stevenson-Hinde, 1991](#)). Similarly Sawyer, Sarris, Baghurst, Cross and Kalucy ([1988](#)) and Stevenson-Hinde et al. found no significant difference between the mean scores of mothers and fathers of adolescents ( $N = 146$  and  $113$  respectively). However, the GFAD scores of parents with preschool age children ( $N = 100$ ) were found to have the highest level of parental agreement ( $r = .52$ ) of all the subscales and mothers tended to rate family functioning healthier than fathers ([Stevenson-Hinde & Akister, 1995](#)).

The differences between GFAD the scores of mothers and fathers are of particular interest because of the practice of summing couple scores to produce a combined rating of family functioning ([Hayden et al., 1998](#); [Miller et al., 1994](#)). Summing parent scores is known to mask the differences in ratings that are commonly found between mothers and fathers and provide a score that is not indicative of either parent's assessment of the family ([Green & Vosler, 1992](#)). Also, analysis of differing scores has revealed that significant differences can be an indication of unhealthy family functioning ([Akister & Stevenson-Hinde, 1991](#)).

Georgiades, et al. ([2008](#)) argued from a family systems perspective for multiple family member reports of whole family functioning, and emphasized the need to identify the shared and non-shared variability of family members' assessments. Epstein et al. ([1983](#)) also acknowledged in the development of the FAD, that different family members and observers of a family are likely to have different points of view regarding family functioning, and suggest that these differences are not primarily errors of measurement. To the contrary, they emphasised the identification of such differences provides useful and important information worthy of further investigation. Likewise Cook and Kenny ([2006](#)) proposed the testing of 'level validity' of family functioning scores. They contend a measure of high-order level family functioning should account for variance that is explained by lower-order levels of dyad functioning or individual functioning. The

#### 4 MOTHER AND FATHER FAMILY FUNCTIONING

differences commonly found in mother and father GFAD scores suggests the influence of unique dyad or individual functioning problems. These differences would be unidentifiable if the assessment relies on only one respondent or if it is a combined score from two respondents.

Controversy regarding the dimensions of the FAD has raised issues of the appropriateness of factor analysis as a test of validity since the FAD was developed with a “rational-theoretical” approach ([Miller, Ryan, Keitner, Bishop, & Epstein, 2000a, 2000c](#); [Ridenour et al., 1999](#); [Ridenour, Daley, & Reich, 2000](#)). This study contributes to the examination of the GFAD by applying a Rasch measurement theory approach to scale evaluation which examines data at an item response level with a unidimensional model ([Rasch, 1966](#)).

#### **Validation of Rating Scale Data**

The goal of psychometric methods is to determine the validity and reliability of scores from instruments of measurement, such as tests and rating scales. Another goal of psychometric methods is to determine if an instrument accurately reflects differences in subgroups of a population. While this is relevant for all instruments used in epidemiological studies, it has particular relevance for the GFAD where it is important to identify if views about family functioning differ between parents and other family members. The Rasch model ([Andrich, 1978](#); [Rasch, 1960](#)) is increasingly used to validate scores from instruments that purport to measure psychological constructs (e.g., [Allison, Baron-Cohen, Wheelwright, Stone, & Muncer, 2011](#); [da Rocha, Chachamovich, de Almeida Fleck, & Tennant, 2013](#)). This use has increased because application of the Rasch model involves the formal testing of a scale against a mathematical model of measurement – a model consistent with the following fundamental requirements for measurement:

- i. *Unidimensionality*: When there is evidence of multidimensionality, the total score is not reflective of a unified construct and the measure should be re-evaluated.

## 5 MOTHER AND FATHER FAMILY FUNCTIONING

- ii. *Independence of responses*: If a response to an item depends on the response to another item, the items are said to be dependent, and one of them can be considered redundant. Response dependent items inflate the reliability of scores giving a false impression of the amount of measurement error ([Marais & Andrich, 2008](#); [Smith, 2002](#)).
- iii. *Response categories operate as intended*: The scoring order of response categories should reflect increasing levels of the latent trait being measured ([Andrich, 2011](#)).
- iv. *Invariant item functioning for subgroups*: Items should retain their meaning for different subgroups of the population, for example across different ages or between females and males. If this is so, the degree to which items are endorsed will be the same for each subgroup.

Since previous studies have produced conflicting results regarding the differences between mothers' and fathers' scores on the GFAD, it is of particular importance to test that items function the same way for mothers' and fathers.

### **The Current Study**

In this study the 12-item GFAD scale was evaluated by two procedures. First mothers' and fathers' responses were analysed separately, and then composite data from both mothers and fathers were conjointly analysed. Application of two procedures to the same data enabled the performance of the GFAD scale to be assessed in two different ways. The research questions investigated in this study are:

- (i) How do mothers and fathers rate family functioning differently on the GFAD?
- (ii) Do scores from a combined 24-item GFAD scale for couples provide a valid measure of the construct?

## **Method**

### **Participants and Procedure**

## 6 MOTHER AND FATHER FAMILY FUNCTIONING

This evaluation of the GFAD used data from the Peel Child Health Study (PCHS; <http://www.peelchildhealthstudy.com.au>). The PCHS is a longitudinal population study, designed to help identify the conditions that provide children with the maximum opportunity for achieving their developmental potential. The project included collection of psychosocial, environmental, biological and genetic data, with a focus on the complexity of individual behaviours in context. Participants of the PCHS were English speaking mothers and fathers living in the Peel Region of Western Australia ([Peel Development Commission, 2012](#)). Over a 3-year period, medical practitioners invited all families in their care with a pregnancy at about the 18th week of gestation to participate in the study. Complete GFAD data were obtained for 237 of 433 couples that completed questionnaires administered at the time of recruitment. Only selected sociodemographic data and responses to the 12-item GFAD were used for this analysis. Ethical approval was obtained from the Curtin University Human Research Ethics Committee.

### **Statistical Analysis**

Four negatively worded GFAD items were scored as strongly agree (0), agree (1), disagree (2) and strongly disagree (3) to satisfy the requirement that scoring starts at 0 in a Rasch analysis. Positively worded items were reverse scored so that higher scores represented *healthier* family functioning, as per recent studies using the FAD ([Aarons, McDonald, Connelly, & Newton, 2007](#); [Georgiades et al., 2008](#)).

Responses were analysed according to the polytomous Rasch model using the RUMM2030 software ([Andrich, Sheridan, & Luo, 2012](#)). The data best fitted the partial credit parameterization of the polytomous model for all the analyses, as indicated by the Likelihood ratio test in RUMM2030. Three analyses were performed:

- (i) Fathers responses only to the 12-item GFAD.
- (ii) Mothers responses only to the 12-item GFAD.

## 7 MOTHER AND FATHER FAMILY FUNCTIONING

- (iii) Father and mothers GFAD responses combined to form a set of 24 items (mother items 1 to 12 and father items 13 to 24).

The following aspects of these scales were investigated:

**Working of response categories.** A threshold is the point on the measurement continuum where the probability of a response in two adjacent response categories is equal. If the order of response categories on the instrument reflects increasing levels of the latent trait, as they should, then the order of the thresholds will be the same as that of the response categories. If not, it indicates that the response categories are not working as intended ([Andrich, 2011](#)).

**Person/Item alignment and reliability.** Because Rasch person and item estimates are on the same scale, the alignment of persons to items can be assessed. An index of reliability, the Person Separation Index (PSI) is reported, which in general is similar in value to Cronbach's alpha ([Andrich, 1982](#)).

**Fit to the model.** A number of statistics indicate data fit to the Rasch model. The item chi-square fit statistic compares the residuals of groups of persons, formed on the basis of their overall estimate. This study reports a summary chi-square statistic - a significant value means that the hierarchical ordering of the items varies across the trait. Item fit is also investigated statistically with the item fit residual statistic, and person fit with the person fit residual statistic. The item and person fit residuals are summary values of individual person by item residuals. If the data fit the model the residual mean will be close to 0 and the SD close to 1 (Andrich et al., 2012).

**Response dependence and multidimensionality.** Marais and Andrich ([2008](#)) considered two ways that the assumption of local independence between items can be violated. The first is a violation of statistical dependence, which they called response dependence and the second is multidimensionality. Items with high residual correlations are considered to be response dependent. Multidimensionality can be assessed in three ways. First, if a principal component analysis of the residuals (PCA) indicates no meaningful pattern in the residuals, the

## 8 MOTHER AND FATHER FAMILY FUNCTIONING

unidimensionality of the scale is supported ([Smith, 2002](#)). Second, if a PCA indicates a meaningful pattern, two subsets of items can be formed, consisting of items loading positively or negatively on the first component. Person scores on the two subtests are compared with a t-test analysis (e.g., [Hagquist, Bruce, & Gustavsson, 2009](#); [Smith, 2002](#)). Third, the relationship between the subtests (or dimensions) is examined with a theoretical correlation between the underlying traits (Andrich et al., 2012).

### Results

#### Descriptive Statistics

Selected sociodemographic characteristics of the parents are summarised in Table B. The parents ranged in age from 16 to 50 years with a mean age of 30 years for mothers and 32 years for fathers. All couples spoke English as their primary language at home and were living in a cohabiting relationship. Sixty per cent of the mothers were expecting their first child.

#### Results of Separate Analyses of Mothers' and Fathers' GFAD

**Response categories.** In both analyses all items were found to have thresholds 1 and 2 reversed, indicating that the response categories did not function as intended. Therefore, as a post hoc solution to this problem ([Andrich & Wright, 1994](#)), all items were rescored such that categories scored 0 and 1 were both scored 0, and the next two categories scored 1 and 2, resulting in three possible scores and two thresholds. Rescoring as described resulted in all items showing ordered thresholds.

**Fit to the Rasch model.** Table C shows a summary of the fit statistics for both analyses after responses were rescored. When analysed separately fathers' responses to the 12-item GFAD fitted the model, with a summary chi square statistic of 48.12 ( $df = 36, p = 0.09$ ). The item fit residual mean was -0.51 (SD = 1.41), and the person fit residual mean was -0.58 (SD = 1.47). Table C shows also that mothers' responses similarly fitted the model when analysed separately.

## 9 MOTHER AND FATHER FAMILY FUNCTIONING

**Person/Item alignment and reliability.** Figure A shows the person-item threshold distributions for mothers' responses (top) and fathers' responses (middle). The mean of the mothers' estimates were 2.16 compared to the mean of the item thresholds, which is constrained to be 0. There was a clear ceiling effect for the mothers' estimates. The mean of the fathers' estimates was also positive, but not as high at 1.82, and the ceiling effect was not as pronounced. Table C also shows values of the index of reliability, which ranged between 0.8 and 0.85.

**Response dependence and dimensionality.** Examination of the item residual correlations showed no evidence of item response dependence for either mother or father GFAD items. A PCA of the residuals showed no evidence of multidimensionality for either mother or father GFAD scales.

**Differences between mothers' and fathers' ratings.** Gender effects were explored in these analyses by investigating the item order for mothers and fathers. Figure B shows the mothers' item locations plotted against the fathers' item locations and makes apparent the ease or difficulty of endorsing particular items. Three items that were *easy to endorse* (most likely to be scored highly even at unhealthy levels of family functioning) for both mothers and fathers were item 4 (We avoid discussing our fears and concerns), 8 (We feel accepted for what we are) and 11 (We don't get on well together). Items that were *difficult to endorse* (most likely to receive a low score even at healthy levels of family functioning) for both mothers and fathers were item 1 (Planning family activities is difficult because we misunderstand each other), 9 (Making decisions is a problem in our family) and 10 (We are able to make decisions about how to solve problems). The correlation between item order for mothers and fathers was high (.84).

Figure B also shows the 95% confidence lines created with the item standard errors (SE). Items plotted inside the SE confidence lines are understood to be invariant ([Bond & Fox, 2007](#)). Items 7 and 5 were only just outside these confidence lines. Most notably aberrant was item 2 (in

## 10 MOTHER AND FATHER FAMILY FUNCTIONING

times of crisis we can turn to each other for support), which was the easiest item to endorse for fathers (location -1.04), and relatively more difficult for mothers (location -0.16).

### **Results of the Combined Analysis of the 24-item GFAD**

**Response categories.** As with the separate analyses all items were found to have thresholds 1 and 2 reversed, so items were rescored as in the separate analyses, which resulted in all items showing ordered thresholds.

**Fit to the Rasch model.** Table C shows that ratings analysed as a combined 24-item scale fit the model, as indicated by a summary chi square statistic of 73.28 ( $df = 72, p = 0.44$ ). The item fit residual mean was -0.11 (SD = 1.31), and the person fit residual mean was -0.44 (SD = 1.73). Smith (2002) has shown that Rasch fit statistics are insensitive to multidimensionality if the dimensions contain the same number of items. That is the case here since 12 items are mothers' ratings items and 12 items are fathers' ratings items. Further investigation through specific tests of unidimensionality is therefore very important before one can conclude that responses to the 24-item couple GFAD fit the Rasch model.

**Person/Item alignment and reliability.** Figure A shows the person-item threshold distributions for the combined analysis (bottom). The mean of the person estimates was positive, 1.80, relative to the arbitrary origin of 0 for the item threshold locations. Even though there was no pronounced ceiling effect, the graph shows very high estimates at the positive end of the scale where there are no thresholds. The PSI, 0.88, indicates good reliability.

**Response dependence.** Examination of the item residual correlations for item response dependence showed a number of item pairs with relatively high correlations. Further analysis should be undertaken to assess the level of dependency (Andrich, Humphry, & Marais, 2012), and redundant items could be considered for removal from the scale or procedures undertaken to account for dependence. Detailed analysis of response dependence is beyond the scope of this study.

## 11 MOTHER AND FATHER FAMILY FUNCTIONING

**Differences between mothers' and fathers' ratings - dimensionality.** Gender effects were explored in this analysis by comparing the means of the item locations of mothers and fathers. The mean of the 12 mother items was -0.18 and mean of the 12 father items was 0.18. On the whole, fathers rated family functioning lower than mothers, that is, the items were more difficult to endorse for fathers than for mothers. This resulted in a higher mean location of items for fathers than for mothers.

Gender differences were further explored in this analysis by looking for evidence of multidimensionality, which was clearly found using a PCA of residuals. Mother items loaded negatively and fathers positively on the first principal component. This first component explained 20.8% of the total variance among residuals with an eigenvalue of 4.99. The next two components were comparatively smaller; only 7.1% (eigenvalue = 1.71) and 6.6% (eigenvalue = 1.58) for components two and three respectively. A paired t-test of person estimates from subsets of mother and father scores differed significantly ( $p < 0.5$ ) for 27% of couples (95% CI: 25% - 30%).

In a third test of unidimensionality, mothers' responses (items 1 to 12) were summed into a subtest and fathers' responses (items 13 to 24) were summed into a subtest. This resulted in two 'items' for each family unit, one summarising mothers' ratings of the family and the other summarising fathers' ratings of the family. When two subtests were formed in this way, the PSI decreased from 0.88 to 0.41, indicating considerable multidimensionality. The estimated correlation between dimensions was low at .47. This is the theoretical correlation between the underlying traits in different subtests, corrected for attenuation because of error.

### **Discussion**

When analysed separately, it was found that mothers' and fathers' GFAD scores were generally valid and reliable and that the underlying constructs being measured were similar. However, the four response categories did not function as expected and serious problems were

## 12 MOTHER AND FATHER FAMILY FUNCTIONING

encountered when mother and father GFAD ratings were combined to produce a combined score for family functioning.

According to the Rasch model each of the four rating response categories of the GFAD items should correspond with a consecutively higher level of the underlying trait. It was found, however, that successive categories of the GFAD items were not responded to as would be expected, and this was the case for both mothers and fathers, when analysed separately or together. A possible reason that this pattern was not observed for the two low scored categories of the GFAD is that there were relatively low frequencies of responses in these categories. It is also possible that parents reporting low family functioning may have difficulty or confusion when distinguishing between low and very low ratings of items.

The finding that mothers' and fathers' GFAD scores showed good reliability and fit to the Rasch model when considered separately, with no evidence of item response dependence or multidimensionality, confirms previous factor analysis results ([Ridenour et al., 1999](#)). These results support the use of the general functioning subscale as a summary measure of family functioning. However, differences were found between the item location order for mothers and fathers. Mothers and fathers endorsed three of the items to a different degree in relation to the other items. The most notable difference was found with item 2 (in times of crisis we can turn to each other for support), which was the easiest item to endorse for fathers and comparatively less so for mothers. The differences found in item location with these three items indicate some slight differences in priority or meaning of some aspects of family functioning for mothers and fathers. Further study with qualitative methods is required to better understand how availability of support in crisis is perceived differently by mothers and fathers in the context of family functioning.

As expected, the combined measure of family functioning was found to have clear evidence of multidimensionality, with the father and mother items accounting for two dimensions that were not strongly correlated ( $r = .47$ ), confirming similar correlations from a previous study of families

## 13 MOTHER AND FATHER FAMILY FUNCTIONING

with young children ([Stevenson-Hinde & Akister, 1995](#)). When this multidimensionality was accounted for, the estimated reliability was very low indicating the initial estimate of reliability was inflated. This finding suggests that the use of a combined mother and father rating is a poor use of the GFAD. While a combined score may be highly correlated with other measures of family functioning ([Stevenson-Hinde & Akister, 1995](#)), such a use of the GFAD masks significant variability in couples' scores that is most likely explained by dyadic or individual characteristics ([Cook & Kenny, 2006](#)).

The variability between mother and father scores was examined with a comparison of mean scores and a paired comparison of couple's scores. For this sample, fathers found the GFAD items generally more difficult to endorse than mothers, which supports the previous findings that fathers tend to report poorer family functioning than mothers ([Stevenson-Hinde & Akister, 1995](#)). This difference is also clearly represented in the high proportion (27%) of couples with significantly different scores. These results indicate both mother and father ratings should be separately considered and compared when evaluating family functioning. The identification of differences in scores is likely to be an indication of unhealthy functioning in addition to exceeding a simple raw score threshold ([Akister & Stevenson-Hinde, 1991](#)).

### **Implications for Practice and Research**

The internal consistency reliability (PSI) of .80 to .88 for the GFAD scores is considered adequate for basic research purposes ([Nunnally, Bernstein, & Berge, 1967](#)). Fisher ([2010](#)) reported that a PSI of between .80 and .90 indicates the possibility of distinguishing between three groups enabling low medium and high levels of the trait to be separated with 95% confidence. Therefore, this study confirms the use of the GFAD for comparing groups or screening to identify families that may be experiencing problems. However, the GFAD is not an instrument that is suitable for quantitatively distinguishing between individual families or assessing change for an individual

## 14 MOTHER AND FATHER FAMILY FUNCTIONING

family in a clinical setting. Given the established theoretical basis of the GFAD ([Miller et al., 2000b](#)), practitioners may consider qualitative use of the GFAD responses as clinically useful.

Despite the fact that family functioning as a whole is greater than the sum of the individual family members' functioning ([Cox & Paley, 1997](#); [Miller et al., 2000b](#)), this study highlights the potential masking of important variation when the GFAD is used as a single informant or combined informant score. This finding confirms previous cautions by Green and Vosler ([1992](#)), and Georgiades et al. ([2008](#)) with clear evidence from Rasch analysis.

Reasons for different reports of family functioning from informants of the same family have been explored qualitatively by Stevenson-Hinde and Akister ([1995](#)). Unhealthy functioning was falsely identified for reasons such as situational stress expressed differently, limited personal development, and poor communication or expression. Conversely, healthy functioning was found to be falsely reported in cases of family dysfunction, such as when a child is used as a scapegoat and other problems are masked. Identifying differences in GFAD scores between multiple informants of a family is a useful screening for these potential issues.

### **Directions for Future Research**

The sample in this research was limited to non-clinical English speaking couples. Ideally this Rasch analysis should be replicated with a sample including families that have been identified as having family functioning difficulties, and also with versions of the GFAD in other languages and with other cultures. Although this study investigated the general performance of the GFAD it did not make comparisons with scores from other measures of individual and dyad functioning or alternative measures of family functioning, which is an important aspect of convergent construct validity. Another useful variation of this current analysis of the GFAD would be to utilise other combinations of family members, such as a parent and an adolescent child rating combined. These suggested directions of research are also applicable for the more comprehensive FAD.

## 15 MOTHER AND FATHER FAMILY FUNCTIONING

The threshold disorder identified with the lower rating categories requires further investigation. Data from a clinical sample would provide higher frequencies in these low categories and provide clearer evidence of response category use. If threshold disorder is confirmed, further study should then be undertaken with a single category for low functioning responses.

### **Conclusions**

Data from the GFAD were found to conform to the requirements of the Rasch measurement model. These findings confirm cautions from previous authors regarding use of a single informant of family functioning or summing mother and father scores. Analysis revealed that over a quarter of couples differed in their rating of their family's functioning. Also, fathers were on average more severe in their rating of family functioning than mothers. Identifying these differences is a source of valuable information regarding assessment of family functioning and an opportunity for better understanding overall family health or pathology.

**References**

- Aarons, G., McDonald, E., Connelly, C., & Newton, R. (2007). Assessment of family functioning in Caucasian and Hispanic Americans: Reliability, validity, and factor structure of the Family Assessment Device. *Family Process, 46*(4), 557-569.
- Akister, J., & Stevenson-Hinde, J. (1991). Identifying families at risk: exploring the potential of the McMaster Family Assessment Device. *Journal of Family Therapy, 13*(4), 411-421.  
doi: 10.1046/j..1991.00437.x
- Allison, C., Baron-Cohen, S., Wheelwright, S. J., Stone, M. H., & Muncer, S. J. (2011). Psychometric analysis of the Empathy Quotient (EQ). *Personality and Individual Differences, 51*(7), 829-835. doi: 10.1016/j.paid.2011.07.005
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika, 43*(4), 561-574.
- Andrich, D. (1982). An index of person separation in latent trait theory, the traditional KR-20 index, and the Guttman scale response pattern. *Education Research and Perspectives, 9*(1), 95-104.
- Andrich, D. (2011). Rating scales and Rasch measurement. *Expert review of pharmacoeconomics & outcomes research, 11*(5), 571-585. doi: 10.1586/erp.11.59
- Andrich, D., Humphry, S. M., & Marais, I. (2012). Quantifying local, response dependence between two polytomous items using the Rasch model. *Applied Psychological Measurement, 34*(4), 309-324
- Andrich, D., Sheridan, B. E., & Luo, G. (2012). *RUMM2030: Rasch Unidimensional Models for Measurement*. Perth, Western Australia: RUMMLaboratory.
- Andrich, D., & Wright, B. D. (1994). Rasch sensitivity and Thurstone insensitivity to graded responses. *Rasch Measurement Transactions, 8*(3), 382.

## 17 MOTHER AND FATHER FAMILY FUNCTIONING

- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: fundamental measurement in the human sciences* (2nd ed.). Mahwah, N.J.: Lawrence Erlbaum.
- Byles, J., Byrne, C., Boyle, M. H., & Offord, D. R. (1988). Ontario Child Health Study: Reliability and validity of the general functioning subscale of the McMaster Family Assessment Device. *Family Process, 27*(1), 97-104.
- Cook, W. L., & Kenny, D. A. (2006). Examining the validity of self-report assessments of Family Functioning: A question of the level of analysis. *Journal of Family Psychology, 20*(2), 209-216. doi: 10.1037/0893-3200.20.2.209
- Cox, M. J., & Paley, B. (1997). Families as systems. *Annual Review of Psychology, 48*(1), 243-267. doi: 10.1146/annurev.psych.48.1.243
- da Rocha, N. S., Chachamovich, E., de Almeida Fleck, M. P., & Tennant, A. (2013). An introduction to Rasch analysis for psychiatric practice and research. *Journal of Psychiatric Research, 47*(2), 141-148. doi: 10.1016/j.jpsychires.2012.09.014
- Epstein, N. B., Baldwin, L., & Bishop, D. S. (1983). The McMaster Family Assessment Device. *Journal of Marital and Family Therapy, 9*(2), 171-180.
- Fisher, W. P. (2010). Reliability, precision, and measurement in the context of data from ability tests, surveys, and assessments. *Journal of Physics: Conference Series*(238), 1-6.
- Georgiades, K., Boyle, M. H., Jenkins, J. M., Sanford, M., & Lipman, E. (2008). A multilevel analysis of whole family functioning using the McMaster Family Assessment Device. [References]. *Journal of Family Psychology, 22*(3), 344-354.
- Green, R. G., & Vosler, N. R. (1992). Issues in the assessment of family practice. *Journal of Social Service Research, 15*(3-4), 1-19. doi: 10.1300/J079v15n03\_01
- Hagquist, C., Bruce, M., & Gustavsson, J. P. (2009). Using the Rasch model in nursing research: An introduction and illustrative example. *International journal of nursing studies, 46*(3), 380-393.

## 18 MOTHER AND FATHER FAMILY FUNCTIONING

- Hayden, L. C., Schiller, M., Dickstein, S., Seifer, R., Sameroff, S., Miller, I., . . . Rasmussen, S. (1998). Levels of family assessment: I. Family, marital, and parent-child interaction. *Journal of Family Psychology, 12*(1), 7.
- Kabacoff, R. I., Miller, I. W., Bishop, D. S., Epstein, N. B., & Keitner, G. I. (1990). A psychometric study of the McMaster Family Assessment Device in psychiatric, medical, and nonclinical samples. [References]. *Journal of Family Psychology, 3*(4), 431-439.
- Marais, I., & Andrich, D. (2008). Formalizing dimension and response violations of local independence in the unidimensional Rasch model. *Journal of Applied Measurement, 9*(3), 200-215.
- Miller, I. W., Kabacoff, R. I., Epstein, N. B., Bishop, D. S., Keitner, G. I., Baldwin, L. M., & van der Spuy, H. I. J. (1994). The development of a clinical rating scale the McMaster model of family functioning. *Family Process, 33*(1), 53-69. doi: 10.1111/j.1545-5300.1994.00053.x
- Miller, I. W., Ryan, C. E., Keitner, G. I., Bishop, D. S., & Epstein, N. B. (2000a). "Factor analyses of the Family Assessment Device," by Ridenour, Daley, & Reich. *Family Process, 39*(1), 141-144. doi: 10.1111/j.1545-5300.2000.39112.x
- Miller, I. W., Ryan, C. E., Keitner, G. I., Bishop, D. S., & Epstein, N. B. (2000b). The McMaster approach to families: Theory, assessment, treatment and research. *Journal of Family Therapy, 22*(2), 168-189. doi: 10.1111/1467-6427.00145
- Miller, I. W., Ryan, C. E., Keitner, G. I., Bishop, D. S., & Epstein, N. B. (2000c). Why fix what isn't broken? A rejoinder to Ridenour, Daley, & Reich. *Family Process, 39*(3), 381-384.
- Nunnally, J. C., Bernstein, I. H., & Berge, J. M. t. (1967). *Psychometric theory* (Vol. 226): McGraw-Hill New York.
- Peel Development Commission. (2012). Peel Profile. Retrieved April 3, 2014, from <http://peel.wa.gov.au/wp-content/uploads/PeelProfileMarch2012.pdf>

## 19 MOTHER AND FATHER FAMILY FUNCTIONING

- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests (reprinted 1980, expanded ed. with forward and afterward by B. D. Wright)*. Chicago: The University of Chicago Press.
- Rasch, G. (1966). An item analysis which takes individual differences into account. *The British Journal of Mathematical and Statistical Psychology*, *19*(1), 49-57.
- Ridenour, T. A., Daley, J., & Reich, W. (1999). Factor analyses of the family assessment device. *Fam Process*, *38*(4), 497-510.
- Ridenour, T. A., Daley, J., & Reich, W. (2000). Further evidence that the family assessment device should be reorganized: Response to Miller and colleagues. *Family Process*, *39*(3), 375-380.
- Sawyer, M. G., Sarris, A., Baghurst, P. A., Cross, D. G., & Kalucy, R. S. (1988). Family Assessment Device: Reports from mothers, fathers, and adolescents in community and clinic families. *Journal of Marital and Family Therapy*, *14*(3), 287-296. doi: 10.1111/j.1752-0606.1988.tb00748.x
- Smith, E. (2002). Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. *Journal of Applied Measurement*, *3*(2), 205-231.
- Stevenson-Hinde, J., & Akister, J. (1995). The McMaster model of family functioning: Observer and parental ratings in a nonclinical sample. *Family Process*, *34*(3), 337-347. doi: 10.1111/j.1545-5300.1995.00337.x
- Stevenson-Hinde, J., Curley, J. P., Chicot, R., & Jóhannsson, C. (2007). Anxiety within families: Interrelations, consistency, and change. *Family Process*, *46*(4), 543-556. doi: 10.1111/j.1545-5300.2007.00231.x

Table A

*General functioning subscale of the FAD (GFAD)*

Item Number	Item Wording
1.	Planning family activities is difficult because we misunderstand each other*
2.	In times of crisis we can turn to each other for support
3.	We cannot talk to each other about sadness we feel*
4.	Individuals (in the family) are accepted for what they are
5.	We avoid discussing our fears and concerns*
6.	We express feelings to each other
7.	There are lots of bad feelings in our family*
8.	We feel accepted for what we are
9.	Making decisions is a problem in our family*
10.	We are able to make decisions about how to solve problems
11.	We don't get on well together*
12.	We confide in each other

---

\* Reverse scored items

## 21 MOTHER AND FATHER FAMILY FUNCTIONING

Table B

*Descriptive Statistics*

	Mean (SD)	Min / Max	Mothers % (N)	Fathers % (N)
Mother Age	30 (5.3)	16 / 45		
Father Age	32 (6.3)	16 / 50		
Mother pregnant with first child			60 (142)	
Education Level				
Not completed year 12			9 (21)	11 (24)
Only year 12 completed			10 (23)	11 (23)
Certificate or Diploma			49 (112)	59 (126)
Degree or higher			32 (73)	18 (39)
Average hours worked per week				
None			38 (91)	5 (13)
Less than 35			36 (85)	3 (8)
35 to 40			20 (47)	32 (75)
More than 40			6 (14)	60 (141)

*Note.* Total number of participants = 474 (237 couples). Percentage figures are of total number of responses to particular question.

## 22 MOTHER AND FATHER FAMILY FUNCTIONING

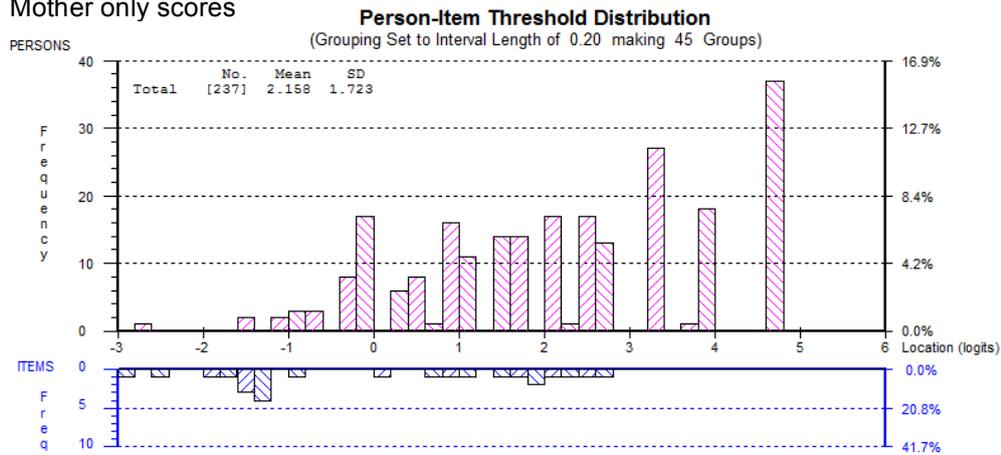
Table C

*Summary of fit statistics for the analysis of only mothers responses, only fathers responses and father/mother responses combined*

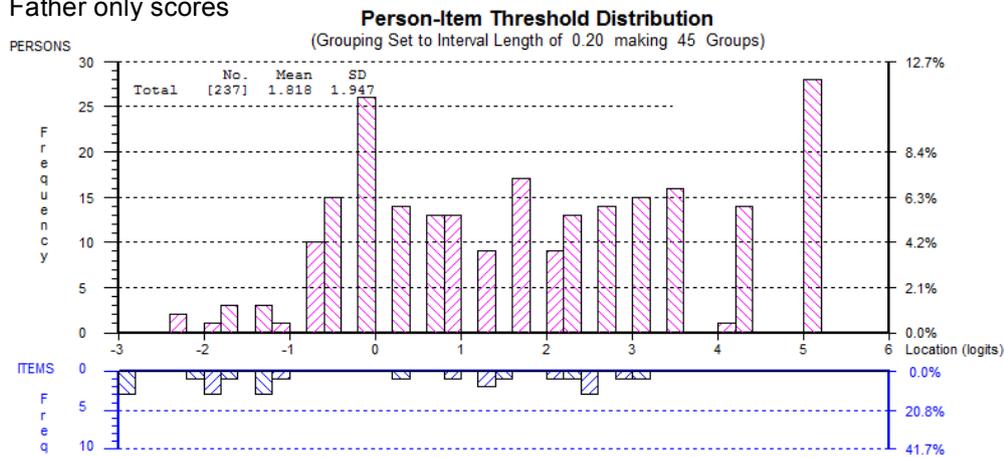
Analysis	Item fit residual		Person fit residual		Chisquare	<i>p</i>	Reliability	Unidimensional
	Mean (SD)		Mean (SD)		Value, df		PSI	t-test (95% CI)
Fathers	-0.51	(1.41)	-0.58	(1.47)	48.12, 36	0.09	0.85	n/a
Mothers	-0.21	(1.14)	-0.52	(1.60)	43.63, 36	0.18	0.80	n/a
Combined	-0.11	(1.31)	-0.44	(1.73)	73.28, 72	0.44	0.88	27% (25% - 30%)

## 23 MOTHER AND FATHER FAMILY FUNCTIONING

### Mother only scores



### Father only scores



### Combined scores

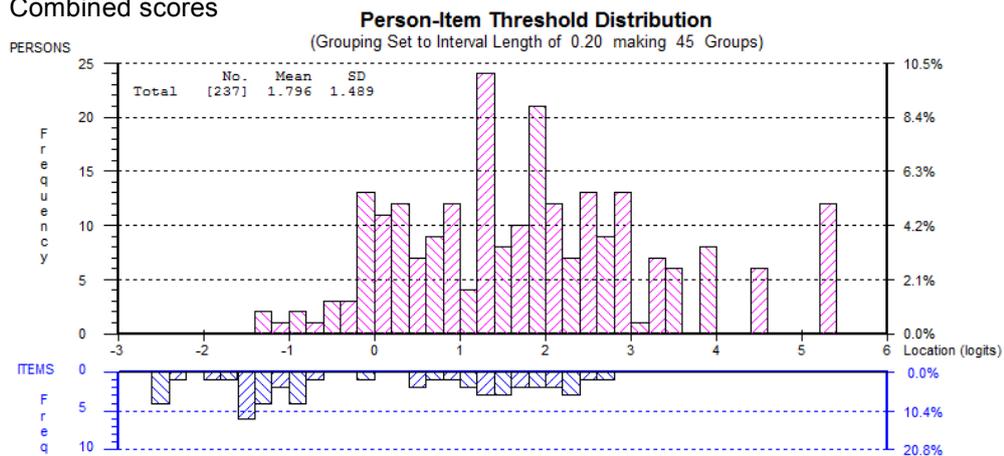


Figure A. GFAD person-item threshold distributions for mother only scores, father only scores and combined scores

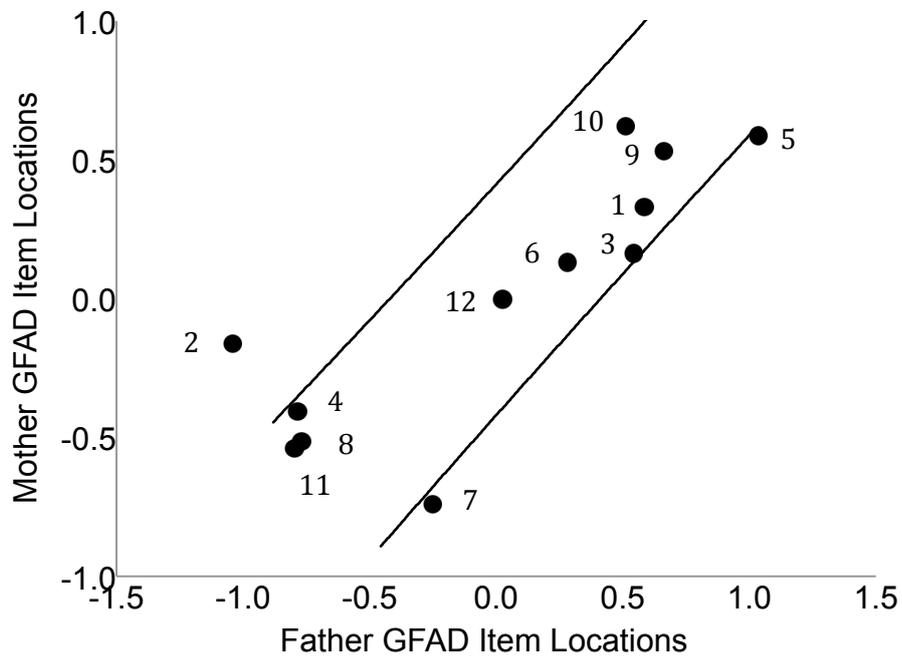


Figure B. Plot of mother and father GFAD item locations with 95% confidence lines