UNIVERSITY OF SOUTHERN QUEENSLAND

Calibration Analysis Within the Cognitive and Personality Domains: Individual Differences in Confidence, Accuracy, and Bias

A Dissertation submitted by

Sandra Florentina Baker, BSc (Hons)

For the award of

Doctor of Philosophy

2010

Abstract

Calibration research is concerned with the accuracy of confidence judgments made by individuals when responding to various cognitive tasks. Individuals are scored as accurate or inaccurate based on the objective criterion of whether their responses are correct. Within the personality domain, judging the accuracy of trait self-ratings is more complicated than in the cognitive area as there are no perfect criteria for evaluating the accuracy of these types of judgements (Colvin & Funder, 1991).

Because cognitive calibration research findings formed the anchor for the current studies, the decision was made to initially scrutinize the cognitive domain for mis-calibration, and whether individual differences in gender, age, personality, and ability, influenced cognitive confidence and bias scores. In order to achieve the aims, of this dissertation, three studies were conducted with a total of 831 individuals being tested. To determine accuracy within the personality domain, the current studies constructed a situation wherein Big Five personality assessments could be scored as accurate or inaccurate. Results showed that when consistency measures were used, accuracy scores for each Big Five trait were reasonably high across Studies 1, 2, and 3. Prior to the studies conducted in this dissertation, no techniques using calibration procedures had been established to assess Big Five confidence or bias.

Within the cognitive domain, calibration research has demonstrated the existence of a trait of self-confidence that appears to be independent of the type of activity being investigated. This result was replicated in Studies 1 and 3. However, the generality of this trait across other domains, such as personality assessments, remains largely unexplored. Two measures were designed to obtain confidence ratings in relation to Big Five personality judgments. Results from three studies showed that the

benchmark for peoples' confidence in Big Five judgments was around 80%. Data from Studies 1, 2, and 3 also demonstrated a one-factor solution when confidence scores for each Big Five trait were factor analysed. In studies 1 and 3 the factorial structure of cognitive and Big Five confidence scores was examined and both studies produced a two-factor solution.

The calibration paradigm also uses bias scores as a measure of how well calibrated individuals are when self-monitoring their performance on various cognitive tasks. Whether people are well-calibrated within the Big Five domain has not been investigated by previous researchers. The current studies examined whether people were mis-calibrated when making Big Five judgments about themselves. The data from two studies indicate that people were well-calibrated for each of the Big Five traits. Factor analyses of Big Five bias scores revealed a one-factor solution. When study 3 examined the factorial structure of cognitive and Big Five bias scores, the analyses showed that bias across these domains were separate but correlated processes. Across all three studies, individual differences in gender and age did not influence Big Five confidence, Big Five accuracy or Big Five bias scores.

One of the most significant implications of this dissertation, for calibration researchers, who are striving to understand the mis-calibration phenomenon, was that the structural analyses of cognitive and Big Five bias scores yielded a two-factor solution (i.e., Personality and Cognitive Bias), that was moderately correlated. Also, in the current studies, simple methods were used to obtain Big Five confidence ratings. These procedures could now be used to investigate the factorial structure of confidence in much more detail, and across other domains such as interests, attitudes and values. Practical implications of the current research within the field of clinical psychology were also discussed.

Certification of Thesis

I certify that the ideas, experimental work, results, analyses, software, and conclusions reported in this dissertation are entirely my own effort, except where otherwise acknowledged. I also certify that the work is original and has not been previously submitted for any other award, except where otherwise acknowledged.

Sandra F Baker

Date

ENDORSEMENT

Prof. Gerard J. Fogarty (Supervisor)

Date

Acknowledgements

To my supervisor Professor Gerard Fogarty thank you for your eternal faith in my ability to complete this document despite many difficulties. Your guidance has been invaluable and very much appreciated. This journey has been a challenge. However, the learning has been remarkable!

Thanks to all of the participants who made this journey possible. Thanks to the Psychology Technical Team who were always available for the technical aspects of data collection. Many thanks to Audrey, the nurse unit manager who has supported me all of the way, and who has been so gracious with allowing me to take time off. To my colleague Sheree who has taken over group therapy for me so that I could finish this project.

To my beautiful children Andrew, Brendon, Sarah, Katherine, Anna and Raleigh thanks for your constant support and love. To Peter, thank you for being there during the difficult journey of the PhD.

To my friends Sandra, Richard, Kerry, Lorna, Mel, and Nancey who have always been there for me. Thanks to Maxine, Liam, and Jen, for their continued support.

I would like to dedicate this PhD to my beloved sister-in-law Julie, a life-long friend, who passed away from cancer during the course of the PhD. To my father who survived a serious stroke. Good on you Dad! To my Mum who has been on the brink of death several times during my PhD journey. Go Mum you are amazing! To my closest Aunty who is stricken with leukaemia. Each one of you has been a rock to me.

Table of Contents

Abstract i

Certification of Thesis iii

Acknowledgements iv

Table of Contents v

List of Tables xiv

List of Figures xviii

List of Appendices xx

Chapter 1 – Introduction 1

1.1 Structure of the Dissertation 1

1.2 Statement of the Problem 1

1.3 Brief Rationale for Including Personality Judgments 4

1.4 Broad Aims and Overview of Current Studies 5

Chapter 2 – Literature Review 7

2.1 Introduction 7

2.2 Confidence and Meta-cognition 7

2.3 Calibration Research 9

2.3.1 The Experimental Paradigm 9

2.3.2 Numerical Measures of Calibration 9

2.3.2.1 Item-by-Item Confidence Scores (Self-Monitoring) 9

2.3.2.2 Self Evaluation Scores (Mis-calibration at the Global Level) 10

2.3.3 Graphical Measures of Calibration 10

2.3.3.1 Calibration Curves 10

2.3.3.2 Item-Specific Scatterplots 11

2.4 Theories of Cognitive Mis-calibration 11

2.4.1 Heuristics and Biases Approach 11

2.4.2 Error Explanations of Mis-calibration 14

2.4.3 Differing Modes of Uncertainty 14

2.4.4 Ecological Approach 15

2.4.5 Summary 18

2.5 Empirical Findings in Calibration Research 19

2.5.1 The Hard-Easy Effect 19

2.5.2 Good Calibration 19

2.5.3 Underconfidence 21

2.5.4 Overconfidence 22

2.5.5 Self-Confidence and the Study of Individual Differences 23

2.5.5.1 Gender Differences in Accuracy, Confidence and Mis-calibration 24

2.5.5.1.1 Gender Differences and Task Accuracy 24

2.5.5.1.2 Gender Differences and Confidence 25

2.5.5.1.3 Gender Differences and Mis-Calibration 26

2.5.5.2 Age Differences in Accuracy, Confidence and Mis-Calibration ... 27

2.5.5.2.1 Age Differences in Accuracy 27

2.5.5.2.2 Age differences in Confidence and Mis-calibration 28

2.5.5.3 Ability and Confidence 29

2.5.5.4 Personality Correlates of Confidence and Mis-Calibration 31

2.5.5.5 A Trait of Self-Confidence within the Cognitive Domain 34

2.5.5.5.1 Gf-Gc Theory and the Calibration Paradigm 35

2.6 Justifying Inclusion of Personality Confidence Judgments 38

2.7 Empirical Relations Between Personality and Ability/Intelligence Measures 40

2.8 Confidence Judgments Across Cognitive and Personality Measures: One Trait or Two? 42
2.8.1 Summary 44
2.9 Measurement Issues and Empirical Findings of Personality Judgements 45
2.9.1 Interpersonal Perception 46
2.9.2 Intrapersonal Perception 46
2.9.2.1.1 Reflected Appraisals 47
2.9.2.1.2 Direct Appraisals...47
2.9.2.1.3 Biases That Can Affect the Accuracy of Self-Perception.....48
2.9.2.1.4 Self-Enhancement 49
2.9.2.1.5 Above-Average Effect 49
2.9.2.1.6 Social Desirability...50
2.9.2.1.7 Are Personality Judgements Error-Prone or Error-Free?......51
2.10 Chapter Summary...53
Chapter 3 - Study 1 55
3.1 Introduction...55
3.2 Mis-calibration for Gf, Gc, & Gv Tasks 56
3.3 Individual Differences in Cognitive Confidence and Mis-calibration.............57
3.3.1 Gender 57
3.3.2 Age 58
3.3.3 Ability 59
3.3.4 Personality...59
3.4 Individual Differences in Big Five Accuracy, Confidence, and Mis-calibration 60

3.4.1 Gender Differences in Big Five Personality Confidence, Accuracy and Bias 60
3.4.2 Age Differences in Big Five Confidence, Accuracy and Bias 62
3.5 The Factorial Structure of Cognitive and Personality Confidence Judgments 63
3.6 Factorial Structure of Cognitive Bias Scores 65
3.6.1 Correlations between the Personality Bias Score and Cognitive Bias Scores 65
3.7 Restatement of Hypotheses 66
3.8 Method 68
3.8.1 Participants 68
3.8.2 Materials 69
3.8.2.1 Cognitive Tests 69
3.8.2.2 Self-report Measures 70
3.8.3 Procedure 75
3.9 Results 76
3.9.1 Data Screening 76
3.9.1.1 Reliability Analysis 77
3.9.1.2 Descriptive Statistics 78
3.9.2 Calibration Results for the Cognitive Tasks 79
3.9.2.1 Reasoning Tasks (Gf) 79
3.9.2.2 Acculturated Knowledge Tasks (Gc) 80
3.9.2.3 Visual Perceptual Task (Gv) 83
3.10 Individual Differences in Cognitive Confidence and Mis-calibration 84
3.10.1 Gender Differences in Cognitive Confidence 84
3.10.2 Age differences in Cognitive Mis-calibration 85

3.10.3 Mis-calibration and Ability 86
3.10.4 Personality Correlates of Cognitive Confidence Judgments 86
3.11 Individual Differences in Big Five Accuracy, Confidence and Mis-calibration 88
3.11.1 Gender Differences in Big Five Confidence and Accuracy 91
3.11.2 Age Differences in Big Five Confidence and Accuracy 92
3.12 The Factorial Structure of Cognitive and Big Five Confidence 93
3.13 Factor Structure of Cognitive Bias Scores 95
3.13.1 Correlations between the Personality Bias Score and Cognitive Bias Scores 96
3.14 Discussion 96
Chapter 4 - Study 2 107
4.1 Introduction 107
4.2 Item-by-Item and Block Personality Confidence Judgments 107
4.3 Accuracy Scores for the Big Five Block Judgments 113
4.4 Bias and Accuracy Scores for the Item-By-Item Big Five Judgments 114
4.5 Gender and Age Differences in Personality Confidence, Accuracy, and Bias 117
4.6 Factorial Structure of Big Five Confidence and Big Five Subscale Scores ... 118
4.7 Hypotheses 118
4.8 Participants 120
4.9 Materials 120
4.10 Procedure 125
4.10.1 Scoring 126
4.11 Results 126

4.11.1 Factorial Structure of the IPIP 127

4.11.2 Factorial Structure of the IPIP Form B 129

4.11.3 Descriptive Statistics and Factorial Structure of Item-level Accuracy Scores 131

4.11.4 Reliability Analysis 132

4.11.4.1 Descriptive Statistics.. 132

4.12 The Factorial Structure of Confidence... 134

4.13 Gender and Age Differences in Personality Confidence 135

4.14 Big Five Accuracy and Bias Scores... 137

4.14.1 Item-by-item Big Five Accuracy and Bias 139

4.14.2 Gender and Age Differences in Accuracy and Bias 139

4.14.3 Factorial Structure of Big Five Bias 140

4.15 Factorial Structure of Big Five Confidence and Big Five Subscale Scores 142

4.16 Discussion... 143

Chapter 5 - Study 3 150

5.1 Introduction.. 150

5.2 Cognitive Calibration and Confidence on Gf Tasks 150

5.2.1 Bias in Relation to Gf Tasks 150

5.2.2 Factorial Structure of Item-Level Cognitive Confidence Ratings and Confidence Ratings Obtained in Relation to PTPEs 151

5.3 Other Individual Differences in Cognition 153

5.3.1 Need for Cognition in Cognition: Gf Tasks 153

5.3.2 Negative Affect in Cognition: Gf Tasks 154

5.4 Replication Analyses in Relation to Personality... 155

5.5 Validity Checks of Big Five Confidence Judgments.................................... 155

5.5.1 Self-Focussed Attention in Relation to Personality 155
5.5.2 Negative Affect in Relation to Personality 156
5.5.3 Positive Affect in Relation to Personality..156
5.5.4 Need for Cognition in Relation to Personality.......................................156
5.6 Factorial Structure of Personality and Gf Confidence Scores 157
5.7 Self-report Intelligence/Abilities Confidence and Personality Confidence Ratings 159
5.8 Factorial Structure of Cognitive and Personality Bias Scores 160
5.9 Hypotheses in Relation to Cognition 160
5.10 Replication Hypotheses in Relation to Personality.....................................162
5.11 New Hypotheses: Study 3 in Relation to Personality 163
5.12 New Hypotheses Study 3 in Cognition and Personality 164
5.13 Method 165
5.13.1 Participants...165
5.13.2 Materials...165
5.13.3 Procedure..169
5.13.4 Scoring 170
5.14 Results..171
5.14.1 Normality and Outliers..171
5.15 Results for Cognitive Tasks 172
5.15.1 Bias in Relation to Gf Tasks 172
5.15.2 Individual Differences for Gf Tasks 176
5.15.2.1 Ability Differences in Relation to Gf Bias Scores.....................177
5.15.2.2 Personality in Relation to Gf Confidence and Bias 178
5.15.2.3 Need for Cognition in Relation to Gf Tasks 179

5.15.2.4 Negative Affect in Relation to Gf Tasks 179
5.15.3 Factorial Structure of the IPIP 179
5.15.4 Factorial Structure of the IPIP Form B 181
5.15.5 Descriptive Statistics and Factorial Structure of Item-level Accuracy Scores 183
5.15.6 Reliability Analysis for Self-Report Measures 184
5.15.7 Descriptive Statistics Self-Report Measures 185
5.16 The Factorial Structure of Big Five Confidence 186
5.16.1 Gender and Age Differences in Big Five Confidence, Accuracy and Bias 187
5.16.1.1 Big Five Accuracy and Bias Scores 190
5.17 Validity Checks in Personality 193
5.17.1 Self-Focussed Attention in Relation to Personality 193
5.17.2 Negative Affect in Relation to Personality 194
5.17.3 Positive Affect in Relation to Personality 194
5.17.4 Need for Cognition in Relation to Personality 194
5.18 Factorial Structure of Big Five, Gf, and Self-Report Intelligence Confidence Scores 195
5.19 Factorial Structure of Cognitive and Big Five Bias Scores 198
5.20 Discussion 200
Chapter 6 General Discussion 212
6.1 Factorial Structure of Big Five Confidence 217
6.2 Gender and Age Differences in Big Five Confidence, Accuracy and Bias Scores 218
6.3 The Factorial Structure of Cognitive and Big Five Confidence 219

6.4 Big Five Bias 221

6.5 Factorial Structure of Cognitive and Big Five Bias Scores 222

6.6 Practical and Other Research Implications 223

6.6.1 Clinical and Counselling Psychology 224

6.7 Other Limitations 225

6.8 Conclusions 226

References 228

List of Tables

Table 1.1 *Brief Overview of Studies* 6

Table 3.1 *Descriptive Statistics for all Dependent Variables (N = 127)* 78

Table 3.2 *Means of Confidence and Bias Scores for Males (N = 40) and Females (N = 87) on Cognitive Tasks in Study 1* 85

Table 3.3 Correlations between Age and Bias Scores 86

Table 3.4 *Mean Bias Scores for First and Fourth Quartiles on Cognitive Tasks* 86

Table 3.5 *Correlations Among IPIP Subscale Scores and Accuracy, Confidence and Bias and Partial Correlations Between Big Five Scores and Accuracy, Confidence and Bias (N = 127)* 88

Table 3.6 *Descriptive Statistics for Big Five Accuracy Scores (N*= 127) 90

Table 3.7 *Correlations among Big Five Accuracy Scores and Cognitive Confidence, Accuracy and Bias Scores* (*N* = 127). 92

Table 3.8 *Mean Big Five Confidence Scores for Males (n* = 40*) and Females (n* = 87) 92

Table 3.9 *Mean Big Five Accuracy Scores for Males (n* = 40*) and Females (n* = 87) 92

Table 3.10 *Correlation Matrix of Cognitive and Personality Confidence Variables (N= 127)* 93

Table 3.11 *Summary of Exploratory Structural Analysis Results for Cognitive Confidence and Big Five Personality Confidence Scores, Using Principal Axis Factoring with Oblique Rotation (N = 127)* 94

Table 3.12 *Correlations between Cognitive Bias Scores (N = 127)* 95

Table 3.13 *Principal Axis Factoring of Cognitive Bias Scores (N = 127)* 95

Table 4.1 *Pattern Matrix of IPIP Items using Principal Components Analysis with Promax Rotation and Kaiser Normalisation (N =* 405*).* 128

Table 4.2 *Correlations Among the IPIP Subscales*...129

Table 4.3 *Pattern Matrix of IPIP Form B Items using Principal Components Analysis with Promax Rotation and Kaiser Normalisation (N = 405).*...........................130

Table 4.4 *Descriptive Statistics for Accuracy Scores Derived from Methods 1, 2, and 3 (N=405)*..131

Table 4.5 *Descriptive Statistics for all Dependent Variables (N = 405)*..................133

Table 4.6 *Correlations among Block and Item-by-Item Confidence Ratings (N = 405)* 134

Table 4.7 *Principal Axis Factoring of the Item-by-Item and Block Confidence Rating Scores (N = 405)*...135

Table 4.8 *Means for Block and Item-by-Item Big Five Confidence Ratings for Males (N = 85) and Females (N = 318) in Study 2.*...136

Table 4.9 *Correlations among Big Five Block Accuracy Scores (N* = 405)..............138

Table 4.10 *Correlations among Big Five Item-Level Accuracy Scores (N* = 405)....138

Table 4.11 *Correlations among Big Five Block and Item-Level Accuracy Scores (N* = 405) 139

Table 4.12 *Means for Global Block and Item-by-Item Big Five Accuracy Scores for Males (N = 85) and Females (N = 318) in Study 2.*..140

Table 4.13 *Correlations among Personality Bias Scores (N* = 405). 141

Table 4.14 *Summary of Exploratory Structural Analysis Results for Big Five Personality Bias Scores, Using Principal Axis Factoring (N = 405)*................141

Table 4.15 *Correlations among IPIP and IPIP Form B Subscale Scores*.................142

Table 4.16 *Pattern Matrix for all IPIP and IPIP Form B Subscale Scores, Using Principal Components Analysis with Oblique Rotation (N = 405)*..................143

Table 5.1 *Descriptive Statistics for Cognitive Variables (N=243)*...........................173

Table 5.2 *Means of Confidence and Bias Scores for Males (n = 61) and Females (n = 182) for Cognitive Tasks in Study 3.* 177

Table 5.3 *Mean Bias Scores for First and Fourth Quartiles on Cognitive Tasks (N = 243)* 178

Table 5.4 *Correlations among IPIP subscale scores and GF Confidence and Gf Bias and Partial Correlations between Big Five Scores and Confidence and Bias (N = 243)* 179

Table 5.5 *Pattern Matrix of IPIP using Principal Components Analysis with Promax rotation and Kaiser Normalisation (N = 243)* 180

Table 5.6 *Pattern Matrix of IPIP Form B using Principal Components Analysis with Promax rotation and Kaiser Normalisation* (*N = 243)* 182

Table 5.7 *Descriptive Statistics for Accuracy Scores Derived from Methods 1, 2, and 3 (N=243)* 184

Table 5.8 *Descriptive Statistics for Self-Report Dependent Variables (N = 243)* 185

Table 5.9 *Correlations among Block and Item-by-Item Confidence Ratings (N = 243)* 187

Table 5.10 *Principal Axis Factoring of Big Five Item-by-Item and Block Confidence Rating Scores (N = 243)* 187

Table 5.11 *Means for Block and Item-by-Item Big Five Accuracy Scores for Males (n = 61) and Females (n = 182) in Study 3.* 188

Table 5.12 *Means for Block and Item-by-Item Big Five Confidence Ratings for Males (N = 61) and Females (N =182) in Study 3.* 188

Table 5.13 *Means for Item-by-Item Big Five Bias Scores for Males (N = 61) and Females (N = 182) in Study 3.* 189

Table 5.14 *Correlations among Big Five Block Accuracy Scores (N* = 243*).* 191

Table 5.15 *Correlations among Big Five Item-Level Accuracy Scores (N* = 243). ...191

Table 5.16 *Correlations among IPIP and IPIP Form B Subscale Scores* 192

Table 5.17 *Pattern Matrix for all IPIP and IPIP Form B Subscale Scores, Using Principal Components Analysis with Oblique Rotation (N = 243)* 193

Table 5.18 *Correlations between Positive Affect and Personality Confidence Scores (N* = 243) 194

Table 5.19 *Correlations Among all Confidence Variables in Study 3 (N = 243).* 196

Table 5.20 *Principal Axis Factoring with Oblique Rotation all Big Five and all Gf Confidence Scores (N = 243)* 198

Table 5.21 *Correlations among Cognitive and Big Five Bias Scores* 199

Table 5.22 *Principal Axis Factorings of Cognitive and Big Five Bias Scores (N = 243).* 199

Table 6.1 *Accuracy Scores for the BFBD and the IPIP-Form B* 214

Table 6.2 *Confidence Scores for the BFBD and the IPIP-Form B* 216

Table 6.3 *Mean Bias Scores across Studies 2 and 3* 217

List of Figures

Figure 2.1: Diagrammatic representation of the multidimensional, hierarchical model of self-concept........ 44

Figure 3.1. Calibration curve for the Letter Series Task. 79

Figure 3.2. Scatterplot of mean confidence rating and mean accuracy scores for the Letter Series Task. 80

Figure 3.3. Calibration curve for the Cattell’s Matrices Task. 80

Figure 3.4. Scatterplot of mean confidence rating and mean accuracy scores for the Cattell’s Matrices Task. 81

Figure 3.5. Calibration curve for the General Knowledge Task........ 81

Figure 3.6. Scatterplot of mean confidence rating and mean accuracy scores for the General Knowledge Task........ 82

Figure 3.7. Calibration curve for the Esoteric Analogies Task. 82

Figure 3.8. Scatterplot of mean confidence rating and mean accuracy scores for the Esoteric Analogies Task. 83

Figure 3.9. Calibration Curve of mean confidence rating and mean accuracy scores for the general Concealed Words Task. 83

Figure 3.10. Scatterplot of mean confidence rating and mean accuracy scores for the general Concealed Words Task. 84

Figure 5.1. Calibration Curve of mean confidence rating and mean accuracy scores for the Letter Series Task. 173

Figure 5.2. Scatterplot of mean confidence rating and mean accuracy scores for the Letter Series Task. 174

Figure 5.3. Calibration curve for the Esoteric Analogies Task. 174

Figure 5.4. Scatterplot of mean confidence rating and mean accuracy scores for the Esoteric Analogies Task. 175

Figure 5.5. Calibration curve for the Word Association task. 175

Figure 5.6. Scatterplot of mean confidence rating and mean accuracy scores for the Word Association Task. 176

List of Appendices

Appendix A Sign Up Sheet Study 1 249

Appendix B Confidence Ratings for Gf/Gc Quickie Tests 250

Appendix C General Knowledge Test 251

Appendix D Letter Series Test 252

Appendix E Esoteric Analogies Test 253

Appendix F Concealed Words Test 255

Appendix G Cattell’s Matrices 256

Appendix H The International Personality Item Pool Five-Factor Personality Scale 259

Appendix I Big Five Self-Rated Personality Block Descriptions 261

Appendix J Informed Consent for Study 1 266

Appendix K Factor Analysis of IPIP Scales and IPIP Block Descriptors-Study 1 ... 267

Appendix L IPIP Form B 269

Appendix M Informed Consent for Study 2 285

Appendix N Factor Analysis of Big Five Accuracy Scores Study 2 286

Appendix O Word Association Test 289

Appendix P Private Self-Consciousness Scale 292

Appendix Q Need for Cognition (short form) 293

Appendix R Positive and Negative Affect Schedule (PANAS) 295

Appendix S Self-report Intelligence and Confidence Questionnaire (SICQ) 296

Appendix T Covering Page and Informed Consent Study 3 299

Appendix U Additional Factor Analyses for Study 3 300

Chapter 1 – Introduction

1.1 Structure of the Dissertation

This dissertation comprises six chapters. This chapter briefly sets the scene for the next five chapters, provides the reader with an overview of the broad aims 39of this PhD research, and concludes with a tabular overview of the three studies carried out in this dissertation. The research literature is reviewed in Chapter 2. Studies 1, 2, and 3 are covered in Chapters 3, 4, and 5 respectively. Chapter 6 forms the general discussion.

1.2 Statement of the Problem

Accurate self-insight into one's cognitive abilities and personality traits is fundamental to navigating daily life. Realistic self-views of personal strengths and weaknesses in both these areas, for example, allows individuals to make appropriate academic, career, and even relationship decisions, that capitalise on talents and at the same time consider areas of difficulty. Because individuals receive extensive feedback from each other and from their natural environments, accurate self-knowledge should be a relatively easy task for most people. Empirical evidence, however, suggests that people lack self-knowledge in some situations, and that there is considerable variation across individuals (e.g., Ackerman, Beier, & Bowen, 2002; Alicke, 1985; Dunning, 2005; Ehrlinger & Dunning, 2003; Hansford & Hattie, 1982; Kleitman, 2008; Kleitman & Stankov, 2001; Pallier et al., 2002; Stankov, 1998, 1999a, 2000a, 2000b; Stankov & Crawford, 1996a, 1996b, 1997).

Research examining the accuracy of self-insight can be split into three main groups. Within the first body of research, cognitive psychologists have examined correlations between participants' self-report ratings of their intelligence and objective scores on tests or tasks of these intellectual abilities. Participants' self-perceptions of abilities are regarded as being accurate if they are highly correlated with performance measures. Earlier research showed that average correlations between these variables ranged between .20 and .30 suggesting that peoples' perceptions of their abilities are not closely tied to actual performance (for reviews see Hansford & Hattie, 1982; Mabe & West, 1982). Later research paints a slightly more flattering portrait with correlations between self-estimated and tested abilities ranging from .35 to .58 (e.g., Ackerman, 1997; Ackerman et al., 2002; Ackerman, Kanfer, & Goff, 1995; Borkenau & Liebler, 1993b; Paulhus & Morgan, 1997).

Correlational techniques have also been used when examining the accuracy of trait self-reports when making personality judgments. Within the personality domain, judging the accuracy of trait self-ratings is more complicated than in the cognitive area as there are no perfect criteria for evaluating the accuracy of these types of judgements (Colvin & Funder, 1991). Inter-judge agreement (consensus) and self-other agreement are the most commonly used criteria for determining the accuracy of trait self-reports. In both types of studies, low to moderate levels of agreement (i.e., self-peer or peer-peer ratings of personality) have been obtained, with mean correlation co-efficients being in the order of .30 (e.g., Funder, 1999; Funder & Colvin, 1997; Funder, Kolar, & Blackman, 1995; John & Robins, 1993; Meyer et al., 2001). Importantly, however, consensus does not necessarily imply accuracy (Blackman & Funder, 1998; Swan & Gill, 1997).

There has been much debate within the psychological literature on how best to operationalise personality accuracy given the complex methodological issues that confront researchers (cf. Funder, 1999). This debate and attendant methodological challenges are reviewed in Chapter 2. The situation can be likened to the long delay that preceded the emergence of the emotional intelligence (EI) construct where researchers grappled for almost a century with the problem of objectively estimating people's awareness of their emotional states. In the end, progress was made when researchers showed a willingness to experiment with innovative measurement methods and to tolerate vigorous challenges to the reliability and validity of their instruments. A major aim of this dissertation was to develop and test various accuracy methods for use with personality judgments within Goldberg's (1997) Five-Factor taxonomic framework of personality structure.

The second body of psychological enquiry encompasses those studies that have investigated the above-average effect, also coined comparative ability mis-calibration. In these studies, participants have typically compared their own abilities with those of their peers. For example, university students may be asked to compare their own academic performance relative to their classmates by endorsing a percentile rank from 0 (*I'm at the very bottom*) to 50 (*I'm exactly average*) to 99 (*I'm at the very top*). Within the cognitive domain, the general findings indicate that individuals in the bottom quartile for various cognitive tasks overestimate themselves as being above average, and those in the top quartile underestimate their performance (e.g., Dunning, Johnson, Ehrlinger, & Kruger, 2003; Kruger, 1999; Kruger & Dunning, 1999).

The third area of empirical enquiry that investigates the accuracy of self-insight encompasses the overconfidence studies wherein individuals provide or select answers to various cognitive test items and then indicate how confident they are that each answer is correct. In general, the research findings suggest that people are overconfident regarding their abilities to quite a large degree. There are instances in the literature where expressed confidence in cognitive performance matches objective accuracy and therefore confidence ratings are realistic/accurate (e.g., Baker, 2001; Keren, 1987;

Stankov & Crawford, 1996a, 1996b, 1997; Stankov & Lee, 2008; Tomassini, Solomon, Romney, & Krostad, 1982; Winman, Juslin, & Bjorkman, 1998). There are also instances where underconfidence has been displayed. However, the evidence for the tendency of people to be overconfident regarding their abilities and skills, is quite substantial (e.g., Allwood & Granhag, 1996; Juslin, 1994; Kleitman & Stankov, 2001; Stankov, 1998; Stankov & Crawford, 1996a, 1996b, 1997; Stankov & Lee, 2008). When individuals are either underconfident or overconfident regarding their skills and abilities, their insight is regarded as being inaccurate (or mis-calibrated). The main overall aim of this dissertation was to contribute to this body of knowledge by extending calibration research into the domain of personality judgments.

1.3 Brief Rationale for Including Personality Judgments

Although investigation of the confidence-accuracy relationship regarding accurate self-monitoring of cognitive performance is an important area of psychological inquiry, Koehler, Brenner, Liberman and Tversky (1996) assert that investigation of intuitive personality judgments is equally important to personality, clinical, organizational, and social psychologists (John & Robins, 1994). For example, within clinical psychology, inaccurate self-insight underlies various psychological disorders such as depression and personality dysfunction (e.g., Dimaggio et al., 2005; Dimaggio, Semerari, Carcione, Procacci, & Nicolo, 2006). On a practical level, accurate personality judgements are vital because they impact on daily life (e.g., do my personality traits suit the job I am applying for), and influence our behaviour in terms of how we think and feel about ourselves and others (e.g., would I choose this person as a partner, or would our personalities clash). To date, however, researchers have not investigated whether mis-calibration occurs within the personality domain. This lack of research is understandable given that determining personality accuracy and mis-calibration is a complex process that is difficult to operationalise. Nevertheless, as noted earlier, one of the major aims of this dissertation was to develop various accuracy protocols (see Chapters 3 and 4) and use them to examine personality mis-calibration (see Chapters 4 and 5). The calibration paradigm (see Chapter 2) was used to assess the confidence-accuracy relationship (i.e., self-monitoring) within both the cognitive and personality arenas.

1.4 Broad Aims and Overview of Current Studies

This dissertation has several broad aims:

1. To investigate mis-calibration within the cognitive realm.
2. To investigate self-confidence and mis-calibration within personality judgments.
3. To investigate the impact of individual differences (e.g., age, need for cognition) variables on cognitive confidence and mis-calibration.
4. To investigate the impact of individual difference variables (e.g., age, private self consciousness, affect) on personality confidence judgments.
5. To investigate the factorial structure of both Big Five Confidence and Big rating Scales.
6. To investigate the factorial structure of both cognitive and personality confidence judgments.
7. To investigate the factorial structure of both cognitive and personality bias judgments.

In order to achieve these aims, three studies were conducted with a total of 831 individuals being tested. An overview of these studies is presented in Table 1.1. In Studies 1 and 3, participants responded to both cognitive and personality measures. The cognitive tasks selected for Studies 1 and 3 were based on Horn and Cattell’s (1966) theory of Fluid and Crystallised Intelligence. In Study 2, only personality measures were used. In Studies 1, 2 and 3, personality judgments were made within the Five-Factor taxonomic framework of personality structure (Goldberg, 1997).

Table 1.1
Brief Overview of Studies

Study 1	Study 2	Study 3
Cognitive and Big Five confidence and bias judgments	Big Five confidence and bias judgments only	Cognitive and Big Five confidence and bias judgments
Face-to-face testing (N = 127)	Web based testing (N= 452)	Web based testing (N= 252)
Examine mis-calibration across Gf, Gc, & Gv tasks		Examine mis-calibration with Gf tasks
Investigate whether individual differences in gender, age, personality and ability impacts on cognitive confidence and mis-calibration		Investigate whether individual differences in gender, age, personality, ability, need for cognition (NFC), and affect impacts on cognitive confidence
Trial method 1 for determining Big Five accuracy	Trial various methods of assessing Big Five accuracy. Obtained Big Five confidence judgments	
Investigate whether individual differences in gender and age impacts on Big Five accuracy, and confidence	Investigate whether individual differences in gender and age impacts on Big Five accuracy, confidence, and mis-calibration	Investigate whether individual differences in gender, age, private self-consciousness, affect, and NFC impacts on Big Five confidence.
	Investigate the factorial structure of item-by-item and Block Big Five confidence judgments	Investigate the factorial structure of item-by-item and block Big Five confidence judgments
	Investigate factorial structure of Big Five confidence and Big Five rating scales	Investigate factorial structure of Big Five confidence and Big Five rating scales
Investigate the factorial structure of cognitive and personality confidence judgments		Investigate the factorial structure of cognitive, personality, and self-report ability confidence judgments
Investigate the factorial structure of cognitive bias scores	Investigate the factorial structure of Big Five bias scores	Investigate the factorial structure of cognitive and Big Five bias scores

Chapter 2 – Literature Review

2.1 Introduction

This chapter is divided into six subsections. The first section defines self-confidence within a meta-cognitive framework. The calibration paradigm and numerical and graphical measures of calibration are discussed in section two. Theoretical explanations of cognitive mis-calibration follow. Section four reviews the calibration research literature. Justification for extending calibration research into the domain of personality is addressed in section five. Measurement issues and empirical findings in the personality judgment literature follow in section six.

2.2 Confidence and Meta-cognition

Item-by-item confidence judgments which were examined in the current studies are thought to represent the important meta-cognitive process of self-monitoring (Kleitman & Stankov, 2007; Stankov, 1999b). Meta-cognition refers to cognition about cognition, and is generally considered to have two elements: knowledge about one's own cognition, and the regulatory sub-processes of meta-cognitive control (Brown, 1987; Schraw & Dennison, 1994; Schraw & Moshman, 1995). Five regulatory sub-processes have been mentioned in the literature: planning, information management, monitoring, debugging, and evaluation (e.g., Artz & Armour-Thomas, 1992). Schraw and Dennison (1994) operationalised these sub-processes as follows:

1. Planning: planning, goal setting, and allocating resources prior to learning.
2. Information management: skills and strategy sequences used on-line to process information more efficiently (e.g., organizing, elaborating, summarizing, selective focusing).

3. Monitoring: assessment of one's learning or strategy use.
4. Debugging: strategies used to correct comprehensions and performance errors.
5. Evaluation: analysis of performance and strategy effectiveness after a learning episode (pp. 474-475).

It is generally argued that the sub-processes of planning, monitoring, and evaluation are important variables in explaining effective learning (Flavell, 1977, 1987), and therefore are an important area of psychological enquiry. Within differential psychology in Australia, researchers have been studying the processes of self-monitoring and evaluation using confidence judgments obtained from tests of human cognitive abilities (e.g., Kleitman & Stankov, 2001; Pallier, 2003; Pallier et al., 2002; Stankov, 1999a, 1999b; Stankov & Crawford, 1996a, 1996b). The current studies will continue this line of enquiry. This area of research is important not only because accurate self-assessment or self-monitoring of our cognitive performance is a fundamental aspect of successful learning (Flavell, 1977; Schraw, Crippen, & Hartley, 2006; Sternberg, 1997a, 1997b), it is also vital for effectual decision-making (Moore & Healy, 2008), and is critical in many work situations. For example, it is crucial that physicians accurately diagnose their patients' illnesses, are confident about their decisions, and are able to evaluate the accuracy of their work. The current research programme also takes the calibration paradigm into new territory by applying it to personality judgements. The importance of this extension will be addressed in a later section. Before introducing the work undertaken in this thesis, the calibration paradigm, and work based on that paradigm will be described.

2.3 Calibration Research

2.3.1 The Experimental Paradigm

Calibration studies evaluate the correctness of individuals' subjective probability ratings, or confidence in their judgements and predictions (Keren, 1991). Self-monitoring is operationalised by self-confidence scores; that is, individuals are asked to express how confident they are in their judgements, answers, or predictions. The paradigm has been discussed at length in the literature and is therefore not repeated here (see Harvey, 1997; Kleitman, 2008; Stankov & Kleitman, 2008).

2.3.2 Numerical Measures of Calibration

2.3.2.1 Item-by-Item Confidence Scores (Self-Monitoring)

There are various ways to evaluate the realism of obtained confidence ratings, with a number of studies having used Brier's quadratic scoring rule (cf. Keren, 1991). However the psychometric properties of the scores obtained from this rule have been shown to be inadequate (Stankov & Crawford, 1996a). Other measures such as signal detection theory and the confidence-judgment accuracy quotient have also been considered inadequate measures of the confidence-accuracy relationship (see Keren, 1991). The simplest and most reliable calibration measure is the bias score, which is the average confidence rating minus the proportion correct score across all items in a task (Stankov & Crawford). A positive bias score suggests overconfidence, whereas a negative score indicates underconfidence. A bias score of greater than ± 10 indicates marked under-or-overconfidence (Stankov, 2000a), if it falls in the range of ± 5 then it represents reasonable calibration (Stankov, 1999b). Bias scores were generated for both the cognitive (Studies 1 & 3) and personality measures (see Studies 2 & 3) used in the current research. Throughout this dissertation, the terms bias and mis-calibration reflect the discrepancy between confidence and accuracy and were used interchangeably.

2.3.2.2 Self Evaluation Scores (Mis-calibration at the Global Level)

The literature also points to the distinction between the meta-cognitive processes of self-monitoring (i.e., item-by-item confidence judgments) and self-evaluation (Schraw & Moshman, 1995). This evaluative aspect of meta-cognition has been measured by asking participants to estimate the proportion of items they have solved correctly after finishing the test (i.e., post test performance estimate or PTPE). A bias score that is similar to the one reported earlier can also be obtained using the PTPE score (Stankov & Crawford, 1996b). This bias score requires that the actual mean percentage of correct responses is subtracted from the estimated percentage of correctly solved items. A negative value PTPE bias score represents under-evaluation whereas a positive value indicates over-evaluation. These bias scores will be calculated for the cognitive tasks in Study 3.

2.3.3 Graphical Measures of Calibration

2.3.3.1 Calibration Curves

Calibration curves are the most common method of representing the relationship between the proportion of correct responses (i.e., accuracy or objective probability estimates) and participants’ confidence ratings. These curves depict the proportion of correct responses associated with various confidence intervals, and are useful for displaying general trends in data sets. When using calibration curves, it is helpful to display the number of observations at each point on the curve as interpretation of calibration curves is difficult when the curves are located in both the underconfidence and overconfidence regions of the graph (Keren, 1991). Frequency-weighted curves are a useful aid, and were inspected for all cognitive tasks used in Studies 1 and 3.

2.3.3.2 Item-Specific Scatterplots

In a further methodological development, Stankov (1999a) argued that calibration curves fail to provide information about individual items that may have contributed to mis-calibration for a particular task. He advocated that item-specific information for each cognitive task be depicted graphically by producing plots of the mean confidence ratings and proportion correct scores (i.e., item difficulties) for each item in a given task. For items that are well calibrated, the proportion correct and mean confidence ratings should be the same. Wide separation of the points indicates mis-calibration for that item. Item-specific plots were produced in addition to the frequency-weighted calibration curves described previously for all the cognitive tasks in the present studies.

2.4 Theories of Cognitive Mis-calibration

Four different theoretical perspectives have sought to explain underconfidence and overconfidence effects in the cognitive domain. They are a) heuristics and biases, b) error, c) differing modes of uncertainty, and d) ecological accounts of mis-calibration respectively.

2.4.1 Heuristics and Biases Approach

According to this approach, confidence judgments in the accuracy of one’s performance (or decisions) are said to follow Bayesian laws of probability and rational decisions follow the axioms of probability theory (see Gilovich, Griffin, & Kahneman, 2002 for a review). Mis-calibration from this perspective reflects a cognitive bias that represents an example of an irrational decision making process (Stankov & Kleitman, 2008). Gigerenzer and colleagues, however, questioned the appropriateness of applying Bayes rule to confidence judgments and mis-calibration (Gigerenzer, 1991, 1992, 1993, 1996a, 1996b, 2000; Gigerenzer, Hoffrage, & Kleinbolting, 1991; Gigerenzer & Murray, 1994). More recently, within differential psychology, Kleitman (2003) investigated whether participants in her studies violated the additivity rule of probability theory. That is, where the sum of probability judgments made about mutually exclusive events (i.e., confidence ratings in her work) should equal one (or 100% on a confidence rating scale). Kleitman defined subadditivity as being when the sum of confidence ratings fell below 100% and superadditivity when the sum of confidence ratings exceeded 100%. Results showed that 60% of her participants deviated from the additivity rule. This finding suggests that a significant number of participants did not use the principles of probability when providing confidence ratings in their answers.

According to the heuristics and biases approach, mis-calibration is due to information processing biases such as heuristics (Brenner, Koehler, Liberman, & Tversky, 1996; Kahneman, Slovic, & Tversky, 1982; Keren, 1991; Klayman, Soll, González-Vallejo, & Barlas, 1999; Koehler, 1994; Koriat, Lichtenstein, & Fischhoff, 1980). Heuristics are the cognitive short-cuts that individuals use when estimating probabilities associated with various problems (Tversky & Kahneman, 1983). These short-cuts can lead to errors because people fail to consider all relevant information. Availability, representativeness, anchoring and adjustment heuristics are some examples mentioned in the research literature (Kahneman et al., 1982). From the information processing perspective, mis-calibration occurs when individuals use simplifying heuristics that lead to wrong answers on cognitive problems, and confidence in those judgments is based on natural assessments that disregard the possibility of an alternative correct response. Overconfidence is predicted if people strongly believe that a response alternative is correct despite being based on a small amount of information. Underconfidence is predicted if individuals weakly believe that a response alternative is correct, despite being based on a large amount of information (Griffin & Tversky, 1992). An example problem from Kahneman and Tversky (1973) is presented below:

> A panel of psychologists have interviewed and administered personality tests to 30 engineers and 70 lawyers, all successful in their respective fields. On the basis of this information, thumbnail descriptions of the 30 engineers and 70 lawyers have been written…For each description, please indicate your probability that the person described is an engineer on a scale from 0 to 100 (p. 241).
>
> Jack is a 45-year-old man. He is married and has four children. He is generally conservative, careful, and ambitious. He shows no interest in political science and spends most of his free time on his many hobbies which include home carpentry, sailing, and mathematical puzzles.
>
> The probability that this man is one of the 30 engineers in the sample of 100 is _____% (p. 241).

Participants in this study were more likely to choose engineer even if the ratio of engineers to lawyers was changed to 70 versus 30. According to probability theory, however, if one takes into consideration the population base rates of lawyers and engineers, it is more likely that Jack was a lawyer. Participants did not consider this when they made their probability judgments. It was therefore argued by Kahneman and Tversky (1973) that participants used the representative heuristic, in that the personality description appeared to be more representative of an engineer than the lawyer.

2.4.2 Error Explanations of Mis-calibration

Error explanations of mis-calibration, on the other hand, stress the importance of random error (e.g., attentional lapses, fatigue, memory lapses) as a contributing factor to biased confidence judgments. According to this position, confidence judgments comprise a true judgment component that is based on cue validities, as well as random error (e.g., Soll, 1996). Soll suggested that random error can affect judgments in multiple ways. Cognitive inconsistency is the first consideration in terms of error. That is, in the absence of practice effects, if an individual made several judgments about the same item over time, the judgments may differ because of idiosyncratic random errors (Soll, 1996). Random error can also occur if an individual has limited ecological experience with the problem or task presented by experimenter. For example, a doctor is presented with a diagnostic vignette and asked to make a diagnosis, and then provide a confidence rating indicating how confident he or she is that the diagnosis is correct. If the doctor is experienced with the particular disease and draws upon this information, the random error component decreases. The opposite is true if the doctor has limited information. Therefore, if random error is great then under-or-over confidence is the likely result (Budescu, Wallsten, & Au, 1997; Erev, Wallsten, & Budescu, 1994). By contrast, another group of researchers asserted that mis-calibration results when differing types of uncertainty influence the judgment process.

2.4.3 Differing Modes of Uncertainty

Juslin and Olsson (1997) argued that poor calibration is due to differing modes of uncertainty influencing the judgement process. Specifically, they suggested that mis-calibration could be due to Thurstonian (after L. L. Thurstone) or Brunswikian (after Egon Brunswick) modes of uncertainty affecting the judgement process. These researchers presented a computational model of confidence in sensory discrimination tasks that involved paired comparisons. Juslin and Olsson contended that the uncertainty which underlies confidence judgements is the result of two factors: noise in the nervous system or incomplete states of knowledge. In other words the uncertainty that underlies perceptual tasks differs from that of other cognitive tasks. For example, a participant is required to discriminate between lifted weights to determine which weight is heavier. Mis-calibration in this case can be attributed to Thurstonian uncertainty (i.e., noise in the sensory system). When an individual provides an answer to a general knowledge question such as, "Which city hosted the Winter Olympic Games in the year 2006?" mis-calibration can be attributed to Brunswikian uncertainty (i.e., incomplete states of knowledge).

According to this account, underconfidence is expected with all perceptual tasks because of sensory noise. However, this expectation is not supported in the research literature as Stankov (Stankov, 1999a) provided evidence where overconfidence was apparent for perceptual tasks within the auditory, olfactory, tactile and gustatory modalities. Similarly, in Pallier et al.'s (2002) research, overconfidence was apparent for the visual perceptual Hidden Figures and Concealed words tasks used in their work.

2.4.4 Ecological Approach

The Probabilistic Mental Model (PMM) originally formulated by Gigerenzer et al. (1991), and later developed by Juslin (1993, 1994), explains mis-calibration in terms of ecological factors that are external to the individual. The fundamental premise of this model is that individuals learn the validities of their own environmental knowledge cues and use these cues when making judgments in their natural environments. That is, individuals have knowledge of both the relative frequencies of effective cues within their natural environments, as well as the numerical values of those validities, and memory provides an excellent store of frequency information from numerous natural environments. Moreover, encoding of frequency information is both automatic and requires minimal attentional resources or effort.

PMM theory (Gigerenzer et al., 1991) argues that mis-calibration results from researchers selecting general knowledge items that are tricky and unrepresentative of the real world. Gigerenzer et al. (1991) reasoned that if researchers selected a representative sample of items, then the overconfidence bias that is apparent with general knowledge tasks is likely to disappear, given that individuals are well calibrated to their natural environments. Thus, these theorists have argued that individuals' knowledge about their natural environments helps them generate cues, which, in turn, are used to answer general knowledge test items. Therefore, mis-calibration occurs when there is a mismatch between the cue validity on a general knowledge test item and its respective ecological validity. In other words, individuals are accurate self-monitors if the answers to general knowledge questions are already stored in their long-term knowledge structures (i.e., a local mental model), or if they can use inductive inferences to solve general knowledge problems (i.e., a probabilistic mental model). The important qualifier, however, is that response options must be ecologically valid (Gigerenzer et al., 1991).

PMM theory (Gigerenzer et al., 1991) also distinguishes between the item-by-item confidence judgments that are given during a cognitive test and the post-test percentage correct estimates that are given at the end of the test. In the latter case, individuals estimate the percentage of items they think they answered correctly. According to ecological theory, these post-test judgments should not correlate with the item-by-item confidence ratings because their respective cue and ecological validities are different. For example, if an individual was asked to estimate how many general knowledge items about Australia he or she answered correctly, then ecological validity may be influenced by base rates of correct performance on previous tests of a similar nature whereas response to a single item about the largest state in Australia would be influenced by ecological validities such as population estimates, geographical size, and so forth. Therefore, it was concluded that item-by-item confidence judgments and global post test percentage correct estimates, are not subserved by the same cognitive processes. Usually, mean post-test estimates are smaller than mean confidence ratings for a test and often display better calibration, or even under-evaluation, compared to confidence bias measures (Gigerenzer et al., 1991; Schneider, 1995; Stankov & Crawford, 1996a, 1996b). When this occurs, it is called the confidence/frequency effect (Gigerenzer et al., 1991).

Empirical evidence in favour of PMM theory has been provided by Gigerenzer (1991), and Juslin (1993, 1994). In these studies, randomly chosen general knowledge questions resulted in good calibration, whereas typical general knowledge questions resulted in overconfidence. Partial support for PMM theory in terms of the distinction between mean item-by-item confidence ratings and post-test performance estimates has also been provided by Stankov and his colleagues where at the factorial level separate factors were obtained representing confidence and post-test percentage correct estimates respectively (Kleitman & Stankov, 2001; Stankov & Crawford, 1996a, 1996b). For example, Kleitman and Stankov obtained both item-by-item confidence ratings and post-test percentage correct judgments for three cognitive tests (i.e., Geography, Raven's Progressive Matrices, and Line Length) and, as predicted by PMM theory, there was a split at the factorial level between these two types of judgments, indicating support for the contention that differing cognitive processes are at play. Other research challenges this theory as overconfidence was demonstrated on tasks where items were selected at random from a representative set of items (Brenner et al., 1996; Griffin & Tversky, 1992; Liberman, 2004; Soll, 1996).

This approach has largely ignored other important factors such as individual differences (cf. Kleitman, 2003, 2008; Stankov & Kleitman, 2008). Moreover, PMM theory (Gigerenzer et al., 1991) predicates that confidence ratings from differing cognitive domains should not be consistently correlated since their respective ecological and test cue validities are likely to be different. However, there exists a large body of research where consistent correlations have been found between confidence ratings from differing cognitive domains (e.g., Baker, 2001; Kleitman, 2003, 2008; Kleitman & Stankov, 2001; Pallier et al., 2002; Stankov, 1999b, 2000a; Stankov & Kleitman, 2008).

Given the underlying assumption of PMM theory (Gigerenzer et al., 1991) is that ecological validity affects mis-calibration, it is reasonable to investigate whether individuals are well-calibrated with judgments that are made on a daily basis. Personality judgments are a case in point as they are implicated in many aspects of daily life (see sections 2.6, 2.7 and 2.8 of this chapter for details). In terms of the personality confidence judgments, it could be argued that Gigerenzer's model offers insight with regard to personality mis-calibration. Specifically, the foundation of PMM theory (Gigerenzer et al., 1991) is that individuals are well calibrated to their natural environments and calibration can be expected where ecological validity is high. It will be argued that personality judgments are of high ecological validity in today's society and will therefore exhibit the good calibration predicted by PMM theory (Gigerenzer et al., 1991).

2.4.5 Summary

In this section four theories of mis-calibration were reviewed. The empirical literature provides evidence both for and against each of theories. None of the theories in isolation, however, can explain all calibration research findings. Moreover, examination of individual differences has usually been ignored by the theories discussed above. The individual differences approach to the study of confidence and mis-calibration has not been neglected, however, and accounts for a large proportion of recent calibration research. The next section reviews what we have learned so far, starting with the hard-easy effect.

2.5 Empirical Findings in Calibration Research

2.5.1 The Hard-Easy Effect

The hard-easy effect has been demonstrated with cognitive tasks wherein individuals appear to exhibit higher levels of overconfidence for difficult tasks and underconfidence or good calibration for easy tasks (Baranski & Petrusic, 1994; Harvey, 1997; Juslin & Olsson, 1997; Keren, 1991). The relationship between task difficulty and underconfidence or overconfidence has been coined the calibration difficulty effect (Griffin & Tversky, 1992) or the hard-easy effect (Gigerenzer et al., 1991). This effect has been demonstrated with vocabulary and general knowledge tasks (e.g., Koriat et al., 1980; Lichtenstein & Fischhoff, 1977; Pulford & Colman, 1997; Schraw & DeBacker Roedel, 1994), with visual perceptual tasks such as the Line Length task (Baranski & Petrusic, 1994; Stankov, 1999a), and with discrimination of American and European handwriting (Lichtenstein & Fischhoff, 1977).

2.5.2 Good Calibration

Winman, Juslin, and Bjorkman (1998) found that individuals were well-calibrated on various tasks that required hindsight bias (i.e., biased judgments of past events after the outcomes of these events are known). Conversely, Granhag, Stromwall, and Allwood (2000) found overconfidence on a task that required hindsight bias, where questions were asked about a filmed kidnapping.

Stankov (1999a) and his collaborators (Kleitman & Stankov, 2001; Pallier et al., 2002, Study 2; Stankov, 1998; Stankov & Crawford, 1996b, 1997) demonstrated that individuals were reasonably well calibrated on the Raven’s Progressive Matrices Test. This test would be novel to participants and perhaps greater attentional resources were given to this task leading to better calibration. However, the literature contains instances where overconfidence was demonstrated for Ravens Progressive Matrices (Pallier et al., 2002, Study 1; Stankov & Dolph, 2000). Random errors (e.g., motivational factors) could explain this inconsistency in Pallier’s first study. That is, military participants may have tried to reduce cognitive dissonance (e.g., “I believe I am bright, therefore my answer must be correct”) by elevating their confidence ratings. This explanation makes sense as these participants were overconfident across a diverse range of cognitive tasks.

Good calibration has also been demonstrated on digit span (Baker, 2001; Crawford & Stankov, 1996; Stankov & Crawford, 1996a, 1996b), and visual memory spatial tasks (Baker, 2001). Good calibration for the memory tasks could occur because all of the information required for task solution was within the span of immediate awareness facilitating better self-monitoring.

Expertise in various domains also leads to better calibration, with good calibration demonstrated by experts in the domains of bridge playing (Keren, 1987), weather forecasting (Murphy & Winkler, 1977), and accountancy (Tomassini et al., 1982). In the sporting domains of golf (Fogarty & Else, 2005) and tennis, however, expertise has not always led to better calibration (Fogarty & Ross, 2007).

2.5.3 Underconfidence

Research indicates that people tend to be underconfident when answering questions about future events (Vreugdenhil & Koele, 1988), when responding to sensory discrimination tasks (Bjorkman, Juslin, & Winman, 1993; Juslin, 1994), with visual perceptual tasks such as discriminating between the length of various lines (Crawford & Stankov, 1996; Kleitman & Stankov, 2001; Pallier et al., 2002, Study 2; Petrusic & Baranski, 1997; Stankov & Crawford, 1996a, 1996b, 1997), and with unfamiliar computer-based tasks (Briggs, Burford, & Dracup, 1998). Moreover, Bjorkman et al. (1993) contended that underconfidence is a pervasive phenomenon in tasks of sensory discrimination. Two theories provide possible explanations for these findings. From the ecological stand point, the aforementioned tasks would not be representative of participants' natural ecology thereby leading to mis-calibration (i.e., underconfidence in these instances). Additionally for the sensory discrimination tasks, noise in the sensory system may have led to mis-calibration.

Stankov (1999a), however, investigated whether underconfidence generalised to other sensory modalities (i.e., auditory, kinaesthetic, gustatory, and olfactory). Another aim of Stankov's research was to investigate whether the tendency of participants to be underconfident on the Line Length task also occurred with other visual perceptual tasks, such as the Square Gap task and the Muller-Lyer Illusion. Only a Tactile Texture and a Line Length task displayed underconfidence. Contrary to expectations, participants were well calibrated on the Square Gap task and were overconfident on the Muller-Lyer Illusion. Also, the weight, gustatory, and olfactory tasks displayed overconfidence, whereas a pitch task displayed perfect calibration. Stankov's overall conclusion was that the pervasive underconfidence effect found with the Line Length task does not generalise to perceptual tasks in other sensory modalities or to the other visual perceptual tasks used in his study. A similar conclusion was reached by Pallier et al. (2002, Study 2) where participants were presented with a wide range of cognitive and perceptual (both visual and sensory) tasks. Two visualization tasks in Study 1 of Pallier's research, however, displayed overconfidence (i.e., the Hidden Figures Task and the Concealed Words Task) when administered to military participants. As noted earlier, this sample may be have been motivated to reduce cognitive dissonance by providing higher confidence ratings thereby resulting in overconfidence. Study 1 investigates whether participants are mis-calibrated on a similar task to that used by Pallier (i.e., a Concealed Words task). Study 1 will contribute further information as to whether underconfidence generalises from the Line Length task to another visual perceptual task.

2.5.4 Overconfidence

In general, research suggests that individuals are often more confident than they are accurate. Overconfidence has been displayed by many different participant groups, including physicians (Christensen-Szalanski & Bushyhead, 1981), medical students (O'Keefe, Wildemuth, & Freidman, 1999), clinical psychologists (Oskamp, 1965), engineers (Kidd, 1970), lawyers (Wagenaar & Keren, 1985), United States Air Force recruits (Pallier et al., 2002), university students (Granhag et al., 2000; Kleitman & Stankov, 2001; Pallier et al., 2002; Pulford & Colman, 1997; Renner & Renner, 2001; Schaefer, Williams, Goodie, & Campbell, 2004; Stankov, 1998, 1999b; Stankov & Crawford, 1997), CIA analysts (Cambridge & Shreckengost, 1978, as cited in Sharp, Cutler, & Penrod, 1988), adolescents (Newman, 1984), children (Allwood, Granhag, & Jonsson, 2006; Newman & Wick, 1987), and tennis players (Fogarty & Ross, 2007).

Overconfidence has also been demonstrated across numerous domains and tasks including: prediction of sports outcomes (Fogarty & Else, 2005; Fogarty, Graham, &

Else, 2001; Ronis & Yates, 1987), prediction of outcomes of past events (Lichtenstein, Fischhoff, & Phillips, 1982), assessment of reading skills (Glenberg & Epstein, 1987; Lin & Zabrucky, 1998), marketing management predictions (Mahajan, 1992), categorical judgement tasks (Schneider, 1995), motor task performance (West & Stanovich, 1997), eye witness memory (Bornstein & Zickafoose, 1999), economic forecasts (Braun & Yaniv, 1992), hindsight bias (Granhag et al., 2000), psychology course related quizzes (Renner & Renner, 2001), tennis knowledge (Fogarty & Ross, 2007), tennis rules (Fogarty & Ross, 2007), and with vocabulary and general knowledge tasks (Allwood & Granhag, 1996; Crawford & Stankov, 1996; Juslin, 1994; Kleitman & Stankov, 2001; Schaefer et al., 2004; Stankov, 1998, 1999b; Stankov & Crawford, 1996a, 1996b, 1997; West & Stanovich, 1997). Essentially, the aforementioned research suggests that the overconfidence phenomenon is a robust finding that has demonstrated external validity across numerous participant groups and across a number of different domains. Some researchers have interpreted overconfidence as a pervasive psychological bias (e.g., Baron, 1994). A general knowledge task has been included in Study 1 and it is expected that participants will be overconfident.

The literature reviewed thus far has focussed on examination of group differences (i.e., calibration curves and bias scores) as opposed to the examination of individual differences, which are discussed next.

2.5.5 Self-Confidence and the Study of Individual Differences

From the individual differences perspective, confidence judgments represent the important meta-cognitive process of self-monitoring, an appraisal process whereby individuals evaluate the accuracy of their performance whilst working through psychological tests items (e.g., Kleitman & Stankov, 2001; Stankov & Crawford, 1996a, 1996b). Furthermore, West and Stanovich (1997) highlighted the fact that there are consistent inter-individual differences apparent in the confidence literature. For example, researchers have found that males tend to be more confident than females on cognitive tasks (e.g., Baker, 2001; Pallier, 2003; Pallier et al., 2002), experts are more confident than novices (Spence, 1996), and older children are more confident than younger children on counting tasks (Newman & Wick, 1987).

The following sections review gender, age and personality differences in terms of accuracy, self-monitoring, and mis-calibration. These are the areas that are explored in this dissertation.

2.5.5.1 Gender Differences in Accuracy, Confidence and Mis-calibration

2.5.5.1.1 Gender Differences and Task Accuracy

Are females more intelligent than males? Halperin and LaMay (2000) tried to answer this question in their critical review of gender differences in the intelligence literature. They concluded that males did not differ from females in terms of general intelligence. However, gender differences appeared for visualisation tasks and for verbal tasks that required retrieval from long-term memory. Males displayed an advantage for the former, and females for the latter. Gender differences in accuracy are not elaborated on in any further detail as it is not the focus of this dissertation.

Rather, the current doctoral research sought to extend the examination of gender differences from the cognitive accuracy domain, an area that has been well reported, into the domain of personality accuracy. Studies 1 and 2 addressed this issue as various accuracy methods were developed and tested.

2.5.5.1.2 Gender Differences and Confidence

To date, empirical research indicates that males are more confident than females on: General Knowledge (Pallier, 2003; Pulford & Colman, 1997), Vocabulary (Stankov, 1998), Line Length (Pallier, 2003; Stankov, 1998), Letter Series (Baker, 2001; Pallier, 2003), Ravens Progressive Matrices (Stankov, 1998), Working Memory (Baker, 2001), Cattell's Matrices (Pallier, 2003), golf tasks (Graham, 2006), and with tests of tennis knowledge (Ross & Fogarty, 2006). Gender-stereotypic socialization patterns provide one explanation for the elevated confidence in males. Within the self-concept literature, it has been reported that males hold more favourable maths and science self-concepts than do females in accordance with the gender stereotype that males will be more proficient at maths and science type questions (for a review see Marsh & Yeung, 1998). It could also be the case that parental socialisation patterns engender higher levels of confidence in males across a range of cognitive abilities thereby explaining why males were more confident than females on the tasks used by the calibration researchers mentioned above. Similarly, if tasks are perceived by participants as being masculine gender-typed (e.g., a quiz on sports figures) or gender-neutral then males also tend to be more confident than females (Beyer, 1990). For example, Ross and Fogarty (2006) found gender differences with females being less confident than males on a test of tennis knowledge, a task that would be considered gender-neutral to tennis trainees. Similarly, in Beyer's (1990) study, males were more confident on some of the gender-neutral (e.g., anagrams) tasks than were their females counterparts. If, however, tasks were perceived as being feminine gender-typed then gender differences disappeared. More recently, however, Stankov and Lee (2008) did not find gender differences in confidence on the Reading and Listening sections of the test of English as a Foreign

Language Internet-Based Test (TOEFLiBT). Studies 1 and 3 explored gender differences for the cognitive tasks used in those studies.

However, the current research programme differs from other calibration research in a further significant aspect: by examining gender differences in Big Five confidence judgments. A discussion of this investigation, carried out in Studies 1, 2, and 3, is deferred to later chapters.

2.5.5.1.3 Gender Differences and Mis-Calibration

Previous research findings of gender differences in mis-calibration in the cognitive arena have been somewhat mixed (Baker, 2001; Beyer & Bowden, 1997; Crawford & Stankov, 1996; Jonsson & Allwood, 2003; Pallier, 2003; Pulford & Colman, 1997; Stankov, 1998; Stankov & Crawford, 1997; Stankov & Lee, 2008). In Baker's study, males were significantly more mis-calibrated than females on only three of the five working memory tasks used in her test battery (i.e., Digits Backward, Visual Memory, and Letter-Number-Sequencing tasks). Pallier, on the other hand, conducted two studies, with Study 1 comprising young adults (mean age = 19.81, N = 185) and Study 2 consisting of older adults (mean age = 22, range 17 to 80, N = 303). Results for Study 1 indicated that males were both significantly more confident and mis-calibrated than their female counterparts on General Knowledge and Line Length tasks. Similarly, Pulford and Colman found that males were more mis-calibrated than females for General Knowledge questions. Beyer and Bowden, however, did not find similar results with sport's trivia questions and females were more mis-calibrated than males. This finding makes sense if females perceived sport's questions as being masculine gender-typed, then lowered their efforts accordingly legitimising that poorer performance results from decreased effort, not lower ability.

In the second study by Pallier (2003), older participants responded to two crystallised intelligence tasks (i.e., General Knowledge and a Synonyms Vocabulary test) and two fluid intelligence tasks (i.e., Letter Series and Cattell's Matrices). Again, results indicated significant differences between the genders, with males endorsing higher levels of confidence in the accuracy of their performance across all four of the cognitive tasks. For the Vocabulary and Letter Series tasks, males were also significantly more mis-calibrated than females. Gender differences in terms of mis-calibration for the General Knowledge task, however, were not significant. Similarly, Jonsson and Allwood (2003) did not find stable gender differences in mis-calibration for either a Word Knowledge task or a Logical Spatial task. More recently, Stankov and Lee (2008) found that males were significantly more mis-calibrated than females on the TOEFLiBT.

To summarise, the literature suggests that on tasks of cognitive abilities, males have been more confident, but not necessarily more calibrated, than females. Gender differences in mis-calibration within the cognitive domain are examined in Studies 1 and 3 of this dissertation. At this point it should be noted that calibration researchers have not previously examined Big Five mis-calibration or gender differences in Big Five mis-calibration, despite these being important topics of empirical investigation. Studies 2 and 3 (Chapters 4 and 5) have examined these gender differences.

2.5.5.2 Age Differences in Accuracy, Confidence and Mis-Calibration

2.5.5.2.1 Age Differences in Accuracy

From the vantage point of Gf/Gc theory (Horn, 1988), accuracy on tasks of fluid intelligence (Gf) decreases with age as fluid intelligence peaks in early adulthood and then declines, whereas accuracy on tasks of crystallised intelligence (Gc) either increases, or stays stable, from age 20 to 65 years (Kaufman & Horn, 1996). As age differences in cognitive accuracy are well established in the literature, it will not be investigated in the current studies. Rather, Studies 1, 2 and 3 examined the hitherto unexplored role of age differences in personality accuracy. See Chapter 3 for further discussion.

2.5.5.2.2 Age differences in Confidence and Mis-calibration

Self-efficacy theory (Bandura, 1977), and cognitive dissonance theory (Festinger, 1957), may provide some insight into what to expect in terms of age differences in cognitive confidence and mis-calibration. Self-efficacy describes people's beliefs and confidence about their own ability to perform in a particular domain (Bandura, 1997). Domain specific self-efficacy beliefs and confidence increase with personal accomplishments within a domain, with the reverse happening when negative experiences occur. For instance, an elderly woman is invited to join a trivia club. If she wins a number of games her self-confidence in her ability to play the game increases, thereby increasing her motivation to try harder. If she often loses, however, both confidence and motivation decline. Blanton, Pelham, DeHart, and Carvallo (2001) related this decrease in motivation to cognitive dissonance theory (Festinger, 1957). This theory posits that individuals prefer that their cognitions, including those about their own actions, be consistent. Dissonance occurs when these cognitions are inconsistent, and the individual is motivated to make them more consistent in order to decrease uneasiness or distress. Blanton et al. (2001) argued that mis-calibration can result from the need to view oneself as a capable and knowledgeable individual who makes competent self-judgments. Accordingly, unwarranted confidence occurs with judgments that challenge a positive view of the self. For example, Mary believes she is a bright woman and then participates in an experiment examining mis-calibration on tasks of cognitive abilities. In the test battery she finds one of the tasks quite difficult, and in order to reduce cognitive dissonance (i.e., I should know the answer to this as I am bright) she provides a higher confidence rating than is warranted resulting in overconfidence.

Few studies have examined age differences in cognitive confidence and mis-calibration (Crawford & Stankov, 1996; Pallier, 2003). Crawford and Stankov (1996) found small but significant correlations (from .22 to .23) between age and overconfidence scores on tests of both Fluid intelligence (i.e., Raven's Progressive Matrices, Letter Number Sequencing, and Animals) and Crystallised intelligence (i.e., Vocabulary, Esoteric Analogies, Proverbs) as well as significant positive correlations between age and scores on the visual perceptual Line Length task. Similarly, Pallier (2003) also reported similar correlation coefficients between age and overconfidence scores for the fluid (r = .23) and crystallised (r = .16) tasks used in his study. However, within the sporting domain age did not co-vary with calibration on golf tasks (Fogarty & Else, 2005).

Given the paucity of studies that have examined the impact of age, a further aim of the current dissertation was to investigate whether age was associated with confidence and mis-calibration for the tasks in Study 1 that spanned various cognitive domains. The current research programme also contributes to the literature by its investigation into age differences in Big Five confidence and mis-calibration.

2.5.5.3 Ability and Confidence

A potential source of individual differences in cognitive confidence and mis-calibration could be ability levels, as research in the area of relative comparisons suggests that individuals of lower ability may have difficulty accurately appraising their cognitive abilities compared with their more competent peers (see also Dunning et al., 2003; Kruger & Dunning, 1999). In these types of studies, participants provide estimates in the form of percentile ranks of how their abilities compare to their peers (i.e., other undergraduate students). University students scoring in the bottom quartile on various cognitive tasks have significantly overestimated their own scores and percentile ranks on tests of Psychology (Dunning et al., 2003), humour (Kruger & Dunning, 1999, Study 1), logical reasoning (Kruger & Dunning, 1999, Study 2), and grammar (Kruger & Dunning, 1999, Study 3) compared with top quartile participants (see also Maki, Jonas, & Kallod, 1994; Moreland, Miller, & Laucka, 1981; Shaughnessy, 1979). In these studies, researchers employ calibration type procedures in terms of relative comparisons although confidence ratings and bias scores were not calculated. For example, in the Kruger and Dunning studies (1999) participants responded to various cognitive tasks and were then asked to provide percentile estimates about their own performance relative to peers. The sample was then divided into quartiles according to their actual performance and, for the first and fourth quartile individuals, percentile estimates were evaluated against their respective actual percentile ranks. Results suggested that people of lower ability overestimate their performance whilst the opposite is true for those of higher ability. Kruger and Dunning (1999) attributed these differences to a lack of metacognitive insight of lower ability participants.

More recent studies, however, have presented data demonstrating that judgments of relative comparisons contain noise such as regression to the mean, and floor and ceiling effects (Ackerman et al., 2002; Burson, Larrick, & Klayman, 2006; Krueger & Mueller, 2002). The more recent studies raise the possibility that Kruger and Dunning’s results need to be interpreted with caution as their findings may have been the result of statistical artefacts. For instance, Ackerman (2002) in a simulation of Kruger and

Dunning's study found that the perceived average percentile rank in their data fell at the 66^{th} percentile thus the division of participants into quartiles guarantees that bottom quartile participants are more mis-calibrated than individuals at top. That is, top quartile participants are closer to the 66^{th} percentile than are individuals in the bottom quartile. Other literature (e.g., Burson et al., 2006; Krueger & Mueller, 2002) indicates that both top and bottom quartile participants can also be wrong about their judgments of relative comparisons. For instance, Burson et al. across three studies demonstrated that top quartile individuals underestimated their relative standing in terms of percentile ranks across 12 cognitive tasks whereas the opposite was true for bottom quartile participants. However, when tasks were moderately difficult, then both top and bottom quartile participants were just as accurate in their judgments of relative comparisons. Moreover for difficult tasks, top quartile participants were less accurate in their judgments than were their low ability counterparts.

The focus of this dissertation is not whether individuals believe that they are better than their average peer. Of interest to the current studies is that confidence judgments are obtained from individuals with regard to the accuracy of their own self-assessment within the cognitive domain, and if those of lower ability believe that they performed better than they actually did (Dunning, 2005), then these individuals should be more overconfident in their *own* performance than their high ability counterparts. This issue was investigated in Studies 1 and 3, where participants responded to various cognitive tasks.

2.5.5.4 Personality Correlates of Confidence and Mis-Calibration

Previous research has yielded interesting findings regarding the personality correlates of confidence and mis-calibration. For example, proactiveness and activity were positively correlated with confidence (Pallier et al., 2002). Highly anxious individuals were less confident than non-anxious individuals on a Line-up Identification task (Nolan & Markham, 1998). Cognitive impulsivity co-varied with confidence and mis-calibration on a Comprehension Monitoring task (Walczyk & Hall, 1989). Clinically depressed people were more overconfident than their non-depressed peers when making judgments about real life events (Dunning & Story, 1991), but individuals with mild depression were better calibrated than their non-depressed counterparts on General Knowledge questions (Stone, Dodrill, & Johnson, 2001). Narcissism was positively correlated with overconfidence on a General Knowledge task (Campbell, Adam, & Joshua, 2004), and self confidence ratings obtained from an Esoteric Analogies task correlated negatively with imposterism (Want & Kleitman, 2006).

The focus of the current studies was to examine the relationships between the Big Five personality dimensions and cognitive confidence and mis-calibration. In terms of previous research using Big Five personality measures, moderately small but significant positive correlations (r = .30) have been found between the Openness/Intellect dimension and Working Memory confidence (Baker, 2001) and between Openness and confidence ratings on a Verbal Reasoning test (Kleitman, 2003). However, because Openness correlates with both the accuracy and confidence rating scores from tasks of cognitive abilities, it is important to partial out the variance that is attributable to accuracy when examining the relationship between this trait and cognitive confidence measures (Schaefer et al., 2004). This practice was adopted in the current research programme.

The relationship between Extraversion and both confidence and mis-calibration remains unclear. Extraversion, as measured by Goldberg's International Personality Item Pool (IPIP) inventory, did not correlate with confidence or overconfidence scores from the five Working Memory tasks in Baker's (2001) study. Similarly, Extraversion, as measured by NEO PI-R, did not correlate with the overconfidence factor that resulted from principal components analysis of a large battery of intellectual tasks in the Pallier et al. (2002) study. Conversely, Schaefer et al. (2004) found that the Extraversion subscale of the IPIP significantly predicted overconfidence for General Knowledge questions.

It is important to consider interpretative issues when examining whether Big Five personality variables co-vary with confidence and mis-calibration on cognitive tasks. Schaefer and his colleagues (2004) argued that interpretation of only simple zero-order correlations between overconfidence and each of the Big Five personality dimensions may be misleading because of the shared variance that exists between the five personality factors (i.e., correlations .30 or greater). They advocated the use of partial correlations when examining the relationships between the Big Five dimensions and accuracy and confidence scores (i.e., mean confidence ratings and bias scores) derived from cognitive tasks. After reporting zero-order correlation coefficients between Big Five subscale scores and bias scores, they also calculated partial correlations between each Big Five trait and accuracy, confidence and overconfidence scores, and partialled out the variance attributable to the other four personality dimensions. This change of analyses provided the clearest conclusions according to these researchers, and some of the significant simple correlations reported in their study failed to reach significance when partial correlations were used.

Schaefer et al. (2004) found that Intellect/Openness significantly predicted accuracy and confidence scores, but not overconfidence on the General Knowledge task used in their study, indicating that those with higher scores on this dimension were also more confident and accurate in their performance. Extraversion predicted overconfidence but not accuracy suggesting that extraverts were significantly mis-calibrated. Agreeableness and Neuroticism did not predict accuracy, confidence, or overconfidence. Finally, Conscientiousness significantly predicted confidence, but not accuracy or overconfidence, suggesting that, whilst conscientious individuals were more confident, this did not influence accuracy or mis-calibration. Other research has also demonstrated that Agreeableness and Conscientiousness do not correlate with bias scores derived from tasks of cognitive abilities, although it should be noted that these researchers did not use partial correlations in their work (Baker, 2001) However research by Kleitman (2003) found small but significant correlations (*rs* = .15) between Conscientiousness and bias scores derived from a Verbal Reasoning test, a Nonsense Syllogisms test, and the Esoteric Analogies test.

In summary, because previous research has not used partial correlations in the way advocated by Schaefer and his colleagues (2004), the role of personality in cognitive confidence and mis-calibration remains unclear, and further research is warranted. The use of partial correlations will be adopted in Studies 1 and 3 of this dissertation to investigate Schaefer et al.'s assertions.

2.5.5.5 A Trait of Self-Confidence within the Cognitive Domain

Another interesting question raised by differential psychologists is whether the item-by-item confidence judgments obtained from cognitive tasks represents a confidence trait. Correlational analyses of confidence ratings and accuracy scores obtained from batteries of tasks across different cognitive domains shows that mean confidence rating scores tend to be highly correlated. Indeed, they are more highly correlated than the accuracy scores obtained from the tasks, an observation that led Stankov and colleagues (see Stankov & Dolph, 2000 for a review) to claim that there is a general self-confidence (or self-monitoring) trait that is independent of accuracy. This confidence factor has been replicated many times using larger test batteries and different populations (Crawford & Stankov, 1996; Kleitman, 2003; Kleitman & Stankov, 2001; Pallier et al., 2002; Stankov, 1998, 1999a, 2000a; Stankov & Crawford, 1996a, 1996b, 1997; Stankov & Dolph, 2000; Stankov & Lee, 2008). Other investigators also provide data in support of a confidence factor that is domain independent (e.g., Schraw, 1994, 1997; Schraw & DeBacker Roedel, 1994; Schraw & Dennison, 1994; Schraw, Dunkle, & Bendixen, 1995; Schraw & Moshman, 1995).

Whilst Stankov and colleagues speculated that the confidence trait was related to metacognitive processes, few studies have examined the relationships between this trait and self-report measures of meta-cognition. Examination of these relations was addressed by Kleitman and Stankov (2007) who investigated the relationship between a self-confidence factor (from a diverse battery of cognitive tasks) and a meta-cognitive processes factor (from three meta-cognitive measures) via confirmatory factor analysis. The correlation between these two factors was $r = .44$. Stankov and Lee (2007) reported similar results when a battery of acculturated knowledge accuracy, confidence, and meta-cognition scores were factor analysed using exploratory factor analysis. The correlation between the meta-cognition and confidence factors was .32. Both studies indicate that confidence and meta-cognition represent separate factors that share some cognitive processes.

2.5.5.5.1 Gf-Gc Theory and the Calibration Paradigm

In each of the studies by Stankov and various colleagues noted above, task selection was guided by Gf-Gc theory (Horn, 1985) which postulates first, second, and third order factors. The two first-order factors being Gf (i.e., Fluid ability) and Gc (i.e., Crystallised ability) and the second and third-order factors being (i.e., Gv (visualisation abilities), Speed, Ga (auditory abilities), SAR (short-term acquisition retrieval), TRS (long-term storage retrieval), Visual Sensory Detectors, and Auditory sensory detectors.

This hierarchical structure of human cognitive abilities is posited on the basis of large factor analytic studies employing both exploratory and confirmatory techniques (Carroll, 1993). The first-order and second-order factors were derived from factor analysing accuracy scores from a diverse battery of cognitive tasks.

Stankov and his collaborators (Stankov, 1998, 1999a) attempted to replicate some of the factors in Gf-Gc theory by factor analysing accuracy and mean confidence rating scores from test batteries that included markers for Gf, Gc, and Gv. For example, Stankov (1998) used a test battery that included one marker test for Gf (i.e., Raven's Progressive Matrices), one marker test for Gc (i.e., a multiple choice synonyms vocabulary test), and one marker test for Gv (i.e., Line Length test). Correlational analyses revealed two findings. Firstly, the accuracy scores from the three tests were correlated but the magnitude of these correlations was low (i.e., .15 to .20). Stankov concluded that the magnitude of these correlations suggested that the accuracy scores do not share much common variance, and that each test measures a different construct in line with the Gf-Gc theory. That is, Gf, Gc, and Gv respectively. Secondly, the correlations between the confidence ratings were much higher (.32 to .52) and did not point to the distinction between Gf, Gc, and Gv that was apparent with the accuracy scores as the confidence ratings across a diverse range of tasks defined a single confidence factor. Stankov therefore concluded that the correlations between confidence ratings suggested the presence of a self-confidence trait that was distinct from accuracy measures. More recently, Stankov and his collaborators have provided further data in support of the factorial separation of confidence, and accuracy factors, obtained from the same tests, across various cognitive domains (Kleitman & Stankov, 2001, 2007; Pallier et al., 2002; Stankov & Lee, 2008). This separation occurs, despite the average correlation between confidence and accuracy scores obtained from the same test, ranging between 0.40 and 0.60 (Stankov & Kleitman, 2008). Data in these studies indicate that the high correlations among the confidence ratings obtained from a diverse range of cognitive tests, leads to the emergence of a confidence factor that is separate from, but correlated with, ability factors (Stankov & Kleitman, 2008).

Collectively, the findings from the aforementioned factor analytic studies indicate the presence of self-confidence factor (trait) that is factorially separate from the cognitive domain being investigated. Study 1 extended this line of investigation by measuring and factor analysing confidence ratings from both ability and non-ability domains (i.e., measures of how confident people are when making judgments about their personality within the Big Five taxonomic framework of personality structure). Study 1 includes two markers for Gc, two markers for Gf, one marker for Gv, and one confidence rating for each of the Big Five personality dimensions. A major aim of this dissertation was to examine the factorial structure of these ratings to determine whether confidence in ability and personality judgments are subserved by the same cognitive processes. A rationale for extending confidence judgments into the domain of personality judgments is presented in sections 2.6 and 2.7.

Returning to the bias scores obtained from ability tasks, Stankov and his colleagues have also factor analysed the bias scores obtained from the various cognitive tasks and found that they also loaded onto one factor. Accuracy scores (i.e., proportion correct scores), however, were not factor analysed with the bias scores in these studies, because these scores are mathematically dependent. As discussed earlier, the bias score is the simple difference between the mean confidence rating score and the proportion correct score both obtained from a cognitive task. The factorial structure of cognitive bias scores will be examined in Study 1. More importantly, however, is that the current research programme extended previous empirical work by examining mis-calibration within the domain of personality (see Studies 1, 2 and 3). Furthermore, the current studies contributed to the extant literature by investigating the factorial structure of cognitive and Big Five bias judgments (see Chapter 5).

Calibration research reviewed in this section has shown a well-established confidence trait that has emerged from tasks spanning differing cognitive domains (e.g., Gf, Gc, Gv). This trait has been argued to lie on the boundaries between abilities and personality (Stankov, 1999b). However, differential psychologists have not yet investigated the generality of this trait in other domains such as personality judgments. Therefore the question of whether there is a general confidence trait that influences judgments across both the personality and abilities domains remains unanswered. That is, are the processes involved in making confidence judgments about aspects of my personality the same as those used when I am asked to provide a confidence rating about whether I answered a test item correctly? Is there a difference between knowing your abilities and knowing your personality? These are important questions that have hitherto not been addressed.

2.6 Justifying Inclusion of Personality Confidence Judgments

There are good reasons for extending calibration research into the domain of personality. Opponents of research in the judgment and decision-making area have argued that calibration research has portrayed both an uncomplimentary and unmerited image of human inferential capabilities because researchers have chosen judgmental tasks that are not familiar to, or maybe even misunderstood by, some participants, therefore leading to mis-calibration (Dunning, Griffin, Milojkovic, & Ross, 1990). Moreover, these judgements are fundamentally different from the important day-to-day decisions that govern people’s lives. Dunning and his colleagues argued that it was important to extend calibration research into more familiar domains.

Dunning et al. (1990) conducted a series of studies in which participants predicted the actions of peers in response to a variety of stimulus situations (e.g., responses to hypothetical predicaments), and were then asked to express their confidence in the accuracy of each prediction on a numerical scale ranging from 50 to 100 percent. Findings indicated that participants were consistently over-confident when predicting the social behaviours of peers with bias scores ranging from 8.7 to 14.9. These studies provide evidence that overconfidence is also present in the area of social prediction. Overconfidence effects were also found in the context of self-predictions of future behaviours (e.g., leisure and social activities) in a study by Vallone, Griffin, Lin and Ross (1990). Dunning et al. and Vallone et al. stated that it is important that researchers continue to investigate whether mis-calibration occurs in other domains that are both familiar and relevant to the decisions people make in daily life. It is argued that accurate personality judgments are very pertinent to the many decisions that individuals make each day.

For example, as argued by previous researchers, accurate personality judgements are an important aspect of daily life (Letzring, Wells, & Funder, 2006) that influences how one behaves, and how one thinks and feels about himself or herself (Kolar, Funder, & Colvin, 1996). These judgments also impact how one chooses relationship partners or friends (Funder, 1999), and has implications for organizational effectiveness in terms of employing the right person for the job (Christiansen, Wolcott-Burnam, Janovics, Burns, & Quirk, 2005; Funder, 1999). Accurate judgments of clients' personalities are critical in clinical psychology so that appropriate interventions can be selected and implemented (Funder, 1999; Funder & Sneed, 1993). Indeed my own clinical experience from working as a psychologist in a psychiatric hospital over several years has been that when patients' lack self-insight into their own personality disorders then psychological interventions are ineffectual as patients are not motivated to work on their issues as they do not believe that there is anything wrong. Also, studying the accuracy of personality judgments is also of theoretical interest to personality and social psychologists (Koehler et al., 1996), as well as to attribution theorists (Funder, 1980). Another reason for examining calibration in personality judgments is the widespread use of self-report measures of personality. A major problem with this form of assessment is faking, whereby respondents deliberately misrepresent their scores on certain personality traits to maximise the chances of a favourable assessment. The research on mis-calibration described in this chapter raises the possibility that these effects may not be due entirely to faking. Consequently it is important to investigate whether individuals are mis-calibrated (i.e., inaccurate) when making personality judgments about themselves. Personality accuracy is operationalised in Chapters 3 and 4.

2.7 Empirical Relations Between Personality and Ability/Intelligence Measures

Over the last ten years research interest in investigating the relationships between abilities (intelligence) and personality has been increasing (see meta-analyses by Ackerman, 1997; Reeve, Meyer, & Bonaccio, 2006). Prior to this, these domains were studied independently despite more than 100 years of scientific interest in these relations (see Reeve et al., 2006, for a discussion). One reason cited in the meta-analyses for this separation has been that intelligence researchers are interested in maximal performance (what I can do) whereas in the personality domain typical performance (what I am generally like) has been the focus of attention (Ackerman, 1997; Ackerman & Heggestad, 1997). Maximal performance is measured by objective tests whereas typical performance is measured by self-report inventories.

Investigators who have studied the relationship between personality and intelligence measures have used measures guided by the Five Factor Model of personality structure as this taxonomy of personality traits is regarded by many researchers as a framework that includes most phenotypic personality attributes (e.g., Ackerman & Heggestad, 1997; Austin, Dreary, & Eber, 2002; Goldberg, 1999; Pallier et al., 2002; Saucier & Goldberg, 1996; Wiggins, 1996).

Results from Ackerman and Heggestad's meta-analytic review of 155 studies showed that psychometric *g* was significantly related to Neuroticism (r = -.15), Extraversion (r = .08) and Openness (r = .33). Similar correlations were found between Gc and these three personality variables. Correlations between Gv and these variables were r = -.04 for Neuroticism, r = .06 for Extraversion, and r = .24 for Openness. Nevertheless, despite the magnitude of these correlations being generally low, Ackerman and colleagues (Ackerman, 1997; Ackerman & Heggestad, 1997) and Demetriou and Kazi (2001) assert that intelligence and personality are actively entwined in determining the chance of success in a particular task domain. For example, within the domain of academic performance, Chamorro-Premuzic and Furnham (2008) used hierarchical regression to examine how much personality (i.e., Big Five personality traits) predicts academic performance over and above the variance explained by intelligence (i.e., the Wonderlic IQ test, and Baddeley's Reasoning test, a measure of Gf). Results indicated that Gf accounted for only 6% of the variance in academic performance; Conscientiousness contributed a further 27% and Openness an additional 3%. Chamorro-Premuzic and Furnham's research was inspired by other empirical evidence which showed that IQ tests infrequently explain more than 50% of the variance in academic performance (e.g., Chamorro-Premuzic & Furhnam, 2004; O'Connor & Paunonen, 2007).

Because of the long standing interest in the relations between personality and intelligence/ability measures, and also because personality works in tandem with ability in determining the success in various domains, the current research investigated the existence of links between confidence judgments obtained from both the abilities and Big Five personality domains.

2.8 Confidence Judgments Across Cognitive and Personality Measures: One Trait or Two?

Do confidence judgments obtained from both the abilities and personality domains load onto a general confidence factor? Because this question has not yet been investigated in the calibration literature, it has been investigated in the current studies. Self-concept theory may shed some light on what to expect in regard to the factorial structure of these confidence ratings.

Shavelson, Hubner, and Stanton’s (1976) model (see Figure 2.1) posits that global self-concept can be divided into academic, social, emotional, and physical self-concepts, and that each of these second-order constructs being further sub-divided into narrower sub-domains. More recent research however, has not demonstrated evidence in favour of this hierarchy and has consistently shown that both academic and non-academic (social, emotional, and physical) self-concept are highly differentiated (see Marsh, 2008, for a review). Moreover, Marsh’s (2008, p. 450) review of his empirical work that demonstrates that “if specific components of self-concept are highly differentiated, then there is much variation in specific components that cannot be explained in terms of a single global component such as self-esteem”. These findings are relevant to the current studies as Stankov and Crawford (1997) suggested that both confidence and self-concept judgments are cognitive appraisals of the self. If self-concept is highly differentiated between the academic and non-academic domains then perhaps the self-confidence trait is also highly differentiated across the abilities and personality domains. It follows then that confidence judgments across cognitive and personality judgments should define different factors at the factorial level. Investigating whether personality confidence splits at the factorial level from cognitive confidence has important implications for calibration theorists who are trying to understand self-monitoring. This question has not been examined by previous researchers.

The current studies have attempted to answer this important question by obtaining confidence measures from both the cognitive and personality domains. As noted earlier the personality variables were selected within the framework of the Big Five Model of personality (Goldberg, 1997). Horn’s (1985) theory of Fluid and Crystallized Intelligence provided the framework from which the psychometric tasks were chosen. The cognitive domains that were assessed include acculturated knowledge (Gc), abstract reasoning (Gf), and visual perception (Gv).

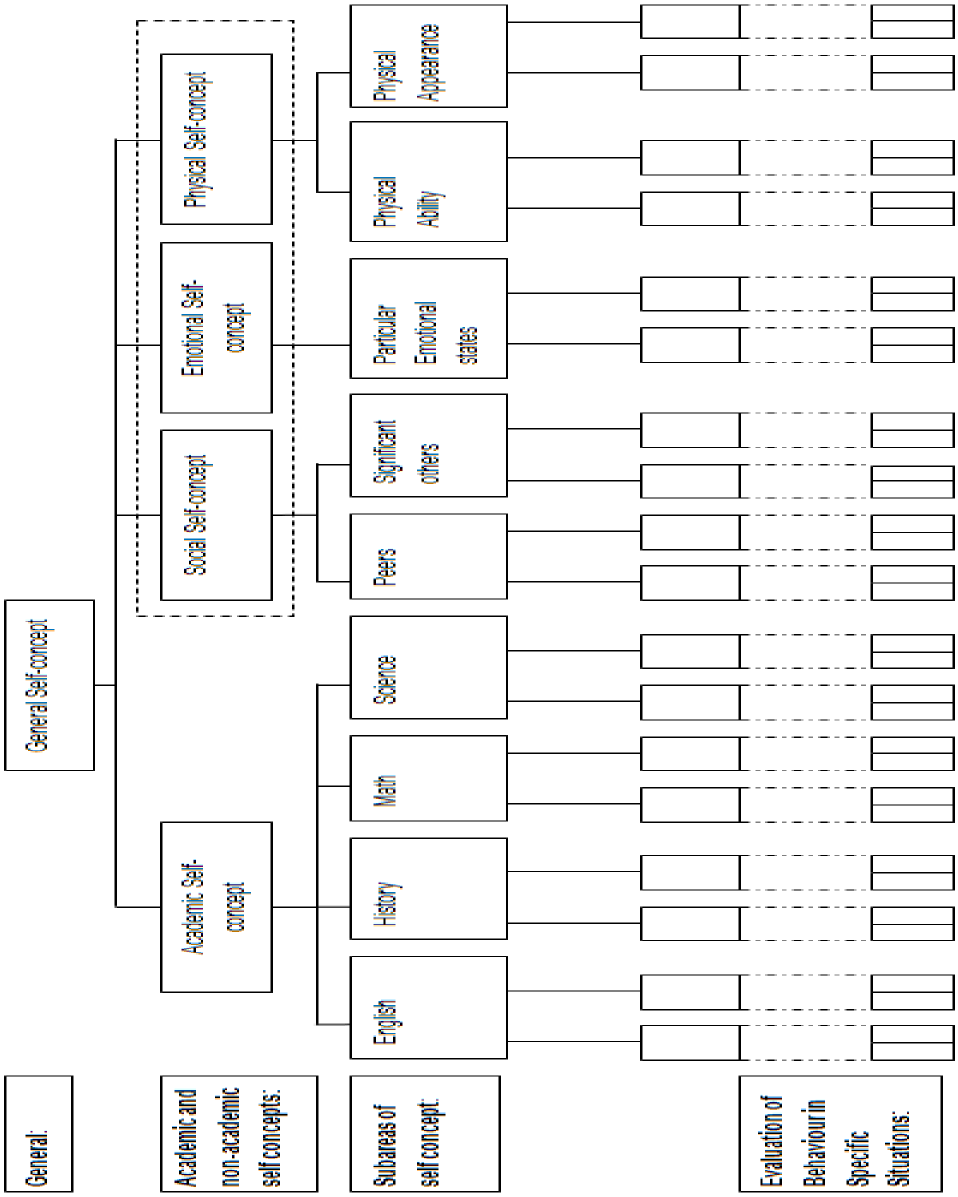


Figure 2.1: Diagrammatic representation of the multidimensional, hierarchical model of self-concept.

2.8.1 Summary

Thus far, this chapter has reviewed calibration research findings mainly from within the cognitive domain, and has provided both theoretical and empirical reasons for conducting calibration type analyses within the personality arena. The next section deals with the measurement issues associated with personality judgments.

2.9 Measurement Issues and Empirical Findings of Personality Judgements

Research examining the accuracy of personality judgments falls into two main groups, that is, interpersonal and intrapersonal perception. It is acknowledged from the outset that studying accuracy within the personality domain is weighed down by methodological complexities, because there is no criterion by which a personality judgment can be appraised as being either true or false. Nevertheless, Funder (1999) argued that these difficulties should not stop accuracy research and advocated that researchers should continue to tackle the topic and frankly acknowledge any methodological problems, because studying accuracy is far too important to be abandoned. He contended that studying the accuracy of personality judgments should be centred on three reasonable premises that must be accepted as a given before conducting this type of investigation:

1) individual differences in personality (personality traits) exist and are important;
2) people sometimes make judgments about these traits;
3) these judgments are sometimes accurate (p. 11).

These premises were accepted for the purposes of the current studies. The accuracy debate concerning personality judgments will be elaborated upon in Chapters 3 and 4.

The major focus of personality research with regard to phenomenological accuracy has dealt with interpersonal perception and this literature is reviewed in the next section.

2.9.1 Interpersonal Perception

Studies have shown moderate correlations between self-ratings and others' ratings of personality (e.g., Funder, 1980, 1999; Meyer et al., 2001; Watson, Hubbard, & Wiese, 2000). It has been argued that these moderate correlations imply both accuracy and convergent validity for intrapersonal self-reported ratings of personality.

Inter-judge agreement or consensus between a number of observers has been the most commonly used criterion for assessing the accuracy of personality judgments (e.g., Albright et al., 1997; Borkenau & Liebler, 1993a, 1993b; Funder & Colvin, 1988; Vogt & Colvin, 2003). However, consensus/agreement does not necessarily imply accuracy, because shared errors or biases can lead to consensus, but not necessarily to accuracy (Blackman & Funder, 1998; Funder, 2001; Swan & Gill, 1997). That is, two or more judges can reach agreement but one or all of them can be inaccurate for a number of reasons. Firstly, observers may not have had enough information about target persons, therefore reducing accuracy. Secondly, observers may have discussed their conclusions about a target and inadvertently influenced each other, so that although they reached consensus in their judgements, they were in fact inaccurate. Judges may also have used their own implicit personality theories to make judgements about the target rather than relying on the information solely provided. Alternatively, observers may make judgements about a target based on population base rates. Because dissertation is concerned with self-judgments, the discussion now addresses the research literature that has examined intrapersonal perception.

2.9.2 Intrapersonal Perception

How do we come to know ourselves? Am I agreeable, conscientious, extraverted, neurotic, or even intellectual, for example? The psychological literature highlights that individuals draw upon at least two self-appraisal processes to respond to these types of questions. These appraisal processes can be grouped as being either reflected or direct (Ochsner et al., 2005).

2.9.2.1.1 Reflected Appraisals

Two theories that exemplify reflected appraisals are the looking glass self theory, and Bem's (1972) self-perception theory. According to the looking glass self theory, self-knowledge is based on what individuals believe others think of them (e.g., Shrauger & Schoeneman, 1979). For example, Joe loves going to parties and being the centre of attention. Joe's family and friends think he is extraverted; therefore, Joe believes he is extraverted. Bem expresses it somewhat differently in his self-perception theory:

> Individuals come to know their own attitudes, emotions and internal states by inferring them from observations of their own behaviour and circumstances in which they occur. When internal cues are weak, ambiguous, or uninterpretable, the individual is in the same position as the outside observer (p. 2).

Therefore, according to Bem (1972), self-knowledge is attained in the same way that one perceives others. That is, individuals make personality judgments about themselves by observing their own behaviour, just as they make judgments about others by observing others' behaviours. In contrast to both the looking glass self theory and Bem's self-perception theory, direct appraisals examine self-knowledge from the vantage point of two types of knowledge structures in memory.

2.9.2.1.2 Direct Appraisals

Cognitive psychologists assert that trait self-knowledge, and recall of specific occasions that involve those traits, stem from two different knowledge structures (e.g., Kihlstrom, Beer, & Klein, 2003; Klein, Loftus, & Burton, 1989; Klein, Loftus, &

Kihlstrom, 1996). These are semantic self-knowledge and episodic personal knowledge. Episodic knowledge represents memory for specific events from one's past in which the self was present, whereas semantic self-knowledge is more generalized, is context-free, and includes general knowledge of the world (Kihlstrom et al.)**.** According to this view, individuals' self-knowledge is based on their general view of themselves, as well as on recalling specific events that provide evidence for trait self-knowledge. For example, Fred believes he is generally conscientious, and then remembers when he worked long hours to finish a marketing project on time. From a cognitive perspective, however, self-knowledge is not limited to semantic and episodic memory structures.

Other cognitive work on self-knowledge has focussed on a number of areas: associative network models of the self, the self-reference effect, priming and self-referent processing, self-relevance and recognition, interactions between episodic and semantic self-knowledge, and neuropsychological approaches to the self as memory (see Kihlstrom et al., 2003 for a review).

2.9.2.1.3 Biases That Can Affect the Accuracy of Self-Perception

Within the psychological literature, researchers argue that many motives influence the process of self-evaluation (for reviews see, Leary, 2007; Sedikides & Strube, 1997; Taylor, Neter, & Wayment, 1995; Tesser, 2003). These motives include, self-enhancement, self-verification, self-serving attributions, self-improvement, need for closure, cognitive consistency, social desirability, the bias blind spot, and uncertainty orientation. However, the emphasis has been on two motives: self-enhancement and social desirability.

2.9.2.1.4 Self-Enhancement

It has been argued that individuals self-enhance to bolster their self-esteem (e.g., Dunning, 2005; Taylor & Brown, 1988; Tesser, 2003), and that self-perceptions are essentially distorted and self-enhancing, and are often more positive compared with the perceptions of others (Taylor & Brown, 1988). Within the personality domain, this self-enhancement bias has been demonstrated by studies in which individuals were found to be less likely to endorse negative personality traits as being characteristic of themselves (e.g., Dunning, Meyerowitz, & Holzberg, 1989; Halperin, Snyder, Shenkel, & Houston, 1976). John and Robins (1994), however, demonstrated that the self-enhancement bias was not a universal law of human behaviour, and that there were marked individual differences ranging from self-enhancement to self-diminishment biases. In their study, 53% of participants were reasonably accurate when asked to estimate their performance on a managerial group-discussion task, with 15% underestimating and 32% overestimating their performance. Gosling, John, Craik, and Robins (1998) conducted a study in which a similar conclusion was reached. Furthermore, in their review of the characteristics of self-enhancers, Robins, Paulhaus, Roberts, and Hogan (2001) concluded that self-enhancement is not universal, and that the conclusion drawn from the few studies in which the percentage of participants who self-enhanced was reported, as well as those who did not, is that self-enhancers are in the minority. These authors critiqued previous research methodologies for not including the percentages of participants who self-enhanced, self-effaced, or who were accurate, so that appropriate conclusions could be made.

2.9.2.1.5 Above-Average Effect

In the area of social cognition, researchers have argued that the above-average effect is also an example of self-enhancement bias. The literature on personality judgments describes studies in which, participants were required to indicate the degree to which various personality traits described themselves, compared with unfamiliar average peers of the same gender (e.g., Alicke, 1985; Alicke & Govorum, 2006; Alicke, Klotz, Breitenbecher, & Yurak, 1995; Silvera & Seger, 2004). Findings indicated that people tended to rate themselves more positively than they rated their average peers. For example, in Alicke et al.'s study in 1995 university students were informed that the purpose of the study was to investigate how they would rate themselves on various personality traits relative to unknown average university students on 20 positive (e.g., responsible, perceptive, creative, and polite) and 20 negative (e.g., meddlesome, complaining, unforgiving, and vain) traits. Results indicated that participants rated themselves as better than average on 38 out of 40 traits. These authors concluded that this effect was a pervasive phenomenon. However, Colvin, Block, and Funder (1995) argued that when individuals are asked to rate their own personality traits relative to unknown peers, there are inevitably times when participants' ratings are accurate, and times when they are inaccurate or self-enhancing. They suggest that there is no way of telling what percentage of participants might fall into each category. Self-enhancement also occurs when individuals respond to psychological test items in a socially desirable manner, a phenomenon discussed in the following section.

2.9.2.1.6 Social Desirability

Psychologists hope clients will respond to self-report inventories in a frank and truthful manner so that valid results can be interpreted. Unfortunately, as previously noted, this does not always happen, and some individuals may respond to items in a socially desirable manner. Social desirability is defined as positive endorsement of socially-acceptable items and negative endorsement of items that are considered socially unacceptable (Edwards, 1953). With regard to personality self-report, some authors have argued that social desirability is not a response bias that causes problems for personality measurement and that psychologists should stop "…kicking the methodological dead horse", and that "…the sacred cow status" of this construct should be re-evaluated (Nevid, 1983, p. 139). McCrae and Costa (1983) concurred with Nevid, and provided empirical data to support this conclusion. Similarly, conclusions are found within the organisational literature. In their meta-analytic review, Ones, Viswesvaran, and Reiss (1996) concluded that social desirability in personality testing for personnel selection was, overall, nothing more than a "…red herring", and that the use of personality inventories should continue.

2.9.2.1.7 Are Personality Judgements Error-Prone or Error-Free?

Another focus of research into intrapersonal perception of personality traits focussed on the instances when individuals make errors in self-judgements about their personality. This research has consistently demonstrated that individuals readily accept general personality statements, supposedly derived from personality tests, as being accurate descriptions of their personality (e.g., Ditto & Boardman, 1995; Greene, Bausom, & Macon, 1980; Jackson & Murray, 1985; Mosher, 1965; Snyder & Larson, 1972; Snyder & Shenkel, 1976; Snyder, Shenkel, & Lowery, 1977). The acceptance of bogus personality feedback has been coined the Barnum effect, so as to stigmatise those personality descriptions that are highly accepted simply because their base rate in the general population is high (Meehl, 1956). An example of a Barnum feedback statement from Marks and Seeman (1962) is:

> You have a tendency to worry at times but not to excess. You do get depressed at times but you couldn't be called moody because you are generally cheerful and rather optimistic. You have a good disposition although earlier in life you have had to struggle with yourself to control your impulses and temper (p. 205).

Accepting of this type of feedback does not necessarily constitute an error in judgment for the following reasons. Firstly, this scenario has a high base rate in the general population; therefore endorsement of this description is not an error per se, but rather a methodological flaw in the studies that have used these kinds of feedback statements. Secondly, researchers who have used Barnum scenarios in their studies have also employed deception as part of the experimental design, with participants being told that this feedback was derived from the psychological inventories they had completed earlier. Hence, it is possible that acceptance of this type of feedback is due to participants not wanting to challenge the psychological interpretation provided by the so-called personality expert. Previous research has demonstrated that the credibility of the interpreter of psychological tests influences participants' acceptance of personality feedback, even if the feedback is negative (Binderman, Fretz, Scott, & Abrams, 1972; Halperin et al., 1976).

Harris and Greene (1984) improved previous research by examining participants' perceptions of actual feedback (their true score), trivial feedback (true for people in general), and inaccurate feedback (the opposite of their actual score) based on their responses to the California Psychological Inventory. Results showed that participants were able to differentiate between the three types of feedback thereby showing that their perceptions of themselves were valid. Similar results were reported by Davies (1997 see experiment 3) where participants were administered the 16PF questionnaire and were asked to rate the accuracy of true and false feedback from the 16 PF manual. True feedback was judged as more accurate than was false feedback. Most recently, Andersen and Nordvik (2002) examined whether a Barnum type effect (that is, participants' endorsement of a false trait profile as being accurate) occurred within the Five Factor Model of personality judgments. Participants responded to the NEO

Personality Inventory-Revised (NEO PI-R) so that their actual personality profiles were calculated for each of the Big Five dimensions. One month later participants were asked to rate the accuracy of a unique NEO PI-R *false* profile that was generated for each participant. These profiles included *T* scores and subjects were given information about the Five Factor Model and were able to read descriptions of each trait from the test manual before giving their accuracy rating. Participants correctly recognized and rejected *T* score profiles that were distant from their actual *T* scores suggesting that the Barnum type effect is not present in Big Five judgments.

In summary, acceptance of Barnum type feedback does not necessarily imply that people are inaccurate when making judgments about themselves. What is important is that general personality descriptions are not useful for judging the accuracy of self-judgments of personality. Therefore, Big Five trait specific personality descriptions were generated for the current research studies (see Study 1 for details).

2.10 Chapter Summary

Accurate self-monitoring is an important area of psychological enquiry. Overconfidence appears to be the most common judgmental error. Underconfidence appears to be less prevalent although there is still some doubt as to whether underconfidence is a pervasive phenomenon in all tasks of sensory discrimination. Study 1 has included a measure of Gv to examine whether participants are mis-calibrated.

Gender differences in confidence and mis-calibration in the cognitive arena has demonstrated that whilst males are more confident, they are not necessarily more mis-calibrated than females. However, previous research findings are mixed therefore the role of gender is re-examined in the current research programme. Few studies have investigated whether age is correlated with cognitive confidence and mis-calibration.

This was examined in Studies 1 and 3. Additionally, obtaining confidence ratings from personality judgments allows examination of age and gender differences in personality confidence which have hitherto not been previously examined. These differences are investigated in Studies 1, 2, and 3.

A robust confidence factor (trait) has been demonstrated within the cognitive domain and Stankov and his collaborators have argued that this trait initiates self-monitoring. The present studies have continued and extended this line of investigation by measuring and factor analysing, the confidence ratings obtained from both the ability and personality domains to examine, whether confidence in ability and personality judgments are subserved by the same cognitive processes. This is important for calibration researchers because it remains unclear whether there is a general monitoring trait that spans differing domains. A robust overconfidence factor is also present in the cognitive domain, and a single overconfidence factor was expected to emerge from the diverse range of cognitive tasks in Studies 1 and 3. The question that remains unexplored, however, is whether overconfidence in personality judgments shares variance with overconfidence in cognitive judgments. Before this could be addressed, accuracy methods for personality judgments were developed in Studies 1 and 2. Study 3 then examines whether there is a general overconfidence factor that spans both the cognitive and personality domains. As this concludes the review of the relevant literature, the focus now shifts to Study 1, which is presented in Chapter 3.

Chapter 3 - Study 1

3.1 Introduction

The domains assessed in Study 1 include Gc, Gf, Gv and confidence ratings from each of the Big Five personality dimensions. Study 1 was designed to:

1. Investigate whether individuals are mis-calibrated across a diverse battery of cognitive tasks.
2. Examine if gender, age, and personality, are each correlated with cognitive accuracy, with confidence, and with mis-calibration.
3. Explore if those of lower ability are more mis-calibrated than those of higher ability.
4. Examine the factorial structure of cognitive confidence judgments.
5. Construct self-rated personality descriptions based on the Big Five trait adjectives by Goldberg (1997).
6. Obtain confidence and accuracy measures for each of the Big Five personality dimensions.
7. Investigate the factorial structure of cognitive and personality confidence judgments.
8. Examine the factorial structure of cognitive bias scores.
9. Examine the relationship between cognitive and personality bias.

In the following sections, relevant research is reviewed briefly and predictions are presented.

3.2 Mis-calibration for Gf, Gc, & Gv Tasks

This section presents the overall trends from the calibration literature for Gf, Gv, and Gc tasks. Individual differences in cognitive confidence mis-calibration are addressed in section 3.3.

Reliable findings have shown that individuals tend to be either overconfident or underconfident when evaluating the accuracy of their cognitive performance and being well calibrated is much less common. Reasonably good calibration has been demonstrated on reasoning (Gf) tasks (e.g., Raven's Progressive Matrices and/or Letter Series) (e.g., Baker, 2001; Crawford & Stankov, 1996; Kleitman & Stankov, 2001; Pallier et al., 2002, Study 2; Stankov, 1998; Stankov & Crawford, 1996a, 1996b). Pallier et al. asserted that good calibration was expected for Gf tasks as these tests do not include misleading items or else the construct validity of these tests would be questionable. It is not possible to select a representative sample of reasoning items from participants' natural ecology, and therefore cue and ecological validities are equal. Some instances in the literature, however, have demonstrated overconfidence for Ravens Progressive Matrices task (Pallier et al., 2002, Study 1; Stankov & Dolph, 2000). Pallier et al. explained this overconfidence in their study by suggesting that military participants were less accurate than were the samples of undergraduate students used in the other studies cited above. In general however, the aforementioned literature concluded that good calibration was evident for Gf tasks. Therefore, for the Gf tasks used in the current study, good calibration was expected, along with bias scores that were close to zero and calibration curves that closely paralleled the perfect calibration line.

Research findings have also been mixed for visual perceptual (Gv) tasks, with some studies demonstrating underconfidence (Crawford & Stankov, 1996; Kleitman &

Stankov, 2001; Pallier et al., 2002, Study 2; Petrusic & Baranski, 1997; Stankov & Crawford, 1996a, 1996b, 1997), and others demonstrating overconfidence (Pallier et al., 2002 Study 1; Stankov, 1999a). Reasons for these discrepancies remain unclear and further research is required. A Gv measure was included in Study 1 to further examine mis-calibration.

For tasks of acculturated knowledge (Gc), on the other hand, overconfidence has consistently been demonstrated in the research literature (Allwood & Granhag, 1996; Crawford & Stankov, 1996; Juslin, 1994; Kleitman & Stankov, 2001; Pallier et al., 2002; Schaefer et al., 2004; Stankov, 1998; Stankov & Crawford, 1996a, 1996b, 1997; West & Stanovich, 1997). One possible explanation for these findings is derived from PMM theory which Gigerenzer (1991) used to argue that items from acculturated knowledge tasks were unrepresentative of participants' natural ecology, thereby leading to overconfidence. Gigerenzer and colleagues asserted that if researchers randomly selected ecologically valid items, overconfidence would disappear. In line with this argument for acculturated knowledge tasks, participants in Study 1 were expected to be overconfident, bias scores were expected to be positive, and calibration curves were expected to display overconfidence.

The next section covers whether individual differences in gender, age, ability, and personality, impact on cognitive confidence and on mis-calibration.

3.3 Individual Differences in Cognitive Confidence and Mis-calibration

3.3.1 Gender

Previous research showed that males were more confident than females (Pallier, 2003; Pulford & Colman, 1997; Ross & Fogarty, 2006; Stankov, 1998; see Stankov &

Lee, 2008, for an exception). However, gender differences in mis-calibration in the cognitive arena have been somewhat mixed (Baker, 2001; Beyer & Bowden, 1997; Crawford & Stankov, 1996; Jonsson & Allwood, 2003; Pallier, 2003; Pulford & Colman, 1997; Stankov, 1998; Stankov & Crawford, 1997; Stankov & Lee, 2008). It was suggested in the last chapter that higher confidence in males may have been due to either gender-stereotypic socialization patterns, or to task characteristics, or to both of these variables (Beyer, 1990; Marsh & Yeung, 1998; Ross & Fogarty, 2006). Males were therefore expected be more confident than females for all five tasks used in Study 1. Gender differences in cognitive mis-calibration were also explored although no specific hypotheses were generated due to conflicting research findings.

3.3.2 Age

Little research has examined whether age differences are linked to individual differences in confidence and mis-calibration within the cognitive domain (Crawford & Stankov, 1996; Pallier, 2003). In terms of confidence and mis-calibration, both self efficacy theory (Bandura, 1977), and cognitive dissonance theory (Festinger, 1957) may shed some light on what to expect. If task completion activates either negative self-efficacy beliefs or cognitive dissonance, confidence levels in older participants may be affected. Specifically, activated negative self-efficacy may produce lower confidence than is warranted by accuracy scores, whereas cognitive dissonance may affect confidence in the opposite direction (see Chapter 2 for the rationale).

Calibration literature has demonstrated evidence of greater mis-calibration for older participants with small but significant positive correlations between age and bias scores that have been obtained from fluid, crystallised, and visual perceptual tasks (Crawford & Stankov, 1996; Pallier, 2003). This same pattern of results was expected for the Gc, Gf, and Gv tasks used in the current study.

3.3.3 Ability

Work by Dunning and his colleagues (Dunning et al., 2003; Kruger & Dunning, 1999) and others (see also Maki et al., 1994; Moreland et al., 1981; Shaughnessy, 1979), suggests that, compared with those of higher ability those of lower ability may have some difficulty in accurately appraising their cognitive abilities. Consequently for the cognitive tasks in Study 1, low scorers (i.e., in the bottom quartile) were expected to be more mis-calibrated than high scorers (i.e., top quartile). See Chapter 2 for a detailed rationale.

3.3.4 Personality

The current study also examined the relationships that exist between each of the Big Five personality dimensions, and both cognitive confidence and mis-calibration. Previous empirical studies using zero order correlations have demonstrated that Openness was significantly positively correlated with cognitive confidence (Baker, 2001; Kleitman, 2003; Pallier et al., 2002) and, Conscientiousness was significantly associated with cognitive bias scores (Kleitman, 2003). The role of personality, however, is somewhat unclear due to the concerns raised by Schaefer and his colleagues (2004) (see Chapter 2). They convincingly argued that researchers need to use partial correlations instead of simple zero-order correlations because of the shared variance that exists between the five personality dimensions. Consideration of the results by Schaefer et al. (2004) led to the expectation that, after controlling for the influence of the other four personality dimensions:

1. Openness/Intellect scores would be significantly related to cognitive confidence scores.
2. Extraversion would be associated with both confidence and bias scores.
3. Conscientiousness would be significantly related to confidence scores.

4. Emotional Stability and Agreeableness scores would not be associated with either cognitive confidence or bias.

3.4 Individual Differences in Big Five Accuracy, Confidence, and Mis-calibration

Despite the methodological difficulties mentioned in Chapter 2, the emphasis placed on personality assessment justifies the extension of the calibration paradigm to the domain of personality judgements. Following Gigerenzer's (1991) assertions that individuals are well calibrated to their natural environments, and coupling this assertion with the proposition that personality assessments are a perfectly natural everyday occurrence, it was expected that individuals would be well calibrated in this domain.

3.4.1 Gender Differences in Big Five Personality Confidence, Accuracy and Bias

The gender similarities hypothesis (Hyde, 2005) provides some insight into what to expect in terms of gender differences in Big Five confidence judgments. This hypothesis posits that males and females are more similar across a wide range of psychological variables, than they are different. Hyde argued that this hypothesis holds true for adults, adolescents and children. It is important to note, however, that Hyde was not arguing that men and women are similar across all psychological dimensions. With regard to effect sizes, Hyde postulated that the majority of psychological gender differences are likely to be in the close-to-zero ($d \leq 0.10$) or small ($0.11 < d < 0.35$) range, that very few would be in the moderate range ($0.36 < d < 0.65$), and that only a small number would be large (d = 0.66-1.00) or very large ($d > 1.00$). Hyde reviewed the results of 46 meta-analyses of studies that investigated gender differences across numerous psychological variables (i.e., cognitive, communication, social, personality, psychological well-being, motor behaviours, moral reasoning, delay of gratification, cheating, computer use and efficacy, and job attribute preferences). Results obtained from her review supported the gender similarities hypothesis with 78% of the 124 effect sizes being close to zero or small. There were, however, some exceptions. For example, motor behaviours, and in particular throwing distance (d = 1.98) and throwing velocity (d = 2.18) produced very large effect sizes in terms of gender differences. However, in her review Hyde did not comment about findings within the Big Five personality domain, which are relevant to the current studies. Nevertheless, inspection of the data Hyde presented in Table 1 of her work, which summarises the effect sizes obtained in Feingold's (1994) meta-analysis of gender differences in personality, upheld the gender similarities hypothesis. That is, most of the effect sizes were zero to small, with only the assertiveness facet of the Extraversion dimension demonstrating a moderate effect size (d = + 0.51). Moreover, as expected by the gender-similarities hypothesis, only one large effect size (-0.91) emerged for the tendermindedness facet of the agreeableness dimension. These findings suggested that males were more assertive and less tenderminded than females.

Gender-stereotypic socialisation patterns may provide one explanation for Hyde's findings within the Big Five personality domain. Assertiveness is often regarded as being a more masculine attribute and is therefore more likely to be reinforced in males. Tendermindedness, on the other hand, is often perceived as more feminine and consequently, encouraged in females. These suggestions are reasonable if one refers to social role theory, proposed by Eagly (1987). This theory contends that, each gender is expected to behave in certain ways that are considered appropriate for their gender and culture (Eagly & Wood, 1991), and that these expected behaviours influence personality (Feingold, 1994). For example, females are expected to have elevated levels of communal qualities such as friendliness, selflessness, concern for others and emotional expressivity (Eagly & Wood, 1991), whereas, males are expected to have elevated levels of agentic attributes such as independence, masterfulness, assertiveness and instrumental competence (Eagly & Wood, 1991). Nonetheless, although the literature has highlighted gender differences in Big Five personality ratings, there was no a priori reason to suspect that males and females would differ when making personality confidence judgments about themselves as it was argued in Chapter 2 that these judgments have high ecological validity. Therefore, it was expected that males and females would not differ in terms of personality confidence and that Hyde's hypothesis would be upheld in Study 1. For the reasons just stated, the same predictions were made for both personality accuracy and mis-calibration.

3.4.2 Age Differences in Big Five Confidence, Accuracy and Bias

According to the five-factor theory of personality (McCrae & Costa, 1999) personality traits develop during childhood and adolescence and then remain stable in adulthood. However, the fact that the traits themselves remain stable in adulthood reveals little about the stability of peoples' perceptions of those traits. Nevertheless, this stability (e.g., I am conscientious), combined with the assumption that personality judgments are made on a daily basis, and therefore have high ecological validity, plausibly leads to the conclusion that confidence, accuracy, and bias judgments about one's personality would also remain stable in adulthood. It is difficult to draw parallels with the cognitive confidence literature wherein individuals make confidence judgments on tests of maximal performance whereas the confidence judgments obtained for the Big Five dimensions were taken from measures of typical performance. Thus it was expected that age would not be associated with Big Five confidence, bias, or accuracy scores.

3.5 The Factorial Structure of Cognitive and Personality Confidence Judgments

Differential psychologists have demonstrated the existence of a trait (factor) of self-confidence that appears independent of the type of cognitive activity being investigated (cf. Stankov & Dolph, 2000). This confidence factor has been replicated using larger test batteries and different populations (Crawford & Stankov, 1996; Kleitman & Stankov, 2001, 2007; Pallier et al., 2002; Stankov, 1998, 1999a, 2000a; Stankov & Crawford, 1996a, 1996b, 1997; Stankov & Dolph, 2000; Stankov & Lee, 2008). Other investigators have also provided data in support of a confidence factor that is domain independent (e.g., Schraw, 1994, 1997; Schraw & DeBacker Roedel, 1994; Schraw & Dennison, 1994; Schraw et al., 1995; Schraw & Moshman, 1995). Moreover, it has been argued that this general self-monitoring/self-confidence trait represents one aspect of meta-cognition that is related to the accuracy of self-assessment in the cognitive domain (Kleitman & Stankov, 2001; Stankov & Dolph, 2000). Stankov and associates (Kleitman & Stankov, 2007; Stankov & Lee, 2008) have since demonstrated that this trait (factor) was moderately associated with a metacognitive factor in each of these studies. These moderate correlations imply that confidence and meta-cognition share common cognitive processes.

Empirical evidence has demonstrated that confidence ratings have consistently displayed very high internal consistency co-efficients (Baker, 2001; Jonsson & Allwood, 2003; Kleitman, 2008; Kleitman & Stankov, 2001, 2007; Liberman & Tversky, 1993; Pallier et al., 2002; Stankov, 1999a, 1999b, 2000a; Stankov & Lee, 2008). However, differential psychologists have not investigated the generality of this trait in other domains such as Big Five confidence judgments. For the purposes of Study 1 the assumption was made that the rating for each Big Five dimension would not be highly correlated with confidence in that rating as Stankov (1999b) has placed the cognitive confidence trait somewhere between the boundaries of intelligence and personality. The question that remains unanswered, however, is whether there is one general factor for both personality and cognitive confidence judgments or whether there are two. The current study examined the factorial structure of cognitive and Big Five confidence ratings.

Self-concept theory (Shavelson et al., 1976) could provide insight into what to expect with regard to the factorial structure of these ratings. Marsh (2008) asserted that academic and non-academic self-concept are highly differentiated, therefore variance in specific areas cannot be elucidated in terms of one general trait. It was therefore argued that the self-confidence trait is also likely to be highly differentiated across the Big Five (non-academic) and cognitive (i.e., abilities) domains, because both self-confidence and self-concept judgments are cognitive appraisals of the self. Investigating whether Big Five personality and cognitive confidence judgments splits at the factorial level has important implications for calibration theorists who are trying to understand self-monitoring. The domains assessed in Study 1 included Gc, Gf, Gv and confidence ratings from each of the Big Five personality domains. Two confidence factors were expected to emerge from the structural analyses of confidence scores obtained from both the cognitive and the personality measures used in Study 1. This expectation was based on the assumption that the self-confidence trait would be similarly differentiated across the cognitive and Big Five domains as the self-concept construct, as both represent cognitive appraisals of the self.

To date, the prediction about the factorial structure of cognitive and Big Five confidence scores has been made with reference to Self-Concept theory (Marsh, 2008; Shavelson et al., 1976) but not to Gigerenzer et al.’s (1991) ecological theory. Even though this theory was developed in relation to general knowledge questions, it is reasonable to assume that the same principles apply to personality judgments; thus separate confidence factors for personality and cognition would emerge, because the cue validities used to endorse personality test items are vastly different from those cues used to answer cognitive test items.

3.6 Factorial Structure of Cognitive Bias Scores

Stankov and his collaborators (Pallier et al., 2002; Stankov, 1998, 1999a) factor analysed bias scores obtained from various combinations of Gc, Gf, and Gv tasks, and found that the bias scores loaded onto one factor. These results indicated that mis-calibration across a diverse range of abilities and perceptual tasks were driven by the same cognitive processes. In view of these findings factor analysis of the bias scores obtained from the Gc, Gf, and Gv tasks used in Study 1 was expected to produce a single bias factor.

3.6.1 Correlations between the Personality Bias Score and Cognitive Bias Scores

As discussed in the previous chapter, an important question that remains unexplored is whether bias scores from the cognitive domain are associated with personality bias. In Study 1, an overall personality bias score was developed across all of the Big Five personality dimensions and cognitive bias scores were obtained from each of the cognitive tasks. The assumption from Gigerenzer's theory (1991) that cues used to answer cognitive items differ from those used to answer personality items, led to the expectation that personality bias would not be associated with cognitive bias scores.

3.7 Restatement of Hypotheses

The hypotheses for the current study are summarised as follows:

1. It was hypothesised that good calibration would be demonstrated for the Gf tasks used in Study 1, with bias scores being close to zero and with calibration curves that align closely to the perfect calibration line. Previous research which demonstrated that good calibration occurs for Gf tasks formed the basis of this hypothesis (e.g., Baker, 2001; Crawford & Stankov, 1996; Kleitman & Stankov, 2001; Pallier et al., 2002, Study 2; Stankov, 1998; Stankov & Crawford, 1996a, 1996b).
2. It was hypothesised that participants would be overconfident on the Gc tasks in Study 1 with positive bias scores, and calibration curves that display overconfidence. This hypothesis was developed from the calibration literature wherein overconfidence has consistently been demonstrated on Gc tasks (Allwood & Granhag, 1996; Crawford & Stankov, 1996; Juslin, 1994; Kleitman & Stankov, 2001; Pallier et al., 2002; Schaefer et al., 2004; Stankov, 1998; Stankov & Crawford, 1996a, 1996b, 1997; West & Stanovich, 1997).
3. It was hypothesised that males would be more confident than females for all five cognitive tasks. Previous research which found that males have demonstrated significantly higher levels of confidence than females on cognitive tasks formed the foundation of this hypothesis (Pallier, 2003; Pulford & Colman, 1997; Ross & Fogarty, 2006; Stankov, 1998).
4. It was hypothesised that age would be positively related to Gf, Gc, and Gv bias scores. Calibration studies that demonstrated evidence of greater mis-calibration for older participants, and small but significant positive correlations between age and bias scores obtained from Gf, Gc, and Gv tasks (Crawford & Stankov, 1996; Pallier,

2003), underpinned this hypothesis. This same pattern of results was expected for the Gc, Gf, and Gv tasks in the current study.

5. It was hypothesised that low scorers (i.e., in the bottom quartile) would be more mis-calibrated than high scorers (i.e., top quartile) for each of the cognitive tasks used in Study 1. Previous research that demonstrated that those of lower ability have some difficulty in accurately appraising their cognitive abilities compared with those of higher ability (Dunning et al., 2003; Kruger & Dunning, 1999; Maki et al., 1994; Moreland et al., 1981; Shaughnessy, 1979), formed the rationale for this hypothesis.
6. Based on the results by Schaefer et al. (2004) it was hypothesised that, after controlling for the influence of the other four personality dimensions, that:
    - 6.1. Openness/Intellect scores would be significantly related to confidence scores.
    - 6.2. Extraversion would be associated with confidence and bias scores.
    - 6.3. Conscientiousness would be significantly related to confidence scores.
    - 6.4. Emotional Stability and Agreeableness scores would not be associated with cognitive confidence or bias.
7. It was hypothesised that good calibration was also expected for the Big Five judgments, based on Gigerenzer’s (1991) theory in which it was argued that individuals are well calibrated to their natural environments.
8. Based on the gender similarities hypothesis (Hyde, 2005) it was hypothesised that males and females would not differ in terms of Big Five confidence, Big Five accuracy, or Big Five bias.
9. Based on the five-factor theory of personality (McCrae & Costa, 1999) combined with the assumption that Big Five judgments are made on a daily basis, and

therefore have high ecological validity it was hypothesised that age would not be associated with Big Five confidence, Big Five accuracy, or Big Five bias.

10. It was hypothesised that two confidence factors would emerge from the structural analyses of confidence scores obtained from the cognitive (Gc, Gf, and Gv) and Big Five Measures. This hypothesis was developed with reference to Self-Concept theory (Marsh, 2008; Shavelson et al., 1976) and PMM theory (Gigerenzer et al., 1991).
11. It was hypothesised that one bias factor would emerge from factor analysing the bias scores obtained from the Gc, Gf, and Gv tasks used in the current study. Work by Stankov and his collaborators (Pallier et al., 2002; Stankov, 1998, 1999a) who factor analysed bias scores obtained from various combinations of Gc, Gf, Gv tasks and found that the bias scores loaded onto one factor, formed the groundwork for this hypothesis.
12. It was hypothesised that cognitive bias would not be associated with the personality bias score. PMM theory (Gigerenzer et al., 1991) provided the basis for this hypothesis.

3.8 Method

3.8.1 Participants

A total of 127 individuals participated in this study. The sample comprised 40 males and 87 females, ranging in age from 17 to 74 years (*M* = 34.42, *SD* = 12.76). The mean age of the males was 33.98 years (*SD* = 12.32 years). The mean age for females was 34.63 (*SD* = 10.02 years). The highest educational level of the sample varied from completion of grade 7, 8, 9, 10, 11, or 12 (*n* = 77), to completion of tertiary studies (*n* = 50). Sixty participants were enrolled in undergraduate Psychology courses at the

University of Southern Queensland and received 1% course credit toward their final grade in return for their participation.

Participants were recruited in two ways. Firstly, a sign-up sheet was placed on the Psychology department notice board at the University of Southern Queensland (see Appendix A). Snow ball sampling techniques were used to obtain a community sample of 67 participants. When community organizations were randomly approached they agreed to provide individuals with the rationale for the study; then respondents recommended it to others they thought may be interested in participating. The community sample came from both metropolitan and regional areas in Queensland, and these participants received the opportunity to enter a draw for cash prizes.

3.8.2 Materials

Demographic questions consisted of items regarding, gender, age, and highest level of education. All participants completed a battery of five cognitive tasks, one self-report personality inventory, and five short descriptions of personality.

3.8.2.1 Cognitive Tests

For each of the cognitive tests, participants provided an answer to every trial, as well as a confidence rating indicating how confident they were that the answer provided was correct. For the open-ended tests (i.e., General Knowledge, Letter Series, and Concealed Words), confidence ranged from 0% (Just guessing) to 100% (Absolutely certain). For the other multiple choice tests (Esoteric Analogies and Cattell's Matrices), the starting point on the confidence scale was 100/k, where k = the number of response alternatives (see Appendix B). There were two markers each for Gc (General Knowledge and Esoteric Analogies) and Gf (Letter Series and Cattell's Matrices). The marker for Gv was the Concealed Words task.

General Knowledge Test (GKT) - (Stankov, 1997). This test (see Appendix C) covered knowledge of diverse areas such as history and geography, and contained 20 open-ended items. For example, “What are BASIC, FORTRAN, and ALGOL?”

Letter Series Test (LST) - (Stankov, 1997). After being presented with a series of letters (e.g., A, D, G, J, ?) participants provided the next letter of the series. They responded to 12 trials within a time limit of four minutes (see Appendix D).

Esoteric Analogies Test (EST) - (Stankov, 1997). Participants chose words that completed verbal analogies for which four response options were provided. For example, LIGHT is to DARK as HAPPY is to GLAD, SAD, GAY, EAGER. A time limit of four minutes was imposed within which respondents were presented with 24 trials (see Appendix E).

Concealed Words Test (CWT) - (Stankov, 1997). Participants identified words in which parts of each letter were degraded. Participants responded to 26 trials within a time limit of two minutes (see Appendix F).

Cattell’s Matrices (CM)- (Stankov, 1997). From among six options, participants chose the design that completed a matrix (see Appendix G). They responded to 11 trials within a four minutes time limit.

3.8.2.2 Self-report Measures

The International Personality Item Pool Five-Factor Personality Scale (IPIP, Goldberg, 1999). The scale consists of 50 statements describing people’s behaviours and comprises five subscales (each consisting of 10 statements). The IPIP assesses personality across the following dimensions: Extraversion (e.g., “Am the life of the party”), Agreeableness (e.g., “Am interested in people”), Conscientiousness (e.g., “Am exacting in my work”), Emotional Stability (e.g., “Seldom feel blue”), and

Intellect (e.g., "Am quick to understand things"). Intellect is the equivalent of Openness in Big Five terminology.

Respondents indicated, on a 5-point Likert scale from 1 (*Very inaccurate*) to 5 (*Very accurate*), the extent to which each statement described them. After the negatively worded items were recoded, a total score for each subscale was calculated by summing the 10 scores. Scores for each of the five subscales can range from 10 to 50, with higher scores indicating higher levels of the particular personality dimension.

Goldberg (1999) reported that the IPIP was internally consistent, with a mean alpha coefficient of .84. Goldberg also reported adequate alpha coefficients for the Extraversion (.87), Agreeableness (.82), Conscientiousness (.79), Emotional Stability (.86), and Intellect (.84) sub-scales.

Reasons for choosing the IPIP for Study 1 included its prior use in other studies of interpersonal and intrapersonal perception, as well as its psychometric properties (e.g., Christiansen et al., 2005; Funder, 1999; Gosling et al., 1998; John & Robins, 1993; John & Robins, 1994; Levesque, 1997; Mount, Barrick, & Strauss, 1994; Watson et al., 2000).

Goldberg's (1999) work provided evidence for the convergent validity of the IPIP scales by showing significant positive correlations (ranging between .64 and .80) between these subscales and other personality measures including Cattell's 16 Personality Factors Questionnaire (Cattell, Eber, & Tatsuoka, 1992), Gough's California Psychological Inventory (Gough, 1996), Costa and McCrae's (1991) NEO-PIR, and Cloninger's Temperament and Character Inventory (Cloninger, Przybeck, Svrakic, & Wetzel, 1994). A copy of the IPIP is presented in Appendix H.

Big Five Block Descriptions of Personality (BFBD) Based on the Trait Adjectives - (Goldberg, 1999). The IPIP item-based personality descriptors described above are often referred to as “objective” measures of personality. It is well-known that they do not provide completely accurate assessments of personality but they do provide the most objective means of assessing this aspect of individual differences and, were therefore used as dependent variables in Study 1.

To continue the parallel with cognitive testing, a situation was constructed wherein these five item-based personality assessments could be scored as accurate or inaccurate. As mentioned in Chapter 2, strict parallels are not possible because, outside the limits imposed by reliability and validity estimates, there is no way to determine whether personality assessments are accurate. However, it is possible to approach this goal by using the notion of consistency/reliability. Consistency does not ensure accuracy, but a lack of consistency implies inaccuracy. Thus, a person who estimates a personality trait at one level using a reliable and valid form of a personality test but obtains a completely different estimate using a different but equally reliable and valid personality test, would be considered to be inaccurate. In this scenario, an assumption is made that the wildly different trait estimates are due to the person making the assessment, not to features of the instrument itself. Parallel forms of the IPIP Big Five measures were therefore developed. Rather than use another set of item-based measures (respondents had just completed one set), participants instead viewed text blocks presenting textual descriptions of the main personality traits, and then rated the extent to which each of these blocks described their own personalities.

The present study utilized five block personality descriptions, which were constructed using the trait adjectives from the 100 item IPIP scale (Goldberg, 1999). The block descriptions entailed grouping together eight to ten adjectives for each personality dimension. However, participants completed the 50 item IPIP in the same testing session as these block descriptions, which meant that practice effects may have been a concern. Therefore, the other 50 items that make up the 100 item scale supplied the trait adjectives for the block descriptions. In addition, these descriptions utilized only positively worded trait descriptions, because the group of items were presented in a block. Finally, in a further attempt to minimise overlap, the instructions and the rating scales were also changed (see below) to suit the purposes of the current research. A confidence measure (see below) was also added to each description. This resulted in 8 trait adjectives being used for the Extraversion dimension, 10 for the Agreeableness dimension, 10 for the Intellect dimension, 9 for the Emotional Stability dimension, and 9 for the Conscientiousness dimension.

For each Big Five personality dimension, participants viewed a block description and rated the extent to which the overall block description generally reflected their personality. They marked their choices on an 11-point scale with end points of -5 (*Not like me*) to + 5 (*Like me*). The Extraversion personality block description, in which eight trait adjectives were grouped, appears below:

> I don’t mind being the centre of attention; I make friends easily; I take charge; I know how to captivate people; I feel at ease with people; I am skilled in handling social situations; I am the life of the party; I start conversations.

In order to continue the analogy with calibration assessment in the cognitive domain, a situation needed to be devised wherein respondents rated the extent to which they were confident about their personality ratings. This is not the first time this has been done. As early as 1920, Lewis Terman began his famous study of gifted children, in which he assessed not only their intelligence but also their personality and social skills. His methodology involved asking parents and teachers to rate children on various traits using a single seven-point scale for each trait, and then asking parents/teachers to immediately indicate how certain they were about the judgment they had just made (very certain, fairly certain, rather uncertain, very uncertain) (Terman, as cited in Friedman & Schustack, 2009). A comparable confidence scale was used for the current study. Another reason for using this form of personality assessment as the parallel form (rather than another set of individual items), is that the block descriptors method lends itself to confidence ratings, because respondents are perfectly clear that they are rating themselves on these personality dimensions. Following each description, participants therefore provided their confidence rating after reading the following:

> Imagine that there was some device that could accurately tell us about your personality. How confident are you that the rating you gave above would correspond with the device's rating? Please rate your confidence on the scale that appears below by circling your level of confidence.

The confidence scale for the personality judgements used 10% intervals and ranged from 0% (*Just guessing*) to 100% *(Absolute certainty*). All five personality block descriptions are presented in Appendix I.

It is noted that within the neurosciences domain, researchers have not entirely neglected the area of individual differences in personality, and that modern textbooks on personality now routinely include descriptions of physiological approaches to measuring personality (Friedman & Schustack, 2009). Because physiological methods/measurements were beyond the scope of this dissertation, the current study developed an alternative method (described above) to obtain confidence ratings. Although the imagined device is not a physiological measure, it still constitutes a useful method of probing confidence in one's self-assessment of personality judgments. Despite the fact that respondents were not expected to experience difficulty bringing to mind such an imagined device, the current study was preceded by a pilot investigation that explored individuals' reactions to the instructions regarding this means of providing their confidence ratings. Individuals consistently reported that the instructions were easy to comprehend and apply.

Finally, to complete what is admittedly a forced analogy, accuracy and bias scores were obtained for the personality assessments by following the steps outlined below:

1. Convert each subscale score from the IPIP item-based personality assessments to a percentage.
2. Convert each of the Big Five block descriptors self-ratings to percentages. That is, -5 = 0%, -4 = 10%, -3 = 20%, -2 = 30%, -1 = 40%, 0 = 50%, 1 = 60%, 2 = 70%, 8 = 80%, 9 = 90%, 10 = 100%.
3. Subtract (2) from (1).
4. Code each score as either accurate or inaccurate. An arbitrary figure of 20% was chosen for the purposes of determining accuracy. That is, if the absolute value of the difference between the parallel forms was ≤ 20, then the score was coded accurate.
5. Calculate a mean accuracy score across the five dimensions.
6. Calculate a mean confidence rating score across the five dimensions.
7. **7.** Subtract (5) from (6).

3.8.3 Procedure

Participants were tested individually on a face-to-face basis after providing a rationale, and explaining that all data would remain confidential. An informed consent sheet was signed (see Appendix J) and participants were told that they could withdraw from the study at any time.

The test battery started with the GKT test. The order of the BFBD descriptors and the IPIP were randomised for each participant, with the only constraint being that each descriptor was followed by a cognitive task. Participants were not informed of their accuracy during testing but were given feedback at the end of the experiment. The battery took approximately one hour to complete.

Scoring: For the cognitive tasks described above, the correctness of every item was recorded (i.e., 1 = correct, 0 = incorrect) as well as the confidence rating for each item. The dependent variables for each cognitive task included:

1. A proportion correct score (mean accuracy score). That is, summing each individual's number correct and dividing this number by the number of items in the task.
2. A mean confidence rating score was calculated for each participant for each objective task (i.e., summing his or her confidence ratings for each task and dividing this number by the number of items in the task).
3. A bias score; that is, the mean confidence rating for each task minus the proportion correct for that task.

Subscale scores were calculated for the IPIP. For the gender variable, 0 = males and 1 = females.

3.9 Results

3.9.1 Data Screening

Prior to statistical analyses, all variables were examined using Version 16 of the Statistical Package for Social Scientists (SPSS, 2009) for accuracy of data input, missing data, and evaluation of the multivariate analysis assumptions of normality, linearity, multicollinearity, singularity, and univariate and multivariate outliers. No missing values were present in the data set. No problems were detected with the assumptions of linearity, multicollinearity, and singularity.

For several variables one or more univariate outliers were detected. Visual inspection of all data provided by these cases, however, suggested that they were plausible responses. These cases were retained as a legitimate variation (Tabachnick & Fidell, 1996). All subscales of the IPIP were normally distributed. Several other self-report and objective measures displayed problems with normality and/or kurtosis. Various transformations were applied which successfully reduced skewness and kurtosis for all variables. Statistical analyses were performed with both the transformed and untransformed data. As the transformed data did not alter the outcome of the multivariate statistical analyses, the untransformed data were retained in line with recommendations made by Tabachnick and Fidell (1996).

3.9.1.1 Reliability Analysis

Prior to statistical analyses, internal consistency reliability estimates (i.e., Cronbach's coefficient alphas) were calculated for all self-report and objective task variables. Descriptive statistics and alpha coefficients for all dependent variables are presented in Table 3.1 showing alpha coefficients ranged from .56 (Cattell's Matrices task) to .90 (Esoteric Analogies confidence score). Cattell's task displayed poor internal consistency but was still considered acceptable for use in experimental research (Gregory, 1996). Reliability coefficients for the IPIP subscales were similar to those reported by Goldberg (1999). Internal consistency estimates for the cognitive confidence variables were similar to those reported by other calibration researchers (e.g., Kleitman, 2003; Stankov & Lee, 2008).

Table 3.1
Descriptive Statistics for all Dependent Variables (N = 127)

Dependent Variables	M	SD	# items	α
Cognitive Accuracy				
GKTAC	40.08	19.65	20	.79
LSTAC	60.84	17.82	15	.74
CWTAC	26.47	11.13	26	.65
ESTAC	60.24	16.34	24	.73
CMAC	76.16	15.32	11	.56
Cognitive Confidence				
GKTC	44.72	19.84	20	.86
LSTC	60.57	15.12	15	.78
CWTC	26.42	12.52	26	.78
ESTC	65.55	16.61	24	.90
CMC	85.70	13.29	11	.86
Cognitive Bias Scores				
GKT bias	4.64	11.34		
LST bias	-.27	15.24		
CWT bias	-.05	6.66		
EST bias	5.31	13.07		
CM bias	9.54	16.01		
Big Five Block Accuracy Scores (BFBD Accuracy 20%)				
CONSCAC20%	71.65	45.25		
EXTRAAC20%	77.95	41.62		
AGREEAC20%	82.68	37.99		
EMOTAC20%	72.44	44.86		
INTAC20%	87.40	33.31		
Big Five Confidence				
CONC	80.08	14.17	1	-
EMOTC	78.11	15.30	1	-
INTELLC	78.11	14.73	1	-
EXTRAC	80.16	11.95	1	-
AGREEC	81.89	13.84	1	-
IPIP Subscales				
ICON	35.78	5.59	10	.73
IEMOT	32.44	7.83	10	.89
IINTELL	36.68	5.93	10	.82
IEXTRA	32.95	6.68	10	.83
IAGREE	41.00	4.64	10	.66

Note. GKT= General Knowledge; LST = Letter Series; EST = Esoteric Analogies; CWT = Concealed Words; CM = Cattell's Matrices; CONC= Conscientiousness; EMOT= Emotional stability; INTELL = Intellect; EXTRA = Extraversion; AGREE= Agreeableness; ICON = IPIP Conscientiousness; IEMOT = IPIP Emotional stability; INTELL = IPIP Intellect; IEXTRA = IPIP Extraversion; IAGREE = IPIP Agreeableness; CONSCAC20% = Conscientiousness Accuracy ≤ 20 %; EXTRAAC20% = Extraversion accuracy ≤ 20 %; AGREEAC20% = Agreeableness Accuracy ≤ 20 %; EMOTAC20% = Emotional Stability Accuracy ≤ 20 %; INTAC20% = Intellect Accuracy ≤ 20 %;.

3.9.1.2 Descriptive Statistics

Means and standard deviations of all the dependent variables are presented in Table 3.1. The table shows, bias scores for the cognitive tasks ranged from –.05 for the CW task to 9.54 for the CM task. Proportion correct scores for all objective measures represented traditional measures of difficulty. Participants found the Concealed Words task the most difficult with approximately 26% of items correctly solved. Respondents found Cattell's task relatively easy, correctly solving 76% of the items. Big Five block accuracy scores ranged from 71.65 for Conscientiousness, to 87.40 for Intellect.

3.9.2 Calibration Results for the Cognitive Tasks

3.9.2.1 Reasoning Tasks (Gf)

The first hypothesis stated that good calibration was expected for the reasoning tasks with bias scores being close to zero. The bias scores for the Letter Series and Cattell's Matrices Tasks were -.27 and 9.54 respectively. The calibration curve for the Letter Series task is presented in Figure 3.1. Inspection of this graph revealed that most of the curve is either close to the perfect calibration line in the overconfidence region or slightly above it. The impact of individual item means is presented in Figure 3.2. As can be seen, most of the items displayed reasonable calibration.

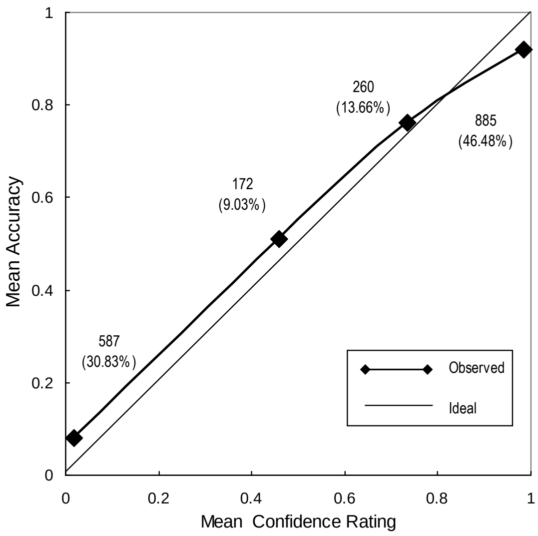


Figure 3.1. Calibration curve for the Letter Series Task.

Figure 3.3 displays the calibration curve for the Cattell's matrices task. Again, visual inspection of this graph without taking into account the relative frequencies may have led to an incorrect interpretation as 90.06% of the observations lie in the overconfidence region of the graph. The scatterplot of individual item means are presented in Figure 3.4 showing that two items at the lower levels of accuracy contributed to the overconfidence for this task.

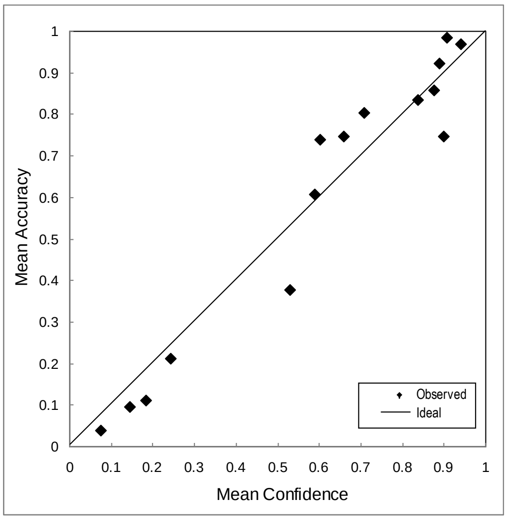


Figure 3.2. Scatterplot of mean confidence rating and mean accuracy scores for the Letter Series Task.

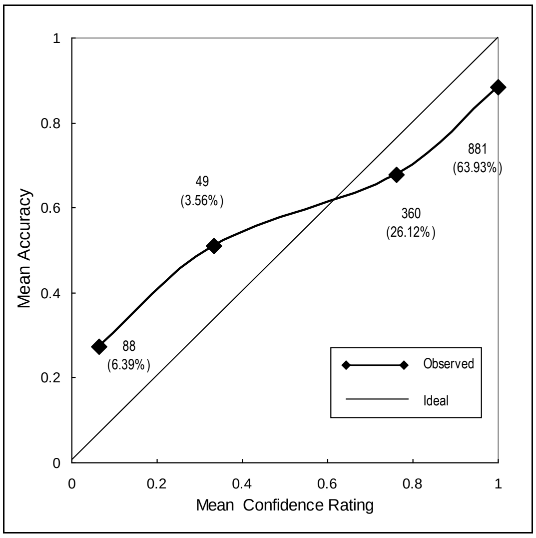


Figure 3.3. Calibration curve for the Cattell's Matrices Task.

3.9.2.2 Acculturated Knowledge Tasks (Gc)

The second hypothesis stated that individuals would be overconfident on tasks of acculturated knowledge. Both the General Knowledge and the Esoteric Analogies task demonstrated overconfidence, with bias scores of 4.64 and 5.31, respectively.

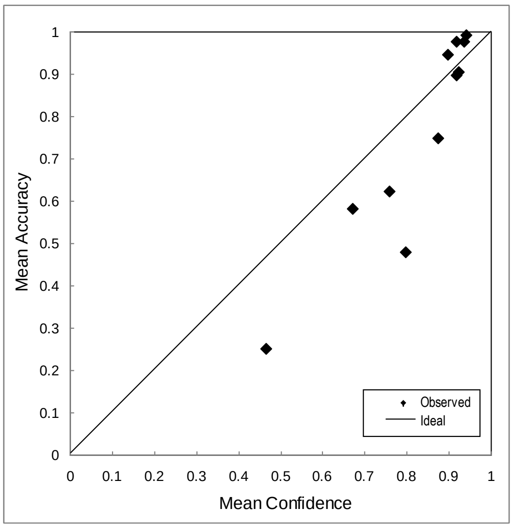


Figure 3.4. Scatterplot of mean confidence rating and mean accuracy scores for the Cattell's Matrices Task.

Figure 3.5 presents the calibration curve for the General Knowledge task. Visual inspection of this figure highlighted overconfidence because 55.61% of the observations are situated below the perfect calibration line, which results in a positive bias score.

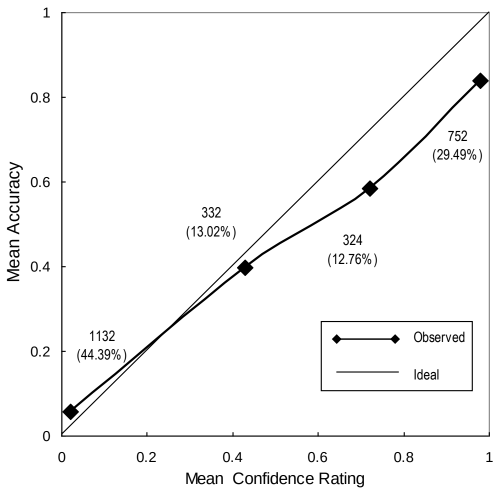


Figure 3.5. Calibration curve for the General Knowledge Task.

Figure 3.6 presents the impact of individual item means for this task, and shows that most of the items come close to the perfect calibration line. Only a few items at the lower levels of accuracy displayed overconfidence.

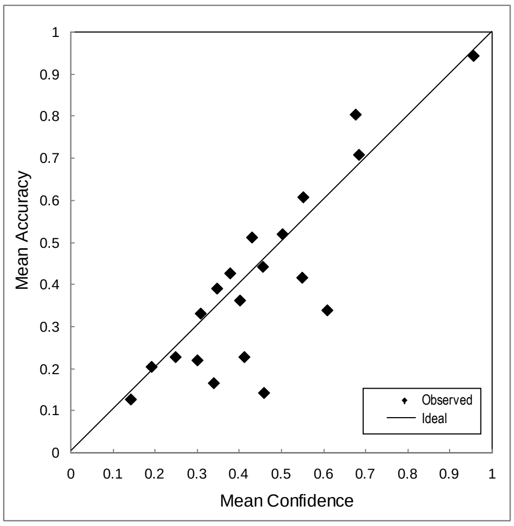


Figure 3.6. Scatterplot of mean confidence rating and mean accuracy scores for the General Knowledge Task.

Figure 3.7 presents the calibration curve for the Esoteric Analogies task. Visual inspection of this graph took the relative frequencies into account, to avoid being misled by the fact that the majority of the curve lies in the underconfidence region of the graph. The occurrence of 61.10% of observations in the overconfidence region, explained the positive bias score.

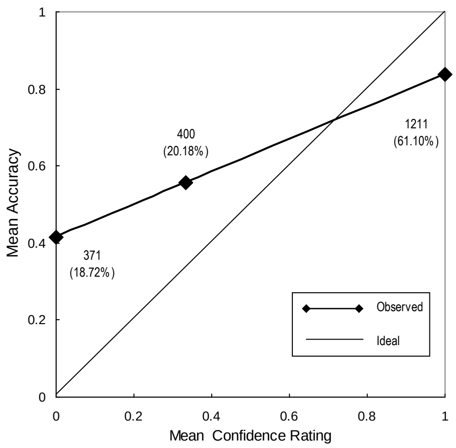


Figure 3.7. Calibration curve for the Esoteric Analogies Task.

Figure 3.8 displays the scatterplot of the item means. Observation of this scatterplot showed that, several items in the overconfidence region of the graph appeared to influence the bias score for this task.

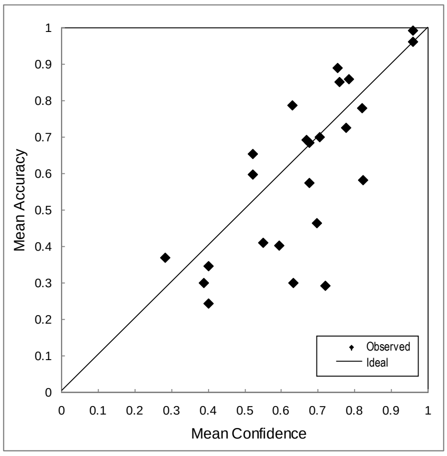


Figure 3.8. Scatterplot of mean confidence rating and mean accuracy scores for the Esoteric Analogies Task.

3.9.2.3 Visual Perceptual Task (Gv)

No hypothesis was made for the Concealed Words task; nevertheless the bias was examined and was found to be -.05, which indicated good calibration. Figure 3.9 presents the calibration curve for this task. Visual inspection of the relative frequencies clearly showed that the majority of observations were situated close to the perfect calibration line, or just above it in the underconfidence region of the graph. The item-specific scatter plot presented in Figure 3.10, shows that, most of the items were close to the perfect calibration line.

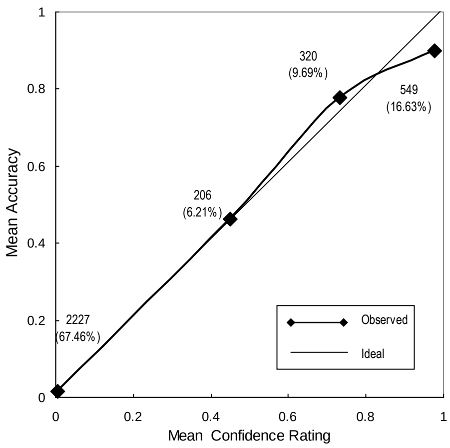


Figure 3.9. Calibration Curve of mean confidence rating and mean accuracy scores for the general Concealed Words Task.

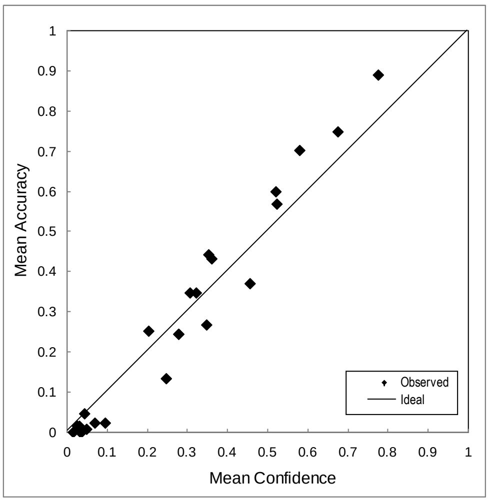


Figure 3.10. Scatterplot of mean confidence rating and mean accuracy scores for the general Concealed Words Task.

3.10 Individual Differences in Cognitive Confidence and Mis-calibration

3.10.1 Gender Differences in Cognitive Confidence

The third hypothesis stated that males would be significantly more confident than females, on tasks of cognitive abilities. No hypothesis was made in terms of gender differences in mis-calibration; nevertheless differences in bias scores were examined. The gender data were subjected to an independent samples *t* test with the results presented in Table 3.2 [1]. A bonferroni adjustment was made to control for family-wise error with the alpha level being set at 0.01. As expected, in terms of General Knowledge

---

[1] The male to female ratio in this study is not ideal, however, it closely resembles other calibration research wherein similar gender ratios were reported (e.g., Pallier, 2003). Also, based on the recommendations made by Howell (2002), each effect size calculation used the mean and the standard deviation for each gender as the denominator. This practice guarantees that **d** is approximated independently of *N*, thereby removing potential concerns regarding unequal sample sizes.

confidence significant differences arose between males and females (see Table 3.2). Contrary to predictions, however, no significant differences arose between the genders in terms of confidence for the other cognitive tasks. There were no significant differences between the genders with regard to cognitive mis-calibration. These results indicated that apart from males being more confident than females on the General Knowledge task, males and females did not differ in terms of cognitive confidence or mis-calibration.

Table 3.2
Means of Confidence and Bias Scores for Males (N = 40) and Females (N = 87) on Cognitive Tasks in Study 1.

	GKTC	GKT BIAS	LSTC	LST BIAS	CWTC	CWT BIAS	ESTC	EST BIAS	CMC	CM BIAS
Male	52.18	6.55	61.72	1.38	29.42	.481	69.41	6.59	88.14	11.32
Female	41.30	3.77	60.04	-1.03	25.04	-.30	63.77	4.72	84.58	8.71
t tests	**2.96****	1.29	.58	.83	1.85	.61	1.79	.75	1.40	.85

Note. gktc= General Knowledge confidence; gktbias= General Knowledge bias; letter series confidence = Letter Series confidence; lstbias = Letter Series bias; cwtc = Concealed Words confidence; cwtbias = Concealed Words bias; estc = Esoteric Analogies confidence; estbias = Esoteric Analogies bias; CMC = Cattell's Matrices confidence; CMBIAS = Cattell's Matrices bias *p* = .00.

3.10.2 Age differences in Cognitive Mis-calibration

Pearson Product Moment correlations were calculated in order to test the fourth hypothesis, which stated that there would be significant positive associations between age and bias scores. To help comparison with previous research, composite bias scores for Gc (average bias score on General Knowledge and Esoteric Analogies) and Gf (average bias score for Cattell's Matrices and Letter Series) were calculated, and were found to be correlated with age (see Table 3.3). For Gv, only one marker was available, therefore a composite variable could not be calculated. The hypothesis was partially supported, as age was significantly positively correlated with Gf and Gc bias scores, suggesting that older people were more mis-calibrated than younger people on Gc and Gf tasks.

Table 3.3
Correlations between Age and Bias Scores

	age	GKTBIAS	LSTBIAS	CWTBIAS	ESTBIAS	CMBIAS
age	1.00					
GCBIAS	**0.23***	1.00				
GFBIAS	**0.28****	0.16	1.00			
CWBIAS	0.15	0.36	0.32	1.00		

Note. GKTBIAS= General Knowledge bias; LSTBIAS = Letter Series bias; CWTBIAS = Concealed Words bias; ESTBIAS = Esoteric Analogies bias; CMBIAS = Cattell's Matrices bias
* $p < .05$. ** $p < .01$.

3.10.3 Mis-calibration and Ability

The fifth hypothesis stated that for the five cognitive tasks used in Study 1, low scorers (i.e., in the bottom quartile) are likely to be more mis-calibrated than high scorers (i.e., top quartile). Based on their accuracy scores, participants were divided into quartiles for each cognitive task. Independent samples t-tests were conducted to compare bias scores scored obtained by top and bottom quartile participants (see Table 3.4). A bonferroni adjustment was made to control for family-wise error with the alpha level being set at 0.01.

Table 3.4
Mean Bias Scores for First and Fourth Quartiles on Cognitive Tasks

	GKTBIAS	LSTBIAS	CWTBIAS	ESTBIAS	CMBIAS
Quartile 1	6.84	10.76	1.04	12.20	27.05
Quartile 4	-1.62	-9.00	0.73	-.87	-.51
t tests	3.65**	3.24**	.17	4.40**	7.68**

Note. gktbias= general knowledge bias; lstbias = letter series bias; cwtbias = concealed words bias; estbias = esoteric analogies bias; cmbias = Cattell's matrices bias
** $p = .00$.

Results indicated that low scorers were significantly more miscalibrated than high scorers, for all cognitive tasks except for the Concealed Words task. The hypothesis was therefore supported for four out of the five cognitive tasks, suggesting that in this sample, those who know more know more about what they know.

3.10.4 Personality Correlates of Cognitive Confidence Judgments

Hypothesis 6.1 proposed that there would be a significant positive association between Intellect and cognitive confidence, and hypothesis 6.2 postulated that positive associations would occur between Extraversion and both cognitive confidence and bias scores. Hypothesis 6.3 suggested a significant positive relationship between Conscientiousness and cognitive confidence. Hypothesis 6.4 proposed that Emotional Stability and Agreeableness would not be associated with cognitive confidence or bias scores.

Both zero order and partial correlations were used to test all these hypotheses (see Table 3.5). To simplify results, composite scores were created for the cognitive confidence (average confidence score across all five cognitive tasks), accuracy (average accuracy score across all five cognitive tasks), and bias variables (average bias score across all five cognitive tasks). Zero order correlations are presented in the top portion of this table and partial correlations are presented below these. Two salient features emerged: Intellect was significantly positively correlated with cognitive confidence and accuracy, and Conscientiousness was significantly positively correlated with cognitive bias. The overall conclusions did not differ, whether zero-order correlations or partial correlations were used to test the hypotheses.

Table 3.5

Correlations Among IPIP Subscale Scores and Accuracy, Confidence and Bias and Partial Correlations Between Big Five Scores and Accuracy, Confidence and Bias (N = 127)

	IEXTRA	IAGREE	ICON	IEMOT	IINTELL [a]	COGCON1	COGAC1	COGBIAS1
IEXTRA	1.00							
IAGREE	0.18	1.00						
ICON	0.10	0.20	1.00					
IEMOT	0.12	-0.01	0.14	1.00				
IINTELL	0.35	0.11	0.03	-0.12	1.00			
COGCON1	0.11	0.01	0.14	0.06	0.21**	1.00		
COGAC1	0.04	-0.02	-0.07	-0.03	0.47**	0.70	1.00	
COBIAS1	0.09	0.04	0.27**	0.13	0.01	0.39	-0.39	1.00
Partial Correl- ations	IEXTRA[1]	IAGREE[1]	ICON[1]	IEMOT[1]	IINTELL[2]			
COGCON1	-0.08	-0.07	0.15	0.13	0.23**			
COGAC1	-0.12	-0.05	-0.08	-0.03	0.49**			
COGBIAS1	0.05	-0.02	0.26**	0.13	-0.08			

Note. cogcon1= Cognitive confidence; cogac1 = Cognitive accuracy; cogbias1 = Cognitive bias; [a] = controlling for accuracy; 1 = controlling for the other four personality dimensions; [2] = controlling for the other four personality variables and accuracy.
* $p < .05$. ** $p < .01$.

3.11 Individual Differences in Big Five Accuracy, Confidence and Mis-calibration

Because this study presented the BFBD measures in a novel format, it was important to first check the psychometric properties of the scales. Beginning with the IPIP item-based measures, the internal consistency reliability estimates have already been reported as being similar to those reported in the literature. The pattern of correlations among the IPIP scales was also similar to that recorded in the literature (see Saucier & Goldberg, 1996), which is not surprising given that the item-based IPIP measures were administered and scored in the usual way. The block descriptors of personality (BFBD), however, were an original feature of this study, and there was a possibility that measures derived from these descriptors might not have behaved in the expected way. Such an outcome would have rendered invalid any attempt to calculate accuracy measures based on differences between what were meant to be parallel tests of personality. To check this possibility, correlations were calculated and no significant differences emerged ($p < .05$) between the correlations obtained using the block descriptors and those obtained from the traditional IPIP item-based method. Factor analysis of a correlation matrix, formed by combining all 10 measures yielded a clear five-factor solution with both measures from each of the two forms pairing to define the Big Five factors (see Appendix K). These outcomes were taken as convincing evidence that the attempt to construct a parallel measure of the Big Five was successful.

The point of constructing this parallel form measure was to allow participants to rate their own personality traits and to express confidence in those ratings. Although interesting in themselves, these confidence ratings also served as one of the components needed to calculate bias scores. The other part of the bias score was the accuracy score, formed by first converting both sets of scores to percentages, and then subtracting the block descriptor scores from the IPIP item-based descriptor scores. Differences greater than 20% were given an accuracy score of zero; smaller differences were scored as correct. To recapitulate the rationale; if individuals gave themselves a rating on a particular trait that was more than 20% above or below the rating obtained via the usual objective method of assessing that trait, they were regarded as being incorrect in that assessment. The large interval (20%) acknowledged the fact that the objective measure itself is not perfectly reliable.

Descriptive statistics for these accuracy scores appear in Table 3.6. The means of these difference scores suggest that, when converted to percentages, the scores across the parallel forms were similar. However, the standard deviation of the difference scores revealed many inaccurate ratings for some traits.

Table 3.6
Descriptive Statistics for Big Five Accuracy Scores (N= 127)

Big Five Accuracy	*M*	*SD*
CONSCAC20%	71.65	45.25
EXTRAAC20%	77.95	41.62
AGREEAC20%	82.68	37.99
EMOTAC20%	72.44	44.86
INTAC20%	87.40	33.31

Note. CONSCAC20% = Conscientiousness Accuracy ≤ 20 %; EXTRAAC20% = Extraversion accuracy ≤ 20 %; AGREEAC20% = Agreeableness Accuracy ≤ 20 %; EMOTAC20% = Emotional Stability Accuracy ≤ 20 %; INTAC20% = Intellect Accuracy ≤ 20 %;

The original intention was to combine the five accuracy scores (percentage correct across the five traits) and use this figure in the bias calculation, in exactly the same way that accuracy estimates are obtained for cognitive tasks by averaging performance on individual items. However, when these accuracy scores were formed for the five traits, it was apparent that, for the most part, they were uncorrelated. In other words, using this experimental measure of accuracy, people who were consistent (accurate) in their scores on one trait were not necessarily consistent in their scores on other traits. The lack of intercorrelations among the accuracy measures meant that an overall bias score obtained from the personality measures was questionable. For this reason, in this study, examination of individual differences in gender and age were restricted to the confidence and accuracy scores for each Big Five dimension. Consequently, hypotheses 7, 8, 9, and 12, related to the personality bias score, were not examined.

Nevertheless, prior to investigating gender and age differences in Big Five confidence and accuracy, correlations between Big Five accuracy scores and scores of cognitive accuracy, confidence, and bias were examined. These correlations appear in Table 3.7. For the sake of brevity, composite scores for Gc (average accuracy/bias/confidence scores on General Knowledge and Esoteric Analogies tasks) and Gf (average accuracy/bias/confidence scores for Cattell’s Matrices and Letter Series tasks) were calculated, and these composite scores were correlated with Big Five accuracy scores. A salient feature was that Emotional Stability accuracy was negatively correlated with Gf bias, suggesting that as Emotional Stability accuracy increased, Gf bias decreased. There were also positive correlations between Intellect accuracy and both Gf and Gc accuracy scores, indicating that as Gf and Gc accuracy increased, so did Intellect accuracy. Similarly, significant positive correlations occurred between Intellect accuracy and both Gf and Gc confidence. Overall, these correlations suggested that individuals who scored higher for accuracy on the Intellect dimension were also more confident and more accurate for both Gf and Gc tasks. Partial correlations between Intellect accuracy and both Gf and Gc confidence were also examined, controlling for Gf and Gc accuracy respectively. These partial correlations showed that Intellect accuracy did not correlate with either Gf or Gc confidence, suggesting that the positive correlations between Intellect accuracy and both Gf and Gc confidence were mediated by Gf and Gc accuracy scores, respectively.

3.11.1 Gender Differences in Big Five Confidence and Accuracy

The next hypothesis proposed that males and females would not differ in terms of Big Five confidence or accuracy and independent samples t-tests were used to test this hypothesis. Bonferroni adjustment was used to keep family-wise error at alpha = .05. Table 3.8 and Table 3.9 present the statistical analyses[2]. Both tables show that none of the t-tests indicated significant gender differences and the effect sizes were in

---

[2] The male to female ratio in this study is not ideal, however, it closely resembles other calibration research wherein similar gender ratios were reported (e.g., Pallier, 2003). Also, based on the recommendations made by Howell (2002), each effect size calculation used the mean and the standard deviation for each gender as the denominator. This practice guarantees that d is approximated independently of N, thereby removing potential concerns regarding unequal sample sizes.

the close-to-zero, or small range, as expected. These results indicate that males did not differ from females in terms of either Big Five confidence or accuracy scores.

Table 3.7
Correlations among Big Five Accuracy Scores and Cognitive Confidence, Accuracy and Bias Scores (*N* = 127).

Variable	CONSCAC20%	EXTRAAC20%	AGREEAC20%	EMOTAC20%	INTAC20%
CONSCAC20%	1.00				
EXTRAAC20%	**0.26****	1.00			
AGREEAC20%	-0.01	**0.21***	1.00		
EMOTAC20%	-0.04	0.10	-0.05	1.00	
INTAC20%	-0.08	0.14	**0.33****	0.08	1.00
GCACC	0.07	-0.01	0.08	-0.01	**0.20***
GFACC	-0.07	-0.01	-0.06	0.07	**0.21***
GFBIAS	0.12	0.04	0.14	**-0.25****	-0.06
GCBIAS	0.15	0.03	0.01	-0.06	-0.08
GFCONF	0.05	0.03	0.08	-.17	**0.18***
GCCONF	0.10	0.01	0.16	0.02	**0.20***

Note. CONSCAC20% = Conscientiousness Accuracy ≤ 20 %; EXTRAAC20% = Extraversion accuracy ≤ 20 %; AGREEAC20% = Agreeableness Accuracy ≤ 20 %; EMOTAC20% = Emotional Stability Accuracy ≤ 20 %; INTAC20% = Intellect Accuracy ≤ 20 %; GCACC = Gc accuracy; GFACC = Gf accuracy; GFBIAS = Gf bias; GCBIAS = Gc bias; GFCONF = Gf confidence; GCCONF = Gc confidence.
* $p < .05$. ** $p < .01$.

Table 3.8
Mean Big Five Confidence Scores for Males (n = 40*) and Females (n* = 87)

Variable	CONC	INTELLC	AGREEC	EXTRAAC	EMOTAC
Male	79.25	79.25	79.50	79.75	76.00
Female	80.46	77.58	82.99	80.34	79.19
t tests	-.45	.59	-1.32	-.26	-1.10
Effect size	-0.09	0.12	-0.26	-0.05	-0.20

Note. CONC= conscientiousness confidence; INTELLC = intellect confidence; AGREEC= agreeableness confidence; EXTRAC = extraversion confidence; EMOTC= emotional stability confidence.

Table 3.9
Mean Big Five Accuracy Scores for Males (n = 40*) and Females (n* = 87)

Variable	CONSC AC20%	INT AC20%	AGREE AC20%	EXTRAC20 %	EMOTC20 %
Male	67.50	82.50	75.00	82.50	82.50
Female	73.56	89.66	86.21	75.86	67.82
t tests	-.70	-1.04	-1.42	.83	1.86
Effect size	-0.13	-0.18	-0.25	0.15	0.33

Note. CONSCAC20% = Conscientiousness Accuracy ≤ 20 %; EXTRAAC20% = Extraversion accuracy ≤ 20 %; AGREEAC20% = Agreeableness Accuracy ≤ 20 %; EMOTAC20% = Emotional Stability Accuracy ≤ 20 %; INTAC20% = Intellect Accuracy ≤ 20 %.

3.11.2 Age Differences in Big Five Confidence and Accuracy

Hypothesis 9 proposed that age would not be associated with either Big Five confidence or Big Five accuracy. Pearson’s Product Moment correlations were calculated in order to test both aspects of this hypothesis, and the only significant correlations were between Intellect confidence and age ($r_{(125)}$ = .19, p < .05), and Intellect accuracy and age ($r_{(125)}$ = -.21, p < .05). These correlations suggest that older participants were more confident than younger participants were when rating their confidence for the Intellect dimension; and were less accurate than were younger participants' when rating themselves on the Intellect dimension. However, the effect sizes were small.

3.12 The Factorial Structure of Cognitive and Big Five Confidence

According to hypothesis 10, two confidence factors were expected to emerge from the structural analysis of the confidence ratings obtained from both the cognitive domain and the Big Five personality domain. Exploratory factor analysis was used to test this hypothesis. Principal Axis factoring with oblique rotation was undertaken with the five Big Five personality confidence scores, as well as with the five cognitive confidence rating scores from the objective tasks. The correlation matrix of these psychometric variables is presented in Table 3.10.

Table 3.10
Correlation Matrix of Cognitive and Personality Confidence Variables (N= 127)

Variable	CONC	INTELLC	AGREEC	EXTRAC	EMOTC	GKTC	LSTC	CWTC	ESTC	CMC
CONC	1.00									
INTELLC	0.38**	1.00								
AGREEC	0.49**	0.54**	1.00							
EXTRAC	0.46**	0.32**	0.36**	1.00						
EMOTC	0.49**	0.42**	0.55**	0.41**	1.00					
GKTC	-0.19*	0.07	-0.09	-0.11	-0.03	1.00				
LSTC	-0.07	0.01	0.04	-0.06	0.02	0.33**	1.00			
CWTC	-0.12	0.11	0.00	-0.04	0.11	0.12	0.15	1.00		
ESTC	0.06	0.15	0.12	0.10	0.10	0.53**	0.33**	0.19*	1.00	
CMTC	0.01	0.22*	0.09	0.09	0.10	0.37**	0.54**	0.23*	0.52**	1.00

Note. CONC= Conscientiousness confidence; INTELLC = Intellect confidence; AGREEC= Agreeableness confidence; EXTRAC = Extraversion confidence; EMOTC= Emotional Stability confidence; GKTC= General Knowledge confidence; LSTC= Letter Series confidence; ESTC = Esoteric Analogies confidence; CWTC = Concealed Words confidence; CMTC = Cattell's Matrices confidence
* p < .05 ** p < .01

The data from Table 3.10 were considered factorable as all assumptions as advocated by Coakes and Steed (1996) were met. A solution employing root one criterion produced two factors, which accounted for 52.80% of the total variance. Inspection of Cattell's Scree Plot supported a two factor solution. The pattern matrix, percent of variance accounted for, eigenvalues, communalities, and factor correlation matrix for the two-factor solution are presented in Table 3.11.

The first factor was labelled Cognitive Confidence because all five confidence scores from the objective tasks loaded on it. The second factor comprised high loadings from all Big Five confidence ratings and was labelled Big Five confidence. The factor correlation matrix presented in Table 3.11 indicated that factors 1 and 2 were not significantly related suggesting that Cognitive Confidence and Personality Confidence are not driven by the same cognitive processes.

Table 3.11
Summary of Exploratory Structural Analysis Results for Cognitive Confidence and Big Five Personality Confidence Scores, Using Principal Axis Factoring with Oblique Rotation (N = 127)

Variable	$h^{2(b)}$	F1[a]	F2
CONC	.51	-.14	.71
INTELLC	.40	.14	.61
AGREEC	.56	.01	.74
EXTRAC	.33	-.04	.57
EMOTC	.50	.04	.70
CWTC	.08	.28	.00
GKTC	.40	.62	-.14
ESTC	.50	.69	.11
CMC	.59	.75	.11
LSTC	.34	.58	-.04
Eigenvalues		2.89	2.40
% of variance		28.84	23.96
Factor Correlation Matrix			
	F1	F2	
F1	1.00		
F2	.05	1.00	

Note. $h^{2(b)}$ = Communalities
[a] F1 = Cognitive Confidence; F2 = Personality Confidence

Big Five ratings were not included in the factor analysis because it was assumed that the rating for each Big Five block dimension would not be highly correlated with confidence in that rating. There were only two significant correlations for the

Agreeableness (r = .62, p < .01) and Emotional Stability (r = .28, p < .01) dimensions between the block rating and its respective confidence rating.

3.13 Factor Structure of Cognitive Bias Scores

Principal axis factoring tested the hypothesis that a single bias factor would emerge from the bias scores obtained from the five cognitive tasks. The correlations among the cognitive bias scores are presented in Table 3.12

A solution employing root one criterion produced one factor, which accounted for 45.57% of the total variance. Table 3.13 presents the percent of variance accounted for, the eigenvalues, and the communalities. The factor extracted from the data set was called Cognitive Bias because all five bias scores loaded highly on it. This finding was in line with expectations and further supported the premise that mis-calibration is not a domain specific phenomenon within the cognitive domain. That is, individuals who are mis-calibrated in one domain are also mis-calibrated across other cognitive domains.

Table 3.12
Correlations between Cognitive Bias Scores (N = 127)

Variable	GKTBIAS	LSTBIAS	CWTBIAS	ESTBIAS	CMBIAS
GKTBIAS	1.00				
LSTBIAS	0.17*	1.00			
CWTBIAS	0.35**	0.32**	1.00		
ESTBIAS	0.28**	0.27**	0.39**	1.00	
CMBIAS	0.21*	0.37**	0.36**	0.43**	1.00

Note. GKTBIAS= General Knowledge bias; LST BIAS= Letter Series bias; ESTBIAS = Esoteric Analogies bias; CWTBIAS = Concealed Words bias; CMBIAS = Cattell's Matrices bias
* *p < .05* ** *p < .01*

Table 3.13
Principal Axis Factoring of Cognitive Bias Scores (N = 127)

Variable	Communalities	F1-Cognitive Bias
GKTBIAS	.19	.56
LSTBIAS	.24	.62
CWTBIAS	.41	.73
ESTBIAS	.39	.72
CMBIAS	.40	.72
Eigenvalues	-	2.29
% of variance	-	45.57

3.13.1 Correlations between the Personality Bias Score and Cognitive Bias Scores

The last hypothesis which proposed that cognitive bias scores would not be associated with an overall Big Five bias score was not investigated because of the lack of correlations among the Big Five accuracy measures.

3.14 Discussion

The present study was designed to: (a) investigate whether individuals are mis-calibrated across a diverse battery of cognitive tasks; (b) examine if gender, age, and personality, are correlated with cognitive accuracy, confidence, and mis-calibration; (c) explore if those of lower ability, are more mis-calibrated than those of higher ability; (d) examine the factorial structure of cognitive confidence judgments; (e) construct self-rated personality descriptions based on the Big Five trait adjectives by Goldberg (Goldberg, 1997); (f) obtain confidence and accuracy measures for each of the Big Five personality dimensions; (g) investigate the factorial structure of cognitive and personality confidence judgments; (h) examine the factorial structure of cognitive bias scores; and (i) examine the relationship between cognitive and personality bias.

Results supported the hypothesis that participants would be overconfident on the General Knowledge and Esoteric Analogies tasks. The bias scores were positive and close enough to five to indicate some overconfidence. These findings were largely consistent with previous calibration research (Allwood & Granhag, 1996; Crawford & Stankov, 1996; Juslin, 1994; Kleitman, 2003; Kleitman & Stankov, 2001; Schaefer et al., 2004; Stankov, 1998; Stankov & Crawford, 1996a, 1996b, 1997; West & Stanovich, 1997). Overconfidence for these tasks appeared to be due to the effect of three or four items that displayed overconfidence at the lower levels of accuracy. The hard-easy effect may provide one reason for the overconfidence on these tasks. These results support Gigerenzer et al.'s (1991) theory which asserts that overconfidence is the product of ecological factors. That is, because the items for both of these tasks were not selected at random, the items are unrepresentative of the participants' ecology, thereby leading to overconfidence.

The next hypothesis expected good calibration for the reasoning tasks. Bias scores provide partial support for this hypothesis. The bias score for the Letter Series task was very close to zero, and was consistent with previous research (e.g., Baker, 2001; Crawford & Stankov, 1996; Kleitman & Stankov, 2001; Stankov, 1998; Stankov & Crawford, 1996a, 1996b). Conversely, the positive bias score for the Cattell's Progressive Matrices test was more moderate in magnitude. The scatterplot of item means for this task indicated that the hard-easy effect may provide one explanation for these findings, as participants were overconfident on two items thereby influencing the bias score. The finding for the matrices task was consistent with studies that also found moderate levels of overconfidence (Pallier, 2003; Stankov & Dolph, 2000).

Although no predictions were made for the concealed words task, the bias score indicated that people were well calibrated in this study. This finding was inconsistent with that of Pallier et al. (2002, Study 1) who found that participants were overconfident. The differing samples may explain this inconsistency. That is, participants in Pallier et al's study were military personnel who may have tried to reduce cognitive dissonance (e.g., I believe I am bright therefore my answer must be correct) by elevating their confidence ratings. This explanation is possible because military participants were overconfident across all of the Gv, Gc, and Gf tasks used in their study.

Males were expected to be more confident than females on tasks of cognitive abilities. The data provided partial support for this hypothesis, as males were significantly more confident than females on General Knowledge questions. This result is consistent with previous research (Pallier, 2003). Contrary to expectations, however, males and females did not differ in terms of cognitive confidence for the other four tasks used in the current study. These findings differ from previous research wherein males were significantly more confident than females (Pallier, 2003; Pulford & Colman, 1997; Ross & Fogarty, 2006; Stankov, 1998). The accuracy scores for males in the current study, however, are lower across all four tasks than those reported by previous calibration researchers, and it is possible that males may have decreased their confidence levels accordingly, thereby eliminating gender differences in confidence ratings. The lack of gender differences in the current study were, however, consistent with more recent work by Stankov and Lee (2008)

Older participants were expected to be more mis-calibrated than younger participants, a hypothesis which was partially supported by the findings that age was positively correlated with Gf and Gc bias scores. For the Gv task, however, age differences did not appear. The findings for the Gf and Gc bias scores are consistent with previous research (Crawford & Stankov, 1996; Pallier, 2003). Mean Gf and Gc bias scores for older participants in the current study demonstrated overconfidence, which indicates that older people were more confident than they were accurate. These findings fit well with cognitive dissonance theory (Festinger, 1957), which posits that individuals prefer their cognitions, including those about their actions, to be consistent with one another. Dissonance occurs when these cognitions are inconsistent, and the individual is motivated to make them more consistent in order to decrease uneasiness or distress. Accordingly, Blanton et al. (2001) argued that unwarranted confidence occurs with judgments that challenge a positive view of the self. It is plausible to speculate that older participants provided higher confidence ratings on Gf and Gc tasks to reduce cognitive dissonance and maintain a positive view of the self. Perhaps older people place higher value on judgments that contain knowledge and reasoning components, because these judgments are more similar to those made in daily life than judgments about perceptual tasks, which are far removed from real-world decision making. This may explain why age did not correlate significantly with the bias score from the Concealed Words task, an outcome that is inconsistent with previous work, in which a significant positive correlation was obtained between age and bias on a Line Length task (Crawford & Stankov, 1996). This inconsistency may also be explained by sample differences. Crawford and Stankov (Crawford & Stankov, 1996) recruited older participants from The University of the Third Age, where individuals attend social gatherings and lectures given by retired professionals. Their convenience sample may not have been representative of the population of older adults thereby limiting the generalisability of their results. To investigate how these types of tasks influence confidence ratings, future researchers could examine age differences on a larger battery of Gv tasks, as well as ask participants to provide think aloud protocols when supplying their confidence ratings.

The hypothesis that low scorers on tasks of cognitive abilities are more mis-calibrated than high scorers is partially supported. Significant differences were found for all tasks except the Concealed Words task. Participants found this exercise very difficult, with both top and bottom quartile participants performing poorly. Data indicate that both groups must have known they had answered questions incorrectly, and decreased their confidence ratings accordingly. As both groups were well calibrated, group differences were eliminated. Alternatively, low reliability for this task may explain the inconsistency. The significant differences between top and bottom quartile participants in terms of mis-calibration for the other four tasks, parallel the work of

Kruger and Dunning (1999) who showed that low scorers overrated their test performance relative to their peers. Ability differences were re-examined in Study 3.

In accordance with expectations and previous research, Intellect was positively associated with cognitive confidence and accuracy (Baker, 2001; Kleitman, 2003; Pallier et al., 2002). However, contrary to expectations, Conscientiousness was positively associated with cognitive bias. Although this result is inconsistent with the findings of Schaefer and colleagues (2004), there are other data that demonstrate small but significant positive correlations between Conscientiousness and cognitive bias scores (Kleitman, 2003). To date, Conscientiousness has not been shown to follow a consistent pattern of associations with bias scores from the cognitive domain. Because its contribution to cognitive bias remains unclear, the role of Conscientiousness was re-examined in Study 3. More importantly the current study utilized both zero order and partial correlations to investigate the claim by Schaefer and his colleagues, that partial correlations provide the clearest conclusions when examining associations between Big Five personality dimensions and both cognitive confidence and bias scores. As shown in Table 3.5, the overall conclusions do not differ whether zero order or partial correlations were examined. To test the legitimacy of these overall conclusions, Study 3 re-examined the personality correlates of cognitive confidence and bias, using both zero-order and partial correlations.

The lack of intercorrelations among the Big Five accuracy measures meant that the bias score obtained from the personality measures was questionable. Hence personality calibration was not investigated. The hypotheses that were related to the personality bias score were also not investigated. The lack of intercorrelations between the Big Five accuracy scores suggested that accuracy for each dimension was driven by differing cognitive processes. Thus, people who were consistent (accurate) in their scores on one trait were not necessarily consistent (accurate) in their scores on other traits. The lack of intercorrelations between the Big Five accuracy scores does, however, make sense. For example, Joe knows he is really conscientious because he works long hours and receives positive feedback from his supervisor and colleagues. Joe is more likely to provide consistent responses across both the Conscientiousness subscale of the IPIP and the Conscientiousness block measure, than an individual who has not thought about him or herself in this way. Thus, for Joe, consistency of responding across two measures of the same personality dimension is influenced by the ecological validity of the Conscientiousness dimension. Moreover, Joe knows he is conscientious and agreeable; however, he is not sure if the intellect, extraversion or emotional stability dimensions describe him, so his ratings for those dimensions may not be consistent. In Joe's case, to assume that all his accuracy scores are going to be significantly intercorrelated, assumes that all Big Five traits have equal amounts of ecological validity. It also assumes that random error (e.g., fatigue, boredom etc) does not affect consistency of responding but this assumption may not be realistic. Accuracy scores were re-examined in Studies 2 and 3, to determine the replicability of the non-significant low correlations between the Big Five accuracy scores.

An interesting pattern of correlations did, however, emerge when Big Five accuracy scores were correlated with accuracy, confidence and bias scores for both the acculturated knowledge (Gc) and reasoning (Gf) domains. Intellect accuracy scores were significantly related to confidence and accuracy for both Gf and Gc. However, partial correlations indicated that when accuracy was partialled out, the only significant correlations occurred between both Gf and Gc accuracy, and Intellect accuracy. These correlations suggested that individuals who were coded as being accurate for the Intellect dimension were also more likely to be more intelligent. The positive correlations between Intellect accuracy and Gf and Gc accuracy scores are largely consistent with previous research within the cognitive domain where positive correlations have been demonstrated between Intellect and Gf and Gc accuracy scores (Pallier et al., 2002). Although findings regarding personality accuracy scores need to be replicated, that there are strong indications in these data that the experimental personality accuracy measures have yielded data that are of psychological interest. The correlations between personality accuracy scores and Gf accuracy scores were re-examined in Study 3.

When the block description method was used, it was not possible to develop a bias score for each of the Big Five personality dimensions because of the lack of correlation among the accuracy scores. This lack of correlation is problematic because a major aim of the current studies was to examine the factorial structure of bias scores obtained from both the Big Five and cognitive domains. Consequently, Study 2 focussed on trying to address this issue, and on developing various accuracy methods. This enabled bias scores for each Big Five personality dimension to calculated, which assisted the investigation of the factorial structure of cognitive and Big Five bias scores in Study 3.

For the reasons stated above, in this study, examination of individual differences in both gender and age were also restricted to the confidence and accuracy scores for each Big Five dimension. The hypothesis that males do not differ from females in terms of Big Five confidence or accuracy, was upheld in the current study. In accordance with the gender similarities hypothesis, it appears that males and females are more similar than different. Furthermore these findings fit well with PMM theory (Gigerenzer et al., 1991) because individuals are likely to be familiar with Big Five judgments, given that similar judgments are made on a daily basis. However, because of the exploratory nature of the current study, it is too early to draw conclusions about gender differences in either Big Five confidence or Big Five accuracy. Studies 2 and 3 returned to the examination of gender differences.

As expected, age was not associated with Big Five confidence scores, with the exception being that age was weakly positively correlated with Intellect confidence with only 4% of the variance explained. This finding may represent a Type I error, and requires replication before speculating why older participants were more confident for this dimension. Also as predicted, age was not associated with Big Five accuracy scores again with one qualification. That is, age was negatively correlated with Intellect accuracy, although the magnitude of this correlation was also small with only 4% of the variance explained by the correlation between age and Intellect accuracy. A Type I error may also be responsible for this finding, which needs to be replicated before any suppositions are made about why younger adults were more accurate for this dimension. Studies 2 and 3 re-examined age differences in Big Five confidence and accuracy. The remainder of the results are in concert with Five-Factor theory of personality (McCrae & Costa, 1999), which posits that personality traits develop during childhood and adolescence, and then remain stable in adulthood. It seems that Big Five accuracy and Big Five confidence remain constant too. These other findings also fit well with PMM theory (Gigerenzer et al., 1991) which argues that people are well calibrated to their natural ecology, therefore age differences should not be expected.

Two confidence factors did indeed emerge, as predicted, from the structural analysis of the confidence ratings obtained from both the cognitive and Big Five personality domains. These findings suggest that self-confidence is similar in factorial structure to the self-concept construct (see Marsh, 2008 for a review) which is highly differentiated across both the academic and non-academic domains. Therefore, variance in specific components cannot be explained in terms of one general trait. Self-confidence and self-concept resemble one another in that they both require cognitive appraisals of the self. It was therefore assumed that the factorial structure of both constructs is similar. The split between cognitive confidence and Big Five confidence at the factorial level, raises doubt about whether confidence judgments are driven by a general confidence trait. However, further conjecture requires replication, which is why the factorial structure of Big Five confidence and cognitive, were scrutinized again in Study 3. The existence of two separate confidence factors also supports PMM theory (Gigerenzer et al., 1991) which was interpreted for the purposes of the current study to mean that cue validities used to endorse personality test items were vastly different from those cues used to answer cognitive test items. Using this logic, it seems that participants do use differing cues to generate confidence judgments across these two domains, as the Big Five confidence factor was not significantly associated with the cognitive confidence factor.

Another potential reason for the lack of correlation between personality confidence and cognitive confidence is that the cognitive confidence judgments were made on tests of maximal performance, whereas personality confidence judgments were elicited from measures of typical performance. It may be that, if participants were given a self-report measure of abilities that also elicited confidence ratings for those abilities, then cognitive self-report confidence may share variance with personality confidence. This possibility was investigated in Study 3.

Calibration researchers may also argue that the Big Five confidence judgments in the current study were of a global nature, and were therefore more similar to post-test evaluative judgments in the cognitive domain. To make these post-test judgments, after completing a test, individuals indicate the percentage of items they believe they answered correctly. In a similar way, participants in the current study provided a global rating as to whether the block of trait adjectives described themselves, and then provided a confidence measure about the accuracy of that global rating. In the cognitive domain, factor analyses of post-test evaluative judgments and item-by-item confidence estimates have resulted in two separate factors. This separation has been explained in terms of PMM theory (Gigerenzer et al., 1991) which argues that these types of judgments are not subserved by the same cognitive processes. Perhaps item-by-item personality confidence ratings share more variance with item-by-item cognitive confidence ratings? This question was further examined in Study 3, wherein participants were asked to make both item-by-item and global confidence judgments within both the cognitive and personality domains.

One cognitive bias factor emerged from the structural analyses of the cognitive bias scores, as expected. This result is consistent with previous research (Pallier et al., 2002; Stankov, 1998, 1999a), which provided further evidence that if individuals were miscalibrated in one domain, then they are also miscalibrated across other domains.

To conclude, this study was exploratory in terms of adding (a) confidence ratings to Big Five personality judgments and (b) developing simple measures of accuracy. Results showed that Big Five confidence ratings were uni-dimensional across traits, with accuracy scores that were themselves uncorrelated, and that these Big Five confidence ratings were unrelated to cognitive confidence scores. The attempt to derive accuracy scores remains a challenge and was investigated further in Study 2. Nevertheless, results for the personality confidence and accuracy measures were meaningful, because all hypotheses except for two were supported. Of particular interest were the findings that both Gf and Gc accuracy were correlated with Intellect accuracy, suggesting that if individuals are accurate in estimating their Intellect scores, they are likely to be more intelligent. Study 3 re-examined the correlations between Big Five accuracy scores and Gf accuracy.

Study 2 focused solely upon Big Five confidence and bias and extended upon the personality confidence data obtained in Study 1 by taking item level-confidence ratings for each Big Five dimension. It also developed other measures of accuracy allowing Big Five bias scores to be examined. A question investigated in Studies 2 and 3 – whether Big Five confidence subscale scores define a separate factor from Big Five subscale ratings – is important, because the answers would provide further evidence that confidence is related to, but distinct from personality.

Chapter 4 - Study 2

4.1 Introduction

This study was designed to (a) obtain both item-by-item and block confidence judgments based on the Big Five taxonomy of personality structure, (b) examine the factorial structure of item-by-item and block personality confidence judgments, (c) develop different accuracy protocols so that a bias score for each Big Five dimension can be calculated, (d) investigate the factorial structure of item-level Big Five accuracy scores, (e) investigate whether individual differences in both gender and in age influence Big Five confidence, accuracy, and bias scores, (f) investigate the factorial structure of item-level Big Five bias scores, and (g) investigate the factorial structure of Big Five confidence subscale scores and Big Five subscale ratings.

4.2 Item-by-Item and Block Personality Confidence Judgments

Findings from Study 1 suggested that confidence judgments obtained from the Big Five block descriptors did not share much variance with item-by-item cognitive confidence judgments. If personality confidence had been measured at the item level, results may have been different, assuming that block and item-by-item personality confidence judgments have the same factorial structure as cognitive confidence judgments. That is, in the cognitive arena, item-by-item confidence judgments and post-test performance estimates (PTPE) scores do not correlate at the factorial level because, according to theorists like Gigerenzer and his colleagues (1991), cue validities people use to rate their confidence on item-by-item questions are different from those they use to make global post-test performance estimates (see section 2.3.2.2). Research has not investigated whether, at the factorial level, the structure of Big Five confidence follows the same pattern.

It is doubtful that the factorial structure of Big Five confidence splits unless, at the item level, individuals were making their confidence judgments with reference to episodic memory, but their block judgments encouraged retrieval from semantic memory. If this were the case then cue validities could possibly be different. For example, Jim is extraverted and went to a party on the weekend. At this party, he did not like some of the people and was rather reserved. On the following Monday he participated in an experiment that examined his level of extraversion at both the item level and the block level. He also provided confidence ratings for these judgments. If this party came to mind only when he answered item-by-item questions but not when he made block confidence judgments, then he used episodic knowledge at the item level and semantic knowledge for the block judgment. However, Jim's example does not fit well with the Five Factor theory of personality, which postulates that personality remains stable over adulthood (McCrae & Costa, 1999). If this is true for confidence judgments as well, it follows that Jim is more likely to make all his confidence judgments from semantic memory. This is logical with regard to the Big Five traits because these traits come from the lexicon of everyday language and therefore, will have high ecological validity for him. However, this statement assumes that semantic knowledge also incorporates self-knowledge of one's personality traits. It is interesting that the social-cognitive literature reveals theoretical and evidential information that demonstrates that semantic memory contains a specialized database that facilitates both the storage and retrieval of personality trait information and a brief overview of the relevant literature follows.

According to Klein's (2004) review, most of the psychological investigation regarding the relation between trait knowledge and memory has focused on episodic memory. This makes sense in view of its operational definition where episodic memory involves a psychological depiction of the self as the "agent or recipient of some action, or as the stimulus or experiencer of some state" (Kihlstrom, 1997, as cited in Klein, 2004, p. 1078). For example, I remember going to the movies last night, and I recall watching Andre Rieu on television on Christmas Eve, and feeling very moved by the music. Episodic memories about oneself obviously record both the instances when one exemplified a particular personality trait (e.g., being extraverted at a work party), and the times when one did not (e.g., having lunch with the mother-in-law) (Klein, Cosmides, Tooby, & Chance, 2001). The same can be said for our episodic memories of others (e.g., my daughter was extraverted at the party on Saturday night). Therefore, episodic memory provides one store of information that individuals can use to make personality judgments about themselves or others (Klein). Perhaps every time individuals make trait judgments they perform a serial search of each instance in which a particular trait was displayed or was not displayed. Such a laborious serial search, however, is impractical, because many decisions in daily life require quick and accurate judgments. Furthermore, it is more beneficial for individuals to have access to a store of precomputed trait summaries or generalizations across many episodes, which can be readily retrieved. For example, I am usually conscientious, or my friend Joe is generally agreeable. Such a store of precomputed trait summaries is more likely to reside in semantic memory. A review of the relevant cognitive literature follows.

Cognitive psychologists provide research data that supports the theory that trait self-knowledge and recall of specific occasions that involves those traits, stem from two different knowledge structures (e.g., Kihlstrom et al., 2003; Klein et al., 1989; Klein et al., 1996). These structures are semantic and episodic knowledge respectively. So far, however, the literature reviewed in this dissertation has not addressed how people make use of both types of knowledge structures when making trait judgments. Memory theorists propose two explanations (cf. Klein, 2004), known as the computational and the abstraction viewpoints, which are discussed below.

From the computational standpoint, individuals make trait judgments by retrieving trait-consistent behavioural exemplars from episodic memory, and then compare their similarity to the trait being judged (cf. Klein, 2004). For example, Kerry decides whether she is conscientious. First, she retrieves trait-relevant behaviours from memory, and then she computes online whether the retrieved exemplars match the trait of conscientiousness.

In contrast, the abstraction viewpoint posits that trait-relevant episodes are not retrieved when making trait judgments (Klein, 2004). Klein, Cosmides, Murray and Tooby (2004) posit that people have a cognitive database of trait generalizations/summaries (e.g., I am usually introverted, my father is generally conscientious, and my son is often stubborn) from which they make personality judgments about themselves and others. Abstraction theorists argue that trait information is abstracted from a delimited number of episodes, which leads to the development of trait summaries/generalizations (e.g., Buss & Craik, 1983; Klein, 2004; Klein & Loftus, 1993b; Klein, Loftus, Trafton, & Fuhrman, 1992). Individuals decide whether they exemplify a particular trait by retrieving trait summary knowledge from memory, as well as trait-inconsistent episodes (e.g., Klein, 2004; Klein, Loftus, Trafton et al., 1992). This makes sense because retrieval of trait-consistent episodes is redundant unless a trait summary does not exist (Klein, 2004). In the absence of a trait summary, episodic memory is the only store of information. For example, John is asked whether he is gregarious. If he does not have a summary store relevant to the trait of gregariousness, then he needs to perform a serial search of episodic memory before he answers the question. Predictions from both the computational and the abstraction viewpoints have been extensively examined within the memory literature (see Klein, 2004 for a review). A very brief summary follows.

As reviewed by Klein (2004), data from priming, encoding specificity, and encoding variability paradigms all support the abstraction viewpoint (e.g., Klein et al., 2001; Klein & Loftus, 1993a, 1993b; Klein et al., 1989; Klein, Loftus, & Plog, 1992; Klein, Loftus, Trafton et al., 1992). For example, in the priming experiments (e.g., Klein et al., 2001; Klein & Loftus, 1990, 1993a, 1993b; Klein et al., 1989; Klein, Loftus, Trafton et al., 1992), participants undertook numerous pairs of tasks, each of which included a certain trait adjective such as conscientious (Klein, 2004). Each adjective was preceded by a prime stimulus. That is, a descriptive judgment (e.g., does conscientious describe you?), a control judgment (e.g., think of the definition of conscientious) or a filler task. The dependent variable was the response latency for the recall task for each trait in question (e.g., try to recall a particular event during which you demonstrated conscientiousness). If the computational view is correct then participants presented with the describe prime should have answered the recall task more quickly than participants in the other two conditions, because trait-consistent episodes have already been activated. This was not the case and, in fact, individuals in all three conditions performed equally quickly when responding to the recall task. The overall conclusion was that individuals make trait self-judgments from a semantic store without activating episodic memory of those traits, which is consistent with the abstraction viewpoint (Klein, 2004).

Similarly, neuropsychological data from cognitively impaired individuals also provides evidence of the functional independence of semantic and episodic memory when individuals make trait self-judgments. Studies of amnesic patients allow psychologists to investigate the contributions of semantic and episodic memory to trait self-knowledge, because these patients often have impaired episodic memory, but intact semantic memory (cf. Klein, 2004 for a review). Five case studies provide compelling evidence for the existence of trait summaries within semantic knowledge structures. The cases of KC, who suffered a motorcycle accident (Tulving, 1993); WJ, who received a blow to her head (Klein et al., 1996); DB, who suffered anoxic encephalopathy following cardiac arrest (Klein, Rozendale, & Cosmides, 2002); RJ, who was autistic (Klein, Chan, & Loftus, 1999); and KR, who had Alzheimer's dementia (Klein, Cosmides, & Costabile, 2003), are particularly relevant. KC, WJ, and DB all suffered brain injuries and were unable to retrieve information from their respective episodic memory stores. Nevertheless, they were still able to provide reliable judgments about their own personalities when significant others were used as the criterion for accuracy. For example, DB's personality ratings correlated moderately (r = .64) with ratings of his personality obtained from his daughter. In the case of KR, who had advanced Alzheimer's dementia combined with severe retrieval deficits in many semantic domains (e.g., cannot name simple objects such as batteries or pencils and so forth), it was demonstrated that, despite her pronounced cognitive deficits, she demonstrated preserved knowledge of her own premorbid personality traits. KR's trait self-ratings correlated moderately with ratings of her premorbid personality obtained from both her daughter (r = .59) and her son-in-law (r = .79). To summarise, the evidence from both the cognitive and neuropsychological domains reviewed, suggests the presence of a substore within semantic memory, which specializes in the storage and retrieval of trait self-knowledge (Klein, 2004).

Perhaps confidence in personality judgments also resides within the same store of semantic memory. This is plausible if it is assumed that the traits under investigation have semantic summaries in place. The Big Five personality traits are expected to fit this criterion, because they come from the lexicon of daily life. If this is true, then all (i.e., item-by-item and block) Big Five personality confidence judgments can be expected to follow abstraction processes rather than computational processes. For example, it is onerous to have to recall specific episodes in which one is conscientious before being able to endorse one’s confidence level for items that measure that construct. It is more logical for people to have summary representations within semantic memory that allow them to make confidence judgments about well-known traits. For this reason, item-by-item and block Big Five personality confidence ratings are expected to load onto one factor because, both types of judgments reside within the same store of semantic memory. Previous research has not examined the factorial structure of Big Five confidence.

4.3 Accuracy Scores for the Big Five Block Judgments

In Study 1, accuracy scores obtained from the block method were largely uncorrelated. Therefore, people who were consistent (accurate) in their scores on one trait were not necessarily accurate in their scores on other traits. In consequence, an accuracy score could not be formed across the Big Five dimensions, and an overall personality bias score was not calculated. The intercorrelations between the accuracy scores were re-examined in Study 2. However, for the reasons discussed at the end of Study 1, it is problematic to assume that personality accuracy can be summed across all the Big Five measures. Any decision to simply add the Big Five accuracy scores requires making the assumption that each Big Five trait has the same ecological validity for every individual in the sample. Therefore, each person should be able to provide consistent responses across the IPIP and its respective block measure. Another assumption that random error does not affect consistency of responding may also not be realistic. It was assumed that properties of the accuracy measures themselves did not lead to the low intercorrelations among the Big Five accuracy scores, but, that the inconsistencies were, for the reasons stated, within the individuals themselves. The associations among the Big Five block accuracy scores, were, therefore, expected to be low. Consequently, bias scores cannot be calculated using the block method. Thus, it was important for accuracy to be calculated in a different way, in order to compute a bias score for each Big Five dimension. Study 2 aimed to develop accuracy methods so that mean accuracy scores were calculated for each Big Five dimension. A detailed discussion of these accuracy scores can be found in the method section of this chapter.

4.4 Bias and Accuracy Scores for the Item-By-Item Big Five Judgments

The calibration paradigm uses bias scores as a measure of how well calibrated individuals are when self-monitoring their performance on various cognitive tasks. Individuals are scored as accurate or inaccurate based on the objective criterion of whether their responses are correct. It is acknowledged that for personality judgements, assessing intra-phenomenological accuracy is difficult, because psychologists do not have access to the kind of objective criteria that exists in the cognitive arena. Nevertheless, psychologists do have self-report measures of personality that have repeatedly demonstrated more than adequate psychometric properties across many previous research studies (Anastasi & Urbina, 1997). Indeed, it has been argued that the

Big Five taxonomy is one of the most parsimonious and efficacious ways of conceptualising the architecture of personality (e.g., Goldberg, 1999; Pallier et al., 2002; Saucier & Goldberg, 1996; Wiggins, 1996) and thus this taxonomy, as measured by the IPIP, has been used as the objective criterion for determining accuracy in Study 2. The rationale follows.

In Study 2, individuals were deemed accurate if they provided consistent endorsements across the two parallel forms of the IPIP that were developed for each personality dimension (see method section). It is argued that without any objective criterion, other than self-report measures within the personality domain, that consistent responding across two measures of the same construct should, at the macro level, imply some level of accuracy. This assumption may appear contentious because reliability is usually considered a property of the test. Nevertheless, the notion that consistent responding implies some level of accuracy can be found within the clinical psychology literature (e.g., Trauma Symptom Inventory (TSI; Briere, 1995); Millon Clinical Multiaxial Inventory-III (MCMI-III; Millon, Davis, & Millon, 1997) ; OMNI-IV Personality Inventory (OMNI; Loranger, 2001), Personality Assessment Inventory (PAI; Morey, 1991), and the Minnesota Multiphasic Personality Inventory (MMPI-2; Hathaway & McKinley, 1989). The PAI and the MMPI-2 are the most widely used personality assessment inventories in clinical settings and normative data is provided for both clinical and non-clinical samples. Both of these inventories are also used outside traditional mental health settings in forensic, medical, neuropsychological and employment contexts. MMPI-2 interpretation rests on the supposition that the individual has, for the most part, endorsed items that are semantically similar in a consistent fashion (Nichols, 2001). To this end, the response consistency scale (i.e., Variable Response Inconsistency Scale [VRIN]) is examined prior to the interpretation of the standard clinical scales (Nichols & Greene, 1997). VRIN is the variable response inconsistency scale which attempts to highlight a random response pattern or an inconsistent pattern of responses. This scale consists of 67 pairs of items that were chosen because of their statistical associations and semantic correspondence (Nichols, 2001). Anomalies in VRIN, combined with elevations on True Response Consistency Scale (TRIN), and the F-scales, can render an individual's protocol invalid. Moreover, Nichols, Greene, and Schmolck (1989) cited Greene (1988), who asserted that "…item endorsements must be consistent if they are to be accurate" (p. 249). However, response consistency does not guarantee accuracy (Nichols et al., 1989). Guaranteeing absolute accuracy in intra-phenomenological judgments is a moot point. In a similar way, the PAI has the Inconsistency (ICN) scale was also designed to determine if test takers endorsed items with similar content in a consistent manner (Morey, 2003). T-scores of 73 or higher suggest that respondents answered items in a completely random fashion rendering the protocol invalid. Morey suggests that whilst there are a number of possible causes for high scores on the ICN scale (e.g., reading problems, uncertainty of test instructions, inattention) interpretation of clinical scales is abandoned at this stage of the analysis. As consistent responding has been shown to be the fundamental basis upon which clinical interpretation rests, for the purposes of the current research, perceived accuracy was assumed if the individual provided consistent responses across both IPIP Form A and IPIP Form B, for each of the Big Five personality dimensions. Three accuracy protocols were developed for Study 2. The standard error of difference scores (Anastasi & Urbina, 1997) was used to determine consistency of responding for two of the methods. Full descriptions of how accuracy was determined for all three methods and how item-level bias scores were calculated is left to the method section. It was expected that the associations among item-level Big Five accuracy scores would also be low for the same reasons as stated in section 4.3 therefore the calculation of separate bias scores for each Big Five dimension is justified.

As the Big Five traits are likely to have high ecological validity, it was expected that individuals would be well calibrated. It is also important to investigate whether Big Five bias scores are correlated. The reason for this is that, within the cognitive domain, bias scores from Gv, Gf, and Gc tasks tend to load onto one factor, suggesting that bias scores across various cognitive tasks use the same cognitive processes. What remains unanswered is whether bias within the personality domain is driven by the same cognitive processes. The factorial structure of Big Five item-level bias scores was examined in Study 2, although no specific hypotheses were made at this exploratory stage.

4.5 Gender and Age Differences in Personality Confidence, Accuracy, and Bias

Findings from Study 1 indicated that males were just as confident and accurate as females were when making confidence judgments about the Big Five personality dimensions. These findings were consistent with the gender similarities hypothesis (Hyde, 2005). Gender differences in Big Five confidence and accuracy were re-examined in Study 2. On the basis of Hyde's hypothesis, males and females were not expected to differ in terms of Big Five confidence, accuracy, or bias.

Results from Study 1 showed that age was essentially uncorrelated with either Big Five confidence or accuracy, with two qualifications. That is, age was positively correlated with both Intellect confidence ($r_{(125)} = .19$, $p < .05$) and Intellect accuracy ($r_{(125)} = -.21$, $p < .05$), however, the effect sizes were small. Correlations between age and Big Five accuracy, Big Five confidence, and Big Five bias were examined in Study

2. Based on the five-factor theory of personality (McCrae & Costa, 1999), combined with the assumption that Big Five judgments are made on a daily basis, and therefore have high ecological validity, it was expected that age would neither be associated with Big Five confidence, nor be associated with Big Five accuracy, or Big Five bias scores.

4.6 Factorial Structure of Big Five Confidence and Big Five Subscale Scores

Research has shown that the confidence trait is exists the borderline between cognitive abilities and personality (Baker, 2001; Kleitman & Stankov, 2001, 2007; Pallier et al., 2002; Stankov, 1998, 1999a, 2000a; Stankov & Crawford, 1996a, 1997; Stankov & Lee, 2008). Furthermore, Stankov and Lee (2008, p. 974) asserted "… that confidence is indeed a psychological trait that is related to, but distinct from both personality and ability traits. Within the structure of all individual differences dimensions, confidence should be located between these two domains" (p. 974). In the aforementioned studies these assertions were based on low correlations between *cognitive* confidence ratings and Big Five personality traits. The claim that confidence is a distinct trait that is related to personality can also be independently investigated by factor analysing all the IPIP and IPIP Form B subscales (i.e., adjective ratings and confidence scores). It was expected that six factors would emerge from this structural analysis. That is, Big Five confidence and the other five factors would correspond to each of the Big Five dimensions.

4.7 Hypotheses

The hypotheses are summarised below:

1. It was hypothesised that one confidence factor would emerge from factor analysing the item-by-item and block confidence judgments obtained from the Big Five

personality dimensions. This hypothesis was developed with reference to abstraction theory within the memory domain (e.g., Buss & Craik, 1983; Klein, 2004; Klein & Loftus, 1993b; Klein, Loftus, Trafton et al., 1992).

2. In view of the findings from Study1, and the gender similarities hypothesis (Hyde, 2005), it was hypothesised that males and females would not differ in terms of Big Five confidence at either the block description or item level.
3. Based on the five-factor theory of personality (McCrae & Costa, 1999) combined with the assumption that Big Five judgments are made on a daily basis, and therefore have high ecological validity it was hypothesised that age would not be associated with Big Five confidence, Big Five accuracy, or Big Five bias judgments obtained from either the block description or item level measures.
4. In view of the findings from Study 1 in which the correlations among the Big Five block accuracy scores were low, it was hypothesised that the associations among the Big Five block accuracy scores, and the item-level accuracy scores will also be low.
5. Good calibration was also expected for the Big Five judgments based on Gigerenzer’s (1991) theory which argues that individuals are well calibrated to their natural environments.
6. In view of the findings from Study 1, and the gender similarities hypothesis (Hyde, 2005), it was hypothesised that males and females would not differ in terms of Big Five accuracy judgments obtained at either the block or item level.
7. Based on previous research, and the claim that confidence is a distinct trait that is related to personality, it was hypothesised that six factors would emerge from the structural analyses of the Big Five confidence and Big Five subscale scores from the IPIP and IPIP Form B.

4.8 Participants

A total of 411 individuals participated in this study. The sample comprised males (*n* = 87) and females (*n* = 322), ranging in age from 18 to 63 years (*M* = 28.65 years, *SD* = 9.25 years). Two people did not indicate their gender. The mean age of the males was 29.74 years (*SD* = 10.04 years) and the mean age for females was 28.03 (*SD* = 9.01 years).

The highest educational level varied from completion of grade 9, 10, 11, or 12 (*n* = 86) to completion of tertiary studies (*n* = 39) and eight participants did not respond to this question. Two hundred and forty nine participants were enrolled in undergraduate Psychology courses at the University of Southern Queensland and received course credit for their participation. Snowball sampling techniques (see method section Study 1) were used to obtain the other 162 participants. In return for taking the time to complete the study, entry in a raffle for cash prizes was offered to these participants.

4.9 Materials

All participants completed a computerised battery of measures that included demographic information and self-report measures of personality. Demographic questions consisted of items regarding, gender, age, and highest level of education.

The test battery comprised IPIP, IPIP Form B (see below), and the BFBD. IPIP Form A was the measure used in the data analyses for the item-level accuracy methods and, was formed from participants' responses to the IPIP. Prior to developing IPIP Form A, the factorial structure of the IPIP was examined (see results) and reliability estimates were consulted for each of the IPIP subscales. The factorial structure of IPIP Form A was also examined as this scale was used as the criterion for the accuracy scores and the factor analysis yielded a clear five factor solution.

The Original International Personality Item Pool Five-Factor Personality Scale (IPIP) - (Goldberg, 1999). See Study 1. The IPIP was the criterion for accuracy for the BFBD described next.

Big Five Block Descriptions of Personality (BFBD) Based on the IPIP Trait Adjectives - (Goldberg, 1999). The same block descriptions used in Study 1 were used in Study 2, and the accuracy scores for these block descriptions were derived using the same steps outlined in the method section of Study 1.

Shortened version of the IPIP Scale based on the Trait Adjectives (IPIP Form A) - (Goldberg, 1999). A shortened version of the IPIP was used as the criterion for accuracy for the item-level judgments. The original IPIP contains 10 items for each of the Big Five personality subscales. The IPIP form B (see below) on the other hand, does not have 10 items for each subscale for the reasons highlighted in the next section. However, the protocols for determining accuracy which are described later necessitate the calculation of differences scores for each personality dimension, and this calculation required that both the IPIP Form A and the IPIP Form B, have the same number of items for each personality subscale. Consequently, items were randomly deleted from the Extraversion (2 items), Conscientiousness (1 items), and Emotional Stability subscales (1 item) of the original IPIP, resulting in a parallel form of the IPIP subscales. Convergent validity co-efficients between IPIP Form A subscales and its respective IPIP subscale were .98 or higher.

Item-by-Item Big Five Self-Rated Personality Descriptions (IPIP Form B) based on the Trait Adjectives by Goldberg (1999). The item-by-item judgments used the same 46 trait adjectives, instructions, and rating scales, as the BFBD described in the last chapter. Thus, 8 trait adjectives assessed the Extraversion dimension, 10 assessed both the Agreeableness and the Intellect dimensions, and 9 assessed both the

Emotional Stability, and the Conscientiousness dimensions. For these judgments, however, participants rated each trait adjective (e.g., "I don't mind being the centre of attention") separately, and also provided an item-by-item confidence rating. As before, participants rated the extent to which each trait adjective generally reflected their personality on an 11-point scale with end points of -5 (*Not like me*) to +5 (*Like me*). For each trait adjective, participants also provided their confidence rating after reading the same confidence instructions as described for the BFBD in Study 1. The five item-by-item personality descriptions are presented in Appendix L.

Three Accuracy Methods

In order to continue the parallel with cognitive testing, three accuracy methods were developed. This was for the purpose of constructing a necessary situation wherein a mean accuracy score was able to be calculated for each Big Five dimension. Prior to describing these methods recoding of the parallel forms is discussed.

As the items across both IPIP Form A and IPIP Form B were scored on different rating scales, the first step involved recoding the IPIP Form B scores back to the same scale as the IPIP Form A. Thus, IPIP Form B scores of -5 and -4 were recoded as 1; scores of -3 and -2 were recoded as 2; scores of -1, 0, and 1 were recoded as 3; scores of 2 and 3 were recoded as 4; and scores of 4 and 5 were coded as 5. It is acknowledged that post hoc transformations to equate the two test forms do not overcome the fact that the original response formats were different. The Big Five subscale scores on IPIP Form A and IPIP Form B were calculated. On each measure the scores for the Big Five dimensions ranged from 8 to 40 for the Extraversion dimension, from 9 to 45 for both Emotional stability and Conscientiousness, and from 10 to 50 for both the Agreeableness and Intellect dimensions.

Method 1

For each Big Five dimension the standard error of measurement was calculated for raw subscale scores on both IPIP Form A and the parallel scales for IPIP Form B. The standard error of difference score was then computed, for each dimension, to determine accuracy.

Anastasi and Urbina (1997) assert that individual scores can be interpreted using the standard error of measurement. However, the current study aimed to find a way of interpreting whether both subscale scores were answered in a consistent manner. Therefore the formula for the standard error of difference scores was used. This score provides a measure of “how large a score difference could be obtained by chance” (Anastasi & Urbina, 1997, p. 110). At the 95% level of confidence, the standard error of the difference between two scores needs to be multiplied by 1.96 whereas at the 99% level of confidence, the difference score needs to be multiplied by 2.58. The current study used the 95% confidence interval as this is more conservative and results in a smaller band of scores around each individual’s difference score.

The following formula was taken from Anastasi and Urbina (1997, p. 111)

$$SE_{diff.} = \sqrt{(SEM_1)^2 + (SEM_2)^2}$$

in which SE_{diff} is the standard error of the difference between the two scores, and SEM_1 and SEM_2 are standard errors of measurement of the separate scores. By substituting $SD\sqrt{1 - r_{11}}$ for SEM_1 and $SD\sqrt{1 - r_{22}}$ for SEM_2 we may rewrite the formula directly in terms of reliability coefficients, as follows:

$$SE_{diff.} = SD\sqrt{2 - r_{11} - r_{22}}$$

An example for the Agreeableness dimension follows.

$$SE_{diff.} = \sqrt{(SEM_{scoreagree1})^2 + (SEM_{scoreagree2})^2} \text{ x } 1.96$$

$SEM_{scoreagree1}$ = SEM score on the IPIP Form A for Agreeableness, $SEM_{scoreagree2}$ = SEM score on the IPIP Form B for Agreeableness. The SEM for any test = $SD_t \sqrt{1 - r_{tt}}$ where SD_t is the standard deviation of the test scores and r_{tt} is the reliability co-efficient, both computed on the same group.

A difference score was calculated by subtracting each individual's IPIP Form B Agreeableness subscale score from his or her Form A Agreeableness subscale score. If the absolute value of the difference between the parallel forms fell within the 95% confidence interval then the score was coded as accurate (1); otherwise it was coded 0. Accuracy scores for the other four dimensions were determined in the same manner.

Method 2

This method involved converting raw subscale scores from IPIP Form A and IPIP Form B, to z scores. Difference scores were then computed between the parallel forms for each dimension. An arbitrary choice of z = 1 was set for determining accuracy. Thus, if the absolute value of the difference between the parallel forms for each dimension was z = 1 then the individual was coded as accurate (1); all else was coded 0.

Method 3

With this method and as a precautionary measure, individual items on IPIP Form A and IPIP Form B were transformed to *z* scores. This was done because item response formats and instructions were different for the parallel forms. Next, the Big Five subscale scores on the IPIP Form A and IPIP Form B were re-calculated using the z-transformed items. For every participant a difference score was then computed for each Big Five dimension by subtracting the zIPIP Form B subscale score from its corresponding subscale score of zIPIP Form A. For each Big Five dimension the standard error measurement was calculated for the z transformed subscale scale scores on both IPIP Form A and IPIP Form B. Then to determine accuracy, the standard error of difference scores for each dimension was computed in the same manner as in method one (see above). If the absolute value of the difference between the parallel forms fell within the 95% confidence interval, then the score was coded as accurate (1); otherwise it was coded as 0.

Big Five Bias Scores for each dimension

The accuracy scores just described were factor analysed to investigate whether a five-factor solution emerged, with each factor comprising the three relevant scores for each dimension (see results section). For each person a mean accuracy score could then be calculated for each dimension, using the accuracy scores obtained from the three methods described above. Then for each participant a mean confidence rating score for each Big Five dimension was computed. This was done by using the confidence ratings taken from IPIP Form B. A bias score could then be calculated for each person for each dimension in the same manner as the cognitive bias scores. That is, mean accuracy expressed as a percentage was taken away from the mean confidence score.

4.10 Procedure

Participants were recruited via the USQ Psychology Experiment Sign Up Database (PESUD) and were able to login from any internet connected computer. All data were submitted electronically and collected by the Psychology technical team. Informed consent was obtained at the beginning of the testing session. Participants read the electronic consent form, and typed in a unique identifying number indicating that they understood that their data would be used for research purposes (see Appendix M). Respondents were informed that they could withdraw from the study at any time without penalty, and were assured of confidentiality and anonymity. To eliminate missing data, participants were required to provide an answer to each question to progress through the test battery. The battery took approximately 1 hour to complete.

Due to programming restrictions, the order in which the measures were presented, was not randomized. Therefore, the test battery was administered in the following order:

1. IPIP
2. Item-by-item Agreeableness judgments (IPIP Form B subscale)
3. Block Agreeableness judgment (BFBD-A)
4. Item-by-item Conscientiousness judgments (IPIP Form B subscale)
5. Block Conscientiousness judgment (BFBD-C)
6. Item-by-item Extraversion judgments (IPIP Form B subscale)
7. Block Extraversion judgment (BFBD-E)
8. Item-by-item Emotional Stability judgments (IPIP Form B subscale)
9. Block Emotional Stability judgment (BFBD-EM)
10. Item-by-item Intellect judgments (IPIP Form B subscale)
11. Block Intellect judgment (BFBD-I)

4.10.1 Scoring

For the gender variable, 1 = males and 2 = females. For the Big Five accuracy variables 0 = inaccurate and 1 = accurate. Big Five bias scores were calculated for the item-by-item judgments and were described above.

4.11 Results

Prior to statistical analyses, all variables were examined through various subprograms from the Statistical Package for Social Scientists (SPSS, Version 16) for accuracy of data input, missing data, and fit between their distributions and the assumptions of multivariate analysis (Tabachnick & Fidell, 1996). For several variables one or more univariate outliers were detected. Visual inspection of the data from these cases revealed that cases 271, 239, 30, 82, 394, and 268 were both univariate and multivariate outliers and were therefore deleted from the data set leaving 405 participants in the study. The other univariate outliers were retained as legitimate variation in line with recommendations made by Tabachnick and Fidell (1996). No other problems were detected.

4.11.1 Factorial Structure of the IPIP

The factorial structure of the IPIP was examined using Principal Components Analysis with Promax rotation and Kaiser normalisation. A solution employing root one criterion produced 12 factors. Cattell's Scree Plot, however, provided support for interpreting a five-factor solution. The pattern matrix, percent of variance accounted for, eigenvalues, and factor correlation matrix are presented in Table 4.1. The five-factor solution accounted for 43.40% of the total variance. As anticipated, for each dimension, all the factors comprised loadings from the expected IPIP items. The five factors were labelled IPIP Emotional Stability, IPIP Extraversion, IPIP Agreeableness, IPIP Intellect,

Table 4.1

Pattern Matrix of IPIP Items using Principal Components Analysis with Promax Rotation and Kaiser Normalisation (N = 405).

Variable	F_1 [a]	F_2	F_3	F_4	F_5	h^2
ipip34	0.81	0.07	-0.11	0.05	0.06	.63
ipip04	0.80	0.04	0.05	-0.06	0.10	.60
ipip39	0.79	0.03	-0.06	0.04	-0.03	.63
ipip44	0.78	0.10	-0.07	0.01	0.05	.57
ipip14	0.76	-0.03	0.18	-0.05	0.10	.58
ipip29	0.74	-0.02	0.15	-0.11	0.04	.57
ipip49	0.72	-0.14	-0.06	0.03	-0.02	.61
ipip24	0.60	0.03	0.00	0.01	-0.10	.38
ipip09	-0.51	0.13	0.13	0.02	-0.10	.34
ipip19	-0.51	0.18	-0.07	-0.11	0.15	.38
ipip18	0.44	0.06	0.10	-0.09	-0.40	.41
ipip38	0.33	0.16	-0.32	-0.02	-0.20	.30
ipip21	0.16	0.76	0.11	-0.04	0.13	.58
ipip01	0.06	0.75	-0.13	-0.04	-0.03	.51
ipip41	0.09	0.73	-0.13	0.06	-0.02	.51
ipip31	-0.06	0.71	0.11	-0.04	-0.02	.55
ipip16	0.14	-0.67	0.19	-0.08	-0.02	.54
ipip06	-0.11	-0.62	-0.20	-0.05	0.03	.46
ipip46	0.15	-0.60	-0.02	0.06	0.06	.42
ipip36	0.11	-0.57	0.22	-0.04	0.08	.39
ipip11	-0.18	0.54	0.22	-0.03	0.10	.50
ipip26	0.06	-0.43	-0.24	-0.25	-0.02	.47
ipip47	0.05	0.35	0.34	0.00	0.10	.30
ipip42	0.08	-0.04	0.66	-0.03	0.02	.42
ipip27	0.13	-0.02	0.66	-0.14	-0.03	.40
ipip22	0.05	-0.07	-0.62	0.01	0.11	.39
ipip17	0.11	-0.05	0.62	0.04	0.00	.39
ipip32	0.08	-0.22	-0.61	-0.03	0.07	.49
ipip07	0.02	0.13	0.60	0.00	-0.01	.40
ipip37	-0.08	-0.06	0.55	0.01	0.06	.33
ipip12	0.31	0.18	-0.47	0.06	-0.07	.33
ipip02	0.06	0.05	-0.40	0.00	0.04	.16
ipip45	0.17	-0.15	0.29	0.21	-0.07	.18
ipip50	-0.01	0.11	-0.05	0.70	0.02	.54
ipip30	0.08	0.05	-0.10	-0.68	0.13	.48
ipip10	0.19	0.08	0.01	-0.66	0.05	.44
ipip05	0.03	-0.02	-0.07	0.66	0.05	.41
ipip25	0.01	0.14	-0.11	0.63	0.15	.48
ipip15	0.23	0.04	0.06	0.58	-0.04	.39
ipip40	0.15	0.02	-0.16	0.57	0.00	.31
ipip20	0.05	0.00	-0.11	-0.55	0.13	.35
ipip35	-0.19	-0.03	-0.01	0.49	0.08	.30
ipip23	0.00	0.11	-0.05	-0.20	0.70	.50
ipip43	0.12	0.07	0.00	-0.04	0.70	.47
ipip03	-0.01	0.11	-0.11	0.05	0.66	.44
ipip33	0.17	-0.10	0.06	0.08	0.64	.44
ipip08	0.09	0.03	0.03	0.14	-0.60	.38
ipip28	0.13	0.01	0.04	0.08	-0.59	.38
ipip48	0.04	-0.10	0.06	0.23	0.48	.32
ipip13	0.08	-0.16	0.14	0.27	0.48	.38
Eigenvalue	8.12	4.65	3.78	2.64	2.51	
% of Variance	16.24	9.31	7.56	5.28	5.02	

Variable	F_1	F_2	F_3	F_4	F_5
F1	1.00				
F2	-0.30	1.00			
F3	-0.09	0.18	1.00		
F4	-0.09	0.29	0.27	1.00	
F5	-0.21	0.06	0.24	0.12	1.00

Note. [a] Factor Labels, F_1 = IPIP Emotional Stability ; F_2 = IPIP Extraversion; F_3 = IPIP Agreeableness; F_4 = IPIP Intellect ; F_5 = IPIP Conscientiousness

and IPIP Conscientiousness respectively. The correlations among the factors were concordant with previous research (Saucier, 2002; Saucier & Goldberg, 2002).

Table 4.2 presents the correlations among the IPIP subscales. These correlations were similar to those reported in the literature (Saucier, 2002; Saucier & Goldberg, 2002).

Table 4.2
Correlations Among the IPIP Subscales

Variable	1	2	3	4	5
IINTELL	1.00				
ICON	0.11*	1.00			
IEXTRA	0.30**	0.08	1.00		
IAGREE	0.26**	0.26**	0.26**	1.00	
IEMOT	0.06	0.25**	0.34**	0.14*	1.00

* $p < .05$ ** $p < .01$

4.11.2 Factorial Structure of the IPIP Form B

The factorial structure of the IPIP Form B items were also examined using Principal Components Analysis with Promax rotation and Kaiser normalisation. A solution employing root one criterion produced 7 factors. Cattell's Scree Plot, however, was indeterminate after five factors thus supporting the interpretation of a five factor solution. Table 4.3 presents the pattern matrix, percent of variance accounted for, eigenvalues, communalities, and the factor correlation matrix. The five-factor solution accounted for 55.04% of the total variance. For each Big Five dimension, all the factors consisted of loadings from the expected IPIP Form B items. The five factors were labelled IPIP Emotional Stability Form B, IPIP Extraversion Form B, IPIP Intellect Form B, IPIP Conscientiousness Form B, and IPIP Agreeableness Form B respectively. The bottom of Table 4.3 presents the component correlation matrix. As shown, low to moderate correlations emerged among the IPIP Form B subscales. It is noted that these correlations among the subscales were higher than the correlations among the IPIP subscales.

Table 4.3

Pattern Matrix of IPIP Form B Items using Principal Components Analysis with Promax Rotation and Kaiser Normalisation (N = 405).

Variable	F_1[a]	F_2	F_3	F_4	F_5	h^2
PDEM03	0.84	-0.06	-0.07	0.03	0.10	.70
PDEM04	0.80	-0.12	-0.08	-0.04	0.21	.65
PDEM05	0.79	-0.08	0.14	0.00	-0.15	.61
PDEM01	0.76	0.06	0.00	-0.09	0.04	.61
PDEM07	0.76	0.10	-0.08	-0.03	0.03	.62
PDEM02	0.75	-0.05	0.09	-0.15	0.06	.57
PDEM08	0.72	0.08	-0.05	0.12	-0.01	.61
PDEM09	0.70	0.05	0.11	-0.02	-0.06	.56
PDEM06	0.68	0.05	0.13	0.14	-0.27	.55
PDE07	0.02	0.91	-0.07	-0.03	-0.13	.71
PDE01	-0.04	0.86	-0.02	-0.02	-0.18	.60
PDE04	-0.05	0.83	0.08	0.03	-0.03	.70
PDE08	-0.05	0.81	-0.03	-0.08	0.09	.65
PDE02	0.04	0.79	-0.10	-0.04	0.16	.69
PDE06	0.04	0.78	0.05	0.00	0.07	.72
PDE03	-0.01	0.68	0.14	0.12	-0.15	.54
PDE05	0.19	0.61	-0.06	0.03	0.23	.68
PDI08	-0.01	-0.06	0.81	-0.11	0.04	.59
PDI07	0.08	-0.11	0.80	-0.07	-0.05	.57
PDI05	0.05	-0.08	0.72	0.02	-0.01	.50
PDI010	-0.19	-0.07	0.71	-0.01	0.16	.48
PDI03	0.08	-0.03	0.67	0.07	-0.04	.50
PDI02	0.13	0.04	0.65	-0.02	0.01	.52
PDI04	0.06	0.07	0.64	0.07	0.02	.54
PDI09	-0.09	0.13	0.63	-0.07	0.04	.44
PDI06	0.18	0.07	0.58	0.10	-0.03	.54
PDI01	-0.02	0.29	0.53	0.00	0.02	.50
PDC01	-0.14	0.00	-0.03	0.84	0.01	.66
PDC03	-0.10	0.09	-0.05	0.82	0.04	.68
PDC04	-0.07	-0.04	-0.04	0.82	0.01	.63
PDC09	0.05	0.05	-0.13	0.77	0.02	.61
PDC07	0.04	0.05	0.05	0.76	-0.04	.62
PDC02	-0.09	-0.07	0.08	0.75	-0.01	.54
PDC06	0.15	0.08	-0.10	0.65	0.03	.51
PDC08	-0.01	-0.17	0.26	0.65	0.06	.54
PDC05	0.18	-0.06	-0.03	0.56	0.03	.37
PDA08	-0.02	-0.08	-0.04	0.02	0.81	.60
PDA010	-0.14	-0.05	0.11	-0.08	0.78	.55
PDA07	0.13	-0.08	-0.11	0.10	0.67	.49
PDA01	-0.21	0.03	0.11	0.02	0.64	.43
PDA05	0.24	0.00	-0.07	0.00	0.62	.51
PDA09	-0.06	0.13	0.12	-0.05	0.56	.40
PDA06	0.02	-0.07	0.05	0.14	0.50	.33
PDA04	0.24	0.06	-0.01	-0.02	0.48	.39
PDA03	0.11	-0.10	-0.03	0.03	0.42	.19
PDA02	-0.12	0.21	0.16	0.08	0.40	.34
Eigenvalue	12.35	4.23	3.38	2.92	2.43	
% of Variance	26.85	9.2	7.35	6.36	5.29	
Factor Correlation Matrix						
	F1	F2	F3	F4	F5	
F1	1.00					
F2	0.43	1.00				
F3	0.34	0.43	1.00			
F4	0.27	0.27	0.29	1.00		
F5	0.33	0.41	0.30	0.37	1.00	

Note. [a] Factor Labels, F_1 = IPIP Emotional Stability Form B ; F_2 = IPIP Extraversion Form B; F_3 = IPIP Intellect Form B; F_4 = IPIP Conscientiousness Form B; F_5 = IPIP Agreeableness Form B.

4.11.3 Descriptive Statistics and Factorial Structure of Item-level Accuracy Scores

Three accuracy methods were developed for the purpose of calculating a mean accuracy score for each Big Five dimension. The descriptive statistics for these scores are presented in Table 4.4, with the most salient feature being that the means for methods 2 and 3 are quite similar for each of the dimensions except for the Agreeableness dimension. The correlations among these accuracy scores are presented in Appendix N. All of the accuracy scores for each Big Five dimension were correlated because the three methodologies were similar. It is worth noting for each dimension, the highest correlations were between methods 2 and 3.

Table 4.4
Descriptive Statistics for Accuracy Scores Derived from Methods 1, 2, and 3 (N=405)

Variable	*M*	*SD*
Intellect Accuracy Method 1	68.40	46.55
Intellect Accuracy Method 2	81.73	38.69
Intellect Accuracy Method 3	85.68	35.07
Conscientiousness Accuracy Method 1	78.77	40.95
Conscientiousness Accuracy Method 2	83.95	36.75
Conscientiousness Accuracy Method 3	88.15	32.36
Extraversion Accuracy Method 1	73.09	44.41
Extraversion Accuracy Method 2	89.63	30.53
Extraversion Accuracy Method 3	87.90	32.65
Agreeableness Accuracy Method 1	82.72	37.86
Agreeableness Accuracy Method 2	72.35	44.78
Agreeableness Accuracy Method 3	85.19	35.57
Emotional Stability Accuracy Method 1	72.35	44.78
Emotional Stability Accuracy Method 2	85.93	34.82
Emotional Stability Accuracy Method 3	81.73	38.69

Prior to calculating mean accuracy scores for each Big Five dimension, the accuracy scores in Table 4.4 were factor analysed which showed five clear factors with the three accuracy scores for each dimension loading on its expected factor. The results from this structural analysis suggested that accuracy was specific to each Big Five domain. So, calculating mean accuracy scores for each dimension was justified. This factor analysis is presented in Appendix N. Please note for the remainder of this results section, item-level accuracy scores refer to the mean accuracy scores for each Big Five dimension.

4.11.4 Reliability Analysis

Internal consistency reliability estimates (i.e., Cronbach's coefficient alphas) were calculated for all variables where applicable. Table 4.5 presents these alpha coefficients, along with descriptive statistics for all dependent variables. All alpha coefficients indicated good internal consistency. The alpha coefficients reported for the original IPIP subscales accorded with those reported in the research literature where the IPIP was administered on a face-to-face basis (e.g., Goldberg, 1999). It is worth noting that the table shows that the subscales of IPIP Form B along with Big Five accuracy scores all demonstrated more than acceptable levels of internal consistency. Also worthy of mention is that the item-by-item Big Five confidence ratings from IPIP Form B demonstrated high internal consistency, with alpha coefficients ranging from .91 to .94.

4.11.4.1 Descriptive Statistics

Descriptive statistics for all dependent variables are presented in Table 4.5. Mean ratings for each of the Big Five dimensions on the original IPIP that was administered via the internet, ranged from 30.50 for the Emotional Stability dimension to 40.86 for the Agreeableness dimension. These mean ratings closely matched those endorsed by Study 1 participants, which used face-to-face test administration. Study 2 respondents also reported block personality ratings similar to those reported in Study 1.

Table 4.5 also shows that across all of the Big Five personality dimensions, individuals were very confident for both the block and item-by-item personality judgments. The fact that the mean confidence rating scores for the Big Five block judgements closely followed those reported by a different group of respondents in Study 1, is worthy of note here.

Table 4.5
Descriptive Statistics for all Dependent Variables (N = 405)

Dependent Variables	*M*	*SD*	# items	α
IPIP Original Subscales				
ICON	35.54	5.76	10	.78
IEMOT	30.50	7.76	10	.90
IINTELL	36.77	5.31	10	.78
IEXTRA	32.19	7.00	10	.86
IAGREE	40.86	4.83	10	.76
IPIP Form A Subscales				
ICONA	32.18	5.04	9	.75
IEMOTA	28.08	7.05	9	.89
IINTELLA	36.77	5.31	10	.78
IEXTRAA	25.07	5.81	8	.84
IAGREEA	40.86	4.83	10	.76
Personality Confidence Item-by-Item Ratings For Each Personality Dimension				
CONCIC	81.88	11.35	9	.93
EMOTIC	80.67	12.06	9	.94
INTELLIC	80.66	11.75	10	.93
EXTRAIC	80.71	11.99	8	.92
AGREEIC	82.52	10.62	10	.91
IPIP Form B Item-by-Item Big Five Adjective Ratings-Recoded [1]				
CONCIR	34.36	6.92	9	.90
EMOTIR	31.13	7.99	9	.91
INTELLIR	40.12	6.56	10	.88
EXTRAIR	29.17	7.00	8	.92
AGREEIR	43.04	4.73	10	.82
Big Five Block Confidence Ratings (BFBDC)				
CONCB	81.01	13.30	1	-
EMOTCB	79.36	14.76	1	-
INTELLCB	80.54	13.87	1	-
EXTRACB	80.47	14.91	1	-
AGREECB	82.54	12.61	1	-
Big Five Block Adjective Ratings (BFBD) [1]				
CONCR	3.76	1.07	1	-
EMOTCR	3.58	1.15	1	-
INTELLCR	4.00	0.90	1	-
EXTRACR	3.62	0.85	1	-
AGREECR	4.40	0.68	1	-
Big Five Block Accuracy Scores (BFBD Accuracy 20%)				
CONCBA20%	72.84	44.53		-
EMOTBA20%	70.12	45.83		-
INTELLBA20%	84.20	36.52		-
EXTRABA20%	57.78	49.45		-
AGREEBA20%	91.11	28.49		-
Big Five Item-Level Accuracy Scores Derived from Methods 1, 2, and 3				
Conscientiousness Accuracy	83.62	31.43	3	.81
Emotional Stability Accuracy	80.00	33.79	3	.81
Intellect Accuracy	78.60	31.60	3	.68
Extraversion Accuracy	83.54	28.78	3	.70
Agreeableness Accuracy	80.08	34.45	3	.84
Big Five Bias Scores				
Conscientiousness Bias	-1.73	34.98		-
Emotional Stability Bias	0.67	36.52		-
Intellect Bias	2.07	34.98		-
Extraversion Bias	-2.83	32.32		-
Agreeableness Bias	2.44	36.48		-

Note. ICON = IPIP Conscientiousness; IEMOT = IPIP Emotional stability; IINTELL = IPIP Intellect; IEXTRA = IPIP Extraversion, IAGREE = IPIP Agreeableness; ICONA = IPIP Form A Conscientiousness; IEMOTA = IPIP Form A Emotional stability; INTELLA = IPIP Form A Intellect; IEXTRAA = IPIP Form A Extraversion, IAGREEA = IPIP Form A Agreeableness; CONIC= Conscientiousness Item-by-item confidence Form B; EMOTIC= Emotional stability Item-by-item confidence Form B; INTELLIC = Intellect Item-by-item confidence Form B; EXTRAIC = Extraversion Item-by-item confidence Form B; AGREEIC= Agreeableness Item-by-item confidence Form B; CONCIR= Conscientiousness Item-by-item adjective rating Form B; EMOTIIR= Emotional stability Item-by-item adjective rating Form B; INTELLIR = Intellect Item-by-item adjective rating Form B; EXTRAIR = Extraversion Item-by-item adjective rating Form B; AGREEIR= Agreeableness Item-by-item adjective rating Form B; CONCB= Conscientiousness Block confidence rating; EMOTCB= Emotional stability Block confidence rating; INTELLCB = Intellect Block confidence rating, EXTRACB = Extraversion Block confidence rating; AGREECB= Agreeableness Block confidence rating; CONCR = Conscientiousness Block rating; EMOTCR= Emotional stability Block rating; INTELLCR = Intellect Block rating; EXTRACR = Extraversion Block rating; AGREECR= Agreeableness Block rating; [1] = These means were calculated after recoding the data from -5 to +5 to 1 to 5 (see method section); CONSCBA20% = Conscientiousness Accuracy ≤ 20 %; EXTRABA20% = Extraversion accuracy ≤ 20 %; AGREEBA20% = Agreeableness Accuracy ≤ 20 %; EMOTBA20% = Emotional Stability Accuracy ≤ 20 %; INTBA20% = Intellect Accuracy ≤ 20 %.

4.12 The Factorial Structure of Confidence

The first hypothesis stated that one confidence factor would emerge from the structural analysis of the item-by-item and block confidence ratings. Exploratory factor analytic techniques were used to test this hypothesis. Principal Axis Factoring with oblique rotation was undertaken with the following variables: five mean confidence rating scores for each Big Five dimension obtained from the item-by-item confidence scores; and the five block confidence rating scores. The correlation matrix of these psychometric variables is presented in Table 4.6.

Table 4.6
Correlations among Block and Item-by-Item Confidence Ratings (N = 405)

Variable	1	2	3	4	5	6	7	8	9	10
1. AGREECB	1.00									
2. CONCB	0.59	1.00								
3. EXTRACB	0.46	0.63	1.00							
4. EMOTCB	0.51	0.56	0.64	1.00						
5. INTELLCB	0.48	0.54	0.56	0.60	1.00					
6. EMOTIC	0.53	0.60	0.68	0.78	0.62	1.00				
7. CONCIC	0.59	0.73	0.62	0.59	0.55	0.74	1.00			
8. EXTRAIC	0.54	0.66	0.75	0.62	0.62	0.79	0.79	1.00		
9. AGREEIC	0.72	0.58	0.52	0.53	0.55	0.68	0.74	0.70	1.00	
10.INTELLIC	0.51	0.58	0.61	0.66	0.82	0.78	0.70	0.76	0.65	1.00

Note. CONCB= Conscientiousness block confidence rating; EMOTCB= Emotional stability block confidence rating: INTELLCB = Intellect block confidence rating, EXTRACB = Extraversion block confidence rating; AGREECB= Agreeableness block confidence rating; CONIC= Conscientiousness Item-by-item confidence IPIP Form B; EMOTIC= Emotional stability Item-by-item confidence IPIP Form B; INTELLIC = Intellect Item-by-item confidence IPIP Form B; EXTRAIC = Extraversion Item-by-item confidence IPIP Form B; AGREEIC= Agreeableness Item-by-item confidence IPIP Form B. All correlations were significant at the .01

The data from Table 4.6 were considered factorable as the assumptions as advocated by Coakes and Steed were met (1996). A solution employing root one criterion produced one factor, which accounted for 67.18% of the total variance. The percent of variance accounted for, eigenvalue, and communalities are presented in Table 4.7.

Table 4.7
Principal Axis Factoring of the Item-by-Item and Block Confidence Rating Scores (N = 405)

Variable	Communalities	Factor –Big Five Confidence
AGREECB	.45	.72
CONCB	.57	.79
EXTRACB	.58	.79
EMOTCB	.58	.79
INTELLCB	.55	.77
EMOTIC	.77	.89
CONCIC	.73	.87
EXTRAIC	.78	.89
INTELLIC	.73	.87
AGREEIC	.62	.81
Eigenvalue	-	6.72
% of variance	-	67.18

Table 4.7 shows that the communality values ranged from .51 for AGREECB to .79 for EXTRAIC. The factor extracted from the data set was called Big Five Confidence as all of the personality confidence scores loaded highly on it. This finding suggests that the cognitive processes that underlie personality confidence judgments at the block level are the same as those used when individuals make item-by-item confidence judgments about their personality.

4.13 Gender and Age Differences in Personality Confidence

The next hypothesis proposed that males and females would not differ in terms of Big Five confidence. The gender data were subjected to an independent samples *t* test with the results presented in Table 4.8 [3]. Bonferroni adjustment was used to keep

[3] The male to female ratio in this study is not ideal, however, it closely resembles other calibration research wherein similar gender ratios were reported (e.g., Pallier, 2003). Also, based on the recommendations made by Howell (2002), each effect size calculation used the mean and the standard

family-wise error at alpha = .05. None of the *t*-tests indicated significant gender differences and the effect sizes were in the close-to-zero, or small range, as expected. These results indicate that males did not differ from females in terms of Big Five confidence. This finding essentially replicates the Study 1 results for the block Big Five confidence judgments.

Table 4.8
Means for Block and Item-by-Item Big Five Confidence Ratings for Males (N = 85) and Females (N = 318) in Study 2.

	AGREE CB	CON CB	EXTRA CB	EMOT CB	INTELL CB	EXTRA IC	AGREE IC	CONC IC	EMOT IC	INTELL IC
Male	81.53	81.65	79.76	78.82	82.24	80.07	80.92	80.21	79.74	81.27
Female	82.89	80.85	80.66	79.50	80.09	80.90	82.97	82.34	80.95	80.53
t tests	-0.83	0.49	-0.48	-0.35	1.22	-0.53	-1.54	-1.41	-0.82	0.52
Effect size	-0.10	0.06	-0.06	-0.04	0.15	-0.06	-0.19	-0.18	-0.10	0.06

Note. CONCB= Conscientiousness block confidence rating; EMOTCB= Emotional stability block confidence rating; INTELLCG = Intellect block confidence rating, EXTRACB = Extraversion block confidence rating; AGREECB= Agreeableness block confidence rating; CONIC= Conscientiousness Item-by-item confidence IPIP Form B; EMOTIC= Emotional stability Item-by-item confidence IPIP Form B; INTELLIC = Intellect Item-by-item confidence IPIP Form B; EXTRAIC = Extraversion Item-by-item confidence IPIP Form B; AGREEIC= Agreeableness Item-by-item confidence IPIP Form B

Hypothesis 3 postulated that age would not be associated with Big Five confidence at either the block or item level. Pearson’s Product Moment correlations were calculated in order to test this hypothesis, and the only significant correlations were between age and item-level Conscientiousness confidence ($r_{(405)} = .10$, $p < .05$), and age and item-level Intellect confidence ($r_{(405)} = .11$, $p < .05$). These correlations suggest that older people were more confident than younger participants when rating their confidence for both the Intellect and Conscientiousness dimensions. However, the effect sizes were small in both cases.

deviation for each gender as the denominator. This practice guarantees that d is approximated independently of N, thereby removing potential concerns regarding unequal sample sizes.

4.14 Big Five Accuracy and Bias Scores

The descriptive statistics for the block description accuracy scores were shown in Table 4.5. The accuracy scores for the Conscientiousness, Emotional Stability, Intellect, Extraversion and Agreeableness dimensions were 72.84, 70.12, 84.20, 57.78, and 91.11 respectively. The standard deviation of the block accuracy scores showed that there were many inaccurate ratings for some traits. It was noteworthy that the descriptive statistics presented in Table 4.5 for the block accuracy scores were similar to those reported in Study 1 which used face-to-face mode of administration.

The same validity checks as undertaken in Study 1 (see section 3.11) were undertaken for the block descriptions of personality and the results were concordant with the checks from Study 1 thus for the sake of brevity were not reported here.

Hypothesis 4 proposed that the associations among the Big Five block accuracy, and among the Big Five item-level accuracy scores would be low. Pearson's Product Moment correlations were calculated to test both aspects of this hypothesis. The correlations amongst the block accuracy scores are presented in Table 4.9, and the correlations among the item-level accuracy scores are presented in Table 4.10. The correlations between the block accuracy scores, and the item-level accuracy scores were low, as expected. These correlations suggest that individuals who were accurate (consistent) in their scores on one trait were not necessarily accurate in their scores for other traits.

Table 4.9
Correlations among Big Five Block Accuracy Scores (N = 405).

Variable	CONSCAC20%	EXTRAAC20 %	AGREEAC20%	EMOTAC20%	INTAC20%
CONCBA20%	1.00				
EXTRABA20%	0.15*	1.00			
AGREEBA20%	0.08	0.16*	1.00		
EMOTBA20%	0.15*	0.19*	0.19*	1.00	
INTELLBA20%	0.02	0.13*	0.15*	0.12*	1.00

Note. CONCBAC20% = Conscientiousness Accuracy ≤ 20 %; EXTRABA20% = Extraversion accuracy ≤ 20 %; AGREEBA20% = Agreeableness Accuracy ≤ 20 %; EMOTBA20% = Emotional Stability Accuracy ≤ 20 %; INTBA20% = Intellect Accuracy ≤ 20 %.
* $p < .05$.

Table 4.10
Correlations among Big Five Item-Level Accuracy Scores (N = 405).

Variables	1	2	3	4	5
INTAC	1.00				
CONAC	0.18*	1.00			
EXTAC	0.22**	0.23**	1.00		
AGAC	0.12*	0.10*	0.10*	1.00	
EMOTAC	0.16**	0.10*	0.26**	0.17**	1.00

Note. CONAC= Conscientiousness accuracy item level; EXTAC = Extraversion accuracy item level; AGAC = Agreeableness accuracy item level; EMOTAC = Emotional Stability accuracy item level; INTAC = Intellect accuracy item level
* $p < .05$ ** $p < .01$

Table 4.11 presents the correlations among the block and item-level accuracy scores that were also examined. As shown, significant positive correlations emerged for all Big Five block accuracy scores and their respective item-level accuracy scores, with the Extraversion dimension being the only exception. It is possible that the lack of correlation between the two extraversion accuracy measures may not reflect a lack of any statistical relationship, but rather could represent unknown confounding variables as this study was conducted via the internet in an environment that was not proctored.

Table 4.11

Correlations among Big Five Block and Item-Level Accuracy Scores (N = 405)

Variable	1	2	3	4	5	6	7	8	9	10
INTAC	1.00									
CONAC	0.18	1.00								
EXTAC	0.22	0.23	1.00							
AGAC	0.12	0.10	0.10	1.00						
EMOTAC	0.16	0.10	0.26	0.17	1.00					
INTELLBA20%	**0.28****	0.03	0.05	0.17	0.08	1.00				
CONCBA20%	0.05	**0.31****	0.07	0.04	0.04	0.02	1.00			
EXTRABA20%	0.06	-0.02	-0.08	0.04	0.03	-0.07	-0.07	1.00		
AGREEBA20%	0.04	0.08	0.04	**0.39****	0.11	0.15	0.08	-0.02	1.00	
EMOTBA0%	0.08	0.08	0.20	0.14	**0.35****	0.12	0.15	-0.01	0.19	1.00

Note. CONAC= Conscientiousness accuracy item level; EXTAC = Extraversion accuracy item level; AGAC = Agreeableness accuracy item level; EMOTAC = Emotional Stability accuracy item level; INTAC = Intellect accuracy item level

** $p < .01$

4.14.1 Item-by-item Big Five Accuracy and Bias

Item-level accuracy scores were high for each of the Big Five personality dimensions. The scores were 83.62, 80.00, 78.60, 83.54, and 80.08 for the Conscientiousness, Emotional Stability, Intellect, Extraversion and Agreeableness dimensions respectively.

The item-level Big Five bias scores presented in Table 4.5 were -1.73, 0.67, 2.07, -2.83, and 2.44 for the Conscientiousness, Emotional Stability, Intellect, Extraversion and Agreeableness dimensions respectively. The magnitude of these scores suggested good calibration for each of the Big Five dimensions, as expected.

4.14.2 Gender and Age Differences in Accuracy and Bias

According to hypothesis 6, gender differences were not expected in relation to Big Five accuracy at either the block or item levels. To test this hypothesis, independent samples t-tests were used and the results are presented in Table 4.12. Bonferroni adjustment was used to keep family-wise error at alpha = .05. Consistent with expectations, no gender differences appeared across either the block or the item-by-item ratings suggesting that males were just as accurate as females when making judgments about the Big Five personality dimensions. This outcome replicated the Study 1 findings for the Big Five block accuracy judgments.

Table 4.12
Means for Global Block and Item-by-Item Big Five Accuracy Scores for Males (N = 85) and Females (N = 318) in Study 2.

	AGREE BA20%	CONC BA20%	EXTRA BA20%	EMOT BA20%	INTELL BA20%	EXTAC	AGAC	CON AC	EMOT AC	INTAC
Male	87.06	75.29	74.12	75.29	85.88	80.39	79.21	88.63	77.25	78.43
Female	92.14	72.33	75.47	68.55	83.65	84.38	80.19	82.18	80.71	78.51
t tests	-1.28	.55	-.26	1.25	.50	-1.13	-.23	1.94	-.84	-.02
Effect size	-0.17	0.06	-0.12	0.15	0.06	-0.13	-0.03	0.22	-0.10	-0.00

Note. CONSCBA20% = Conscientiousness Accuracy ≤ 20 %; EXTRABA20% = Extraversion accuracy ≤ 20 %; AGREEBA20% = Agreeableness Accuracy ≤ 20 %; EMOTBA20% = Emotional Stability Accuracy ≤ 20 %; INTBA20% = Intellect Accuracy ≤ 20 %; CONAC= Conscientiousness accuracy item level; EXTAC = Extraversion accuracy item level; AGAC = Agreeableness accuracy item level; EMOTAC = Emotional Stability accuracy item level; INTAC = Intellect accuracy item level

Established support for the hypotheses that gender differences do not occur in relation to Big Five confidence or accuracy, led to the expectation that gender differences in Big Five bias also do not occur. Therefore the non-significant results were not reported.

Age was not expected to be associated with either block or item-level accuracy scores. Nor was age expected to be associated with Big Five Bias scores. Pearson's Product Moment correlations were calculated to test both aspects of this hypothesis. As expected all correlations between age and accuracy, and age and bias were non-significant with one qualification. There was one significant negative correlation between age and the Conscientiousness block accuracy score ($r_{(405)} = -.10$, $p < .05$), however, the effect size was small. The correlation suggests that as age increased accuracy for the Conscientiousness dimension decreased.

4.14.3 Factorial Structure of Big Five Bias

No hypothesis was formed in terms of the factorial structure of the personality bias scores. Nevertheless, preliminary examination of the correlations among the bias scores ascertained whether or not these scores were factorable. Table 4.13 presents these correlations. The Kaiser-Meyer-Olkin (KMO) value was .7, exceeded the recommended value of .6; Bartlett's test resulted in a value of 218.18, which was statistically significant; and the off-diagonal partial correlations of the anti-image matrix revealed mainly small values thus supporting the factorability of the correlation matrix (Tabachnick & Fidell, 1996). In addition, all measures of sampling adequacy exceeded the recommended value of .5 proposed by Tabachnick and Fidell (1996).

Table 4.13
Correlations among Personality Bias Scores (N = 405).

Variable	1	2	3	4	5
Intellect Bias	1.00	0.26	0.33	0.19	0.29
Conscientiousness bias	0.26	1.00	0.34	0.20	0.23
Extraversion bias	0.33	0.34	1.00	0.21	0.35
Agreeableness bias	0.19	0.20	0.21	1.00	0.24
Emotional Stability bias	0.29	0.23	0.35	0.24	1.00

Note. All correlations significant at the .01 level.

Table 4.14
Summary of Exploratory Structural Analysis Results for Big Five Personality Bias Scores, Using Principal Axis Factoring (N = 405)

Variable	h2(b)	F1a
Extraversion bias	.27	.52
Emotional Stability bias	.29	.45
Agreeableness bias	.14	.33
Intellect Bias	.27	.41
Conscientiousness bias	.25	.46
Eigenvalues		2.07
% of variance		41.32

Note. $h^{2(b)}$ = Communalities [a] F1 = Big Five bias.

A solution employing root one criterion produced one factor, which accounted for 41.32% of the total variance. Table 4.14 presented the pattern matrix, percent of variance accounted for, eigenvalue, and communalities for the one-factor solution. The factor was labelled Personality Bias as all Big Five bias scores loaded on it. This one-factor solution suggests that the cognitive processes that underlie bias for each of the Big Five dimensions were the same.

4.15 Factorial Structure of Big Five Confidence and Big Five Subscale Scores

The last hypothesis proposed that a six-factor solution would emerge from the structural analysis of all subscales from both the IPIP and IPIP Form B. Principal Components Analysis with oblique rotation was undertaken to test this hypothesis. The correlation matrix of these psychometric variables is presented in Table 4.15.

The data from Table 4.15 were considered factorable as all assumptions as advocated by Coakes and Steed (1996) were met. Root one criterion produced a 5 factor solution. However Cattell's Scree plot provided support for a six-factor solution, and the sixth eigenvalue was .92 which was high enough to warrant the interpretation of a six-factor solution (Carroll, 1993)

Table 4.15
Correlations among IPIP and IPIP Form B Subscale Scores

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
INTELLIR	1.00														
CONCIR	0.30**	1.00													
EXTRAIR	0.46**	0.27**	1.00												
AGREEIR	0.38**	0.41**	0.43**	1.00											
EMOTIR	0.40**	0.27**	0.44**	0.38**	1.00										
INTELLIC	0.54**	0.19**	0.28**	0.31**	0.25**	1.00									
CONCIC	0.32**	0.28**	0.25**	0.31**	0.16**	0.70**	1.00								
EXTRAIC	0.32**	0.16**	0.31**	0.28**	0.19**	0.76**	0.79**	1.00							
AGREEIC	0.32**	0.21**	0.28**	0.52**	0.25**	0.65**	0.74**	0.70**	1.00						
EMOTIC	0.29**	0.19**	0.24**	0.27**	0.23**	0.78**	0.74**	0.79**	0.68**	1.00					
IINTELL	0.68**	0.05	0.26**	0.19**	0.08	0.41**	0.24**	0.22**	0.23**	0.21**	1.00				
ICON	0.18**	0.72**	0.11*	0.27**	0.18**	0.16**	0.25**	0.11*	0.23**	0.16**	0.11*	1.00			
IEXTRA	0.36**	0.06	0.78**	0.28**	0.31**	0.22**	0.20**	0.25**	0.24**	0.20**	0.30**	0.08	1.00		
IAGREE	0.19**	0.16**	0.18**	0.58**	0.08	0.23**	0.24**	0.21**	0.43**	0.19**	0.26**	0.26**	0.26**	1.00	
IEMOT	0.29**	0.17**	0.31**	0.27**	0.77**	0.22**	0.18**	0.16**	0.23**	0.20**	0.06	0.25**	0.34**	0.14**	1.00

* $p < .05$ ** $p < .01$

Table 4.16 presents the pattern matrix, percent of variance accounted for, eigenvalues, and factor correlation matrix. The six-factor factor solution accounted for 85.32% of the total variance. The first factor was labelled Big Five Confidence and comprised loadings from all IPIP Form B confidence subscale scores. The second, third, fourth, fifth, and sixth factors were labelled Emotional Stability, Conscientiousness, Intellect, Agreeableness, and Extraversion respectively. As anticipated, for each dimension, the last five factors all comprised loadings from the expected IPIP, and IPIP Form B subscale scores. There were low to moderate correlations among the factors. This six-factor solution was in accordance with expectations, and suggested that Big Confidence and Big Five ratings are separate but correlated processes.

Table 4.16
Pattern Matrix for all IPIP and IPIP Form B Subscale Scores, Using Principal Components Analysis with Oblique Rotation (N = 405)

Variable	$h^{2(b)}$	F1[a]	F2[c]	F3[d]	F4[e]	F5[f]	F6[g]
INTELLIR	0.86				.80		
CONCIR	0.89			.94			
EXTRAIR	0.91						-.92
AGREEIR	0.76					-.70	
EMOTIR	0.90		.92				
INTELLIC	0.84	.79			.32		
CONCIC	0.82	.89					
EXTRAIC	0.86	.94					
AGREEIC	0.81	.74				-.36	
EMOTIC	0.83	.92					
IINTELL	0.88				.94		
ICON	0.82			.90			
IEXTRA	0.88						-.93
IAGREE	0.86					-.94	
IEMOT	0.86		.94				
Eigenvalues		5.64	2.14	1.62	1.35	1.13	.92
% of variance		37.62	14.30	10.78	8.98	7.53	6.11
Factor Correlation Matrix							
	F1	F2	F3	F4	F5	F6	
F1	1.00						
F2	0.23	1.00					
F3	0.21	0.26	1.00				
F4	0.29	0.17	0.13	1.00			
F5	-0.28	-0.17	-0.25	-0.19	1.00		
F6	-0.25	-0.38	-0.14	-0.31	0.25	1.00	

Note. $h^{2(b)}$ = Communalities [a] F1 = Big Five Confidence, [c]F2 = Emotional Stability, F3[d] = Conscientiousness, F4[e] = Intellect, F5[f] = Agreeableness, F6[g] = Extraversion. The cut-off for suppression was .20.

4.16 Discussion

The present study investigated personality confidence judgments in more detail by obtaining both item-by-item and block personality confidence judgments. This area of investigation is important because people make many personality judgments in daily life. Previous calibration researchers, however, have not investigated whether individuals are biased when making personality judgments about themselves. Internal consistency co-efficients for the item-by-item confidence ratings were high (i.e., >.90), attesting to the reliability of the confidence scores. These findings were consistent with the cognitive literature reported in Chapter 2, wherein cognitive confidence judgments have also demonstrated high internal consistency (e.g., Kleitman, 2003, 2008; Stankov & Kleitman, 2008).

Study 2 explored various ways of determining personality accuracy. Three methods were developed. For each dimension the means and standard deviations for methods 2 and 3 were reasonably similar. There were significant correlations for each Big Five dimension, among all three methods. It is worthy of note, that within each Big Five domain, the highest correlations were between methods 2 and 3. Factor analysis of all Big Five accuracy scores obtained using three new methodologies, revealed a five-factor solution, suggesting that accuracy was specific to a domain. Thus calculation of mean accuracy scores for each dimension was justified. These methods were re-employed in Study 3. For the rest of the discussion, it is important to note that item-level accuracy scores refer to the mean accuracy scores for each Big Five dimension.

The reliability estimates of the item-level accuracy scores for each Big Five dimension were encouraging. Nevertheless, these scores need to be viewed with caution until the replicability of these findings is demonstrated. Study 3 re-examined Big Five item-level accuracy scores with a different sample. Also promising, was the factorial structure of IPIP Form B was in accordance with expectations. This finding too requires replication and will be re-examined in Study 3.

Results supported the hypothesis that one confidence factor exists with loadings from both the block and item-by-item confidence ratings. This outcome suggests that the cognitive processes that underlie both block and item-by-item confidence judgments are the same. This finding was consistent with the argument presented in the introductory section of this chapter, that all Big Five confidence judgments have trait summaries in place because they come from the lexicon of daily life, and therefore follow abstraction processes. This finding needs to be replicated before tentative conclusions can be made. Study 3 will re-visit this issue.

The next hypothesis reasoned that males and females would be equally confident when making personality judgments about themselves, both the block and the item-by-item-level. As expected, males and females did not differ in terms of personality confidence. The results for the block confidence judgments replicate the findings from Study 1. The lack of gender differences for both block and item-level confidence judgments, was consistent with Hyde’s (2005) gender similarities hypothesis, which states that males and females are more alike than they are different. The effect sizes found in the current study were also in the small and close-to-zero range which is consistent with Hyde’s work. Gender differences in personality confidence judgments were-examined in Study 3, so that preliminary conclusions can be drawn from the data derived from three studies.

It was hypothesised that individual differences in age would not be associated with Big Five confidence at either the block or the item-by-item level. This hypothesis was supported for the item-by-item confidence judgments but not for the block confidence scores. That is, for block confidence judgments of both Agreeableness and Conscientiousness, age was significantly positively correlated with these dimensions with *rs* of .11 and .10 respectively. The effect sizes however, were small.

There were low correlations among the Big Five block accuracy scores, as expected. The lack of intercorrelations in Study 2 replicates the pattern found in Study 1, which used a different sample of participants and a different mode of testing. These results suggest that at the block level, personality accuracy needs to be measured for each personality dimension, and that an overall Big Five accuracy score cannot be calculated. Nevertheless, to verify these findings further replication is required, and Study 3 therefore re-examined the intercorrelations between block level accuracy scores, so that conclusions can be drawn from the data derived from three studies. As was the case in Study 1, block level bias scores were not calculated. Similarly, correlations among the Big Five item-level accuracy scores were also low. It appears that at both the item and block levels, individuals who were accurate (consistent) for one trait were not necessarily accurate (consistent) across other Big Five traits.

The hypothesis that individuals would be well-calibrated when making personality judgments about the Big Five domains, because these traits are likely to have high ecological validity, was supported. The absolute value of the bias scores for each Big Five domain was close to five suggesting good calibration. The results for the bias scores was consistent with Gigerenzer et al.'s (1991) Ecological theory, which posits that individuals are well-calibrated to their natural environments. Item-level bias scores were re-examined in Study 3.

The hypothesis that males and females would not differ in terms of personality accuracy at either the block level or item-by-item level, was supported. The findings for the block judgments replicate the results from Study 1. The overall findings were consistent with the gender similarities hypothesis as advocated by Hyde (2005). These findings, however, need further replication before conclusions can be drawn, and were therefore re-examined in Study 3.

Individual differences in age were not expected to be associated with Big Five accuracy at either the block level or the item-by-item level. This hypothesis was supported, with the qualification that age was negatively associated with block Conscientiousness accuracy ($r = .10$). Nevertheless, the effect size was small.

The hypothesis that individual differences in age would not be associated with item-level bias was supported. The overall findings for age are largely consistent with the Five Factor theory of personality (McCrae & Costa, 1999), which posits that personality traits develop during childhood and adolescence and then remain stable in adulthood. It appears that Big Five accuracy, confidence and bias remains constant too. Taken as a whole, the findings for age also fit well with PMM theory (Gigerenzer et al., 1991) which argues that people are be well calibrated to their natural ecology, therefore age differences were not expected. It was therefore argued in Studies 1 and 2 that confidence also remains stable in adulthood, and consequently age does not correlate with Big Five confidence, accuracy or bias scores. It is also interesting to note that the small but significant positive correlation between age and Intellect confidence found in Study 1 does not emerge from the data in Study 2. Age differences in personality confidence were be examined again in Study 3.

Although the factorial structure of Big Five bias scores was examined in the current study, no hypothesis was formulated. Results showed that one Personality bias factor emerges from the structural analysis of the five item-level bias scores. This one-factor solution suggests that the cognitive processes that underlie bias for each of the Big Five dimensions are the same. The factorial structure of item-level Big Five bias scores was re-examined in Study3. The one-factor solution for the Big Five item-level bias scores is concordant with previous research from the cognitive domain which shows that if individuals are miscalibrated in one domain then they are also miscalibrated across other domains (Pallier et al., 2002; Stankov, 1998, 1999a). The question that remains unanswered is whether personality bias shares variance with cognitive bias. Study 3 examined the factorial structure of personality and cognitive bias scores.

A six-factor solution was expected from the structural analyses of all the IPIP and IPIP Form B Big Five subscale scores (i.e., adjective ratings and confidence scores). This hypothesis was supported and suggests that Big Confidence and Big Five ratings, are separate but correlated processes. This six-factor solution also provides evidence that confidence is related to, but is distinct from personality which is in agreement with research in the cognitive domain which has demonstrated that the confidence trait is on the borderline between cognitive abilities and personality (Baker, 2001; Kleitman & Stankov, 2001, 2007; Pallier et al., 2002; Stankov, 1998, 1999a, 2000a; Stankov & Crawford, 1996a, 1997; Stankov & Lee, 2008). This result requires replication and was re-examined in Study 3.

One of the shortcomings of this study was that it examined the impact of a limited number of individual differences variables (i.e., age and gender) on Big Five confidence, accuracy, and bias. It is important that research continues to investigate other individual differences variables that may influence confidence, accuracy and bias in the Big Five domain. Study 3 extended this investigation by examining the relationships between personality confidence and (a) affect, (b) private self-conscientiousness and (c) need for cognition. The rationale for including these variables is left until Study 3.

Another limitation of this study was that due to technical restrictions, the ordinal position of the measures could not be randomised across participants. Consequently, the block ratings for each Big Five dimension were made after the item-level ratings for that dimension had been made. Practice effects may have influenced the results. In an attempt to overcome this limitation, in Study 3, the order of presentation of the IPIP, and the other Big Five block and item-level measures, changed.

In summary, Study 2 aimed to examine the personality confidence ratings in more detail and develop methods by which mean accuracy scores can be calculated for each of the Big Five dimensions. In brief, findings suggest that individuals in the current sample used the same cognitive processes when making both block and item-level confidence ratings about their personality. This conclusion needs to be viewed with caution until the factorial structure of block and item-by-item Big Five confidence is re-examined in Study 3. Individuals in Study 2 were also well calibrated for each of the Big Five dimension. Mean accuracy scores were high and the reliability of these scores was encouraging. Also, a number of other findings require replication before preliminary conclusions can be drawn and Study 3 re-examined this. Nevertheless, at this stage, it appears that neither age nor gender influence personality accuracy, confidence, or bias. For the block confidence ratings in particular, data from Studies 1 and 2 do provide some evidence in favour of this tentative conclusion. The finding that Big Five bias scores load onto one factor suggests that the same cognitive processes underlie these scores. Study 3 re-examined the factorial structure of Big Five bias scores so that conclusions can be drawn from two studies. However, a major question that remains unclear is whether Big Five bias is related to cognitive bias. This particular issue was examined in Study 3, wherein participants provided confidence ratings at both the item-by-item, and block levels, across both the personality and cognitive domains.

Chapter 5 - Study 3

5.1 Introduction

Study 3 builds on Studies 1 and 2, and aimed to (a) examine bias (item and PTPE) with Gf tasks; (b) investigate whether individual differences in age, gender, ability, personality, need for cognition, and negative affect, influence cognitive confidence scores; (c) confirm previous findings from Studies 1 and 2 for Big Five confidence, accuracy and bias judgments; and (d) examine both the discriminant and convergent validity of Big Five confidence scores in relation to self-focussed attention, affect, and need for cognition.

Following this, Study 3 investigated the factorial structure of both personality confidence scores and cognitive confidence scores. Additional confidence ratings were taken from a measure of self-report abilities and were factor analysed with the Big Five and Gf confidence measures. Finally, the factorial structure of both cognitive bias scores and Big Five bias scores were examined.

5.2 Cognitive Calibration and Confidence on Gf Tasks

5.2.1 Bias in Relation to Gf Tasks

Conflicting findings have resulted from calibration research that uses reasoning tasks to examine bias (i.e., Esoteric Analogies and Letter Series), with some studies demonstrating good calibration (Crawford & Stankov, 1996; Kleitman, 2003) and others demonstrating overconfidence (Pallier, 2003; Stankov, 2000a). In Study 1, the Esoteric Analogies task demonstrated overconfidence and loaded with Gc tasks rather than Gf tasks. This Esoteric Analogies task can load onto either Gf or Gc factors, depending on the composition of tasks used in the test battery. For Study 3, three reasoning tasks (Esoteric Analogies, Letter Series, and Word Association) were selected. Because the findings from pervious research are conflicting, no specific hypotheses were formulated for the item-level bias scores. These scores, however, were examined and reported, along with their calibration curves and item-specific scatterplots. For the post-test performance estimate (PTPE see Chapter 2) bias scores , previous research most often demonstrates better calibration with the PTPE bias scores than with item-level bias scores (Kleitman & Stankov, 2001; Stankov, 2000a; Stankov & Crawford, 1996a, 1996b, 1997). This same pattern was expected in Study 3.

Stankov and his collaborators (Pallier et al., 2002; Stankov, 1998, 1999a) factor analysed bias scores obtained from various combinations of Gc, Gf, and Gv tasks, and found that the bias scores loaded onto one factor. These results indicated that mis-calibration across a diverse range of abilities and perceptual tasks were driven by the same cognitive processes. In view of these findings, a single bias factor was expected to emerge from factor analysing the item-level bias scores obtained from the Gf tasks used in the current study.

Study 3 also investigated whether individual differences in age, gender, ability and personality influenced cognitive confidence and bias. As the rationale for these analyses was presented in Study 1, it is not repeated here. Hypotheses two to five in section 5.9 detail the expectations.

5.2.2 Factorial Structure of Item-Level Cognitive Confidence Ratings and Confidence Ratings Obtained in Relation to PTPEs

Previous researchers have argued that item-level confidence ratings initiate the meta-cognitive process of self-monitoring, and have obtained these ratings as individuals work through the items of a cognitive task (e.g., Kleitman & Stankov, 2007; Stankov, 1999b). However the literature also points to the distinction between the metacognitive processes of self-monitoring (i.e., item-by-item confidence judgments) and self-evaluation (Schraw & Moshman, 1995). The evaluative aspect of meta-cognition is measured after the completion of each test, when participants estimate the percentage of items they have solved correctly (i.e., post-test performance estimate or PTPE). A bias score similar to an item-level bias score can also be obtained using the PTPE score (Stankov & Crawford, 1996b). This is achieved by subtracting the actual mean percentage of correct responses from the estimated percentage of correct responses.

Previous researchers, however, have not yet asked participants the question, “How confident are you that your percentage estimate is correct?”. Study 3 explored this new territory by obtaining confidence ratings in relation to post-test performance estimates. Some researchers may well argue that the PTPE estimate itself is a confidence type rating, but Kleitman and Stankov (2001) demonstrated that PTPE scores for the tasks used in their study (Geography, Line Length, an Raven’s Progressive Matrices), did *not* load onto a separate item-level confidence factor. Rather, these scores defined factors which also had high loadings from accuracy, speed and expectancy measures. An expectancy measure was defined as a pre-test performance estimate (i.e., what percentage of items do you expect to answer correctly). As PTPE estimates did not load with item-level confidence, they concluded that self-monitoring and evaluation were distinct factorially, and that evaluation was not distinct from accuracy measures. Consequently, for the purposes of Study 3, the PTPE was simply an estimate, used to take the current research further than previous research studies had gone, by taking a confidence rating in that estimate. Two assumptions were made; that confidence judgments initiate self-monitoring, and that confidence in relation to the PTPE represents a construct similar to the evaluative aspect of meta-cognition as operationalised by Schraw and Dennison (1994, see section 2.2). The question remains, are confidence judgments made in relation to post-test performance estimates driven by the same cognitive processes as the item-level confidence ratings? Two correlated factors were expected to emerge from the structural analyses of item-level and PTPE confidence scores because; these measures could theoretically represent the two aspects of meta-cognition- self-monitoring and self-evaluation. Because the factorial structure of Big Five and cognitive confidence scores were examined (see Section 5.6 for the rationale) in Study 3, obtaining an evaluative cognitive confidence measure was important.

5.3 Other Individual Differences in Cognition

5.3.1 Need for Cognition in Cognition: Gf Tasks

A paucity of research has examined whether individual differences in need for cognition (NFC) influence cognitive confidence and mis-calibration. Wolfe and Grosch (1990) found significant positive correlations between NFC and confidence ratings on both an information task ($r = .25$, $p < .01$) and a person prediction task ($r = .25$, $p < .01$). Jonsson and Allwood (2003) found that NFC was also significantly positively related to confidence ratings for both a word knowledge task ($r = .46$, $p < .01$) and a logical spatial ability task ($r = .47$, $p < .01$). Regarding mis-calibration, however, NFC was not correlated with bias scores. More recently, Blais, Thompson, & Baranski, (2005) investigated whether cognitive styles (i.e., NFC, Personal Need for Structure, and Personal Fear of Invalidity) accounted for individual differences in both cognitive confidence and mis-calibration on a general knowledge task, a vocabulary task and a perceptual task. Their results indicated that NFC did not significantly influence confidence or mis-calibration. None of the studies used Gf tasks, however, so for Gf tasks, the role of NFC in cognitive confidence and mis-calibration remains unexplored.

The findings that were very relevant to Study 3 are that high NFC individuals scored higher on tasks of verbal reasoning (Cacioppo, Petty, Kao, & Rodriguez, 1986), endorsed themselves as being more effective problem solvers (Heppner, Reeder, & Larson, 1983), processed information with greater efficiency, and devoted greater effort in decision making tasks, which in turn has led to greater accuracy (Levin, Huneke, & Jasper, 2000) than their low NFC counterparts. Therefore, in view of the findings by Cacioppo et al. (1986) and Levin, Huneke, and Jasper (2000), it was assumed that confidence in the accuracy of self-assessment would also increase. For these reasons NFC was expected to be positively associated with Gf confidence.

5.3.2 Negative Affect in Cognition: Gf Tasks

Few studies have investigated the impact of negative affect (NA) with regard to confidence in the accuracy of self-assessment. Using Levin and Stokes's Negative Affectivity scale, Wolfe and Grosch (1990), found significant negative correlations between negative affect and confidence on both a factual information task ($r = -.16$, $p < .05$) and a writing discrimination task ($r = -.16$, $p < .05$). Moreover, these correlations remained significant even when the effects of accuracy were partialled out. Conversely, for general knowledge questions, Allwood and Bjorhag (1991) did not find associations between depressed mood and either confidence or mis-calibration. More recently, also for general knowledge questions, Allwood, Granhag, and Jonsson (2002), used music and film to create a happy mood for half their participants, and a sad mood for the rest, to investigate the impact of mood on cognitive confidence and mis-calibration. Allwood et al. successfully induced a happy mood but not a sad one, therefore the influence of sad mood in relation to cognitive confidence and mis-calibration went unanswered. However, positive affect (PA) did not influence cognitive confidence or bias scores. Because the effect of NA on cognitive confidence remains uncertain, it was examined in Study 3. For the Gf tasks, used in the current study, NA was expected to be negatively associated with cognitive confidence scores. However, there was no reason to speculate that positive affect (PA) shares variance with cognitive confidence and bias measures. Therefore, PA will not be discussed further.

5.4 Replication Analyses in Relation to Personality

Study 3 investigated the factorial structure of Big Five confidence judgments (both item and block-levels), and also whether individual differences in gender, and age influenced Big Five confidence, accuracy and bias. Because the rationale for these analyses was presented in Study 2, it was not restated here. Hypotheses 8 to 17 in section 5.10 clarify the expectations for Study 3.

5.5 Validity Checks of Big Five Confidence Judgments

The discriminant validity of Big Five confidence judgments is discussed below in relation to the conceptually distinct constructs of self-focussed attention (PrSc) and NA. The convergent validity of Big Five confidence judgments is then explored in regard to the potentially related constructs of PA and NFC.

5.5.1 Self-Focussed Attention in Relation to Personality

Chapter 4 presented the argument that personality confidence judgments follow abstraction processes, and that individuals would not undertake a serial search of episodic memory for the Big Five personality dimensions because these traits have high ecological validity. The abstraction view-point implies that individuals do not constantly focus their attention inwardly when making Big Five personality confidence judgments about themselves. It follows then that, confidence in rating one’s personality traits would not be associated with self-focussed attention, labelled as private self-consciousness by Fenigstein, Scheier, and Buss (1975).

5.5.2 Negative Affect in Relation to Personality

Negative affect (NA) is defined as a mood-dispositional dimension that encompasses a range negative mood states including apprehension, anger, derision, disgust and guilt (Watson, Clark, & Tellegen, 1988). Conceptually, it is reasonable to assume that confidence in Big Five judgments is distinct from this negative mood state. It was therefore expected that Big Five confidence scores would not be associated with NA.

5.5.3 Positive Affect in Relation to Personality

Watson et al. (1988) argued that at the factorial level, NA and positive affect (PA) represent two orthogonal factors of affective structure. Trait PA was defined as the degree to which an individual reflects enthusiasm, high energy and concentration (Watson et al., 1988), with high PA reflecting positive characteristics like joy and self-confidence (Fromson, 2006; Watson et al., 1988). The fact that self-confidence has been identified as one aspect of PA, led to speculation about the relationship between trait PA and personality confidence judgements. The associations between PA and Big Five confidence scores were subsequently investigated, and anticipated to be positive.

5.5.4 Need for Cognition in Relation to Personality

Individuals who are high in NFC actively seek and enjoy cognitive activities that are challenging (Cacioppo & Petty, 1982; Cacioppo, Petty, Feinstein, & Jarvis, 1996). High NFC individuals have been called chronic cognizers whereas, their low NFC counterparts have been labelled as cognitive misers (see Cacioppo et al., 1996, for a review). Of interest is that the construct definition of NFC sounds similar to that of the Big Five Intellect dimension. Individuals high on this dimension are characterized by terms such as curious, creative, innovative, and inquisitive (Hampson, Goldberg, Vogt, & Dubanoski, 2006), and by the readiness to consider new information (Sadowski & Cogburn, 1997). The research by Sadowski and Cogburn investigated the relationship between NFC and Openness to Experience (Intellect) and found that there was a significant positive correlation (r = .50) between these two variables. Thus people who are high on NFC may be likely to be more confident for this dimension. Positive associations between NFC and Intellect confidence at both the block and item levels were therefore expected.

The prospect that NFC is positively associated with Conscientiousness confidence was also feasible. Conscientiousness is measured by adjectives such as hardworking, organised and task oriented (Hampson et al., 2006; Sadowski & Cogburn, 1997) which also sounds similar to the construct definition of the NFC. A short form of the NFC scale for use with Australians was selected for the current study (Forsterlee & Ho, 1999).

5.6 Factorial Structure of Personality and Gf Confidence Scores

In Study 1, Big Five block confidence ratings and item-level cognitive confidence ratings obtained across a diverse range of cognitive tasks, defined separate factors that were not significantly correlated (r = .05). The factorial structure of item-level cognitive and Big Five confidence ratings were re-examined in Study 3. Two confidence factors were expected to emerge from the structural analyses of confidence scores obtained from the Gf and Big Five confidence measures. This expectation was developed with reference to Self-Concept theory (Marsh, 2008; Shavelson et al., 1976);

PMM theory (Gigerenzer et al., 1991), and the results from Study 1 (see section 3.5 for the rationale).

A question that remains unexplored is whether Big Five confidence shares variance with confidence ratings made in relation to PTPEs within the cognitive domain. At the theoretical level, item-level confidence judgments within the cognitive domain have been argued to initiate self-monitoring whereas a novel aspect of this study was to obtain a confidence rating in relation to the PTPE score. In other words, individuals considered how confident they were that their post-test estimate was accurate. Section 5.2.2 explained the PTPE has been operationalised as the evaluative aspect of meta-cognition; therefore a confidence rating in that estimate is more evaluative again. Assuming then, that trait summaries in semantic memory are also evaluative, it is possible that a confidence rating in the PTPE shares more variance with Big Five confidence. That is, trait summaries are the product of numerous self-evaluations, and others' evaluations of one's personality, that have developed over childhood and adolescence, and have remained reasonably stable over the life-span. If these assumptions are true, three factors would emerge from the structural analyses of the Big Five and Gf confidence scores. That is, a Big Five confidence factor defined by the block and item-level Big Five confidence ratings, a Gf item-level self-monitoring confidence factor defined by the item-by-item Gf confidence ratings, and a Gf self-evaluative confidence factor defined by the three confidence ratings made in relation to the PTPEs.

In addition to obtaining confidence scores from objective cognitive tasks, confidence ratings were taken from a measure of self-report abilities. This was done in order to investigate whether these scores shared variance with either objective or personality confidence ratings, or with both.

5.7 Self-report Intelligence/Abilities Confidence and Personality Confidence Ratings

The results from Study 1 showed that when cognitive and personality confidence scores were factor analysed, there was a lack of correlation (r = .05) between the personality confidence and the cognitive confidence factors. One possible explanation for this lack of correlation posited was that cognitive confidence judgments were made with reference to tests of maximal performance, whereas personality confidence judgments were elicited from measures of typical performance. It may be that, if participants were given a self-report measure of their general intelligence/abilities that also elicited confidence ratings for those abilities, then cognitive self-report confidence might share variance with personality confidence. This possibility was investigated in the current study.

To date researchers have not examined whether confidence ratings from a measure of self-report general intelligence (abilities), uses the same cognitive processes as those employed when making personality confidence judgments. Confidence in rating self-beliefs about one's general abilities could well share variance with personality confidence because participants would be asked to rate how they typically are in the abilities domain where confidence ratings would not be tied to objective performance. It is surmised that individuals could engage in a similar process to that which has been argued in relation to personality confidence judgments. That is, it was argued in Chapter 4 that individuals would make their personality confidence ratings from a store within semantic memory that contains a delimited number of trait generalizations (please refer to abstraction theory in section 4.2). To this end a measure of general abilities was chosen and adapted for the purposes of the current study. Confidence ratings for items such as “I am intelligent” and “I have a good vocabulary” could feasibly also reside in a store within semantic memory which contains an individual's self-beliefs about how one typically is ability-wise. It is for this reason that it was expected that the cognitive processes underlying confidence judgments about self-report abilities, would differ from those used to make confidence judgments about cognitive test items. If this is true, then self-report confidence in general abilities can be expected to share more variance with personality confidence judgments than with objective cognitive confidence scores. It was also expected to load onto the personality confidence factor when all of the Big Five (i.e., item and block), cognitive, and self-report general intelligence confidence scores were factor analysed in Study 3.

5.8 Factorial Structure of Cognitive and Personality Bias Scores

The unanswered question is whether Big Five bias shares variance with cognitive bias. If one considers PMM theory (Gigerenzer et al., 1991), bias across the personality and cognitive domains should not be highly correlated because the cues used to respond to cognitive tasks would differ from those used to respond to personality items. Moreover, considering self-concept theory (see Chapter 2), and assuming that bias in cognitive and personality judgments separate across both the abilities and personality (non-ability) domains, it is reasonable to expect that personality and cognitive bias scores would define two separate factors. Investigating whether at the factorial level, personality bias splits from cognitive bias, has important implications for calibration theorists trying to understand self-monitoring, yet this topic has not been examined previously.

5.9 Hypotheses in Relation to Cognition

The hypotheses for the Gf tasks are summarised as follows:

1. It was hypothesised that PTPE bias scores would demonstrate better calibration than item level bias scores. This hypothesis was formulated from previous research findings (Kleitman & Stankov, 2001; Stankov, 2000a; Stankov & Crawford, 1996a, 1996b, 1997).
2. It was hypothesised that age would be positively related to Gf item-level bias scores. Calibration studies that demonstrated small but significant positive correlations between age and bias scores obtained from Gf, Gc, and Gv tasks (Crawford & Stankov, 1996; Pallier, 2003), and from the results obtained in Study 1 which showed a positive association between age and Gf bias scores, underpinned this hypothesis
3. It was hypothesised that males would be more confident than females for all three Gf tasks used in the current study. The basis for this hypothesis came from previous research, which indicated that for cognitive tasks, males have demonstrated significantly higher levels of confidence than females (Pallier, 2003; Pulford & Colman, 1997; Ross & Fogarty, 2006; Stankov, 1998).
4. For the Gf tasks, it was hypothesised that low scorers (i.e., in the bottom quartile) would be more mis-calibrated than high scorers (i.e., top quartile). The rationale for this hypothesis was based on previous research, which demonstrated that those of lower ability have some difficulty in accurately appraising their cognitive abilities compared with those of higher ability (Dunning et al., 2003; Kruger & Dunning, 1999; Maki et al., 1994; Moreland et al., 1981; Shaughnessy, 1979) and also on findings from Study 1.
5. Based on the results by Schaefer et al (2004) it was hypothesised that, after controlling for the influence of the other four personality dimensions, that:

5.1. Openness/Intellect scores would be significantly related to Gf confidence scores.

5.2. Extraversion would be associated with Gf confidence and bias scores.

5.3. Conscientiousness would be significantly related to confidence scores.

6. Based on the findings by Wolfe and Grosch (1990) coupled with the research that showed that individuals high on NFC were more accurate (Levin et al., 2000), it was expected that confidence in the accuracy of self-assessment would also increase. It was hypothesised that NFC would be positively associated with Gf confidence scores.

5.10 Replication Hypotheses in Relation to Personality

The replication hypotheses based on findings from Studies 1 and 2 in relation to Big Five confidence, accuracy and bias scores, are summarised as follows:

7. One confidence factor would emerge from factor analysing the item-by-item and block confidence judgments obtained from the Big Five personality dimensions.
8. Males and females would not differ in terms of Big Five accuracy judgments obtained from either the item-level method or the block description method.
9. Males and females would not differ in terms of Big Five confidence judgments obtained from either the item-level method or the block description method.
10. Males and females would not differ in terms of Big Five bias.
11. Age would not be associated with Big Five confidence scores obtained from either the item-level method or the block description method.
12. Age would not be associated with Big Five accuracy scores obtained from either the item-level method or the block description method.
13. Age would not be associated with Big Five bias.

14. The associations among the Big Five block description accuracy scores and item-level accuracy scores would be low.
15. Good calibration was expected for the Big Five item-level bias scores.
16. Based on previous research, the results from Study 2, and the claim that confidence is a distinct trait that is related to personality, it was hypothesised that six factors would emerge from the structural analyses of the Big Five confidence and Big Five subscale scores from the IPIP and IPIP Form B.

5.11 New Hypotheses: Study 3 in Relation to Personality

In order to achieve the other aims of Study 3 the following hypotheses were tested:

17. As the abstraction view-point implies that individuals do not constantly focus their attention inwardly when making Big Five personality confidence judgments about themselves it was hypothesised that confidence in rating one's personality traits would not be associated with PrSc at either the block description level or item level.
18. As NA is defined as subjective distress it was hypothesised that trait NA would not be associated with Big Five confidence at either the block description level or item-level.
19. Because high PA reflects positive characteristics such as joy, enthusiasm, interest, and self-confidence (Fromson, 2006; Watson et al., 1988), and self-confidence has been identified as one aspect of PA, it was hypothesised that positive associations would emerge between PA and Big Five confidence scores.
20. High NFC individuals are defined as actively seeking and enjoying cognitive activities that are challenging (Cacioppo & Petty, 1982; Cacioppo et al., 1996), characteristics that share striking similarities with the Big Five dimensions of both Intellect and Conscientiousness. Previous research has indeed demonstrated that

NFC was moderately correlated with both Intellect and Conscientiousness (Sadowski & Cogburn, 1997). It can be expected then people who are high on NFC will also be more confident for both these dimensions at both the block description level and the item-level.

5.12 New Hypotheses Study 3 in Cognition and Personality

21. It was hypothesised that three confidence factors would emerge from the structural analyses of the confidence scores obtained from factor analysing Big Five (item and block), Gf item-level, and Gf evaluative confidence measures. This hypothesis was developed with reference to Self-Concept theory (Marsh, 2008; Shavelson et al., 1976), Gigerenzer’s (1991) PMM theory, and abstraction theory within the memory domain (e.g., Buss & Craik, 1983; Klein, 2004; Klein & Loftus, 1993b; Klein, Loftus, Trafton et al., 1992).
22. Based on abstraction theory within the memory domain (e.g., Buss & Craik, 1983; Klein, 2004; Klein & Loftus, 1993b; Klein, Loftus, Trafton et al., 1992), cognitive processes that underlie confidence judgments on self-report abilities were expected to differ from those used to make confidence judgments about cognitive test items. This is because individuals are likely to have summaries about their abilities stored in semantic memory. It was hypothesised that, after all of the Big Five (i.e., item and block) and Gf confidence scores were factor analysed that the self-report intelligence confidence score would load on the personality confidence factor.
23. Based on Self-Concept theory (Marsh, 2008; Shavelson et al., 1976) and Gigerenzer’s (1991) PMM theory, it was hypothesised that two factors would emerge from the structural analyses of the item-level cognitive bias scores and the item-level Big Five bias scores.

5.13 Method

5.13.1 Participants

Two hundred and forty three individuals participated in this study. The sample comprised males (*n* = 61) and females (*n* = 182), ranging in age from 17 to 62 years (*M*= 29.04 years, *SD* = 9.80 years). The mean age of the males was 27.70 years (*SD* = 8.45 years). The mean age for females was 29.46 (*SD* = 10.21 years).

The highest educational level of the sample varied from completion of grade 10, 11, or 12 (*n* = 32), to completion of tertiary studies (*n* = 42). The rest of the sample comprised undergraduate students. Two participants did not respond to this question. Due to various difficulties involved in recruiting participants, emails were sent to all faculty heads at the University of Southern Queensland, asking permission to send out global emails to students in their courses via their respective distribution lists. Snowball sampling techniques were also employed. Participants enrolled in undergraduate Psychology courses at the University of Southern Queensland received course credit for their participation. Other participants received the opportunity to enter a raffle for cash prizes in return for taking the time to complete the study.

5.13.2 Materials

Demographic questions consisted of items regarding gender, age, and highest level of education. All participants completed a computerised battery of three cognitive tasks and a number of self-report inventories. Each measure is described below.

Cognitive Tests: For each cognitive test participants provided an answer to every trial, as well as a confidence rating indicating how confident they were that the answer provided was correct. For the open ended Letter Series and Word Association tests, confidence ranged from 0% (Just guessing) to 100% (Absolutely certain). For the multiple choice Esoteric Analogies test, the starting point on the confidence scale was 100/k, where k = the number of response alternatives. At the end of each cognitive test described below, participants were presented with the following instructions:

> At the end of the test, please provide an estimate of the percentage of items you think you answered correctly. It is IMPORTANT that you provide this estimate IMMEDIATELY after completing the test. After providing your percentage estimate, I want you click on how confident you are that your percentage estimate is correct using the confidence scale that appears on the screen.

The confidence scale for the PTPE estimate ranged from 0% (Just guessing) to 100% (Absolutely certain) using increments of ten percentage points.

Letter Series Test (LST) - (Stankov, 1997). See study 1 for details.

Esoteric Analogies Test (EST) - (Stankov, 1997). Details are the same as reported in Study 1.

Word Association test (WAT) - (Stankov, 1997). Participants were presented with a two-word stimulus (e.g., “number” and “nobility”) and were asked to provide one word that was associated with both stimulus words. In the example of number and nobility, a correct answer would be count. Participants responded to 10 trials within a time limit of four minutes (see Appendix O).

IPIP based measures: The test battery comprised IPIP, IPIP-Form B, and the BFBD. IPIP Form A (see below) was the measure used in the data analyses for the item-level accuracy methods and, was formed from participants’ responses to the IPIP. Prior to developing IPIP Form A, the factorial structure of the IPIP was examined (see results) and reliability estimates were consulted for each of the IPIP subscales. Because the IPIP Form A was used as the criterion against which the accuracy scores were measured, its factorial structure was also examined. Due to the constraints imposed by word limits, these results, which were in accordance with expectations, were not reported.

The International Personality Item Pool Five-Factor Personality Scale (IPIP) - (Goldberg, 1997). See Study 1 for a detailed description. The IPIP was the criterion for accuracy for the BFBD.

Big Five Block Descriptions of Personality (BFBD) Based on the Trait Adjectives - (Goldberg, 1999). Refer to Study 1 for further detail.

Item-by-Item Big Five Self-Rated Personality Descriptions (IPIP Form B) Based on the Trait Adjectives - (Goldberg, 1997). Study 2 provides all relevant details.

Shortened version of the IPIP Scale based on the Trait Adjectives (IPIP Form A) - Goldberg (1997). See Study 2 for details.

Other self-report measures:

Private Self-Consciousness scale (PrSC) - (Fenigstein et al., 1975). The scale consists of eight items measuring individual differences in self-focused attention (e.g., “I’m generally attentive to my inner feelings”). Participants rated themselves on a 5-point scale from 0 (*extremely uncharacteristic)* to 4 (*extremely characteristic)*. The original scale by Fenigstein et al. comprised 10 items, however, subsequent research demonstrated that items 3 and 9 were unreliable, and therefore were not used in the current research (e.g., see Burnkrant & Page, 1984). Burnkrant and Page (1984) reported acceptable internal consistency for the shortened version ($\alpha = .75$). Scores can range from 0 to 32 with higher scores indicating higher levels of private self-consciousness. The PrSC scale is presented in Appendix P.

The Need for Cognition Scale short version (NFC) – (Cacioppo, Petty, & Kao, 1984). The 18 item short version measures individual differences in the “tendency to engage in and enjoy effortful cognitive endeavours" (Cacioppo et al., 1984, p. 306). For example, "I would prefer complex to simple problems". Participants rated each item on a 5-point Likert-type scale that ranged from -2 (*Very strong disagreement*) to 2 (V*ery strong agreement*). Scores could potentially range from -36 to + 36 with higher scores indicative of a greater level of NFC. Acceptable internal consistency co-efficients (α = .81) have been reported in the research literature (e.g., Forsterlee & Ho, 1999) when the short form was used with an Australian sample. The need for cognition scale is presented in Appendix Q.

Positive and Negative Affect Schedule (PANAS) - (Watson et al., 1988). This 20 item scale measures individual differences in positive and negative affect. Each subscale comprised 10 items (e.g., "inspired"), and respondents rated each item on a 5-point scale from 1 (V*ery slightly or not at all)* to 5 (E*xtremely)*. Varying time-frames can be specified for the instructions of the PANAS, in Study 3, respondents rated each emotion in terms of how they generally feel. Scores for each subscale can potentially range from 10 to 50. The PANAS is a psychometrically reliable and valid measure of affect with research reporting internal consistency coefficients of at least 0.85 (e.g., Fogarty et al., 1999). The PANAS is presented in Appendix R.

Self-report Intelligence and Confidence Questionnaire (SICQ): The SICQ is a 10 item scale adapted from the Self-report Intelligence Questionnaire (SRIQ) of Gignac, Stough, and Loukomitis (2004). All of the items in Gignac et al's scale were retained. One extra item ("I am good at being able to perceive patterns in a series of numbers or letters") was included because the Letter Series test in the current study required participants to perceive the pattern in each string of letters (e.g., A, D, G, J, ?), then provide the next letter in the series. In addition, to fulfil one of the aims of the current study, which was to investigate the relationship between self-report cognitive confidence (SICQ) and objective cognitive confidence (i.e., confidence ratings obtained from the LST, EAT, and WAT tests), the instructions and rating scales of the SRIQ were changed, and confidence ratings were added. SICQ is presented in Appendix S.

5.13.3 Procedure

Participants were recruited via PESUD (see method section study 2). Again, all data were submitted electronically and collected by the Psychology technical team, who then forwarded the completed data set for data screening. Informed consent was obtained at the beginning of the testing session when participants read the electronic consent form and were asked to type in a unique identifying number indicating that they understood that their data would be used for research purposes (see Appendix T for covering page for Study 3). Respondents were informed that they could withdraw from the study at any time without any penalty and were assured of confidentiality and anonymity. For the cognitive tasks described above, participants were able to skip items they were unsure of within each test by clicking the next button, and returned to missed questions if time limits allowed. The rest of the test battery, however, required an answer to be provided for each test item before respondents were able to progress to the next question. The battery took approximately two hours to complete.

Due to programming restrictions it was not possible to randomise the order in which the measures were presented. Therefore, the test battery was administered in the following order:

1. Letter Series Test –LST (15 items)
2. Word Association Test-WAT (10 items)
3. Esoteric Analogies Test-EAT (24 items)
4. Block Conscientiousness judgment (BFBD-C)
5. Item-by-item Intellect judgments (IPIP Form B subscale)

6. Block Extraversion judgment (BFBD-E)
7. Item-by-item Emotional Stability judgments (IPIP Form B subscale)
8. Block Agreeableness judgment (BFBD-A)
9. Item-by-item Conscientiousness judgments (IPIP Form B subscale)
10. Block Intellect judgment (BFBD-I)
11. Item-by-item Extraversion judgments (IPIP Form B subscale)
12. Block Emotional Stability judgments (BFBD-EM)
13. Item-by-item Agreeableness judgments (IPIP Form B subscale)
14. PANAS (20 items)
15. Need For Cognition (18 items)
16. Private Self-consciousness scale (8 items)
17. SICQ (10 items)
18. IPIP (50 items)

The test battery also contained a number of measures designed for the current study which assessed participants' implicit self-theories about each Big Five dimension, their personality as a whole, and their reasoning abilities. These measures were developed with reference to implicit theories and empirical research undertaken by Dweck and her colleagues (e.g., Benenson & Dweck, 1986; Dweck, Chiu, & Hong, 1995a, 1995b; Dweck & Leggett, 1988; Hong, Chiu, Dweck, & Sacks, 1997). These measures are not discussed here because the scales did not share variance with either confidence or bias scores for personality or cognition, and because word limit constraints prevented their inclusion.

5.13.4 Scoring

For the gender variable, 1 = males and 2 = females. For the accuracy variables 0 = inaccurate and 1 = accurate. Big Five accuracy scores at the block level were calculated in the same way as for Study 1. Big Five accuracy scores at the item-level were calculated using the three accuracy methods outlined in Study 2. As was the case in Study 2, the accuracy scores derived from the three methodologies were factor analysed (see results section) before mean accuracy scores were computed for each Big Five dimension. Item-level bias scores for the cognitive and personality judgments were calculated in the same way as described in Studies 1 and 2. This PTPE bias score for each cognitive test required that the actual mean percentage of correct responses be subtracted from the estimated percentage of correctly solved items.

5.14 Results

5.14.1 Normality and Outliers

Prior to statistical analyses, all variables were examined through various subprograms from the Statistical Package for Social Scientists (SPSS, Version 16) for accuracy of data input, missing data, and fit between their distributions and the assumptions of multivariate analysis (Tabachnick & Fidell, 1996). There were some problems with missing data for the last five trials of the LST task. These trials were removed prior to data analyses and reliability of the measure improved. There were no problems in terms of missing data for the self-report variables.

Univariate and multivariate outlier checks revealed the presence of several univariate and multivariate outliers. Multivariate analyses were conducted with and without these outliers. As there were no noticeable differences in the outcome of the analyses these outliers were retained in accordance with recommendations made by Tabachnick and Fidell (1996).

Several self-report and objective variables displayed problems with skewness and/or kurtosis. Various transformation were applied to normalise the data in accordance with recommendations made by Tabachnick and Fidell (1996). Statistical analyses were performed with both the transformed and untransformed data. As the transformed data did not alter the outcome of these analyses, the untransformed data were retained.

5.15 Results for Cognitive Tasks

Means and standard deviations of the Gf dependent variables are presented in Table 5.1. As can be seen, item-level bias scores for the cognitive tasks ranged from 8.86 for the LST task to 23.79 for the WAT task. Participants found the WAT task to be the most difficult, correctly solving approximately 28% of items. The WAT task displayed lower internal consistency but was still considered acceptable for use in experimental research (Gregory, 1996). Reliability for the confidence variables was high and similar, to reports by other calibration researchers (e.g., Kleitman, 2003).

5.15.1 Bias in Relation to Gf Tasks

No hypotheses re bias scores were made for the LST, EAT, and WAT tasks. Nevertheless bias scores, calibration curves, and item-specific scatterplots were examined. The bias score for the Letter Series task was 8.86 indicating overconfidence. Figure 5.1 presents the calibration curve for this task. Visual inspection of the relative frequencies shows that the majority of observations were situated close to the perfect calibration line with approximately 23% of the ratings in the underconfidence region. The item-specific scatter plot is presented in Figure 5.2 and most of the items were close to the perfect calibration line.

Table 5.1
Descriptive Statistics for Cognitive Variables (N=243)

Dependent Variables	M	SD	# items	α
Cognitive Accuracy				
LSTAC	79.80	19.85	10	.86
WATAC	30.70	21.08	10	.64
ESTAC	62.86	19.27	24	.71
Cognitive Confidence				
LSTC	88.66	15.63	10	.87
WATC	54.48	25.95	10	.87
ESTC	78.88	14.88	24	.92
Item-Level Cognitive Bias				
LSTBIAS	8.86	19.62	-	-
WATBIAS	23.79	25.74	-	-
ESTBIAS	16.02	20.25	-	-
Cognitive Post-test Performance Percentage Correct Estimate				
LSTGE	73.76	20.17	1	-
WATGE	35.12	25.24	1	-
ESTGE	61.44	20.18	1	-
Cognitive Post-test Performance Percentage Correct Estimate Confidence Rating				
LSTGC	76.28	21.21	1	-
WATGC	64.50	30.50	1	-
ESTGC	69.71	22.42	1	-
PTPE Bias				
LSTGBIAS	1.97	23.61	1	-
WATGBIAS	4.30	24.20		
ESTGBIAS	-1.42	22.83		

Note. LST = Letter Series; WAT = Word Association; EAT = Esoteric Analogies.

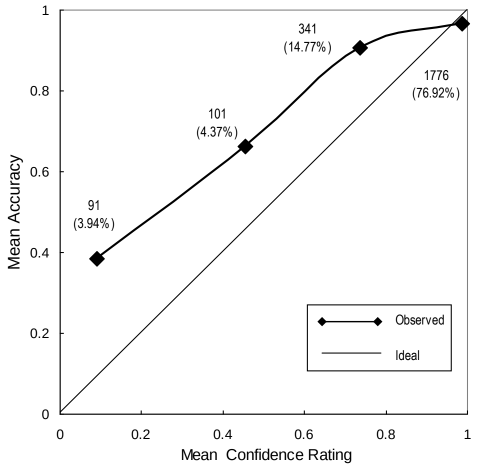


Figure 5.1. Calibration Curve of mean confidence rating and mean accuracy scores for the Letter Series Task.

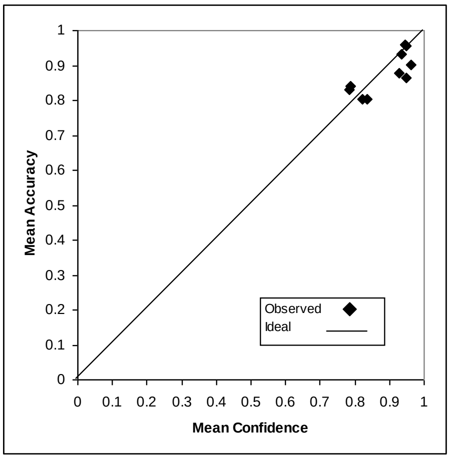


Figure 5.2. Scatterplot of mean confidence rating and mean accuracy scores for the Letter Series Task.

The bias score for the Esoteric Analogies task was 16.02 indicating marked overconfidence. Figure 5.3 presents the relevant calibration curve. Observation of relative frequencies shows that over 50% of the observations were in the overconfidence region, meaning that participants gave themselves a rating of 100% confident despite their accuracy being around 75%. The item-specific scatter plot is presented in Figure 5.4 and numerous items were in the overconfidence region.

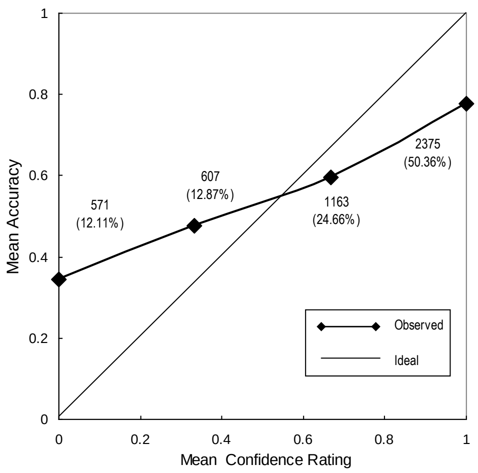


Figure 5.3. Calibration curve for the Esoteric Analogies Task.

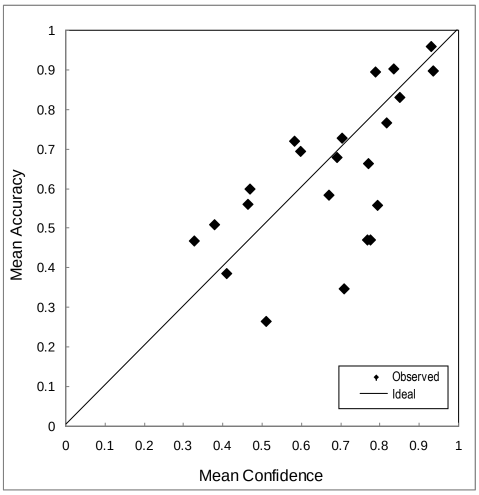


Figure 5.4. Scatterplot of mean confidence rating and mean accuracy scores for the Esoteric Analogies Task.

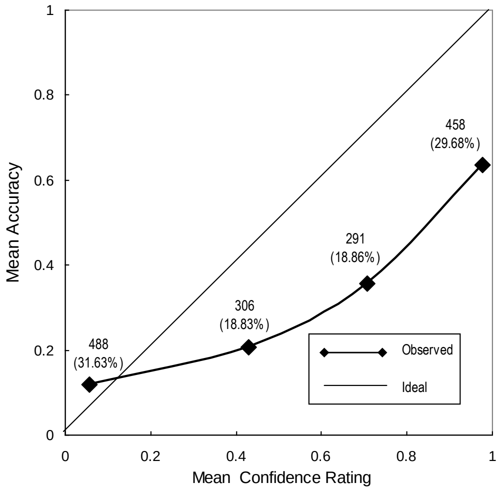


Figure 5.5. Calibration curve for the Word Association task.

At 23.79, the bias score for the Word Association task indicated marked overconfidence. The calibration curve for the Word Association task is presented in Figure 5.5 and the relative frequencies show that approximately 67% of the observations were in the overconfidence region. The scatterplot of item means for the Word Association task is presented in Figure 5.6 and shows that, in this case, several items at the lower levels of accuracy were responsible for the overconfidence effect.

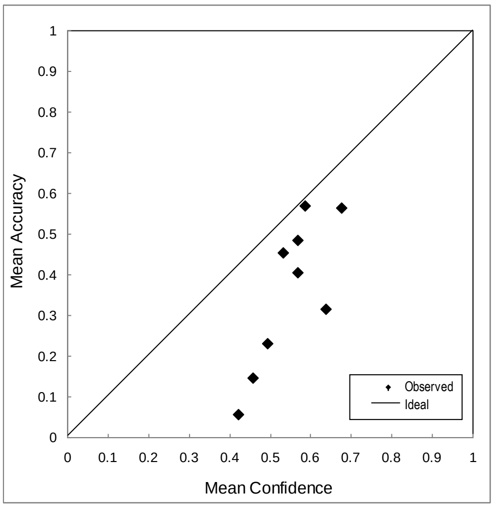


Figure 5.6. Scatterplot of mean confidence rating and mean accuracy scores for the Word Association Task.

Hypothesis 1 stated that individuals would be better calibrated for the PTPE bias scores than for the item-level bias scores. The PTPE bias scores for the Letter Series, Esoteric Analogies, and Word Association tasks were 1.97, -1.42, and 4.30 respectively, scores that indicated good calibration, as expected.

5.15.2 Individual Differences for Gf Tasks

To simplify results for some individual differences analyses using Gf tasks, (including those undertaken later on with the Big Five variables), composite variables were formed for Gf accuracy, Gf bias, and Gf confidence. All composite variables entailed calculating an average score across the three Gf tasks.

The outcome that age was not correlated with the Gf bias composite score ($r = .07$, $p > .05$), was contrary to the premise of hypothesis 2, which assumed a positive association between age and Gf bias.

The next hypothesis reasoned that males would be significantly more confident than females on Gf tasks. No hypothesis was made in terms of gender differences in item-level bias; nevertheless gender differences in item-level bias scores were examined. Independent samples *t* tests were used to test the hypothesis and results appear in Table

5.2 [4]. To control for family-wise error, a bonferroni adjustment was made with the alpha level being set at 0.01. Another unexpected outcome occurred. Males and females did not differ in terms of item level confidence for Gf tasks; nor were there gender differences in mis-calibration. Overall these results indicated that males and females did not differ in terms of cognitive confidence or mis-calibration.

Table 5.2
Means of Confidence and Bias Scores for Males (n = 61) and Females (n = 182) for Cognitive Tasks in Study 3.

	LSTC	LST BIAS	ESTC	EST BIAS	WATC	WAT BIAS
Male	87.80	11.05	81.34	17.38	57.65	28.47
Female	88.94	8.12	78.05	15.56	53.42	22.22
t tests	-.49	1.01	1.50	.61	1.10	1.66
Effect size	-0.06	0.15	0.23	0.09	0.05	0.24

Note. LST = Letter Series; WAT = Word Association; EAT = Esoteric Analogies

5.15.2.1 Ability Differences in Relation to Gf Bias Scores

Considering hypothesis 4, low scorers (i.e., in the bottom quartile) were expected to be more mis-calibrated than high scorers (i.e., top quartile) for the three Gf tasks used in this study. Based on their accuracy scores for each cognitive task, participants were divided into quartiles. This data was subjected to an Independent samples t-test with the results presented in Table 5.3. Once again, to control for family-wise error a bonferroni adjustment was made, with the alpha level set at 0.01.

The hypothesis received support for all of the Gf tasks indicating that low scorers were significantly more mis-calibrated than high scorers were.

---

[4] The male to female ratio in this study is not ideal, however, it closely resembles other calibration research wherein similar gender ratios were reported (e.g., Pallier, 2003). Also, based on the recommendations made by Howell (2002), each effect size calculation used the mean and the standard deviation for each gender as the denominator. This practice guarantees that d is approximated independently of N, thereby removing potential concerns regarding unequal sample sizes.

Table 5.3
Mean Bias Scores for First and Fourth Quartiles on Cognitive Tasks (N = 243)

	LSTBIAS	ESTBIAS	WATBIAS
Quartile 1	24.49	33.93	34.18
Quartile 4	1.86	-.53	10.68
t tests	5.48**	10.48**	6.62**

Note. lstbias = letter series bias; estbias = esoteric analogies bias; watbias = word association bias
** $p = .00$.

5.15.2.2 Personality in Relation to Gf Confidence and Bias

Hypotheses 5.1 and 5.3 stated that, after controlling for the influence of the other four personality dimensions, there would be significant positive associations between Gf confidence and both Intellect and Conscientiousness respectively. Hypothesis 5.2 dealt with Extraversion, and proposed a positive association between Gf confidence and bias. Partial correlations were used to test all three hypotheses (see Table 5.4). Zero order correlations are presented in the top portion of this table and partial correlations appear below these. The first two expectations were met: Intellect and Gf confidence were positively correlated, and when partial correlations were examined, Conscientiousness and Gf confidence were also positively associated.

The use of these partial correlations also revealed an outcome that was unexpected; that is, an association between Intellect and Gf bias. Also, anticipated were the findings that Extraversion was not correlated with either Gf confidence or Gf bias.

Overall, for Study 3, there was no difference between conclusions drawn when zero-order correlations were used, and those drawn when partial correlations were used, with two qualifications. To re-state these, when partial correlations were used to test the hypotheses, (a) Intellect was correlated with Gf bias and, (b) the significant zero-order correlation between Emotional Stability and Gf confidence became insignificant.

Table 5.4
Correlations among IPIP subscale scores and GF Confidence and Gf Bias and Partial Correlations between Big Five Scores and Confidence and Bias (N = 243)

	IEXTRA	IAGREE	ICON	IEMOT	IINTELL [a]	Gfcon	Gfbias
IEXTRA	1.00						
IAGREE	0.17*	1.00					
ICON	0.10	0.10	1.00				
IEMOT	0.24**	0.12	0.36**	1.00			
IINTELL	0.32**	0.21**	0.12	0.23**	1.00		
Gfcon	0.04	0.03	0.21**	0.21**	0.29**	1.00	
Gfbias	0.10	-0.10	0.18**	0.11	0.04	0.59**	1.00
Partial Correlations							
	IEXTRA[1]	IAGREE[1]	ICON[1]	IEMOT[1]	IINTELL[2]		
Gfcon	-0.09	-0.04	0.15*	0.11	0.18**		
Gfbias	0.09	-0.14*	0.17*	0.04	0.17**		

Note: Gfcon= Gf confidence composite variable; Gfbias = Gf bias composite variable; [a] = controlling for accuracy; 1 = controlling for the other four personality dimensions; [2] = controlling for the other four personality variables and accuracy.
* $p < .05$. ** $p < .01$.

5.15.2.3 Need for Cognition in Relation to Gf Tasks

Hypothesis 6, that NFC would be positively associated with Gf confidence score was supported by the data. A significant correlation emerged between NFC and the Gf composite confidence variable ($r_{(243)} = .28$, $p < .01$) suggesting that as NFC increased, so did Gf confidence.

5.15.2.4 Negative Affect in Relation to Gf Tasks

No formal hypothesis was formed in relation to the potential association between NA and Gf confidence. NA was negatively associated with the Gf composite confidence variable ($r_{(243)} = -.19$, $p <$ 01) suggesting that as NA increased Gf confidence decreased.

5.15.3 Factorial Structure of the IPIP

The factorial structure of the IPIP was examined using Principal Components Analysis with Promax rotation and Kaiser normalisation. A solution employing root one criterion produced 12 factors. Cattell's Scree Plot, however, provided support for interpreting a five factor solution. The pattern matrix, percent of variance accounted for, eigenvalues, and factor correlation matrix are presented in Table 5.5. The five factor solution accounted for 50.11% of the total variance. As anticipated, all the factors comprised loadings from the expected IPIP items for each dimension. The five factors were labelled Emotional Stability, Extraversion, Agreeableness, Conscientiousness, and Intellect respectively. Although a couple of complex variables were noted, they did not cause concern because the overall structure was consistent with expectations. The correlations among the factors were concordant with previous research (Saucier, 2002; Saucier & Goldberg, 2002).

Table 5.5

Pattern Matrix of IPIP using Principal Components Analysis with Promax rotation and Kaiser Normalisation (N = 243)

Variable	F_1 [a]	F_2	F_3	F_4	F_5	h^2
PIP29	0.83	0.02	0.04	0.07	-0.15	.71
IPIP44	0.83	0.06	-0.14	0.05	0.04	.67
IPIP39	0.81	0.04	-0.02	-0.07	0.05	.67
IPIP04	0.81	-0.02	0.00	0.12	-0.06	.63
IPIP49	0.76	-0.15	0.04	-0.07	0.02	.68
IPIP14	0.74	-0.08	0.18	0.15	-0.03	.56
IPIP34	0.73	0.12	-0.05	-0.07	0.03	.55
IPIP24	0.66	0.11	0.09	-0.09	-0.08	.48
IPIP09	-0.60	0.19	0.09	-0.16	0.04	.45
IPIP19	-0.53	0.15	-0.05	0.21	0.01	.45
IPIP01	0.06	0.85	-0.09	0.01	-0.03	.68
IPIP31	0.01	0.81	0.06	-0.03	0.05	.71
IPIP41	0.12	0.79	-0.13	-0.01	0.08	.63
IPIP36	0.03	-0.76	0.28	0.04	0.05	.59
IPIP16	0.12	-0.75	0.12	-0.04	-0.03	.62
IPIP21	-0.04	0.73	0.19	-0.02	0.06	.65
IPIP06	-0.13	-0.71	-0.14	-0.10	0.10	.50
IPIP46	0.04	-0.69	-0.04	-0.03	0.09	.47
IPIP11	-0.13	0.69	0.25	-0.03	-0.05	.62
IPIP26	0.05	-0.63	-0.13	-0.02	0.00	.46
IPIP17	0.05	-0.13	0.79	0.08	0.01	.62
IPIP42	0.19	-0.02	0.77	0.02	0.07	.63
IPIP37	0.02	0.04	0.75	0.06	-0.11	.55
IPIP07	-0.03	0.14	0.68	-0.02	0.09	.56
IPIP27	0.09	-0.06	0.64	0.02	-0.17	.41
IPIP32	0.07	-0.15	-0.60	0.11	-0.03	.43
IPIP22	0.07	-0.06	-0.60	0.03	-0.01	.38
IPIP47	-0.09	0.31	0.51	-0.04	0.02	.43
IPIP02	0.14	0.02	-0.50	0.12	0.14	.25
IPIP12	0.42	0.30	-0.45	0.00	0.09	.40
IPIP45	0.21	-0.15	0.43	-0.02	0.29	.31
IPIP33	0.10	-0.11	-0.03	0.76	0.12	.57
IPIP23	0.08	0.10	-0.04	0.72	-0.09	.50
IPIP43	0.00	0.04	0.00	0.70	-0.07	.49
IPIP28	0.13	-0.09	0.05	-0.67	0.18	.53
IPIP03	0.05	0.07	-0.10	0.65	0.21	.47
IPIP08	0.10	0.01	0.12	-0.63	0.17	.46
IPIP38	0.25	0.19	-0.20	-0.52	0.03	.47

Variable	F1 [a]	F2	F3	F4	F5	h^2
IPIP48	0.17	-0.01	0.13	0.51	0.21	.31
IPIP18	0.46	-0.01	0.04	-0.47	-0.06	.58
IPIP13	0.11	0.09	0.18	0.34	0.22	.24
IPIP50	-0.02	0.14	0.04	-0.01	0.65	.53
IPIP25	0.02	0.16	-0.03	0.01	0.64	.49
IPIP10	0.20	0.08	0.21	0.04	-0.63	.43
IPIP20	0.18	0.23	0.05	0.12	-0.61	.37
IPIP40	0.13	-0.09	-0.14	0.00	0.60	.32
IPIP15	0.10	0.06	0.12	-0.03	0.59	.41
IPIP05	-0.02	-0.03	0.00	0.09	0.58	.34
IPIP30	0.13	-0.01	-0.16	0.05	-0.56	.43
IPIP35	-0.09	0.04	-0.02	0.19	0.50	.35
Eigenvalue	9.13	5.58	4.42	3.17	2.76	
% of Variance	18.25	11.16	8.84	6.34	5.51	

Variable	F1	F2	F3	F4	F5
F1	1.00				
F2	-0.20	1.00			
F3	-0.06	0.14	1.00		
F4	-0.30	0.05	0.08	1.00	
F5	-0.19	0.34	0.21	0.07	1.00

Note. a Factor Labels, F1 = IPIP Emotional Stability ; F2 = IPIP Extraversion; F3 = IPIP Agreeableness; F4 = IPIP Conscientiousness; F5 = IPIP Intellect

Because the correlations among the IPIP subscales closely resembled those reported in the literature (Saucier, 2002; Saucier & Goldberg, 2002), the correlation matrix was not presented.

5.15.4 Factorial Structure of the IPIP Form B

The factorial structure of the IPIP Form B was also examined using Principal Components Analysis with Promax rotation and Kaiser normalisation. A solution employing root one criterion produced seven factors. Cattell's Scree Plot, however, was indeterminate after five factors thus supporting interpretation of a five-factor solution. The pattern matrix, percent of variance accounted for, eigenvalues, communalities and factor correlation matrix are presented in Table 5.6. The five-factor solution accounted for 58.75% of the total variance. As expected, all the factors consisted of loadings from the expected IPIP Form B items for each dimension. The five factors were labelled IPIP Agreeableness Form B, IPIP Conscientiousness Form B; IPIP Extraversion Form B; IPIP Emotional Stability Form B; IPIP Intellect Form B respectively. As shown in the bottom portion of the table, there were low to moderate correlations among the IPIP Form B subscales.

Table 5.6

Pattern Matrix of IPIP Form B using Principal Components Analysis with Promax rotation and Kaiser Normalisation (*N = 243)*

Variable	F_1[a]	F_2	F_3	F_4	F_5	h^2
pda010	0.89	-0.02	-0.01	-0.19	0.00	.72
pda08	0.86	0.00	0.00	-0.10	-0.05	.67
pda01	0.83	-0.06	0.01	-0.02	0.09	.70
pda06	0.77	0.04	-0.06	0.04	0.10	.65
pda07	0.76	0.24	-0.06	-0.02	-0.17	.62
pda09	0.74	0.00	0.14	-0.12	0.12	.67
pda02	0.70	-0.01	0.12	-0.01	0.07	.61
pda05	0.62	0.05	-0.02	0.30	-0.13	.55
pda04	0.52	-0.04	0.14	0.25	0.02	.51
pda03	0.44	0.02	-0.03	0.03	0.01	.20
pdc04	-0.05	0.87	-0.03	-0.06	-0.06	.65
pdc09	-0.01	0.86	0.07	-0.04	-0.17	.65
pdc01	-0.02	0.83	0.06	0.00	0.04	.70
pdc03	-0.01	0.81	0.04	-0.01	0.00	.66
pdc02	0.06	0.76	-0.07	-0.11	0.13	.61
pdc05	0.13	0.74	-0.06	0.11	-0.06	.63
pdc07	0.07	0.74	0.04	0.04	0.05	.66
pdc06	0.01	0.72	0.00	0.15	-0.11	.56
pdc08	0.24	0.45	-0.17	0.00	0.33	.51
pde07	-0.10	-0.06	0.88	0.07	-0.05	.70
pde06	0.09	0.03	0.86	-0.02	-0.06	.77
pde01	-0.20	0.03	0.86	-0.05	0.04	.66
pde08	0.20	-0.07	0.82	-0.07	-0.05	.73
pde04	-0.01	-0.04	0.81	-0.06	0.09	.68
pde05	0.16	0.05	0.77	0.14	-0.12	.76
pde02	0.23	-0.14	0.74	0.13	-0.07	.72
pde03	-0.14	0.27	0.53	-0.19	0.28	.53
pdi01	0.02	0.02	0.43	-0.09	0.38	.46
pdem01	-0.01	-0.12	-0.03	0.87	0.00	.67
pdem03	0.04	0.01	-0.08	0.87	-0.01	.73
pdem04	0.16	-0.13	-0.14	0.85	0.00	.68
pdem02	0.08	-0.08	-0.15	0.81	0.04	.60
pdem05	-0.12	0.10	0.01	0.72	0.00	.55
pdem07	-0.15	0.03	0.18	0.69	0.09	.62
pdem06	-0.19	0.13	0.10	0.60	0.08	.47
pdem09	0.02	0.11	0.12	0.57	0.04	.49
pdem08	-0.12	0.21	0.20	0.56	-0.02	.53
pdi08	-0.04	-0.19	-0.02	-0.06	0.76	.46
pdi07	0.05	-0.03	-0.08	0.00	0.75	.52
pdi05	-0.02	0.11	-0.10	0.04	0.68	.50
pdi010	-0.01	0.11	-0.05	-0.01	0.67	.49
pdi06	-0.09	0.13	0.07	0.08	0.64	.55
pdi03	0.05	0.07	-0.05	0.08	0.63	.48
pdi09	0.15	-0.19	-0.04	0.14	0.61	.41
pdi02	-0.06	-0.04	0.17	0.04	0.57	.42
pdi04	0.07	-0.10	0.15	0.01	0.54	.37
Eigenvalue	13.32	4.39	3.73	3.37	2.23	
% of Variance	28.95	9.53	8.10	7.33	4.84	
Factor Correlation	1	2	3	4	5	
1	1.00					
2	0.30	1.00				
3	0.38	0.27	1.00			

Variable	F_1[a]	F_2	F_3	F_4	F_5	h^2
4	0.31	0.40	0.37	1.00		
5	0.30	0.43	0.45	0.30	1.00	

Note. [a] Factor Labels, F_1 = IPIP Agreeableness Form B ; F_2 = IPIP Conscientiousness Form B; F_3 = IPIP Extraversion Form B; F_4 = IPIP Emotional Stability Form B; F_5 = IPIP Intellect Form B

The correlations among the factors for the IPIP Form B were approximately .30 higher than those reported for IPIP Form A. One possible explanation for this is that the IPIP Form B differed from the IPIP from A in that participants were asked to provide confidence ratings in relation to each item endorsement. Asking participants to provide this confidence rating may have affected the orthogonality of the factors. Non-orthogonality of the IPIP factors, however, is not new and it has been argued by Saucier, (2002) to represent “not the Big Five factors themselves, but rather scale construction procedures used by the developers of measures” (p. 28).

5.15.5 Descriptive Statistics and Factorial Structure of Item-level Accuracy Scores

The three accuracy methods that were developed in Study 2 for the purpose of calculating a mean accuracy score for each Big Five dimension were again used in Study 3. Table 5.8 presents the descriptive statistics for these scores. The most salient feature being that the means for methods 2 and 3 are quite similar for each of the dimensions except for the Intellect dimension. The correlations among these accuracy scores are presented in Appendix U. Because the three methodologies were similar, all of the accuracy scores for each Big Five dimension were correlated. It is worth noting for each Big dimension, the highest correlations were between methods 2 and 3 with two exceptions. That is for both the Conscientiousness and Emotional Stability dimensions, the correlations between methods 1 and 2 were the highest. Overall, the correlations between the accuracy scores for each dimension, and the means for each dimension were similar to those reported in Study 2.

Table 5.7
Descriptive Statistics for Accuracy Scores Derived from Methods 1, 2, and 3 (N=243)

Variable	*M*	*SD*
Intellect Accuracy Method 1	81.48	38.92
Intellect Accuracy Method 2	80.66	39.58
Intellect Accuracy Method 3	88.89	31.49
Conscientiousness Accuracy Method 1	85.60	35.18
Conscientiousness Accuracy Method 2	88.48	32.00
Conscientiousness Accuracy Method 3	86.83	33.88
Extraversion Accuracy Method 1	70.37	45.76
Extraversion Accuracy Method 2	88.89	31.49
Extraversion Accuracy Method 3	83.13	37.53
Agreeableness Accuracy Method 1	81.89	38.59
Agreeableness Accuracy Method 2	79.01	40.81
Agreeableness Accuracy Method 3	80.66	39.58
Emotional Stability Accuracy Method 1	78.60	41.10
Emotional Stability Accuracy Method 2	84.77	36.00
Emotional Stability Accuracy Method 3	79.42	40.51

As was done in Study 2, prior to calculating mean accuracy scores for each Big Five dimension, the accuracy scores in Table 5.7 were factor analysed. A five-factor solution emerged with the three accuracy scores for each dimension loading on its expected factor. The results from this structural analysis suggested that accuracy was specific to each Big Five domain thereby replicating Study 2 results. Therefore, calculating mean accuracy scores for each dimension was again justified. This factor analysis is presented in Appendix U. Please note for the remainder of this results section, item-level accuracy scores refer to the mean accuracy scores for each Big Five dimension.

5.15.6 Reliability Analysis for Self-Report Measures

Internal consistency reliability estimates (i.e., Cronbach's coefficient alphas) were calculated for all self-report variables where applicable. These alpha coefficients are presented in Table 5.8 along with descriptive statistics for self-report dependent variables. All alpha coefficients indicated good internal consistency. The internal consistency coefficients reported for IPIP and IPIP Form B were consistent with those reported in Study 2. Noteworthy was that all self-report confidence ratings also displayed high internal consistency. The alpha co-efficient for the modified self-report intelligence scale (.79) was higher than that of the original scale SRIQ scale (.73) developed by Gignac et al. (2004).

5.15.7 Descriptive Statistics Self-Report Measures

Mean ratings for each of the Big Five dimensions on the original IPIP administered via the internet ranged from 31.58 for Extraversion to 40.79 for Agreeableness. These mean ratings closely resembled those reported by the participants in Study 1, in which the test was administered face-to-face, and closely resembled those reported in Study 2. This similarity to the first two studies, continued for the block personality ratings, as well for the overall pattern of confidence rating endorsements. Their confidence levels also remained high for self-ratings of intelligence.

Table 5.8
Descriptive Statistics for Self-Report Dependent Variables (N = 243)

Dependent Variables	M	SD	# items	α
IPIP Original Subscales				
ICON	34.27	6.85	10	.83
IEMOT	32.13	8.71	10	.91
IINTELL	36.45	5.58	10	.78
IEXTRA	31.58	8.79	10	.91
IAGREE	40.97	5.80	10	.83
IPIP Form A				
ICONA	31.04	5.97	9	.80
IEMOTA	29.52	7.93	9	.89
IINTELLA	36.45	5.58	10	.78
IEXTRAA	24.40	7.39	8	.91
IAGREEA	40.97	5.80	10	.83
Personality Confidence Item-by-Item For Each Personality Dimension From IPIP Form B				
CONCIC	83.51	12.91	9	.95
EMOTIC	83.60	12.44	9	.94
INTELLIC	82.42	12.41	10	.92
EXTRAIC	82.91	13.98	8	.95
AGREEIC	86.70	12.50	10	.95
IPIP Form B-Item-by-Item Big Five Adjective Ratings Recoded[1]				
CONCIR	31.35	7.59	9	.92
EMOTIR	29.79	7.65	9	.90
INTELLIR	38.17	6.18	10	.84
EXTRAIR	27.83	7.08	8	.92
AGREEIR	42.03	6.27	10	.90
Big Five Block Confidence Ratings (BFBDC)				
CONCB	79.88	16.87	1	-
EMOTCB	82.72	15.51	1	-
INTELLCB	81.73	14.61	1	-
EXTRACB	83.50	16.38	1	-

AGREECB	86.26	14.18	1	-
Big Five Block Adjective Ratings (BFBD) [1]				
CONCR	3.51	1.17	1	-
EMOTCR	3.39	1.20	1	-
INTELLCR	3.70	1.00	1	-
EXTRACR	3.34	1.21	1	-
AGREECR	4.20	.90	1	-
Big Five Block Accuracy Scores (BFBD Accuracy 20%)				
CONCBA20%	74.49	43.68		-
EMOTBA20%	71.19	45.38		-
INTELLBA20%	79.01	40.81		-
EXTRABA20%	76.13	42.72		-
AGREEBA20%	84.36	36.40		-
Big Five Item-Level Accuracy Scores Derived from Methods 1,2, & 3				
Conscientiousness Accuracy	86.97	31.19	3	.92
Emotional Stability Accuracy	80.93	37.57	3	.95
Intellect Accuracy	83.67	32.12	3	.84
Extraversion Accuracy	80.80	29.64	3	.65
Agreeableness Accuracy	80.52	36.93	3	.92
Big Five Bias Scores				
Conscientiousness Bias	-3.46	34.97		
Emotional Stability Bias	2.67	40.60		
Intellect Bias	-1.25	35.63		
Extraversion Bias	2.12	32.77		
Agreeableness Bias	6.19	37.76		
Other Self-Report Measures				
PrSC	24.49	7.65	8	.84
PA	35.69	6.38	10	.86
NA	20.86	8.38	10	.92
NFC	9.33	11.33	18	.89
Intelligence Self-report confidence Ratings	82.05	12.24	10	.92
Intelligence Self-report Ratings	15.08	13.42	10	.79

Note. ICON = IPIP Conscientiousness; IEMOT = IPIP Emotional stability; IINTELL = IPIP Intellect; IEXTRA = IPIP Extraversion, IAGREE = IPIP Agreeableness; ICONA = IPIP Form A Conscientiousness; IEMOTA = IPIP Form A Emotional stability; INTELLA = IPIP Form A Intellect; IEXTRAA = IPIP Form A Extraversion, IAGREEA = IPIP Form A Agreeableness; CONIC= Conscientiousness Item-by-item confidence Form B; EMOTIC= Emotional stability Item-by-item confidence Form B; INTELLIC = Intellect Item-by-item confidence Form B; EXTRAIC = Extraversion Item-by-item confidence Form B; AGREEIC= Agreeableness Item-by-item confidence Form B; CONCIR= Conscientiousness Item-by-item adjective rating Form B; EMOTIIR= Emotional stability Item-by-item adjective rating Form B; INTELLIR = Intellect Item-by-item adjective rating Form B; EXTRAIR = Extraversion Item-by-item adjective rating Form B; AGREEIR= Agreeableness Item-by-item adjective rating Form B; CONCB= Conscientiousness Block confidence rating; EMOTCB= Emotional stability Block confidence rating; INTELLCB = Intellect Block confidence rating, EXTRACB = Extraversion Block confidence rating; AGREECB= Agreeableness Block confidence rating; CONCR = Conscientiousness Block rating; EMOTCR= Emotional stability Block rating; INTELLCR = Intellect Block rating; EXTRACR = Extraversion Block rating; AGREECR= Agreeableness Block rating; [1] = These means were calculated after recoding the data from -5 to +5 to 1 to 5 (see method section); CONSCBA20% = Conscientiousness Accuracy ≤ 20 %; EXTRABA20% = Extraversion accuracy ≤ 20 %; AGREEBA20% = Agreeableness Accuracy ≤ 20 %; EMOTBA20% = Emotional Stability Accuracy ≤ 20 %; INTBA20% = Intellect Accuracy ≤ 20 %; PrSC = Private Self-Consciousness; PA = Positive Affect; NA = Negative Affect; NFC = Need For Cognition.

5.16 The Factorial Structure of Big Five Confidence

According to the seventh hypothesis, one confidence factor was expected from the structural analysis of the item-by-item and block confidence ratings. Exploratory factor analytic techniques investigated this hypothesis, and all Big Five confidence scores underwent principal Axis Factoring with oblique rotation. The data from Table 5.9, which presents the correlation matrix of psychometric variables, were considered factorable as all assumptions as advocated by Coakes and Steed (1996) were met.

A solution employing root one criterion produced one factor, which accounted for 67.48% of the total variance. The percent of variance accounted for, eigenvalue, and communalities are presented in Table 5.10. The factor extracted from the data set was called Big Five Confidence as all of the personality confidence scores loaded highly on it. This finding infers that the same cognitive processes underlie all personality confidence judgments, and has therefore successfully replicated the results from Study 2.

Table 5.9
Correlations among Block and Item-by-Item Confidence Ratings (N = 243)

Variable	1	2	3	4	5	6	7	8	9	10
INTELLIC	1									
EMOTIC	0.74	1.00								
CONCIC	0.74	0.87	1.00							
EXTRAIC	0.72	0.80	0.84	1.00						
AGREEIC	0.68	0.76	0.80	0.80	1.00					
AGREECB	0.56	0.70	0.70	0.63	0.74	1.00				
EMOTCB	0.57	0.74	0.71	0.74	0.67	0.55	1.00			
INTELLCB	0.62	0.63	0.65	0.67	0.60	0.51	0.56	1.00		
CONCB	0.66	0.62	0.68	0.51	0.55	0.45	0.49	0.41	1.00	
EXTRACB	0.61	0.59	0.56	0.63	0.52	0.39	0.52	0.49	0.47	1.00

Note. All correlations were significant at the .01 level.

Table 5.10
Principal Axis Factoring of Big Five Item-by-Item and Block Confidence Rating Scores (N = 243)

Variable	Communalities	Factor –Big Five Confidence
AGREECB	.53	.73
CONCB	.44	.66
EXTRACB	.43	.65
EMOTCB	.60	.77
INTELLCB	.51	.71
EMOTIC	.84	.92
CONCIC	.87	.93
EXTRAIC	.81	.90
INTELLIC	.67	.82
AGREEIC	.74	.86
Eigenvalue	-	6.75
% of variance	-	67.49

5.16.1 Gender and Age Differences in Big Five Confidence, Accuracy and Bias

For all gender differences analyses in this section, independent samples t-tests were used. Hypothesis eight surmised that males would be just as accurate as females when making both block level and item-by-item personality judgments about themselves, and as Table 5.11 [5] shows, this hypothesis was upheld. Results from Study 3 have again successfully replicated those from Study 2.

Table 5.11
Means for Block and Item-by-Item Big Five Accuracy Scores for Males (n = 61) and Females (n = 182) in Study 3.

	AGREEB A20%	CONCB A20%	EXTRAB A20%	EMOTB A20%	INTELLB A20%	EXTRA AC	AGREE AC	CON SCACC	EMOT AC	INT AC
Male	85.25	77.05	83.61	80.33	78.69	80.87	79.24	88.52	89.07	79.24
Female	84.07	73.63	73.63	68.13	79.12	80.76	80.95	86.45	78.21	85.16
t tests	.22	.53	1.72	1.97	-.07	0.02	-0.31	0.45	2.29	-1.25
Effect size	0.03	0.07	0.24	0.28	-0.01	0.00	-0.04	0.06	0.31	-0.18

Hypotheses nine speculated that males would be just as confident as females when making both block and item-by-item personality judgments about themselves. Results appear in Table 5.12. Bonferroni adjustment was used to keep family-wise error at alpha = .05.

Table 5.12
Means for Block and Item-by-Item Big Five Confidence Ratings for Males (N = 61) and Females (N =182) in Study 3.

	AGREE CB	CON CB	EXTRA CB	EMOT CB	INTELL CB	EXTRA IC	AGREE IC	CONC IC	EMOT IC	INTELL IC
Male	84.59	79.84	82.13	83.77	83.11	82.60	85.81	81.95	83.50	85.04
Female	86.81	79.89	83.69	82.36	81.26	83.02	87.01	84.03	83.64	81.55
t tests	-1.06	-.02	-.75	.61	.88	-.20	-.65	-1.09	-.08	1.92
Effect size	-0.15	-0.003	-0.11	0.09	0.13	-0.03	-0.09	-0.15	-0.01	0.30

As no gender differences arose across either the block or the item-by-item ratings, it appears that males were just as confident as females when making judgments

[5] The male to female ratio in this study is not ideal, however, it closely resembles other calibration research wherein similar gender ratios were reported (e.g., Pallier, 2003). Also, based on the recommendations made by Howell (2002), each effect size calculation used the mean and the standard deviation for each gender as the denominator. This practice guarantees that d is approximated independently of N, thereby removing potential concerns regarding unequal sample sizes.

about the Big Five personality dimensions. These findings essentially replicate the results from Studies 1 and 2.

Hypothesis 10 addresses gender differences in Big Five bias. The results are displayed in Table 5.13. Bonferroni adjustment was used to keep family-wise error at alpha = .05. In accordance with hypothesis 11, no gender differences were apparent for the Big Five item-level bias scores. This finding, that males and females did not differ in terms of Big Five bias, replicates results from Study 2.

Table 5.13
Means for Item-by-Item Big Five Bias Scores for Males (N = 61) and Females (N = 182) in Study 3.

	AGREE BIAS	CON BIAS	EXTRA BIAS	EMOT BIAS	INTELL BIAS
Male	6.57	-6.58	1.73	-5.57	5.81
Female	6.06	-2.42	2.25	5.43	-3.62
t tests	0.09	-0.80	-0.11	-2.07	1.80
Effect size	0.01	-0.12	-0.02	-0.29	0.26

The assumptions of hypotheses 11, 12 and 13- that age would not be associated with Big Five confidence (either block or item-level), Big Five accuracy (either block or item-level), or Big Five item-level bias- were tested by calculating Pearson's Product Moment correlations. The outcomes largely met expectations. No significant correlations were found between age and any of the block or item-by-item personality confidence ratings, nor between age and Big Five bias scores. Age was also not significantly correlated with either item-level or block accuracy except for one significant correlation between age and block accuracy for the Emotional Stability dimension ($r_{(243)}$ = .13, p < .05). The effect size however, was small. Overall, the aforementioned findings replicate results from Studies 1 and 2.

Please note: Pearson's product moment correlations were computed to test all subsequent hypotheses unless otherwise stated.

5.16.1.1 Big Five Accuracy and Bias Scores

The descriptive statistics for the block description accuracy scores were shown in Table 5.8 . Scores for the Conscientiousness, Emotional Stability, Intellect, Extraversion and Agreeableness dimensions were 74.49, 71.19, 79.01, 76.13 and 84.36, respectively. The standard deviation of the block accuracy scores showed that there were many inaccurate ratings for some traits. It was noteworthy that observation of the descriptive statistics presented in Table 5.8 revealed a salient similarity to those reported in Studies 1 and 2, but with one qualification worth mentioning. The block accuracy score for the Extraversion dimension was much lower in Study 2.

Hypothesis 14 postulated that the associations among the Big Five block and item level accuracy scores would be low. The correlations are presented in Table 5.14 and Table 5.15. As expected, the correlations between the block and item-level accuracy scores were low suggesting that individuals who were consistent (accurate) in their scores on one trait were not necessarily accurate in their scores for other traits.

The block descriptions of personality underwent the same validity checks that were used in Study 1 (see section 3.11), and because the results also agreed with those from Study 1, they were not reported here.

Table 5.14
Correlations among Big Five Block Accuracy Scores (N = 243).

Variable	CONCBA20%	EXTRABA20%	AGREEBA20%	EMOTBA20%	INTELLBA20%
CONCBA20%	1.00				
EXTRABA20%	-0.02	1.00			
AGREEBA20%	0.14*	0.13*	1.00		
EMOTBA20%	0.13*	0.07	0.05	1.00	
INTELLBA20%	-0.02	0.21**	0.08	0.10	1.00

* $p < .05$. ** $p < .01$.

Table 5.15
Correlations among Big Five Item-Level Accuracy Scores (N = 243).

Variables	CONAC	EMOTAC	INTAC	EXTAC	AGAC
CONAC	1.00				
EMOTAC	0.14*	1.00			
INTAC	0.31**	0.10	1.00		
EXTAC	0.23**	-0.02	0.29**	1.00	
AGAC	0.05	0.13*	0.21**	0.24**	1.00

CONSCAC= Conscientiousness accuracy item level; EXTRAAC = Extraversion accuracy item level; AGREEAC = Agreeableness accuracy item level; EMOTAC = Emotional Stability accuracy item level; INTAC = Intellect accuracy item level .
* $p < .05$. ** $p < .01$.

Good calibration was expected for the item-level bias scores according to hypothesis 15, with these scores being previously presented in Table 5.8. As anticipated, these scores were -3.46, 2.67, -1.25, 2.12, and 6.19 for the Conscientiousness, Emotional Stability, Intellect, Extraversion and Agreeableness dimensions respectively. It is important to note the similarity between these scores and the item-level bias scores from Study 2.

Considering hypothesis 16, six factors were expected to emerge from the structural analysis of all subscales from the IPIP, and IPIP Form B. Principal Components Analysis with oblique rotation was undertaken to test this hypothesis, and the correlation matrix of these psychometric variables is presented in Table 5.16.

Table 5.16
Correlations among IPIP and IPIP Form B Subscale Scores

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
INTELLIR	1.00														
CONCIR	0.43**	1.00													
EXTRAIR	0.52**	0.27**	1.00												
AGREEIR	0.36**	0.38**	0.44**	1.00											
EMOTIR	0.38**	0.43**	0.39**	0.34**	1.00										
INTELLIC	0.36**	0.17**	0.27**	0.34**	0.13*	1.00									
CONCIC	0.22**	0.29**	0.23**	0.39**	0.13*	0.74**	1.00								
EXTRAIC	0.21**	0.18**	0.31**	0.35**	0.15*	0.72**	0.84**	1.00							
AGREEIC	0.19**	0.22**	0.27**	0.52**	0.10	0.68**	0.80**	0.80**	1.00						
EMOTIC	0.19**	0.15*	0.20**	0.33**	0.09	0.74**	0.87**	0.80**	0.76**	1.00					
IINTELL	0.66**	0.17**	0.32**	0.19**	0.16*	0.11	0.00	0.02	0.02	-0.02	1.00				
ICON	0.30**	0.80**	0.21**	0.27**	0.34**	0.12	0.23**	0.14*	0.15*	0.09	0.12	1.00			
IEXTRA	0.36**	0.05	0.82**	0.32**	0.24**	0.17**	0.13*	0.24**	0.18**	0.11	0.32**	0.10	1.00		
IAGREE	0.12	0.09	0.20**	0.59**	0.06	0.20**	0.16*	0.22**	0.35**	0.11	0.21**	0.10	0.17**	1.00	
IEMOT	0.30**	0.34**	0.29**	0.23**	0.77**	0.08	0.08	0.09	0.05	0.04	0.23**	0.36**	0.24**	0.12	1.00

* $p < .05$ ** $p < .01$

The data from Table 5.16 were considered factorable as all assumptions as advocated by Coakes and Steed (1996) were met. Root one criterion produced a 5 factor solution. However Cattell's Scree plot provided support for a 6 factor solution and the sixth eigenvalue was .90 which was high enough to warrant the interpretation of a six factor solution (Carroll, 1993). The pattern matrix, percent of variance accounted for, eigenvalues, and factor correlation matrix for the six factor solution are presented in Table 5.17. The six-factor solution accounted for 86.94% of the total variance. The first factor was labelled Big Five Confidence and comprised loadings from all IPIP Form B confidence subscale scores. The second, third, fourth, fifth, and sixth factors were labelled Emotional Stability, Conscientiousness, Agreeableness, Intellect, and Extraversion respectively. As anticipated, for each dimension, the last five factors all comprised loadings from the expected IPIP, and IPIP Form B subscale scores. There were low to moderate correlations among the factors. Because of the emergence of this anticipated six-factor solution, it can be surmised that Big Five Confidence and Big Five ratings are separate, but correlated processes. This structural analysis replicates the six-factor solution obtained in Study 2.

Table 5.17

Pattern Matrix for all IPIP and IPIP Form B Subscale Scores, Using Principal Components Analysis with Oblique Rotation (N = 243)

Variable	$h^{2(b)}$	F1a	F2[c]	F3[d]	F4[e]	F5[f]	F6[g]
INTELLIR	0.86					.78	
CONCIR	0.91			-.92			
EXTRAIR	0.92						-.90
AGREEIR	0.79				.70		
EMOTIR	0.89		.92				
INTELLIC	0.77	.87					
CONCIC	0.90	.93					
EXTRAIC	0.85	.90					
AGREEIC	0.84	.81			.26		
EMOTIC	0.87	.97					
IINTELL	0.88					.94	
ICON	0.87			-.95			
IEXTRA	0.92						-.99
IAGREE	0.89				.96		
IEMOT	0.89		.97				
Eigenvalues		5.35	2.84	1.66	1.20	1.09	.90
% of variance		35.67	18.93	11.04	8.01	7.27	6.03
Factor Correlation Matrix							
	F1	F2	F3	F4	F5	F6	
F1	1.00						
F2	0.11	1.00					
F3	-0.21	-0.40	1.00				
F4	0.27	0.13	-0.16	1.00			
F5	0.09	0.25	-0.20	0.16	1.00		
F6	-0.24	-0.31	0.18	-0.24	-0.34	1.00	

Note. Note. $h^{2(b)}$ = Communalities [a] F1 = Big Five Confidence, [c]F2 = Emotional Stability, F3[d] = Conscientiousness, F4[e] = Agreeableness, F5[f] = Intellect, F6[g] = Extraversion. CONIC= Conscientiousness Item-by-item confidence Form B; EMOTIC= Emotional stability Item-by-item confidence Form B; INTELLIC = Intellect Item-by-item confidence Form B; EXTRAIC = Extraversion Item-by-item confidence Form B; AGREEIC= Agreeableness Item-by-item confidence Form B; CONCIR= Conscientiousness Item-by-item adjective rating Form B; EMOTIIR= Emotional stability Item-by-item adjective rating Form B; INTELLIR = Intellect Item-by-item adjective rating Form B; EXTRAIR = Extraversion Item-by-item adjective rating Form B; AGREEIR= Agreeableness Item-by-item adjective rating Form B; ICON = IPIP Conscientiousness; IEMOT = IPIP Emotional stability; IINTELL = IPIP Intellect; IEXTRA = IPIP Extraversion, IAGREE = IPIP Agreeableness. The cut-off for suppression was .20.

5.17 Validity Checks in Personality

5.17.1 Self-Focussed Attention in Relation to Personality

Hypothesis 17 formulated that PrSc would not be associated with Big Five confidence at either the block or item levels, and the expected outcome was produced. All correlations were theoretically zero, so the correlation matrix was not presented in tabular form.

5.17.2 Negative Affect in Relation to Personality

NA was not expected to be associated with Big Five confidence at either the block or item levels, according to hypothesis 18. The expected outcome was produced but with one exception. NA was significantly negatively correlated with Intellect block description confidence (r = -.17, $p < .01$), indicating that as Intellect block confidence increases, NA decreases.

5.17.3 Positive Affect in Relation to Personality

Table 5.18 reproduces the correlations that were calculated to test hypothesis 19: that PA would be positively associated with Big Five confidence scores at both the block and item-levels. Perusal of the table revealed significant positive correlations between PA and confidence at both the block and item levels for Agreeableness, Intellect, and Conscientiousness, as expected. In addition, PA was also correlated with item level Extraversion confidence.

Table 5.18
Correlations between Positive Affect and Personality Confidence Scores (N = 243)

Variable	PA
PA	1.00
Agreeableness block confidence rating	0.27**
Intellect block confidence rating	0.21**
Conscientiousness block confidence rating	0.14*
Extraversion block confidence rating	0.08
Emotional Stability block confidence rating	0.12
mean Intellect confidence	0.20**
mean Emotional Stability confidence	0.13
mean Conscientiousness confidence	0.24**
mean Extraversion confidence	0.22**
mean Agreeableness confidence	0.24**

* $p < .05$ ** $p < .01$

5.17.4 Need for Cognition in Relation to Personality

Hypotheses 20 proposed that NFC would be positively associated with both Intellect confidence, and Conscientiousness confidence at both the block and item levels. As expected NFC was positively correlated with both block Intellect confidence ($r_{(243)}$

= .19, $p < .01$) and item-level Intellect confidence ($r_{(243)} = .24$, $p < .01$). The next part of the hypothesis was also supported, with NFC being positively associated with both block Conscientiousness confidence ($r_{(243)} = .19$, $p < .01$) and item-level Conscientiousness confidence ($r_{(243)} = .16$, $p < .01$). Unexpectedly, NFC was also significantly positively related with item-level Extraversion confidence ($r_{(243)} = .13$, $p < .01$). Taken as a whole, these findings suggested that as NFC increased so did confidence for the Intellect, Conscientiousness, and item-level Extraversion dimensions.

5.18 Factorial Structure of Big Five, Gf, and Self-Report Intelligence Confidence Scores

Hypothesis 21 reasoned that three confidence factors are likely to emerge from the structural analyses of the confidence scores obtained of the Big Five (item and block), Gf item-level, and Gf evaluative confidence measures. The next hypothesis dealt with the self-report intelligence confidence score, which was expected to load on the personality confidence factor when all Big Five and Gf confidence scores were factor analysed. First, all confidence scores underwent Principal Axis Factoring with oblique rotation, and the correlation matrix of these psychometric variables is presented in Table 5.19. The data were considered factorable as all the assumptions proposed by Coakes and Steed (1996) were met.

Table 5.19
Correlations Among all Confidence Variables in Study 3 (N = 243).

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
1.LSTGC	1.00																
2.WATGC	0.42**	1.00															
3.ESTGC	0.47**	0.57**	1.00														
4.CONCB	0.40**	0.29**	0.42**	1.00													
5.EXTRACB	0.18**	0.15*	0.22**	0.47**	1.00												
6.AGREECB	0.06	0.06	0.05	0.45**	0.39**	1.00											
7.INTELLCB	0.25**	0.18**	0.24**	0.41**	0.49**	0.51**	1.00										
8.EMOTCB	0.16*	0.15*	0.19**	0.49**	0.52**	0.55**	0.56	1.00									
9.LSTC	0.54**	0.20**	0.23**	0.20**	0.13*	0.02	0.11	0.06	1.00								
10.ESTC	0.32**	0.24**	0.49**	0.23**	0.21**	-0.01	0.16*	0.12	0.33**	1.00							
11.WATC	0.26**	0.24**	0.30**	0.17**	0.06	0.09	0.17**	0.08	0.33**	0.49**	1.00						
12.INTELLIC	0.36**	0.31**	0.43**	0.66**	0.61**	0.56**	0.62**	0.57**	0.19**	0.26**	0.21**	1.00					
13.EMOTIC	0.19**	0.24**	0.26**	0.62**	0.59**	0.70**	0.63**	0.74**	0.08	0.19**	0.13	0.7**4	1.00				
14.CONCIC	0.25**	0.21**	0.27**	0.68**	0.56**	0.70**	0.65**	0.71**	0.12	0.17**	0.16*	0.74**	0.87**	1.00			
15.EXTRAIC	0.27**	0.21**	0.23**	0.51**	0.63**	0.63**	0.67**	0.74**	0.16*	0.15*	0.12	0.72**	0.80**	0.84**	1.00		
16.AGREEIC	0.18**	0.15*	0.21**	0.55**	0.52**	0.74**	0.60**	0.67**	0.10	0.10	0.15*	0.68**	0.76**	0.80**	0.80**	1.00	
17.SICQ	0.31**	0.22**	0.29**	0.54**	0.50**	0.57**	0.66**	0.63**	0.12	0.19**	0.15*	0.72**	0.78**	0.81**	0.80**	0.72**	1.00

* $p < .05$ ** $p < .01$

A solution employing root one criterion produced the three anticipated correlated factors, which accounted for 67.57% of the total variance. The percent of variance accounted for, eigenvalues, and communalities are presented in Table 5.20. Self-Report Confidence was the first factor extracted from the data set, so named because all of the Big Five confidence scores plus the intelligence self-report confidence score, all loaded highly on it. This outcome provided support for the notion that individuals' use the same cognitive processes to make either item-by-item confidence judgments or block confidence judgments regarding their personality. Study 2 findings were therefore successfully replicated, as anticipated. Furthermore, self-report intelligence confidence also appears to employ similar cognitive processes to those used when making Big Five confidence judgments, again fulfilling expectations. The second factor extracted from the data was labelled Gf Self-Evaluative Confidence because its major loadings came from all of the Gf PTPE confidence scores, as anticipated. The Letter Series PTPE confidence score was complex as it demonstrated loadings on both factors 2 and 3. In accordance with expectations, the third factor was labelled Gf Self-Monitoring Confidence because its loadings came from the item-level Gf confidence variables. Conscientiousness block confidence rating was a complex and loaded on factors 1 and 2. The correlation matrix showed that the three factors used separate but correlated cognitive processes. Overall the factor analysis of the confidence variables shows that Personality Confidence, Gf Evaluative Confidence, and Gf Self-Monitoring confidence are separate but correlated constructs, as presumed by the relevant hypotheses.

Accuracy scores were not included in the above factor analysis as these scores were not the focus of this analysis because it was expected that confidence in PTPEs and cognitive accuracy scores would define two separate factors at the structural level.

This expectation was upheld when a two factor solution emerged with a low correlation between (r = .29) the factors (see Table U2 in Appendix U).

Table 5.20
Principal Axis Factoring with Oblique Rotation all Big Five and all Gf Confidence Scores (N = 243)

Variable	Communalities	F1 – Personality Confidence	F2 -Gf Evaluative Confidence	F3- Gf- Self-Monitoring Confidence
LSTGC	0.49	0.09	0.27	0.50
WATGC	0.36	0.05	0.55	0.05
EATGC	0.85	0.01	0.94	-0.04
CONCB	0.51	0.57	0.26	0.06
EXTRACB	0.41	0.62	0.04	0.03
AGREECB	0.58	0.80	-0.18	-0.05
INTELLCB	0.52	0.71	0.02	0.05
EMOTCB	0.60	0.79	-0.02	-0.06
LSTC	0.69	-0.01	-0.17	0.91
EATC	0.34	0.01	0.36	0.32
WATC	0.24	0.02	0.19	0.36
INTELLIC	0.73	0.74	0.23	0.04
EMOTIC	0.84	0.92	0.05	-0.08
CONCIC	0.87	0.93	0.01	0.00
EXTRAIC	0.83	0.92	-0.08	0.07
AGREEIC	0.75	0.89	-0.08	0.00
SICQ	0.73	0.83	0.05	0.02
Eigenvalues		7.94	2.52	1.02
% of variance		46.71	14.84	6.02

Factor correlation Matrix

Factor	F1	F2	F3
F1	1.00		
F2	.30	1.00	
F3	.22	.49	1.00

Note. ACC= Academic self-report confidence.

5.19 Factorial Structure of Cognitive and Big Five Bias Scores

The final hypothesis of this dissertation stated that two factors would emerge from the structural analyses of the item-level cognitive and item-level Big Five bias scores. Please note that the PTPE bias scores were not included because of the experimental dependency that exists between these scores and the cognitive item-level bias scores. Principal Axis Factoring with oblique rotation was applied to these bias scores and the correlations among these scores are presented in Table 5.21. The correlation matrix met all of the assumptions for factor analysis as recommended by

Coakes and Steed (1996), and a solution employing root one criterion produced the two expected factors, which accounted for 45.34% of the total variance.

Table 5.21
Correlations among Cognitive and Big Five Bias Scores

Variable	1	2	3	4	5	6	7	8
LSTBIAS	1.00							
WATBIAS	0.25**	1.00						
EATBIAS	0.24**	0.25**	1.00					
Intellect Bias	0.08	0.25**	0.17**	1.00				
Conscientiousness Bias	0.12	0.18**	0.21**	0.27**	1.00			
Extraversion Bias	0.07	0.11	0.08	0.37**	0.31**	1.00		
Agreeableness Bias	0.07	0.20**	0.11	0.26**	0.19**	0.33**	1.00	
Emotional Stability Bias	0.08	0.19**	0.19**	0.28**	0.31**	0.12	0.23**	1.00

** $p < .01$.

Table 5.22 displays the percent of variance accounted for, eigenvalues, and communalities. Factor 1 was labelled Personality Bias because all of the Big Five Bias scores loaded highly on it. Gf Bias was the label given to factor 2. The factor correlation matrix shows that the two factors were moderately correlated. Collectively, the results indicate that Gf and Big Five bias are separate but correlated constructs.

Table 5.22
Principal Axis Factorings of Cognitive and Big Five Bias Scores (N = 243).

Variable	$h^{2(b)}$	F1[a]	F2
LSTBIAS	0.19	-0.05	<u>0.45</u>
WATBIAS	0.28	0.12	<u>0.47</u>
EATBIAS	0.28	0.02	<u>0.52</u>
Intellect Bias	0.36	<u>0.56</u>	0.07
Conscientiousness Bias	0.26	<u>0.43</u>	0.15
Extraversion Bias	0.40	<u>0.69</u>	-0.17
Agreeableness Bias	0.23	<u>0.47</u>	0.02
Emotional Stability Bias	0.20	<u>0.32</u>	0.20
Eigenvalues		2.41	1.21
% of variance		30.16	15.18

Factor Correlation Matrix

	F1	F2
F1	1.00	
F2	.43	1.00

Note. $h^{2(b)}$ = Communalities; [a] F1 = Personality Bias; F2 = Gf Bias.

5.20 Discussion

The aims of this study were to a) examine bias (item and PTPE) with Gf tasks; (b) investigate whether individual differences in age, gender, ability, personality, need for cognition, and negative affect, influence cognitive confidence scores; (c) confirm previous findings from Studies 1 and 2 for Big Five confidence, accuracy and bias judgments; and (d) examine both the discriminant and convergent validity of Big Five confidence scores in relation to self-focussed attention, affect, and need for cognition.

Following this, Study 3 investigated the factorial structure of both personality confidence scores and cognitive confidence scores. Further confidence ratings were taken from a measure of self-report abilities and were factor analysed with the Big Five confidence and Gf confidence measures. Finally, the factorial structure of cognitive and Big Five bias scores were examined.

No specific hypotheses were formulated in relation to the Gf item-level bias scores, but results indicated that participants were overconfident for all three tasks. These results were consistent with previous research in which overconfidence was demonstrated (Pallier, 2003; Stankov, 2000a). For the Esoteric Analogies and Word Association tasks, the scatterplots of item means clearly showed that the hard-easy effect is one explanation for the observed overconfidence (see Figures 5.4 and 5.6). This was not the case for the Letter Series task, wherein the scatterplot showed that items were close to the perfect calibration line. However, the percentage of missing data suggested that participants may have run out of time and the last five items were subsequently deleted from the analyses. Some difficulties may have arisen from the fact that responses were provided in uncontrolled environments. That is, participants completed the battery from any computer that had access to the internet, including their home computers. Conducting this study in a controlled environment was unfortunately not viable, because many of the participants were not university students. Despite this limitation, it was encouraging that alpha co-efficients for the accuracy and confidence scores obtained from the Gf tasks were high, and also consistent with previous research, in which the environment for test administration was controlled (e.g., Kleitman, 2003). Moreover, collecting cognitive and self-report data online was a practical and viable option, with past research indicating that web-based data typically yield results comparable in reliability and validity to that of traditional data collection methods (Gosling, Vazire, Srivastava, & John, 2004; McGraw, Tew, & Williams, 2000).

The first hypothesis, that PTPE bias scores would demonstrate better calibration than item-level bias scores, was upheld by the findings as well as concordant with previous research (Kleitman & Stankov, 2001; Stankov, 2000a; Stankov & Crawford, 1996a, 1996b, 1997). However, the next hypothesis, that age would be positively related to Gf item level bias, was neither supported nor in agreement with previous research (Crawford & Stankov, 1996; Pallier, 2003), or with the findings from Study1. This inconsistency may represent a Type I error.

Addressing gender differences in confidence, some previous researchers found that, males were more confident than females for various cognitive tasks (Pallier, 2003; Pulford & Colman, 1997; Ross & Fogarty, 2006; Stankov, 1998). The Study 3 findings did not agree with the research just mentioned, but dovetailed instead with the findings from Study 1, and with Stankov and Lee’s more recent work (Stankov & Lee, 2008) which indicated that males and females were equally confident for the TOEFL internet-based test.

The next hypothesis concentrated on ability levels and assumed that those who attain low scores on tasks of cognitive ability are likely to be more mis-calibrated.

This hypothesis was supported, and replicated the results from Study 1 which also showed that individual differences in ability led to significant differences in mis-calibration. The findings for the Gf tasks in Studies 1 and 3, parallel the work of Kruger and Dunning (1999), who found that low scorers overrated their test performance relative to their peers. Overall these findings suggest that those who know, do know more about what they know with one *very* important caveat. Such a conclusion is, at this early stage of investigation, is strictly limited to the tasks used in Studies 1 and 3. Also worth mentioning is that top quartile participants experienced ceiling effects. That is, because scores for both accuracy and confidence were high for these participants, and due to the way confidence is measured, it is not possible to achieve high bias scores. As Stankov and Kleitman (Stankov & Kleitman, 2008, p. 557) more recently noted, "…it is probably best to ignore the tweak-your-nose interpretation" that high scorers on cognitive tasks have more insight than low scorers do, because of the measurement problems highlighted above.

Work by Schaefer et al. (2004) who advocated the use of partial correlations to examine the relations between personality and both cognitive confidence and cognitive bias scores, formed the basis of the next tested hypothesis. In accordance with expectations, Intellect was positively associated with cognitive confidence and accuracy. An unexpected relationship however, was discovered when Conscientiousness was positively associated with cognitive confidence and cognitive bias. Although this is inconsistent with the findings of Schaefer and colleagues (2004), other research using zero-order correlations has demonstrated small but significant positive correlations between Conscientiousness, and cognitive confidence and cognitive bias scores (Kleitman, 2003). Taken collectively across Studies 1 and 3, simple zero-order correlations provide similar information to that provided by partial correlations, with one qualification. Because Intellect is significantly correlated with both accuracy and confidence, it is prudent to partial out the influence of accuracy when correlations between Intellect and both cognitive confidence and cognitive bias are being examined.

The next hypotheses stated that NFC would be positively associated with Gf confidence and that NA would be negatively associated with Gf confidence. Both hypotheses were supported and results were consistent with previous research by Wolfe and Grosch (1990) who investigated whether individual differences in NA and NFC influenced cognitive confidence. Although the findings for NFC also agreed with results obtained by Jonsson and Allwood (2003), they conflicted with findings from Blais et al. (2005). Task differences (general knowledge versus reasoning) are a probable explanation for this discrepancy given that previous research showed that high NFC individuals scored higher on reasoning tasks than did their low NFC counterparts (Cacioppo et al., 1986). It follows that confidence in the accuracy of self-assessment is also likely to increase for the reasoning tasks in Study 3. The negative correlations between NA and Gf confidence were inconsistent with Allwood (2002) and Allwood (1991) who did not find associations between NA and confidence and mis-calibration for general knowledge questions. It may be that the positive correlation found between NA and Gf confidence in Study 3 was due to the pervasive nature of the NA variable, which has a reputation, in the occupational stress literature for creating unwanted associations (e.g., Brief, Burke, George, Robinson, & Webster, 1988; Burke, Brief, & George, 1993; Elliott, Chartrand, & Harkins, 1994; Fogarty et al., 1999). Overall, the results for NFC and NA need to be viewed with caution until they are replicated.

The next step in Study 3 involved testing whether one confidence factor, with loadings from both the Big Five block description and Big Five item-by-item confidence ratings, emerged from the structural analysis of the data set. This one-factor solution did indeed emerge, suggesting that the same cognitive processes underlie both block description and item-by-item confidence judgments. This finding was consistent with the argument presented in Chapter 4 that all Big Five confidence judgments would have trait summaries in place because they are derived from the lexicon of daily life, and are therefore likely to follow abstraction processes. The one-factor solution from Study 3 replicated the one-factor solution from Study 2, which used a different sample of participants. In view of these outcomes, it is reasonable to speculate that individuals used a cognitive database of trait generalizations or summaries which reside within semantic memory, from which they made all Big Five confidence judgments about themselves. This assertion was consistent with abstraction theory within the memory domain (e.g., Buss & Craik, 1983; Klein, 2004; Klein & Loftus, 1993b; Klein, Loftus, Trafton et al., 1992). Logic dictates that computational processes are to onerous to be practical for making the many decisions presented in daily life. At this stage, these conclusions are limited to Big Five confidence judgments, and to investigate the veracity of these findings, results need to be replicated using other Big Five measures, different samples, and in other cultures. The Big Five traits are not the only personality constructs likely to display high ecological validity, and other constructs present a worthwhile basis for future research. Are there trait summaries in place, for instance, for other traits like friendliness, assertiveness, nurturance, and fairness?

Further hypotheses stated that individual differences in gender would not influence Big Five confidence, accuracy or bias. All of these hypotheses were supported thus replicating findings across three studies for the confidence and accuracy variables. With regard to Big Five bias, the absence of gender differences in Study 3 paralleled findings from Study 2. Overall, this lack of gender differences for Big Five confidence, Big Five accuracy and Big Five bias scores were consistent with Hyde's (2005) gender similarities hypothesis which stated that males and females are more alike than they are different. Again, conclusions are limited to the Big Five domain require replication using different samples. Personality variables that are more strongly stereotyped with regard to gender roles (e.g., social boldness, expressiveness, gentleness, and diligence) pose an interesting question for future researchers.

Correlations among block description Big Five accuracy scores were low. The result for the block-description accuracy scores replicated findings from Studies 1 and 2, with all studies evincing low correlations among the block description scores. It seems then that calculating an overall personality bias score from block level judgments does not make sense. The low correlations among the Big Five item-level accuracy scores in the current study, match the low correlations also found in Study 2. Moreover factor analyses of the accuracy scores developed from methods 1, 2, and 3 demonstrated that accuracy was specific to each Big Five domain in both Studies 2 and 3. Therefore, calculation of separate bias scores for each personality dimension was warranted. These result for the accuracy scores need to be viewed with caution as there is no perfect criterion by which to determine accuracy within the personality domain. Suggestions for future research in relation to personality accuracy are presented in the general discussion of this dissertation.

Several of the hypotheses developed for Study 3, emphasised that age is not likely to be associated with Big Five confidence (either block or item-level), accuracy (either block or item-level), or item-level bias. As expected, age did not correlate with

Big Five confidence, accuracy or bias, but with one qualification. A significant correlation arose between age and the block description accuracy score for Emotional Stability dimension (r = .13), however, the effect size was small.

This finding, however, requires replication before speculating why older participants were more accurate for this dimension. At this stage, the overall results for Study 3 suggest that individual differences in age do not influence confidence, accuracy or bias for the Big Five dimensions. This outcome largely mirrors findings from Studies 1 and 2, and is consistent with Five-Factor theory of personality (McCrae & Costa, 1999) which posits that personality traits develop during childhood and adolescence, then remain stable in adulthood. It seems that Big Five accuracy, confidence, and bias remain constant too. The overall results for age in the current studies also fit well with PMM theory (Gigerenzer et al., 1991) which reasons that people are well calibrated to their natural ecology rendering age differences unlikely. Because the current study was limited to the Big Five dimensions, it is left to future researchers to establish the generalisability of conclusions to other personality dimensions, and of course using different samples.

Results supported the hypothesis that individuals would be well-calibrated for the Big Five item-level bias scores. This finding of good calibration for Big Five judgments in Study 3 replicates the results from Study 2 where individuals were also well-calibrated. These results were consistent with Gigerenzer's (1991) theory where it was argued that individuals are well calibrated to their natural environments. Naturally, this conclusion is also limited to Big Five judgments and awaits future research before it can be determined if good calibration generalises to other personality domains. Results from Studies 2 and 3 give investigators sufficient reason to expect good calibration with other personality judgments that have high ecological validity. However, because of the level of variability in the bias data, the methods used to calculate bias in Studies 2 and 3 may need further refinement. Nevertheless, these scores represent a viable starting point for other research endeavours, particularly as the results obtained from these scores were replicated across two studies.

The next step in the current studies constituted examination of whether a six-factor solution emerged from the structural analysis of all IPIP, and IPIP Form B subscales. The hypothesis that a six-factor solution was likely to emerge was upheld by outcomes from both Studies 2 and 3. This result is very interesting as it has not been investigated before. Perhaps psychologists have made the assumption that rating one's personality is the same as confidence in that rating. Data from Studies 2 and 3, however, provided evidence contrary to this assumption because these scores defined separate (although) correlated factors at the structural level. At this stage, these conclusions still need to be viewed cautiously as the IPIP Form B was a newly developed measure for the purposes of the current research. Replication with other samples and other cultures appears warranted.

This six-factor solution also provides evidence that confidence is related to, but is distinct from personality which is in agreement with research in the cognitive domain which has demonstrated that the confidence trait is on the borderline between cognitive abilities and personality (Baker, 2001; Kleitman & Stankov, 2001, 2007; Pallier et al., 2002; Stankov, 1998, 1999a, 2000a; Stankov & Crawford, 1996a, 1997; Stankov & Lee, 2008).

The reliabilities of all the IPIP Form B subscales developed for use in Studies 2 and 3, were very high for both the ratings (.82 to .92) and the confidence scores (.91 to .95). These high internal consistency estimates for the Big Five confidence scores were similar to the high internal consistency scores reported for confidence ratings in the cognitive domain (Baker, 2001; Jonsson & Allwood, 2003; Kleitman, 2008; Kleitman & Stankov, 2001, 2007; Liberman & Tversky, 1993; Pallier et al., 2002; Stankov, 1999a, 1999b, 2000a; Stankov & Lee, 2008). Convergent and discriminant validity checks, provide more information about the construct validity of Big Five confidence scores.

To test the validity of Big Five confidence judgments, discriminant validity was first examined in relation to the constructs of PrSc and NA. As expected, Big Five confidence did not correlate with PrSc and NA scores, with the exception that NA was significantly negatively correlated with Intellect block description confidence. Given the pervasive nature of NA this correlation may be spurious. Next the convergent validity of Big Five confidence judgments was explored in relation to the potentially related constructs of PA and NFC. It was expected that PA would be associated with Big Five confidence scores, a hypothesis that was partially supported, with PA demonstrating positive associations with Agreeableness, Conscientiousness, Intellect and Extraversion confidence scores. The anticipated positive associations between NFC and confidence for both Intellect and Conscientiousness were also established, along with a significant positive relationship between NFC and item-level Extraversion confidence. Overall, these validity checks were in line with expectations but need to be viewed with caution until future researchers endeavour to expand upon the construct and validity data for the Big Five confidence judgments obtained in the current studies.

Overall, the results for the IPIP Form B confidence scores were encouraging but even so, needs to be reproduced using other samples, and by investigating whether confidence across other personality dimensions all loads onto one factor in the way it does for the Big Five dimensions in Studies 2 and 3. Extending this confidence research into the areas of interests, attitudes and values, may well benefit individual differences psychologists in their efforts to understand the structure of the confidence trait across other domains.

Three correlated confidence factors were expected from the structural analyses of Big Five (item and block), Gf (item-level and post-test confidence in the PTPE), and self-report intelligence confidence scores. The relevant hypotheses were supported. The first factor was labelled Personality Confidence which comprised all the confidence scores from the Big Five personality dimensions plus the self-report intelligence confidence measure. The second factor was labelled Gf Self-Monitoring Confidence, and the third factor was labelled Gf Evaluative Confidence. This three-factor solution was consistent with a combination of Meta-cognitive theory (Schraw & Dennison, 1994), Self-Concept theory (Marsh, 2008; Shavelson et al., 1976), and Abstraction theory within the memory domain (e.g., Buss & Craik, 1983; Klein, 2004; Klein & Loftus, 1993b; Klein, Loftus, Trafton et al., 1992). From the meta-cognitive vantage point, one expects the Gf confidence factors to be similar to the constructs of self-monitoring and evaluation respectively.

As was argued in Chapter 2, in terms of Self-Concept theory, confidence judgments are thought to be differentiated across the ability and non-ability domains, because both self-concept and self-confidence are cognitive appraisals of the self. The emergence of the Personality Confidence factor, with loadings from both the block and item-level confidence ratings, was consistent with the argument that Big Five confidence judgments have trait summaries that reside within semantic memory. It follows that abstraction processes were likely to have been used. On the other hand, the three-factor solution did not conform to a strict application of PMM theory

(Gigerenzer, 1991) because it surmises that the cues used to answer different types of questions will differ across the domains. Therefore the three factors should not be associated. As with other aspects of the current exploratory research, such as the taking of a confidence rating in the PTPE scores, the three-factor solution must be contemplated with caution, until findings are reproduced by future researchers.

Regarding cognitive bias and Big Five bias, two factors were expected to emerge from the structural analyses of these scores. This hypothesis was supported and was consistent with Self-Concept theory (Marsh, 2008; Shavelson et al., 1976) which posits that ability and non-ability factors split at the factorial level. These two factors were moderately correlated, which indicates that Gf and Big Five bias are separate but correlated constructs. However definite conclusions must be left until the findings have been replicated. At this stage it is not clear why the factors were correlated but the magnitude of this correlation is certainly encouraging, and suggests that Big Five self-report bias scores share approximately 18.5% of common variance with objective bias measures from the cognitive domain.

The limitations of this study need to be taken into account and addressed by researchers. The ordinal position of the measures, for instance, could not be randomised across participants due to programming restrictions. Nevertheless, the order of presentation of the IPIP based measures were changed from Study 2 and overall results were by in large the same (see general discussion).

Study 3 was also restricted by the small number of cognitive tasks it employed, leading to a lack of clarity about whether Big Five item-level confidence and bias share variance with other ability measures. This limitation may be overcome by including a larger number of cognitive tasks from different ability domains. The current research also used only IPIP based instruments to measure the Big Five personality dimensions, and a prudent next step for future researchers might entail replication of confidence and bias results, using instruments other than the IPIP.

Chapter 6 General Discussion

The journey of a thousand miles begins with a first step

Chinese Sage Lao-tze

The overall aim of this dissertation was to take the first step in extending the calibration paradigm into the domain of personality judgments. Before delving into personality appraisals, mis-calibration in the cognitive domain was examined, along with whether individual differences in gender, age, personality, and ability influenced these scores. These well-established findings from the cognitive domain have been discussed in detail in previous research, and in the preceding chapters. They are therefore not included here. Instead this chapter is directed toward an overview of the *new* findings in relation to cognition and personality, as well as the implications of these findings for future research and psychological practice. Before presenting a summary of these findings, however, some issues in relation to Big Five accuracy, confidence and bias are briefly discussed.

As outlined in Chapters 1 and 2, strict parallels with cognitive accuracy were not possible because, outside the bounds imposed by reliability and validity estimates, as there is no way to determine whether personality assessments are accurate. However, it is possible to approach this goal by using the notion of consistency/reliability. Consistency does not guarantee accuracy, but a lack of consistency implies inaccuracy Parallel forms of the IPIP Big Five measures were therefore developed in Studies 1 (BFBD) and 2 (IPIP Form A and IPIP Form B). The psychometric properties of IPIP Form B were more than acceptable. Two points are worth highlighting. First, the factorial validity of the adjective ratings from IPIP Form B were in accordance with the Five Factor theory of personality (McCrae & Costa, 1999), and second, a five-factor solution with significant loadings from the expected IPIP Form B items, demonstrated that participants comprehended what they were asked to do. Furthermore the scales of the IPIP Form B were internally consistent across both Studies 2 and 3.

It is acknowledged that although differences in scores on parallel forms of less than perfectly reliable tests occur for a number of reasons (e.g., fatigue, failure to follow test instructions and so forth), one of these reasons, as yet largely unexplored, involves an individual's knowledge of his or her personality traits, or the lack there of.

Despite possible concerns regarding the methodologies used in determining personality accuracy in the current studies, perusal of Table 6.1 shows that meaningful results have emerged for accuracy scores at both the block description and item-levels. The data reproduced in Table 6.1 demonstrates that use of the block description method and the 20% rule, produced accuracy scores for Conscientiousness and Emotional Stability that were strikingly similar across all three studies. For the other three personality dimensions, accuracy scores were more varied. Yet for each trait, the similarities in block accuracy scores were encouraging. Moreover, it cannot be denied, that the continued study of the accuracy of self-insight, has important implications for clinical and counselling psychology for example, particularly because effective psychotherapy relies heavily on clients' insight into their own problems and disorders. Failure of this insight is a major stumbling block to the implementation of effective interventions. Study 2 continued the investigation into personality accuracy for this very reason.

Table 6.1
Accuracy Scores for the BFBD and the IPIP-Form B

BFBD 20% Accuracy Scores	Study 1	Study 2	Study 3
Conscientiousness	71.65	72.84	74.49
Extraversion	77.95	57.78	76.13
Agreeableness	82.68	91.11	84.36
Emotional Stability	72.44	70.12	71.19
Intellect	87.40	84.20	79.01
IPIP Form B Mean Accuracy Scores		Study 2	Study 3
Conscientiousness	-	83.62	86.96
Extraversion	-	83.54	80.80
Agreeableness	-	80.08	80.52
Emotional Stability	-	80.00	80.93
Intellect	-	78.60	83.67

The similarity of the IPIP Form-B item-level mean accuracy scores for each trait across Studies 2 and 3, is evident in the bottom portion of Table 6.1. The findings for the block description and item-level accuracy scores, were largely consistent with PMM theory (Gigerenzer et al., 1991), which claims that individuals are well calibrated to their natural ecology. It follows then that individuals would be reasonably accurate for traits that have high ecological validity.

Remembering that there is no perfect criterion for determining personality accuracy, results should be treated with caution. Despite the attainment of reasonably consistent results across three studies for the block description ratings, and across two studies for the item-level accuracy scores. Definitive conclusions must wait until researchers replicate these findings using differing samples, and using other measures of accuracy that have been highlighted in the personality accuracy literature (e.g., Albright et al., 1997; Borkenau & Liebler, 1993a, 1993b; Funder & Colvin, 1988; Vogt & Colvin, 2003). Perhaps researchers could use agreement measures (e.g., self-other) and physiological methods, in conjunction with the self-report accuracy measures developed in the current studies. This would assist in establishing whether accuracy across these different techniques leads to the same conclusions.

Concerns about measurement are not merely confined to the study of personality accuracy, however. The psychological literature provides evidence, that other areas of psychology have spent years grappling with the quest to find appropriate assessment methods. The debate is ongoing. Emotional intelligence, social intelligence, and multiple intelligences are some of the popular constructs on which the current debate about measurement issues, are focussed.

Prior to the studies conducted in this dissertation, no techniques using calibration procedures had been established to assess Big Five confidence or bias. The studies in this dissertation represent a tentative exploration with regard to an important area of psychological investigation.

Progressing to the measurement of confidence, the BFBD (see Study 1) and the IPIP Form B (see Study 2) were designed to obtain confidence ratings in relation to the Big Five personality judgments. In Study 1, 127 participants were tested face-to-face basis and reported no difficulties in understanding that they were required to provide a confidence rating that corresponded with a (hypothetical) rating given by an imagined device that accurately knew their personality.

A limitation of the BFBD, however, was that a mean confidence score could not be calculated. The development of IPIP-Form B addressed this limitation by obtaining item-level confidence ratings, from which a mean confidence rating score was calculated for each Big Five dimension. Confidence scores obtained from both these measures are presented in Table 6.2. Confidence in Big Five personality judgments was around 80%, irrespective of the trait in question. Because there was no available benchmark with which to compare this percentage, the current studies have, of necessity, defined that benchmark. It is not surprising that people are decidedly confident about their personality self-assessments, given the high ecological validity of the Big Five dimensions. There are however other possibilities that could explain why confidence ratings were consistently high across the Big Five traits across all studies. For example, participants may have responded without thinking, or they may have been trying to create a positive impression. Think aloud protocols are one way that could be used in future research to elucidate the reasons why participants were so confident.

Whether this benchmark remains high across the sub-facets of the Big Five dimensions, as well as whether it holds true when confidence in personality self-assessments is measured using personality constructs not examined in this dissertation, are both subjects for future empirical investigation.

Table 6.2

Confidence Scores for the BFBD and the IPIP-Form B

BFBD Confidence Scores	Study 1	Study 2	Study 3
Conscientiousness	80.08	81.88	81.01
Intellect	78.11	80.66	80.54
Agreeableness	81.89	82.52	82.54
Extraversion	80.16	80.70	83.50
Emotional Stability	78.19	80.67	79.36
IPIP Form B Mean-Item Level Confidence Scores		Study 2	Study 3
Conscientiousness	-	81.88	83.51
Intellect	-	80.67	82.43
Agreeableness	-	82.52	86.75
Extraversion	-	80.71	82.91
Emotional Stability	-	80.67	83.60

Having established a method of calculating mean personality accuracy and mean confidence scores, for each personality dimension, a bias score was then computed (see Table 6.3). For each trait across Studies 2 and 3, these bias scores were shown to be comparable, although a higher degree of overconfidence emerged for the Agreeableness dimension in Study 3. From these results it was evident that people were reasonably well-calibrated in relation to their personality judgments. Overall, the bias findings are generally consistent with PMM theory (Gigerenzer et al., 1991), which argues that people are well calibrated to their natural environments.

Table 6.3
Mean Bias Scores across Studies 2 and 3

Big Five Mean Item Level Bias Scores	Study 2	Study 3
Conscientiousness	-1.74	-3.46
Intellect	2.07	-1.24
Agreeableness	2.43	6.19
Extraversion	-2.83	2.11
Emotional Stability	0.67	2.67

6.1 Factorial Structure of Big Five Confidence

Results from Studies 2 and 3 demonstrated that a one-factor solution resulted when the block and item-by-item confidence ratings were factor analysed. This one-factor solution suggests that the cognitive processes that underlie both block and item-by-item confidence judgments are the same. This finding reinforces the argument presented in Chapter 4 that all Big Five confidence judgments would have trait summaries in place because they come from the lexicon of daily life, and would therefore follow abstraction processes.

At this stage, it appears reasonable to speculate that individuals used a cognitive database of trait generalizations, or summaries, that reside within semantic memory from which they made Big Five confidence judgments about themselves. This assertion was consistent with abstraction theory within the memory domain (e.g., Buss & Craik, 1983; Klein, 2004; Klein & Loftus, 1993b; Klein, Loftus, Trafton et al., 1992). It makes sense that computational processes are to onerous and impractical for day-to-day decision making. Although at this stage, conclusions are confined to Big Five confidence judgments, to investigate the veracity of these findings, results need to be replicated using other Big Five measures, different samples, and in other cultures. It would also be worthwhile for future researchers to investigate these findings using other personality constructs that are likely to have high ecological validity and therefore, would have trait summaries in place. For example, constructs such as friendliness, assertiveness, nurturance and fairness could be useful in this regard.

The one-factor solution for the Big Five confidence scores obtained across Studies 1, 2, and 3 also mirrors findings within the cognitive domain. That is, independent of the types of tasks employed (e.g., Gf, Gc, Gv), structural analyses of cognitive confidence scores have mostly resulted in a one-factor solution (Crawford & Stankov, 1996; Kleitman & Stankov, 2001; Pallier et al., 2002; Stankov, 1998, 1999a, 2000a; Stankov & Crawford, 1996a, 1996b, 1997; Stankov & Dolph, 2000; Stankov & Lee, 2008).

6.2 Gender and Age Differences in Big Five Confidence, Accuracy and Bias Scores

Across all three studies the data has demonstrated that individual differences in gender and age do not influence Big Five confidence, accuracy or bias scores. Despite findings being consistent with expectations derived from Hyde's work (Hyde, 2005) and the Five-Factor theory of personality (McCrae & Costa, 1999), further empirical investigation is required before generalising these conclusions outside the Big Five domain. Such investigation requires employing a variety of samples and personality dimensions not evaluated in the current studies.

6.3 The Factorial Structure of Cognitive and Big Five Confidence

Studies 1 and 3 assessed the factorial structure of cognitive and Big Five confidence scores and both studies produced a two-factor solution. In Study 1 the factors were uncorrelated (r = .05) with the opposite being true for Study 3 (r = .30. It is possible that the lack of correlation between the Big Five and cognitive confidence factors in Study 1 reflects not a lack of any statistical relationship, but rather an inability to validly measure personality confidence using a one-item scale for each Big Five dimension. To overcome this possibility, the IPIP Form B was developed for the purposes of Studies 2 and 3. The reliability of the confidence scores obtained from this measure was very high across both studies which were concordant with research in the cognitive domain where confidence ratings have also demonstrated high internal consistency (e.g., Kleitman, 2008; Kleitman & Stankov, 2001, 2007; Pallier et al., 2002; Stankov, 1998, 2000a; Stankov & Crawford, 1996a, 1997; Stankov & Kleitman, 2008; Stankov & Lee, 2007). Another potential reason why Big Five and cognitive confidences did not load onto the one factor is that these constructs may be conceptually different. One option is that Big Five confidence may be related to Kleitman's (2008) Sureness measure that assesses respondents' assuredness that various opinion statements will occur in the future albeit that those individuals are aware that correct answers may never become available. This possibility requires further empirical investigation.

Nevertheless, the factorial structure of Big Five and cognitive confidence in Study 3 provided further evidence that the *confidence trait* is related to, but is distinct from both personality and abilities confidence. Similarly, factor analyses of the IPIP and IPIP Form B subscales across Studies 2 and 3 provided convergent evidence that strengthens Stankov's contention that *confidence* is separate from, but related to personality.

The structural independence of personality ratings and confidence in those ratings from the IPIP and IPIP Form B across Studies 2 and 3 suggests that rating one's personality and expressing confidence in those ratings, are separate but correlated processes. This finding lays the foundation for future researchers to investigate whether this structural independence also occurs for other personality dimensions.

It is still unclear whether confidence across other personality dimensions shares variance with Big Five confidence, a worthwhile area of further investigation. Asking participants for confidence ratings regarding personality constructs perceived as negative remains a challenge, given the valid concerns about self-report data in relation to issues such as self-enhancement biases. Future research could examine the impact of self-enhancement by including measures of social desirability in their test batteries, and by educating participants that people tend toward self-enhancement. In other words, people need to understand that while answering psychological tests items in such a way as to maintain a positive view of themselves, is a natural human tendency, which often operates at the subconscious level to protect individuals from anxiety. As a psychologist who has facilitated over 400 group therapy sessions, in an acute mental health ward, providing patients with this information has facilitated greater meta-cognitive insight which has enhanced the therapy process. If participants understand the meta-cognitive processes that could well underlie their reluctance to endorse negative traits as being characteristic of themselves, then obtaining confidence ratings about these traits could be possible, and is a fruitful area of further research investigation.

The simple way that confidence ratings were obtained using the IPIP Form B opens up the way for future researchers to explore the confidence trait in other domains such as interests, attitudes and values. This area of psychological enquiry has expanded over the last ten years, particularly with the work of Ackerman and his colleagues (e.g., Ackerman, 1996; Ackerman, 1997, 2003; Ackerman & Beier, 2003; Ackerman & Heggestad, 1997). A number of constructs within the field of positive psychology offer a useful guide for future research investigation (e.g., zest, appreciation of beauty, citizenship team-work, gratitude, optimism, and leadership).

As expected, confidence ratings obtained from a measure of self-report general intelligence (abilities) shared more variance with personality confidence than with Gf confidence as it loaded on the personality confidence factor. The fact that the SICQ confidence did not share as much variance with Gf self-monitoring confidence as it did with Gf evaluative confidence, was probably due to the nature of the items in the inventory which were more evaluative in nature. It would be useful for future researchers to include a self-report measure of abilities that included several items for each ability domain, and a larger battery of cognitive tasks for each ability domain (e.g., Gc, Gf, Gv and SAR). The results for the self-report intelligence measure cannot be attributed to concerns about internal consistency because the alpha co-efficients were more than acceptable. Nonetheless, the SICQ was a newly adapted measure and, as with all new measures, reliability and validity needs to be re-examined through future research before commenting.

6.4 Big Five Bias

The data from two studies indicate that people were well calibrated for Big Five judgments. These results were in accordance with expectations and were consistent with PMM theory (Gigerenzer et al., 1991), which asserts that individuals are well calibrated to their natural environments. However, because of the exploratory nature of the method used to examine bias in Big Five judgments, and also because personality accuracy was a component of the bias scores, the current findings need to be viewed with caution. The findings of good calibration would need to be replicated with other samples and with other cultures before drawing conclusions. Moreover, these results are confined to Big Five judgments as operationalised by the IPIP based measures. It would be useful for other investigators to examine bias using different Big Five self-report questionnaires.

6.5 Factorial Structure of Cognitive and Big Five Bias Scores

Study 3 examined the factorial structure of cognitive and Big Five bias scores. The structural analyses showed that bias across these domains were separate but correlated processes. That is, Personality Bias and Gf bias respectively. The two-factor solution was consistent with Self-Concept theory (Marsh, 2008; Shavelson et al., 1976) where ability and non-ability factors split at the factorial level. It was argued in Chapter 2 that this theory provides insight into what could be expected as both self-confidence and self-concept are cognitive appraisals of oneself. Results for the Gf bias factor were in accordance with findings from the cognitive domain where Stankov and his collaborators (Pallier et al., 2002; Stankov, 1998, 1999a) factor analysed bias scores obtained from various combinations of Gc, Gf, Gv tasks and found that the bias scores loaded onto one factor.

The moderate correlation (r = .43) between the two bias factors is encouraging, as it suggests, that despite the methodological difficulties of measuring Big Five bias, the factor scores were reliably correlated. The correlation between these two factors indicates that the processes that underlie cognitive (objective) bias are also partially involved in Big Five bias. This finding requires replication before speculating too widely as to why this moderate correlation between the Big Five and Gf factors emerged. Nevertheless, this result has significant implications for calibration researchers, who have been tyring to understand the mis-calibration phenomenon. From a conceptual vantage-point one possibility why Big Five and Gf bias did not load onto one factor, is that Big Five bias could be an internal consistency bias, which is conceptually distinct from Gf bias. This suggestion merits further empirical attention.

6.6 Practical and Other Research Implications

Throughout this chapter, a number of suggestions have already been made regarding continued research into personality confidence, accuracy and bias. Another interesting implication of this dissertation is for differential psychologists who are striving to understand the structure of the confidence trait. In the current studies, simple methods were used to obtain confidence ratings. These procedures could now be used to investigate the factorial structure of confidence in much more detail, and across other domains such as interests, attitudes and values.

The current research assumed that all individuals used abstraction processes when making confidence judgments for each Big Five trait. However, it is important for future researchers to verify this assumption, by asking individuals, via either think aloud protocols or qualitative commentaries, to elucidate whether or not they did require the use of computational processes to make their Big Five confidence judgments. This is important so that future researchers are aware of the underlying processing mechanism/s that individuals use when making confidence judgments re personality self-assessments.

6.6.1 Clinical and Counselling Psychology

The split at the factorial level between rating one's personality and expressing confidence in that rating (see Studies 2 and 3) has important implications in the field of clinical psychology. Earlier in this chapter it was argued that this factorial split makes sense, because rating one's personality and expressing confidence in those ratings are likely to be separate but correlated processes, given that the latter involves metacognitive processing. For example, Jane rated herself highly on an item measuring Conscientiousness. When she provided a confidence judgment in that rating, she was in fact monitoring and appraising her thoughts in relation to her original rating.

Within the clinical domain, it would be useful for clinicians to initially investigate whether rating one's worry or rumination, for example, were factorially distinct from confidence in that rating as this could inform clinicians about important aspects of therapy. For instance, Susan rates the item "worry will drive me mad" highly, and if she provides a high confidence rating (e.g., 80%), then therapy would need to address this maladaptive worry. If, however, her confidence in that rating was 30%, the focus of therapy is more usefully directed toward other worries, about which she expressed a higher confidence level.

It is incumbent on future research endeavours to continue measuring and refining methods for determining the accuracy of self-insight, particularly within the area of mental health. The prevalence of mental illness in Australia and other western countries is high (Department of Families, Housing, Community Services and Indigenous Affairs, 2006). That is, approximately one in five adults will experience symptoms of mental illness during a 12 month period. One of the main stumbling blocks to effective psychotherapy is that many clients fail to have adequate metacognitive insight into their problems and disorders (Dimaggio et al., 2005; Dimaggio et al., 2006). By contributing toward a valid attempt to operationalise intra-phenomenological accuracy within the area of personality, the current studies represent an important first step.

6.7 Other Limitations

One aim of Study 3 was to investigate the construct validity of personality confidence judgments. It was a limitation that this could only be achieved in relation to a limited number of psychological variables as participants had already responded to a large battery of objective and self-report measures. Self-concept measures investigating individuals' self-assessment of personality may prove to be a fruitful area of further research investigation.

Randomisation of the tests batteries was not possible for Studies 2 and 3 due to programming restrictions. Consequently, the impact of ordinality is unknown and future research could address this by randomising the presentation of measures in their test batteries.

Study 3 contained only a small number of cognitive tasks that measured Gf abilities, and only the Big Five dimensions. It would be useful for future research to include more tasks that measured other cognitive abilities and other personality constructs as well as the Big Five traits. Test-retest estimates showing the stability of Big Five confidence ratings over time are also needed.

The mean accuracy scores for each Big Five trait across Studies 2 and 3 were calculated from the three accuracy protocols developed in Study 2. It is a limitation that the three protocols for each dimension were significantly correlated suggesting that random variation could not be captured. Nonetheless, consistent results were obtained for each trait across two studies. It is beyond the scope of this dissertation to resolve the accuracy debate. Nevertheless, the methods used in this dissertation could be viewed as a spring board for future research, which could be directed toward refinement of the accuracy measures used in the current studies.

6.8 Conclusions

The main aim of this dissertation was to take calibration procedures into the domain of personality judgements. Interesting findings have emerged from this foray into the Big Five domain. The results from three studies showed that the benchmark for peoples' confidence in Big Five judgments was around 80%.

Accuracy for each trait was also reasonably high across the studies when consistency measures were used. High accuracy makes sense, because people normally have extensive feedback from their natural environments. Of course, there would be some people who do not benefit from this feedback because their insight is inadequate. Operationalising personality accuracy however, remains somewhat elusive because of the absence of any objective criterion against peoples' self-assessments can be judged. This same problem exists in other areas of psychological enquiry. There has, and continues to be, intense debate about the inherent problems with the development of valid and reliable measures of social, emotional, practical and interpersonal intelligences, which are also more complicated in terms of determining the accuracy of self-insight. Nonetheless, researchers have both persevered, and have continually refined the measures of such constructs. In determining the accuracy of self-assessments within the personality domain and using objective self-reports as the criterion for accuracy, the same sorts of issues and debates will likely arise. Nevertheless, psychologists need to ask themselves the following question. Are continued steps on a thousand mile journey worthwhile? Or does the psychological profession discontinue efforts towards examining a complicated issue that is valuable, and is of vital importance to individuals, for whom the cost of inaccurate self-insight is high.

References

Ackerman, P. L. (1996). A theory of adult intellectual development: Process, personality, interests, and knowledge. *Intelligence, 22*(2), 227-257.

Ackerman, P. L. (1997). Personality, self-concept, interests, and intelligence: Which construct doesn't fit? *Journal of Personality, 65*(2), 171-204.

Ackerman, P. L. (2003). Aptitude complexes and trait complexes. *Educational Psychologist, 38*(2), 85-93.

Ackerman, P. L., & Beier, M. E. (2003). Intelligence, personality, and interests in the career choice process. *Journal of Career Assessment, 11*(2), 205-218.

Ackerman, P. L., Beier, M. E., & Bowen, K. R. (2002). What we really know about our abilities and our knowledge. *Personality & Individual Differences, 33*(4), 587.

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligence, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin, 12*(2), 219-245.

Ackerman, P. L., Kanfer, R., & Goff, M. (1995). Cognitive and noncognitive determinants and consequences of complex skill acquisition. *Journal of Experimental Psychology: Applied, 1*(4), 270-304.

Albright, L., Malloy, T. E., Dong, Q., Kenny, D. A., Fang, X., & Winquist, L. (1997). Cross-cultural consensus in personality judgments. *Journal of Personality and Social Psychology*, *72*(3), 558-569.

Alicke, M. D. (1985). Global self-evaluation as determined by the desirability and controllability of trait adjectives. *Journal of Personality and Social Psychology, 49*(6), 1621-1630.

Alicke, M. D., & Govorum, O. (2006). The better-than-average effect. In M. D. Alicke, M. L. Klotz, D. Dunning & J. Krueger (Eds.), *The self in social judgment*. Philadelphia, PA: Psychology Press.

Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., & Yurak, T. J. (1995). Personal contact, individuation, and the better-than-average effect. *Journal of Personality and Social Psychology, 68*(5), 804-825.

Allwood, C. M., & Bjorhag, C. G. (1991). Mood and the realism of confidence judgments of one's own answers to general knowledge questions. *Scandinavian Journal of Psychology, 32*(4), 358-371.

Allwood, C. M., & Granhag, P. A. (1996). Realism in confidence judgments as a function of working in dyads or alone. *Organizational Behavior and Human Decision Processes, 66*(3), 277-289.

Allwood, C. M., Granhag, P. A., & Jonsson, A. (2002). Does mood influence the realism of confidence judgments? *Scandinavian Journal of Psychology, 43*(3), 253-260.

Allwood, C. M., Granhag, P. A., & Jonsson, A. (2006). Child witnesses' metamemory realism. *Scandinavian Journal of Psychology, 47*(6), 461-470.

Anastasi, A., & Urbina, S. (1997). *Psychological testing*. Crawfordsville, IN: Prentice-Hall.

Andersen, P., & Nordvik, H. (2002). Possible Barnum effect in the Five Factor model: Do respondents accept random NEO Personality Inventory-Revised scores as their actual trait profile? *Psychological Reports, 90*(2), 539-545.

Artz, A. F., & Armour-Thomas, E. (1992). Development of a cognitive-metacognitive framework for protocol analysis of mathematical problem solving in small groups. *Cognition & Instruction, 9*(2), 137-175.

Austin, E. J., Dreary, I. J., & Eber, H. W. (2002). Interactions between intelligence and personality: Results from two large samples. *Personality and Individual Differences, 29*, 405-427.

Baker, S. (2001). Self-confidence within the calibration paradigm: An investigation of group and individual differences utilising memory and visual-spatial tasks. Unpublished Honours Thesis, University of Southern Queensland, Toowoomba.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review, 84*, 191-215.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Baranski, J. V., & Petrusic, W. M. (1994). The calibration and resolution of confidence in perceptual judgements. *Perception and Psychophysics, 55*(4), 412-428.

Baron, J. (1994). *Thinking and deciding* (2nd ed.). Cambridge: Cambridge University Press.

Bem, D. J. (1972). Self-perception theory. Advances in Experimental Social Psychology, 6.

Benenson, J. F., & Dweck, C. S. (1986). The development of trait explanations and self-evaluations in the academic and social domains. *Child Development, 57*(5), 1179-1187.

Beyer, S. (1990). Gender differences in the accuracy of self-evaluations of performance. *Journal of Personality and Social Psychology, 59*(5), 960-970.

Beyer, S., & Bowden, E. M. (1997). Gender differences in self-perceptions: Convergent evidence from three measures of accuracy and bias. *Personality and Social Psychology Bulletin*, *23*(2), 157-172.

Binderman, R. M., Fretz, B. R., Scott, N. A., & Abrams, M. H. (1972). Effects of interpreter credibility and discrepancy level of results on responses to test results. *Journal of Consulting Psychology, 19*(3), 399-403.

Bjorkman, M., Juslin, P., & Winman, A. (1993). Realism of confidence in sensory discrimination: The underconfidence phenomenon. *Perception and Psychophysics, 54*(1), 75-81.

Blackman, M. C., & Funder, D. C. (1998). The effect of information on consensus and accuracy in personality judgment. *Journal of Experimental Social Psychology, 34*, 164-181.

Blais, A., Thompson, M. M., & Baranski, J. V. (2005). Individual differences in decision processing and confidence in comparative judgment tasks: The role of cognitive styles. *Personality and Individual Differences, 38*, 1701-1703.

Blanton, H., Pelham, B. W., DeHart, T., & Carvallo, M. (2001). Overconfidence as dissonance reduction. *Journal of Experimental Social Psychology, 37*(5), 373-385.

Borkenau, P., & Liebler, A. (1993a). Consensus and self-other agreement for trait inferences from minimal information. *Journal of Personality, 61*(4), 477-496.

Borkenau, P., & Liebler, A. (1993b). Convergence of stranger ratings of personality and intelligence with self-ratings, partner ratings, and measured intelligence. *Journal of Personality and Social Psychology, 65*(3), 546-553.

Bornstein, B. H., & Zickafoose, D. J. (1999). "I know I know it, I know I saw it": The stability of the confidence-accuracy relationship across domains. *Journal of Experimental Psychology: Applied, 5*(1), 76-88.

Braun, P. A., & Yaniv, J. F. (1992). A case study of expert judgement: Economist's probabilities versus base-rate model forecasts. *Journal of Behavioral Decision Making, 5*, 217-231.

Brenner, L. A., Koehler, D. J., Liberman, V., & Tversky, A. (1996). Overconfidence in probability and frequency judgements: A critical examination. *Organizational Behavior and Human Decision Processes, 65*(3), 212-219.

Brief, A. P., Burke, M. J., George, J. M., Robinson, B. S., & Webster, J. (1988). Should negative affectivity remain an unmeasured variable in the study of job stress? *Journal of Applied Psychology, 73*(2), 193-198.

Briere, J. (1995). *Trauma Symptom Inventory*. Odessa, FL: Psychological Assessment Resources.

Briggs, P., Burford, B., & Dracup, C. (1998). Modelling self-confidence in users of a computer-based system showing unrepresentative design. *International Journal of Human-Computer Studies, 49*, 717-742.

Brown, A. (1987). Metacognition, executive control, self-regulation and other more mysterious mechanisms. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation and understanding* (pp. 65-116). Hillsdale, NJ: Lawrence Erlbaum Associates.

Budescu, D. V., Wallsten, T. S., & Au, W. (1997). On the importance of random error in the study of probability judgment. Part II: Using the stochastic judgment model to detect systematic trends. *Journal of Behavioral Decision Making, 10,* 173-188.

Burke, M. J., Brief, A. P., & George, J. M. (1993). The role of negative affectivity in understanding relations between self-reports of stressors and strains: A comment on the applied psychology literature. *Journal of Applied Psychology, 78*(3), 402-412.

Burnkrant, R. E., & Page, T. J. (1984). A modification of the Fenigstein, Scheier, and Buss Self-Consciousness Scales. *Journal of Personality Assessment, 48*(6), 629-637.

Burson, K. A., Larrick, R. P., & Klayman, J. (2006). Skilled or unskilled, but still unaware of it: How perceptions of difficulty drive mis-calibration in relative comparisons. *Journal of Personality and Social Psychology, 90*(1), 60-77.

Buss, D. M., & Craik, K. H. (1983). The act frequency approach to personality. *Psychological Review, 90*(2), 105-126.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116-131.

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119*(2), 197-253.

Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*(3), 306.

Cacioppo, J. T., Petty, R. E., Kao, C. F., & Rodriguez, R. (1986). Central and peripheral routes to persuasion: An individual differences perspective. *Journal of Personality and Social Psychology, 51,* 1032-1043.

Campbell, W. K., Adam, S. G., & Joshua, D. F. (2004). Narcissism, confidence, and risk attitude. *Journal of Behavioral Decision Making, 17*(4), 297-311.

Carroll, J. B. (1993). Human cognitive abilities: A survey of factor-analytical studies. Cambridge: Cambridge University Press.

Cattell, R. B., Eber, H. W., & Tatsuoka, M. M. (1992). *Handbook for the 16PF.* Champaign, IL: Institute for Personality and Ability Testing.

Chamorro-Premuzic, T., & Furhnam, A. (2004). A possible model for explaining the personality-intelligence interface. *British Journal of Psychology, 95*, 249-264.

Chamorro-Premuzic, T., & Furnham, A. (2008). Personality, intelligence and approaches to learning as predictors of academic performance. *Personality and Individual Differences, 44,* 1596-1603.

Christensen-Szalanski, J. J., & Bushyhead, J. B. (1981). Physician's use of probabilistic information in a real clinical setting. *Journal of Experimental Psychology: Human Perception and Performance, 7*(4), 928-935.

Christiansen, N. D., Wolcott-Burnam, S., Janovics, J. E., Burns, G. N., & Quirk, S. W. (2005). The good judge revisited: Individual differences in the accuracy of personality judgments. *Human Performance, 18*(2), 123-149.

Cloninger, C. R., Przybeck, T. R., Svrakic, D. M., & Wetzel, R. D. (1994). *The temperament character inventory (TCI): A guide to its development and use*. St Louis, MO: Washington University.

Coakes, S. J., & Steed, L. G. (1996). *SPSS for windows: Analysis without anguish*. Brisbane, QLD: Wiley.

Colvin, C. R., Block, J., & Funder, D. C. (1995). Overly positive self-evaluations and personality: Negative implications for mental health. *Journal of Personality and Social Psychology, 68*(6), 1152-1162.

Colvin, C. R., & Funder, D. C. (1991). Predicting personality and behavior: A boundary on the acquaintanceship effect. *Journal of Personality and Social Psychology, 60*(6), 884-894.

Costa, P. T., Jr., & McCrae, R. R. (1991). *Manual of revised NEO personality inventory and NEO Five-Factor Inventory*. Odessa, FL: Psychological Assessment Resources.

Crawford, J. D., & Stankov, L. (1996). Age differences in the realism of confidence judgements: A calibration study using tests of fluid and crystallized intelligence. *Learning and Individual Differences, 8*(2), 83-103.

Davies, M. F. (1997). Positive test strategies and confirmatory retrieval processes in the evaluation of personality feedback. *Journal of Personality and Social Psychology, 73*(3), 574-583.

Demetriou, A., & Kazi, S. (2001). Unity and modularity in the mind and the self: Studies on the relationships between self-awareness, personality, and intellectual development from childhood to adolescence. London: Routledge.

Dimaggio, G., Carcione, A., Petrilli, M., Procacci, M., Semerari, A., & Nicolo, G. (2005). Poor metacognition in narcissistic and avoidant personality disorders: Analysis of four psychotherapy patients. *Clinical Psychology and Psychotherapy, 14,* 386-401.

Dimaggio, G., Semerari, A., Carcione, A., Procacci, M., & Nicolo, G. (2006). Toward a model of self-pathology underlying personality disorders: Narratives, metacognition, interpersonal cycles and decision making processes. *Journal of Personality Disorders, 20*, 597-617.

Ditto, P. H., & Boardman, A. F. (1995). Perceived accuracy of favourable and unfavourable psychological feedback. *Basic and Applied Social Psychology, 16*, 137-157.

Dunning, D. (2005). Self-insight roadblocks and detours on the path to knowing thyself. New York: Psychology Press.

Dunning, D., Griffin, D. W., Milojkovic, J. D., & Ross, L. (1990). The overconfidence effect in social prediction. *Journal of Personality and Social Psychology, 58*(4), 568-581.

Dunning, D., Johnson, K., Ehrlinger, J., & Kruger, J. (2003). Why people fail to recognize their own incompetence. *Current Directions in Psychological Science, 12*(3), 83-87.

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (1989). Ambiguity and self-evaluation: The role of idiosyncratic trait definitions in self-serving assessments of ability. *Journal of Personality and Social Psychology, 57*(6), 1082-1090.

Dunning, D., & Story, A. L. (1991). Depression, realism, and the overconfidence effect: Are the sadder wiser when predicting future actions and events? *Journal of Personality and Social Psychology, 61*(4), 521-532.

Dweck, C. S., Chiu, C., & Hong, Y. (1995a). Implicit theories and their role in judgments and reactions: A world from two perspectives. *Psychological Inquiry, 6*(4), 267-285.

Dweck, C. S., Chiu, C., & Hong, Y. (1995b). Implicit theories: Elaboration and extension of the model. *Psychological Inquiry, 6*(4), 322-333.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95*(2), 256-273.

Eagly, A. H. (1987). Sex differences in social behavior: A social-role interpretation. Hillsdale, NJ: Erlbaum.

Eagly, A. H., & Wood, W. (1991). Explaining sex differences in social behavior: A meta-analytic perspective. *Personality and Social Psychology Bulletin, 17*(3), 306-315.

Edwards, A. L. (1953). The relationship between the judged desirability of a trait and the probability that the trait will be endorsed. *Journal of Applied Psychology, 37*(2), 90-93.

Ehrlinger, J., & Dunning, D. (2003). How chronic self-views influence (and potentially mislead) estimates of performance. *Journal of Personality and Social Psychology, 84*(1), 5-17.

Elliott, T. R., Chartrand, J. M., & Harkins, S. W. (1994). Negative affectivity, emotional distress, and the cognitive appraisal of occupational stress. *Journal of Vocational Behavior, 45*(2), 185-201.

Erev, I., Wallsten, T. S., & Budescu, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgement processes. *Psychological Review, 101*(3), 519-527.

Feingold, A. (1994). Gender differences in personality: A meta-analysis. *Psychological Bulletin, 116*(3), 429-456.

Fenigstein, A., Scheier, M. F., & Buss, A. H. (1975). Public and private self-consciousness: Assessment and theory. *Journal of Consulting and Clinical Psychology, 43*(4), 522-527.

Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row and Peterson.

Flavell, J. H. (1977). *Cognitive development*. Englewood Cliffs, NJ: Prentice-Hall.

Flavell, J. H. (1987). Speculations about the nature and development of metacognition. In F. E. Weinert & R. H. Kluwe (Eds.), *Metacognition, motivation, and understanding* (pp. 21-29). Hillsdale, NJ: Lawrence Erlbaum Associates.

Fogarty, G., & Else, D. (2005). Performance calibration in sport: Implications for self-confidence and metacognitive biases. *International Journal of Sport and Exercise Psychology, 3*(1), 41-57.

Fogarty, G., Graham, C., & Else, D. (2001). Self-efficacy, confidence judgments, and self-monitoring in golfers. In P. Thomas (Ed.), *Optimising performance in golf* (pp. 300-310). Brisbane: Australian Academic Press.

Fogarty, G., Machin, M. A., Albion, M., Sutherland, L. F., Lalor, G. I., & Revitt, S. (1999). Predicting occupational strain and job satisfaction: The role of stress, coping, personality, and affectivity variables. *Journal of Vocational Behavior, 54*(3), 429-452.

Fogarty, G., & Ross, A. (2007, September). *Calibration in tennis: The role of feedback and expertise.* Paper presented at the 42nd Annual Australian Psychological Society Conference, Brisbane.

Forsterlee, R., & Ho, R. (1999). An examination of the short form of the need for cognition scale applied in an Australian sample. *Educational and Psychological Measurement, 59*(3), 471-480.

Friedman, H. S., & Schustack, M. W. (2009). *Personality: Classic theories and modern research* (4th ed.). New York: Allyn & Bacon.

Fromson, P. M. (2006). Self-discrepancies and negative affect: The moderating roles of private and public self-consciousness. *Social Behavior & Personality: An International Journal, 34*(4), 333-349.

Funder, D. C. (1980). On seeing ourselves as others see us: Self-other agreement and discrepancy in personality ratings. *Journal of Personality, 48*(4), 473-493.

Funder, D. C. (1999). On the accuracy of personality assessment: A realistic approach. San Diego, CA: Academic Press.

Funder, D. C. (2001). Accuracy in personality assessment: Some research and theory concerning an obvious question. In B. W. Roberts & R. Hogan (Eds.), *Personality Psychology in the Workplace* (pp. 121-140). Washington, DC: American Psychological Association.

Funder, D. C., & Colvin, C. R. (1988). Friends and strangers: Acquaintanceship, agreement, and the accuracy of personality judgment. *Journal of Personality and Social Psychology, 55*(1), 149-158.

Funder, D. C., & Colvin, C. R. (Eds.). (1997). *Congruence of others' and self-judgments of personality*. New York: Academic Press.

Funder, D. C., Kolar, D. W., & Blackman, M. C. (1995). Agreement among judges of personality: Interpersonal relations, similarity, and acquaintanceship. *Journal of personality and Social Psychology, 69*(4), 656-672.

Funder, D. C., & Sneed, C. D. (1993). Behavioral manifestations of personality: An ecological approach to judgmental accuracy. *Journal of Personality and Social Psychology, 64*(3), 479-490.

Gigerenzer, G. (1991). Does the environment have the same structure as Bayes theorem? *Behavioural and Brain Sciences, 14*(3), 495-495.

Gigerenzer, G. (1992). Cognitive illusions illusory? Rethinking judgment under uncertainty. *International Journal of Psychology, 27*(3-4), 172-172.

Gigerenzer, G. (1993). The bounded rationality of probabilistic mental models. In K. I. Manktelow & D. E. Over (Eds.), *Rationality: Psychological and Philosophical Perspectives* (pp. 284-313). London, UK: Routledge.

Gigerenzer, G. (1996a). On narrow norms and vague heuristics: A reply to Kahneman and Tversky. *Psychological Review, 103*(3), 592-596.

Gigerenzer, G. (1996b). Why do frequency formats improve Bayesian reasoning? Cognitive algorithms work on information, which needs representation. *Behavioural and Brain Sciences, 19*(1), 23-45.

Gigerenzer, G. (2000). Ecological rationality. *International Journal of Psychology, 35*(3-4), 284-284.

Gigerenzer, G., Hoffrage, U., & Kleinbolting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review, 98*(4), 506-528.

Gigerenzer, G., & Murray, D. J. (1994). Two revolutions - Cognitive and probabilistic: Is the mind a Bayesian? In H. Gutfreund & G. Toulouse (Eds.), *Biology and computation: A physicist's choice. Advanced series in neuroscience* (Vol. 3). Singapore: World Scientific Publishing Co.

Gignac, G. E., Stough, C., & Loukomitis, S. (2004). Openness, intelligence, and self-report intelligence. *Intelligence, 32*(2), 133-143.

Gilovich, T., Griffin, D. W., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive assessment*. New York: Cambridge University Press.

Glenberg, A. M., & Epstein, W. (1987). Inexpert calibration of comprehension. *Memory and Cognition, 15*(1), 84-93.

Goldberg, L. (1999). A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt & F. Ostendorf (Eds.), *Personality Psychology in Europe* (Vol. 7, pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.

Goldberg, L. (Ed.). (1997). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. The Netherlands: Tilburg University Press.

Gosling, S. D., John, O. P., Craik, K. H., & Robins, R. W. (1998). Do people know how they behave? Self-reported act frequencies compared with on-line codings by observers. *Journal of Personality and Social Psychology, 74*(5), 1337-1349.

Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist, 59*(2), 93-104.

Gough, H. G. (1996). *CPI Manual: Third Edition*. Palo Alto, CA: Consulting Psychologists Press.

Graham, C. (2006). Calibration in sport: The role of feedback, age and gender. Unpublished masters thesis. University of Southern Queensland, Toowoomba.

Granhag, P. A., Stromwall, L. A., & Allwood, C. M. (2000). Effects of reiteration, hindsight bias, and memory on realism in eyewitness confidence. *Applied Cognitive Psychology, 14*(5), 397-420.

Greene, R. L., Bausom, D. H., & Macon, R. S. (1980). Students' acceptance of high and low generalized personality interpretations. *Journal of Clinical Psychology, 36*(1), 166-170.

Gregory, R. J. (1996). Psychological testing: History, principles, and applications. Bostyn: Allyn & Bacon.

Griffin, D., & Tversky, A. (1992). The weighting of evidence and the determinants of confidence. *Cognitive Psychology, 24*, 411-435.

Halperin, K., Snyder, C. R., Shenkel, R. J., & Houston, B. K. (1976). Effects of source status and message favourability on acceptance of personality feedback. *Journal of Applied Psychology, 61*(1), 85-88.

Halpern, D. F., & LaMay, M. L. (2000). The smarter sex: A critical review of sex differences in intelligence. *Educational Psychology Review, 12*(2), 229-246.

Hampson, S. E., Goldberg, L. R., Vogt, T. M., & Dubanoski, J. P. (2006). Forty years on: Teachers' assessments of children's personality traits predict self-reported health behaviors and outcomes at midlife. *Health Psychology, 25*(1), 57-64.

Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research, 52,* 123-142.

Harris, M. E., & Greene, R. L. (1984). Students' perception of actual, trivial, and inaccurate personality feedback. *Journal of Personality Assessment, 48*(2), 179-183.

Harvey, N. (1997). Confidence in judgement. *Trends in Cognitive Science, 1*(2), 78-82.

Hathaway, S., & McKinley, J. (1989). *Minnesota Multiphasic Personality Inventory (MMPI-2).* Columbus, OH: Merrill/Prentice-Hall.

Heppner, P. P., Reeder, B. L., & Larson, L. M. (1983). Cognitive variables associated with personal problem solving: Implications for counselling. *Journal of Counselling Psychology, 30*, 537-545.

Hong, Y., Chiu, C., Dweck, C. S., & Sacks, R. (1997). Implicit theories and evaluative processes in person cognition,. *Journal of Experimental Social Psychology, 33*(3), 296-323.

Horn, J. L. (1985). Remodelling old models of intelligence. In B. B. Wolman (Ed.), *Handbook of intelligence: Theories, measurements and applications* (pp. 267-300). NY: Wiley.

Horn, J. L. (1988). Thinking about human abilities. In J. R. Nesselroade (Ed.), *Handbook of multivariate experimental psychology* (pp. 645-685). New York: Academic Press.

Horn, J. L., & Cattell, R. B. (1966). Refinement of the theory of fluid and crystallized intelligence. *Journal of Educational Psychology, 57*, 253-270.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*(6), 581-592.

Jackson, D. E., & Murray, B. S. (1985). Predicting accuracy and liking ratings for bogus and real personality feedback. *The Journal of Psychology, 119*(5), 495-503.

John, O. P., & Robins, R. W. (1993). Determinants of interjudge agreement: The big five personality domains, observability, evaluativeness, and the unique perspective of the self. *Journal of Personality, 61*(4), 521-551.

John, O. P., & Robins, R. W. (1994). Accuracy and bias in self-perception: Individual differences in self-enhancement and the role of narcissism. *Journal of Personality and Social Psychology, 66*, 206-219.

Jonsson, A., & Allwood, C. M. (2003). Stability and variability in the realism of confidence judgments over time, content domain, and gender. *Personality and Individual Differences, 34*(4), 559-574.

Juslin, P. (1993). An explanation of the hard-easy effect in studies of realism of confidence in one's general knowledge. *European Journal of Cognitive Psychology, 5*(1), 55-71.

Juslin, P. (1994). The overconfidence phenomenon as a consequence of informal experimenter-guided selection of almanac items. *Organizational Behavior and Human Decision Processes, 57*, 226-246.

Juslin, P., & Olsson, H. (1997). Thurstonian and Brunswikian origins of uncertainty in judgement: A sampling model of confidence in sensory discrimination. *Psychological Review, 104*(2), 344-366.

Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgments under uncertainty: Heuristics and biases*. Cambridge, UK: Cambridge University Press.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. *Psychological Review*, *80*(4), 237-251.

Kaufman, A. S., & Horn, J. L. (1996). Age changes on tests of fluid and crystallized ability for women and men on the Kaufman Adolescent and Adult Intelligence Test (KAIT) at ages 17-94 years. *Archives of Clinical Neuropsychology, 11*(2), 97-121.

Keren, G. (1987). Facing uncertainty in the game of bridge. *Organizational Behavior and Human Decision Processes, 39*, 98-114.

Keren, G. (1991). Calibration and probability judgments: Conceptual and methodological issues. *Acta Psychologica, 77*, 217-273.

Kidd, J. B. (1970). The utilization of subjective probabilities in production planning. *Acta Psychologica, 34*, 338-347.

Kihlstrom, J. F., Beer, J. S., & Klein, S. B. (2003). Self and identity as memory. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 68-90). New York, NY: Guilford Press.

Klayman, J., Soll, J. B., González-Vallejo, C., & Barlas, S. (1999). Overconfidence: It depends on how, what, and whom you ask. *Organizational Behavior and Human Decision Processes, 79*(3), 216-247.

Klein, S. B. (2004). The cognitive neuroscience of knowing one's self. In M. S. Gazzaniga (Ed.), *The cognitive neurosciences* (3 rd. ed., pp. 1077-1089). Cambridge, MA, US: MIT Press.

Klein, S. B., Chan, R. L., & Loftus, J. (1999). Independence of episodic and semantic self-knowledge: The case from autism. *Social Cognition, 17,* 413-436.

Klein, S. B., Cosmides, L., & Costabile, K. A. (2003). Preserved knowledge of self in Alzheimer's dementia. *Social Cognition, 21*, 157-165.

Klein, S. B., Cosmides, L., Murray, E. M., & Tooby, J. (2004). On the acquisition of knowledge about personality traits: Does learning about the self engage different mechanisms than learning about others? *Social Cognition, 22*(4), 367-390.

Klein, S. B., Cosmides, L., Tooby, J., & Chance, S. (2001). Priming exceptions: A test of the scope hypothesis in naturalistic trait judgments. *Social Cognition, 19*(4), 443-468.

Klein, S. B., & Loftus, J. (1990). The role of abstract and exemplar-based knowledge in self-judgments: Implications for a cognitive model of the self. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition* (Vol. 3, pp. 131-139). Hillsdale, NJ: Erlbaum.

Klein, S. B., & Loftus, J. (1993a). Behavioural experience and trait judgments about the self. *Personality and Social Psychology Bulletin, 19*, 740-745.

Klein, S. B., & Loftus, J. (1993b). The role of abstract and exemplar-based knowledge in self-judgments: Implications for a cognitive model of the self. In T. K. Srull & R. S. Wyer (Eds.), *Advances in social cognition* (Vol. 5). Hillsdale, NJ: Erlbaum.

Klein, S. B., Loftus, J., & Burton, H. A. (1989). Two self-reference effects: The importance of distinguishing between self-descriptiveness judgments and autobiographical retrieval in self-referent encoding. *Journal of Personality and Social Psychology, 56*(6), 853-865.

Klein, S. B., Loftus, J., & Kihlstrom, J. F. (1996). Self-knowledge of an amnesic patient: Toward a neuropsychology of personality and social psychology. *Journal of Experimental Psychology: General, 125*(3), 250-260.

Klein, S. B., Loftus, J., & Plog, A. E. (1992). Trait judgments about the self: Evidence from the encoding specificity paradigm. *Personality and Social Psychology Bulletin, 18*, 730-735.

Klein, S. B., Loftus, J., Trafton, J. G., & Fuhrman, R. W. (1992). Use of exemplars and abstractions in trait judgments: A model of trait knowledge about the self and others. *Journal of Personality and Social Psychology, 63*(5), 739-753.

Klein, S. B., Rozendale, K., & Cosmides, L. (2002). A social-cognitive neuroscience analysis of the self. *Social Cognition, 20*, 105-135.

Kleitman, S. (2003). Self confidence and self-monitoring aspects of metacognition, their nature and their status within rationality debates: An individual differences perspective. Unpublished doctoral dissertation. University of Sydney.

Kleitman, S. (2008). *Metacognition in the rationality debate: Self-confidence and its calibration*. Saarbrucken, Germany: VDM Verlag Dr Muller Aktiengesellschaft & Co. KG.

Kleitman, S., & Stankov, L. (2001). Ecological and person-oriented aspects of metacognitive processes in test-taking. *Applied Cognitive Psychology, 15*(3), 321-341.

Kleitman, S., & Stankov, L. (2007). Self-confidence and metacognitive processes. *Learning and Individual Differences, 17*(2), 161-173.

Koehler, D. J. (1994). Hypothesis generation and confidence in judgment. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 20*(2), 461-469.

Koehler, D. J., Brenner, L. A., Liberman, V., & Tversky, A. (1996). Confidence and accuracy in trait inference: Judgment by similarity. *Acta Psychologica, 92*(1), 33-57.

Kolar, D. W., Funder, D. C., & Colvin, C. R. (1996). Comparing the accuracy of personality judgments by the self and knowledgeable others. *Journal of Personality, 64*(2), 311-337.

Koriat, A., Lichtenstein, S., & Fischhoff, B. (1980). Reasons for confidence. *Journal of Experimental Psychology: Human Learning and Memory, 6*(2), 107-118.

Krueger, J., & Mueller, R. A. (2002). Unskilled, unaware, or both? The better-than-average heuristic and statistical regression predict errors in estimates of own performance. *Journal of Personality and Social Psychology, 82*(2), 180-188.

Kruger, J. (1999). Lake Wobegone be gone! The 'below-average effect' and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology, 77*(2), 221-232.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology, 77*(6), 1121-1134.

Leary, M. R. (2007). Motivational and emotional aspects of the self. *Annual Review of Psychology, 58*(1), 317-344.

Letzring, T. D., Wells, S. M., & Funder, D. C. (2006). Information quantity and quality affect the realistic accuracy of personality judgment. *Journal of Personality and Social Psychology, 91*(1), 111-123.

Levesque, M. J. (1997). Meta-accuracy among acquainted individuals: A social relations analysis of interpersonal perception and metaperception. *Journal of Personality and Social Psychology, 72*(1), 66-74.

Levin, I. P., Huneke, M. E., & Jasper, J. D. (2000). Information processing at successive stages of decision making: Need for cognition and inclusion-exclusion effects. *Organizational Behavior and Human Decision Processes, 82*, 171-193.

Liberman, V. (2004). Local and global judgments of confidence. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*(3), 729-732.

Liberman, V., & Tversky, A. (1993). On the evaluation of probability judgements: Calibration, resolution, and monotonicity. *Psychological Bulletin, 111*(1), 162-173.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance, 20*, 159-183.

Lichtenstein, S., Fischhoff, S., & Phillips, L. (Eds.). (1982). *Calibration of probabilities: The state of the art in 1980.* Cambridge: Cambridge University Press.

Lin, L., & Zabrucky, K. M. (1998). Calibration of comprehension: Research and implications for education and instruction. *Contemporary Educational Psychology, 23*, 345-391.

Loranger, A. W. (2001). *OMNI Personality Inventories*. Odessa, FL: Psychological Assessment Resources.

Mabe, P. A., & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology, 67*(3), 280-296.

Mahajan, J. (1992). The overconfidence effect in marketing management predictions. *Journal of Marketing Research, 29*(3), 329-342.

Maki, R. H., Jonas, D., & Kallod, M. (1994). The relationship between comprehension and metacomprehension ability. *Psychonomic Bulletin and Review, 1*, 126-129.

Marks, P. A., & Seeman, W. (1962). On the Barnum effect. *The Psychological Record, 12*, 203-208.

Marsh, H. W. (2008). A multidimensional, hierarchical model of self-concept. In G. J. Boyle, G. Matthews & D. Saklofske (Eds.), *The Sage Handbook of Personality Theory and Assessment* (pp. 447-469): Sage publications.

Marsh, H. W., & Yeung, A. S. (1998). Longitudinal structural equation models of academic self-concept and achievement: Gender differences in the development of math and english constructs. *American Educational Research Journal, 35*(4), 705-738.

McCrae, R. R., & Costa, P. T. (1983). Social desirability scales: More substance than style. *Journal of Consulting and Clinical Psychology, 51*(6), 882-888.

McCrae, R. R., & Costa, P. T., Jr. (1999). A five-factor theory of personality. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality theory and research* (Vol. 2, pp. 139-153). New York: Guilford Press.

McGraw, K. O., Tew, M. D., & Williams, J. E. (2000). The integrity of web-delivered experiments: Can you trust the data? *Psychological Science, 11*(6).

Meehl, P. E. (1956). Wanted - A good cookbook. *American Psychologist, 41*, 26-272.

Meyer, G. J., Finn, S. E., Eyde, L. D., Kay, G. G., Moreland, K. L., Dies, R. R., et al. (2001). Psychological testing and psychological assessment: A review of evidence and issues. *American Psychologist, 56*(2), 128-165.

Millon, T. R., Davis, R., & Millon, C. (1997). *Millon Clinical Multiaxial Inventory-III manual (MCMI-III)*. Minneapolis, MN: Pearson Assessments.

Moore, D. A., & Healy, P. J. (2008). The trouble with overconfidence. *Psychological Review, 115*(2), 502-517.

Moreland, R., Miller, J., & Laucka, F. (1981). Academic achievement and self-evaluations of academic performance. *Journal of Educational Psychology, 73,* 335-344.

Morey, L. C. (1991). *Personality Assessment Inventory: Professional manual*. Odessa, FL: Psychological Assessment Resources.

Morey, L. C. (2003). *Essentials of PAI Assessment*. Hoboken, NJ: John Wiley & Sons.

Mosher, D. L. (1965). Approval motive and acceptance of fake personality test interpretations which differ in favorability. *Psychological Reports, 17*, 395-402.

Mount, M. K., Barrick, M. R., & Strauss, J. P. (1994). Validity of observer ratings of the big five personality factors. *Journal of Applied Psychology, 79*(2), 272-280.

Murphy, A. H., & Winkler, R. L. (1977). Can weather forecasters formulate reliable probability forecasts of precipitation and temperature? *National Weather Digest, 2*, 2-9.

Nevid, J. S. (1983). Hopelessness, social desirability, and construct validity. *Journal of Consulting and Clinical Psychology, 51*(1), 139-140.

Newman, R. S. (1984). Children's numerical skill and judgements of confidence in estimation. *Journal of Experimental Child Psychology, 37*, 107-123.

Newman, R. S., & Wick, P. L. (1987). Effect of age, skill, and performance feedback on children's judgements of confidence. *Journal of Educational Psychology, 79*(2), 115-119.

Nichols, D. S. (2001). *Essentials of MMPI-2 Assessment*. New York, NY: John Wiley & Sons.

Nichols, D. S., & Greene, R. L. (1997). Dimensions of deception in personality assessment: The example of the MMPI-2. *Journal of Personality Assessment, 68*(2), 251-266.

Nichols, D. S., Greene, R. L., & Schmolck, P. (1989). Criteria for assessing inconsistent patterns of item endorsement on the MMPI: Rationale, development, and empirical trails. *Journal of Clinical Psychology, 45*(2), 239-250.

Nolan, J., & Markham, R. (1998). The accuracy-confidence relationship in an eyewitness task: Anxiety as a modifier. *Applied Cognitive Psychology, 12*, 43-54.

O'Connor, M., & Paunonen, S. V. (2007). Big five personality predictors of post-secondary academic performance. *Personality and Individual Differences, 43*, 971-990.

O'Keefe, K. M., Wildemuth, B. M., & Freidman, C. P. (1999). Medical students' confidence judgments using a factual database and personal memory: A comparison. *Journal of the American Society for Information Science, 8*, 698-708.

Ochsner, K. N., Beer, J. S., Robertson, E. R., Cooper, J. C., Gabrieli, J. D. E., Kihlstrom, J. F., et al. (2005). The neural correlates of direct and reflected self-knowledge. *NeuroImage, 28*(4), 797-814.

Ones, D. S., Viswesvaran, C., & Reiss, A. D. (1996). Role of social desirability in personality testing for personnel selection: The red herring. *Journal of Applied Psychology, 81*(6), 660-679.

Oskamp, S. (1965). Overconfidence in case-study judgments. *Journal of Consulting Psychology, 29*(3), 261-265.

Pallier, G. (2003). Gender differences in the self-assessment of accuracy on cognitive tasks. *Sex Roles, 48*(5-6), 265-276.

Pallier, G., Wilkinson, R., Danthiir, V., Kleitman, S., Knezevic, S., Stankov, L., et al. (2002). The role of individual differences in the realism of confidence judgements. *The Journal of General Psychology, 129*, 257-299.

Paulhus, D. L., & Morgan, K. L. (1997). Perceptions of intelligence in leaderless groups: The dynamic effects of shyness and acquaintance. *Journal of Personality and Social Psychology, 72*(3), 581-591.

Petrusic, W. M., & Baranski, J. V. (1997). Context, feedback, and the calibration and resolution of confidence in perceptual judgements. *American Journal of Psychology*, *110*(4), 543-572.

Pulford, B. D., & Colman, A. M. (1997). Overconfidence: Feedback and item difficulty effects. *Personality and Individual Differences, 23*(1), 125-133.

Reeve, C. L., Meyer, R. D., & Bonaccio, S. (2006). Intelligence-personality associations reconsidered: The importance of distinguishing between general and narrow dimensions of intelligence. *Intelligence, 34*(4), 387-402.

Renner, C. H., & Renner, M. J. (2001). But I thought I knew that: Using confidence estimation as a debiasing technique to improve classroom performance. *Applied Cognitive Psychology, 15*, 23-32.

Robins, R. W., Paulhus, D. L., Roberts, B. W., & Hogan, R. (2001). The character of self-enhancers: Implications for organizations. In B. W. Roberts & R. Hogan (Eds.), *Personality psychology in the workplace* (pp. 193-219). Washington, DC: American Psychological Association.

Ronis, D. L., & Yates, J. F. (1987). Components of probability judgement accuracy: Individual consistency and effects of subject matter and assessment method. *Organisational Behavior and Human Decision Processes, 61*, 148-157.

Ross, A., & Fogarty, G. (2006). *Calibration in tennis: The role of feedback, gender, and expertise.* University of Southern Queensland, Toowoomba.

Sadowski, C. J., & Cogburn, H. E. (1997). Need for cognition in the Big-Five factor structure. *Journal of Psychology: Interdisciplinary and Applied, 131*(3), 307-312.

Saucier, G. (2002). Orthogonal markers for orthogonal factors: The case of the big five. *Journal of Research in Personality, 36*(1), 1-31.

Saucier, G., & Goldberg, L. (2002). Assessing the Big Five: Applications of 10 psychometric criteria to the development of marker scales. In B. de Raad & M. Perugini (Eds.), *Big Five Assessment*. Goettinger, Germany: Hogrefe & Hubner.

Saucier, G., & Goldberg, L. (Eds.). (1996). The language of personality: Lexical perspectives on the Five-Factor model. New York: Guilford Press.

Schaefer, P. S., Williams, C. C., Goodie, A. S., & Campbell, W. K. (2004). Overconfidence and the Big Five. *Journal of Research in Personality, 38*(5), 473-480.

Schneider, S. L. (1995). Item difficulty, discrimination, and the confidence-frequency effect in a categorical judgement task. *Organizational Behavior and Human Decision Processes, 61*, 148-167.

Schraw, G. (1994). The effect of metacognitive knowledge on local and global monitoring. *Contemporary Educational Psychology, 19*(2), 143-154.

Schraw, G. (1997). The effect of generalised metacognitive knowledge on test performance and confidence judgements. *The Journal of Experimental Education, 65*(2), 135-146.

Schraw, G., Crippen, K. L., & Hartley, K. (2006). Promoting self-regulation in science education: Metacognition as part of a broader perspective on learning. *Research in Science Education, 36*, 111-139.

Schraw, G., & DeBacker Roedel, T. (1994). Test difficulty and judgement bias. *Memory & Cognition, 22*(1), 63-69.

Schraw, G., & Dennison, R. S. (1994). Assessing metacognitive awareness. *Contemporary Educational Psychology, 19*, 460-475.

Schraw, G., Dunkle, M. E., & Bendixen, L. D. (1995). Does a general monitoring skill exist? *Journal of Educational Psychology*, *87*(3), 433-444.

Schraw, G., & Moshman, D. (1995). Metacognitive theories. *Educational Psychology Review, 7*(4), 351-371.

Sedikides, C., & Strube, M. J. (1997). Self-evaluation: To thine own self be good, to thine own self be sure, to thine own self be true, and to thine own self be better. *Advances in Experimental Social Psychology*, *29*, 209-269.

Sharp, G. L., Cutler, B. L., & Penrod, S. D. (1988). Performance feedback improves the resolution of confidence judgements. *Organizational Behavior and Human Decision Processes, 42,* 271-283.

Shaughnessy, J. J. (1979). Confidence judgment accuracy as a predictor of test performance. *Journal of Research in Personality*, *13*, 505-514.

Shavelson, R. J., Hubner, J. J., & Stanton, G. C. (1976). Self-concept: Validation of construct interpretations. *Review of Educational Research, 46*, 407-441.

Shrauger, J. S., & Schoeneman, J. (1979). Symbolic view of self-concept: Through the looking glass darkly. *Psychological Bulletin, 86*(3), 549-573.

Silvera, D. H., & Seger, C. R. (2004). Feeling good about ourselves: Unrealistic self-evaluations and their relation to self-esteem in the United States and Norway. *Journal of Cross-Cultural Psychology, 35*(5), 571-585.

Snyder, C. R., & Larson, G. R. (1972). A further look at student acceptance of general personality interpretations. *Journal of Consulting and Clinical Psychology, 38*(3), 384-388.

Snyder, C. R., & Shenkel, R. J. (1976). Effects of "favorability," modality and relevance on acceptance of general personality interpretations prior to and after receiving diagnostic feedback. *Journal of Consulting and Clinical Psychology, 44*(1), 34-41.

Snyder, C. R., Shenkel, R. J., & Lowery, C. R. (1977). Acceptance of personality interpretations: The "Barnum Effect" and beyond. *Journal of Consulting and Clinical Psychology, 45*(1), 104-114.

Soll, J. B. (1996). Determinants of overconfidence and mis-calibration: The roles of random error and ecological structure. *Organizational Behavior and Human Decision Processes*, *65*(2), 117-137.

Spence, M. T. (1996). Problem solver characteristics affecting the calibration of judgements. *Organizational Behavior and Human Decision Processes, 67*(3), 271-279.

SPSS. (2009). Statistical Package for Social Sciences for Windows (Version 16). Chicago: SPSS.

Stankov, L. (1997). The Gf/Gc Quickie Test Battery. Unpublished test battery. School of Psychology, University of Sydney.

Stankov, L. (1998). Calibration curves, scatterplots and the distinction between general knowledge and perceptual tasks. *Learning and Individual Differences, 10*(1), 29-50.

Stankov, L. (1999a). Accuracy, confidence, and speed in perceptual tasks: Calibration effects and individual differences. Unpublished manuscript. School of Psychology, University of Sydney.

Stankov, L. (1999b). Mining on the "no man's land" between intelligence and personality. In P. L. Ackerman, P. C. Kyllonen & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants* (pp. 314-347). Washington, DC: American Psychological Society.

Stankov, L. (2000a). Complexity, metacognition, and fluid intelligence. *Intelligence, 28*(2), 121-143.

Stankov, L. (2000b). Structural extensions of a hierarchical view on human cognitive abilities. *Learning and Individual Differences, 12*(1), 35-51.

Stankov, L., & Crawford, J. D. (1996a). Confidence judgements in studies of individual differences. *Personality and Individual Differences, 21*(6), 971-986.

Stankov, L., & Crawford, J. D. (1996b). Confidence judgements in studies of individual differences support the 'confidence/frequency effect'. In C. Latimer & J. Mitchell (Eds.), *At once scientific and philosophic: A festchrift in honour of J. P. Sutcliffe* (pp. 215-213). Sydney: University of Sydney Press.

Stankov, L., & Crawford, J. D. (1997). Self-confidence and performance on tests of cognitive abilities. *Intelligence, 25*(2), 93-109.

Stankov, L., & Dolph, B. (2000). Metacognitive aspects of test-taking and intelligence. *Psychologische-Beitrage, 42*, 213-227.

Stankov, L., & Kleitman, S. (2008). Processes on the borderline between cognitive abilities and personality: Confidence and its realism. In G. J. Boyle, G. Matthews & D. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment* (pp. 541-555): Sage.

Stankov, L., & Lee, J. (2007). *Confidence and cognitive test performance. Research Report 3*. Princeton, NJ.: Educational Testing Service.o. Document Number)

Stankov, L., & Lee, J. (2008). Confidence and cognitive test performance. *Journal of Educational Psychology, 100*(4), 961-976.

Sternberg, R. J. (1997a). *Thinking styles*. Cambridge, MA: University Press.

Sternberg, R. J. (1997b). The triarchic theory of intelligence. In D. P. Flanagan, J. L. Genshaft & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues*. New York: Guilford Press.

Stone, E. R., Dodrill, C. L., & Johnson, N. (2001). Depressive cognition: A test of depressive realism versus negativity using general knowledge questions. *Journal of Psychology, 135*(6), 583-602.

Swan, W. B., & Gill, M. J. (1997). Confidence and accuracy in person perception: Do we know what we think we know about our relationship partners? *Journal of Personality and Social Psychology, 73*(4), 474-757.

Tabachnick, B. G., & Fidell, L. S. (1996). *Using multivariate statistics* (3rd ed.). New York, NY: HarperCollins.

Taylor, S. E., & Brown, J. D. (1988). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103*(2), 193-210.

Taylor, S. E., Neter, E., & Wayment, H. A. (1995). Self-evaluation processes. *Personality and Social Psychology Bulletin, 21*(12), 1278-1287.

Tesser, A. (2003). Self-evaluation. In J. F. Kihlstrom, J. S. Beer & S. B. Klein (Eds.), *Handbook of self and identity*. New York, NY: Guilford.

Tomassini, L. A., Solomon, I., Romney, M. B., & Krostad, J. L. (1982). Calibration of auditors' probabilistic judgements: Some empirical evidence. *Organisational Behavior and Human Performance, 30*, 391-406.

Tulving, E. (1993). Self-knowledge of an amnesic individual is represented abstractly. In T. K. Srull & R. S. Wyer (Eds.), *Advances in Social Cognition* (Vol. 5, pp. 147-156). Hillsdale, NJ: Erlbaum.

Tversky, A., & Kahneman, D. (1983). Extensional versus intuitive reasoning: The conjunction fallacy in probability judgment. *Psychological Review, 90*, 293-315.

Vallone, R. P., Griffin, D. W., Lin, S., & Ross, L. (1990). Overconfident predictions of future actions and outcomes by self and others. *Journal of Personality and Social Psychology, 58*(4), 583-592.

Vogt, D. S., & Colvin, C. R. (2003). Interpersonal orientation and the accuracy of personality judgments. *Journal of Personality, 71*(2), 267-295.

Vreugdenhil, H., & Koele, P. (1988). Underconfidence in predicting future events. *Bulletin of the Psychonomic Society, 26*(3), 236-237.

Wagenaar, W. A., & Keren, G. B. (1985). Calibration of probability assessments by professional blackjack dealers, statistical experts, and lay people. *Organizational Behavior and Human Decision Processes, 36*, 406-416.

Walczyk, & Hall. (1989). Is the failure to monitor comprehension an instance of cognitive impulsivity? *Journal of Educational Psychology, 81*, 294-298.

Want, J., & Kleitman, S. (2006). Imposter phenomenon and self-handicapping: Links with parenting styles and self-confidence. *Personality and Individual Differences*, *40*(5), 961-971.

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS Scales. *Journal of Personality and Social Psychology, 54*(6), 1063-1070.

Watson, D., Hubbard, B., & Wiese, D. (2000). Self-other agreement in personality and affectivity: The role of acquaintanceship, trait visibility, and assumed similarity. *Journal of Personality and Social Psychology, 78*(3), 546-558.

West, R. F., & Stanovich, K. E. (1997). The domain specificity and generality of overconfidence: Individual differences in performance estimation bias. *Psychonomic Bulletin and Review, 4*(3), 387-392.

Wiggins, J. S. (1996). *The five-factor model of personality*. New York: Guilford.

Winman, A., Juslin, P., & Bjorkman, M. (1998). The confidence-hindsight mirror effect in judgement: An accuracy-assessment model for the knew-it-all-along phenomenon. *Journal of Experimental Psychology, 24*(2), 415-431.

Wolfe, R. N., & Grosch, J. W. (1990). Personality correlates of confidence in one's decisions. *Journal of Personality, 58*(3), 515.

Appendix A

Sign Up Sheet Study 1

Self-Confidence in Cognition and Personality

Area of Investigation:	Self-confidence in Cognition and Personality
Credit/Raffle	2%
Tasks:	Respond to 5 cognitive tasks and several personality measures
People required:	17 years and older
Name of Experimenter:	Sandra Baker
Contact Number:	0402070056 or 46311613
Supervisor:	Professor Gerry Fogarty
Ethics Approval #:	H02STU198
Book:	Click here to book

More Information: You are invited to participate in an experiment examining self-confidence in both cognition and personality. You will be asked to respond to three reasoning tasks and at the end of each trial you will be asked to provide a confidence rating. You will also be asked to provide confidence ratings for the personality measures. Specific instructions will be provided at the time of testing. Although your participation should give you no cause for concern, you may withdraw at any stage and will not be penalized. Upon completion of testing you will be provided with further information and any questions will be answered. Your identity as a participant in this research will remain confidential with respect to any publication of the results of the study. Any information that can identify you as a participant will be stored in a secured place, with the information available only to the investigator. The results will only be reported in their aggregate form. The USQ Human Research Ethics committee has approved this study and the approval number is listed above. This experiment should take two hours to complete. For participating, you will receive either course credit (2%) or a ticket in the departmental raffle. If you have any questions at any time regarding this research you may contact Sandra Baker on (07) 46311613 or Professor Gerry Fogarty on (07) 46312379

Appendix B

Confidence Ratings for Gf/Gc Quickie Tests

Confidence Ratings for Gf/Gc Quickie Tests (Stankov, 1997)

1) General Knowledge Test
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

2) Letter Series Test
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

3) Concealed Words Test
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

4) Esoteric Analogies
25% 50% 75% 100%

5) Cattell's Matrices Test
15% 30% 45% 60% 75% 90% 100%

6) Word Associations Test
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%

Appendix C

General Knowledge Test

General Knowledge Test (Stankov, 1997)

Directions:

Say “This is a test of your general knowledge. I will ask you a series of questions and I want you to give me the answers as best you can. After each answer I want you to say how confident you are that your answer is correct. In this test there are no multiple choice answers provided, so a guess would correspond to 0% confidence. If you are absolutely certain your answer is correct then you would say 100% confident. Please make your choice from the ratings provided on your sheet”.

If a response is incomplete or unclear, ask the participant to explain more fully, but if the response is “I Don’t know” record that as the answer.

You may repeat the question but do not spell out any words or alter the wording.

Write the answers on the person’s score sheet for later marking.

Questions:

1. How many weeks are there in a year?
2. What are BASIC, FORTRAN, and ALGOL?
3. What nationality was Picasso?
4. What is the capital of Austria?
5. At what temperature does water freeze?
6. Who wrote the Odyssey?
7. Where is Libya?
8. In Chemistry, what letter does the letter S stand for?
9. What was the nationality of Beethoven?
10. Who was the author of The Origin of the Species?
11. Name three kinds of blood vessels found in the human body?
12. Who was Confucius?
13. Who wrote King Lear?
14. Who was the President of the United States at the end of the Vietnam War?
15. What language is spoken in Brazil?
16. What is today’s name for the ancient city of Constantinople?
17. What is the main religion in Malaysia?
18. Who wrote Crime and Punishment?
19. Who was the founder of psychoanalysis?
20. Name one of the main languages spoken in the country formerly known as Yugoslavia?

Appendix D

Letter Series Test

(Stankov, 1997)

Directions:

Say " I am going to show you some letters. Your task is to say what should be the next letter in the series. Here are two examples". Show the participant the examples but cover the test items with a sheet of paper.

Say "You can see that in the first example the next letter in the series would be 'G'. In the second example the next letter would be 'A' because you must follow a rule that after 'Z' the alphabet starts again at 'A'. Do you understand the rule and what you must do?"

If the person says "No" repeat the instructions, otherwise **say** "After each answer I want you to tell me how confident you are that you are correct. A guess corresponds closely to 0% confidence so you should give this as your rating. Absolute certainty corresponds to 100% confidence. Please make your choice from the ratings provided on the sheet. You will have four minutes to complete this test. Please work as quickly and accuratley as you can". Remove the cover sheet, start timing and record the participant's responses on their score sheet.

After four minutes **say** "Please stop now, the time limit for this test is up".

see For example:

Example 1:	A	B	C	D	E	F	?
Example 2:	U	V	W	X	Y	Z	?

Trials:

1. J K L M N O P Q ?
2. C C Z C C Y C C X C C ?
3. P Q Q R R R S S S S ?
4. T R A T R B T R C T R ?
5. B C C D E E F G ?
6. O P Q O P Q R S T R S T U ?
7. L O M P N ?
8. A D G B E H C F ?
9. A X A Y B X B Y C X C Y ?
10. A M B C M D E F M G H I J ?
11. A B C R S T D E F Q R S G H I ?
12. R C R C S T C T U C ?
13. Z A X Z Z X Z Y X Z X X Z ?
14. C E B D A C Z B ?
15. X F H Z J L B N P ?

Appendix E

Esoteric Analogies Test

Esoteric Analogies Test (Stankov, 1997)

Directions:

Show the participant the example, but cover the test items with a sheet of paper.

Say “In this test you will be shown three words in the same manner as the example. There is always a relationship between the first two words. In the example, **LIGHT** and **DARK** are opposites. Your task is to find this relationship and then choose, from the alternatives on the right of the sheet, the word which best shares that relationship with the third word. In the example you would look for a word that was opposite to **HAPPY** and in this case the best answer would be **SAD.**

Ask the participant if they understand the procedure, if they answer no repeat the instructions.

When the person understands, **say** “After each answer I want you to tell me how confident you are that you are correct. There are four alternative answers, so a guess corresponds to 25% confidence, and absolute certainty corresponds to 100% confidence. Please make your choice from the ratings provided. You will have only four minutes to complete this task, so please work as quickly and accurately as you can. There are questions on both sides of the sheet, so please turn over when you have finished the first side”. Uncover the test items, start timing and record the person’s score.

After four minutes have elapsed, say “Please stop now, the time limit for this test is up”.

Example: LIGHT is to DARK as HAPPY is to
GLAD SAD GAY EAGER

Trials:

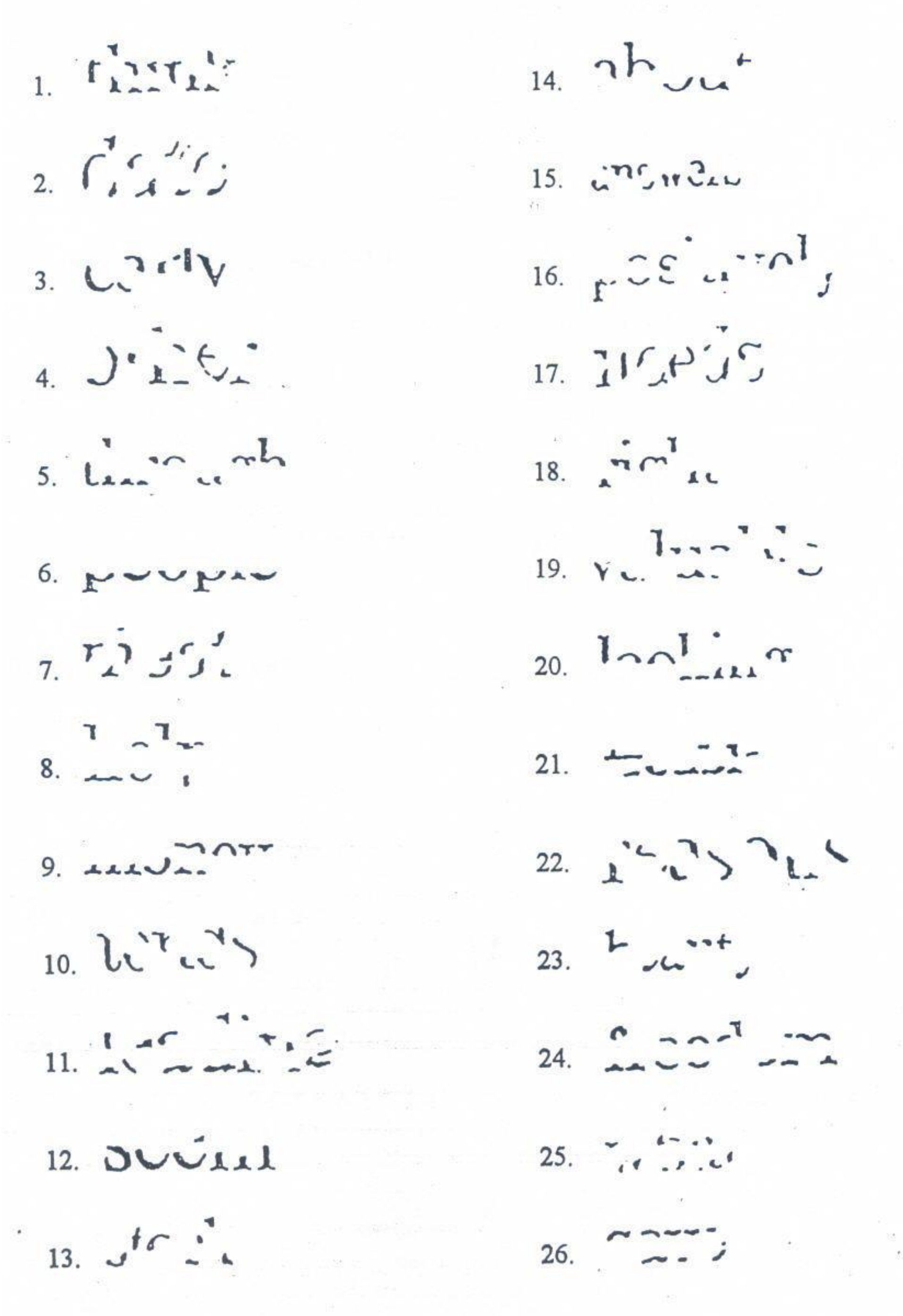
1. FIRE is to HOT as ICE is to
   POLE COLD CREAM WHITE
2. LOVE is to HATE as FRIEND is to
   LOVER PAL OBEY ENEMY
3. STATUE is to SHAPE as SONG is to
   BEAUTY PIANO TUNE NOTE
4. GROUND is to FOOT as RAIL is to
   WHEEL TRAIN IRON STATION
5. FLAME is to HEAT as ROSE is to
   LEAVES SCENT THORN PETALS
6. SPACE is to POINT as TIME is to
   CLOCK ETERNAL MOMENT POTION

7. RAIN is to HAIL as DEW is to
   SNOW WATER CLOUD FROST
8. MANY is to FEW as OFTEN is to
   FREQUENT NEVER ALWAYS
9. BETTER is to WORST as SLOWER is to
   FAST RAPID QUICKEST BEST
10. SURPRISE is to STRANGE as FEAR is to
    ANXIOUS TERRIBLE WEAK QUICK
11. SOON is to NEVER as NEAR is to
    NOWHERE FAR AWAY SOMEWHERE
12. WIN is to JOY as LOSE is to
    FUN SADNESS FAIL DREAM
13. FOX is to WOLF as GOAT is to
    DOG SHEEP TIGER RAT
14. GANDER is to GOOSE as HOG is to
    COW ROOT SOW PIG
15. MAP is to GEOGRAPHY as BLUEPRINT is to
    HOUSE ARCHITECTURE FOUNDATION GEOLOGY
16. FORE is to AFT as BOW is to
    STERN DECK BOAT ARROW
17. HOMOCIDE is to LAW as OEDEMA is to
    ACTING PEDAGOGY THEOLOGY MEDICINE
18. CAT is to FELINE as HORSE is to
    CANINE VULPINE EQUINE CARNIVORE
19. THREE is to TRIANGLE as FIVE is to
    HEXAGON PENTAGON CIRCLE TRAPEZOID
20. ARMADILLO is to ANIMAL as CHARD is to
    VEGETABLE DRINK FISH LIZARD
21. CONSTELLATION is to STAR as ARCHIPELAGO is to
    PENINSULAR ISLAND CONTINENT COUNTRY
22. LENORE is to POE as ALICE is to
    WHITMAN SHAKESPEARE CARROL BYRON
23. GUSTATORY is to TASTE as OLFACTORY is to
    TOUCH SMELL FEEL BALANCE
24. VIRGIL is to AENID as MATTHEW is to
    PSALMS MARK GOSPEL JESUS

Appendix F

Concealed Words Test

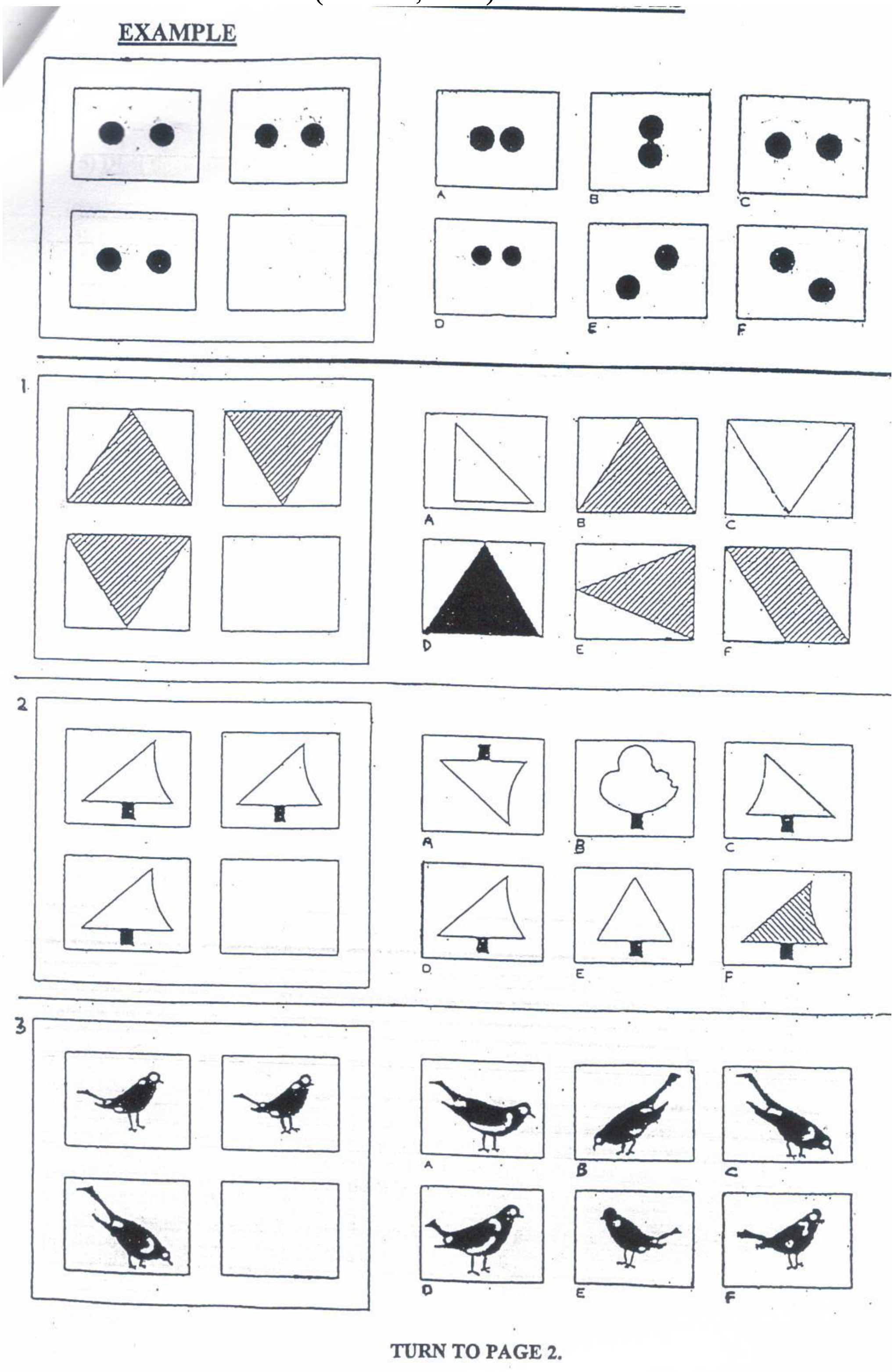
(Stankov, 1997)

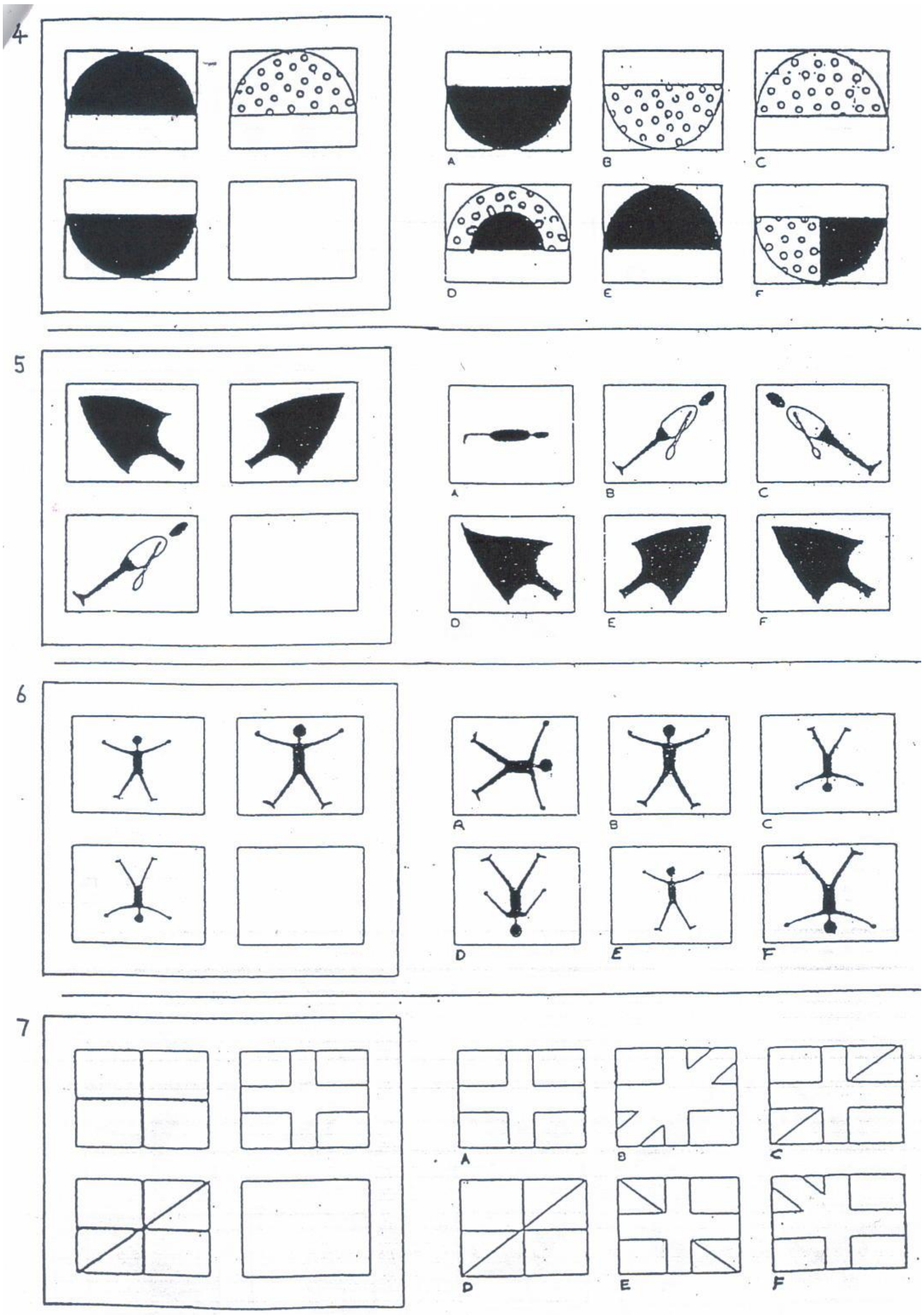


Appendix G

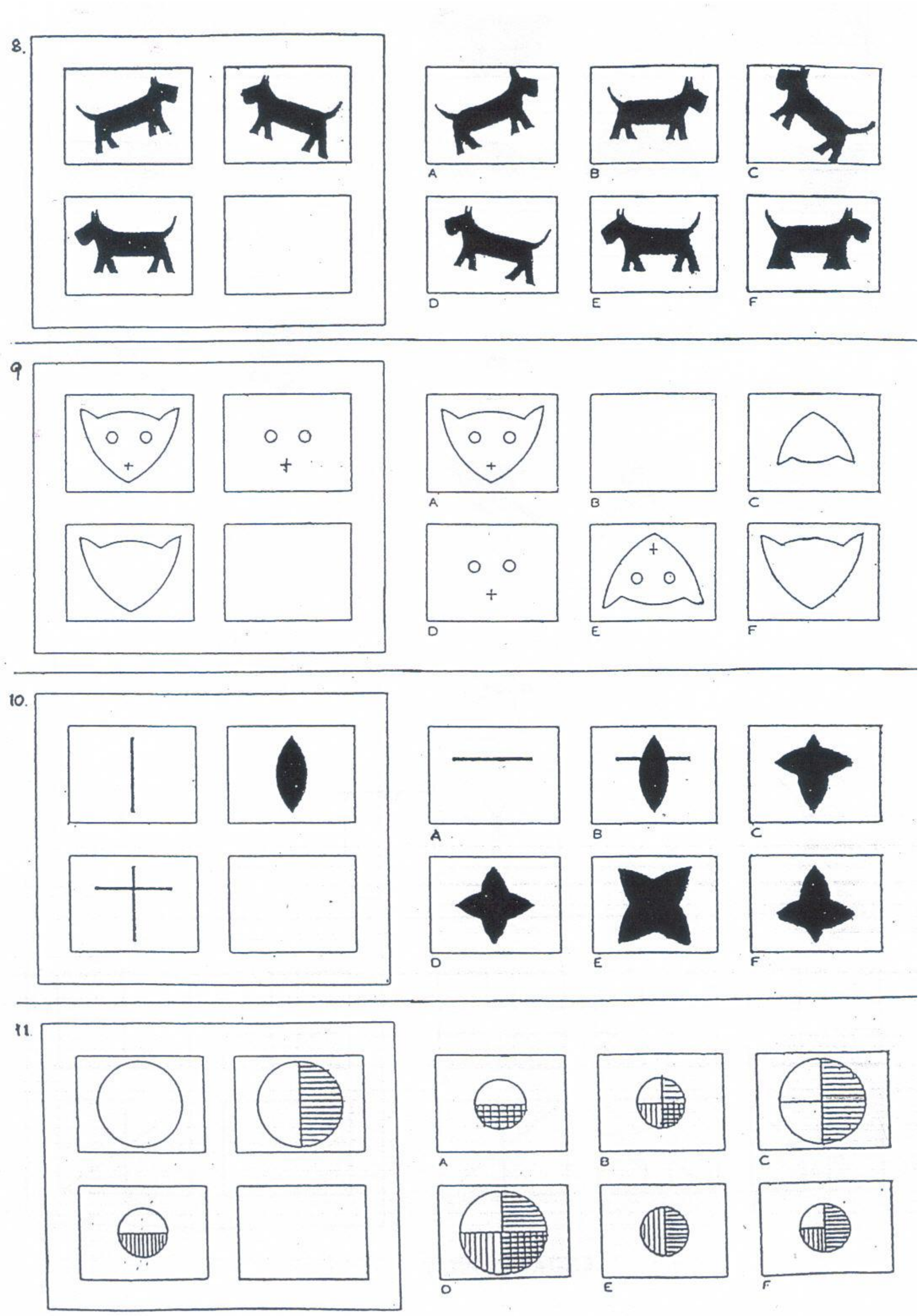
Cattell's Matrices

(Stankov, 1997)





TURN TO PAGE 3.


8.
A
B
C
D
E
F
9
A
B
C
D
E
F
10.
A
B
C
D
E
F
11.
A
B
C
D
E
F

Appendix H

The International Personality Item Pool Five-Factor Personality Scale

(Goldberg, 1997)

On the following pages, there are phrases describing people's behaviors. Please use the rating scale below to describe how accurately each statement describes ***you***. Describe yourself as you generally are now, not as you wish to be in the future. Describe yourself as you honestly see yourself, in relation to other people you know of the same sex as you are, and roughly your same age. So that you can describe yourself in an honest manner, your responses will be kept in absolute confidence. Please read each statement carefully, and then click on the number that corresponds to your rating.

Response Options

1: Very Inaccurate
2: Moderately Inaccurate
3: Neither Inaccurate nor Accurate
4: Moderately Accurate
5: Very Accurate

1	Am the life of the party	1	2	3	4	5
2	Feel little concern for others	1	2	3	4	5
3	Am always prepared	1	2	3	4	5
4	Get stressed out easily	1	2	3	4	5
5	Have a rich vocabulary	1	2	3	4	5
6	Don’t talk a lot	1	2	3	4	5
7	Am interested in people	1	2	3	4	5
8	Leave my belongings around	1	2	3	4	5
9	Am relaxed most of the time	1	2	3	4	5
10	Have difficulty understanding abstract ideas	1	2	3	4	5
11	Feel comfortable around people	1	2	3	4	5
12	Insult people	1	2	3	4	5
13	Pay attention to details	1	2	3	4	5
14	Worry about things	1	2	3	4	5
15	Have a vivid imagination	1	2	3	4	5
16	Keep in the background	1	2	3	4	5
17	Sympathise with others’ feelings	1	2	3	4	5
18	Make a mess of things	1	2	3	4	5
19	Seldom feel blue	1	2	3	4	5
20	Am not interested in abstract ideas	1	2	3	4	5
21	Start conversations	1	2	3	4	5

22	Am not interested in other people’s problems	1	2	3	4	5
23	Get chores done right away	1	2	3	4	5
24	Am easily disturbed	1	2	3	4	5
25	Have excellent ideas	1	2	3	4	5
26	Have little to say	1	2	3	4	5
27	Have a soft heart	1	2	3	4	5
28	Often forget to put things back in their proper place	1	2	3	4	5
29	Get upset easily	1	2	3	4	5
30	Do not have a good imagination	1	2	3	4	5
31	Talk to a lot of different people at parties	1	2	3	4	5
32	Am not really interested in others	1	2	3	4	5
33	Like order	1	2	3	4	5
34	Change my mood a lot	1	2	3	4	5
35	Am quick to understand things	1	2	3	4	5
36	Don’t like to draw attention to myself	1	2	3	4	5
37	Take time out for others	1	2	3	4	5
38	Shirk my duties	1	2	3	4	5
39	Have frequent mood swings	1	2	3	4	5
40	Use difficult words	1	2	3	4	5
41	Don’t mind being the centre of attention	1	2	3	4	5
42	Feel others’ emotions	1	2	3	4	5
43	Follow a schedule	1	2	3	4	5
44	Get irritated easily	1	2	3	4	5
45	Spend time reflecting on things	1	2	3	4	5
46	Am quiet around strangers	1	2	3	4	5
47	Make people feel at ease	1	2	3	4	5
48	Am exacting in my work	1	2	3	4	5
49	Often feel blue	1	2	3	4	5
50	Am full of ideas	1	2	3	4	5

THANKYOU

Appendix I

Big Five Self-Rated Personality Block Descriptions

Extraversion block personality description

Please read the following personality description carefully. Please rate with reference to the 11-point rating scale below the extent to which the **OVERALL** description **GENERALLY** reflects your personality.

I don't mind being the centre of attention; I make friends easily; I take charge; I know how to captivate people; I feel at ease with people; I am skilled in handling social situations; I am the life of the party; I start conversations.

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

You will now be asked to provide a confidence rating. Please read the following instructions carefully.

> Imagine that there was some device that could accurately tell us about your personality. How confident are you that the rating you gave above would correspond with the device's rating? Please rate your confidence on the scale that appears below by circling your level of confidence.

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									**ABSOLUTELY CERTAIN**	

Agreeableness block personality description

Please read the following personality description carefully. Please rate with reference to the 11-point rating scale below the extent to which the **OVERALL** description **GENERALLY** reflects your personality.

> I enquire about others' well being; I know how to comfort others; I love children ; I am on good terms with nearly everyone ; I have a good word for everyone; I show my gratitude; I think of others first; I love to help others; I am interested in people; I sympathise with others' feelings.

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

You will now be asked to provide a confidence rating. Please read the following instructions carefully.

> Imagine that there was some device that could accurately tell us about your personality. How confident are you that the rating you gave above would correspond with the device's rating? Please rate your confidence on the scale that appears below by circling your level of confidence.

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									**ABSOLUTELY CERTAIN**	

Emotional stability block personality description

Please read the following personality description carefully. Please rate with reference to the 11-point rating scale below the extent to which the **OVERALL** description **GENERALLY** reflects your personality.

> I am not easily bothered by things; I do not easily take offense; I rarely get irritated; I seldom get mad; I rarely panic; I am not easily overwhelmed by emotions; I am relaxed most of the time; I seldom feel blue; I don't feel threatened easily.

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

You will now be asked to provide a confidence rating. Please read the following instructions carefully.

> Imagine that there was some device that could accurately tell us about your personality. How confident are you that the rating you gave above would correspond with the device's rating? Please rate your confidence on the scale that appears below by circling your level of confidence.

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									**ABSOLUTELY CERTAIN**	

Conscientiousness Block personality description

Please read the following personality description carefully. Please rate with reference to the 11-point rating scale below the extent to which the **OVERALL** description **GENERALLY** reflects your personality.

I do things according to plan; I continue until everything is perfect; I make plans and stick to them; I love order and regularly like to tidy up; I seldom neglect my duties; I seldom waste my time; I am always prepared; I pay attention to details; I get chores done right away.

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

You will now be asked to provide a confidence rating. Please read the following instructions carefully.

> Imagine that there was some device that could accurately tell us about your personality. How confident are you that the rating you gave above would correspond with the device's rating? Please rate your confidence on the scale that appears below by circling your level of confidence.

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									**ABSOLUTELY CERTAIN**	

Intellect Block Personality Description

Please read the following personality description carefully. Please rate with reference to the 11-point rating scale below the extent to which the **OVERALL** description **GENERALLY** reflects your personality.

> I carry the conversation to a higher level; I catch on to things quickly; I can handle a lot of information; I love to think up new ways of doing things; I love to read challenging material; I am good at many things; I have a rich vocabulary; I am interested in abstract ideas; I have a good imagination; I will probe deeply into a subject.

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

You will now be asked to provide a confidence rating. Please read the following instructions carefully.

> Imagine that there was some device that could accurately tell us about your personality. How confident are you that the rating you gave above would correspond with the device's rating? Please rate your confidence on the scale that appears below by circling your level of confidence.

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									**ABSOLUTELY CERTAIN**	

Appendix J

Informed Consent for Study 1

You are being invited to participate in this study which investigates self-confidence in personality and cognitive judgments. It is anticipated that the results of this study will provide useful information for other researchers and will also help us understand the trait of self-confidence in a more meaningful way.

If you agree to participate in this study, you will be asked to take some time to complete a number of cognitive tasks and personality measures. Completing this test battery should require no longer than 1 hour of your time and a 1% credit applies for those students whose course allows experimental time to be counted toward their final grade. Or you may wish to enter a draw for cash prizes.

Participation in this study is completely voluntary and you will suffer no penalties should you choose not to participate. You are also free to withdraw from the study at any time.

Your identity as a participant in this research will remain confidential with respect to any publication of the results of the study. Any information that can identify you as a participant will be stored in a secured place, with the information available only to the investigator.

If you have any questions at any time regarding this research you may contact **Professor Gerry Fogarty** on **(07) 4631 2379**.

I have fully read the above information, and understand the nature and purpose of this research. I understand that my participation is completely voluntary and that I may withdraw at any time. I understand that the results of this study will be treated with confidentiality. The results will be reported only in their aggregate form and I will not be identified individually.

Name____(please print)

Signature____

Date________________________

Appendix K

Factor Analysis of IPIP Scales and IPIP Block Descriptors-Study 1

The factorial structure of IPIP subscales and IPIP block descriptors were examined using Principal Components Analysis with Oblique rotation. A solution employing root one criterion produced 4 factors. Cattell's Scree plot, however, provided support for a 5 factor solution. Also, the eigenvalue for the fifth factor was .97 providing support for interpreting a five factor solution (see Carroll, 1993). The pattern matrix, percent of variance accounted for, eigenvalues, and factor correlation matrix are presented in Table K1. The five factor solution accounted for 80.71% of the total variance. The first factor was labelled Extraversion which had high loadings from both the IPIP Extraversion subscale and the Extraversion block descriptor. The second factor was labelled Conscientiousness which comprised loadings from both the IPIP and block descriptor measures of Conscientiousness. The third factor was labelled Emotional Stability which had high loadings from both measures. The fourth factor was labelled Agreeableness with high loadings from both Agreeableness measures. The last factor was called Intellect with high loadings from both measures of this dimension.

Table K1

Principal Components Analysis of IPIP Subscale Scores and Big Five Block Descriptors in Study 1 (N = 127).

Variable	$h^{2(b)}$	F1[a]	F2	F3	F4	F5
CONCR	.82		.91			
INTELLCR	.80					.89
AGREECR	.68				-.79	
EXTRACR	.86	.92				
EMOTCR	.86			-.94		
IPIPEXTRA	.82	.88				
IPIPAGREE	.77				-.88	
IPIPCONSC	.81		.89			
IPIPEMOT	.84			-.88		
IPIPINTELLECT	.82					.90
Eigenvalues		2.35	2.01	1.52	1.21	.97
% of variance		23.54	20.10	15.25	12.12	9.69

Factor Correlation Matrix	F1	F2	F3	F4	F5
F1	1.00				
F2	.05	1.00			
F3	-.16	-.14	1.00		
F4	-.14	-.17	.06	1.00	
F5	.30	-.04	.09	-.06	1.00

Note. $h^{2(b)}$ = Communalities; [a] F1 = Personality Bias; F2 = Gf Bias. The cut-off for suppression was .20.

Appendix L

IPIP Form B

IPIP Form B-Agreeableness

Please read the following personality statements carefully. You will be asked to do two things with reference to each statement. Firstly, please rate (on the rating scale from -5 not like me to +5 like me) the extent to which each statement reflects your personality.

After this you will be asked to provide a confidence rating ranging from 0% just guessing to 100% absolutely certain. When making your confidence rating: Imagine that there was some device that could accurately tell us about your personality. How confident are you that the rating you gave would correspond with the device's rating? Please rate your confidence on the scale that appears below by clicking on your level of confidence.

Please read each statement carefully, and then click the number that corresponds to each of your ratings

Statement 1. **I enquire about others' well being**

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 2. I know how to comfort others

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 3. I love children

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 4. I am on good terms with nearly everyone

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 5. I have a good word for everyone;

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 6. I show my gratitude

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 7. I think of others first

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 8. I love to help others

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 9. I am interested in people

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 10. I sympathise with others' feelings

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

IPIP Form B -Conscientiousness

Please read the following personality statements carefully. You will be asked to do two things with reference to each statement. Firstly, please rate (on the rating scale from -5 not like me to +5 like me) the extent to which each statement reflects your personality. After this you will be asked to provide a confidence rating ranging from 0% just guessing to 100% absolutely certain.

When making your confidence rating: Imagine that there was some device that could accurately tell us about your personality. How confident are you that the rating you gave would correspond with the device's rating? Please rate your confidence on the scale that appears below by clicking on your level of confidence.

Please read each statement carefully, and then click on the number that corresponds to each of your ratings

Statement 1. I do things according to plan

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 2. I continue until everything is perfect

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 3. I make plans and stick to them

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 4. I love order and regularly like to tidy up

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 5. I seldom neglect my duties

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 6. I seldom waste my time

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 7. I am always prepared

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 8. I pay attention to details

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 9. I get chores done right away.

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

IPIP Form B -Extraversion

Please read the following personality statements carefully. You will be asked to do two things with reference to each statement. Firstly, please rate (on the rating scale from -5 not like me to +5 like me) the extent to which each statement reflects your personality. After this you will be asked to provide a confidence rating ranging from 0% just guessing to 100% absolutely certain.

When making your confidence rating: Imagine that there was some device that could accurately tell us about your personality. How confident are you that the rating you gave would correspond with the device's rating? Please rate your confidence on the scale that appears below by clicking on your level of confidence.

Please read each statement carefully, and then click on the number that corresponds to each of your ratings

Statement 1. I don't mind being the centre of attention

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 2. I make friends easily

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 3. I take charge

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 4. I know how to captivate people

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 5. I feel at ease with people

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 6. I am skilled in handling social situations

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 7. I am the life of the party

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 8. I start conversations

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

IPIP Form B -Emotional Stability

Please read the following personality statements carefully. You will be asked to do two things with reference to each statement. Firstly, please rate (on the rating scale from -5 not like me to +5 like me) the extent to which each statement reflects your personality. After this you will be asked to provide a confidence rating ranging from 0% just guessing to 100% absolutely certain.

When making your confidence rating: Imagine that there was some device that could accurately tell us about your personality. How confident are you that the rating you gave would correspond with the device's rating? Please rate your confidence on the scale that appears below by clicking on your level of confidence.

Please read each statement carefully, and then click on the number that corresponds to each of your ratings

Statement 1. I am not easily bothered by things

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 2. I do not easily take offense

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 3. I rarely get irritated

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 4. I seldom get mad

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 5. I rarely panic

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 6. I am not easily overwhelmed by emotions

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 7. I am relaxed most of the time

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not										Like
like										me
me										

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST									ABSOLUTLEY	
GUESSING									CERTAIN	

Statement 8. I seldom feel blue

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not										Like
like										me
me										

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST									ABSOLUTLEY	
GUESSING									CERTAIN	

Statement 9. I don't feel threatened easily.

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not										Like
like										me
me										

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST									ABSOLUTLEY	
GUESSING									CERTAIN	

IPIP Form B –Intellect

Please read the following personality statements carefully. You will be asked to do two things with reference to each statement. Firstly, please rate (on the rating scale from -5 not like me to +5 like me) the extent to which each statement reflects your personality. After this you will be asked to provide a confidence rating ranging from 0% just guessing to 100% absolutely certain.

When making your confidence rating: Imagine that there was some device that could accurately tell us about your personality. How confident are you that the rating you gave would correspond with the device's rating? Please rate your confidence on the scale that appears below by clicking on your level of confidence.

Please read each statement carefully, and then click on the number that corresponds to each of your ratings

Statement 1. I carry the conversation to a higher level

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 2. I catch on to things quickly

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 3. I can handle a lot of information

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 4. I love to think up new ways of doing things

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 5. I love to read challenging material

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 6. I am good at many things

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 7. I have a rich vocabulary

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 8. I am interested in abstract ideas

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 9. I have a good imagination

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Statement 10. I will probe deeply into a subject

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									ABSOLUTLEY CERTAIN	

Appendix M

Informed Consent for Study 2

You are being invited to participate in this study which investigates self-confidence in personality and cognitive judgments. It is anticipated that the results of this study will provide useful information for other researchers and will also help us understand the trait of self-confidence in a more meaningful way.

If you agree to participate in this study, you will be asked to take some time to complete a number of self-report personality measures as well as three cognitive tasks. Completing this test battery should require no longer than 3 hour of your time and a 3% credit applies for those students whose course allows experimental time to be counted toward their final grade. Or you may wish to enter a draw for cash prizes.

Participation in this study is completely voluntary and you will suffer no penalties should you choose not to participate. You are also free to withdraw from the study at any time.

Your identity as a participant in this research will remain confidential with respect to any publication of the results of the study. Any information that can identify you as a participant will be stored in a secured place, with the information available only to the investigator.

If you have any questions at any time regarding this research you may contact **Professor Gerry Fogarty** on **(07) 4631 2379**.

I have fully read the above information, and understand the nature and purpose of this research. I understand that my participation is completely voluntary and that I may withdraw at any time. I understand that the results of this study will be treated with confidentiality. The results will be reported only in their aggregate form and I will not be identified individually.

I declare that I am at least 18 years of age and I hereby give my consent to participate in this study by inserting the number from the bottom left-hand corner of the survey into the Consent ID box below.

123456	

Appendix N

Factor Analysis of Big Five Accuracy Scores Study 2

The factorial structure of the Big Five accuracy scores (i.e., methods 1, 2 and 3) were examined using Principal Components Analysis with Promax rotation and Kaiser Normalization. The correlation matrix of these scores is presented in Table N1.

A solution employing root one criterion produced six factors. However, Cattell's Scree plot was indeterminate after five factors which supported a five factor solution. The pattern matrix, percent of variance accounted for, eigenvalues, and factor correlation matrix for the five-factor solution are presented in Table N2. The five factor solution accounted for 72.07% of the total variance. The first factor was labelled Agreeableness accuracy which comprised loadings from all the Agreeableness accuracy scores. The second, third, fourth and fifth factors were labelled Emotional Stability Accuracy, Conscientiousness Accuracy, Extraversion Accuracy, and Intellect Accuracy respectively, with each of the factors comprising loadings from the appropriate methods for each dimension. The low correlations between the factors suggested that accuracy was domain specific and therefore mean accuracy scores were calculated.

Table N1

Correlations Among Big Five Accuracy Scores Obtained From Methods 1, 2, and 3 for Study 2 (N = 405)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
IntellectaccM1	1.00														
ConscientiousnessaccM1	0.11*	1.00													
ExtraversionaccM1	0.20**	0.17**	1.00												
AgreeablenessaccM1	0.12*	0.15**	0.11*	1.00											
EmotionalStabilityaccM1	0.17**	0.12*	0.23**	0.20**	1.00										
IntellectaccM2	0.27**	0.13*	0.07	0.12*	0.08	1.00									
ConscientiousnessaccM2	0.02	0.53**	0.10*	0.10*	0.00	0.16**	1.00								
ExtraversionaccM2	0.08	0.22**	0.34**	0.08	0.15**	0.24**	0.14**	1.00							
AgreeablenessaccM2	0.00	0.04	0.05	0.70	0.14**	0.09	0.05	0.10*	1.00						
EmotionalStabilityaccM2	0.12*	0.07	0.14**	0.08	0.48**	0.08	0.06	0.19**	0.10*	1.00					
IntellectaccM3	0.28**	0.17**	0.09	0.11*	0.08	0.81**	0.15**	0.16**	0.08	0.08	1.00				
ConscientiousnessaccM3	0.08	0.58**	0.09	0.10*	0.10*	0.12*	0.71**	0.10*	0.01	0.07	0.16**	1.00			
ExtraversionaccM3	0.12*	0.27**	0.32**	0.11*	0.16**	0.20**	0.19**	0.82**	0.09	0.24**	0.13**	0.15*	1.00		
AgreeablenessaccM3	0.03	0.07	-0.02	0.54**	0.15**	0.14**	0.12*	0.06	0.67**	0.13**	0.11*	0.04	0.06	1.00	
EmotionalStabilityaccM3	0.12*	0.07	0.12*	0.10*	0.51**	0.09	0.07	0.17**	0.11*	0.86**	0.10*	0.08	0.22**	0.11*	1.00

* p < .05 ** p < .01

Table N-2

Pattern Matrix for Study2 Accuracy Scores Using Principal Components Analysis with Promax Rotation and Kaiser Normalisation (N = 405)

Variable	h2(b)	F1a	F2c	F3d	F4e	F5f
IntellectaccM1	0.29					.51
ConscientiousnessaccM1	0.65			.77		
ExtraversionaccM1	0.35				.57	
AgreeablenessaccM1	0.73	.84				
EmotionalStabilityaccM1	0.55		.71			
IntellectaccM2	0.85					.92
ConscientiousnessaccM2	0.78			.89		
ExtraversionaccM2	0.85				.93	
AgreeablenessaccM2	0.84	.92				
EmotionalStabilityaccM2	0.84		.92			
IntellectaccM3	0.86					.93
ConscientiousnessaccM3	0.82			.92		
ExtraversionaccM3	0.83				.91	
AgreeablenessaccM3	0.72	.84				
EmotionalStabilityaccM3	0.86		.93			
Eigenvalues		3.53	2.13	2.00	1.67	1.48
% of variance		23.54	14.20	13.34	11.13	9.87

Factor Correlation Matrix	F1	F2	F3	F4	F5
F1	1.00				
F2	0.14	1.00			
F3	0.10	0.09	1.00		
F4	0.11	0.23	0.21	1.00	
F5	0.14	0.11	0.19	0.21	1.00

Note. acc= accuracy; M1 = method 1; M2 = method 2; M3 = method 3 . $h^{2(b)}$ = Communalities [a] F1 = Agreeableness Accuracy, [c]F2 = Emotional Stability Accuracy, F3[d] = Conscientiousness Accuracy, F4[e] = Extraversion Accuracy, F5[f] = Intellect Accuracy. The cut-off for suppression was .20.

Appendix O

Word Association Test

(Stankov, 1997)

In this test your task is to think of a word which is associated with two given words. Its relationship with the two given words may be different. For the first example the two words are:

NUMBER…………………….NOBILITY

A correct answer for this example would be **COUNT** because **COUNT** is a word related to numbers and is a title given to members of nobility. For the second example the two words are:

COST…………………………ATTACK

In this case a correct answer could be **CHARGE.** The cost of something is how much you are charged for it, and to charge is a form of attack.

After each answer I want you to type how confident you are that your answer is correct with reference to the confidence scale that is presented below. On the confidence scale 0% represents just guessing and 100% represents that you are absolutely certain your answer is correct. You can make your confidence rating by using the mouse cursor to click on your confidence rating. There are 10 trials in this test. You will be given 3 minutes to complete this test, so if you find a question difficult then move on to the next item and return to unanswered questions if you have time. You can monitor the amount of time that you have taken for the task by referring to the clock in the top left hand corner of the computer screen. Additionally, a counter will appear in the bottom left hand corner of the computer screen so that you can monitor how many trials that you have completed.

At the end of the test, please provide an estimate of the percentage of items you think you answered correctly. It is <u>IMPORTANT</u> that you provide this estimate <u>IMMEDIATELY</u> after completing the test. After providing your percentage estimate, I want you to type how confident you are that your percentage estimate is correct using the confidence scale that appears on the screen.

FINGER……………..HAMMER

How confident are you that your answer is correct?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									**ABSOLUTLEY CERTAIN**	

CARD………………...SHIP

How confident are you that your answer is correct?
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
JUST GUESSING **ABSOLUTLEY CERTAIN**

CLOTHES……………LAW

How confident are you that your answer is correct?
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
JUST GUESSING **ABSOLUTLEY CERTAIN**

MINUTE………………FIRST

How confident are you that your answer is correct?
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
JUST GUESSING **ABSOLUTLEY CERTAIN**

RIVER………………..MONEY

How confident are you that your answer is correct?
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
JUST GUESSING **ABSOLUTLEY CERTAIN**

ACCOUNT……………DUCK

How confident are you that your answer is correct?
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
JUST GUESSING **ABSOLUTLEY CERTAIN**

FISH……………………FILM

How confident are you that your answer is correct?
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
JUST GUESSING **ABSOLUTLEY CERTAIN**

MUSIC…………………CLIMB

How confident are you that your answer is correct?
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
JUST GUESSING **ABSOLUTLEY CERTAIN**

DRESS………………….EDGE

How confident are you that your answer is correct?
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
JUST GUESSING **ABSOLUTLEY CERTAIN**

PIG………………………EYE

How confident are you that your answer is correct?
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
JUST GUESSING **ABSOLUTLEY CERTAIN**

END OF TEST

Please estimate the percentage of items you think you answered correctly:……………………………….

How confident are you that your percentage estimate is correct?
0% 10% 20% 30% 40% 50% 60% 70% 80% 90% 100%
JUST GUESSING **ABSOLUTLEY CERTAIN**

Appendix P

Private Self-Consciousness Scale

(Fenigstein et al., 1975)

Answer the following questions as honestly and accurately as possible on a scale from 0 (Extremely uncharacteristic) to 5 (Extremely characteristic).

Rating Scale

0	1	2	3	4
Extremely Uncharacteristic				Extremely Characteristic

Item 1
I'm generally trying to figure myself out.

Item 2
I reflect about myself a lot.

Item 3
I'm often the subject of my own fantasies.

Item 4
I'm generally attentive to my inner feelings.

Item 5
I'm consistently examining my motives.

Item 6
I sometimes have the feeling that I'm off somewhere watching myself.

Item 7
I'm alert to changes in my mood.

Item 8
I'm aware of the way my mind works when I work through a problem.

Appendix Q

Need for Cognition (short form)

(NFC Cacioppo et al., 1984)

Below are a number of statements. For each statement you are asked to indicate your level of agreement using the following rating scale: Please click on the number that represents your rating.

Rating Scale

-2	-1	0	1	2
Very Strong Disagreement				Very Strong Agreement

Item1
I would prefer complex to simple problems.

Item 2
I would like to have the responsibility of handling a situation that requires a lot of thinking.

Item 3
Thinking is not my area of fun.

Item 4
I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.

Item 5
I try to anticipate and avoid situations where there is likely chance I will have to think in depth about something.

Item 6
I find satisfaction in deliberating hard and for long hours.

Item 7
I only think as hard as I have to.

Item 8
I prefer to think about small, daily projects to long-term ones.

Item 9
I like tasks that require little thought once I’ve learned them.

Item 10
The idea of relying on thought to make my way to the top appeals to me.

Item 11
I really enjoy a task that involves coming up with new solutions to problems.

Item 12
Learning new ways to think doesn't excite me very much.

Item 13
I prefer my life to be filled with puzzles that I must solve.

Item 14
The notion of thinking abstractly is appealing to me.

Item 15
I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.

Item 16
I feel relief rather than satisfaction after completing a task that required a lot of mental effort.

Item 17
It's enough for me that something gets the job done; I don't care how it works.

Item 18
I usually end up deliberating about issues even when they do not affect me personally.

Appendix R

Positive and Negative Affect Schedule (PANAS)

(Watson et al., 1988)

Read each item and then click the ***box*** that best indicates to what extent you generally feel this way, that is, how you feel on average	***Very Slightly or not at all***	***A little***	***Moderately***	***Quite a bit***	***Extremely***
1. interested	☐	☐	☐	☐	☐
2. distressed	☐	☐	☐	☐	☐
3. excited	☐	☐	☐	☐	☐
4. upset	☐	☐	☐	☐	☐
5. strong	☐	☐	☐	☐	☐
6. guilty	☐	☐	☐	☐	☐
7. scared	☐	☐	☐	☐	☐
8. hostile	☐	☐	☐	☐	☐
9. enthusiastic	☐	☐	☐	☐	☐
10. proud	☐	☐	☐	☐	☐
11. irritable	☐	☐	☐	☐	☐
12. alert	☐	☐	☐	☐	☐
13. ashamed	☐	☐	☐	☐	☐
14. inspired	☐	☐	☐	☐	☐
15. nervous	☐	☐	☐	☐	☐
16. determined	☐	☐	☐	☐	☐
17. attentive	☐	☐	☐	☐	☐
18. jittery	☐	☐	☐	☐	☐
19. active	☐	☐	☐	☐	☐
20. afraid	☐	☐	☐	☐	☐

Appendix S

Self-report Intelligence and Confidence Questionnaire

(SICQ)

Please read the following statements about mental abilities carefully. You will be asked to do two things with reference to each statement. Firstly, please rate each statement using the scale from -5 (not like me) to +5 (like me).

After this you will be asked to provide a confidence rating ranging from 0% (just guessing) to 100% (absolutely certain). When making your confidence rating: Imagine that there was some device that could accurately tell us about your mental abilities. How confident are you that the rating you gave would correspond with the device's rating?

Please rate your confidence on the scale that appears below by clicking on your level of confidence.

Please read each statement carefully, and then click the number that corresponds to each of your ratings

Item 1. **I have a good vocabulary**

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									**ABSOLUTLEY CERTAIN**	

Item 2 **I know a lot of worldly facts**

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									**ABSOLUTLEY CERTAIN**	

Item 3 **I have a good short-term memory**

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									**ABSOLUTLEY CERTAIN**	

Item 4 **I am good at arithmetic**

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									**ABSOLUTLEY CERTAIN**	

Item 5 **I have poor visual-spatial skills**

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									**ABSOLUTLEY CERTAIN**	

Item 6 **I have a poor ability at detecting what two concepts have in common**

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device's rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									**ABSOLUTLEY CERTAIN**	

Item 7 **I have good perceptual skills**

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device’s rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									**ABSOLUTLEY CERTAIN**	

Item 8 **I am poor at solving logical problems**

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device’s rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									**ABSOLUTLEY CERTAIN**	

Item 9 **I am intelligent**

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device’s rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									**ABSOLUTLEY CERTAIN**	

Item 10 **I am good at being able to perceive patterns in a series of numbers or letters**

-5	-4	-3	-2	-1	0	+1	+2	+3	+4	+5
Not like me										Like me

How confident are you that the rating you gave above would correspond with the device’s rating?

0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
JUST GUESSING									**ABSOLUTLEY** CERTAIN	

Appendix T

Covering Page and Informed Consent Study 3

DEPARTMENT OF PSYCHOLOGY
FACULTY OF SCIENCES
UNIVERSITY OF SOUTHERN QUEENSLAND
Before starting the survey, close down any menu bars or other programs that may be reducing your screen size. You should be able to read the information on the screen without having to scroll from left to right.

You are being invited to participate in this study which investigates self-confidence in personality and cognitive judgments. It is anticipated that the results of this study will provide useful information for other researchers and will also help us understand the trait of self-confidence in a more meaningful way.

If you agree to participate in this study, you will be asked to take some time to complete a number of self-report personality measures as well as three cognitive tasks. Completing this test battery should require no longer than 2 hours of your time and a 2% credit applies for those students whose course allows experimental time to be counted toward their final grade. Or you may wish to enter a draw for cash prizes. It is important that you try to complete the battery in a single session if possible.

Participation in this study is completely voluntary and you will suffer no penalties should you choose not to participate. You are also free to withdraw from the study at any time.

Your identity as a participant in this research will remain confidential with respect to any publication of the results of the study. Any information that can identify you as a participant will be stored in a secured place, with the information available only to the investigator.

If you have any questions at any time regarding this research you may contact Professor Gerard Fogarty on (07) 4631 2379.

I have fully read the above information, and understand the nature and purpose of this research. I understand that my participation is completely voluntary and that I may withdraw at any time. I understand that the results of this study will be treated with confidentiality. The results will be reported only in their aggregate form and I will not be identified individually.

If you have any technical concerns or difficulties accessing the Survey please contact Sandra Baker, Department of Psychology on 0402070056 or 46311613 or contact Ross Bool, University of Southern Queensland, on 07 4631 2388, or email bool@usq.edu.au.

I declare that I am at least 18 years of age and I hereby give my consent to participate in this study by inserting the number from the bottom left-hand corner of the survey into the Consent ID box below.

Appendix U

Additional Factor Analyses for Study 3

Factor Analysis of Big Five Accuracy Scores Study 3

The factorial structure of the Big Five accuracy scores (i.e., methods 1, 2 and 3) were examined using Principal Components Analysis with Promax rotation and Kaiser Normalization. The correlation matrix of these scores is presented in Table U1.

A solution employing root one criterion produced five factors. The pattern matrix, percent of variance accounted for, eigenvalues, and factor correlation matrix are presented in Table U2. The five factor solution accounted for 81.13% of the total variance. The first factor was labelled Emotional Stability Accuracy which comprised loadings from all the Emotional Stability accuracy scores. The second, third, fourth and fifth factors were labelled Agreeableness Accuracy, Conscientiousness Accuracy, Intellect Accuracy, and Extraversion Accuracy respectively, with each of the factors comprising loadings from the appropriate methods for each dimension. The low correlations between the factors suggested that accuracy was domain specific and therefore mean accuracy scores were calculated.

Table U1

Correlations Among Big Five Accuracy Scores Obtained From Methods 1, 2, and 3 for Study 3 (N = 243)

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
IntellectaccM1	1.00														
ConscientiousnessaccM1	0.14	1.00													
ExtraversionaccM1	0.15	0.04	1.00												
AgreeablenessaccM1	0.19	0.08	0.09	1.00											
EmotionalStabilityaccM1	0.19	0.16	-0.01	0.09	1.00										
IntellectaccM2	0.65	0.13	0.12	0.18	0.13	1.00									
ConscientiousnessaccM2	0.09	0.70	0.02	0.00	0.09	0.15	1.00								
ExtraversionaccM2	0.17	0.12	0.23	0.24	-0.06	0.29	0.12	1.00							
AgreeablenessaccM2	0.14	0.08	0.13	0.81	0.17	0.16	0.07	0.20	1.00						
EmotionalStabilityaccM2	0.18	0.18	0.03	0.13	0.81	0.14	0.10	-0.04	0.20	1.00					
IntellectaccM3	0.61	0.12	0.20	0.14	0.13	0.69	0.12	0.21	0.14	0.14	1.00				
ConscientiousnessaccM3	0.13	0.91	0.07	0.07	0.12	0.12	0.74	0.13	0.07	0.14	0.13	1.00			
ExtraversionaccM3	0.12	0.16	0.26	0.24	-0.02	0.22	0.15	0.78	0.23	0.02	0.19	0.18	1.00		
AgreeablenessaccM3	0.12	0.10	0.12	0.77	0.15	0.13	0.08	0.22	0.82	0.20	0.13	0.09	0.25	1.00	
EmotionalStabilityaccM3	0.20	0.17	0.00	0.10	0.98	0.14	0.10	-0.08	0.16	0.83	0.14	0.13	-0.04	0.16	1.00

Note. * p < .05 ** p < .01

Table U-2

Pattern Matrix for Study 3 Accuracy Scores Using Principal Components Analysis with Promax Rotation and Kaiser Normalisation (N = 243)

Variable	$h^{2(b)}$	F1[a]	F2[c]	F3[d]	F4[e]	F5[f]
IntellectaccM1	0.75				.86	
ConscientiousnessaccM1	0.89			.94		
ExtraversionaccM1	0.25					.48
AgreeablenessaccM1	0.86		.93			
EmotionalStabilityaccM1	0.95	.98				
IntellectaccM2	0.79				.88	
ConscientiousnessaccM2	0.76			.88		
ExtraversionaccM2	0.83					.91
AgreeablenessaccM2	0.89		.94			
EmotionalStabilityaccM2	0.84	.91				
IntellectaccM3	0.77				.87	
ConscientiousnessaccM3	0.92			.96		
ExtraversionaccM3	0.86					.93
AgreeablenessaccM3	0.86		.92			
EmotionalStabilityaccM3	0.96	.98				
Eigenvalues		3.99	2.57	2.36	1.94	1.30
% of variance		29.59	17.13	15.76	12.96	8.69

Factor Correlation Matrix	F1	F2	F3	F4	F5
F1	1.00				
F2	0.17	1.00			
F3	0.15	0.09	1.00		
F4	0.18	0.17	0.15	1.00	
F5	-0.04	0.27	0.17	0.25	1.00

Note. acc= accuracy; M1 = method 1; M2 = method 2; M3 = method 3 . $h^{2(b)}$ = Communalities [a] F1 = Emotional Stability Accuracy, [c]F2 = Agreeableness Accuracy, F3[d] = Conscientiousness Accuracy, F4[e] = Intellect Accuracy, F5[f] = Extraversion Accuracy. The cut-off for suppression was .20.

Factor Analysis of Gf Accuracy Scores and Confidence in PTPEs

The factorial structure of the Gf accuracy scores and confidence in PTPEs were examined using Principal Axis Factoring with Oblique rotation. A solution employing root one criterion produced two factors. The pattern matrix, percent of variance accounted for, eigenvalues, and factor correlation matrix are presented in Table U3. The two factor solution accounted for 55.82% of the total variance. As anticipated, the two factors comprised loadings from the expected confidence and accuracy scores. The first factor was labelled Gf Evaluative Confidence which had high loadings from the confidence in PTPE scores. The second factor was labelled Gf accuracy which comprised loading from the Gf accuracy scores. The factors were not correlated highly suggesting that Gf Evaluative Confidence and Gf Accuracy are distinct but correlated processes.

Table U3

Principal Axis Factoring of Gf Evaluative and Gf Accuracy Scores (N = 243).

Variable	$h^{2(b)}$	F1[a]	F2
ESTCG	.64	.79	
LSTGC	.37	.62	
WATGC	.48	.69	
WATAC	.24		.51
ESTAC	.23		.45
LSTAC	.11		.27
Eigenvalues		2.12	1.23
% of variance		35.27	20.55

Factor Correlation Matrix

	F1	F2
F1	1.00	
F2	.29	1.00

Note. $h^{2(b)}$ = Communalities; [a] F1 = Personality Bias; F2 = Gf Bias