SIMULTANEOUS NORMALISATION AS AN APPROACH TO ESTABLISH EQUIVALENCE IN CROSS-CULTURAL MARKETING RESEARCH

Catharina Strasheim
9505703T

A thesis submitted to the Faculty of Commerce, Law and Management,
University of the Witwatersrand,
in fulfillment of the requirements for the degree of
Doctor of Philosophy

Supervisor: Dr Anthony G. Stacey

May, 2008

ABSTRACT

Since bias threatens the validity of a study, it should be avoided where possible. Across all phases of a research project, bias could be introduced, and in most situations the researcher has reasonable control over processes that may be the source of bias. However, within a quantitative research context in social sciences, where the opinions, attitudes and intentions of people are often sought, response styles patterns due to cultural background, for example, are not within the control of the researcher. Typical response style patterns include acquiescence bias, a tendency to be agreeable to statements, which could be more prevalent in certain cultural groups than other. Another response style pattern is extremity ratings, where respondents tend to avoid the middle categories and mark the scale extremes.

When practitioners sample respondents from different cultural groups, it is difficult, and depending on the research design, sometimes impossible to know whether significant differences are an artefact of substantive differences, or of differences in response styles. Adjusting scores for bias has a significant effect on the interpretation of research findings. To correct for bias, the method most commonly used to adjust scores within each cultural group is standardisation. In this research, SIMNORM, a target distribution estimation approach was used for the simultaneous estimation of a class of non-linear transformation functions that transform the *composite* scores within each cultural group to a standard normal distribution. SIMNORM was found to perform better than standardisation to obtain equivalence across cultural groups when *composite* scores are used.

In addition, SIMITNORM, an *item* normalisation approach was developed, which is a simultaneous non-linear transformation of *item* scores to a standard normal target distribution. The results of seven nested SIMITNORM models were compared to raw item scores and standardised scores, using a multi-group confirmatory factor analysis approach, a method that is suitable to test for construct equivalence, metric equivalence and scalar equivalence. SIMITNORM had significant advantages over standardisation as an approach to obtain equivalence over *items* in a set of data where bias is present.

DECLARATION

I, Catharina Strasheim, declare that this research report is my own, unaided work, except as indicated in the acknowledgements, the text and the references. It is being submitted in fulfillment of the requirements for the degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg. It has not been submitted before, in whole, or in part for any degree or examination at any other University.

_____

Catharina Strasheim

May 9, 2008

DEDICATION

This research report is dedicated to my husband De Wet, in acknowledgement of his ongoing support, love and encouragement during the period of my studies and to my children, Christine, Wilke and Conrad. Thank you for everything that you have given up, especially time.

This is also a dedication to my parents and late father-in-law, for their prayers, encouragement, love and interest in all of us.

My deepest humility and thanks are before God Almighty, my solid rock at all times.

ACKNOWLEDGEMENTS

I would like to express my sincere thanks and appreciation to the following people, whose encouragement in this research was notable:

my research supervisor, Dr Anthony Stacey, for his support and constructive guidance, for very helpful pointers and for being always available;

Professor Mike Bendixen, who accepted me on the PhD programme at WBS;

my personal mentor and friend, Professor Leyland Pitt, who gave the initial idea for this research;

Professor Michael Browne and Dr Stephen du Toit, whose guidance with the development of SIMNORM during 1985-1989 brings back very happy memories;

Erik du Plessis, from Milward Browne and Christopher Darryl, who provided me with the data used in this research;

Prof Carel van Aardt and my colleagues at Unisa BMR for their encouragement and for allowing me the time to complete this work;

Unisa SBL, my previous employer, where I had wonderful opportunities to develop as an academic;

Bernice McNeil, who edited this work; and finally,

the National Research Foundation and the Unisa contributions for a Thuthuka Grant and a Mobility Grant, that made it possible for me to attend a workshop in Chicago in LISREL, and to buy the software required for this research.

TABLE OF CONTENTS

CHAPTER 1

1	BIAS IN CROSS-CULTURAL RESEARCH	1
1.1	INTRODUCTION	1
1.2	SIMULTANEOUS NORMALISATION DEFINED	3
1.3	THE CROSS-CULTURAL RESEARCH CONTEXT	5
1.4	PROBLEM STATEMENT	5
1.4.1	The sub-problems	5
1.5	IMPORTANCE OF THE STUDY	6
1.6	DELIMITATIONS	9
1.7	ASSUMPTIONS	10
1.8	LAYOUT OF THIS THESIS	11

CHAPTER 2

2	LITERATURE REVIEW	12
2.1	STRUCTURE OF THE LITERATURE REVIEW	12
2.2	BIAS IN DIFFERENT PHASES OF CROSS-CULTURAL RESEARCH	13
2.2.1	Measurement bias	15
2.2.2	Construct bias	15
2.2.3	Response bias	16
2.2.4	Item bias	19
2.2.5	Measurement bias summarised	19
2.3	BIAS/NON-INVARIANCE AND EQUIVALENCE/INVARIANCE	20
2.3.1	Uniform and non-uniform bias	22
2.3.2	Detection of uniform and non-uniform bias	23
2.4	MEASUREMENT FORM A CLASSICAL TEST THEORY PERSPECTIVE	26
2.4.1	Measurement error	28

2.5	STATISTICAL TECHNIQUES USED IN CROSS-CULTURAL RESEARCH	30
2.5.1	Preliminary analyses	30
2.5.2	Standardisation methods	31
2.6	ALTERNATIVE APPROACHES TO THE NORMALISATION PROBLEM	33
2.6.1	The problem of non-normality	33
2.6.2	The Stacey (2005) normalisation approach	34
2.6.3	Structure-oriented techniques: Exploratory factor analysis	35
2.6.4	Level-oriented statistical techniques	35
2.7	CONFIRMATORY FACTOR ANALYSIS	36
2.7.1	Introduction to CFA in a SEM-framework	36
2.7.2	The general CFA measurement model (single group case)	37
2.7.3	The general CFA model: Covariance structure (single group case)	38
2.7.4	Measures of fit	38
2.7.5	The general CFA measurement model (multiple group case)	43
2.7.6	The general CFA model: Covariance structure (multiple group case)	44
2.7.7	Different levels of invariance: The general multi-group CFA model	45
2.7.8	Summary of the MG CFA approach	53
2.8	ITEM RESPONSE THEORY	55
2.8.1	The Graded Response Model	55
2.8.2	A comparison of IRT and SEM approaches in multi-group analyses	59
2.8.3	Application limitation of IRT method within the scope of this research	64
2.9	ADVERTISING RESPONSE SCALES	64
2.9.1	The Schlinger (1979) Viewer Response Profile	65
2.9.2	Schlinger, (1982)	71
2.9.3	Wells, Leavitt & McConville, (1971)	72
2.9.4	Zinkhan & Fornell, (1985)	73
2.9.5	Lastovicka, (1983)	74
2.9.6	Zinkhan & Burton, (1989)	76
2.9.7	Moldovan, (1985)	77
2.9.8	Aaker & Bruzonne, (1985)	78
2.9.9	Summary of Advertising Response Scales	80
2.10	SUMMARY	80

CHAPTER 3

3	SIMULTANEOUS NORMALISATION: SIMNORM AND SIMITNORM	89
3.1	NORMALISATION IN MARKETING	89
3.1.1	Normalisation in psychometric testing	89
3.1.2	Standardisation	91
3.1.3	Normalisation	93
3.1.4	The simultaneous normalisation of four simulated non-normal data sets	95
3.1.5	Normalisation in cross-cultural marketing research	101
3.2	SIMULTANEOUS NORMALISATION (SIMNORM)	102
3.2.1	An example of SIMNORM across two cultural groups	102
3.2.2	The target distribution estimation method	104
3.2.3	The transformation function for a single group	106
3.2.4	The density and the likelihood function for a single group	110
3.2.5	Extension of the model to include more groups	112
3.2.6	An example of SIMNORM for two cultural groups	114
3.2.7	Parameter estimation	119
3.3	SIMULTANEOUS ITEM NORMALISATION (SIMITNORM)	121
3.3.1	The mathematical model for the normalisation of individual items	121
3.3.2	Meanings associated with item parameters	123
3.3.3	Some examples of the application of SIMITNORM	128
3.4	PROPOSTIONS IN THIS RESEARCH	133
3.5	APPLICATION OF SIMITNORM IN THIS RESEARCH	134

	CHAPTER 4	135
4	RESEARCH METHODOLOGY	135
4.1	INTRODUCTION TO METHODOLOGY	135
4.2	THE PURPOSE OF THE STUDY	135
4.2.1	Problem statements	135
4.2.2	The sub-problems	135
4.2.3	Assumptions	136
4.3	THE DATA USED IN THIS RESEARCH	137
4.3.1	The nature of this study	137
4.3.2	The COMMAP data	137
4.3.3	Background of the COMMAP database	138
4.3.4	The selection of commercials from COMMAP	139
4.3.5	Variables available on the COMMAP database	140
4.4	TESTING PROPOSITIONS	141
4.4.1	Propositions related to SIMNORM	141
4.4.2	Propositions related to SIMITNORM	142
4.5	METHODOLOGY ROADMAP	144
4.6	CONSISTENCY MATRIX	145

CHAPTER 5

5	SIMNORM RESEARCH RESULTS	150
5.1	THE COMMAP DATABASE	150
5.1.1	The key assumptions	151
5.1.2	The biographic distribution in the data set	152
5.1.3	The descriptive statistics of the research variables in the data set	154
5.2	ITEM BIAS ANALYSIS USING ANOVA	156
5.3	PSYCHOMETRIC PROPERTIES OF THE 32-ITEM SCHLINGER VRP	164
5.3.1	Comparison with similar advertising response research	164
5.3.2	Statements of VRP32	165
5.3.3	The structure of the VRP – Exploratory Factor Analysis of 32 items	166
5.3.4	The structure of the VRP - Confirmatory Factor Analysis of 32 items	167
5.4	THE 20 ITEM SCHLINGER VRP	171
5.5	RELIABILITY AND VALIDITY OF VRP32 AND VRP20	175
5.5.1	The Reliability of VRP32 and VRP20	175
5.5.2	Nomological validity	176
5.5.3	Convergent and discriminant validity	178
5.5.4	Criterion related validity	179
5.5.5	Concurrent validity of VRP32 and VRP20	179
5.5.6	Construct Validity	181
5.5.7	Content Validity	182
5.5.8	Summary of Reliability and Validity Assessment	183
5.6	RESULTS OF SIMNORM	184
5.6.1	SIMNORM of aggregate scores of the Schlinger (1979) VRP20 for four cultural groups	184
5.6.2	The effect of SIMNORM on level-orientated analyses	201
5.6.3	Significance of main effects and interaction effects in ANOVA using the biographic variables as predictors	203
5.6.4	Significance of main effects and interaction effects in ANOVA using the advertising variables as predictors	206
5.6.5	The effect of SIMNORM on post-hoc comparisons in ANOVA using the biographic variables as predictors	212
5.6.6	The effect of SIMNORM on post-hoc comparisons in ANOVA using the advertising variables as predictors	217
5.6.7	Correlations between raw scores, standardised scores and SIMNORM scores	226
5.6.8	Results of SIMNORM on regression analyses	228
5.6.9	Summary of SIMNORM results	229

CHAPTER 6

6	SIMITNORM RESULTS	231
6.1	LAYOUT OF THIS CHAPTER	231
6.2	THE SIMITNORM PROCEDURE	232
6.3	DESCRIPTIVE MEASURES	245
6.4	RELIABILITY ANALYSIS AND CORRELATION BETWEEN SCORES	266
6.5	EXPLORATORY FACTOR ANALYSIS	269
6.6	SINGLE GROUP CONFIRMATORY FACTOR ANALYSES	271
6.7	MULTI-GROUP CONFIRMATORY FACTOR ANALYSES	275
6.7.1	Presentation of Hypotheses 1 to 6	275
6.7.2	Hypothesis 1: Configural Invariance	275
6.7.3	Hypothesis 2a: Metric Invariance	276
6.7.4	Hypothesis 2b: Equal Intercepts	276
6.7.5	Hypothesis 3: Factor Covariance and Variance Invariance	277
6.7.6	Hypothesis 4: Error Variance Invariance	278
6.7.7	Hypothesis 5: Latent Mean Invariance	279
6.7.8	Hypothesis 6: Latent Mean Invariance, scalar invariance, equal intercepts and unequal error variances	280
6.7.9	Implications of Hypotheses 1 to 6 in terms of Equivalence	281
6.8	THE EFFECT OF SIMITNORM ON ITEM BIAS	284

CHAPTER 7

7	DISCUSSION OF FINDINGS	296
7.1	LAYOUT OF THIS CHAPTER	296
7.2	PROPOSITIONS RELATED TO SIMNORM	296
7.3	PROPOSITIONS RELATED TO SIMITNORM	299
7.4	IMPLICATIONS OF RESEARCH RESULTS	302
7.5	SUGGESTIONS FOR FURTHER RESEARCH REGARDING SIMITNORM	303
7.6	SUGGESTIONS FOR FURTHER RESEARCH REGARDING EQUIVALENCE	304
7.7	CONCLUSION	309

REFERENCES 310

APPENDIX A 325

APPENDIX B 328

APPENDIX C 339

APPENDIX D 357

APPENDIX E 391

APPENDIX F 405

APPENDIX G 409

APPENDIX H 412

APPENDIX I 421

LIST OF TABLES

Table		Page
	CHAPTER 1	1
1.1	The contribution of this research	8

Table		Page
	CHAPTER 2	12
2.1	Response styles in marketing resulting in systematic error	17
2.2	Different sources of bias in cross-cultural research	19
2.3	The difference between uniform and non-uniform bias	22
2.4	Summary of five hierarchical multi-group CFA hypotheses	53
2.5	Comparison of attention in scholarly literature to CFA/IRT	60
2.6	Differences between multi-group CFA and multi-group IRT in measurement equivalence applications	62
2.7	Similarities between multi-group CFA and multi-group IRT in measurement equivalence applications	63
2.8	Similar constructs in advertising response scales	79
2.9	Summary of literature review and implications for researchers in cross-cultural research contexts	81

Table		Page
	CHAPTER 3	89
3.1	Three randomly generated data sets	92
3.2	Differences between standardisation and normalisation	93
3.3	Descriptive statistics of raw and normalised scores	96
3.4	Descriptive statistics of normalised scores for two cultural groups	118

Table		Page
	CHAPTER 4	135
4.1	Distribution of cultural groups in the data set used in this study	141
4.2	Assessment of Proposition related to SIMNORM	141
4.3	Assessment of Propositions related to SIMITNORM	142
4.4	Consistency Matrix	145

Table		Page
	CHAPTER 5	150
5.1	The language distribution in the data set	152
5.2	The age distribution in the data set	153
5.3	The gender distribution in the data set	153
5.4	The income distribution in the data set	153
5.5	The distribution according to whether the commercial has been seen before	153
5.6	The distribution according to whether the advertised product is being used	154
5.7	The distribution according to the General Enjoyment of commercials	154
5.8	Descriptive statistics of the 32 VRP items across four cultural groups	155
5.9	Significance Levels of ANOVA analysis by score level for each of the 32 items across two population groups	157
5.10	Descriptive statistics of the 32 VRP items (n=4800)	165
5.11	Results of Exploratory Factor Analysis on 32 item Schlinger Scale (n=4800)	167
5.12	Estimated covariances between latent variables for VRP32	168
5.13	Estimated covariances between latent variables for VRP20	172
5.14	Fit measures of VRP32 versus VRP20 for primary and secondary samples	173
5.15	Reliability of the 32- and 20- item VRP	176
5.16	Results of Exploratory Factor Analysis for VRP20	177
5.17	Convergent and Discriminant Validity assessment for VRP32 and VRP20	178
5.18	Regression of *Likeability* (Dependent variable) with the seven dimensions of VRP32 and VRP20 as independent variables	180
5.19	Construct validity of the VRP: Cross tabulation of the seven VRP dimensions against the 10-point *Likeability* scale	182
5.20	Descriptive measures for raw scores, standardised scores and normalised scores over four cultural groups for *Overall Schlinger VRP20_10*	192
5.21	Variables used in SIMNORM	201
5.22	SIMNORM results on ANOVA, *Likeabilty* as dependent variable with biographic predictors	203
5.23	SIMNORM results on ANOVA, V20_8 (*Positive Mood*) as dependent variable with biographic predictors	204
5.24	SIMNORM results on ANOVA, R20_9 (*Negative Mood*) as dependent variable with biographic predictors	205
5.25	SIMNORM results on ANOVA, V20_10 (*Overall Schlinger VRP20*) as dependent variable with biographic predictors	206
5.26	SIMNORM results on ANOVA, *Likeability* as dependent variable with advertising variables as predictors	207

Table		Page
5.27	SIMNORM results on ANOVA, V20_8 (*Positive Mood*) as dependent variable with advertising variables as predictors	208
5.28	SIMNORM results on ANOVA, R20_9 (*Negative Mood*) as dependent variable with advertising variables as predictors	210
5.29	SIMNORM results on ANOVA, V20_10 (*Overall Schlinger VRP20*) as dependent variable with advertising variables as predictors	211
5.30	Results of SIMNORM on Post-Hoc Comparisons in ANOVA, with *Likeability* as dependent variable, biographic variables	213
5.31	Results of SIMNORM on Post-Hoc Comparisons in ANOVA, with V20_8 (*Positive Mood*) as dependent variable, biographic variables	214
5.32	Results of SIMNORM on Post-Hoc Comparisons in ANOVA, with R20_9 (*Negative Mood*) as dependent variable, biographic variables	215
5.33	Results of SIMNORM on Post-Hoc Comparisons in ANOVA, with V20_10 (*Overall Schlinger VRP20*) as dependent variable, biographic variables	216
5.34	Results of SIMNORM on Post-Hoc Comparisons in ANOVA, with *Likeability* as dependent variable, advertising variables	218
5.35	Results of SIMNORM on Post-Hoc Comparisons in ANOVA, with V20_8 (*Positive Mood*) as dependent variable, advertising variables	220
5.36	Results of SIMNORM on Post-Hoc Comparisons in ANOVA, with R20_9 (*Negative Mood*) as dependent variable, advertising variables	222
5.37	Results of SIMNORM on Post-Hoc Comparisons in ANOVA, with V20_10 (*Overall Schlinger VRP20*) as dependent variable, advertising variables	224
5.38	Pearson Correlation Coefficients between raw scores, standardised scores and SIMNORM scores	226
5.39	Fit Statistics of five regression models	228
5.40	Parameter estimates of five regression models	229
5.41	Significance of the estimates of five regression models	229

Table		Page
	CHAPTER 6	231
6.1	SIMITNORM models, assumptions and practical implications	235
6.2	Item Frequencies across cultural groups	237
6.3	SIMITNORM Score levels and distances between adjacent score categories: Model A	238
6.4	SIMITNORM Score levels and distances between adjacent score categories: Model B	239
6.5	SIMITNORM Score levels and distances between adjacent score categories: Model BB	240
6.6	SIMITNORM Score levels and distances between adjacent score categories: Model BBB	241
6.7	SIMITNORM Score levels and distances between adjacent score categories: Model D	242
6.8	SIMITNORM Score levels and distances between adjacent score categories: Model DD	243
6.9	SIMITNORM Score levels and distances between adjacent score categories: Model DDD	244
6.10	Measures of fit for SIMITNORM models	245
6.11	Descriptive statistics of *Raw Scores, Standarised Z-Scores* and *SIMITNORM Scores*	246
6.12	Normality Tests for *Raw Scores, Standardised Z-Scores* and *SIMITNORM Scores*	263
6.13	Cronbach Alpha Reliability Coefficient	266
6.14(a)	Pearson Correlation Coefficients between Raw Scores, Standardised Scores, SIMNORM and SIMITNORM Scores for English Group	267
6.14(b)	Pearson Correlation Coefficients between Raw Scores, Standardised Scores, SIMNORM and SIMITNORM Scores for Afrikaans Group	267
6.14(c)	Pearson Correlation Coefficients between Raw Scores, Standardised Scores, SIMNORM and SIMITNORM Scores for Nguni Group	268
6.14(d)	Pearson Correlation Coefficients between Raw Scores, Standardised Scores, SIMNORM and SIMITNORM Scores for Sotho Group	268
6.15	Varimax Rotated Factor Pattern: Standardised Items	270
6.16	Varimax Rotated Factor Pattern: SIMITNORM Model DDD	270
6.17	Goodness-of-Fit Measures of Single Group Confirmatory Factor Analyses	274
6.18	Root Mean Square Error of Approximation (RMSEA)	282
6.19	Other fit measures	283
6.20	Fit measures based on Chi-Square value	284

Table		Page
6.21	Significance of *Popgroup* main effect and *Popgroup*Score Level* interaction effect in ANOVA bias detection for items with uniform bias	287
6.22	Significance of *Popgroup* main effect and *Popgroup*Score Level* interaction effect in ANOVA bias detection for items with non-uniform bias	288
6.23	Classification of items in terms of the success of SIMITNORM in removing bias, considering the interaction term Popgroup by Score Level, using the ANOVA approach	290

Table		Page
	CHAPTER 7	296
7.1	Potential of SIMITNORM as an alternative approach in measurement equivalence applications	306

Table		Page
	APPENDIX B	328
B.1	Descriptions of all total scores used for the SIMNORM results	328
B.2	Means and standard deviations of all raw total scores	330
B.3	Frequency distribution of *Likeability* across cultural groups	330
B.4	Estimated parameters using SIMNORM for *Likeability*	331
B.5	Fit statistics of SIMNORM for Likeability	331
B.6	Descriptive statistics of raw scores, standardised scores and normalised scores using SIMNORM for Likeability	332
B.7	Estimated parameters using SIMNORM for *V20_8: Positive Mood* VRP20 variable	333
B.8	Fit statistics of SIMNORM for *V20_8: Positive Mood* VRP20 variable	333
B.9	Descriptive statistics of raw scores, standardised scores and normalised scores using SIMNORM for *V20_8: Positive Mood* VRP20 variable	334
B.10	Estimated parameters using SIMNORM for *R20_9: Negative Mood VRP20 variable*	335
B.11	Fit statistics of SIMNORM for *R20_9: Negative Mood* VRP20 variable	335
B.12	Descriptive statistics of raw scores, standardised scores and normalised scores using SIMNORM for *R20_9: Negative Mood* VRP20 variable	336
B.13	Estimated parameters using SIMNORM for *V20_10: Overall Schlinger VRP20*	337
B.14	Fit statistics of SIMNORM for *V20_10: Overall Schlinger* VRP20	337
B.15	Descriptive statistics of raw scores, standardised scores and normalised scores using SIMNORM for *V20_10: Overall Schlinger* VRP20	338

Table		Page
	APPENDIX C	339
C.1	Item Frequencies	339
C.2	Item Percentages	340
C.3	Descriptive Statistics of Raw Scores	341
C.4	Estimated Parameters: Model A	342
C.5	Descriptive Statistics of SIMITNORM Scores: Model A	343
C.6	Estimated Parameters: Model B	344
C.7	Descriptive Statistics of SIMITNORM Scores: Model B	345
C.8	Estimated Parameters: Model D	346
C.9	Descriptive Statistics of SIMITNORM Scores: Model D	347
C.10	Estimated Parameters: Model BB	348
C.11	Descriptive Statistics of SIMITNORM Scores: Model BB	349
C.12	Estimated Parameters: Model BBB	350
C.13	Descriptive Statistics of SIMITNORM Scores: Model BBB	351
C.14	Estimated Parameters: Model DD	352
C.15	Descriptive Statistics of SIMITNORM Scores: Model DD	353
C.16	Estimated Parameters: Model DDD	354
C.17	Descriptive Statistics of SIMITNORM Scores: Model DDD	355
C.18	Fit Measures of SIMITNORM Functions, Models A, B, D	356
C.19	Fit Measures of SIMITNORM Functions, Models BB, BBB, DD, DDD	356

Table		Page
	APPENDIX E	391
E.1(a)	SIMITNORM Scores: Model A	391
E.1(b)	Differences between adjacent SIMITNORM Score categories: Model A	392
E.2(a)	SIMITNORM Scores: Model B	393
E.2(b)	Differences between adjacent SIMITNORM Score categories: Model B	394
E.3(a)	SIMITNORM Scores: Model D	395
E.3(b)	Differences between adjacent SIMITNORM Score categories: Model D	396
E.4(a)	SIMITNORM Scores: Model BB	397
E.4(b)	Differences between adjacent SIMITNORM Score categories: Model BB	398
E.5(a)	SIMITNORM Scores: Model BBB	399
E.5(b)	Differences between adjacent SIMITNORM Score categories: Model BBB	400
E.6(a)	SIMITNORM Scores: Model DD	401
E.6(b)	Differences between adjacent SIMITNORM Score categories: Model DD	402
E.7(a)	SIMITNORM Scores: Model DDD	403
E.7(b)	Differences between adjacent SIMITNORM Score categories: Model DDD	404

Table		Page
	APPENDIX F	405
F.1	Varimax Rotated Factor Pattern: Raw Items	405
F.2	Varimax Rotated Factor Pattern: SIMITNORM Model A	406
F.3	Varimax Rotated Factor Pattern: SIMITNORM Model B	406
F.4	Varimax Rotated Factor Pattern: SIMITNORM Model D	407
F.5	Varimax Rotated Factor Pattern: SIMITNORM Model BB	407
F.6	Varimax Rotated Factor Pattern: SIMITNORM Model BBB	408
F.7	Varimax Rotated Factor Pattern: SIMITNORM Model DD	408

Table		Page
	APPENDIX G	409
G.1	Lambda-X Completely Standardised Solutions for *Raw Scores, Standardised Z-Scores* and SIMITNORM Models	409
G.2	Estimated PHI Matrix Parameters for *Raw Scores, Standardised Z-Scores* and SIMITNORM Models	410
G.3	THETA-EPS Estimated Parameters for *Raw Scores, Standardised Z-Scores* and SIMITNORM Models	411

Table		Page
	APPENDIX H	412
H.1	Group Fit Measures: Multi-group CFA on *Raw Scores*	412
H.2	Group Fit Measures: Multi-group CFA on *Standardised Z-Scores*	413
H.3	Group Fit Measures: Multi-group CFA on *SIMITNORM A-Scores*	414
H.4	Group Fit Measures: Multi-group CFA on *SIMITNORM B-Scores*	415
H.5	Group Fit Measures: Multi-group CFA on *SIMITNORM D-Scores*	420
H.6	Group Fit Measures: Multi-group CFA on *SIMITNORM BB-Scores*	417
H.7	Group Fit Measures: Multi-group CFA on *SIMITNORM BBB-Scores*	418
H.8	Group Fit Measures: Multi-group CFA on *SIMITNORM DD-Scores*	419
H.9	Group Fit Measures: Multi-group CFA on *SIMITNORM DDD-Scores*	420

LIST OF FIGURES

Figure		Page
	CHAPTER 2	12
2.1(a)	Unbiased item	25
2.1(b)	Item with uniform bias	25
2.1(c)	Item with non-uniform bias	25
2.2	Random error increases the variability of the distribution, but the location remains the same	27
2.3	Systematic error does not affect the variability of the distribution, but there is a difference in the location of the distribution	28
2.4	Operating Characteristic Curves for a five-category item using the graded response model	57
2.5	Categorical response curves of the graded response model	58
2.6	Item categories and item thresholds	59

Figure		Page
	CHAPTER 3	89
3.1	Standardisation is a linear transformation	92
3.2	Hierarchy of SIMNORM models	95
3.3	Model A fitted to four simulated non-normal data sets	97
3.4	Model B fitted to four simulated non-normal data sets	98
3.5	Model D fitted to four simulated non-normal data sets	99
3.6	Four simulated distributions before and after normalisation	100
3.7	Histogram of the raw total scores of the 32-item Schlinger scale for two cultural groups	103
3.8	Cumulative density of the total score of the 32-item Schlinger scale for two cultural groups	104
3.9	Q-Q plot for two cultural groups	106
3.10	Inverse Q-Q plot of total score for two cultural groups	107
3.11	The general shape of the Richards family of growth curves	107
3.12	General shape of the transformation function	108
3.13	The SIMNORM functions using Model A for two cultural groups	115
3.14	The SIMNORM functions using Model B for two cultural groups	116
3.15	The SIMNORM functions using Model D for two cultural groups	117
3.16	Distribution of normalised total scores for two cultural groups	119
3.17	Hierarchy of nested models for SIMITNORM	123
3.18	Simulated target density functions for SIMITNORM with different parameter values	126
3.19	Simulated target transformation functions for SIMITNORM with different parameter values	127
3.20	Examples of the item normalisation functions	129
3.21	Examples of the normalised item density distributions	130
3.22	The sum of standardised items for two cultural groups	132
3.23	The sum of SIMITNORM items for two cultural groups	132

Figure		Page
	CHAPTER 4	135
4.1	Methodology Roadmap	144

Figure		Page
	CHAPTER 5	150
5.1	The nature of bias in the 32 items of Schlinger (1979) VRP	158
5.2	Estimated First-Order, Seven-Factor Schlinger VRP32 Model	169
5.3	Estimated First-Order, Seven-Factor Schlinger VRP20 Model	174
5.4	Histograms of the sum of raw scores of the VRP20 Schlinger (1979) scale for four cultural groups	184
5.5	Cumulative density of the total score of the VRP20 scale for four cultural groups	185
5.6	Q-Q plot for four cultural groups	186
5.7	The SIMNORM Model A functions for four cultural groups	187
5.8	The SIMNORM Model B functions for four cultural groups	188
5.9	The SIMNORM Model D functions for four cultural groups	189
5.10	Distribution of standardised total scores for four cultural groups	191
5.11	Distribution of SIMNORM Model A scores for four cultural groups	193
5.12	Distribution of SIMNORM Model B scores for four cultural groups	194
5.13	Distribution of SIMNORM Model D scores for four cultural groups	195
5.14	Box-and-Whisker Plot of Raw Total VRP20	196
5.15	Box-and-Whisker Plot of Standardised Total VRP20	197
5.16	Box-and-Whisker Plot of Model A Total VRP20	198
5.17	Box-and-Whisker Plot of Model B Total VRP20	199
5.18	Box-and-Whisker Plot of Model D Total VRP20	200

Figure		Page
	CHAPTER 6	231
6.1	Hierarchy of nested models for SIMITNORM in this research	234
6.2	Box-and-Whisker Plot of Raw Scores	247
6.3	Box-and-Whisker Plot of Z-Scores	248
6.4	Box-and-Whisker Plot of SIMITNORM A-Scores	249
6.5	Box-and-Whisker Plot of SIMITNORM B-Scores	240
6.6	Box-and-Whisker Plot of SIMITNORM D-Scores	251
6.7	Box-and-Whisker Plot of SIMITNORM BB-Scores	252
6.8	Box-and-Whisker Plot of SIMITNORM BBB-Scores	253
6.9	Box-and-Whisker Plot of SIMITNORM DD-Scores	254
6.10	Box-and-Whisker Plot of SIMITNORM DDD-Scores	255
6.11	Histograms of Raw Scores, Standardised Scores and SIMITNORM Scores	257
6.12	Three-dimensional histograms of Standardised Scores and SIMITNORM Scores	260
6.13	Normal Probability Plot of Raw Scores	264
6.14	Normal Probability Plot of SIMITNORM Scores Model DDD	265
6.15	Basic Path Model for Confirmatory Factor Analysis	272
6.16	Item Bias Analysis – Item 16	292
6.17	Item Bias Analysis – Item 18	293
6.18	Item Bias Analysis – Item 25	294

Figure		Page
	APPENDIX D	357
D1	Cumulative Density Functions of 20 VRP items	357
D2	Q-Q plot of 20 VRP items over four cultural groups	359
D3.1	SIMITNORM functions for Item 1 over four cultural groups	361
D3.2	SIMITNORM functions for Item 2 over four cultural groups	361
D3.3	SIMITNORM functions for Item 3 over four cultural groups	362
D3.4	SIMITNORM functions for Item 4 over four cultural groups	362
D3.5	SIMITNORM functions for Item 8 over four cultural groups	363
D3.6	SIMITNORM functions for Item 9 over four cultural groups	363
D3.7	SIMITNORM functions for Item 10 over four cultural groups	364
D3.8	SIMITNORM functions for Item 14 over four cultural groups	364
D3.9	SIMITNORM functions for Item 15 over four cultural groups	365
D3.10	SIMITNORM functions for Item 16 over four cultural groups	365
D3.11	SIMITNORM functions for Item 17 over four cultural groups	366
D3.12	SIMITNORM functions for Item 18 over four cultural groups	366
D3.13	SIMITNORM functions for Item 20 over four cultural groups	367
D3.14	SIMITNORM functions for Item 21 over four cultural groups	367
D3.15	SIMITNORM functions for Item 22 over four cultural groups	368
D3.16	SIMITNORM functions for Item 24 over four cultural groups	368
D3.17	SIMITNORM functions for Item 25 over four cultural groups	369
D3.18	SIMITNORM functions for Item 27 over four cultural groups	369
D3.19	SIMITNORM functions for Item 28 over four cultural groups	370
D3.20	SIMITNORM functions for Item 29 over four cultural groups	370

Figure		Page
D4.1	Seven SIMITNORM model score distribution for Item 1 by cultural group	371
D4.2	Seven SIMITNORM model score distribution for Item 2 by cultural group	372
D4.3	Seven SIMITNORM model score distribution for Item 3 by cultural group	373
D4.4	Seven SIMITNORM model score distribution for Item 4 by cultural group	374
D4.5	Seven SIMITNORM model score distribution for Item 8 by cultural group	375
D4.6	Seven SIMITNORM model score distribution for Item 9 by cultural group	376
D4.7	Seven SIMITNORM model score distribution for Item 10 by cultural group	377
D4.8	Seven SIMITNORM model score distribution for Item 14 by cultural group	378
D4.9	Seven SIMITNORM model score distribution for Item 15 by cultural group	379
D4.10	Seven SIMITNORM model score distribution for Item 16 by cultural group	380
D4.11	Seven SIMITNORM model score distribution for Item 17 by cultural group	381
D4.12	Seven SIMITNORM model score distribution for Item 19 by cultural group	382
D4.13	Seven SIMITNORM model score distribution for Item 20 by cultural group	383
D4.14	Seven SIMITNORM model score distribution for Item 21 by cultural group	384
D4.15	Seven SIMITNORM model score distribution for Item 22 by cultural group	385
D4.16	Seven SIMITNORM model score distribution for Item 24 by cultural group	386
D4.17	Seven SIMITNORM model score distribution for Item 25 by cultural group	387
D4.18	Seven SIMITNORM model score distribution for Item 27 by cultural group	388
D4.19	Seven SIMITNORM model score distribution for Item 28 by cultural group	389
D4.20	Seven SIMITNORM model score distribution for Item 29 by cultural group	390

Figure		Page
	APPENDIX I	421
I.1	Item Bias Analysis - Item 1	421
I.2	Item Bias Analysis - Item 2	422
I.3	Item Bias Analysis - Item 3	423
I.4	Item Bias Analysis - Item 4	424
I.5.	Item Bias Analysis - Item 8	425
I.6	Item Bias Analysis - Item 9	426
I.7	Item Bias Analysis - Item 10	427
I.8	Item Bias Analysis - Item 14	428
I.9	Item Bias Analysis - Item 15	429
I.10	Item Bias Analysis - Item 16	430
I.11	Item Bias Analysis - Item 17	431
I.12	Item Bias Analysis - Item 18	432
I.13	Item Bias Analysis - Item 20	433
I.14	Item Bias Analysis - Item 21	434
I.15	Item Bias Analysis - Item 22	435
I.16	Item Bias Analysis - Item 24	436
I.17	Item Bias Analysis - Item 25	437
I.18	Item Bias Analysis - Item 27	438
I.19	Item Bias Analysis - Item 28	439
I.20	Item Bias Analysis - Item 29	440

LIST OF ABBREVIATIONS

AGFI	Adjusted Goodness of Fit Index
AIC	Akaike Information Criterion
ANOVA	Analysis of Variance
CC	Coefficient of Congruency
CFA	Confirmatory Factor Analysis
CTT	Classical Test Theory
CFI	Comparative Fit Index
ECVI	Expected Cross-Validation Index
EFA	Exploratory Factor Analysis
GFI	Goodness-of-fit Index
GLS	Generalised Least Squares
LTT	Latent Trait Theory
ML	Maximum Likelihood
MTMM	Multi-Trait Multi-Method
NCP	Non-centrality Parameter
NFI	Normed Fit Index
NNFI	Non-normed Fit Index
RML	Robust Maximum Likelihood
RMSEA	Root Mean Square Error of Approximation
SEM	Structural Equation Modelling
SIMITNORM	Simultaneous Item Normalisation
SIMNORM	Simultaneous Normalisation (of composite scores)
VRP	Viewer Response Profile

CHAPTER 1

BIAS IN CROSS-CULTURAL RESEARCH

1.1 INTRODUCTION

Cross-cultural studies involve several challenges, and are susceptible to measurement, analysis and data interpretation problems. Although all phases of a study are threatened by bias, the conceptualisation stage, the research design and the statistical analysis phases of the study are particularly vulnerable to the introduction of bias.

During the conceptualisation of the research, when the theoretical constructs and the research propositions and hypotheses are being formulated, difficulties may arise due to the fact that the construct could not be similarly defined across cultural groups. During the design phase, where the choice and administration of instruments and selection of items are being decided upon, the instrument design, data collection method or item wording may introduce cross-cultural bias.

During the data analysis phase, bias may be introduced by ignoring measurement artefacts that may arise due to cultural differences. Researchers and practitioners should understand the limitations and acknowledge the existence of factors that may threaten the statistical interpretation of their research findings, especially when distinctive methodologies and or distinctive groups or sub-populations are involved in the study. During the data analysis phase of any cross-cultural research study, the challenge for the researcher is to ascertain whether observed differences (or the absence of differences) are due to measurement artefacts or due to valid cross-cultural differences (Van de Vijver and Leung, 1997). Although the method of multi-group confirmatory factor analysis has existed for a long time, McGaw and Jöreskog (1971), it has been more recently suggested as an approach to deal with bias in cross-cultural

studies, see for example Marsh and Hovecar (1985); Bollen (1989); Byrne (1998), Little (1997); Steenkamp and Baumgartner (1998); Vandenberg and Lance (2000).

Another branch of methods that has been proposed more recently to deal with bias is item response theory methods (IRT) (Embretson and Reise, 2000; Reise, Widaman and Pugh, 1993; Van de Vijver and Leung, 1997). Various methods of standardisation have also been applied as a method to eliminate bias in studies (Fischer, 2004). One of the consequences of bias is systematic error. Although the sources and effects of response bias (Steenkamp and Baumgartner, 1998; Cheung and Rensvold, 2000) in particular have been well covered, methods to estimate systematic error have not received much attention in the literature so far. An early attempt at estimating systematic error was recently suggested by Baumgartner and Steenkamp (2006).

According to Reise *et al* (1993), since the CFA and IRT approaches have been proposed as methods to deal with the invariance problem, frequent application of these methods are sometimes ignored in situations where it could be fruitful. Also, since these methods have been proposed, the topic of invariance and methods dealing with invariance has been largely ignored in the psychometric literature, and new approaches have not been forwarded recently. A primary purpose of this research is to investigate whether the method of simultaneous normalisation may be suitable as an alternative approach.

This study compares two simultaneous normalisation approaches, namely SIMNORM and SIMITNORM, which are both non-linear data transformation methods, to standardisation and to the multi-group confirmatory factor analysis approach (MG CFA) in the analysis of cross-cultural data in which bias is present. The effect of SIMNORM is compared to standardisation using level-orientated data analysis methods such as ANOVA and regression, while the effect of SIMITNORM is compared to standardisation, using multi-group confirmatory factor analysis, which is mainly a structure-orientated statistical method.

It is expected that simultaneous normalisation may offer an alternative method of removing systematic bias. The principles of simultaneous normalisation arise from methodologies used to equate aptitude tests applied to groups with inherent differences.

1.2 SIMULTANEOUS NORMALISATION DEFINED

It is well known that different cultures respond in different ways to psychometric instruments, of which marketing research surveys and questionnaires are an example. Some of these differences may arise from the instrument. Some may arise from interview method bias. Some bias may arise from the way that specific cultures use and interpret a scale.

A simple case of the problem: one cultural group might tend to be very neutral in responses, and be reluctant to mark the very high or low points on scales, while others may perceive it to be improper or offensive to be disagreeable, and may mark “agree”, “strongly agree” on item scales more often than another cultural group. The issue that faces researchers engaged in cross-cultural research is that of accommodating cultural difference in the statistical interpretation of the research results. The basic question is: does Culture A really respond in a different way from Culture B to a specific instrument, because of true cultural differences, or are the differences merely due to the way the cultures use the scale?

Van de Vijver and Leung (1997) suggest that, when cross-cultural researchers find significant cultural differences in their results, these differences could be a mixture of valid cross-cultural differences and method bias effects. The presence of method bias often shows itself as a systematic variation of stimuli across cultures, and requires very specific analytical approaches.

Various methods for equating scores have been used in other disciplines, such as educational psychology, in order to overcome the problem of inherent differences. For example, when comparing results on an aptitude test, the same test form is typically applied to children from various age groups. The older children tend to achieve higher scores than the younger children due to the

former's longer exposure to schooling. Theoretically speaking, children in a younger age group should have a lower average ability than children in a higher age group. Test equating methods are continually being used to transform the scores obtained from an aptitude test to a SAT-score, for example, which allows valid comparisons across age levels. The specific method, SIMNORM, that was used and extended in this research, was developed for the purpose of test equating over various age groups. SIMNORM is a simultaneous normalisation approach (Strasheim, 1989), where parameters can be expressed as functions of age, or where normalisation can be applied over independent groups, such as gender groups. SIMNORM was developed to deal with data at aggregate level, where the items of a scale or sub-scale are combined to form a total score, and where the scores that are normalised can be assumed to be a continuous variable.

This study proposes an extension of SIMNORM, which will be called SIMITNORM, which allows items that are measured on an ordinal scale, to be normalised. This study examines and compares the plausibility of both SIMNORM and SIMITNORM as methods of obtaining equivalence in cross-cultural data, by comparing them to existing methods that have been proposed in the literature. The major evaluation method that was used is hierarchical hypothesis testing within the method of confirmatory factor analysis in the structural equation modelling framework.

A working definition of simultaneous normalisation for the purposes of this study is: the simultaneous non-linear mathematical transformation using a target distribution estimation method (Elphinstone, 1983; 1985) of data across cultures to eliminate culture-specific differences and maintain individual differences in the data. The mathematical functions (Strasheim, 1989) ensure a smooth transformation, so that if two individuals in the same cultural group were to have different scores on a measurement instrument, the one with the lower score would obtain a lower normalised score. Therefore, normalisation does not affect the ordering of individuals within a cultural group. The simultaneous transformation functions have mathematical properties that ensure a smooth, mathematical progression from observed scores to equivalent scores. Mathematically, if X_i denotes the score of individual i within a specific group, and

$t(X_i)$ denotes the normalised score, and if $x_1 \leq x_2$, then $t(x_1) \leq t(x_2)$. The data is normalised in such a way that the distribution of scores are theoretically normal, with equal means and equal variances across the different groups.

1.3 THE CROSS-CULTURAL RESEARCH CONTEXT

An increasing academic interest in cross-cultural comparison is evident in various disciplines, and particularly in the discipline of marketing. Business practitioners experience an increasing need to compare research results across cultures or nations with the increased growth in globalisation and economic development. Because of the fact that bias may exist due to culture, there is a need for data analysts and researchers to view their data from different angles to understand the deeper dimensions on the one hand. On the other hand, there is a need to develop more robust theories in order to applicability in different contexts.

1.4 PROBLEM STATEMENT

The purpose of this study is to investigate whether two simultaneous normalisation approaches (SIMNORM and SIMITNORM) can offer improvements over the standardisation approach that has so far been mostly in use (Fischer, 2004), and to establish whether the methods are suitable as alternative methods in the analysis of cross-cultural research data in which response bias pose a problem.

1.4.1 The sub-problems

The first sub-problem is to examine whether total score simultaneous normalisation (SIMNORM) is better than standardisation to obtain equivalent scores across groups.

The second sub-problem extends the first sub-problem, and is to investigate how statistical results are affected when SIMNORM-scores, standardised scores and observed scores (based on the same set of empirical data) are used to analyse

data, using level-orientated data analysis methods such as regression analysis and analysis of variance (ANOVA).

The third sub-problem is to examine whether simultaneous item normalisation (SIMITNORM) is better than standardisation at item level to obtain equivalent item scores across groups.

The fourth sub-problem extends the third, and investigates how the statistical results are affected when SIMITNORM scores, standardised scores and observed scores (based on the same set of empirical data) are subjected to structure-orientated data analysis methods, such as exploratory factor analysis (EFA) and multi-group confirmatory factor analysis (MG CFA) within in the structural equation modelling (SEM) framework.

1.5 IMPORTANCE OF THE STUDY

The purpose of this study is to compare SIMNORM and SIMITNORM, two non-linear data transformation approaches, to existing methods (standardisation) used in the analysis of cross-cultural data in which bias is present. The suitability of SIMNORM and SIMITNORM is examined and they are compared to existing approaches to deal with equivalence problems across cultures.

As has been pointed out by Van Herk, Poortinga, and Verhallen (2005), with the steady progress in globalisation, multi-national companies increasingly have to take notice of cultural, economic and geographic differences in the markets they serve. Differences between the home market and markets of other countries may influence how potential customers react to marketing efforts designed in the home country. It is therefore crucial that market researchers pay serious attention to issues that my introduce bias in research findings. As have been described by Appiah (2001), ethnic identity has a significant effect on how consumers evaluate the content of advertisements.

The need for using explicit procedures when bias is present in data has been highlighted by several authors, e.g. Byrne, Shavelson and Muthén (1989); Steenkamp and Baumgartner (1998); Sin, Cheung and Lee (1999); Little, (2000); Smith and Reynolds (2002); Cheung and Rensvold (2000); (Fischer 2004).

" ... what may appear to be cross-cultural differences (or similarities) might just be reflecting cross-cultural differences in response style; some method of detection and removal of their influence must be employed if cross-cultural comparisons .. are to be valid." (Smith and Reynolds, 2002:466).

"The rapid growth of international business in general, and international marketing in particular, has created a need for cross-cultural consumer research" Sin, Cheung and Lee, (1999:76). Day and Montgomery (1999) regard the search for new approaches to deal with marketing challenges as well as cross-cultural research as important and beneficial research. This research addresses these issues, since it is aimed at offering a solution to a fundamental problem in cross-cultural studies.

Several authors have pointed to the need to establish equivalence in data when researchers have cross-cultural research data (Bollen, 1989; Byrne, 1989; Steenkamp and Baumgartner, 1998; Cheung and Rensvold (2000); Fischer 2004).

According to Fischer (2004: 278), "Dealing with cross-cultural response patterns is arguably one of the most challenging issues in cross-cultural survey research."

"Ensuring construct comparability when testing for between-group differences is of paramount importance in cross-cultural research" (Little, 2000:213).

Van de Vijver and Leung (1997) distinguish between methods for eliminating bias at various stages of research. Various approaches are available to detect and deal with bias in a cross-cultural research context at different stages in the research process.

Methods that have been adopted are data transformation techniques, such as standardisation (Van den Vijver and Leung, 1997). Structural methods have been used frequently, and have become widely accepted as valuable tools (Bollen, 1989; Byrne, 1998; Du Toit and Du Toit, 1999; Du Toit and Du Toit, 2001). All of these methods have unique mathematical approaches and certain limitations apply to these techniques. The limitations are often to be found in the assumptions underlying the techniques.

This research demonstrates that SIMNORM could be used to transform non-normal distributions to a normal distribution. Often the normality assumption that is required in most multivariate methods is ignored, especially in the application of level-orientated techniques. It will also be shown that SIMITNORM, used in conjunction with SEM in structure-orientated studies, could help researchers not to violate the assumption of normally distributed data.

A summary of the statistical techniques used in cross-cultural studies with the contribution of this study is provided in Table 1.1.

Table 1.1: The contribution of this research

Type of statistical technique	**Existing Statistical technique to deal with equivalence problems in data**	**Contribution**
Transformation techniques	Standardisation (Fischer, 2004)	SIMNORM and SIMITNORM
Level-orientated techniques	ANOVA, Multiple Regression	SIMNORM
Structure-orientated techniques	Exploratory Factor Analysis (EFA)	SIMITNORM
Structure-orientated techniques	Confirmatory Factor Analysis (CFA) using Structural Equation Modelling (SEM) Bollen 1989; Byrne 1989; Steenkamp and Baumgartner, 1998.	SIMITNORM

1.6 DELIMITATIONS

It may be possible to find a family or class of transformation functions that can be used to normalise data, but this type of research is more in the domain of the pure mathematical and/or statistical sciences. In this research the focus is on a specific function that was proved to be useful in the transformation of data obtained from ability and mental tests, (Strasheim, 1989). The basic SIMNORM model was extended to SIMITNORM, to equate data at item level within the scope of this study. Therefore, the need for intervention in the presence of bias is the fundamental argument followed in this research, and a promising alternative method is suggested.

The researcher did not endeavour to explain theoretical reasons for observed differences between the cultural groups in the data sets analysed, but focused specifically on SIMNORM and SIMITNORM as alternative data analysis methods as compared to standardisation, to achieve equivalence in cross-cultural studies. The purpose of this research was therefore not to explain substantive differences between cultural groups, but rather the focus was to investigate an alternative method to deal with bias and to compare this method with existing methods.

SIMNORM and SIMITNORM are not aimed at providing a solution to poorly constructed, weakly designed and poorly executed research, which implies that this research did not address the problems associated with construct bias.

The researcher did not attempt to improve or suggest new methods or alternative approaches based on the existing methodologies such as SEM or IRT in cross-cultural research, but only demonstrated that SIMNORM and/or SIMITNORM can serve as a supplementary method.

This research excluded the use of item response theory methods used in cross-cultural research, for reasons that will be covered in the literature review presented in Chapter 2.

1.7 ASSUMPTIONS

The first key assumption is that the instrument being selected to demonstrate the use of SIMNORM and SIMITNORM is a valid and reliable instrument across the cultural groups. It is crucial that this assumption is tested before the research proceeds, since SIMNORM and SIMITNORM are not aimed at correcting problems arising from poorly designed survey instruments.

Construct equivalence is the second key assumption and pre-condition for the use of SIMNORM and SIMITNORM. Construct equivalence, or structural equivalence, implies that the theoretical constructs being measured have comparative meaning across cultures. The use of SIMNORM and SIMITNORM on instruments that do not posses construct equivalence is invalid. It is also crucial that this assumption be tested before the use of SIMINORM or SIMITNORM.

The third key assumption is that bias arising from physical conditions of administration, interviewer effects, differential familiarity with physical administration, and communication problems and other unknown sources of bias, is negligible and can be assumed to be equal across cultural groups.

The fourth key assumption is that item bias due to item wording, formulation and appropriateness of item content is negligible and can be assumed to be equal across cultural groups.

The fifth key assumption is that differences observed between cultural groups can be attributed mainly to cultural differences, or response style differences.

1.8 LAYOUT OF THIS THESIS

This chapter gives the background, definitions, and context of the research. The research problem, sub-problems, purpose, delimitations and key assumptions are also covered in Chapter 1.

The literature review in Chapter 2 provides the background to methods currently used to analyse cross-cultural data, and presents a brief background to alternative normalisation methods. The literature on the development of instruments to gauge the emotional response of television advertising is also covered. One of the scales, namely the Schlinger Viewer Response Profile (VRP) (Schlinger, 1979), was used in this research.

Chapter 3 introduces the reader to the principles and method of simultaneous normalisation. The mathematical background of SIMNORM and SIMITNORM as data transformation techniques is given and the methods are both demonstrated using the data set selected for this research.

Chapter 4 describes the methodology and the rationale of this study and provides a consistency matrix of the research.

Chapter 5 firstly tests the first and second key assumptions, and then provides the results of SIMNORM.

In Chapter 6, the results of SIMITNORM are presented.

Chapter 7 concludes with a discussion of the research results, implications of the research and suggestions for further research. In this chapter the research results are evaluated, based on an assessment of the research results against the stated research propositions.

CHAPTER 2

LITERATURE REVIEW

2.1 STRUCTURE OF THE LITERATURE REVIEW

All studies and all phases of studies have potential for bias to be introduced. However, according to Van de Vijver and Leung (1997), in cross-cultural research, the most important sources of bias are in the following stages: (1) the conceptualisation of the theoretical constructs and the formulation of the research questions and hypotheses; (2) the design of the study; and (3) the data analysis phase of the study.

In quantitative studies, where structured questionnaires are administered, this exposition unfortunately excludes another stage of the research process, which happens during the few minutes when actual measurement takes place. This bias is introduced in the data in the moments when participants interpret the wording of items, and respond to them using their own interpretive frameworks. It is often almost impossible for the researcher to control for bias at this stage and it is very difficult and often impossible, to quantify this bias. However, during the data analysis phase of a study, there are several approaches that have been forwarded to deal with bias.

The focus of this study is on examining new approaches to deal with *measurement bias* in the analysis, and to compare and contrast the new approaches using a widely used approach that has become the accepted norm of treating data where measurement bias may pose a problem.

In Section 2.2, the reader can find a general background about bias, and how bias can result in error, showing the need for equivalence in cross-cultural research. In Section 2.3, other normalisation methods, excluding SIMNORM and SIMITNORM are discussed. The SIMNORM and SIMITNORM procedures are covered in detail in Chapter 3. The third part of the review (Section 2.4) covers current approaches

to test for equivalence in the analysis of cross-cultural research data, and includes the Item Response Theory approach. Section 2.5 considers the development of the Schlinger (1979) VRP, which is used extensively in this research, and contrasts the VRP with other available measurement instruments to measure the emotional response of viewers to television commercials.

2.2 BIAS IN DIFFERENT PHASES OF CROSS-CULTURAL RESEARCH

Since bias is a major threat in cross-cultural studies, the control and understanding of bias is extremely important. Bias is pervasive in research, and in quantitative research designs, explicit efforts should be taken to eliminate bias at all phases of a study. When bias is ignored, the validity of research findings is threatened. Several different types of bias could be present in any study, for example, design bias, measurement bias, sampling bias, procedural bias and problem bias (Anonymous, 2008), to name some of the most important sources of bias.

Design bias may be introduced when the researcher fails to address or explore issues that may threaten the validity of the study. The presence of research bias in a study tends to be an issue when researchers neglect to discuss the implications of a specific research design, and can almost never be completely avoided.

Measurement bias can be present, for example when respondents tend to give socially desirable responses, (Bernardi, 2006). Measurement bias can also be introduced by other types of response patterns or problems that may be present such as extremity ratings, where respondents display a tendency to use both the high or low end of the scales and avoid the categories in between, or acquiescence/disacquiescence bias, where certain respondents have a tendency to stick to either the higher end or lower end of the scale, thereby causing a shift in the scale location. Baumgartner and Steenkamp (2001) give a thorough exposition of problematic response styles in marketing research, as will be covered in more detail later in this chapter.

Sampling bias may be introduced when the sampling procedure favours the inclusion of a group of respondents that may be biased towards a specific opinion. An example of bias introduced during sampling is non-response bias, where for example in a postal survey, the respondents that choose not to participate in a survey could be those biased towards a specific viewpoint.

Procedural bias or method bias may be introduced when different methods of data collection are used in the same study, for example when, within a single study, one group completes an online survey, and another group receives a paper-based survey. Method bias can arise from the particular characteristics of the instrument or in the way that it is administered. Differences in physical conditions during administration can also lead to method bias, for example, if one group is examined in a very noisy environment. Communication barriers between the respondents and the interviewer are also potential sources for method bias, (Van de Vijver and Leung, 1997).

Cole, Bedeian and Feild (2006) used multi-group confirmatory factor analysis (MG CFA) to test the equivalence of paper-based versus web-based surveys in measures of transformational leadership. Method bias was also studied by Holbrook, Green and Krosnick (2003), where differences across national samples using face-to-face interviewing, and random digit dialing telephonic interviewing were examined using MG CFA.

In most research endeavours, covariances or correlations between constructs are of interest, as well as differences in mean levels, (Little, 1997). It is of primary concern to marketing practitioners and scholars to know whether variables of interest are perceived differently across different ethnicities in order to have more focused marketing efforts, and whether the associations between constructs are similar across groups. The presence of bias may artificially inflate or deflate correlations, and bias may also seriously distort means. It is therefore of paramount importance that that researchers detect and understand bias, and when bias is suspected to be present, to apply analytical methods that take this bias into account.

2.2.1 Measurement bias

There is some confusion with respect to the terminology used in the literature, for example “method bias” as described by Van de Vijver and Leung (1997) includes bias introduced by research design, procedural bias and response style bias.

According to Van de Vijver and Leung (1997), method bias implies a difference over the entire instrument being used. Often, method bias is present in the data when means over cultural groups in a t-test, for example, are significant. It may also happen that there is a significant main effect of cultural group in an analysis of variance (assuming that the method bias is sufficiently large to reach statistical significance). “We do not adhere to the frequently adopted practice of treating such significant effects ... invariably as indicators of valid differences. In our view, significant cross-cultural differences can be a mixture of valid cross-cultural differences and bias effects, in particular method bias" (Van de Vijver & Leung, 1997:16).

Therefore, absence of statistically significant differences between cultures is also not necessarily an indication of similar attitudes, since it is very difficult, and often impossible to estimate the offset between scales.

Baumgartner and Steenkamp (2001) refer to measurement bias, as the types of bias that are directly associated with differences in response styles. Of importance in this study are three types of bias that are directly related to *measurement bias*, an umbrella term that includes construct bias, response bias and item bias.

2.2.2 Construct bias

Construct bias is present when the construct is not identical across cultural groups. An example is the concept of "intelligence". In non-Western societies the concept is generally broader and may include reasoning and social skills. A very good example of construct bias over different cultures, is given by Cheung and Rensvold (2000), where the construct or notion of “filial piety”, a general respect

for parents and ancestors, is wider in scope in the Chinese culture than it is in the West, requiring researchers to employ more items in the measurement instrument to be used in the Chinese culture in order to tap its full dimensionality.

Construct bias may be introduced where there is a lack of overlap in the behaviours associated with the construct between the cultural groups being studied (Van de Vijver and Leung, 1997), or when the designed instrument does not cover the construct sufficiently, as may be the case in shortened scales. As noted by Myers, Calantone, Page and Taylor (2000:108), "... cross-national researchers must approach the measurement of relevant constructs with caution and realize that these constructs can be perceived uniquely across country populations." Unfortunately, there are no analytical methods to test for construct bias, as noted by Cheung and Rensvold (2000).

This researcher is of the opinion that construct bias is not a major problem in studies reported in peer-reviewed journals. Construct bias may be a problem when the researchers are very inexperienced, the measurement instrument is not adequately developed, and the study is poorly executed. [Note that one of the delimitations mentioned in Section 1.6 is that this research will not address problems associated with construct bias.]

2.2.3 Response bias

The importance of taking cognisance of the potential effects of response bias in research across cultural groups has been highlighted by several researchers (Hui and Triandis, 1989; Cheung and Rensvold, 2000). "Response styles are a source of concern in both domestic and international marketing research because they threaten the validity of empirical findings by contaminating respondents' answers to substantive questions" (Baumgartner and Steenkamp, 2001:143).
Differential response styles such as acquiescence and extremity ratings can introduce method bias in a study.

" ... what may appear to be cross-cultural differences (or similarities) might just be reflecting cross-cultural differences in response style; some method of detection

and removal of their influence must be employed if cross-cultural comparisons ... are to be valid." (Smith and Reynolds, 2002:466).

A primary source of bias in cross-cultural research is differential response styles, which are distinguished from bias arising from content of questions. Baumgartner and Steenkamp (2001) have identified seven important response styles that can have an effect on the results obtained in a study. The seven styles and an explanation of each style is summarised in Table 2.1.

Table 2.1: Response styles in marketing resulting in systematic error

Response Style	Definition and Synonyms	Reasons for response style
ARS (Acquiescence Response Style)	A tendency to agree with questions regardless of content. Also mentioned in literature as agreement tendency, yea-saying or positivity.	Stimulation seeking extroverts who accept statements impulsively. Uncritical endorsement of statements. Common for ambiguous items, or where respondents are uncertain about item content. Respondents may lack cognitive resources due to distractions, time pressure, and so on.
DARS (Disacquiescence Response Style)	A tendency to disagree with questions, regardless of content. Also called disagreement tendency, nay-saying or negativity.	Characteristic of controlled and reflective introverts who try to avoid stimulation.
NARS (Net acquiescence Response Style)	A tendency to show greater tendency for yea-saying than for nay-saying, also called directional bias.	A summary measure based on ARS and DARS.
ERS (Extreme Response Style)	A tendency to endorse the most extreme response categories regardless of content.	Personalities with rigidity, intolerant of ambiguity, associated with high levels of anxiety. Also present when respondents have different cognitive structures and underdeveloped schemas.
RR (Response Range)	A tendency to either use a wide or narrow range of response categories around the mean response.	ERS and RR often seen as very closely related.
MPR (Midpoint responding)	A tendency to use the middle of the scale, regardless of content.	Due to evasiveness, indecision or indifference.
NCR (Noncontingent responding)	A tendency to respond to items carelessly, randomly, or nonpurposefully.	Lack of interest in topic and lack of motivation to read item content.

Source: Baumgartner and Steenkamp (2001)

Bernardi (2006) studied the associations between social desirability response bias and how they affect the outcomes on cultural construct over 12 countries and gender groups, and found that social desirability response bias decreases as the individualism in the cultural group increases. According to the description given by Bernardi (2006), social desirability response bias is similar to acquiescence and disacquiescence response bias.

Cheung and Rensvold (2000) demonstrated the effect of response styles in the multi-group confirmatory factor analysis context, and they focused on ERS (extreme response style) and ARS (acquiescence response style), which is a more holistic treatment of the first six of the seven styles identified by Baumgartner and Steenkamp (2001). In the Cheung and Rensvold (2000) study, ERS, RR and MPR were treated as two sides of the same coin, and similarly ARS, DARS and NARS were treated as another coin with two sides. Cheung and Rensvold (2000) referred to “High ERS” and Low “ERS” to describe the ERS style of responding, and suggested an approach to test for the presence of this response style using multi-group CFA. Similarly they referred to “High ARS” and “Low ERS” to describe the ARS response pattern, and also suggested an approach to deal with this in the multi-group CFA framework. Therefore, within the CFA context, the problem of six of the seven different response styles covered by Baumgartner and Steenkamp (2001) can be simplified to a situation where the presence of these styles can be identified by using a series of hierarchical hypotheses. The method suggested by Cheung and Rensvold (2000) is basically the same method suggested by Steenkamp and Baumgartner (1998). Cheung and Rensvold (2000) cover the topic using a very thorough exposition of the response style problem, and suggest approaches on how to treat the data when only one or a few items within a construct are affected by the problem.

The consequences of response bias should not be underestimated. As Baumgartner and Steenkamp (2001:153) mentioned: “The findings suggest that stylistic responding can have systematic and substantial effects on correlations between scales and that researchers should be concerned about the deleterious effects of stylistic responding on research results.”

2.2.4 Item bias

Item bias as described by Van de Vijver and Leung (1997) could be indicated when specific items show bias, due to translation problems or when an item represents an additional trait in one or more cultures in the investigation. Item bias or *differential item functioning* is usually a problem that affects a few items in the scale. Item bias may be introduced when item content is inappropriate, for example when item formulation or item translation is inadequate, and when there is a difference in the understanding of the item content between cultural groups. Differential item functioning is something that can be specifically tested in an IRT framework. Also, in SEM, partial invariance (Byrne, Shavelson and Muthén, 1989) can be tested where one or two item parameters are not constrained in the sequential hypothesis testing approach.

2.2.5 Measurement bias summarised

The key aspects of measurement bias are summarised in Table 2.2.

Table 2.2: Different sources of bias in cross-cultural research

Potential sources of bias
Construct bias (Presented as very different factor structures, the constructs are different over cultural groups) The concept has a different meaning across cultures. The correlation structure between items is different across cultural groups. This is a very serious source of bias, and the measurement instrument cannot be used to compare concepts over cultures.
Method bias (Presented over the entire instrument and most items) Differences due to the way that the instrument is administered, for example: physical environment; interviewer styles and communication problems; differences in response styles such as acquiescence bias, [acquiescence bias: the tendency of the respondents from one cultural group to agree more strongly or disagree more strongly with statements]; and extremity ratings, the tendency of a certain cultural groups to mark the extremes of the scale.
Item bias (Presented over single items) Differences due to for example: the appropriateness of the item content, where the item content has a sensitive other (even improper) connotation; inadequate item formulation such as complex wording of the item; and translation problems.

Source: Summarised from Van de Vijver and Leung (1997)

2.3 BIAS/NON-INVARIANCE AND EQUIVALENCE/INVARIANCE

“Equivalence and bias are important issues that need to be addressed before meaningful cross-cultural comparisons can be made” (Fischer, 2004:263).

The terms bias, equivalence and invariance are used to describe different aspects of *measurement bias*. Bias is more concerned with all aspects that could threaten the validity of a study. The terms equivalence and invariance (or non-invariance) have become more commonly applied in the context where measurement bias may be present. The terms “invariance” and “equivalence” are frequently used in the application of multiple group confirmatory factor analysis. These terms have also been used in the context of cross-cultural research where item response theory is applied (see for example Reise *et al*, 1993). Although bias and equivalence may be viewed as opposites, the terms equivalence and invariance have become more frequently used in contexts where the measurement model is expressed as an outcome of an underlying latent variable or latent trait.

"Equivalence is a function of the characteristics of an instrument and of the cultural groups involved" (Van de Vijver and Leung 1997:9). At least three levels of equivalence are relevant. The first level is *construct equivalence* or *structural equivalence.* This can be confirmed using separate exploratory factor analyses for each cultural group, and, if similar factor patterns emerge, structural equivalence can be assumed. Another method is to use multi-group confirmatory factor analysis and assess the fit of a model across different cultures.

The second and higher level of equivalence is *measurement unit equivalence.* The measurement units are identical, but the origins of the scales are not. An attempt to determine an offset of scales in cross-cultural studies has been proposed by Baumgartner and Steenkamp (2006), with an alternative approach in Baumgartner and Steenkamp (2001).

The third and highest level of equivalence is called *scalar* or *full score comparability.* This level of equivalence can be achieved when the measurement instrument is on the same ratio scale for each cultural group, and shares the same origin for the groups studied.

In empirical studies, the level of equivalence is usually unknown. It is therefore essential that in all cross-cultural studies, researchers should not merely assume equivalence, but should identify the type of bias that may be present, and adjust the subsequent analyses accordingly.

Since measurement bias is often unavoidable, the only option for the researcher is to test for equivalence after data collection. During research design and data collection, all potential sources of bias can be considered beforehand. Where possible, sources of bias can be eliminated. If bias is suspected to be present in data, since it is very often impossible to completely avoid bias, methods to deal with bias should be applied. It is not appropriate to assume that bias does not exist. Researchers need to follow explicit procedures to deal with bias in research, (Reise *et al,* 1993; Smith and Reynolds, 2002; Fischer, 2004; Van Herk *et al.,* 2005)).

Cheung and Rensvold (2000) offer an exceptionally good explanation of the need to test data for equivalence, and shows clearly how it should be tested with the multi-group confirmatory factor analysis method. They offer useful advice of how to proceed if only parts of an instrument can be assumed to be equivalent. When scores are equivalent or invariant across groups, it is valid to compare score means and covariances or correlations. Equivalence is, therefore, achieved in a study when bias is absent, or can be accounted for, or can, after assessment using explicit procedures, be assumed to be negligible for all groups studied. Equivalence and/or invariance are a result of the absence of measurement bias.

1. The presence of any bias between cultural groups destroys at least scalar equivalence. The origins of the scales are different and the offset between

the scales is unknown. The estimation of the offset of the scales is difficult or impossible, depending on research design.

2. Construct bias destroys all levels of equivalence and no comparisons are permitted.
3. The complete absence of all sources of bias leads to full equivalence, and comparisons can be made between groups without data transformations.

2.3.1 Uniform and non-uniform bias

In the case of method and item bias, a distinction has to be made between uniform and non-uniform bias. *Uniform bias* is present when the bias is more or less the same across score levels. "Uniform bias will not threaten measurement unit equivalence" (Van de Vijver and Leung, 1997:19). *Non-uniform bias* refers to influences that are unequal for all score levels. (Score levels are explained in more detail in Section 2.3.2). "Non-uniform bias will destroy equivalence to a considerable extent because the measurement units in the two groups are no longer the same. When several items show this kind of bias, cross-cultural comparisons are likely to produce incorrect results" (Van de Vijver and Leung, 1997:19). Table 2.3 presents the researcher's interpretation about the differences between uniform and non-uniform bias, as described in Van de Vijver and Leung (1997).

Table 2.3: The difference between uniform and non-uniform bias

Uniform bias The differences are more or less equal across score levels for the cultural groups. This means that there exists an unknown offset between the scales, but the scale lengths are comparable. Uniform bias can be compared to the use of rulers for each group, with equal interval lengths over the tick marks that indicate the scale points, but with both the origins of the rulers and the relative offset of the rulers unknown.
Non-uniform bias The differences are not equal across score levels for the cultural groups. This could be compared to using rulers for each group which have been distorted, (as if they had been exposed to extreme heat) so that the comparable interval lengths over the tick marks do not exist and the origins of the rulers are also unknown. Both the offset of the scales and the interval lengths are unknown.

Source: Researcher's interpretation from Van de Vijver and Leung (1997)

2.3.2 Detection of uniform and non-uniform bias

Since this study only considers the transformation of numerical scores such as ratio, interval or ordinal, (and not nominal or dichotomous scores), only the method that is used for numerical scores is discussed here.

"An item is an unbiased measure of a theoretical construct, say, dominance, if persons from different cultural groups who are equally dominant have the same average score on the item. In more formal terms, persons with an equal standing on the theoretical construct underlying the instrument should have the same expected score on the item, irrespective of group membership" (Van de Vijver and Leung, 1997:62).

The method used for item-bias detection is an extension of the use of an analysis of variance method proposed by Cleary and Hilton (1968). The method will be described using the same terms as explained by Van de Vijver and Leung, (1997). Consider a 20-item questionnaire using a 5-point Likert scale that was administered to two samples of 500 each from Culture A and Culture B. Assume that the data can be considered to be on the interval scale. Score levels are computed as in the description which follows.

Reverse negative phrased items and compute the total score on the instrument, which is the sum of all the item scores. The minimum score is $20 \times 1 = 20$, and the maximum score is $20 \times 5 = 100$. These two score groups are ignored, because they are identical across cultural groups. The remaining 79 score groups with scores (21, 22, ..., 99), provide the information necessary for item bias analysis. Score levels are concatenated so that there are no fewer than 50 persons in each score group. It is recommended that the number of groups in the combination effort is not too small to allow for a statistically powerful analysis of item bias, and score groups of least 50 persons in a group are recommended as a minimum. (Clauser, Mazor and Hambleton, 1994). The score widths can be unequal to allow for a grouping so that the numbers of subjects in a group are as similar as possible.

The analysis of variance model fitted is done independently for each item, and uses the item score as the dependent variable. The independent variables are the score level, the cultural group as well as the interaction between the score levels and the cultural group. In the analysis, the significance of three effects is tested. The first is the significance of the score level, the second is the main effect of culture and the third is the interaction effect between culture and score level. The first effect of score group is seldom of interest. These are almost always significant, indicating that the higher levels of the score are associated with higher levels on the item. (If the score level is not a significant predictor for the item, it indicates that the item does not discriminate across the score levels, and may need to be dropped.) When both the main effect and the interaction effect are non-significant, the item is taken to be unbiased. The typical pattern is shown in Figure 2.1(a), where the vertical axis presents the difference between the item means for the cultural groups at each score level. Unbiased items will have patterns that do not differ from the zero line in any systematic way. In the case of item bias, the departure from zero will be systematic. A significant main effect for culture is indicative of item bias over cultural group, since the individuals from one cultural group have higher scores on the item than individuals from the other cultural group, even though they have the same total score. A significant main effect of culture implies that the difference in item means plotted against score group, is consistently above or below the zero line, as shown in Figure 2.1(b). This indicates uniform bias.

A significant interaction effect between culture and score grouping indicates that non-uniform bias is present. The differences between cultural groups vary over score levels, and the item discriminates better in one group than it does in the other on the total construct. The pattern is shown in Figure 2.1(c), and differences in average scores depart systematically from zero, but they are not uniform across score levels. There is a tendency to increase or decrease systematically, or it may show any other systematic pattern.

Figure 2.1 (a): Unbiased item

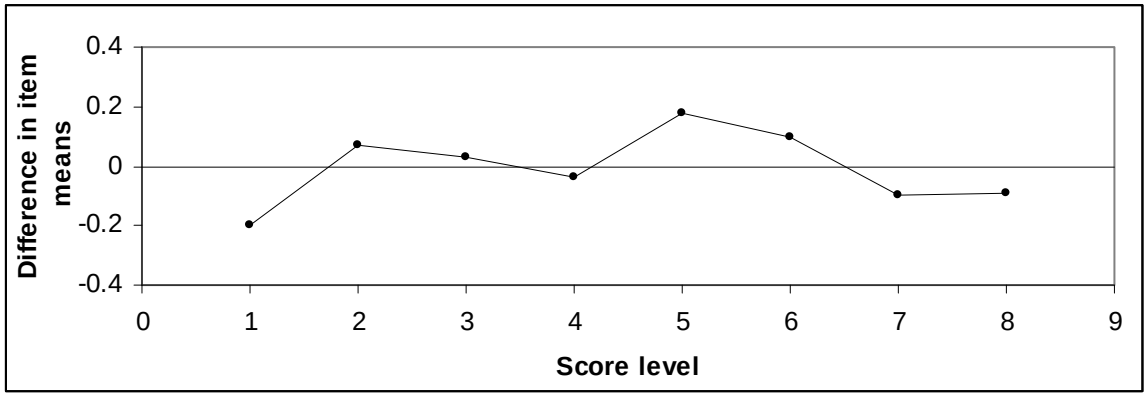


Figure 2.1 (b): Item with uniform bias

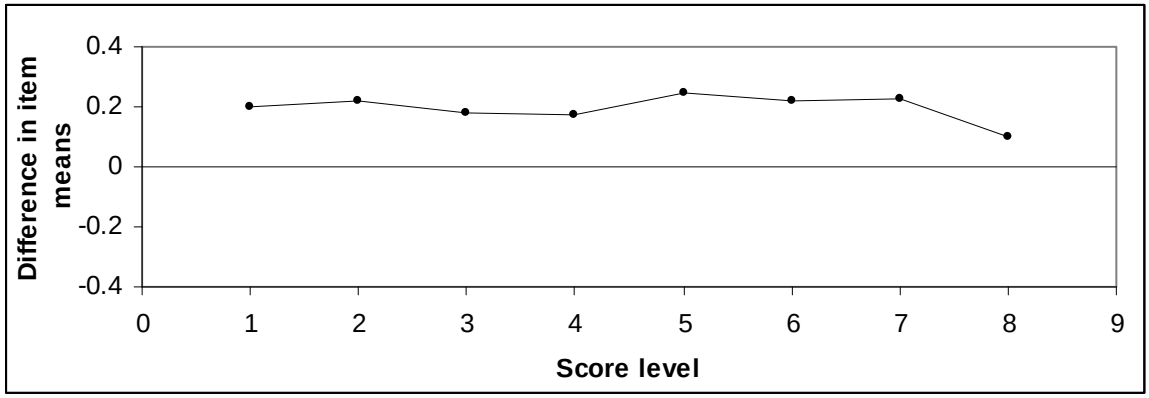
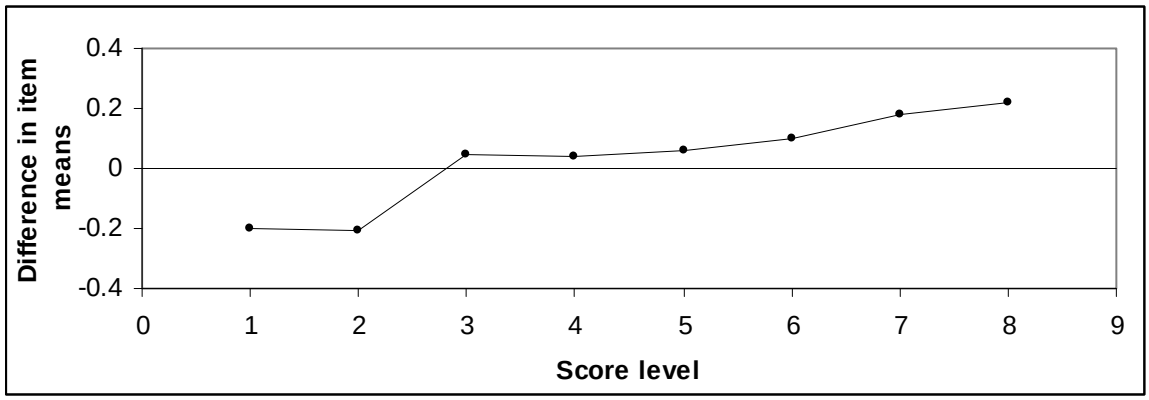


Figure 2.1 (c): Item with non-uniform bias



Source: Van de Vijver and Leung (1997)

When bias is present, the level of equivalence is lowered (Van de Vijver and Leung, 1997). When construct bias is present, comparisons across cultural groups

are not valid, and should be avoided. Uniform bias does not threaten measurement unit equivalence, and the source of bias is basically differences that are reasonably constant over score levels. Non-uniform bias destroys both measurement unit and scalar equivalence. This is because the measurement units in the groups are no longer the same length, and the differences vary across score levels. When several items in the cross-cultural study show non-uniform bias, cross-cultural score comparisons are invalid and will produce incorrect results.

1. Uniform bias is less serious than non-uniform bias, and transformations such as standardisation that align the scales are permitted. An aim of this research is to investigate whether SIMITNORM is better than standardisation in the presence of uniform bias.
2. According to Van de Vijver and Leung (1997), comparisons are not generally permitted in the presence of non-uniform bias, and this research will investigate whether SIMITNORM can be used to adjust scores that present non-uniform bias.

2.4 MEASUREMENT FROM A CLASSICAL TEST THEORY PERSPECTIVE

In cross-cultural research, the attitudes of respondents are often measured as a construct, for example customer satisfaction or service quality, which may constitute underlying sub-dimensions. Often, several items are used to indicate the dimensions. In the framework of classical test theory, one can assume that at any given point in time, every respondent has a “true” level of customer satisfaction, denoted by $X_{T.}$ In ideal circumstances, the observed score X_O, is equal to the true score, X_T. However, rarely are researchers so fortunate. More typically it is found that the measurements observed, X_O, also reflects other characteristics, such as mood, willingness to express true feelings, issues related to situational factors that may be different for subgroups in the sample, differences in how cultures interpret and use and interpret the scales in the questionnaire items, differences in methods to administer the measurement instrument, different interpretations of the formulation of questions or items, and so on. Although not all these factors are

present in every cross-cultural study, their impact may be predictable or systematic, and they tend to distort the observed scores away from the true scores (Trochim, 2006). The relationship can be functionally expressed as:

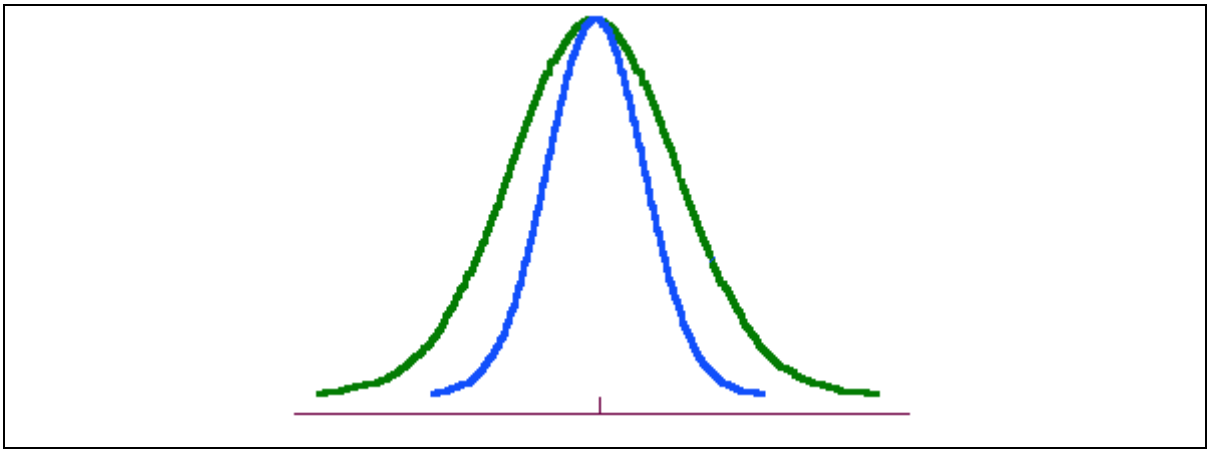
X_O = $X_T + X_S + X_R$ with

X_S = systematic sources of error (such as response bias) which affect the true score, and

X_R = random sources of error such as transient personal factors which affect the true score.

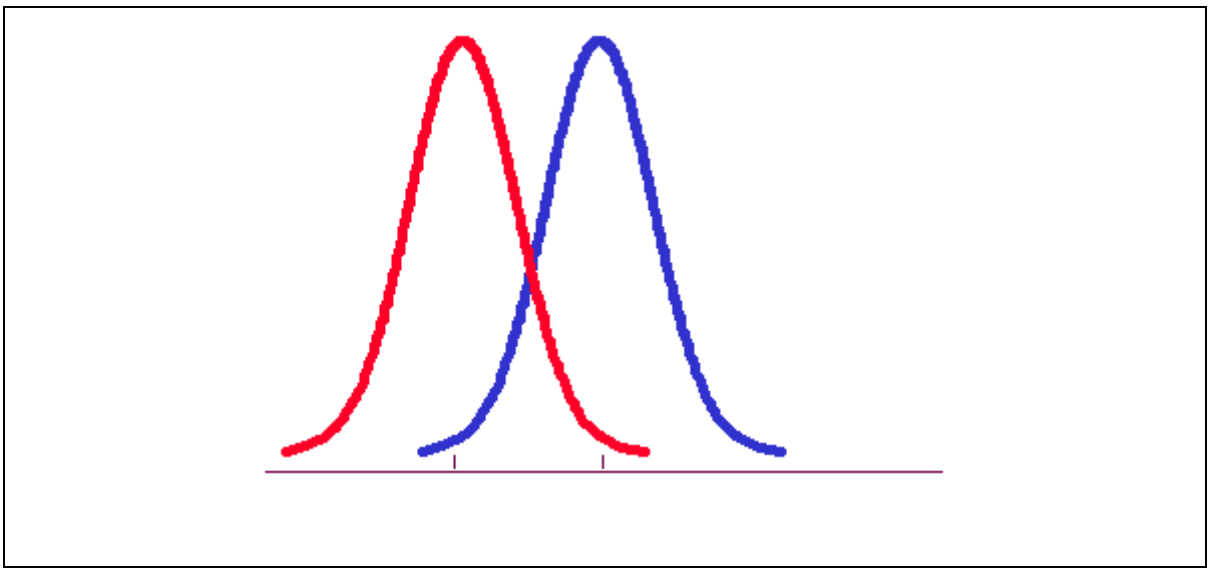
Random error is assumed not to have a consistent effect over the entire sample. Therefore, if it were possible to quantify random error, it would add up to zero across the entire sample. The only effect of random error is that it adds variability to the data, but it does not affect the average. Random error does not have consistent effects across the entire sample, as shown in Figure 2.2.

Figure 2.2: Random error increases the variability of the distribution, but the location remains the same.



Systematic error, in constrast, is caused by any factor that systematically affect the measurement of a sample across the sample. Systematic errors tend to be consistently negative or positive, and systematic error is often presents itself as *measurement bias*, provided that all other potential souces of error were carelfully eliminated. As shown in Figure 2.3, systematic error affects the location of the sample distribution, but the variability within the sample remains fixed. This is typical of an item with uniform bias. If an item has non-uniform bias, both the location and spread parameters will be affected.

Figure 2.3: Systematic error does not affect the variability of the distribution, but there is a difference in the location of the distribution



2.4.1 Measurement error

The term measurement error is commonly found in the structural equation modelling / confirmatory factor analysis (SEM/CFA) literature, and requires clarification. Measurement error appears in physical sciences and social sciences. For example, in the physical sciences, if the lung capacity of a patient is measure through a device, it is often necessary to take three to five measures for the same person, since the method to measure lung capacity is not very reliable. Taking several measures allows the physician to obtain a "more or less true" estimate, by taking the average of the measures. In the context of the social sciences, measurement error can be reduced by pilot testing research instruments, ensuring consistent data collection and interviewer methods, making sure data is captured accurately, using multiple indicators of the same construct (have at least two questions that refer to a construct) or use statistical procedures that allow for measurement error.

Baumgartner and Steenkamp (2001) pointed out that although it is well known that measurement error, which is assumed to be a random type of error, can attenuate

or disattenuate correlations between constructs, it is not well known that systematic error can also artificially inflate or deflate correlations. It is also possible that these errors can lead to estimated correlation coefficients with incorrect signs.

Very few attempts to isolate systematic error, or bias have been reported in the literature. One main problem arises from the fact that estimation of bias in measurement requires a very specific research design with repeated measures, such as a panel study (c.f. Baumgartner and Steenkamp, 2006), or instruments that have a reasonable balance between positively phrased and negatively phrased items for the same construct, as has been demonstrated by (Baumgartner and Steenkamp, 2001).

In an extensive study, Baumgartner and Steenkamp (2001) have used a data set with 10 477 observations over eleven countries in the European union, using a large variety of constructs consisting of 60 items, which included consumer behaviour issues, the media behaviour of respondents, and a few demographical variables. The 60 items were all measured on a five-point Likert scale, and included constructs such as attitude towards advertising in general, attitude towards the past, change seeking, consumer ethnocentricity, environmental consciousness, price consciousness, deal proneness, environmental consciousness, exploratory acquisition of products, exploratory information seeking, health consciousness, and quality consciousness. All these items were used from validated scales, and hierarchical linear modelling was applied to test for the presence of certain types of response styles. The study demonstrated that certain response style biases have had an effect of introducing weak, but significant correlations between variables that were difficult to explain from a theoretical perspective, but once the data had been corrected for bias, the estimated correlations were not significant, which was more in line with theory.

In subsequent work, using panel data, Baumgartner and Steenkamp (2006) have suggested a SEM/CFA approach to estimate systematic error. They also reiterate the need to consider equivalence: “Several authors have discussed the necessary conditions of measurement equivalence (e.g., Steenkamp and Baumgartner 1998), but it is rare for authors to conduct the required invariance tests before they

embark on the substantive comparisons of interest" (Baumgartner and Steenkamp, 2006:433).

"Corrected covariance or correlation matrices based on purified scale scores can serve as input to tests of structural equation models. Failure to apply these corrections may lead to spurious results when analyses are based on contaminated scale scores or correlations between scales." (Baumgartner and Steenkamp, 2001:155).

2.5 STATISTICAL TECHNIQUES USED IN CROSS-CULTURAL RESEARCH

The analysis of cross-cultural data often consists of two stages. The first stage involves the psychometric adequacy of the instrument by assessing the reliability and validity of the instrument. This could also involve an assessment to detect item bias and the nature of the bias (uniform or non-uniform). The second stage of the study often involves the main issues by exploring the research questions and testing hypotheses. The statistical techniques may be structure-orientated or level-oriented. Structure-orientated techniques such as exploratory or confirmatory factor analysis, structural equation modelling, multi-dimensional scaling or cluster analysis can be used. Level-orientated techniques may include techniques such as t-tests, analysis of variance, regression analysis and analysis of covariance.

2.5.1 Preliminary analyses

Before cross-cultural comparisons can be made, it is important to investigate the psychometric properties of the instruments and items across groups.

If significant differences in the reliability coefficients between the cultural groups are found, the items should be examined individually. This can be done using item-total correlations for the cultural groups separately. Differences in the shapes of the score distributions across cultural groups may also point to various problems with the instrument or how it was administered. These methods are all

widely available in most common statistical packages, and are an essential the first step in the analysis of cross-cultural data (Van de Vijver and Leung, 1997).

Steenkamp and Van Trijp (1991) suggested the use of confirmatory factor analysis in the validation of marketing constructs. The advantage of this method is that it provides a powerful and versatile tool for a more detailed and critical examination of construct validity. This method will be employed in Chapter 5 to asses the construct validity of the data used in the study.

2.5.2 Standardisation methods

Leung and Bond (1989) proposed different standardisation procedures. Standardisation is the computation of z-scores using the well-known standardisation transformation $z = [\ X - M\ ] / S$, where X is the score to be standardised, and M and S are the mean and standard deviation of the cultural group. The aim of standardisation is the reduction and the elimination of unwanted cross-cultural differences. If standardisation is used, it requires proper justification, because the eliminated differences may include valid differences. Often the researcher assumes equality of averages across cultures, and if it cannot be justified, a more prudent approach would be to analyse both standardised and raw data to compare how the structures are affected by the standardisation (Van de Vijver and Leung, 1997).

Fischer (2004) provides an extensive review of different standardisation methods that have been applied in several psychological studies in order to adjust scores for response bias in cross-cultural research. The implications of standardisation methods, and how these affect interpretation of results are also were also covered by Fischer (2004). As has been mentioned by Van de Vijver and Leung (1997); Byrne and Campbell (1999) and Cheung and Rensvold (2000), equivalence and bias issues cannot be ignored when meaningful comparisons need to be made across groups. According to Fischer (2004), Hofstede (1980) was one of the first researchers who used standardisation to correct for response tendencies that

resulted in bias. In contrast to this view, Smith (2004) as well as Van Hemert, Van de Vijver, Poortinga and Georgas (2002) argue that certain response styles are manifestations of culture, and standardisation would remove variation that is substantial and culture related. Despite these opposing viewpoints, depending on the research purpose and other methods required for analysis, Fisher (2004) noted that standardisation can be very useful to adjust scores. However, the rationale for the adjustments in scores, a justification for the type of standardisation method used and the implications of the adjustments should be provided when research results are interpreted.

Fischer further (2004) provided a classification of methods that can be used to standardise data. Some methods only involve adjustments at mean level, thereby producing deviation scores. Other measures allow adjustments only using standard deviations, whilst other approaches use both means and standard deviations. The standardisation methods also vary according to the level at which adjustment is done. Adjustment could be done either for individuals, for groups, or both. This is achieved by either using the means across all variables for individual adjustments, or across individuals within a group, or both. An adjustment of the individual scores from their own mean across items is also known as ipsatisation scores (Hicks, 1970). When responses are adjusted according to the group means and standard deviations, the adjusted scores within a cultural group will have a mean of zero and standard deviation of one. This is the method that was used as a benchmark in this research, since it is best suited to remove bias within a group.

Fischer (2004) extended his study by conducting a review of all articles that appeared in the Journal of Cross-Cultural Psychology between 1970 and 2002 to explore the number of studies that have used standardisation, the types of standardisation methods used, the nature of the statistical analyses employed after standardisation, and the rationale provided by the researchers for the standardisation method used. He found 35 articles using standardisation as a method to adjust scores over this period, showing a gradual increase over the period. The statistical methods after employing standardised scores were mostly MANCOVA/MANOVA or ANCOVA/ANOVA; whilst a few studies also applied

factor analysis, correlations, discriminant analysis and non-parametric measures. Very few researchers provided theoretical explanations for the rationale behind the standardisation method. Group mean centering/standardisation implies that subsequent comparisons of means across groups have to take the method of standardisation into account. “Therefore, standardization can be necessary and useful for researchers, and the importance lies in the appropriate interpretation of the results obtained” (Fischer, 2004:277).

2.6 ALTERNATIVE APPROACHES TO THE NORMALISATION PROBLEM

2.6.1 The problem of non-normality

The statistical literature describes two main approaches towards dealing with the problem of non-normality. One mainstream approach is to accept that the data are not normal, and develop parameter estimation methods that are less sensitive for departures from non-normality (see Browne, 1984), however, there is no clarity in the literature on the best estimation methods. “If the variables are highly non-normal, it is still an open question whether to use ML (or GLS) or WLS with a general weight matrix. … Previous Monte Carlo studies have not given a clear-cut answer as to when it is necessary to use WLS rather than ML” (Jöreskog and Sörbom, 1996a: 250). Another alternative to the problem of non-normality includes the use of polychoric correlations and using weighted least squares estimation (Jöreskog, 1994). This method assumes that the data are ordinal.

The second approach to non-normality that can be found in the literature is to transform the data through normalisation, an option in the LISREL software (see Jöreskog and Sörbom, 1996b), which uses the approach suggested by Johnson and Kotz (1970). This method ensures that the weighted mean of the normal scores is zero. The main difference between this method compared to SIMNORM and SIMITNORM is that the method is only aimed at transforming ordinal scores to normalised scores, and has not been developed with the aim of equating scores across groups. It may be possible to use the method across groups, but the

limitation would be that it would only be able to equate the groups so that the weighted mean of each group would be equal to zero, and not that the items are individually equated.

2.6.2 The Stacey (2005) normalisation approach

The Stacey (2005) method, which is based on a distribution fitting algorithm that minimises the χ^2- goodness-of-fit measure for estimating the parameters used in normalisation, has been developed as an alternative to rescaling ordinal level survey data. Stacey (2005) compared his method to the Bendixen and Sandler (1995) correspondence analysis approach for rescaling data, and against the method that assumes an interval level scale, which assumes that the data at ordinal scale level is at interval scale level. Stacey (2005) found the distribution fitting approach to be accurate (reliable) and valid and that the method was superior to alternative approaches.

The Stacey (2005) method considers the frequency distributions over a number of items that were used in a perception survey, for example, using an ordinal response scale, such as the Likert scale. The parameter estimation method includes the estimation of threshold values for the entire set of items, as well as two additional location and spread parameters for each item. The implied resrictions in the estimation procedure are that the overall mean across items is equal to zero, and the combined total variance across items is one. Furthermore, the mean and standard deviation parameters of each item, including the thresholds that are held fixed across items, are estimated using a linear programming approach, and by minimisation of the χ^2- goodness-of-fit measure. The estimation uses “Solver” in Microsoft Excel.

A disadvantage of the Stacey (2005) approach is that the χ^2- goodness-of-fit measure has been combined across dependent “samples” (items). This is open for critique from a mathematical point of view. Furthermore, the issue of the degrees of freedom used in the estimation procedure is also a matter that should be resolved.

The major advantage of the approach by Stacey (2005) is its simplicity and the ease of use. Excel is widely available, and the approach is intuitively appealing and easily understood. The approach is to normalise across all items simultaneously, with the constraint that the combined mean across items is equal to zero, and the combined variance across items is equal to 1. This "standardisation" is also appealing, and it could easily be applied across cultural groups (or any other grouping or combination of grouping variables) to find comparable scores across groups. It would be interesting to examine this methodology in comparison with the SIMNORM and SIMITNORM methods in more detail.

2.6.3 Structure-orientated techniques: Exploratory factor analysis

Structure orientated techniques are often used to examine construct equivalence. Exploratory factor analysis (Lawley and Maxwell, 1971) is a multivariate method that can identify the structure of a set of variables, and it is useful in the process of data reduction. One of the underlying assumptions in factor analysis is that the variables are normally distributed. Factor analysis is more easily understood when one considers a correlation matrix, and where the extracted factors contain groups of variables that are among themselves highly correlated, and between the factors are low or nearly zero correlations. Factor analysis often involves a rotation of the principle axes, in order to obtain a more clearly defined factor pattern. The most popular and most widely used rotation method is Varimax rotation, which assumes that the factors are uncorrelated. (Hair, Anderson, Tatham and Black, 1998). The extracted factors often represent the constructs being used in a measurement instrument.

2.6.4 Level-orientated statistical techniques

Multivariate level-oriented statistical techniques are used to model relationships between variables and differences between subgroups. Well-known linear modelling approaches include multiple regression, and analysis of variance

(ANOVA) (Draper and Smith, 1981; Hair *et al.*, 1998). Since these methods are well-known and widely available in statistical packages, the methods are not further described in this literature review.

2.7 CONFIRMATORY FACTOR ANALYSIS

2.7.1 Introduction to confirmatory factor analysis (CFA) in a structural equation modeling (SEM) framework

A technique that is more frequently used for group comparisons in cross-cultural studies is confirmatory factor analysis (CFA) using structural equation modelling (see McGaw and Jöreskog, 1971). The advantage of this approach is that a wide variety of hypotheses of cross-cultural differences and similarities can be tested. This approach is useful to test for construct, measurement unit and scalar equivalence by testing the tenability of a series of increasingly restrictive models, using goodness-of-fit tests. The first hypothesis tested using this method, is that there are an equal number of factors across groups. The validity of a k-factor structure is tested across groups. If this hypothesis is tenable, the second hypothesis tests whether the factor loadings and intercepts are invariant (equal). If this hypothesis is not rejected, the third hypothesis tests whether factor covariances are invariant, and the fourth hypothesis whether item-pair reliabilities are invariant. (Byrne, 1998). When all four these hypotheses are tenable, it can be concluded that the measurement instrument is invariant (and by implication equivalent) across cultural groups. Byrne (1998) shows in considerable detail how these hypotheses may be tested using LISREL.

The multi-group method of invariance testing was popularised by the thorough treatment of the topic by Steenkamp and Baumgartner (1998) in the marketing discipline. Vandenberg and Lance (2000) conducted an extensive study on the various levels of hypotheses that can be tested and they present an excellent theoretical background revealing what each model restriction implies in terms of equivalence. Since CFA is such a core aspect in this research, the mathematical model is discussed in more detail.

2.7.2 The general CFA measurement model (single group case)

The basic relationship between the observed variable and the hypothesised underlying constructs for a single variable y_i to an item i ($i = 1, ..., p$) is represented as a linear function of a latent construct η_j ($j = 1, ... m$), an intercept τ_i and a stochastic error term δ_i .

Thus the first-order measurement model for the i-th indicator or observed variable is

$$y_i = \tau_i + \lambda_{ij} \eta_j + \delta_i \qquad (2.1)$$

with λ_{ij} the slope of the regression of y_i on η_j. Sorböm (1974) defines λ_{ij} as the metric of measurement, since it gives the amount of change in y due to a unit change in η_j .

If there are p items or variables and m latent constructs, the measurement model in matrix notation is

$$\boldsymbol{y} = \boldsymbol{\tau} + \boldsymbol{\Lambda}_x \boldsymbol{\eta} + \boldsymbol{\delta} \qquad (2.2)$$

where $\boldsymbol{y}$ is a $p \times 1$ vector of observed variables
$\boldsymbol{\tau}$ is a $p \times 1$ vector of intercepts
$\boldsymbol{\Lambda}_x$ is a $p \times m$ LAMBDA-X matrix of factor loadings
$\boldsymbol{\eta}$ is an $m \times 1$ vector of latent variables
$\boldsymbol{\delta}$ is a $p \times 1$ vector of error terms.

It is further assumed that $E(\boldsymbol{\delta}) = \mathbf{0}$, $(p \times 1)$ and that $E(\boldsymbol{\eta}, \boldsymbol{\delta}') = \mathbf{0}$ $(m \times p)$, and the expected mean of the latent variables is defined as $E(\boldsymbol{\eta}) = \boldsymbol{\kappa}_\eta$.

The following variance covariance parameter matrices are defined:
$E[(\boldsymbol{\eta} - \boldsymbol{\mu}_\eta)(\boldsymbol{\eta} - \boldsymbol{\mu}_\eta)'] = \boldsymbol{\Phi}_\eta$ as the $m \times m$ matrix (PHI) of variances and covariances between the first-order latent variables; and

$E[\delta\delta'] = \Theta_\delta$ as the $p \times p$ matrix (THETA-DEL) of error variances and error covariances of the observed indicators $\boldsymbol{y}$ (usually a diagonal matrix, unless errors are correlated).

2.7.3 The general CFA model: Covariance structure (single group case)

In addition to the means structure set out in equation (3), the covariance structure needs to be specified. Ignoring the intercept part of the model, the model implied variance-covariance matrix $\Sigma(\theta)$ of $\boldsymbol{y}$ is expressed by

$$\Sigma(\theta) = E(\boldsymbol{y}, \boldsymbol{y}') \qquad (2.4)$$

$$= E[(\Lambda_y \eta + \delta), (\eta' \Lambda_y' + \delta')]$$

$$= E[(\Lambda_y \eta + \delta), (\eta' \Lambda_y' + \delta')]$$

$$= E[\Lambda_y \eta, \eta' \Lambda_y'] + COV[\Lambda_y \eta, \delta'] + COV[\delta, \eta' \Lambda_y'] + E[\delta, \delta']$$

$$= \Lambda_y E[\eta, \eta'] \Lambda_y' + \Lambda_y COV[\eta, \delta'] + COV[\delta, \eta'] \Lambda_y' + E[\delta, \delta']$$

$$= \Lambda_y \Phi_\eta \Lambda_y' + \Theta_\delta \qquad (2.5)$$

where Φ_η is the $m \times m$ (PHI) matrix of variances and covariances of the latent variables η, and

Θ_δ is the $p \times p$ (THETA-DELTA) variance-covariance matrix of the error terms δ, which is usually constrained to be a diagonal matrix.

2.7.4 Measures of fit

In SEM, several measures of fit are important to consider. Apart from the fit measures, it is imperative that the researcher uses his/her substantive knowledge about the theory tested in the model at all times during model specification. Further, once the model has converged and parameter solutions are obtained, each estimate should be examined. The magnitude and the signs of the parameters should be as expected by theory. Further, the standard errors of the parameters and the significance of the estimates should be evaluated. (Raykov and Marcoulides, 2000)

Model fit is assessed by the discrepancy between the observed covariance matrices **S,** and $\boldsymbol{\Sigma}(\hat{\boldsymbol{\theta}})$, denoted by $\hat{\boldsymbol{\Sigma}}$, the model implied covariance matrix evaluated at the estimate of $\boldsymbol{\theta}$ which minimises the estimation function F**.** Virtually all measures of fit involve some function of **S** and $\hat{\boldsymbol{\Sigma}}$, to evaluate the closeness between these two matrices. The residual matrix **S** - $\hat{\boldsymbol{\Sigma}}$, is perhaps the easiest and simplest assessment of model fit (Bollen, 1989).

In the SEM approach to CFA, the null hypothesis is stated in a way that is not the norm in most statistical analyses. The null hypothesis is that the model fits the data exactly, and in most cases, the researcher is interested in not rejecting the null hypothesis (Raykov and Marcoulides, 2000). Since this hypothesis is very restricting, it is not unusual to reject the null hypothesis. A further problem is that, in contrast to the case in most other statistical methods, rejection of the null hypothesis is more likely as the sample size increases. This fundamental logic of SEM does pose specific challenges in the evaluation of model fit. It is therefore important to follow a variety of approaches when model fit is evaluated.

The Chi-Square value is defined as, $T = (N - 1)\, F_{min}$, with N the sample size and F_{min} the minimal value of the fit function, depending on the estimation procedure used. There are basically three estimation methods used in SEM, and these are: Maximum Likelihood (ML), Generalised Least Squares (GLS), and Robust Maximum Likelihood (RML). The Chi-Square value, T is used as an inferential goodness-of-fit measure to test the null hypothesis that the model fits the observed covariance matrix perfectly. The model is rejected when the p-value is smaller than a preset significance value. (This is the opposite of, for example, regression modelling, where the model is accepted when the overall F-value is significant).

According to Raykov and Marcoulides (2000), this approach is compatible with Popperian logic, which states that the primary interest of research is to reject models, rather than confirming them, since there is no scientific way of proving the validity of a proposed model. Therefore there is a preference for dealing with models with a large number of degrees of freedom. “This is because an intuitive meaning of the notion of degree of freedom is a dimension along which the model

can be rejected. Thus, the more the degrees of freedom, the more dimensions there are along which one can reject the model, and hence the higher the likelihood of rejecting it when it is tested against the data. This is a desirable feature of the testing process because, according to Popperian logic, empirical science can only disconfirm and not confirm models. … Hence, Popperian logic, which maintains that a goal of empirical science is to formulate theories that can be falsified, is facilitated by an application of the parsimony principle. If a more parsimonious model is found to be acceptable, then one may also place more trust in it, since it has withstood a higher chance of rejection than a less parsimonious model" (Raykov and Marcoulides 2000:36-37).

The advantage of the Chi-Square value is that when two models are compared on the same set of data, the Chi-Square difference test can be applied. It involves the calculation of the difference between the Chi-Square values, and the difference in the corresponding degrees of freedom. The difference follows the Chi-Square distribution with degrees of freedom equal to the difference in degrees of freedom. It is therefore relatively easy to determine whether a more parsimonious model resulted in a significant improvement, by evaluating the p-value associated with the Chi-Square difference test. If the Chi-Square difference test is significant, the more parsimonious model offers a significant improvement over the more complex model (Byrne, 1998).

One primary difficulty with the Chi-Square test is that with large samples, there is a tendency to obtain very large Chi-Square values, which in turn are associated with very small p-values. If one therefore only uses the p-value as an indicator of model fit, there will be an artificial tendency to reject models that are only marginally inconsistent with the data. These limitations have led to a new plethora of other fit indices which all have their own merit. There are broadly two additional families of fit indices that have been proposed and implemented in SEM software, namely the *descriptive fit indices*, and the *alternative fit indices*.

In the family of descriptive fit indices, the first that was proposed in the SEM literature, was the goodness-of-fit index (GFI). It can be interpreted as the proportion of variance and covariance that the model is able to explain. This is

very similar to R-square in regression analysis. The adjusted goodness-of-fit index, (AGFI), which is similar to the adjusted R-square in regression, takes model complexity into account. GFI and AGFI range between 0 and 1, and when models fit well, these indices will be closer to 1. The cut-off point for when a model is regarded as fitting well, is in the 0.90 or above (Hu and Bentler, 1999).

Two other descriptive indices that are useful are the normed fit index (NFI) and the non-normed fit index (NNFI) proposed by Bentler and Bonnet (1980). These indices are compared to a model in which absolutely no interrelationships are assumed among any of the variables. This is referred to as the independence or null model. The NNFI takes model complexity in to account, similar to the AGFI, and the same cut-points are used as for the GFI and AGFI (Hu and Bentler, 1999). Apart from these indices, there are several other indices. A problem of these indices is that they are all some form of function of the Chi-Square value, and therefore could be overly strict. This has triggered the development of the alternative fit indices (Raykov and Marcoulides, 2000).

Another very useful criterion is the Bayesian Information Criterion (BIC) with BIC = $T_m - df_m \ln(N)$, Schwartz (1978) and Raftery (1995). If the BIC is equal to zero, it can be said that the model fits as well as the saturated model. If it is negative, it fits better. The smaller BIC (meaning larger negative numbers) the better the model fit. If BIC is positive, the model fits worse than the saturated model, and the larger the positive value, the worse the fit. An advantage of the BIC is that it can be used to compare different models, and it does not require the models to be nested as is necessary with chi-square difference tests.

The alternative fit indices are based on the non-centrality parameter (NCP), denoted by δ. The NCP reflects the extent to which a model does not fit the data. If the data comes from a normal distribution, and if the model is correct, and if the sample is large, T follows the central Chi-Square distribution. But if the model is incorrect, T follows the non-central Chi-Square distribution. A non-central Chi-Square distribution can be viewed as the central Chi-Square distributions with

values shifted δ units to the right. The NCP can be viewed as the degree to which the model fails to fit the data, so that the larger NCP, the worse the model.

A very popular and widely accepted index of model fit was proposed by Browne and Cudeck (1993). The Root Mean Square Error of Approximation (RMSEA), which is based on the NCP, takes model complexity into account, but has less rigid requirements for degree of fit. The primary principle of the RMSEA is that it evaluates the extent to which the model fails to fit the data. Another reason for the support in the literature for the RMSEA is that it is sample independent, and sample size does not have an effect on the RMSEA. Further, the RMSEA is not dependent on characteristics of the sample distribution. It is generally recommended that RMSEA should be less than 0.05 (Raykov and Marcoulides, 2000). RMSEA should be less than 0.05 for the fitted model to indicate a good approximation. Values between 0.05 and 0.08 indicate acceptable fit, values between 0.08 and 0.10 marginal fit, and values above 0.10 poor fit. A very useful aspect of the RMSEA is that confidence intervals for the measure can be calculated. If the confidence interval is not too wide, and both limits are below 0.05, it can be argued that the model is a plausible means of describing the data (Fabrigar, Wegener, MacCallum and Strahan, 1999). Raykov and Marcoulides (2000) indicate that it is important that the lower limit of the RMSEA confidence interval be considerably less than 0.05.

The RMSEA is not the only index that can be obtained as a direct function of the NCP. The comparative fit index (CFI), which ranges between 0 and 1, is also recommended to be in the mid 0.90s or above to indicate a good fit (Bentler, 1990). The expected cross-validation index (ECVI) was also introduced as a function of the NCP by Browne and Cudeck (1993). "It presents a measure of the degree to which one would expect a proposed model to replicate in another sample from the same population. In a set of several proposed models of the same phenomenon, a model is preferred if it minimises the value of the ECVI relative to other models" (Raykov and Macoulides, 2000:41). The Akaike Information Criterion (AIC) (Akaike, 1987) is an index of fit that takes both model complexity and measure of fit into account. In general, models with smaller ECVI and AIC values are preferable to models with higher ECVI and AIC values. In

small samples the ECVI is better than the RMSEA, since it is not desirable to have overly complex models in small samples, and the RMSEA tends to ignore sample size (Raykov and Marcoulides, 2000).

This discussion of the most important fit indices leads to the conclusion that, when CFA models are considered in a SEM framework, several fit indices should be considered when the fit of a model is evaluated.

2.7.5 The general CFA measurement model (multiple group case)

In multiple group analysis, which is usually of interest when the researcher is interested in testing invariance across groups, the models can be defined by extending the single group case. If there are p items and m latent constructs, applied over G groups, the measurement model for a specific group g ($g = 1, ..., G$) is

$$\boldsymbol{x}^{(g)} = \boldsymbol{\tau}^{(g)} + \boldsymbol{\Lambda}_{\boldsymbol{x}}{}^{(g)}\boldsymbol{\eta}^{(g)} + \boldsymbol{\delta}^{(g)} \qquad (2.6)$$

where $\boldsymbol{x}^{(g)}$ is a $p \times 1$ vector of observed variables in group g

$\boldsymbol{\tau}^{(g)}$ is a $p \times 1$ vector of item intercepts in group g

$\boldsymbol{\Lambda}_{\boldsymbol{x}}{}^{(g)}$ is a $p \times m$ LAMBDA-X matrix of factor loadings in group g

$\boldsymbol{\eta}^{(g)\backslash}$ is an $m \times 1$ vector of latent variables in group g

$\boldsymbol{\delta}^{(g)}$ is a $p \times 1$ vector of error terms in group g.

It is assumed that $E(\boldsymbol{\delta}^{(g)}) = \mathbf{0}$, ($p \times 1$) and that $COV(\boldsymbol{\eta}^{(g)}, \boldsymbol{\delta}^{(g)'}) = \mathbf{0}$ ($m \times p$ matrix), and that, furthermore, the expected mean of the latent variables is defined as

$E(\boldsymbol{\eta}^{(g)}) = \boldsymbol{\kappa}_{\eta}{}^{(g)}$, $E[\boldsymbol{\eta}^{(g)}, \boldsymbol{\eta}^{(g)\prime}] = \boldsymbol{\Phi}^{(g)}{}_{\eta}$ and $E[\boldsymbol{\delta}^{(g)}, \boldsymbol{\delta}^{(g)\prime}] = \boldsymbol{\Theta}_{\delta}{}^{(g)}$.

When the expectations of (2.6) are taken, the relationship between observed means and item means is expressed as

$$\boldsymbol{\mu}_{x}{}^{(g)} = \boldsymbol{\tau}^{(g)} + \boldsymbol{\Lambda}_{\boldsymbol{x}}{}^{(g)}\boldsymbol{\kappa}_{\eta}{}^{(g)} \qquad (2.7)$$

where $\boldsymbol{\mu}_x^{(g)}$ is the $p \times 1$ vector of item means for group g

$\boldsymbol{\kappa}_\eta^{(g)}$ is the $m \times 1$ (KAPPA) vector of means of the latent variables $\boldsymbol{\eta}^{(g)}$ for group g.

Equation 2.8 implies that the observed scores on the p items may not be comparable across groups when the scale metrics λ_{ij} , and intercepts τ_i , are different across groups. Since $\boldsymbol{\tau}$ and $\boldsymbol{\kappa}_\eta$ cannot be estimated simultaneously, (Sörbom, 1982), it implies that there is no definite origin for the latent variables. In order to solve this problem of indeterminacy, it is necessary to impose restrictions on the parameters to lead to model identification. In order to identify the model, Steenkamp and Baumgartner (1998) based on Cudeck (1989) suggest that one factor loading for each latent variable is set equal to 1, and that it is done for the same variable in each group.

There are basically two ways to impose these restrictions. The first is to use a marker item for each latent variable, and set it equal to one in each group analysed. For this method, it is important to fix the same items across groups. The second method is to set the vector of latent means all equal to zero in one group, and to constrain one intercept per factor to be equal across groups. Both methods lead to exactly identified models, which makes estimation possible.

2.7.6 The general CFA model: Covariance structure (multiple group case)

In addition to the means structure given in equation (2.8), the covariance structure needs to be specified. Ignoring the intercept part of the model, the model implied variance covariance matrix $\boldsymbol{\Sigma}^{(g)}(\theta)$ of $\boldsymbol{y}$ in group g is given by

$$\boldsymbol{\Sigma}^{(g)}(\theta) = E(\boldsymbol{y}^{(g)}, \boldsymbol{y}^{(g)\prime}) \quad (2.8)$$

$$= \boldsymbol{\Lambda}_y^{(g)} \boldsymbol{\Phi}_\eta^{(g)} \boldsymbol{\Lambda}_y^{(g)'} + \boldsymbol{\Theta}_\delta^{(g)} \quad (2.9)$$

where $\boldsymbol{\Phi}_\eta^{(g)}$ is the $m \times m$ (PHI) matrix of variances and covariances of the latent variables $\boldsymbol{\eta}^{(g)}$ in group g; and

$\boldsymbol{\Theta}_\delta^{(g)}$ is the $p \times p$ (THETA-DELTA) variance-covariance matrix of the error terms $\boldsymbol{\delta}^{(g)}$, that are usually constrained to be a diagonal matrix.

Model fit is assessed by the discrepancy between the observed covariance matrices $\Sigma^{(g)}$ and the model implied covariance matrices $\Sigma^{(g)}(\theta)$, as well as the discrepancy between the observed vectors of means (of each group) and the model implied vectors of means (of each group). The most useful measures of fit are the global fit measures of which the RMSEA is one of the most important (Fabrigar *et al*, 1999; Hu and Bentler 1999). The Chi-Square difference test is useful to assess the differences between increasingly restrictive models (Byrne, 1999).

2.7.7 Different levels of invariance: The general multi-group CFA model

In cross-cultural studies, a popular method of addressing the issue of equivalence is the use of multi-group confirmatory factor analysis using structural equation modelling. The advantage of this approach is that a wide variety of hypotheses relating to cross-cultural differences and similarities can be tested. The approach is useful to test for construct, scalar and metric equivalence by testing the tenability of a series of increasingly restrictive models, using goodness-of-fit tests.

The SEM-approach is more easily understood when it is expressed as a series of five questions, which represent a series of ever increasing restrictions on the CFA-model. This approach can be seen as a hierarchy of five increasingly restrictive hypotheses, namely:

Hypothesis 1: Configural Invariance;

Hypothesis 2a: Metric Invariance;

Hypothesis 2b: Intercept Invariance;

Hypothesis 3: Latent variance covariance invariance;

Hypothesis 4: Error variance invariance; and

Hypothesis 5: Latent mean invariance.

Hypothesis 1: Configural Invariance

The most important and pertinent question is: “Do the different groups have a similar factorial structure?”. This question can be addressed by the first hypothesis, which tests for *configural invariance* across groups, and is expressed as a null hypothesis as set out below.

H1: The factorial structure is similar across groups (which is tested by constraining salient factors in $\Lambda_y^{(g)}$ equal to zero across groups).

The approach for this hypothesis as described by Baumgartner and Steenkamp (1998) is based on Thurstone’s principle of simple structure (Horn, McArdle and Mason, 1983) as well as Horn and McArdle (1992). The principle is based on zero or near zero loadings on non-salient factors and significant non-zero loadings on salient factors which define the structure of the model. In CFA, it is customary to constrain non-salient factor loadings to be equal to zero. Configural invariance is supported if (1) the models with zero loadings on the non-salient factors fit the data well in all groups; (2) when all target loadings are significant and substantially different from zero; and (3) when the correlations between the latent variables are significantly below 1. The third requirement is necessary for discriminant validity.

If the hypothesis is tenable, it implies that the validity of a similar factor structure across groups is acceptable. When the first hypothesis is rejected, it means that the instrument is not valid across groups, and the best option for the researcher is to analyse the results separately for each group. Comparison across groups would be invalid. Only when the first hypothesis is tenable, would it make sense to move on to the second question and hypothesis of metric invariance (Baumgartner and Steenkamp, 1998; Cheung and Rensvold, 2000).

Hypothesis 2a: Metric Invariance

The second question asks: “Do the items in a measurement instrument operate equivalently across different subgroups?” Expressed as the second hypothesis, it corresponds to testing whether the values of the factor loadings are invariant (equal) across groups. To test this hypothesis, the factor loadings in the CFA-model are constrained to be equal for all the groups, and can be expressed as given in H2a. When H2a is tested in this form, both H1 and H2a are actually tested simultaneously.

H2a: $\Lambda_y^{(1)} = \Lambda_y^{(2)} = \ldots = \Lambda_y^{(G)}$ (factor loadings or LAMBDA’s equal across groups);

and

H1: The factorial structure is similar across groups (salient factor loadings equal to zero).

If the fit for this constrained model is poor, the hypothesis is rejected. This can be interpreted as meaning that the measurement instrument operates differently across groups, and the scales are measuring differently. Comparisons between groups would be invalid.

In the H2a model, configural invariance is forced by means of imposed constraints. Metric invariance is supported if (1) the model with equal loadings fits the data well in all groups; (2) when the target loadings are significant and substantially different from zero; and (3) when the correlations between the latent variables are significantly below 1, as was the case in Hypothesis 1.

Metric invariance (Baumgartner and Steenkamp, 1998) is a stricter form of invariance. It tests whether the factor loadings and intercepts are equal for corresponding items across groups. If this model holds, or fits the data reasonably well overall and within each group, it can be assumed that the scale intervals are equal across groups, or that the concept of equal metrics is tenable.

If the model fits well with some LAMBDA's allowed to be unique across groups, the instrument can be seen as partially metric invariant, (Byrne, Shavelson and Muthén, 1989). If these efforts do not lead to a satisfactory fit, the last resort for the researcher would be to analyse the subgroups separately. Comparisons across groups would not be valid. It would also be prudent to investigate the reasons that might have contributed to the non-equivalence, by investigating other issues that might have introduced bias in the study (see Van de Vijver and Leung, 1997; Cheung and Rensvold, 2000). An interesting perspective is provided by Cheung and Rensvold (2000) who distinguish between uniform and non-uniform extreme response style (ERS) differences. If all items show ERS, it is labelled as uniform ERS, and if only a few items display ERS, it is labelled as non-uniform ERS. If H2a is tenable, non-uniform ERS can be "ruled out", but uniform ERS may still be present. In order to "rule out" uniform ERS, it is necessary to proceed to H3 (Cheung and Rensvold, 2000).

Cheung and Rensvold (2000) suggested that in order to test for the presence of acquiescence response style (ARS), H2a is further extended to also include H2b by setting

H2b: $\tau^{(1)} = \tau^{(2)} = ... = \tau^{(G)}$ (item intercepts equal across groups);

and

H2a: $\Lambda_y^{(1)} = \Lambda_y^{(2)} = ... = \Lambda_y^{(G)}$ (factor loadings or LAMBDA's equal across groups);

and

H1: The factorial structure is similar across groups (salient factor loadings equal to zero).

If H2b, H2a and H1 are tenable hypotheses, it will be valid to compare latent means across all constructs (Cheung and Rensvold, 2000). If H2b results in a significantly poorer fit when compared to H2a, the option of partial invariance is plausible, (Byrne *et al.* 1989)and only constructs or latent variables for which non-invariance has been confirmed, can be further used to test for equal latent means.

If H2b is tenable, ARS is ruled out, if the intercept hypothesis (H2b) is rejected it implies that ARS is suggested, but not necessarily indicated (Cheung and Rensvold, 2000).

Hypothesis 3: Factor Covariance and Variance Invariance (the structural invariance hypothesis)

Once the previous suppositions have been confirmed, the researcher may be interested in asking: “Is the factorial structure of the instrument or theoretical construct equivalent across populations?” Differences in the covariances of the latent variables often reveal aspects that are not evident in level orientated analyses. Invariance of the structural model is of primary interest to many researchers, because, if both H1 and H2 hold, H3 will test whether the covariances and variances between the latent variables are equal.

The third hypothesis, therefore, tests whether factor variances and covariances between the latent variables are equal, in addition to the hypothesis that the factor loadings and factor structure are equal. This level of invariance can be tested by imposing more additional constraints upon those tested in H1, H2a and H2b, by also setting the covariances and variances between the latent variables equal across the cultural groups in the model. The additional constraints are given in H3, and the set of hypotheses that are tested in this restricted model are set out below.

H3: $\boldsymbol{\Phi}_{\eta}^{(1)} = \boldsymbol{\Phi}_{\eta}^{(2)} = \ldots = \boldsymbol{\Phi}_{\eta}^{(G)}$ (factor variances and covariances are equal across groups or PHI-matrices are equal across groups).

H2b: $\tau^{(1)} = \tau^{(2)} = \ldots = \boldsymbol{\tau}^{(G)}$ (item intercepts equal across groups).

H2a: $\boldsymbol{\Lambda}_{y}^{(1)} = \boldsymbol{\Lambda}_{y}^{(2)} = \ldots = \boldsymbol{\Lambda}_{y}^{(G)}$ (factor loadings equal across groups).

H1: The factorial structure is similar across groups (salient factor loadings equal to zero).

Since the $\boldsymbol{\Phi}_{\eta}$ matrices are symmetrical, only the diagonal elements (variances) and the either of the non-diagonal elements (covariances) are specified in the model.

The third hypothesis implicitly also tests the first and both the second hypotheses. If the fit for this very constrained model is poor for all groups, hypothesis 3 is rejected. This implies that factor structure is similar, the metrics are the same, but at least one of the covariances or variances of the latent variances are not equivalent.

In H3 configural and metric invariance is enforced by means of imposed constraints. Structural invariance is supported if (1) the model with equal variances and covariances fits the data well in all groups; (2a) when the target loadings are significant and substantially different from zero; (2b) when the item intercepts are equal and (3) when the correlations between the latent variables are significantly below 1.

It is fairly rare for this third hypothesis to be tenable, since it is very restricted. If H3 were to be rejected, one could still proceed to test H5, which tests differences between the latent means. Alternatively it would also be appropriate to consider modification indices of the PHI-matrix, to identify the specific pair of latent variables for which the covariance or variance is different across groups. When these parameters are subsequently set free, and a good fit is obtained, it can be inferred that the instrument is partially structural invariant (Byrne *et al*, 1989). If H3 with all its restrictions is not rejected, H4 can be tested.

Hypothesis 4: Error Variance Invariance
(also known as complete invariance)

The final invariance test that can be implied on the structural model is that the amount of measurement error is invariant across groups. It is achieved by implying another additional set of restrictions:

H4: $\boldsymbol{\Theta}_{\delta}^{(1)} = \boldsymbol{\Theta}_{\delta}^{(2)} = \ldots = \boldsymbol{\Theta}_{\delta}^{(G)}$.

In this hypothesis, indicator error variances are set equal across groups which implies that the measurement errors of items are constrained to be equal for each variable across groups; and

H3: $\boldsymbol{\Phi}_{\eta}^{(1)} = \boldsymbol{\Phi}_{\eta}^{(2)} = ... = \boldsymbol{\Phi}_{\eta}^{(G)}$ (factor variances and covariances are equal across groups or PHI-matrices are equal across groups); and

H2b: $\tau^{(1)} = \tau^{(2)} = ... = \boldsymbol{\tau}^{(G)}$ (item intercepts equal across groups).

H2a: $\boldsymbol{\Lambda}_{y}^{(1)} = \boldsymbol{\Lambda}_{y}^{(2)} = ... = \boldsymbol{\Lambda}_{y}^{(G)}$ (factor loadings equal across groups); and

H1: The factorial structure is similar across groups (salient factor loadings equal to zero).

The fourth hypothesis implicitly also tests the preceding hypotheses. If the fit for this totally constrained model is poor, H4 is rejected.

In this model, configural, metric, scalar and error invariance is enforced by means of imposed constraints. Complete invariance is supported if the model with equal error variances fits the data well in all groups. When all four these hypotheses are tenable, it can be concluded that the measurement instrument is completely structurally invariant (and by implication completely equivalent) across cultural groups. According to Steenkamp and Baumgartner (1998), if the items are all configurally (H1), metrically (H2), and structurally (H3) invariant, as well as when the error variance invariance hypotheses (H4) are not rejected, it can be inferred that the items are equally reliable across groups. This is a very restrictive hypothesis, and it is extremely rare for H4 to be tenable. If Hypothesis 4 is rejected, one can proceed to test Hypothesis 5.

In summary, H1 to H4 are all concerned about the equivalence in the covariance matrix. Using equation (29), the covariance models tested within each of these set of hierarchical hypotheses are set out as shown below.

H1: $\Sigma^{(g)}(\theta) = \Lambda_y{}^{(g)}\, \Phi_\eta{}^{(g)}\, \Lambda_y{}^{(g)'} + \Theta_\delta{}^{(g)}$ } H1

H2a: $\Sigma^{(g)}(\theta) = \Lambda_y\, \Phi_\eta{}^{(g)}\, \Lambda_y{}' + \Theta_\delta{}^{(g)}$ } H2a

H2b: $\Sigma^{(g)}(\theta) = \tau^{(1)} = \tau^{(2)} = \ldots = \tau^{(g)}$ } H2b

} H3

H3: $\Sigma^{(g)}(\theta) = \Lambda_y\, \Phi_\eta\, \Lambda' + \Theta_\delta{}^{(g)}$

} H4

H4: $\Sigma^{(g)}(\theta) = \Lambda_y\, \Phi_\eta\, \Lambda_y{}' + \Theta_\delta$

It should be noted that H4 is not the same as analysing a single group. The covariance matrices for each group are still given separately, but a model with all parameters equal is fitted simultaneously to the group covariance matrices.

Hypothesis 5: Latent mean invariance (equal means of latent variables)

H1 to H4 all relate to covariance structures across groups, but researchers may also be interested in asking: “Are the latent means of the constructs in a model different across groups?” If the latent means are not equal, it implies that some systematic sources of error may be present across groups, and mere interpretation of mean differences across groups needs to be very carefully considered, and may lead to incorrect results (Byrne, 1998). It may be necessary to use methods of identifying the nature of bias (Van de Vijver and Leung, 1997) and to adjust for the differences in the latent means before level-orientated analyses are used (also see Steenkamp and Baumgartner, 1989; Cheung and Rensvold, 2000). Methods that can be considered to adjust the data for these systematic patterns are standardisation, or a data transformation method that “removes” the bias before further level-orientated methods are used.

Scalar invariance (H5) refers to the intercept terms of the latent means, $\kappa^{(g)}$. Configural (H1) and metric (H2) invariance are based on the covariance structure in the observed variables across groups, where scalar (H5) invariance refers to equality of means across groups. Even if a scale is based on the same metrics (equal lambdas and intercepts), the means could be systematically different across groups, due to some source of systematic bias.

Therefore, H5 can be stated as

H5: $\boldsymbol{\kappa}_{\eta}^{(1)} = \boldsymbol{\kappa}_{\eta}^{(2)} = ... = \boldsymbol{\kappa}_{\eta}^{(G)}$

as well as including the relevant constraints as specified in the preceding hypotheses (H1 and H2); (H1, H2 and H3); or very rarely (H1, H2, H3 and H4). Before H5 can be tested, at least H1 and H2a and H2b should be tenable, otherwise it will be invalid to test H5, since the metrics of the scales cannot be assumed to be similar across groups. If the model as restricted under H5 does not fit the data well, the researcher could also consider freeing some parameters as indicated by modification indices, to identify the specific latent means that are unique across groups. A summary of the five hierarchical hypotheses is presented in Table 2.4.

Table 2.4: Summary of five hierarchical multi-group CFA hypotheses

Hypothesis 1: Configural Invariance (same factor structure) across groups
Salient factor loadings in Λ_x equal to zero
Hypothesis 2a: Metric Invariance (equal factor loadings) across groups
$(\boldsymbol{\Lambda}_x^{(1)} = \boldsymbol{\Lambda}_x^{(2)} = ... = \boldsymbol{\Lambda}_x^{(G)})$
Hypothesis 2b: Metric Invariance (equal item intercepts) across groups
$(\tau^{(1)} = \boldsymbol{\tau}^{(2)} = ... = \boldsymbol{\tau}^{(G)})$
Hypothesis 3: Factor variance and covariance invariance across groups
$(\boldsymbol{\Phi}_{\eta}^{(1)} = \boldsymbol{\Phi}_{\eta}^{(2)} = ... = \boldsymbol{\Phi}_{\eta}^{(G)})$
Hypothesis 4: Error variance invariance across groups
$(\boldsymbol{\Theta}_{\delta}^{(1)} = \boldsymbol{\Theta}_{\delta}^{(2)} = ... = \boldsymbol{\Theta}_{\delta}^{(G)})$
Hypothesis 5: Scalar or latent mean invariance (equal latent means) across groups
$\boldsymbol{\kappa}_{\eta}^{(1)} = \boldsymbol{\kappa}_{\eta}^{(2)} = ... = \boldsymbol{\kappa}_{\eta}^{(G)}$

2.7.8 Summary of the MG CFA approach

Within the statistical and psychometric literature, the methodological aspects of the multiple group structural equation modelling approach to equivalence/invariance dates back to McGaw and Jöreskog (1971); Sörbom (1974); Marsh and Hovecar

(1985); Bollen (1989); Byrne (1989); Poortinga (1989); Little (1997); Rensvold and Cheung (1998) and Van de Vijver and Leung (2000), to name a few.

Important methodological papers to be noted outside the marketing discipline, for example in organisational research, are the work of Vandenberg and Lance (2000) and Vandenberg (2002), which are important and widely cited.

In the marketing discipline, the notable scholars in bias and invariance are Steenkamp and co-workers, with papers by Steenkamp and Van Trijp (1991); Steenkamp and Baumgartner (1998); Baumgartner and Steenkamp (2001) and Baumgartner and Steenkamp (2006), suggesting innovative approaches to deal with bias problems in marketing research.

Other methodological papers about the equivalence/invariance problem with applications in marketing include Sin, Cheung and Lee (1999); Myers, Calantone, Page Jr. and Taylor (2000); Mackenzie (2001); Smith and Reynolds (2001); Van Herk, Poortinga and Verhallen (2005); and the list is certainly not exhaustive.

Outside the discipline of marketing, application and methodological papers concerning the equivalence/invariance approach using the multi-group CFA method have appeared in almost every other social sciences discipline. In the health sciences, Bullinger, Anderson, Cella and Aaronson, 1993 argue for the use of the method. In the strategic management literature, Bensaou, Coyne and Venkatraman (1999); in psychology, Byrne and Shavelson (1987); Van Hemert, Van de Vijver, Poortinga and Georgas (2002); and in organisational research, Cole, Bedeian and Feild (2006) apply the method and make important theoretical contributions in their respective fields.

With the growth in the popularity of the multi-group confirmatory factor analysis (CFA) invariance testing approach which has become widely accepted, the importance of the equivalence/invariance debate has recently been raised to a new level. Recently, additional attention has been paid to the item response theory (IRT) approach to equivalence/invariance as an alternative to the multi-group confirmatory factor analysis approach. Comparisons of the methods in the methodological literature can be found in Reise, Widaman and Pugh (1993); Fan (1998); Embretson and Reise (2000) and Raju, Laffitte and Byrne (2002). More

specifically, in the marketing literature, the applications of IRT in comparison with CFA with methodological issues and empirical examples can be found in Salzberger, Sinkovics and Schlegelmilch (2000); Meade and Lautenschlager (2004); as well as Ewing, Salzberger and Sinkovics (2005).

It is clear that the MG CFA approach has become the state-of-the art method to deal with measurement bias. The IRT method has only recently received attention, and some aspects of IRT are covered in more detail in the section that follows.

2.8 ITEM RESPONSE THEORY

Item Response Theory (IRT) methods (Embretson and Reise, 2000) have been used mainly in large-scale educational assessments and have become more popular in applications of psychometric testing in recent years. Several different IRT models exist, and the relevant model is dependent on the situation and the assumptions that can be made within a specific application (Fan, 1998).

Most IRT applications are used in psychometric test assessments, where the response is dichotomous. The item could be answered correctly or incorrectly by the respondent. Item Characteristic Curves (ICC) are estimated with the purpose to quantify the probability of responding to an item in a specified manner for given levels of the latent trait, and IRT can be used to identify items that do not operate equivalently across groups. More recently, IRT models were developed to be used in applications where the variables of interest could be nominal or ordinal, as opposed to being dichotomous (Hui and Triandis, 1985).

2.8.1 The graded response model

Within the scope of this study, where the interest is Likert-type questions, the graded response model described by Samejima (1969) is applicable (Reise, Widaman and Pugh, 1993). The probability of responding above category k, given the person has a level equal to θ on the trait is modelled.

Using the notation and guidelines provided by Raju, Laffitte and Byrne (2002), the discussion in this section will be focused on the graded response model, which is the only model that can be reasonably applied in the context of ordinal variables such as the Likert scale. An IRT model specifies a non-linear relationship between the underlying construct and the item or subscale, where in SEM, the relationship is assumed to be linear. Within the context of cross-cultural research, where opinions and/or attitudes are mainly of interest and ordinal scales are often used, the relevant model is the graded-response model of Samejima (1969). The principle in this model is to arrange the response categories in an order that represents increased intensity of an attitude, and the response categories can be viewed as a direct indicator of the position of the respondent on the continuum of the underlying latent construct (Raju *et al*, 2002).

Using similar notation as Embretson and Reise (2000), the graded response model is discussed hereunder. If an item has K_i categories (with $m_i + 1 = K_i$), the probability of responding above the k-th category in the graded response model for a respondent with a given trait level θ on item i can be expressed as:

$$P_{ik}^{*}(\text{choosing above a response category} > k|\theta) = P_{ik}^{*}(\theta) = \frac{1}{1+e^{-a_i(\theta - b_{ik})}}$$

Therefore,

$$P_{i1}^{*} = P_{i1}^{*}(x > 1|\,\theta) = \frac{1}{1+e^{-a_i(\theta - b_{i1})}} \quad \text{and}$$

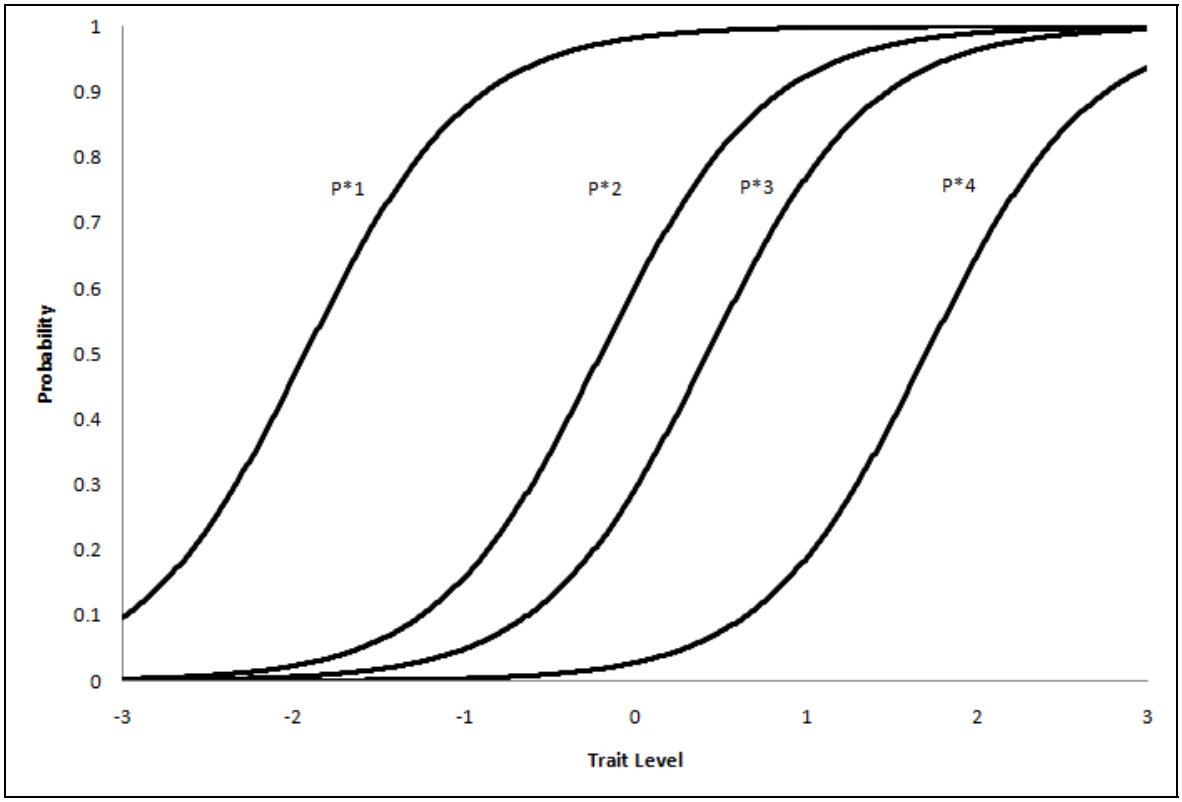
$$P_{i2}^{*} = \frac{1}{1+e^{-a_i(\theta - b_{i2})}}$$

$$\vdots$$

$$P_{i(m_i-1)}^{*} = \frac{1}{1+e^{-a_i(\theta - b_{i(m_i-1)})}}$$

From this formula, an example of the operating characteristic curves for a five-category example for the graded response model (with five parameters) is shown in Figure 2.4.

Figure 2.4: Operating characteristic curves for a five-category item using the graded response model



Following from this, the probability of choosing a specific category for a given trait θ, can be expressed as

$$P_{i1} = P(x = k \mid \theta) = \frac{1}{1+\exp[-a(\theta - b_{j-1})]} - \frac{1}{1+\exp[-a(\theta - b_j)]}$$

Therefore, if $m_i = 5$, then

$$P_{i1} = 1 - P_{i1}^{*}\ ; \qquad P_{i2} = P_{i1}^{*} - P_{i2}^{*}; \qquad P_{i3} = P_{i2}^{*} - P_{i3}^{*}; \qquad P_{i4} = P_{i3}^{*} - P_{i4}^{*} \quad \text{and} \quad P_{i5} = P_{i4}^{*}$$

In general, therefore, the probability of selecting a specific category is

$$P_{i1} = 1 - P_{i1}^{*}$$

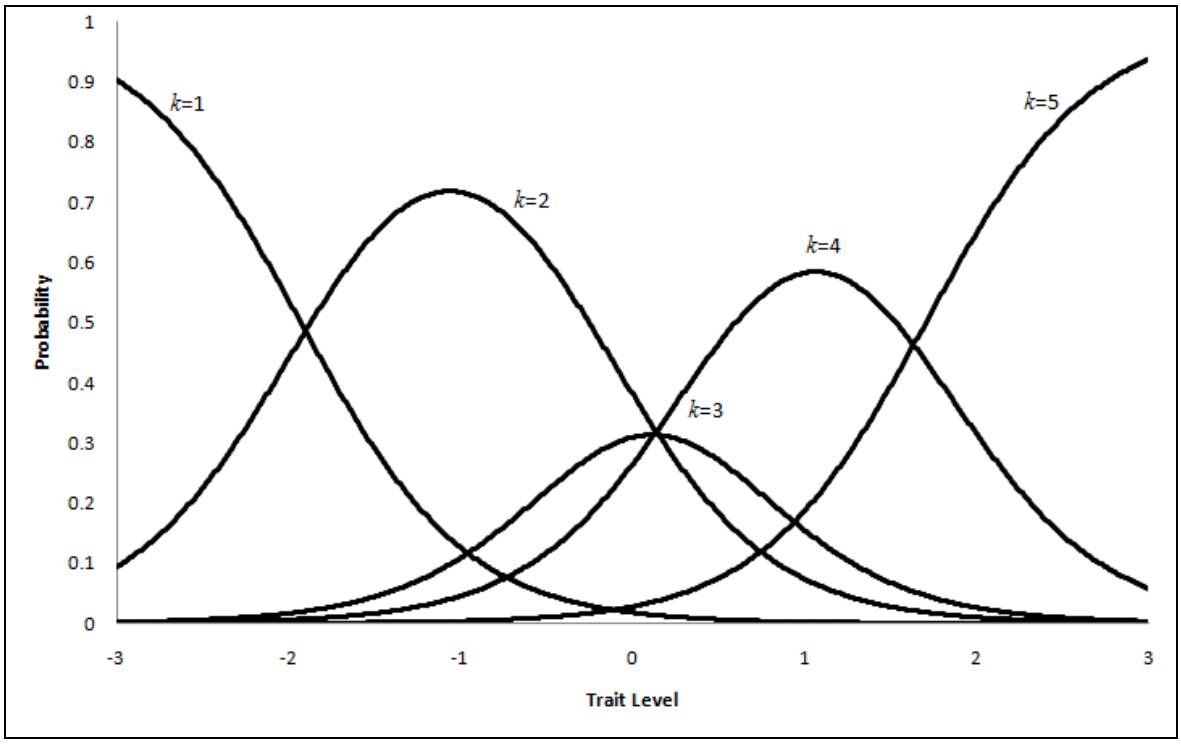
$$P_{i(m_i-1)} = P_{i(m_i-2)}^{*} - P_{i(m_i-1)}^{*}$$

$$P_{i(m_i)} = P_{i(m_i-1)}^{*} - 0$$

where P_{ik} represents the probability of getting a score k, for a given θ. Note that the sum of $(P_{i1} \cdots P_{i(m_i)})$ is always 1. The item characteristic curves or boundary response functions (as shown in Figure 2.4) and category response functions (as shown in Figure 2.5) for a specific item depends on the item parameters a, b and on θ, the level of the trait.

The category response curves as given in Figure 2.5 shows the probability of selecting a specific category, for persons at different trait levels.

Figure 2.5: Category response curves of the graded response model

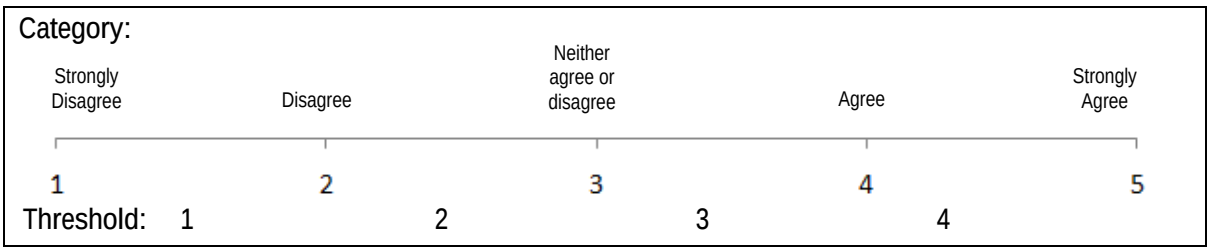


The expected raw score of the true score on an item i for a respondent (with a given θ) can be expressed as

$$t_i(\theta) = 1P_{i1} + 2P_{i2} + (m_i - 1)P^*_{i(m_i-1)} + (m_i)P^*_{i(m_i)}$$

Note that in the graded response model, each item has as many parameters as it has categories, so if an item has five response categories, it will have one item slope parameter a, which is associated with item difficulty (or how difficult is for respondents to give a higher score than the one chosen), and four threshold location parameters, b_{i1}, b_{i2}, ..., b_{i4}. For an explanation of the item threshold, see Figure 2.6.

Figure 2.6: Item categories and item thresholds



Measurement equivalence within the IRT framework is established when the item parameters, a_i, b_{i1}, b_{i2}, ..., b_{im_i} can be assumed to be equal across groups, which implies that there is no differential item functioning (DIF) across groups.

2.8.2 A comparison of IRT and SEM approaches in multi-group analysis

In the context of cross-cultural research, IRT has several limitations, the most limiting is the assumption of uni-dimensionality, which means that analyses could only be performed on subscale level, or for one latent variable at a time. If this assumption is violated, the item parameters may not be accurately estimated. Another limitation is that a large number of items and large samples are required for accurate estimation of item parameters (Hui and Triandis, 1985).

The fundamental difference between CFA models and IRT models is that in CFA models, the model is based on covariances between observed measures, whereas in IRT models, the model is based on the actual responses to the item by respondents. An IRT-model uses the item response function, a nonlinear monotonic increasing function (as in Figure 2.4) to describe the relationship between the level of the respondent on a latent trait, denoted by θ, and the probability of a particular item response. Key assumptions in IRT models include unidimensionality, which means that the items all relate to one underlying latent trait, and that the items are locally independent, which means that for a given level of the trait, there are no correlations between the items (Reise, Widaman and Pugh, 1993). Table 2.5 provides a summary of scholarly attention to the CFA/IRT debate.

Table 2.5: Comparison of attention in scholarly literature to CFA/IRT

Focus	**Multi-group CFA**	**Multi-group IRT**
Methodological papers in marketing	Steenkamp and Baumgartner (1998); Mackenzie (2001)	Procedures not well described in marketing research methods literature
Methodological papers in other disciplines	Well described in other management disciplines (e.g. Vandenberg and Lance, 2000, Vandenberg, 2002)	Hui and Triandis (1985); Reise *et al.* (1993); Fan (1998) and Raju *et al.* (2002).
Applications in marketing	Steenkamp and Baumgartner (1998); Sin *et al.* (1999); Myers *et al.* (2000); Mackenzie (2001); Smith and Reynolds (2001)	Salzberger, Sinkovics and Schlegelmilch (2000) and Ewing, Salzberger, and Sinkovics, (2005) have argued that it could be useful
Applications in other disciplines	Bullinger *et al.* (1993); Bensaou *et al.* (1999); Byrne and Shavelson (1987); Van Hemert *et al.* (2002); Cole *et al.* (2006)	King, Murray, Salomon and Tandon (2004) used IRT to identify DIF in an application in political sciences review

"With regard to specifying a model for a set of data, procedures for CFA seem rather more advanced, simpler and more user-friendly than those developed for IRT" (Reise *et al*, 1993:563). When IRT are used in a multiple group setup, the only software packages that allows the user to fit the graded response model to multiple groups is MULTILOG. The practical application requires the user to

specify the model as if the groups to be compared are subgroups in a larger group. This requires the user to specify an input matrix with (# of groups × # of items) columns, and N rows, where N is the size of all the groups added together. A missing code is then used when the values of the next group are entered. This method is very cumbersome and inexperienced users can have difficulty with it.

A further problem is that the chi-square statistic for model fit is not a simple measure of how well the model fits the data, but a function of the observed and expected response proportions. The degrees of freedom is a function of the number of possible response patterns, minus the number of parameters estimated. “Basically, the MULTILOG program presumes that the program user has a fairly sophisticated understanding of IRT theory and methods. If not, the user may err seriously in the specification of the model and the subsequent evaluation of model fit.” (Reise *et al*, 1993:564).

Compared to CFA, IRT is a limiting approach. IRT methods do not take measurement error into account in the model, and only items related to a uni-dimensional construct can be used in the analysis. IRT methods exclude the modelling of covariances, which is often of primary interest to researchers.

The SEM framework is derived from the principles of Classical Test Theory (CTT), and has a longer and stronger history than IRT. A major advantage of CTT is that it does not pose very strong theoretical assumptions, which makes it easy to apply in different situations. IRT is has a stronger theoretical grounding from Latent Trait Theory (LTT) and models the probability of success at item level. The focus of CTT is on item-level information in contrast to LTT which focuses on test-level information (Fan, 1998). A disadvantage of CTT-models is what Fan (1998) mentions as its circular dependency, which means that the item parameter estimates are both item and person/sample dependent, whereas, at least in theory, LTT models produce item statistics that are independent of examinee samples. The major differences between multi-group CFA and multi-group IRT are summarised in Table 2.6, with similarities given in Table 2.7.

Table 2.6: Differences between multi-group CFA multi-group IRT in measurement equivalence applications

Classical Test Theory and CFA in an SEM framework	**Latent Trait Theory and IRT**
The modelled relationship between the latent variable and the items are linear in nature Raju *et al.,* (2002).	The modelled relationship between the items and the latent traits are non-linear Raju *et al.,* (2002).
The methodology is thoroughly developed and software and procedures are more available to handle multiple constructs and multiple groups simultaneously in an analysis Raju *et al.,* (2002).	The differential item functioning (DIF) methodology can only be applied on uni-dimensional constructs Raju *et al.,* (2002).
CFA allows the user to test whether item error variances are equal across groups although it is generally regarded as an overly strict requirement in most practical applications Raju *et al.,* (2002).	In the IRT framework, testing for item level error variances has not been fully explored by statisticians, and it is only possible to test for group-based aggregate standard errors Raju *et al.,* (2002).
CFA are more geared towards studying smaller sets of continuous observed variables assumed to be normally distributed (Embretson and Reise, 2000).	IRT methods are more useful to detect item bias when a large pool (more than 40) dichotomous and polytomous items that are uni-dimensional (relate to a single latent variable) (Embretson and Reise, 2000).
Hypothesis by hypothesis invariance testing in CFA has been well developed and is widely applied by researchers (Reise *et al.*, 1993; Embretson and Reise, 2000).	In the IRT-framework, investigation of item bias has developed into a branch called differential item functioning (DIF) (Embretson and Reise, 2000).
Can be used to test for two major response bias patterns, namely extreme response style and acquiescence response style (Cheung and Rensvold, 2000).	A particular problem that arises is that DIF is often indicated when the items relating to the latent construct are marginally not uni-dimensional. Very few constructs are so narrowly defined, and DIF may be indicated falsely, since DIF could be a function of the combination of items, and not a result of item bias. (Embretson and Reise, 2000).

Table 2.7: Similarities between multi-group CFA multi-group IRT in measurement equivalence applications

According to Embretson and Reise (2000), multi-group IRT modelling and multi-group CFA have several similarities, despite the major differences in the underlying theory.	
It is required that the researcher use a strategy to specify certain parameters as fixed in order to identify the model, and the usual way is to either set one of the path coefficients equal to one for each latent variable, or set the error variances equal to one (Reise *et al.*, 1993).	As in CFA, the researcher needs to specify a distribution of the latent trait variable θ (often specified as standard normal), in order for the model to be identified. In the multiple group setting, the researcher has to specify an anchor item in which the parameters are constrained to be equal across groups, in order to allow the researcher to compare the values of the other parameters across groups (Reise *et al.*, 1993).
Items paths need to be set equal to 1 in order to assign a scale to the latent variable and in order for the CFA model to be identified (Embretson and Reise, 2000).	Items that have no DIF (no bias) need to be identified to provide the "anchor" items that are assumed to be equal across all groups, thereafter the item parameters across all groups are freely estimated.
Successive increasingly restrictive hypotheses can be tested, by restricting parameters to be equal across groups (Steenkamp and Baumgartner, 1998).	Items are held constant one by one, and if the restriction is found tenable, those items can be viewed as having no DIF, which could be a very tedious process (Embretson and Reise, 2000).
It is possible to identify partial measurement invariance (Byrne, *et al.,* 1989)	It is possible to identify partial measurement invariance using differential item functioning (Reise *et al.*, 1993; Embretson and Reise, 2000; Raju *et al.*, 2002).
Statistical fit indices are available for CFA measures, namely the chi-square test. The procedure allows for chi-square difference tests when the models are nested by restraining parameters to be equal across groups, thereby allowing the testing of increasingly restrictive hypotheses (Reise *et al.*, 1993).	As in CFA, it is possible to use the chi-square difference tests for nested models to assess the plausibility of increasingly restrictive models (Reise *et al.*, 1993).
Practical fit indices are available over and above the chi-square values, which allow the researcher to assess model fit, supplementing the chi-square measures of fit that may indicate poor fit in the case of large samples, even if the model deviates slightly from the covariance matrix.	Practical fit indices have not yet been developed, and according to Reise *et al.* (1993), it is a matter that should be the focus for future research.
Modification indices are calculated that helps the researcher to identify the source of poor fit in models.	The development of modification indices in IRT is also identified as a potential area for future research (Reise *et al.*, 1993).

Reise *et al* (1993) mentioned that the most important area for future research is to do a more in depth comparison of the differential utility of CFA and IRT models, in order to provide a coherent framework to help the researcher decide which

framework is most appropriate in a specific circumstance. Comprehensive research of this nature has not yet become available.

2.8.3 Application limitation of IRT method within the scope of this research

Both CFA-methods in the SEM-framework and DIF-methods in the IRT-framework can be used to test for item bias. The methodology in the SEM-framework is strongly established and it is often applied in the marketing and other disciplines (see Table 2.5). The IRT-framework is cumbersome, and only uni-dimensional constructs or subscales can be handled. The procedures dealing with polytomous items, such as Likert scales, are clumsy and computer programmes are not easily obtainable.

IRT software programmes only deal with categorical data. Once the data has been transformed to a continuous scale, as is necessary in this research, the currently available software will not read continuous data (Du Toit, 2003), and therefore the merits of SIMITNORM cannot be assessed using currently available IRT software.

IRT software packages are not available in generally available statistical software such as SPSS, SAS, Statistica, NCSS or STATA, but SPSS has an associated package called AMOS, on which multi-group CFA can be done. Although SAS and STATA have procedures included that can run SEM-models, these packages have not yet introduced multi-group modelling. LISREL, one of the best known packages for SEM, is distributed by Scientific Software International in Chicago, and they also distribute IRT software, and MULTILOG is one of their products, (Scientific Software International, n.d.).

2.9 ADVERTISING RESPONSE SCALES

The proposed methodology for this research is covered in Chapter 4. The rationale for using a specific advertising response scale, namely the Schlinger (1979) Viewer Response Profile (VRP) has been set out in considerable detail in that chapter. This section of the Literature Review covers the various advertising

response scales that have been reported in marketing and advertising journals. Most of these scales appear in Bearden, Netemeyer and Mobley (1993).

2.9.1 The Schlinger (1979) Viewer Response Profile

The research by Schlinger (1979) can be divided into ten different phases, most of them following on one another. These phases spanned more than five years of research, and involved over 5 000 respondents, and roughly 1 000 commercials and story boards. The data was collected in practice using actual consumers as respondents.

Phase 1

This qualitative research phase involved 400 viewers and 14 commercials and story boards, where the open-ended verbatim feedback from customers was recorded and resulted in 600 different statements. When duplications were eliminated, 150 statements remained. These statements were compared to theories about advertising, and after adding and deleting, 139 items remained.

Phases 2 to 7 involved several phases of data collection and refinements of the item set, using five factor analyses (here referred to as EFA1 to EFA5) and three using analyses of variance (here referred to as Anova1 to Anova3). The aim of the factor analysis was to achieve data reduction, and Anova was performed to retain items that discriminated between commercials.

Phase 2

The 139 items that were identified in Phase 1 were used in a questionnaire using 25 commercials and 20 different respondents to each commercial, resulting in a sample of 500 responses. These responses were subjected to exploratory factor analysis (EFA1) in which the main dimensions of the VRP were identified. The 139 items were also used in an analysis of variance to test which items discriminated between the commercials (Anova1).

Phase 3

Phase 3 repeated the exercise of Phase 2, using 139 items, 50 subjects and 10 of the 25 commercials used in Phase 2, resulting in a sample of 500 respondents. Once again exploratory factor analysis was performed (EFA2), and analysis of variance (Anova2) on each item over the 10 commercials. In this phase, the 139 items were reduced to 70 items, by retaining items with factor loadings higher than 0.5 and items that discriminated between the commercials. At this stage the differences were interpreted "to see if they make sense" (Schlinger 1979:38), which can be interpreted as an attempt to make a judgement about the face validity of the scale.

Phase 4

A data set consisting of the responses of 1 504 men and women on 42 untested commercials was obtained, with between 32 and 50 responses per commercial. Exploratory factor analysis (EFA3) was applied and compared to the results in EFA1 and EFA2. Based on these results, all items that consistently showed low loadings were ignored, and, after the addition of a few other items thought to be important, 49 items were retained.

Phase 5

The 49 item VRP was copy-tested over four years studying 377 different story boards and commercials, each with a quota sample of 30 to 50 consumers. From this data bank, the 49 item VRP was factor analysed (EFA4) using 1871 respondents and 40 different commercials. The commercials represented a variety of styles and products, and included no more than two commercials for each product. (This approach can be interpreted as an attempt by Schlinger (1979) to remove the product-effect).

Phase 6

A fifth factor analysis (EFA5) was performed using the average responses to the 377 commercials described in Phase 5. This eliminated the differences between individuals within a commercial.

By reflecting on the five consecutive factor analyses (EFA1-EFA5), and with some flexibility, only those items that loaded above 0.5 were retained.

Phase 7

Based on the previous phases of the research, this phase also involved several consecutive analyses of variance (Anova3), using various types of commercials, and retaining items that discriminated between commercials within pools of similar brands, resulted in the 32-item Schlinger VRP. This item pool was described as presenting four stable factors, namely Entertainment, Confusion, Relevant News and Brand Reinforcement, and three less stable dimensions namely Empathy, Familiarity and Alienation. Although Schlinger (1979) did not state it explicitly, in this phase of the study, groupings of the commercials of those from a similar brand are given, can be interpreted as an attempt to remove the brand effect in the commercial.

Phase 8

In this phase, the test-retest reliability was estimated with smaller samples using three commercials, three story boards and 30 respondents each. This resulted in 180 responses, using a 34-item VRP. These analyses provided sufficient basis for the test-retest reliability of the VRP, even though Schlinger (1979) did not explicitly mention reliability testing in her article.

Phase 9

This phase of Schlinger's (1979) study can be interpreted as a qualitative assessment of the scale. Five different cases are reported where the Schlinger VRP was applied successfully in practice as a diagnostic tool. In the first case, it indicated that a campaign idea failed, with evidence that pointed to the predictive validity of the VRP. The results were correlated with Burke's Recall measure. In the second case, the VRP was useful in selecting between two styles of execution. In the third case, the VRP helped the advertiser to select between two methods to present food in a commercial. In case 4, the VRP was used to correct elements that were irritating, offensive and distracting in a commercial. In the fifth case, twelve commercials were tested for clutter awareness using the McCollum/Spielman scores. The factor that accounted for the greatest amount of

variation in this score was the "Confusion" dimension, which is also an indication of the predictive validity of the VRP.

Phase 10

In this phase, Schlinger integrated the eight factor dimensions developed by Wells, Leavitt and McConville (1971) with the Schlinger dimensions. A combination of items from both scales, and a few additional items was used, resulting in 63 items. The combined scale was tested using 500 commercials and story boards, and between 30 and 50 respondents for each stimulus. A six-point Likert-type scale was used. The mean scores of each commercial was used (eliminating individual differences) and exploratory factor analysis (EFA6) was applied, examining between two and nine factor solutions. The seven factor solution was thought to make the best sense, and the factor structure agreed with the original seven factors of the VRP.

The VRP focuses on the emotional or affective side of response to a specific advertisement. The feelings of the viewers, rather than their knowledge are being gauged. The scale items are statements on which the respondents evaluate how they felt about various aspects of the advertisement.

An interesting observation by Schlinger (1979) was that product users tend to rate all the VRP items more favourably than non-users. Four of the seven dimensions, namely entertainment, confusion, empathy and familiarity are associated with the executional style of the commercial. Two of the seven dimensions, namely relevant news and brand reinforcement relate to attitudes towards the advertised product. Alienation is a dimension that relates to both style of execution and the advertised product. This suggests that higher order factors may be present in the VRP.

The seven constructs / subscales or dimensions of the Schlinger (1979) scale are as follows:

Entertainment	V1-V7
Confusion	V8-V11
Relevant News	V12-V16
Brand Reinforcement	V17-V18
Empathy	V19-V23
Familiarity	V24-V26
Alienation	V27-V32.

The items of the instrument appears in Appendix A.

Strengths of the VRP

The overall strength of the VRP is evidenced in the step-by-step process that was followed in the development of the scale that spanned more than five years. Although Schlinger reported no reliability measures such as Cronbach Alpha, some evidence of predictive validity (in phases 3 and 9) was reported. Phase 8 was an attempt to provide evidence of test-retest reliability. The approach to compare the open-ended verbatim responses with theories about advertising (in phase 1) was an attempt to obtain a theoretical basis for the VRP. The several iterations where items were discarded and again added in, also ensured that no important dimensions were left out, and that the VRP was comprehensive in measuring the constructs that are important to discriminate between commercials. The procedure suggested by Churchill (1979) for developing better marketing measures has been adhered to by Schlinger (1979) to a large degree.

Weaknesses of the VRP

The weaknesses in Schlinger's research was that she did not make use of reliability measures such as Cronbach Alpha, and item-total correlations (Nunnally and Bernstein, 1994) could have been useful in discarding unnecessary items. She also did not make use of confirmatory factor analysis in the subsequent phases of the study, which could have been applied after Phase 7, and which would have been more appropriate (Churchill, 1979).

A better method of analysis for the last data set reported in Phase 10, would have been to make use of the multitrait-multimethod (Campbell and Fiske, 1959) matrix in order to provide evidence for validity (Churchill, 1979).

Another drawback of the 32-item Schlinger scale is that the various dimensions are not equally well presented by items. The items per factor vary between two and seven, so that some of the factors have too many items and some have too little. (There were seven items for Entertainment, four for Confusion, five for Relevant News, two for Brand Reinforcement, five for Empathy, three for Familiarity, and six for Alienation).

Even though the Schlinger (1979) VRP has been found to be a very useful instrument from a practitioner's point of view (Du Plessis, 1994), it was severely criticised by Zinkhan and Burton (1989), with a result that not much reference to the Schlinger VRP can be found in the latest academic literature.

In the Zinkhan and Burton (1989) study, which was a follow-up on a previous study (Zinkhan and Fornell, 1985), only four of the seven dimensions of Schlinger were included in the study. The aim of the Zinkhan and Burton (1989) study was to assess the reliability and validity of three advertising response scales, namely the Well's (1964) Reaction Profile, the Leavitt (1970) Commercial Profile and the Schlinger (1979) VRP. Zinkhan and Burton (1989) found that the Schlinger (1979) VRP was not as reliable or as stable as the Leavitt (1970) Commercial Profile, and therefore recommended the use of the Leavitt (1970) scale rather than the Schlinger (1979) scale. The major weakness in the argument of Zinkhan and

Burton (1989), is the methodology that they followed. The sample consisted of repeated measures, and the assumption of independent observations was actually ignored in the statistical analyses. Further details of the two Zinkhan studies are given in Sections 2.9.4 and 2.9.6.

2.9.2 Schlinger (1982)

In this article, Mary Jane Schlinger (1982) used the VRP and related the affective responses to consumer demographics and situational variables, such as brand preferences and product usage. The purpose was to identify possible sources of "noise" in the measurement of responses. "Noise is caused by factors external to the advertising itself which affect viewer reaction. If such factors are significant and uncontrolled, they reduce the reliability and validity of copy test results" (Schlinger, 1982:29). Her rationale was based on several articles that were published about the effect of demographic variables and situational factors on commercial recall, and she investigated whether similar trends could be observed when attitudinal scales are used.

The research reported involved a 37-item VRP, and 4380 respondents were exposed to more than 140 different commercials. The purpose was to "find out what, if any, respondent or situational characteristics affect copy-test responses across different brands, product types, and executional styles" (Schlinger, 1982:31). Six factor dimensions were used in this research, namely, Relevant News, Brand Reinforcement, Stimulation (Enjoyment or Entertainment), Empathy, Familiarity and Confusion.

Multiple regression was used, using the six factor scores as dependent variables, as well as using an overall total score as a seventh dependent variable. The predictor variables were the consumer demographics, product usage variables and situational variables.

The results indicated that viewers' attitudinal responses were significantly, but not strongly related to certain demographic and usage variables. Age, sex and brand

usage were significantly related to the dependent variables, and that brand usage was more strongly related to the attitudes than demographic or situational variables.

This study demonstrated that differences in attitudes using the VRP due to demographic, product usage or situational factors are congruent with the factors that affect commercial recall. This rigorous study once again demonstrated the usefulness of the Schlinger VRP.

2.9.3 Wells, Leavitt and McConville (1971)

This research was one of the first methods suggested to obtain subjective reactions to advertisements in a way that the attitudes of viewers can be used to compare advertisements, that is easy and simple to apply, and can be understood by the general public. The purpose was two-fold, the measurements should be both comprehensive and useful in order to discriminate between commercials.

The method followed to develop the scale was thorough, starting from qualitative research with an exhaustive list of possible responses, and several rounds of exploratory factor analyses and analyses of variance were used to reduce the dimensionality and retain items that discriminated between advertisements. (If one takes into account the limitations that these researchers battled with such as slow computers and punch card systems, these contributions can only be admired.)

A weakness of the Wells *et al* (1971) study is that the article lacks details on several phases of the study. It does not give sample sizes or the number of commercials used in subsequent phases of the study. Also, no reliabilities were reported, and no use was made of item-total correlations to eliminate non-contributing items. There is no evidence of reincorporating eliminated items, and there is little evidence of references to any theories of advertising.

The descriptive words listed in the Appendix of Wells *et al* (1971), date strongly to values and norms of that era. This researcher can think of several words that are very descriptive of commercials currently used, such as sexy, passionate, radical,

arrogant, to name a few. These words do not appear in the Wells *et al* (1971) list. Three decades ago, communities were less explicit, and therefore it can be argued that the types of responses obtained then, would have been very different from those of today.

2.9.4 Zinkhan and Fornell (1985)

The purpose of this research was to confirm the dimensionality of the Wells (1964) reaction profile and the Leavitt (1970) commercial profile, and to examine the validity and reliability of these profiles. For each of the profiles, 20 commercials and 20 different subjects were used, a sample of 400 independent individuals for each profile. Confirmatory factor analysis and target rotation were used to examine the factor structures. For the Reaction Profile of Wells (1964), the coefficient of congruence was (CC=0.943), which is very high and this indicated that the factor structure was largely confirmed. For the Commercial Profile, the coefficient of congruence was (CC=0.58), which indicated that the factor pattern suggested by Leavitt (1970) was not satisfactorily confirmed. In terms of predictive validity, the Reaction Profile of Wells (1964) was also found to be better than the Commercial Profile of Leavitt (1970).

The study by Zinkhan and Fornell (1985) made an important contribution since it was one of the first attempts to use confirmatory factor analysis in testing for the construct validity of the dimensions proposed by advertising response scales. It also took one step further towards examining the predictive validity of the scales. The subjects were actual consumers and the predictive validity of the scales was assessed by measuring two additional measures, namely purchase intention (PI) using Juster's (1964) probability approach and attitude towards the brand (Petty, Cacioppo and Schumann, 1983).

Zinkhan and Fornell (1985) did not include the Schlinger (1979) scale in their research, although they did refer to it (Zinkhan and Fornell, 1985:447). This weakness was probably pointed out by a reviewer, and was corrected in subsequent research (Zinkhan and Burton, 1989).

2.9.5 Lastovicka (1983)

Lastovicka (1983) was the first researcher who provided evidence for the convergent and discriminant validity for an advertising response scale using the multitrait-multimethod (MTMM) (Campbell and Fiske 1959) approach. In this study, 634 students were exposed to one of six different 60-second television commercials in a classroom setting where approximately 30 students were present. The commercials were relevant to the students, and the branded products were used or purchased by 83% of the sample. A sixteen item structured questionnaire, which was a combination of eight items from the Wells, Leavitt and McConville (1971) scale and eight items from the Schlinger (1979) VRP, was one method used by Lastovicka (1983). The summated scales were:
Relevance – X1 = sum of 6 items (one reverse coded);
Confusion – X2 = sum of 5 items (one reverse coded); and
Entertainment – X3 = sum of 5 items (two reverse coded).

Scale reliabilities reported were: Relevance, 0.846; Confusion, 0.731 and Entertainment, 0.872. The sixteen item questionnaire with the Likert scale was one of the two methods required for the MTMM analysis.

The second measurement method used was open questioning. The open-ended response was coded by two independent coders on a three-point scale, with
0 = none of the response;
1 = some of the response; and
2 = much of this response;
and with all 634 verbatims coded by one of the coders, and 300 responses coded by a second coder. The simple product-moment correlation between the coders was 0.70. Only the first coder's scores were then used to obtain three summated scales which are listed below.

Relevance: X4 = sum of counter-argument, source rejection and message acceptance.
Confusion: X5 = sum of confusion and a reverse coded comprehension score.
Entertainment: X6 = sum of an entertainment and reverse coded irritation score.

The method of Kalleberg and Kluegel's (1975) to analyse the multitrait-multimethod data with confirmatory factor analysis was used to assess convergent and discriminant validity. The analysis displayed a clear support for the convergent and discriminant validities of the three multiple-item scales, X1, X2 and X3. Random measurement variation or unreliability was reported for the Confusion structured measure (X2), and it was found that the dimensions were related to one another in a consistent and expected manner.

Lastovicka (1983) used a very rigorously designed study, adopting the multitrait-multimethod approach suggested by Campbell and Fiske (1959) to demonstrate the convergent and discriminant validity of the of the three dimensions Relevance, Confusion and Entertainment, as reported by Schlinger (1979) and Wells, Leavitt and McConville (1971). In this method the "external validation" was obtained by using two methods, namely the Likert-type structured scales as suggested by Schlinger (1979) and Wells *et al* (1971), and an open questioning method in which viewers had to list retrospectively the thoughts they had while viewing the commercials. The concepts used in the analysis was based in Krugman's (1967) Three-Exposure Theory, which was one of the first attempts to ground affective or emotional responses to advertising in theory.

The sample consisted of 634 undergraduate students from a university business school, which may not be representative of how consumers would behave in a natural non-forced exposure setting. The study was limited to only three concepts, and only two methods, and there may be many more concepts that are relevant and important to advertisers. A more rigorous approach would be to employ more than two methods, and where the methods included other measures than self-report measures, such as unobtrusive and involuntary measures. These limitations were reported by Lastovicka (1983).

This study actually confirmed that at least three of the seven constructs suggested by Schlinger (1979) possess convergent and discriminant validity in a fairly realistic setting.

2.9.6 Zinkhan and Burton (1989)

In this study, Zinkhan and Burton (1989) partially duplicated an earlier study (Zinkhan and Fornell (1985), which was covered in Section 2.9.4), but this time they included the Schlinger (1979) VRP, and used Procrustean rotation in the confirmatory factor analysis. The advertising response scales that were compared are the Wells (1964) Reaction Profile, the Leavitt (1970) Commercial Profile and the Schlinger (1979) Viewer Response Profile. Unfortunately, only the so-called four 'stable' dimensions of Schlinger (1979), namely Entertainment, Confusion, Relevant News and Brand Reinforcement, were used, leaving three dimensions of the scale untested. This study also included choice behaviour measurement. The Wells (1964) Reaction Profile and the Leavitt (1970) Commercial profile were found to have satisfactory psychometric properties, whereas the Schlinger (1979) VRP was seen as "not as successful" (Zinkhan and Burton, 1989:10) in the establishment of the psychometric soundness of the scale. There were problems with at least three of the four dimensions. It was partially successful in that the first three dimensions, (Entertainment, Confusion and Relevant News) emerged to some degree and the Confusion dimension was seen as the most stable dimension. However, the Schlinger (1979) factor solution was not regarded as being as stable as the Leavitt (1970) solution.

This study is an important contribution to the work of comparing scales of affective responses to advertising, and evaluating their psychometric properties, which had not been reported before in the advertising literature. This article also emphasised to the importance of a more precise conceptual basis or theoretical base for explaining certain dimensions, and referred to Krugman's Three Exposure Theory as examined by Lastovicka (1983).

A major weakness of this study was the fact that the observations were not independent, in the sense that subjects had to judge all the commercials or print advertisements to which they were exposed. Only 20 subjects $\times$ (each exposed to 25 commercials) = 500 advertisement exposures were used for the Leavitt Commercial Profile, and only 25 subjects $\times$ (each exposed to 25 commercials) =

625 advertisement exposures were used for the Schlinger (1979) VRP. One wonders how the subjects felt after they had had to use the same instrument 25 times over! Over and above this, the samples obtained actually consisted of repeated measures. The assumption of independent observations was actually ignored in the statistical analyses. Another weakness is that only four of the seven Schlinger dimensions were considered in this study. The criticisms against the Leavitt (1970) and the Schlinger (1979) scales could be too harsh, or even not legitimate, because of the limitations of the sample and the violations of the assumption of independent observations. The researchers did however admit the limitations of the findings due to the small sample of advertisements and respondents.

2.9.7 Moldovan (1985)

Moldovan (1985) used data obtained from hundreds of commercials, thousands of respondents and 37 variables or items with the aim of determining which elements were necessary (although not sufficient) for successful commercials. Moldovan (1985) made the assumption that persuasive commercials were successful.

The variables measured

The first three items were related to comprehension of the advertisement, the main idea, the brand name and the theme. The next 20 items were related to the perceived involvement of the viewer, which was developed over a time period, starting with qualitative research, and then identification of items that discriminated between commercials. The items were representative of a comprehensive list of the various typical responses that one could expect from a broad spectrum of viewers exposed to a broad spectrum of commercials for a variety of products using different styles of execution. Feelings were also considered to be important, and eleven items were used to assess the degree and nature of respondents' positive and negative reactions.

A persuasion score was obtained by using three items. The first item represented competitive superiority, a measure of the perception of the uniqueness of the

brand. The second item, purchase intent, measured the degree to which the commercial has stimulated an interest in buying the product. The third item, coupon choice, measured the degree to which the respondent was interested in receiving a discount coupon for the advertised product.

Factor analysis and regression analyses

Exploratory factor analysis was used with oblique rotation (which allows for dependence between factors extracted) on all 37 variables, with thousands of respondents and hundreds of different commercials. Six factors were identified, namely the Persuasion factor, which was defined as the dependent variable in the study. The other five factors were Clarity, Tastefulness, Empathy/Self-Involvement, Stimulation and Credibility.

All five factors correlated very highly with the Persuasion factors, which meant that they all contributed positively to the success of a commercial. When used in step-wise regression, the variable that entered first and was the most essential element in predicting advertising success, was the credibility factor. It was therefore reasoned that credibility was a necessary condition for persuasive advertising. However, it was admitted that this is not enough, and the other elements namely stimulation, tastefulness, empathy/self-involvement and clarity were also important, but in different degrees, depending on the product category. This study has been an important contribution in understanding the factors that are absolutely essential in creating successful commercials. It also demonstrated a certain overlap with the construct dimensions of Wells, Leavitt and McConville (1971) and Schlinger (1979), as shown in Table 2.8.

2.9.8 Aaker and Bruzonne (1985)

Aaker and Bruzonne (1985) focused on the causes of irritation in advertising, and considered the effects of product class, audience group and style of execution approaches. It was found that certain product classes are invariably perceived to be very irritating, such as feminine hygiene products.

Table 2.8: Similar constructs in advertising response scales

Wells *et al* (1971)	**Schlinger (1979)**	**Schlinger (1982)**	**Moldovan (1985)**
	Brand Reinforcement	Brand Reinforcement	Persuasion
Confident, consistent, professional (unstable dimension)			Credibility
Humor /Vigour	Entertainment	Stimulation	Stimulation
Irritation		Confusion	Clarity
Irritation	Alienation		Tastefulness
Sensuousness	Empathy	Empathy	Empathy / self-involvement
Personal relevance	Relevant news	Relevant news	
Familiar (unstable)	Familiarity	Familiarity	
Uniqueness (perhaps opposite of familiarity)			

Factors that were found to increase irritation were, if the product was sensitive and showed the product; if the situation was unbelievable, phony or contrived; or if a person was put down in either appearance, knowledge or sophistication. Other factors that contributed to irritation were: when the story threatened an important relationship, such as mother-daughter, husband-wife; or when there was a graphical display of physical discomfort; tension situations; unattractive or unsympathetic characters were used; a suggestive scene was used or where there were poor casting or execution.

Factors that reduced irritation were when there was perceived to be a good casting and execution, a positive happy mood, tender phrases such as love and care and a warm and wonderful mood, an appropriate credible spokesperson, and when amusing and informative commercials were used. Irritation levels were significantly higher for higher levels of socio-econonomic status, younger respondents and non-users of products.

2.9.9 Summary of Advertising Response Scales

In summary all the papers covered in Sections 2.9.1 to 2.9.8 have contributed to the knowledge about the dimensionality of the affective responses in advertising. These studies provided researchers and practitioners with sound tools to measure affective responses that measure a broad comprehensive spectrum of responses, and these can be effectively be used to discriminate between commercials on a broad range of concepts. A subsection of this research revisits the Schlinger (1979) VRP and examines its reliability and validity and contrast this with the findings already noted in the advertising literature. This is covered in Chapter 5 of this thesis.

2.10 SUMMARY

This literature review dealt with bias and equivalence issues that are pertinent in the data analysis phase of a study. It briefly covered the literature on available methods of dealing with non-equivalence in cross-cultural data analysis. Methods to adjust data for bias include standardisation (Fischer, 2004) and normalisation (Stacey, 2005). An approach to deal with data in which bias may be present, arises from the theory of classical test theory (CTT). The multi-group confirmatory factor analysis (MG CFA) method has been applied by several authors dealing with cross-cultural research problems. The mathematical background of CFA and the relevant measures of fit, as well as the set of hierarchical hypotheses used in SEM, were covered in considerable detail. A more recent approach to study equivalence is the IRT approach. The CFA and IRT approaches were compared and limitations of the IRT approach were noted. Finally, since the empirical section of this study uses data obtained from an advertising company, it is therefore appropriate to cover the substantive literature on the topic of emotional response to advertising in sufficient detail. Chapter 3 deals with the methodology and mathematical background of SIMNORM and SIMITNORM, whilst the research methodology is described in Chapter 4. A summary of the literature review is presented in Table 2.9.

Table 2.9 Summary of literature review and implications for researchers in cross-cultural research contexts

Errors, bias and equivalence concerns in different phases of quantitative research projects and potential approaches				
PHASE	**TYPE OF BIAS**	**Description of the specific bias problem**	**Consequences of bias**	**Approaches**
ALL PHASES	All bias in general	Bias a concern during all phases of research	Ignoring potential sources of bias may seriously threaten the validity of a study	Be much more aware during all phases of research to eliminate bias. Apply appropriate qualitative methods to ensure constructs have face validity across groups. Use appropriate methods in research design, instrument development, data collection and data analysis. Be careful with interpretation of research results. This issue highlights the importance of this research.
Conceptualisation and problem definition	Problem bias	Problem definition done inappropriately Wrong propositions and/or hypotheses pursued.	Waste of research effort	More thorough exploration of problem and literature prior to instrument design and data collection. Thorough and careful prior observation and exploration of phenomenon and context in research design.
Research Design	Design bias	Researcher does not address or explore issues in the design phase of the study that may threaten the validity of the research	Research lead to invalid/wrong conclusion due to design problems Internal validity threatened	Thoroughly describe and explore implications of chosen research design.

PHASE	TYPE OF BIAS	Description of the specific bias problem	Consequences of bias	Approaches
Research Instrument development and administering the instrument	Measurement bias From a classical test theory perspective, the basis of the CFA approach, measurement bias is further classified as construct bias, method or response bias and item bias	What may appear as significant differences in mean values may only be response style effects, or may be a mixture of response style and substantive differences. Absence of a significant difference in mean values may wrongly obscure true substantial differences, due to differences in response style. May artificially inflate or deflate covariances between constructs	Worst case: Construct bias Constructs not equivalent Comparisons over constructs invalid	After data collection, from a diagnostic perspective, test the invariance/equivalence across groups using CFA and/or IRT methods, where the measurement model assumes an underlying latent variable or latent trait. Invariance / equivalence / non-invariance diagnosed and/or corrected through explicit procedures.

PHASE	TYPE OF BIAS	Description of the specific bias problem	Consequences of bias	Approaches
Development of research instrument, construct definition	Construct bias An artefact of the construct or the operationalisation of the construct within the group or both	Constructs not similarly defined across groups included in the study Constructs may have different meanings in different groups Use H1 in the MG CFA approach. When H1 is rejected, reasons for construct bias should be investigated. When H1 is tenable, it is still possible that the construct is not fully defined in one or more groups. This cannot be tested statistically. Qualitative methods an/or literature review may be options.	Comparing apples with pears and find a difference or no difference – conclusions are invalid (Van de Vijver and Leung, 1997)	Analytical methods to test for construct equivalence may not be sufficient to indicate lack of construct bias, especially when constructs are not fully defined across cultures, or when construct domain is ill defined and not adequately covered by sample of items. Construct equivalence can be evaluated using multi-group CFA. If this test is not tenable, construct bias is a problem and construct equivalence is ruled out. Comparisons cannot be made if it is a serious problem. Partial measurement invariance (Byrne, Shavelson and Muthén, 1998) can be an option if only a few items are affected. If the test is tenable, it does not rule out construct inadequacy. Further multi-group CFA tests are necessary before comparisons can be made. Qualitative methods should be applied to investigate construct validity across groups. If construct bias can be argued not to be an issue, construct equivalence may be assumed.

PHASE	TYPE OF BIAS	Description of the specific bias problem	Consequences of bias	Approaches
During instrument completion or measurement. Artefact of participant's predisposition and the group affiliation[1].	Response bias	When different groups in the study interpret the instruments differently, or display differential response styles.	Acquiescence bias; disacquiescence bias; net acquiescence bias; extremity ratings; response range; midpoint responding; non-contingent responding (Baumgartner and Steenkamp, 2001)	ANOVA method to identify type of bias Multi-group CFA applied to test for equal factor loadings and equal intercepts across groups. If this hypothesis is tenable, metric equivalence may be assumed (Cheung and Rensvold, 2000).
During instrument development or administration	Item bias	Single items have bias Lack of understanding from the researcher about item content interpretation within cultural groups	Item translation Item indicates additional/different latent constructs in one or more groups Item content inappropriate or even improper connotation	Translation and back-translation of items Explore cultural sensitivities to terminology and issues covered in item content using qualitative approaches and pilot testing (Van de Vijver and Leung, 1997)

1.Group affiliation could be cultural group or gender group

PHASE	TYPE OF BIAS	Description of the specific bias problem	Consequences of bias	Approaches
During completion of research instrument. An artefact of the method of administering the questionnaire, or an artefact of the response style of the participant, or of the group affiliation of the participant.	Uniform bias	A bias pattern that is consistent across all items in the instrument Acquiescence or disacquiescence bias A consistent higher or lower score across all levels	Consistent over-or under-estimation of differences in mean levels Covariances consistently attenuated or disattenuated Mainly a shift in location between groups Challenge is which part of the difference (or no difference) is due to substantive differences, and which is due to response style differences Issue: difficult and often impossible to estimate systematic error and random error separately	Use ANOVA method to test for type of bias present in the study (Van de Vijver and Leung, 1997; Clauser, Mazor and Hambleton; 1994) Use standardisation or ipsatisation to adjust score levels and take this into account during interpretation of results (Fischer, 2004) Use multi-group CFA to test for presence of bias and attempt to establish the type of bias present in the data. Attempts to estimate systematic error are only beginning to appear in literature, e.g. questionnaires with a good balance between positively phrased and negative phrased items for each construct (Baumgartner and Steenkamp, 2001) or methods suggested so far require panel data (Baumgarter and Steenkamp, 2006). SIMITNORM in this research investigated as a possible approach to deal with uniform bias.

PHASE	TYPE OF BIAS	Description of the specific bias problem	Consequences of bias	Approaches
During completion of research instrument. An artefact of the method of administering the questionnaire, or an artefact of the response style of the participant, or of the group affiliation of the participant.	Non-uniform bias	A bias pattern that is not consistent across all items or all score levels in the instrument when groups are compared, such as extreme response style, response range and midpoint responding (Baumgartner and Steenkamp, 2001)	Consistent over-or under-estimation of differences in mean levels Covariances consistently attenuated or disattenuated A shift in both location and spread across groups. Excessive kurtosis and skewness in some or all groups possible.	Almost no attempts in the literature to address this issue. In the CFA context, which can be applied over multi-dimensional constructs, partial measurement invariance (Byrne, Shavelson and Muthén, 1989) has become an accepted approach, if only a few items are affected by non-uniform bias. In the IRT context, which only allows for uni-dimensional constructs (a serious limitation), differential item functioning may be a very useful approach. SIMITNORM in this research is investigated as an alternative approach to deal with non-uniform bias.

PHASE	TYPE OF BIAS	Description of the specific bias problem	Consequences of bias	Approaches
Sampling design	Sampling bias	When the sample have an under- or over presentation of persons that distort the research results Non-response error	Non-response error Wrong sampling design Generalisations to population/similar context and/or similar times not valid. This bias introduces a lack of external validity	Efforts to improve sampling design Specific attempts to use probability sampling rather than non-probability sampling
Data gathering	Procedural bias Method bias	A bias that is introduced when different methods of interviewing or data collection is used for different groups in the study	Different methods of instrument administration or different contexts during data collection could introduce procedural bias	Explicit attempts should be pursued to ensure that different procedures during data collection do not introduce bias. When there is no alternative but to use different procedures, equivalence over methods should be tested using multi-group CFA to test for equivalence in procedure before further analysis may be attempted

PHASE	TYPE OF BIAS	Description of the specific bias problem	Consequences of bias	Approaches
Data Analysis and interpretation	Ignorance bias	Researchers typically assume that bias does not exist, or ignore bias issues in research	Explicit procedures should be followed to deal with bias (Reise et al, 1993; Smith and Reynolds, 2002; Fischer, 2004; Baumgartner and Steenkamp, 1998)	SIMNORM and/or SIMITNORM investigated in this research as potential additional procedure and compared to standardisation (Fischer, 2004) using multi-group CFA (Byrne, 1989; Bollen, 1989; Baumgartner and Steenkamp, 1998), a widely accepted procedure.
Conclusions and recommendations	Conclusion bias	Researchers wrongly interpret results and draw wrong conclusions	Waste of research effort Invalid conclusions Internal validity threatened	Researchers should improve skills in data analysis and interpretation and challenge own conclusions

CHAPTER 3
SIMULTANEOUS NORMALISATION: SIMNORM AND SIMITNORM

3.1 NORMALISATION IN MARKETING

This chapter covers the mathematical background to simultaneous normalisation across aggregate scores (SIMNORM) and across items (SIMITNORM). Section 3.1 sets out a brief background to normalisation; Section 3.2 deals with the mathematical principles and parameter estimation methods of the simultaneous normalisation (SIMNORM) technique; whilst in Section 3.3, the mathematical background and properties of several of simultaneous item normalisation (SIMITNORM) models are given.

3.1.1 Normalisation in psychometric testing

The SIMNORM method (Strasheim, 1989) that was used and extended in this research was originally developed for the purpose of test equating over various age groups. SIMNORM is a non-linear data transformation technique that can be used to transform an unknown distribution to a normal distribution. The technique involves the estimation of a non-linear transformation function that will take the unknown distribution to the normal distribution.

When more than one group is involved, SIMNORM can be used for the simultaneous transformation of scores over age groups, or gender groups. SIMNORM was originally developed to deal with data at aggregate level, where the items of a scale or sub-scale are combined to form a total score, and where the scores that are normalised can be assumed to be a continuous variable.

Normalisation was first developed for test equating in psychometric testing. Often test batteries are designed and applied to children from various age

groups. The test equating problem arises from the fact that children from different age levels, (aged 11, 12, 13 and 14 years), for example, are subjected to the same test battery. A test consists of multiple choice questions, where only one of the options offered is correct. The test battery outcome is calculated as the total number of correct items. The test is designed to discriminate between children with different abilities. The dilemma is that, on an aggregate level, younger children generally obtain lower scores and older children higher scores, due to the effects of experience and schooling. The usual approach is to transform the data in such a way that for each age group, the aptitude scores have equal distributions, with respect to shape, location and spread. The normal distribution is often used as a target distribution, which is then expressed as a stanine with a mean of 5 and standard deviation of 2, or IQ-scores such as SAT which is then transformed which has a normal distribution with a mean of 100 and a standard deviation of 15.

Various methods have been used to equate test scores to normalised scores. Techniques that are often used include equipercentile equating or linear equating methods (Angoff, 1971), and item response theory methods (Lord, 1977). Kolen (1981) provides a comparison between traditional methods (equipercentile and linear equating) with several item response theory methods to equate forms of similar difficulty (horizontal equating) as well as dissimilar difficulty (vertical equating).

Various alternatives to the traditional equipercentile method have been suggested (Angoff and Robertson, 1987). Lindsay and Prichard (1971) have suggested an analytical linear interpolation procedure to smooth the obtained distributions and develop equivalent scores. One method is to fit continuous curves to the empirical distribution functions by specifying some functional equation. Specific functions, for example the logistic (Marco, 1977) and quadratic (Lindsay and Prichard, 1971), have been suggested as alternatives to linear interpolation for computing percentile ranks. Cubic spline methods (Kolen, 1983; Jones and Sabers, 1984) were found to be effective in the smoothing of functions. SIMNORM (Strasheim, 1989) offers an alternative approach to these methods, since a set of smooth simultaneous transformation functions is

estimated. It has been used successfully in equating test scores on several sets of data (Strasheim, 1989).

SIMNORM is a set of smooth, mathematical transformations of raw test scores to normal scores with mean zero and variance one. The mathematical function in SIMNORM was developed to meet the following two constraints:

1) if two children from the same age group have received raw scores of x_1 and x_2 respectively, and $x_1 \leq x_2$; the transformed normalised scores $t(x_1) \leq t(x_2)$; and
2) if two children from different age groups received the same raw score, the SIMNORM procedure will ensure that the younger child does not obtain a lower normalised score than the older child.

The purpose of SIMNORM is to obtain test scores for the different groups of children, which are equivalent across the age groups. Within an age group, all the distributions should be similar and normally distributed, and the means and the standard deviations should be equal across age groups. The normalisation procedure has been mathematically designed to transform non-equivalent test scores to equivalent scores across different age levels.

3.1.2 Standardisation

The familiar standardisation formula, $Z = (X - \mu)/\sigma$, is a linear transformation, where the Z-score has a zero mean and a standard deviation of 1. Standardisation is linear, and relationships (correlations) that existed before the standardisation are not affected. The shape of the distribution does not change after standardisation.

Consider the three normally distributed randomly generated data sets of 100 observations each as presented in Table 3.1.

Table 3.1: Three randomly generated data sets

Raw scores				Standardised Scores			
Variable	Mean	Std dev	Distribution	*Z*-Variable	Mean	Std dev	Distribution
X_1 (blue)	50	10	normal	Z_1 (blue)	0	1	Normal
X_2 (red)	80	20	normal	Z_2 (red)	0	1	Normal
X_3 (green)	100	10	normal	Z_3 (green)	0	1	Normal

Figure 3.1 shows the results of standardisation as a transformation function when the above three data sets are standardised, with the original values X on the horizontal axis, and the standardised values Z on the vertical axis. The transformation only involves changes in locality and spread, and leaves the underlying shape of the distribution intact. Correlations between X_1, X_2 and X_3 are not affected by standardisation, and the same correlations will exist between Z_1, Z_2 and Z_3.

Figure 3.1: Standardisation is a linear transformation

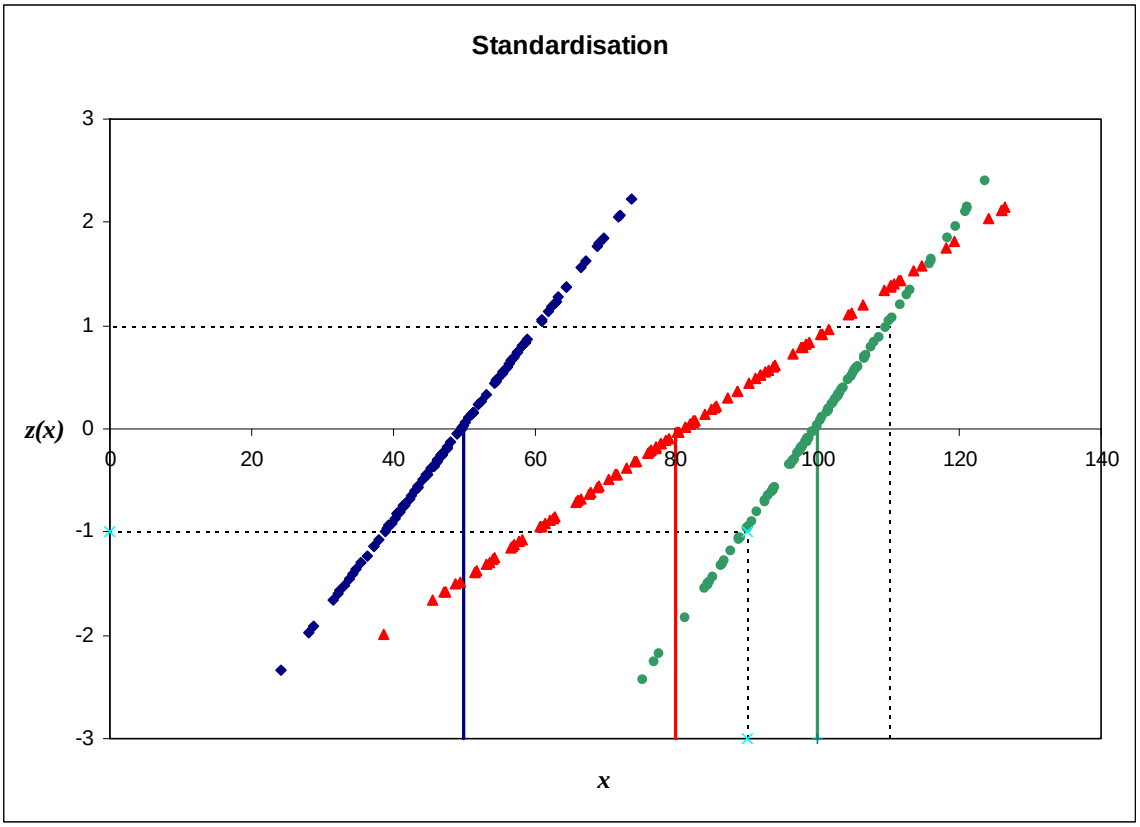


Table 3.2 summarises the major differences between standardisation and normalisation.

Table 3.2: Differences between standardisation and normalisation

Standardisation	Normalisation
Mean = 0	Mean = 0
Standard deviation = 1	Standard deviation ≈ 1 (depending on normalisation model used)
Function linear	Function nonlinear
Correlations remain the same	Correlations change
The shape of the distribution stays the same after standardization	The shape of the distribution becomes approximately a normal distribution
Outliers remain extreme	Outliers become less extreme

3.1.3 Normalisation

The purpose of normalisation is to estimate the function that will transform an unknown distribution to a distribution that has a normal distribution, with a mean of zero and a standard deviation of one. The normalisation transformation function (Strasheim, 1989) is based on the inverse of the Gompertz (Du Toit, 1979) growth function, and is defined as

$$t(X) = \frac{Y - \phi_1}{\phi_2} \qquad (3.1.1)$$

with

$$Y = -\ln\left(\left(\frac{U}{X}\right)^{\gamma} - 1\right) \qquad (3.1.2)$$

and with U > the maximum value of X.

Three nested simultaneous normalisation models have been developed. Model A is the least restrictive model, and all simultaneous normalisations are applied to each data set individually, except that the upper bound U, which is specified by the researcher, is the same for all the data sets. When Model A is used for normalisation, the results would be the same if three independent normalisation procedures were used. If g denotes the number of groups or data sets involved in the simultaneous normalisation, then the number of parameters p to be estimated in model A is equal to

$$p = 3g \ . \qquad (3.1.3)$$

In model B, one of the parameters, γ is restricted to be equal for all the subgroups, but the values of ϕ_1 and ϕ_2 are allowed to be unique for each group. The upper bound U, is specified to be equal to the maximum value in the data set. Model B is more parsimonious than Model A and the number of parameters is equal to

$$p = 2g + 1. \qquad (3.1.4)$$

In Model D, two parameters, γ and ϕ_2, are restricted to be equal for all subgroups, and only the parameter ϕ_1 is allowed to vary for each group. Model D is the most restrictive, but also the most parsimonious model, with the number of parameters

$$p = g + 2. \qquad (3.1.5)$$

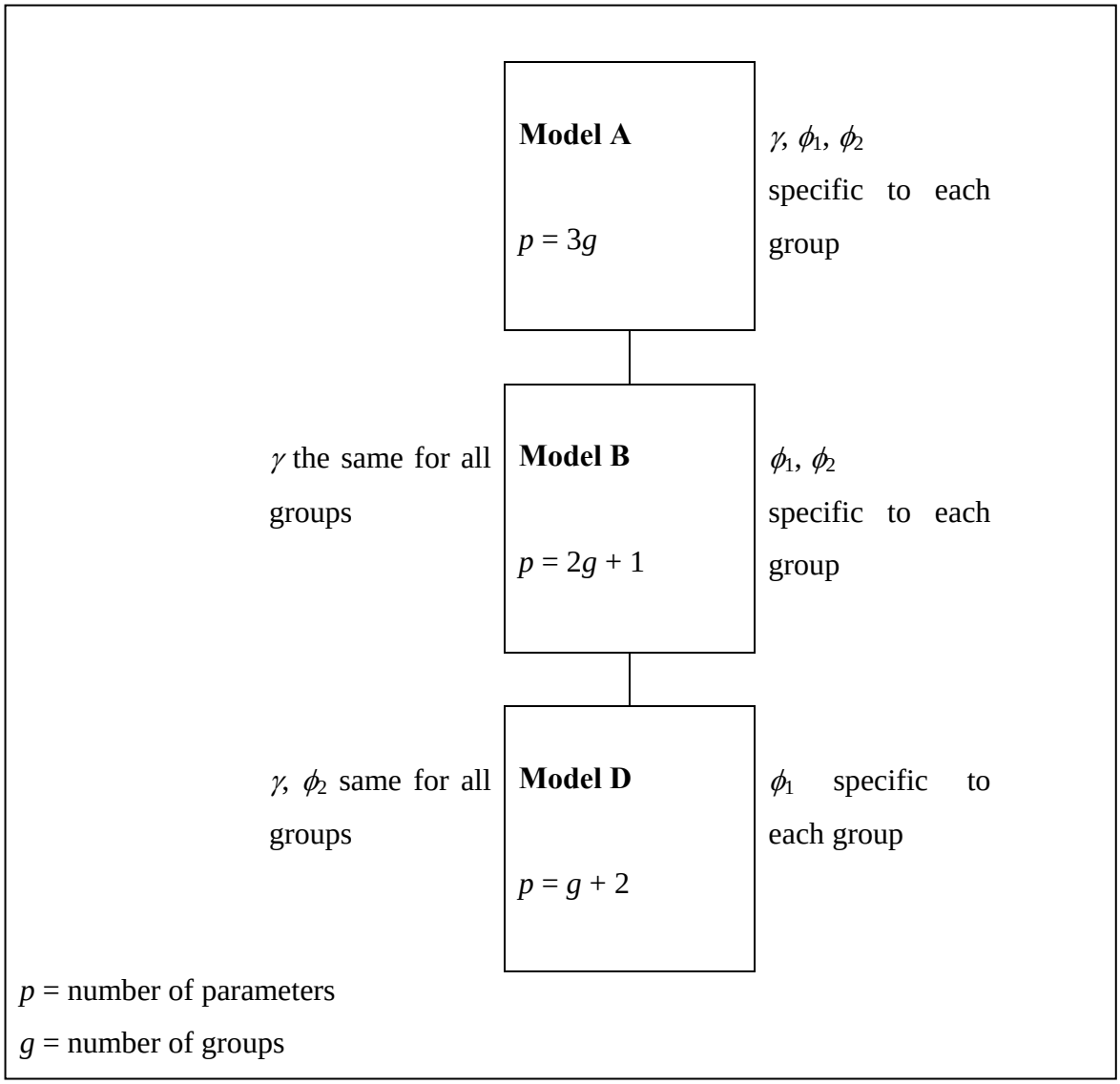
The hierarchy of models is shown in Figure 3.2.

For all three models, the normalised scores approximate a normal distribution with a mean of zero and a standard deviation of one. The normalised score X_T can be transformed into a new target mean μ_T and a new target standard deviation σ_T, using the following transformation formula:

$$X_T = t(x) \times \sigma_T + \mu_T \qquad (3.1.6)$$

with $t(x)$ the normalised score.

Figure 3.2: Hierarchy of SIMNORM models



Source: Strasheim (1989)

3.1.4 The simultaneous normalisation of four simulated non-normal data sets

For illustrative purposes, normalisation is demonstrated by applying it simultaneously to four simulated non-normal data sets. All three normalisation models A, B and D are used to demonstrate the differences between the models. The effects of parameter restrictions on shapes of the transformation functions are also demonstrated graphically. The first data set is a uniform distribution, the second distribution is bimodal, the third distribution is negatively

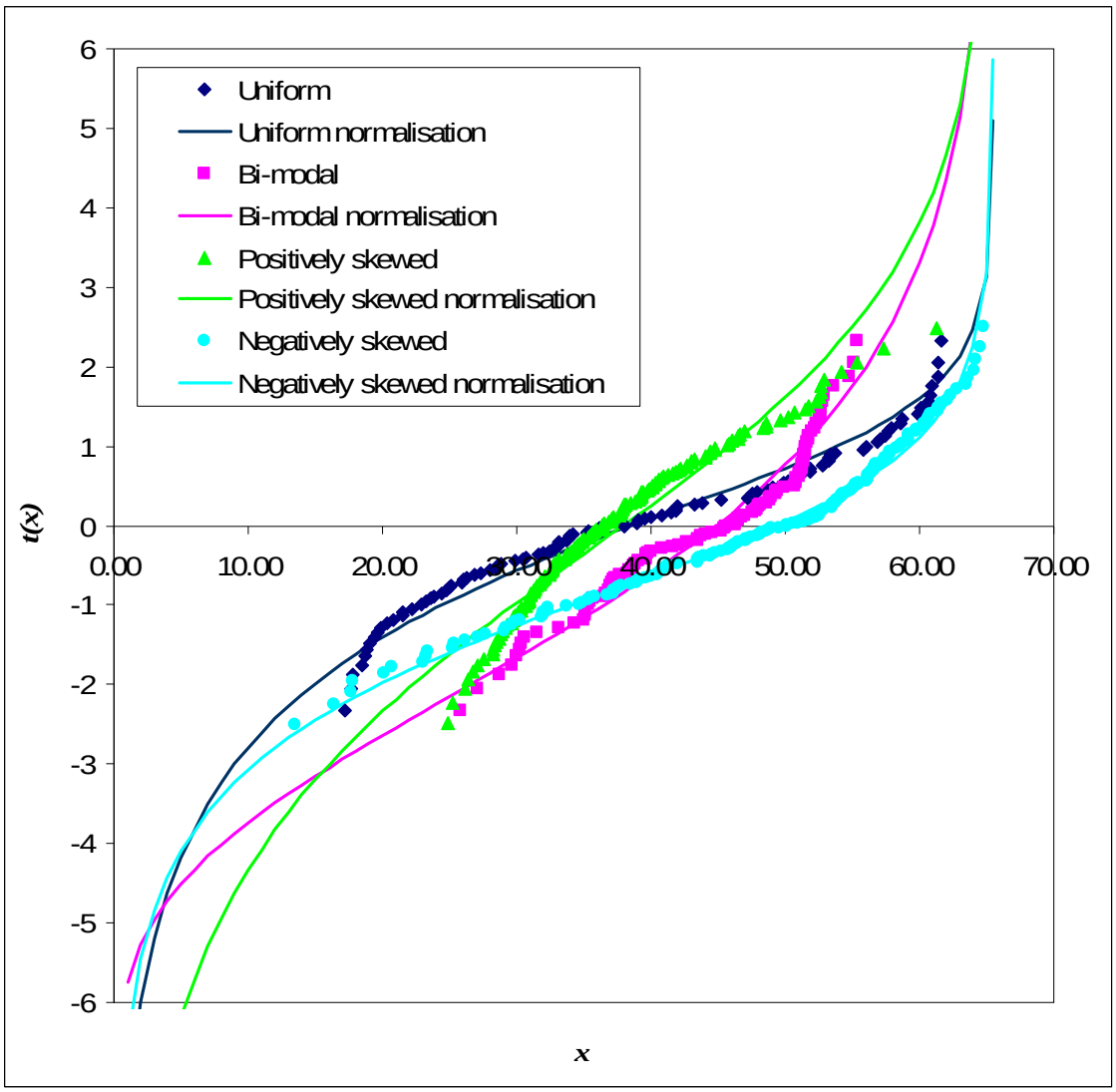
skewed and the fourth distribution is positively skewed. The results and descriptive statistics of the raw scores and the normalised scores for each of the models are summarised in Table 3.3.

Table 3.3: Descriptive statistics of raw and normalised scores

	U = 65.51	Uniform Distribution	Bimodal distribution	Positively skewed distribution	Negatively skewed distribution
	Sample size	100	100	152	160
Raw scores	Mean	38.320	43.430	37.105	46.669
	Std dev	13.891	7.648	7.389	11.946
	Min	17.22	25.85	24.93	13.43
	Max	61.61	55.29	61.27	64.75
Standardised Scores	Mean	0.000	0.000	0.000	0.000
	Std dev	1.000	1.000	1.000	1.000
	Min	-1.5190	-2.2986	-1.6477	-2.7824
	Max	1.6766	1.5507	3.2704	1.5136
Model A	Mean	0.000	0.000	0.000	0.000
	Std dev	1.005	1.005	1.003	1.003
	Min	-1.7067	-2.0693	-1.6256	-2.6178
	Max	1.8422	1.8217	4.3150	2.8420
Model B	Mean	0.000	0.000	0.000	0.000
	Std dev	1.005	1.005	1.003	1.003
	Min	-1.6854	-2.4560	-1.7986	-3.0070
	Max	1.8691	1.5924	3.8469	2.5398
Model D	Mean	0.000	0.000	0.000	0.000
	Std dev	1.326	0.651	0.647	1.210
	Min	-2.2660	-1.6138	-1.1812	-3.7462
	Max	2.4134	1.0164	2.4179	2.9628

The histograms of the raw scores and normalised scores can be seen in Figure 3.6. Graphical displays of the simultaneous normalisation transformation functions using Models A, B and D for the four simulated data sets are given in Figures 3.3, 3.4 and 3.5 respectively. Model A is shown in Figure 3.2, and the simultaneous normalisation functions S1, S2, S3 and S4 for each of the four data sets are, except for the upper bounds and lower bounds, unrestricted. They cross each other at either of the upper and lower bounds, as well as in the middle. Model A is the least restrictive model, and has the most parameters. In this example, four normalisation functions, all with the same upper bound (U = 65.51) were used, each with three parameters, involving the estimation of twelve parameters.

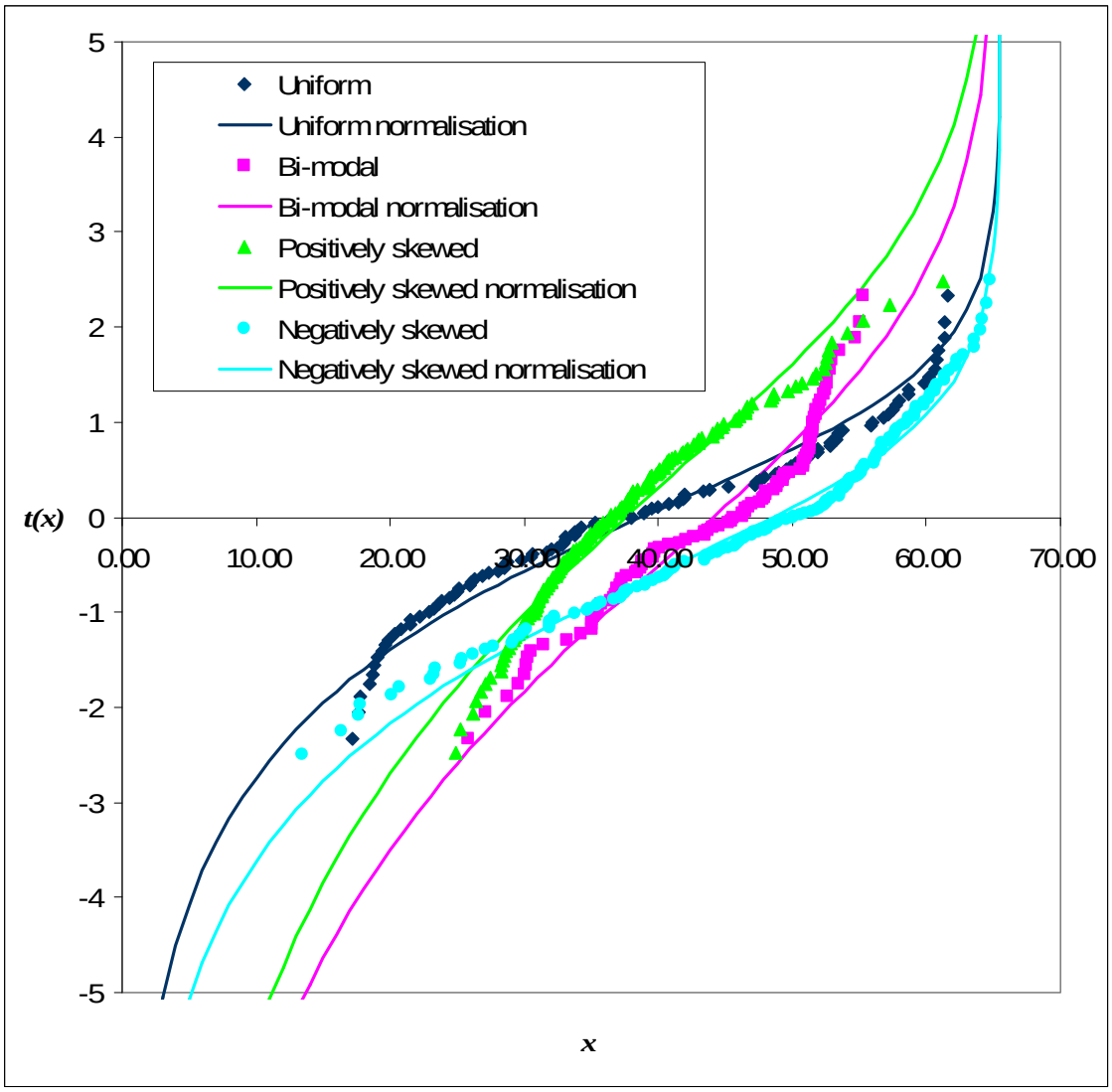
Figure 3.3: Model A fitted to four simulated non-normal data sets.



It should be noted that in Figure 3.3, the normalisation functions could cross each other more than once, since none of the estimated parameters were restricted. This is very clear if one considers the dark blue and the pink curves.

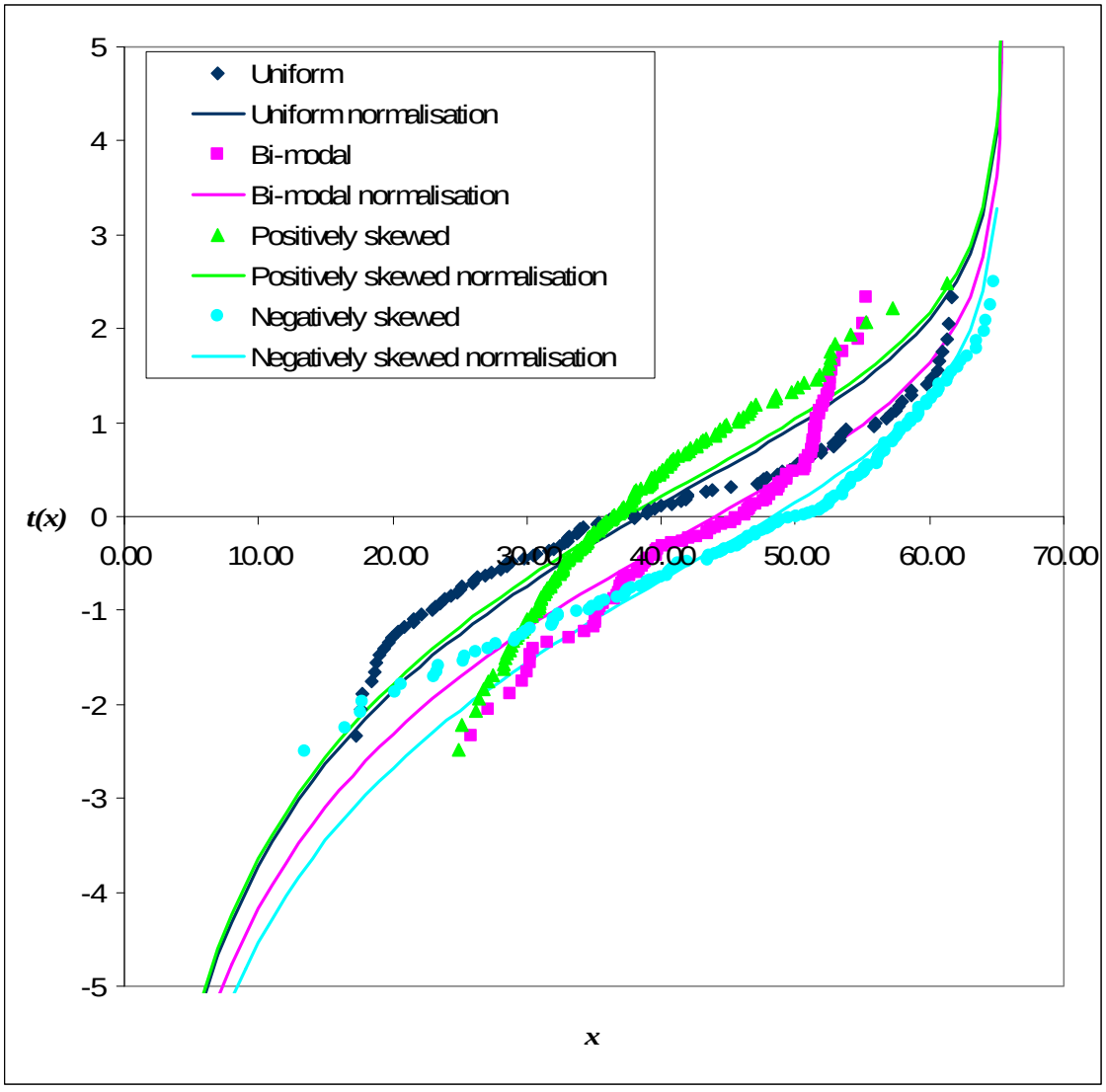
Model B is more restricted than Model A, so that the normalisation functions cross each other once only, as can be seen in Figure 3.4.

Figure 3.4: Model B fitted to four simulated non-normal data sets.



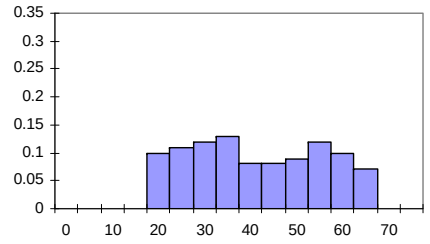
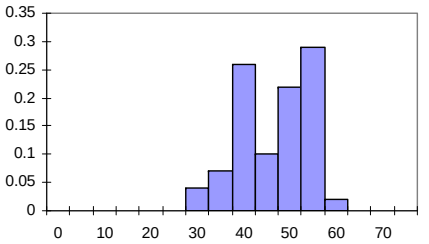
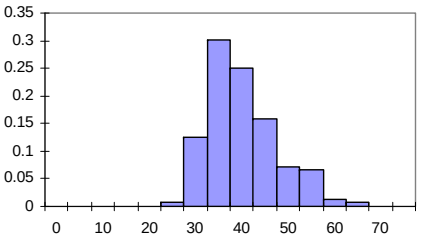
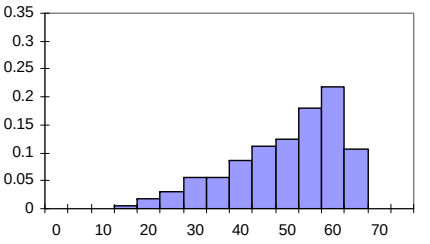
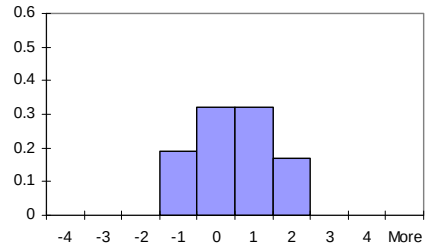
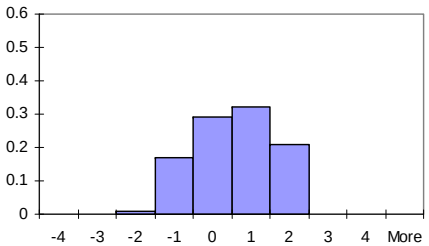
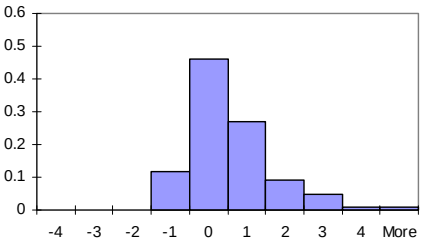
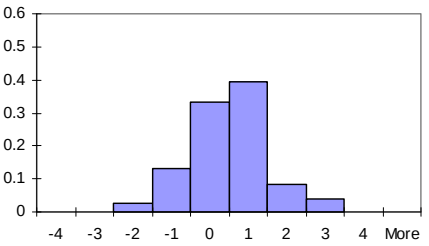
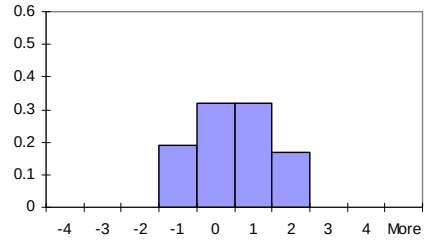
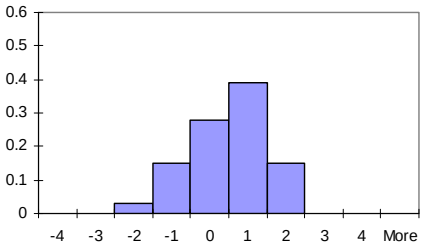
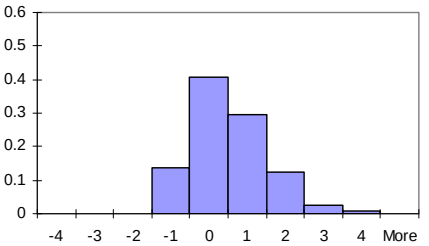
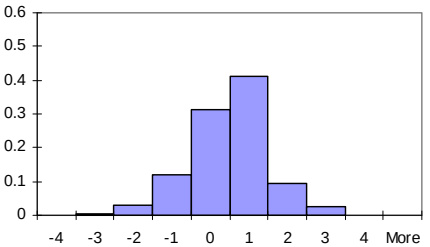
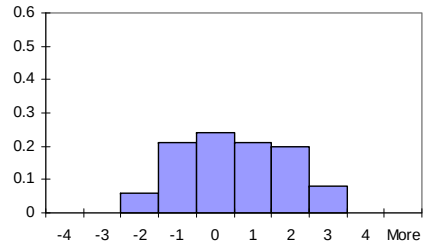
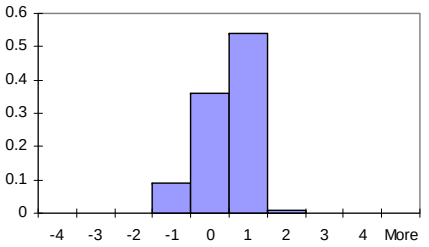
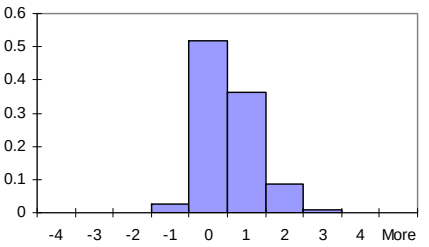
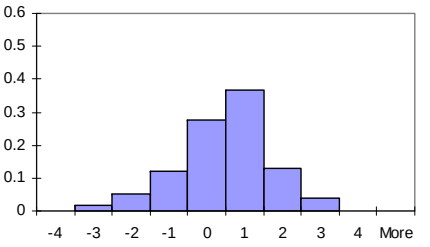
Model D is the most restricted model, and none of the lines are crossing, as can be seen in Figure 3.5.

Figure 3.5: Model D fitted to four simulated non-normal data sets



The graphical displays in Figures 3.3 to 3.5 show how the models become increasingly restrictive as they become more parsimonious. However, if the resulting normalised scores are studied in Figure 3.6, the differences between the models are hardly visible.

Figure 3.6: Four simulated distributions before and after normalisation

	Uniform distribution	Bi-modal distribution	Positively skewed distribution	Negatively skewed distribution
Distribu-tions of four simulated data sets				
	Normalised distributions	Normalised distributions	Normalised distributions	Normalised distributions
MODEL A				
MODEL B				
MODEL D				

3.1.5 Normalisation in cross-cultural marketing research

In marketing research it often happens that an instrument that has been developed in a specific setting or culture is applied in another culture, without much thought on the part of the researcher about the applicability of the scale in the new setting. The researcher might find that the results are significantly different for the cultural groups being studied. The difficulty is then to establish whether the observed differences are real cultural differences, or are caused by differential response styles (acquiescence bias) across cultures.

This research proposes that normalisation can be used to simultaneously normalise the key research variables across the cultural groups. The research variables in many studies involve the calculation of total aggregate scores, also known as factor scores, principal components or subscales. This allows the researcher to equate scores over the various cultural groups, thereby eliminating cultural bias from the data set. The normalised scores allow the researcher to study the relationships between subgroups, without the 'contamination' of cultural bias. Simultaneous normalisation offers an approach for adjusting the data for cultural differences, by equating the research variables across the cultural groups where bias is considered to be problematic.

It is further proposed that an advantage of normalisation is that the normalised scores will allow the researcher to use statistical approaches where normally distributed data is assumed, with more confidence. Simultaneous normalisation (SIMNORM) is an alternative to standardisation, and may offer additional advantages that are not possible with standardisation.

3.2 SIMULTANEOUS NORMALISATION (SIMNORM)

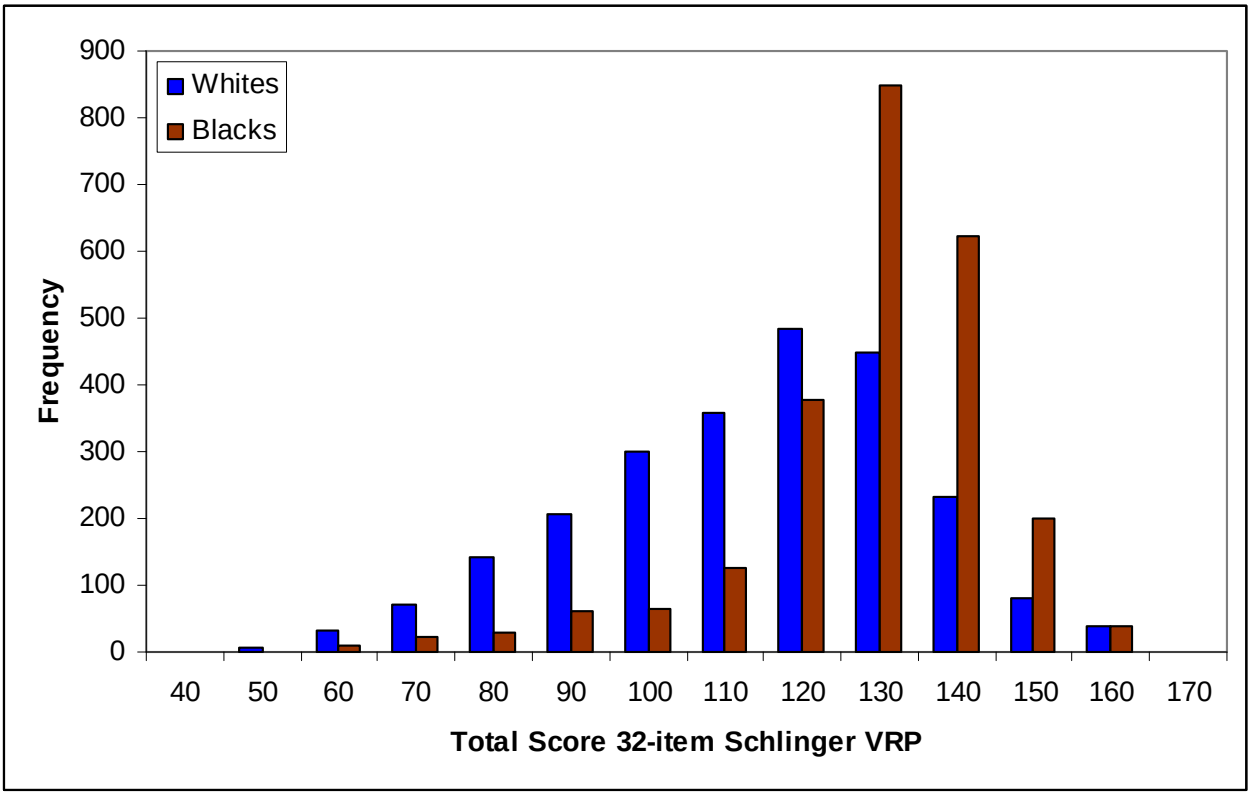
Normalisation (Strasheim, 1989) is a technique that can be applied in order to equate aptitude test scores using a non-linear transformation function that allows for a smooth, mathematical progression of normal transformation functions.

The normalisation technique is a simple approach which develops a formula for deriving equivalent normalised scores. This normalisation approach involves the estimation of a normalisation transformation function that will take an observed raw score as an input, and yield a normalised score. Simultaneous normalisation differs from existing procedures mainly in the mathematical theory that forms the basis of the smoothing procedure, which permits the simultaneous transformation of test scores from an unknown distribution to a distribution of normal scores in a maximum likelihood framework.

3.2.1 An example of SIMNORM across two cultural groups

In this section, for the sake of simplicity, only the two major cultural groups, namely the white and the black groups are considered to demonstrate the use of SIMNORM. In Chapter 5, SIMNORM is applied across four cultural groups. Figure 3.7 displays the total mean scores of 32 items of the Schlinger (1979) scale. The results are from a sample of 4800 whites and blacks, incorporating responses on 24 different commercials, each viewed by 100 whites and 100 blacks. It is clear that the overall responses on the Schlinger (1979) scale are very different between the two major cultural groups.

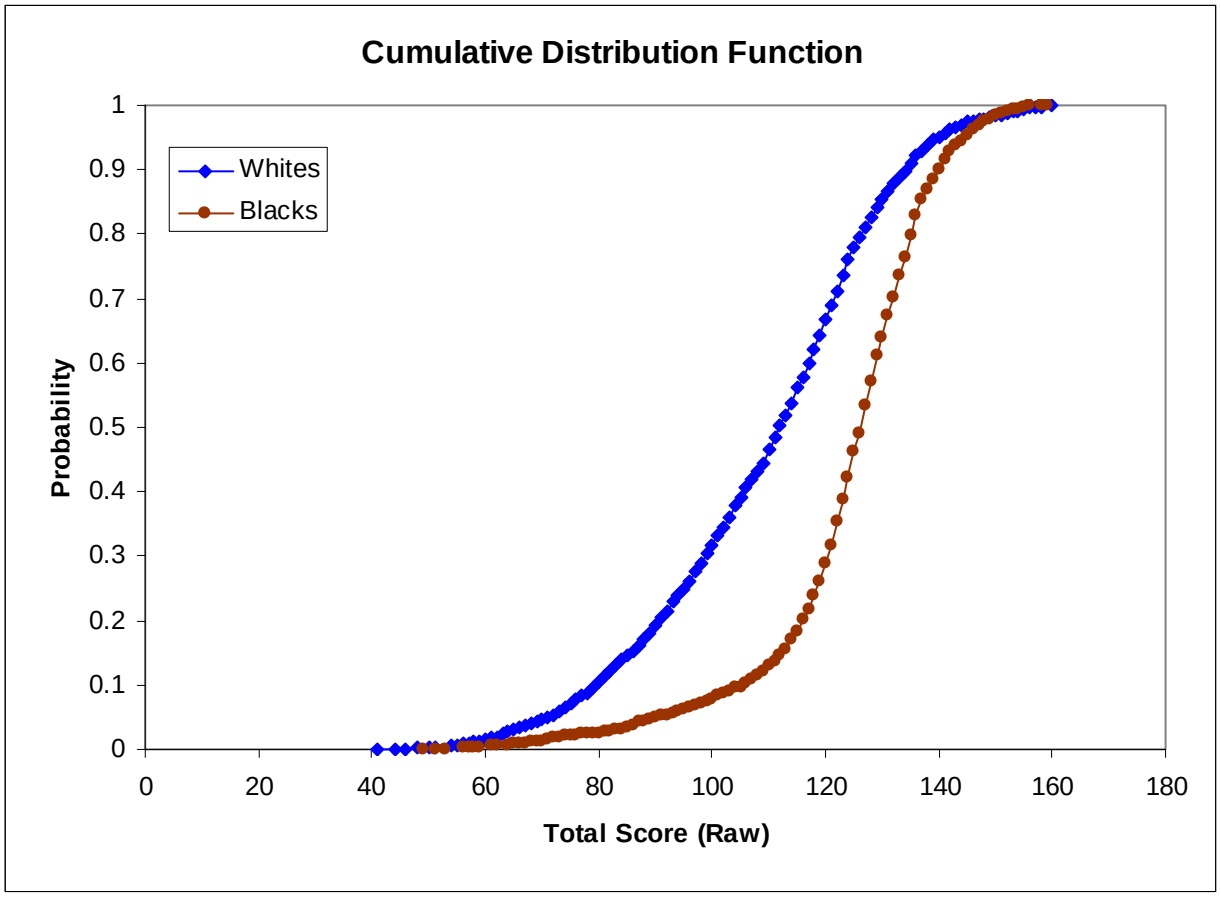
Figure 3.7: Histogram of the raw total scores of the 32 Schlinger (1979) scale for two cultural groups



The cumulative densities are shown in Figure 3.8. The cumulative density functions clearly show that there is not only a locality difference between the two cultural groups, but also a difference in both spread and shape, as shown in the cumulative density functions in Figure 3.8.

Sections 3.2.2 to 3.2.4 deal with the general function that will allow the transformation of non-normal scores to normally distributed scores for a single group. In Section 3.2.5, the transformation is expanded so that scores can be normalised simultaneously over a number of subgroups. Section 3.2.6 shows the results of the simultaneous normalisation of the total scores of the example given in Section 3.2.1.

Figure 3.8: Cumulative density of the total score of the 32-item Schlinger scale for two cultural groups



3.2.2 The target distribution estimation method

Let $F(x)$ denote an unknown continuous distribution function of the variable X, from which a sample $x_1, x_2,..., x_N$ has been observed. Let $H(z)$ be the target distribution. Since $F(x)$ and $H(z)$ denote two non-defective distribution functions, that is, they have a value of 0 at minus infinity, and a value of 1 at plus infinity, there exists (Kendall and Stuart, 1977) a transformation function $z = t(x)$ that can be defined as

$$z = t(x) = H^{-1}[\, F(x)\,] \qquad (3.2.1).$$

Since both $F(x)$ and $H(z)$ are continuous, the transformation function $t(x)$ will be non-decreasing and continuous and its inverse will exist (Kendall and Stuart, 1977).

The target distribution estimator (Elpinstone, 1985) can be written as

$$F(x) = H[t(x)] \qquad (3.2.2)$$

and the target density estimator as

$$f(x) = h\left[t(x)\right] \cdot \left| \frac{\partial\, t(x)}{\partial x} \right| . \qquad (3.2.3)$$

The estimation method is concerned with approximating the transformation function $t(x)$ which takes the observed distribution, $F(.)$, into the target distribution, $H(.)$, based on data observed from $F(.)$.

For the purpose of estimating the transformation function, which takes the observed distribution to the normal distribution, the target distribution is in this case the standard normal distribution. The estimated transformation function can be used to transform the observed distribution to the standard normal distribution with mean zero and variance one.

The density of the standard normal distribution is

$$h(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2} \quad . \qquad (3.2.4)$$

The target density estimator (Elphinstone, 1983; 1985) with the normal distribution as target can be expressed as

$$\begin{aligned} f(x) &= h\left[t(x)\right] \times \left| \frac{\partial\, t(x)}{\partial\, x} \right| \\ &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}[t(x)]^2} \times \left| \frac{\partial\, t(x)}{\partial\, x} \right| \end{aligned} \quad . \qquad (3.2.5)$$

The only parameters that require estimation in (3.2.5), when the standard normal is chosen as the target distribution, are the parameters of the transformation function $z = t(x)$.

3.2.3 The transformation function for a single group

The shape of the transformation function is suggested by the Q-Q plot (Du Toit, Steyn and Stumpf, 1986). It is useful to compare an empirical cumulative distribution function to a pre-specified distribution, which in this study is the standard normal distribution function. If the Q-Q plot does not form a straight line, it suggests the non-linear transformation function that can be used in (3.2.5). The transformation functions for two cultural groups are suggested by the Q-Q plot in Figure 3.9. The specific shape is not well known, but its inverse, as shown in Figure 3.10, can be closely approximated by the Richards (1959) family of growth curves.

The typical shape of the Richards growth function is depicted in Figure 3.11, and the inverse of the Richards growth function is shown in Figure 3.12.

Figure 3.9: Q-Q plot for two cultural groups

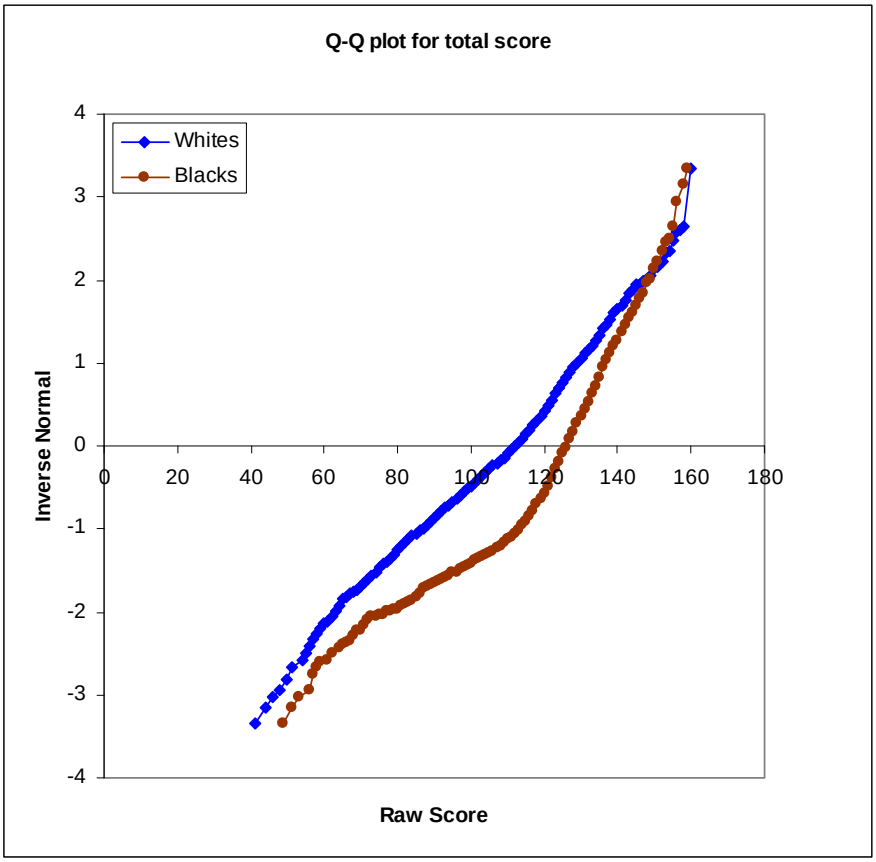


Figure 3.10: Inverse Q-Q plot of raw total score for two cultural groups



Figure 3.11: The general shape of the Richards family of growth curves

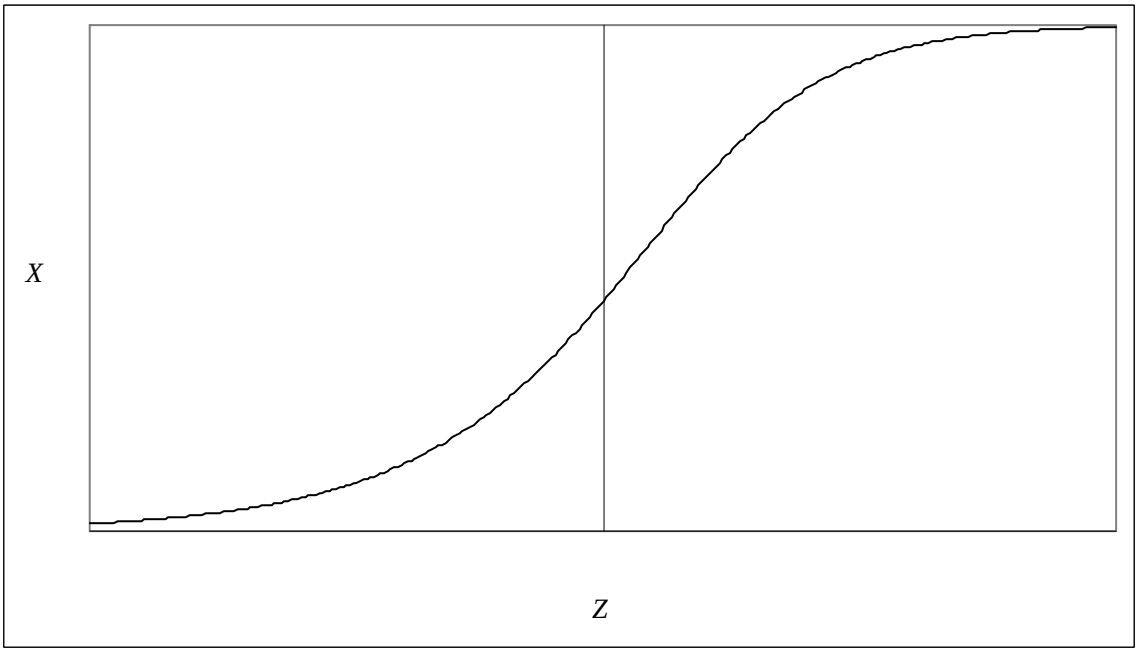
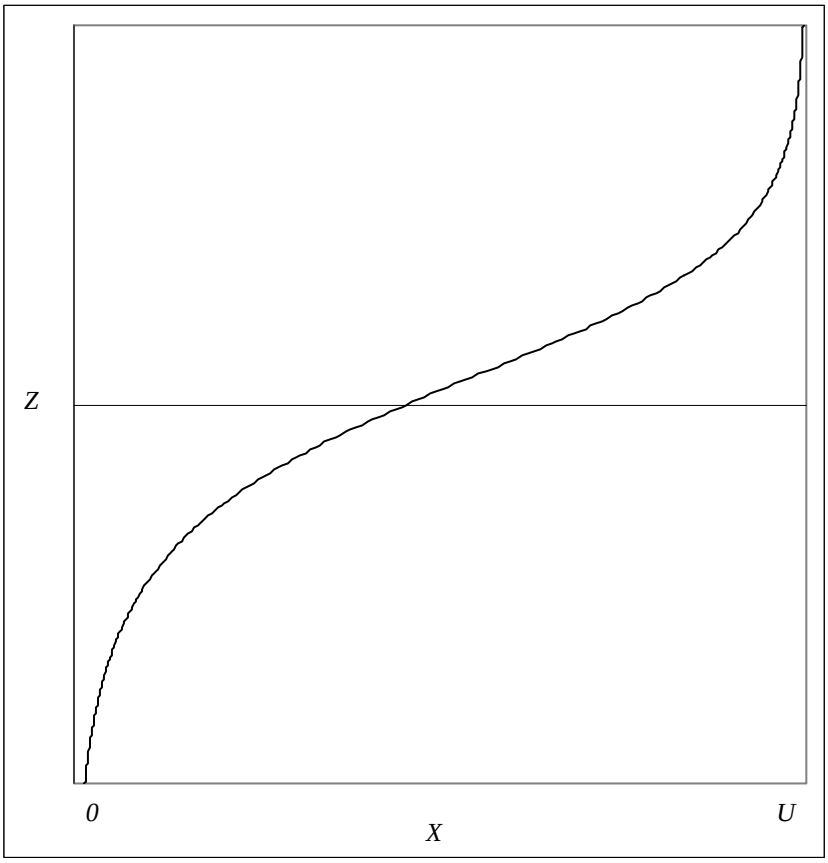


Figure 3.12: General shape of the transformation function



The transformation function as shown in Figure 3.12 shows the desirable properties for the purposes of this research. This research problem involves the estimation of a transformation function which transforms a distribution of bounded raw scores $0 \leq x \leq U$, to the normal distribution with infinite range, $-\infty < z < \infty$, and the shape of the function in Figure 3.12 will permit this.

The monotonically increasing Richards (1959) growth function as parameterised by Du Toit (1979:2.1-2.3) is defined as:

$$x = \alpha(1 + \beta\rho^{z})^{-\lambda} \qquad (3.2.6)$$

where $-\infty < z < \infty$ and with the following constraints on the parameters (α, β, λ, ρ):

$$\alpha > 0$$
$$\beta > 0$$
$$\lambda > 0 \text{ and} \qquad (3.2.7)$$
$$0 < \rho < 1.$$

The parameters of the Richards (1959) family of growth curves have physical meanings, which are helpful in the application of the function. The parameter α represents the asymptotic maximum value of the characteristic being measured. It therefore makes sense to set $\alpha = U$, the theoretical maximum total score that can be obtained. For the 32-item Schlinger (1979), using a 5-point scale, if a total composite score is calculated, after reversing negatively phrased items, the maximum possible score is $U = 32 \times 5 = 160$.

Since the Richards (1959) growth functions (3.2.6) are non-linear in the parameters, an iterative procedure is required for parameter estimation. Since most algorithms for estimation work better when parameters are unconstrained, and also for the sake of simplicity, suitable transformations should be done as described by Box, Davies and Swann (1969:42). Reparameterisation gives the results set out below:

$$\text{Let} \quad \beta = e^{-\phi_1}, \text{ with } -\infty < \phi_1 < \infty \qquad (3.2.8)$$

then β remains positive for any value of ϕ_1.

$$\text{Let} \quad \rho = e^{-\phi_2}, \text{ with } \phi_2 > 0 \qquad (3.2.9)$$

If $\phi_2 \to \infty$ then $\rho \to 0$ and if $\phi_2 \to 0$ then $\rho \to 1$.

$$\text{Let } \lambda = \frac{1}{\gamma} \text{ and } \gamma = e^{\phi_3}, \text{ with } -\infty < \phi_3 < \infty \qquad (3.2.10)$$

If $\lambda > 0$ then $\gamma > 0$, and $\gamma = e^{\phi_3} > 0$ for all values of ϕ_3 .

The Richards (1959) function, (3.2.6), which is the inverse of the transformation function, (denoted as t^{-1}) can be rewritten using (3.2.8), (3.2.9) and (3.2.10) to represent the relationship between the observed scores x and the target normalised scores z. The

result is

$$x = t^{-1}(z) = \alpha\,(1 + e^{-(\phi_1 + \phi_2 z)})^{-\frac{1}{\gamma}}. \qquad (3.2.11)$$

Since the parameter α can be substituted by U, the maximum score, the transformation function, $t(x)$, that maps the bounded raw scores x, $0 < x < U$ to unbounded normalised scores z, $-\infty < z < \infty$ for a single cultural group is given by the inverse of (3.2.11) and can be written as:

$$z = t(x) = \frac{y - \phi_1}{\phi_2}$$

where $y = -\ln\left(\left(\frac{U}{x}\right)^{\gamma} - 1\right)$ with $\gamma > 0$ (3.2.12)

and with

$$-\infty < \phi_1 < \infty$$
$$\phi_2 > 0 \qquad (3.2.13)$$
$$-\infty < \phi_3 < \infty$$

where $\gamma = e^{\phi_3} > 0$ for all values of ϕ_3.

3.2.4 The density and likelihood function for a single group

From equations (3.2.5) and (3.2.12) the density estimator with the normal target distribution is

$$f(x, \boldsymbol{\theta}) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t(x,\boldsymbol{\theta})^2} \times \left|\frac{\partial\, t(x, \boldsymbol{\theta})}{\partial\, x}\right| \qquad (3.2.14)$$

with $\boldsymbol{\theta}' = (\phi_1, \phi_2, \phi_3)$ and the partial derivative of the transformation function with respect to x expressed as

$$t'(\boldsymbol{\theta}) = \frac{\partial\, t(x, \boldsymbol{\theta})}{\partial\, x} = \frac{\frac{1}{\phi_2}\left(\frac{U}{x}\right)^{\gamma}\frac{\gamma}{x}}{\left[\left(\frac{U}{x}\right)^{\gamma} - 1\right]} \text{ with } \gamma = e^{\phi_3} \quad . \qquad (3.2.15)$$

When a single sample of N observations is considered, the log-likelihood function can be written as set out below.

$$\begin{aligned}\ln L &= \sum_{i=1}^{N} \ln f(x_i, \boldsymbol{\theta}) \\ &= N \ln \frac{1}{\sqrt{2\pi}} - \sum_{i=1}^{N} \left\{ \frac{1}{2} t(x_i, \boldsymbol{\theta})^2 - \ln \left| \frac{\partial\, t(x_i, \boldsymbol{\theta})}{\partial\, x_i} \right| \right\}\end{aligned} \quad . \qquad (3.2.16)$$

Let $F = -\frac{1}{N} \ln L + \ln \frac{1}{\sqrt{2\pi}}$. We may minimise

$$F(\phi_1, \phi_2, \phi_3) = \frac{1}{N} \sum_{i=1}^{N} \left\{ \frac{1}{2} t(x_i, \boldsymbol{\theta})^2 - \ln \left| t'(x_i, \boldsymbol{\theta}) \right| \right\} \qquad (3.2.17)$$

with respect to the parameters $\boldsymbol{\theta}' = (\phi_1, \phi_2, \phi_3)$ and with $\gamma = e^{\phi_3}$. The solution provides the maximum likelihood estimates of the parameters.

The transformation function is then

$$\begin{aligned} t(x_i, \boldsymbol{\theta}) &= \frac{y_i - \phi_1}{\phi_2} \text{ with} \\ y_i &= -\ln \left\{ \left(\frac{U}{x_i} \right)^{\gamma} - 1 \right\} \end{aligned} \quad . \qquad (3.2.18)$$

The conditional maximum likelihood equation of ϕ_1 for given γ is

$$\frac{\partial F}{\partial \phi_1} = \frac{1}{N} \sum_{i=1}^{N} \left\{ -\frac{1}{\phi_2} t(x_i, \boldsymbol{\theta}) \right\}. \qquad (3.2.19)$$

Set $\frac{\partial F}{\partial \phi_{1\,|\phi_1 = \hat{\phi}_1}} = 0$ to obtain $\hat{\phi}_1 = g_1(\phi_3) = \frac{1}{N} \sum_{i=1}^{N} y_i = \bar{y}$ (3.2.20)

where $\hat{\phi}_1$ is the conditional maximum likelihood estimate of ϕ_1, given γ.

Similarly, the conditional maximum likelihood equation of ϕ_2 for given γ is

$$\frac{\partial F}{\partial \phi_2} = \frac{1}{N} \sum_{i=1}^{N} \left\{ -\frac{1}{\phi_2} (t^2(x_i, \boldsymbol{\theta}) - 1) \right\}. \qquad (3.2.21)$$

Set $\frac{\partial F}{\partial \phi_{2\,|\phi_2=\hat{\phi}_2}} = 0$ to obtain $\hat{\phi}_2 = g_2(\phi_3) = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})^2}$ (3.2.22)

where $\hat{\phi}_2$ is the conditional maximum likelihood estimate of ϕ_2, given γ.

The maximum likelihood equation for ϕ_3 is

$$\frac{\partial F}{\partial \phi_3} = \frac{1}{N}\sum_{i=1}^{N} - \left\{ \frac{1}{\phi_2} Q\, t(\theta) + \gamma \ln\left(\frac{U}{x}\right) + 1 - Q) \right\}$$

with $Q = \ln\left(\frac{U}{x}\right)\gamma\left(\frac{U}{x}\right)^{\gamma}\left[\left(\frac{U}{x}\right)^{\gamma} - 1\right]^{-1}$ (3.2.23)

and $\gamma = e^{\phi_3}$

In (3.2.20) and (3.2.22), ϕ_1 and ϕ_2 were expressed in terms of $\gamma = e^{\phi_3}$, so that ϕ_1 and ϕ_2 can be regarded as functions of ϕ_3.

Solutions for ϕ_1 and ϕ_2 in (3.2.17) can be obtained explicitly, given ϕ_3, while ϕ_3 can be obtained by means of a successive quadratic interpolation technique, or a one-dimensional search technique with first derivative information.

3.2.5 Extension of the model to include more groups

The previous section dealt with the development of a general model which maps the bounded scores x; $0 \le x \le U$, to unbounded scores z; $-\infty < z < \infty$. A general normalisation function $t(x)$ has been developed and this section demonstrates how the transformation function can be expanded to cater for more than one sample or group. The simplest case is to consider the groups separately and do separate normalisations for each group, so that the normalisation functions each have three parameters.

A group of nested models (see Figure 3.2) is used with Model A as the least restrictive where normalisations are performed separately for each group. Model B is a special case of Model A, where one of the parameters, namely ϕ_3, is restricted to be equal for

all the groups. Model D is a special case of Model B, with two parameters, ϕ_3 and ϕ_2, restricted to be equal for all sub-groups. The number of parameters for models A, B and D depends on the number of sub-groups studied.

Extending from (3.2.12), a general normalisation function for more than one group can be defined for the *j* -th group at the *i* -th total score as

$$t(x_{ij},\phi_{3j}) = \frac{y_{ij} - g_{1j}(\phi_{3j})}{g_{2j}(\phi_{3j})} \qquad (3.2.24)$$

with $y_{ij} = -\ln\left\{\left(\frac{U}{x_{ij}}\right)^{\gamma_j} - 1\right\}$

and $\gamma_j = e^{\phi_{3j}}$

where $g_{1j}(\phi_{3j})$ and $g_{2j}(\phi_{3j})$ are defined in (3.2.27) and (3.2.28).

The partial derivative of (3.2.24) with respect to x is

$$t'(x_{ij},\phi_{3j}) = \frac{\partial\, t(x_{ij},\phi_{3j})}{\partial\, x_{ij}} = \frac{(\frac{U}{x_{ij}})^{\gamma_j}\,\frac{\gamma_j}{x_{ij}}}{g_{2j}(\phi_{3j})[(\frac{U}{x_{ij}})^{\gamma_j} - 1]}\,. \qquad (3.2.25)$$

We may minimise

$$\begin{aligned} F &= \sum_{j=1}^{g} f_j(g_{1j}(\phi_{3j}), g_{2j}(\phi_{3j}), \phi_{3j}) \\ &= \frac{1}{N}\sum_{j=1}^{g}\sum_{i=1}^{h_j}\left\{\frac{1}{2}t(x_{ij},\phi_{3j})^2 - \ln\left|\, t'(x_{ij},\phi_{3j})\right|\right\} \end{aligned} \qquad (3.2.26)$$

where g is the number of independent groups and h_j is the number of observations in the *j* -th group.

The conditional likelihood equations of ϕ_{1j} and ϕ_{2j} for given γ_j are

$$\hat{\phi}_{1j} = g_{1j}(\phi_{3j}) = \frac{1}{h_j}\sum_{i=1}^{h_j} y_{ij} = \bar{y}_j \qquad \text{and} \qquad (3.2.27)$$

$$\hat{\phi}_{2j} = g_{2j}(\phi_{3j}) = \sqrt{\frac{1}{h_j}\sum_{i=1}^{h_j}(y_{ij} - \bar{y}_j)^2}\,. \qquad (3.2.28)$$

The maximum likelihood estimate $\hat{\phi}_{3j}$ can not be written explicitly.

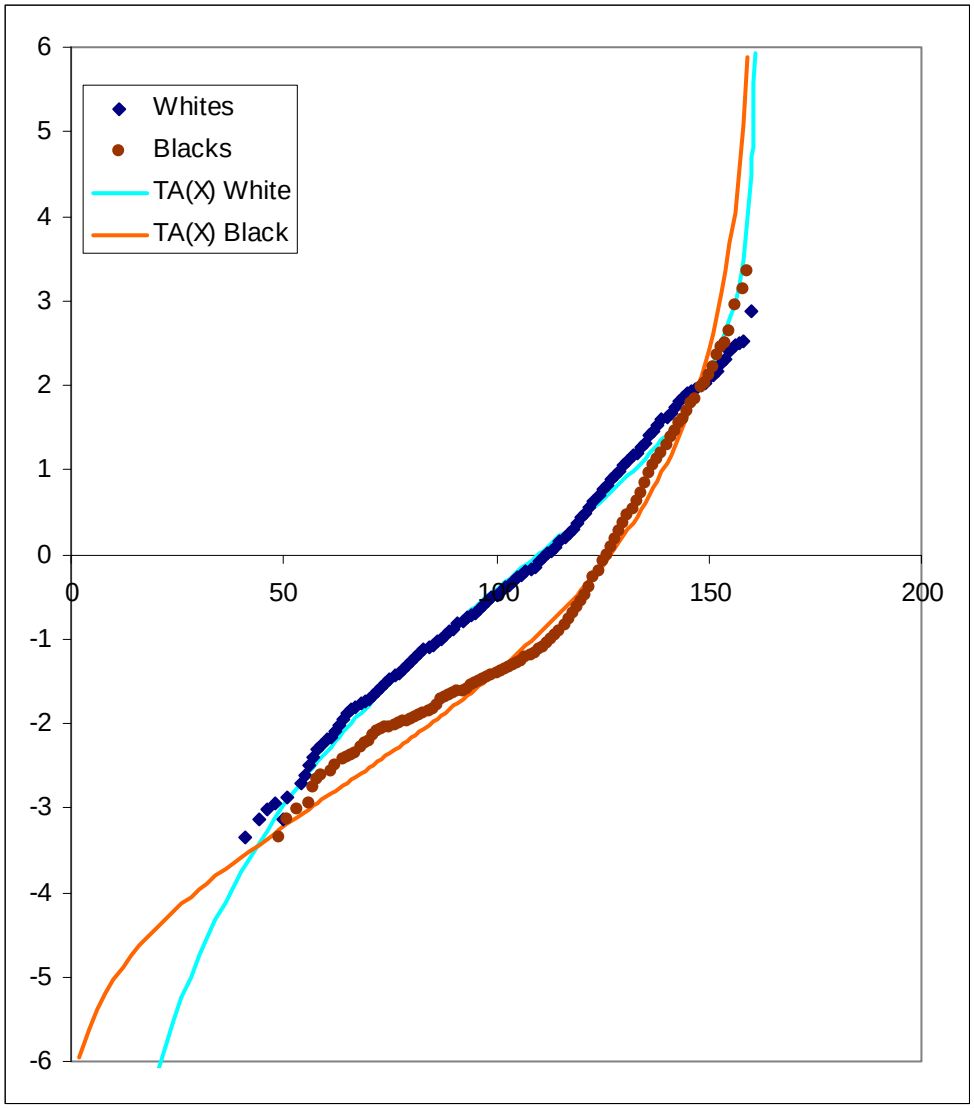
Since the conditional maximum likelihood estimates $\hat{\phi}_{1j}$ and $\hat{\phi}_{2j}$ are functions of ϕ_{3j}, the discrepancy function F in (3.2.26) is reduced to a concentrated discrepancy function of a single parameter ϕ_{3j}. The parameters can therefore also be estimated using a numerical search in one dimension.

3.2.6 An example of SIMNORM for two cultural groups

Model A

Model A is the least restrictive model and is in fact separate normalisations for each group, with U, the theoretical maximum of the total score, fixed for all the groups. The parameters ϕ_{1j}, ϕ_{2j} and ϕ_{3j} are estimated independently for each group, so that there are $p = 3g$ parameters, resulting in six parameters to be estimated. The estimated normalisation functions are depicted in Figure 3.13.

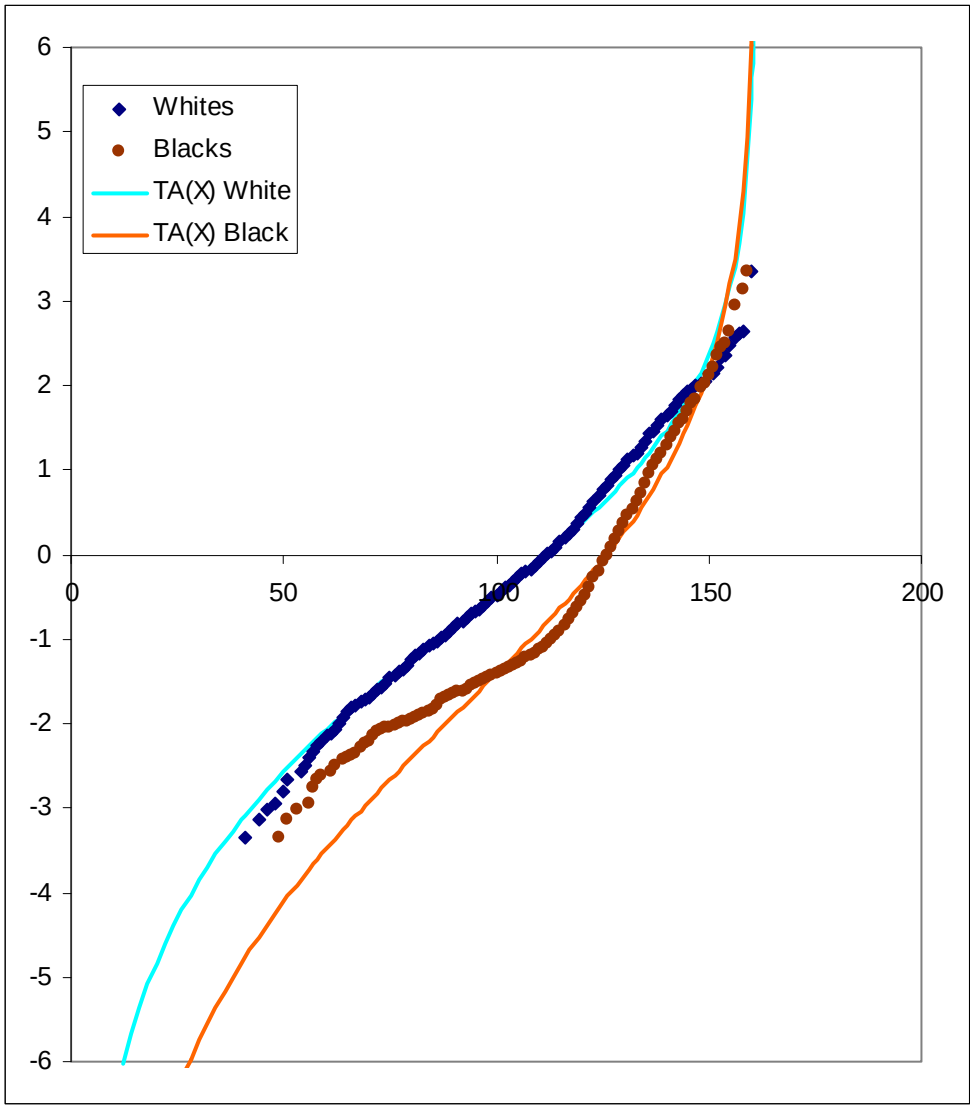
Figure 3.13: The SIMNORM functions using Model A for two cultural groups



Model B

Model B is a special case of Model A, with γ equal for all groups. The number of parameters for Model B is $p = 2g + 1$, resulting in five parameters estimated when two groups are normalised. The estimated simultaneous normalisation functions are given in Figure 3.14.

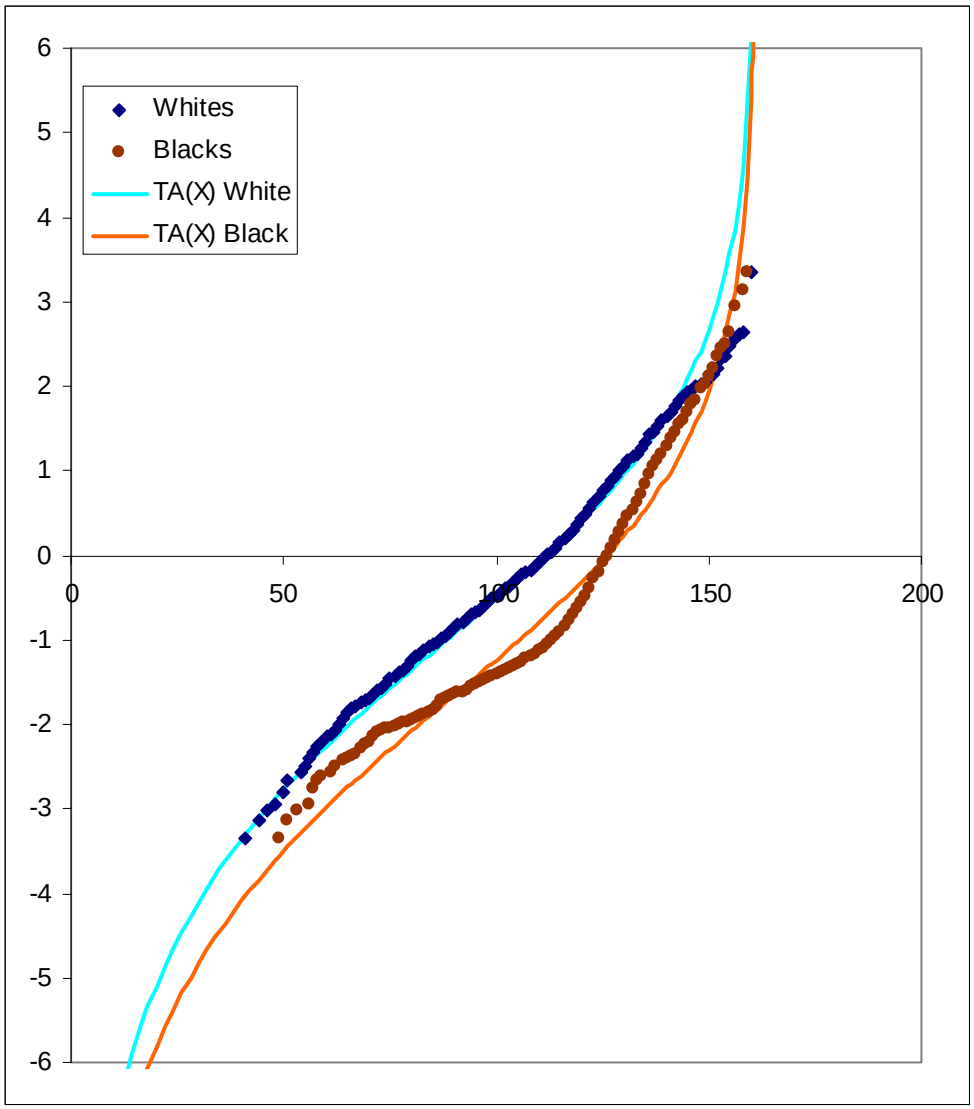
Figure 3.14: The SIMNORM functions using Model B for two cultural groups



Model D

Model D is a special case of Model B, with γ and ϕ_2 restricted to be equal for all groups, so that the number of parameters is $p = g + 2$, which is four parameters in this example. The estimated normalisation functions are plotted in Figure 3.15.

Figure 3.15: The SIMNORM functions using Model D for two cultural groups



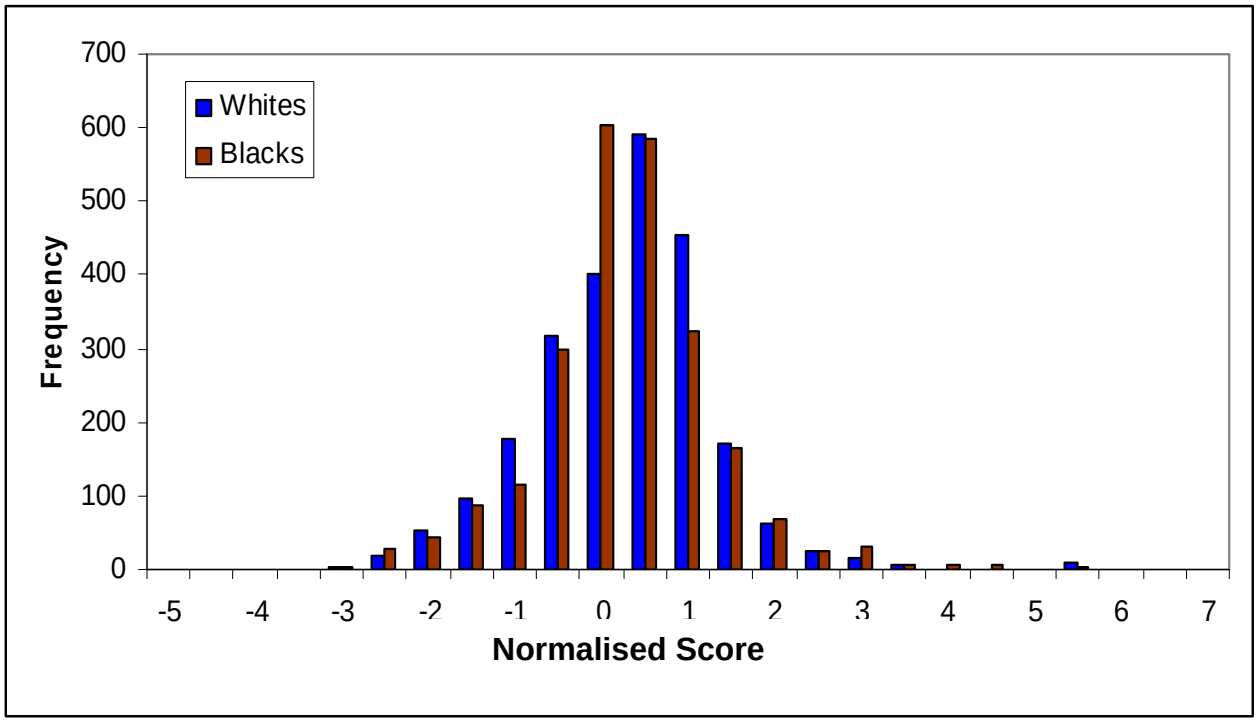
The estimated parameters were used to obtain normalised scores for each of the three models. The descriptive statistics for the normalised scores can be seen in Table 3.4.

Table 3.4: Descriptive statistics of normalised scores for two cultural groups

Statistic	**White group**	**Black group**
Sample size	2400	2400
Mean	0.0000	0.0000
Median	0.0878	0.0111
Mode	0.4258	0.0111
Q1	-0.577	-0.476
Q3	0.601	0.525
Minimum	-3.672	-3.257
Maximum	4.493	5.879
Standard deviation	1.0004	1.0004
Skewness	0.0559	0.3135
Kurtosis	1.5427	2.5008
Range	8.1648	9.1362
Interquartile Range	1.1784	1.0007
Tests for Normality		
Kolmogorov-Smirnov D Pr > D	0.05523 <0.0100	0.077655 <0.0100
Cramer-von-Mises W-Sq Pr > W-Sq	1.746451 <0.0050	4.289025 <0.0050
Anderson-Darling A-Sq Pr > A-Sq	10.53407 <0.0050	24.7193 <0.0050

The Kolmogorov-Smirnov test, the Cramer-von-Mises and the Anderson-Darling tests for normality were applied for the normalised scores. The null hypothesis in each case is that the scores follow a normal distribution. The null hypothesis is rejected for both the white and black groups. However, since these tests could be over-sensitive in the case of large samples, it is necessary also to do a visual inspection of the distributions. The distributions of the total scores that were normalised are shown in Figure 3.16 for the two cultural groups. It is noteworthy how similar the distributions are and that they do closely represent normal distributions.

Figure 3.16: Distribution of normalised total scores for two cultural groups



3.2.7 Parameter estimation

For all the SIMNORM models, a one-dimensional search method is sufficient. The conditional maximum likelihood equations (3.2.20) and (3.2.22) provide conditional solutions for $\hat{\phi}_1$ and $\hat{\phi}_2$. The best estimate for $\hat{\gamma}$ can be obtained using a suitable one-dimensional search technique.

According to Ross (1970), the basic method of nested minimisation is first to differentiate the likelihood with respect to the parameters and obtain the standard normal equations, and then to eliminate from the likelihood function any parameters conveniently expressed in terms of others.

Since it is possible to express the maximum likelihood estimates of ϕ_1 and ϕ_2 as functions of the estimate of ϕ_3, the discrepancy function reduces to a concentrated function of ϕ_3 alone, since $\gamma = e^{\phi_3}$, the estimation requires the search for a solution of ϕ_3, which has no restrictions. The result is that the number of dimensions in which a

numerical search is required is reduced to one. The numerical search involves a recursive use of the minimisation routine.

In the present case, the function F in (3.2.26) is parabolic near to the minimum. It is possible to find three values a, b and c of ϕ_3 so that $a < b < c$ bracket a minimum, $F(b) < F(a)$ and $F(b) < F(c)$, with F defined in (3.2.26). The FORTRAN subroutine MNBRAK (cf. Press, Teukolsky, Vetterling and Flannery, 1996:1201) is applied to find values a, b and c for ϕ_3 bracketing a minimum.

A parabola is fitted through the triplet of points a, b and c and the abscissa x of the minimum of the parabola is found. Then the function (3.2.26) is evaluated at the minimum x and a new bracketing triplet is found. The process is repeated, and converges when the minimum of the parabola is less than a specified distance TOL from a point already evaluated. The FORTRAN function BRENT (Press *et al*, 1996:1204) was applied in the computer program to find the minimum F and the abscissa x, which will be the initial estimate of ϕ_3.

The one-dimensional search technique with first derivatives is a method proposed by Brent (Press *et al*, 1996:1205). It accomplishes the same as the successive quadratic interpolation method described in the paragraph above. It also starts with a triplet of abscissas $a < b < c$ bracketing a minimum, but it utilises an additional capability to compute the function's first derivative and its value. The sign of the derivative at the central point b of the bracketing triplet indicates uniquely whether the next test point should be taken in the interval (a, b) or in the interval (b, c). The FORTRAN function DBRENT is utilised for this purpose.

3.3 SIMULTANEOUS ITEM NORMALISATION (SIMITNORM)

Section 3.2 covered the development of the mathematical model for the simultaneous normalisation of total scores over cultural groups. This section extends the SIMNORM model to the SIMITNORM model, which allows for item-by-item normalisation of test scores over different cultural groups.

Perception and attitude measurements are often observed by administering a questionnaire containing items with statements that pertain to the concepts being measured. Often these scales are ordinal in nature. This section shows how individual ordinal items can be normalised along the same principles as those followed for total score normalisation.

3.3.1 The mathematical model for the normalisation of individual items

Extending from (3.2.12), a general normalisation function for more than one group can be defined for the k-th item of the j-th group at the i-th observed ordinal item score as

$$t(x_{ijk}, \phi_{3jk}) = \frac{y_{ijk} - g_{1jk}(\phi_{3jk})}{g_{2jk}(\phi_{3jk})} \qquad (3.3.1)$$

with $y_{ijk} = -\ln\left\{\left(\frac{U}{x_{ijk}}\right)^{\gamma_{jk}} - 1\right\}$

and $\gamma_{jk} = e^{\phi_{3jk}}$

and U = the maximum score of the ordinal item.

In (3.3.1), $g_{1jk}(\phi_{3jk})$ and $g_{2jk}(\phi_{3jk})$ are extended from (3.2.26) and (3.2.27) to yield the conditional likelihood equations of ϕ_{1j} and ϕ_{2j} for given γ_j, so that

$$\hat{\phi}_{1jk} = g_{1jk}(\phi_{3jk}) = \frac{1}{h_{jk}} \sum_{i=1}^{h_{jk}} y_{ijk} = \bar{y}_{jk} \qquad \text{and} \qquad (3.3.2)$$

$$\hat{\phi}_{2jk} = g_{2jk}(\phi_{3jk}) = \sqrt{\frac{1}{h_{jk}}\sum_{i=1}^{h_{jk}}(y_{ijk} - \bar{y}_{jk})^2}\ . \qquad (3.3.3)$$

The maximum likelihood estimate $\hat{\phi}_{3jk}$ can not be written explicitly.

Parallel with (3.2.24), the partial derivative of (3.2.23) with respect to x is

$$t'(x_{ijk}, \phi_{3jk}) = \frac{\partial\, t(x_{ij}k, \phi_{3jk})}{\partial\, x_{ijk}} = \frac{(\frac{U}{x_{ikj}})^{\gamma_{jk}}\ \frac{\gamma_{jk}}{x_{ijk}}}{g_{2jk}(\phi_{3jk})[(\frac{U}{x_{ijk}})^{\gamma_{jk}} - 1]} \qquad (3.3.4)$$

We may minimise

$$\begin{aligned} F &= \sum_{k=1}^{m}\sum_{j=1}^{g} f_{jk}(g_{1jk}(\phi_{3jk}), g_{2jk}(\phi_{3jk}), \phi_{3jk}) \\ &= \frac{1}{N}\sum_{k=1}^{m}\sum_{j=1}^{g}\sum_{i=1}^{h_{jk}}\left\{\frac{1}{2}t(x_{ijk}, \phi_{3jk})^2 - \ln\left|\, t'(x_{ijk}, \phi_{3jk})\right|\right\} \end{aligned} \qquad (3.3.5)$$

where m is the number of items, g is the number of independent groups and h_{jk} is the number of observations in the j-th group responding to the k-th item.

The conditional likelihood equations of ϕ_{1jk} and ϕ_{2jk} for given γ_{jk} are

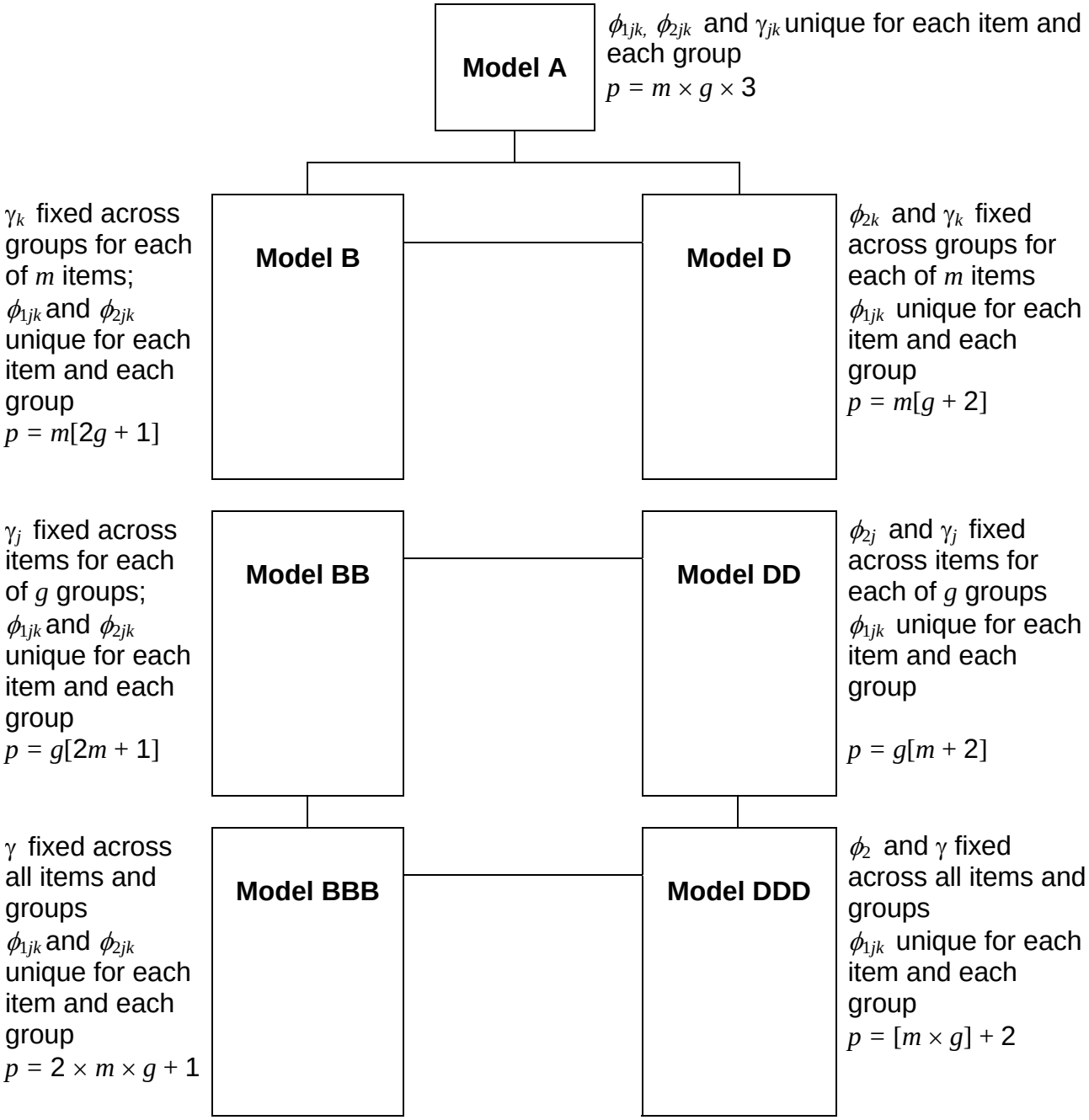
$$\hat{\phi}_{1jk} = g_{1jk}(\phi_{3jk}) = \frac{1}{h_{jk}}\sum_{i=1}^{h_j} y_{ijk} = \bar{y}_{jk} \qquad (3.3.6)$$

$$\hat{\phi}_{2jk} = g_{2jk}(\phi_{3jk}) = \sqrt{\frac{1}{h_{jk}}\sum_{i=1}^{h_{kj}}(y_{ijk} - \bar{y}_{jk})^2} \qquad (3.3.7)$$

The maximum likelihood estimate $\hat{\phi}_{3jk}$ can not be written explicitly.

Since the conditional maximum likelihood estimates $\hat{\phi}_{1j}$ and $\hat{\phi}_{2j}$ are functions of ϕ_{3j}, the discrepancy function F in (3.3.5) reduces to a concentrated discrepancy function of a single parameter ϕ_{3j}. The parameters can also be estimated using a numerical search in one dimension. The hierarchy of nested models is slightly more complex, as can be seen in Figure 3.17.

Figure 3.17: Hierarchy of nested models for SIMITNORM



p = number of parameters
m = number of items
g = number of cultural groups

3.3.2 Meanings associated with item parameters

Simulated data was used to study the effect of changes in parameter values on target density estimators and target transformation functions. The definition of a new term is appropriate at this stage. Since the SIMITNORM model is only considered for items that

are measured on an ordinal response scale, such as the Likert scale, an agreeability or amiability parameter is required. For the application of SIMITNORM, it is assumed that all negatively phrased items within a construct are reverse-coded, so that all items measure in the same direction. Secondly, the item codes should be coded in such a way that a low score is associated with disagreement, and a high score is associated with agreement. When these conditions are met, the parameter γ can be defined as the agreeability or amiability parameter. Amiability or agreeability refers to the degree to which the item will evoke an amicable response from the respondent under similar conditions, or when exposed to the same stimulus. The willingness to agree or give consent to the item statement is reflected in the amiability parameter.

Amiable (amiable, n.d.)
"having or showing pleasant, good-natured personal qualities; affable…"
"friendly; sociable …"
"agreeable; willing to accept the wishes, decision, or suggestion of another or others"

Affable (affable, n.d.)
"Easy to be spoken to or addressed; receiving others kindly and conversing with them in a free and friendly manner; courteous; sociable."

Amicable, Friendly (amicable, n.d.)
"Neither of these words denotes any great warmth of affection, since friendly has by no means the same strength as its noun friendship. It does, however, imply something of real cordiality; while amicable supposes very little more than that the parties referred to are not disposed to quarrel. Hence we speak of amicable relations between two countries, an amicable adjustment of difficulties."

Agreeable (agreeable, n.d.)
"to one's liking; pleasing …"
"willing or ready to agree or consent …"
"suitable; conformable … to…"

A very interesting perspective of the potential of the SIMITNORM models emerges when simulated data is used, and the effect of parameter changes are investigated whilst other parameters are fixed. The effect of changes in the item location parameter, ϕ_1, can clearly be seen in the first columns of Figures 3.18 and 3.19. When the item spread parameter and item amiability (ϕ_2 and γ) parameters are held constant, changes in the location parameters very clearly mainly affects the location of the resulting functions.

The second columns in Figures 3.18 and 3.19 display the effects of keeping the item location parameters and the item amiability parameters constant, and allowing the item spread parameters to vary. Changes in the spread parameter clearly affect mainly the spread of target density functions, with slight changes to the symmetrical aspects of the distributions.

The last column of Figures 3.18 and 3.19 clearly shows the effects if only the item amiability parameters are varied. With higher the parameter values, there is an inclination for the distribution to shift towards the positive side of the graph.

Since both the location and spread parameters are functions of the amiability parameter in the SIMITNORM models, which are also to be expected from a logical perspective, changes in this parameter will inherently also affect the location and the spread of the functions, as is clear in the last column of Figure 3.18.

Figure 3.18: Simulated target density functions for SIMITNORM with different parameter values

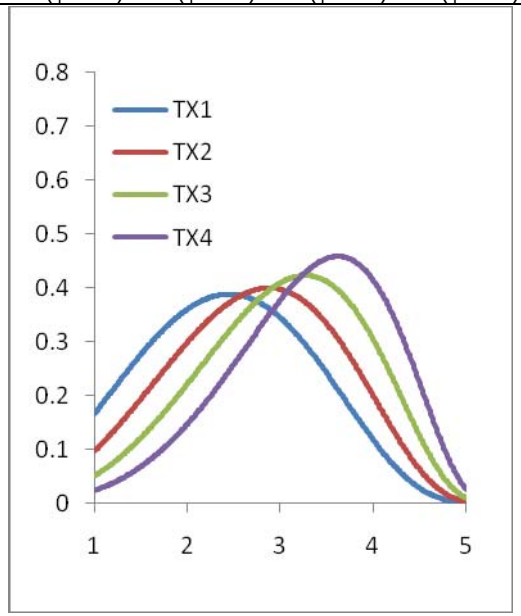
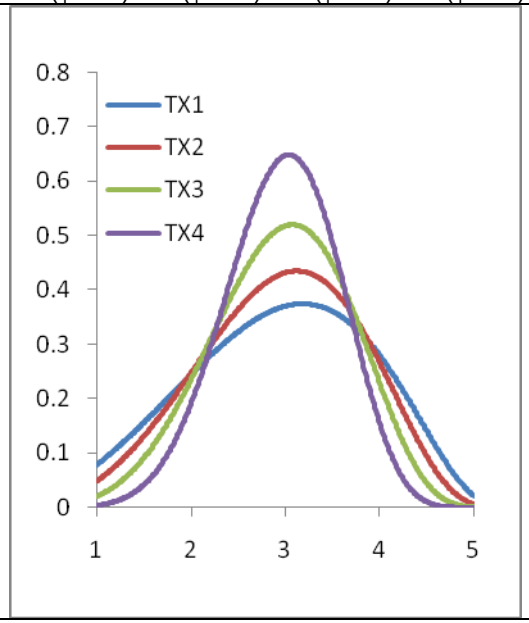
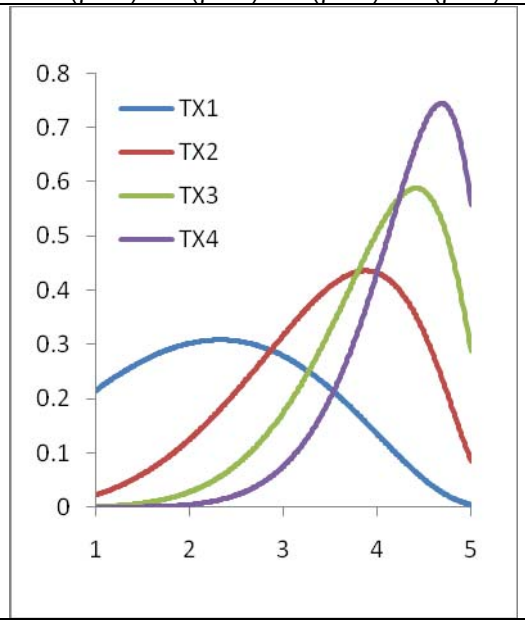
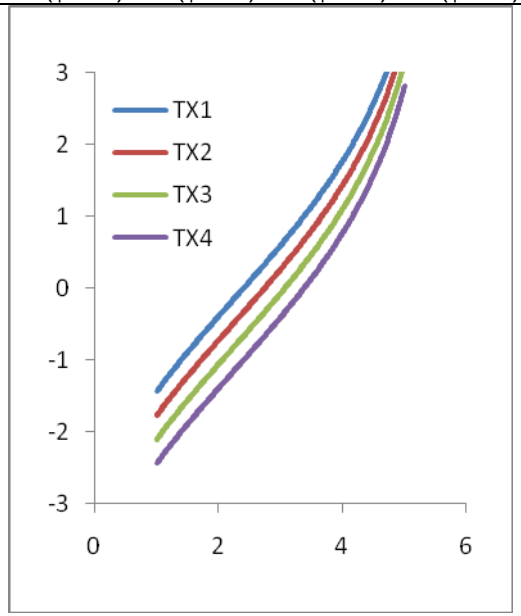
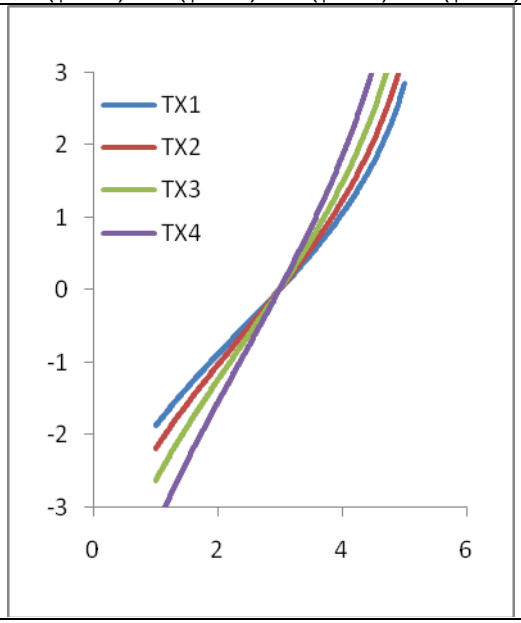
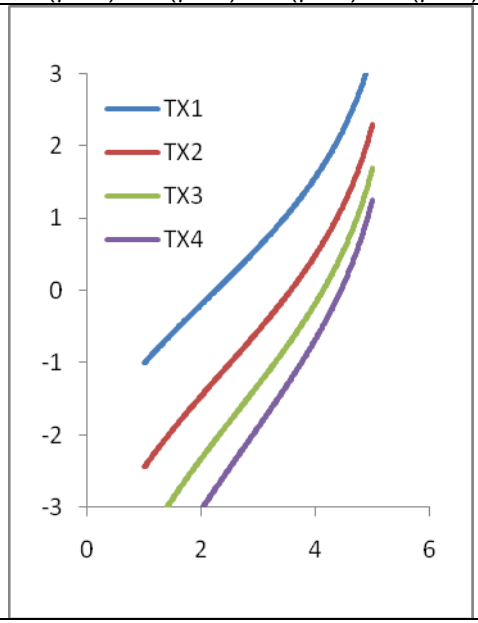
	Effect of varying location parameter	Effect of varying spread parameter	Effect of varying amiability parameter
	Fixed Parameters (ϕ_2=0.6; γ=0.6)	Fixed Parameters (ϕ_1=0.6; γ=0.8)	Fixed Parameters (ϕ_1=1.0; ϕ_2=0.7)
	TX1 (ϕ_1=0.6) TX2 (ϕ_1=0.8) TX3 (ϕ_1=1.0) TX4 (ϕ_1=1.2)	TX1 (ϕ_2=0.7) TX2 (ϕ_2=0.6) TX3 (ϕ_2=0.5) TX4 (ϕ_2=0.4)	TX1 (γ=0.4) TX2 (γ=0.8) TX3 (γ=1.2) TX4 (γ=1.6)
Target density estimation			

Figure 3.19: Simulated target transformation functions for SIMITNORM with different parameter values

	Effect of varying location parameter	Effect of varying spread parameter	Effect of varying amiability parameter
	Fixed Parameters (ϕ_2=0.6; γ=0.6)	Fixed Parameters (ϕ_1=0.6; γ=0.8)	Fixed Parameters (ϕ_1=1.0; ϕ_2=0.7)
	TX1 (ϕ_1=0.6) TX2 (ϕ_1=0.8) TX3 (ϕ_1=1.0) TX4 (ϕ_1=1.2)	TX1 (ϕ_2=0.7) TX2 (ϕ_2=0.6) TX3 (ϕ_2=0.5) TX4 (ϕ_2=0.4)	TX1 (γ=0.4) TX2 (γ=0.8) TX3 (γ=1.2) TX4 (γ=1.6)
Target trans-formation function			

3.3.3 Some examples of the application of SIMITNORM

The item cumulative density distribution functions are depicted in the first row of Figure 3.20, where the densities are given for the Black Afro-centric and White Euro-centric cultural groups for two items, namely items 1 and 2. The Q-Q plots and fitted normalisation functions are also given. The final analyses are discussed in detail when applied across four cultural groups in Chapter 6.

Figure 3.21 gives the observed and normalised item densities for a sample of 4 800 observations for items 1, 2, 15 and 27. The change in the shape of the distributions is clear. It is also clear that the distances between categories of the response scale are not equal after normalisation. Normalisation can therefore also be viewed as a method of transforming an ordinal scale to an interval scale, as was the case with the normalisation approach developed by Stacey (2005).

After the items have been normalised, a total score can be obtained. The sum of normalised items should be a normal distribution. The shapes of the distributions of white and black respondents after standardisation and after normalisation are shown in Figures 3.22 and 3.23. It is clear from Figure 3.23 that the shape of the distribution of the total scores of normalised items approximates a normal distribution.

Figure 3.20: Examples of the estimated item normalisation functions

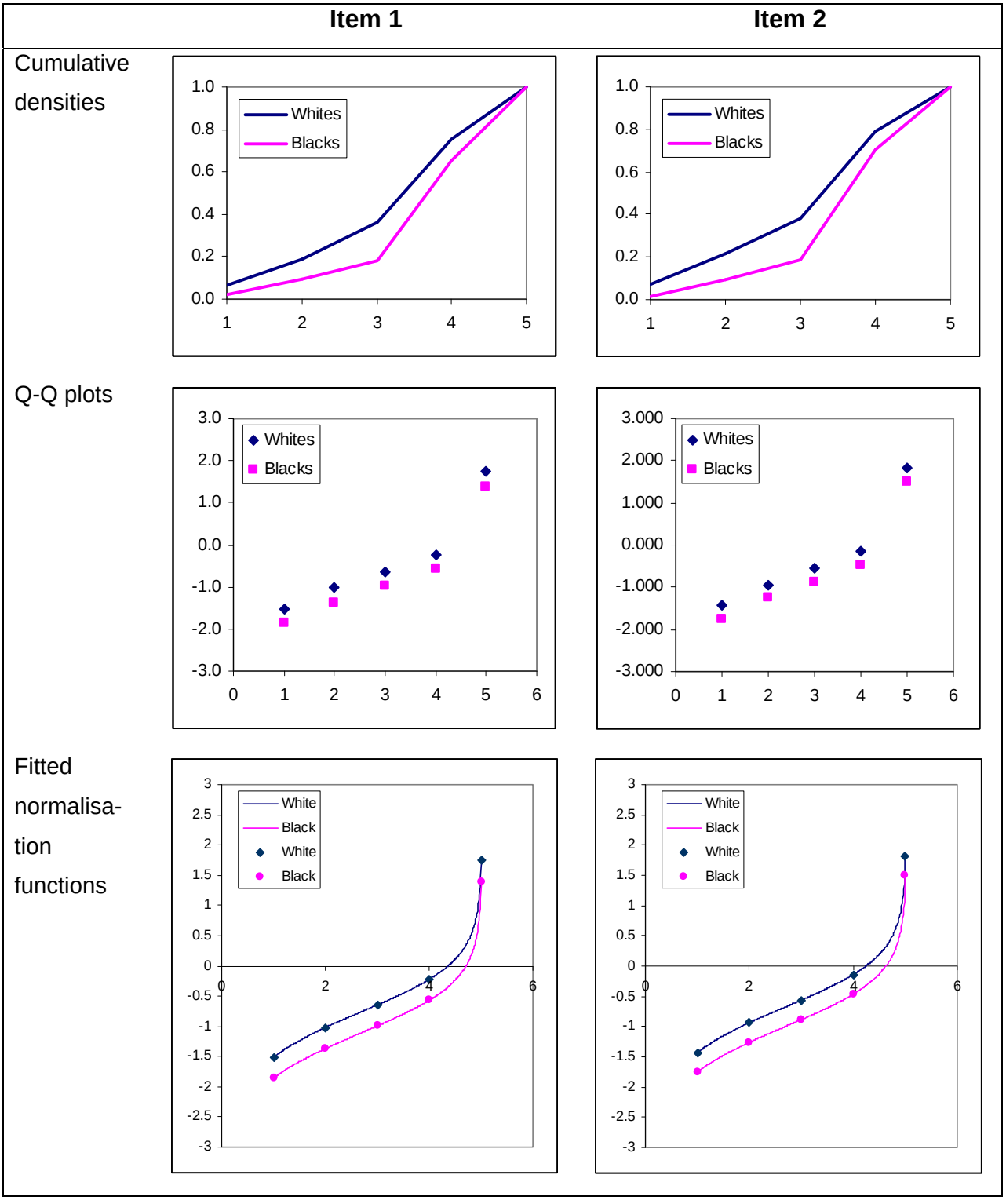


Figure 3.21: Examples of the normalised item density distributions

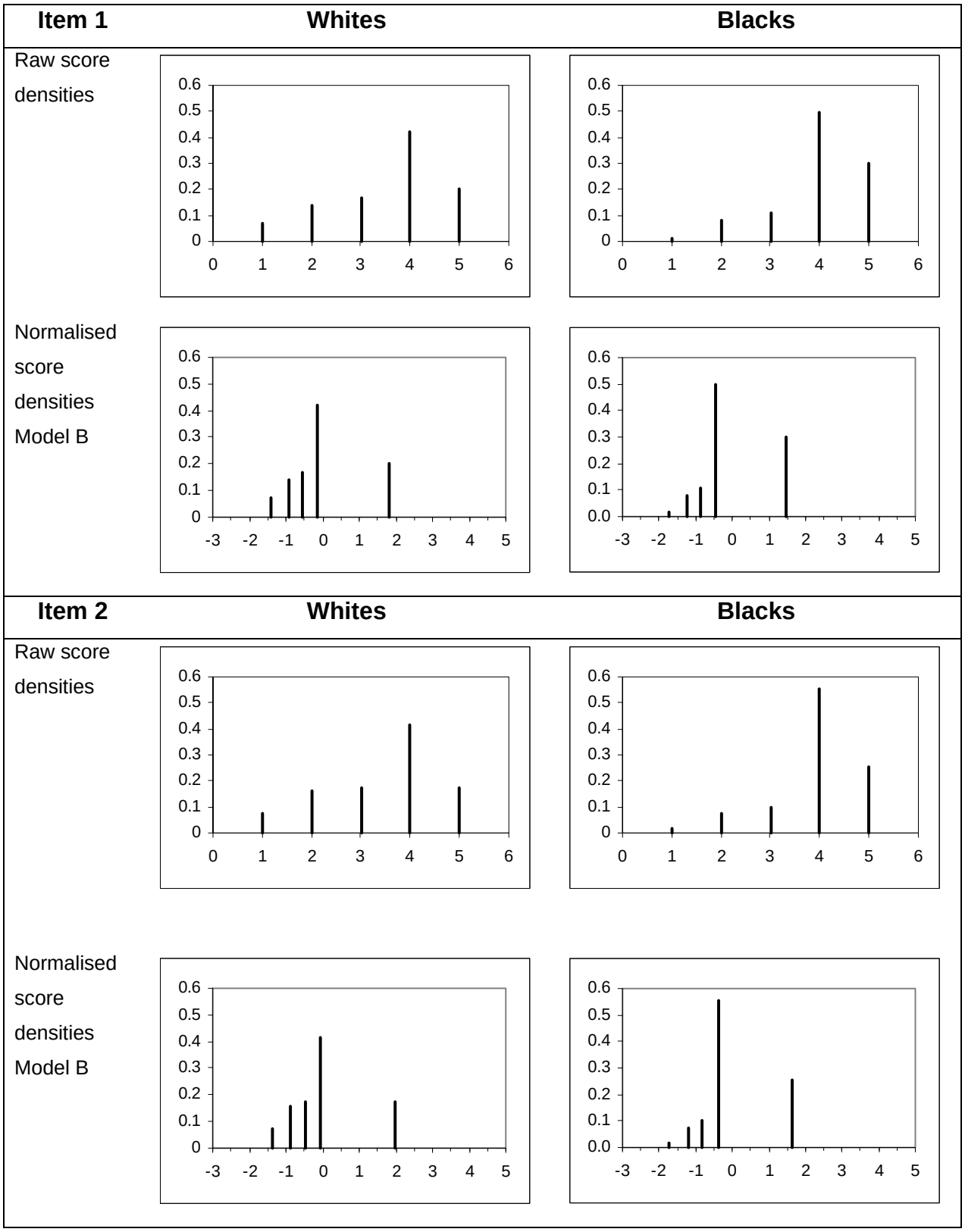


Figure 3.21(continued): Examples of the normalised item density distributions

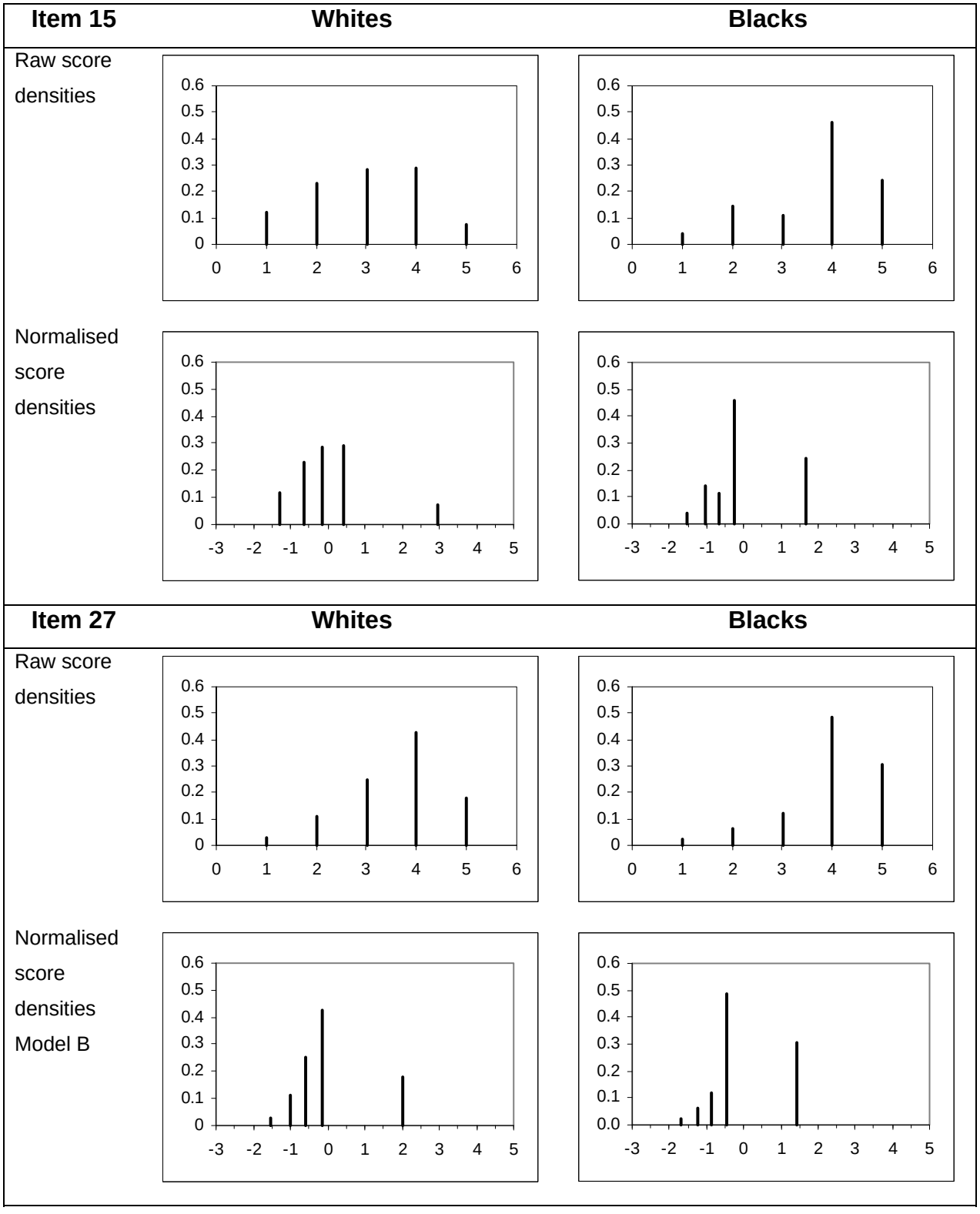


Figure 3.22: The sum of standardised scores for two cultural groups

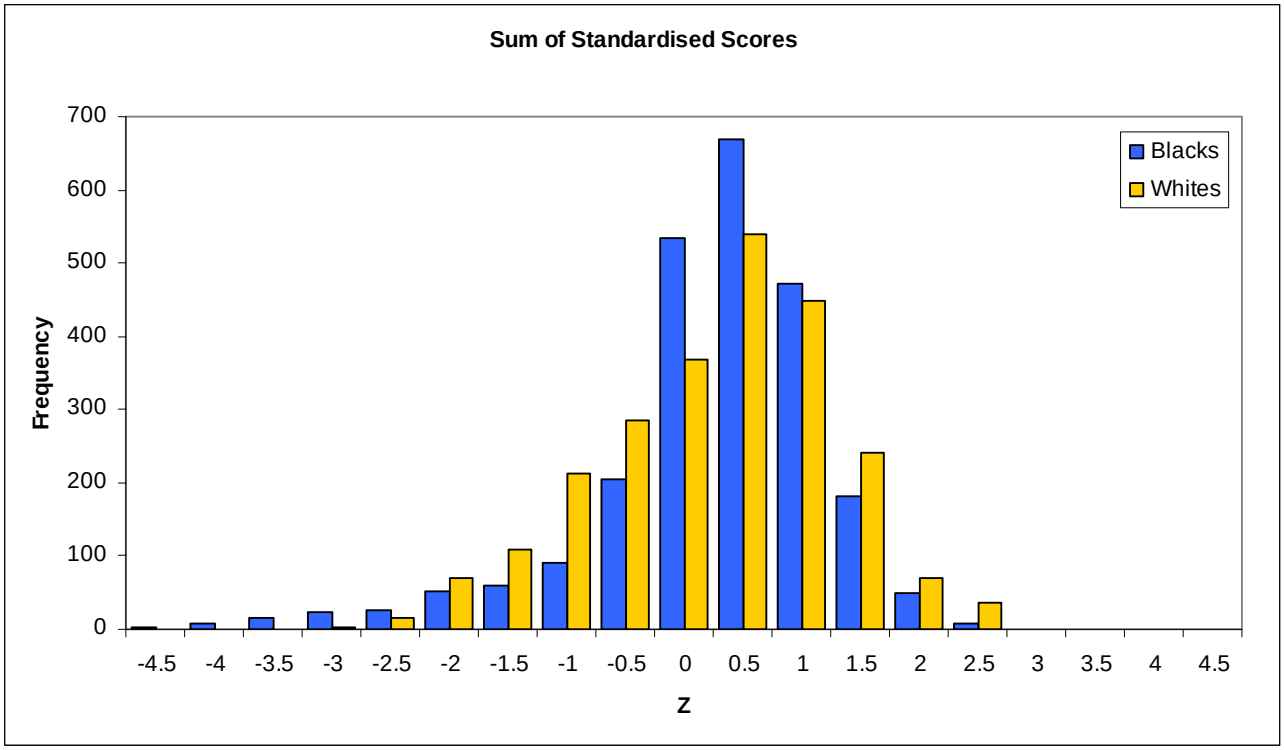
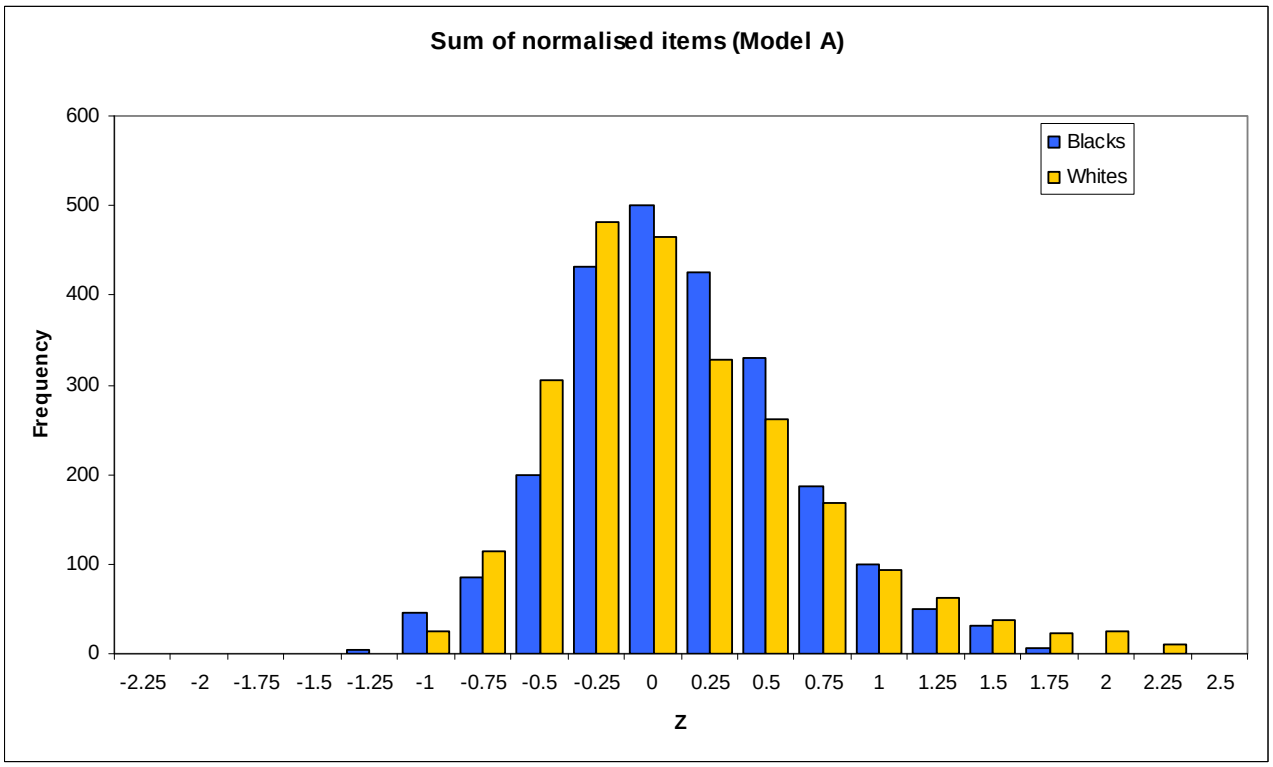


Figure 3.23: The sum of SIMITNORM items for two cultural groups



3.4 PROPOSITIONS IN THIS RESEARCH

Proposition 1:	Since SIMNORM affects scale location, scale dispersion and scale symmetry, it is better than standardisation which only affects scale location and scale dispersion, to remove systematic bias which may present itself as uniform and non-uniform item bias in cross-cultural research.
Proposition 2:	SIMNORM is a non-linear transformation method that has the largest effect on scale location and spread, and a smaller effect on scale symmetry. Standardisation is a linear transformation with the largest effect on scale location and spread, and no effect on scale symmetry. When the results of SIMNORM are compared with the results of standardisation, using level-oriented statistical methods such as ANOVA and regression, very similar results are expected, with minor differences due to the adjustments in the symmetry of the distributions when SIMNORM is used.
Proposition 3:	Simultaneous normalisation of item scores (SIMITNORM) will affect item location, item spread and item symmetry. Item standardisation only affects item location and item spread. It is therefore expected that SIMITNORM will offer a better approach than item standardisation to equate scores over cultural groups, by removing systematic bias that may present itself as uniform and non-uniform bias.
Proposition 4:	SIMITNORM will not have a significant effect on the reliability coefficient Cronbach Alpha when compared to the Cronbach Alpha of the raw items or to the Cronbach Alpha for standardised scores.
Proposition 5:	SIMITNORM will be better than standardisation to ameliorate uniform and non-uniform bias in the total scores and sub-scales of instruments where construct bias is not present, and construct equivalence can be assumed. Interpretation of the results subsequent to SIMITNORM should take the specific method used into account.

Proposition 6:	SIMITNORM used in conjunction with EFA will not affect construct equivalence (the factor structure and factor loadings will remain the same).
Proposition 7:	SIMITNORM used in conjunction with CFA will not affect construct equivalence, metric equivalence or scalar equivalence.
Proposition 8:	SIMITNORM with SEM will result in improved model fit, since the assumption of normally distributed items can be made with more confidence.

3.5 APPLICATION OF SIMITNORM IN THIS RESEARCH

The potential applications of SIMITNORM are expected to be wider and more useful than SIMNORM, since normalisation is achieved at the item level. Several statistical methods applied in cross-cultural research, such as level-orientated statistical analyses and structure-orientated methods such as EFA, CFA and SEM assume that the data (items) have a normal distribution. Also, since the SIMITNORM models allow that certain parameters can be fixed, it allows more rigorous applications of the method.

Structure-orientated methods are based on the covariance matrices of the items, and, since the normalisation functions are non-linear transformations of these scores, the covariances will be affected by the transformation.

The results of SIMNORM and SIMITNORM over four cultural groups are set out in Chapters 5 and 6 of this thesis and the implications of these transformations are also discussed in more detail in these two respective chapters.

CHAPTER 4

RESEARCH METHODOLOGY

4.1 INTRODUCTION TO METHODOLOGY

In order to achieve the purpose of the study and to test the stated propositions, the method proposed in this chapter was followed for the research. For the sake of convenience, the primary research problem and the sub-problems are provided.

4.2 PURPOSE OF THIS RESEARCH

4.2.1 Problem Statement

The purpose of this study is to investigate whether two simultaneous normalisation approaches (SIMNORM and SIMITNORM) can offer improvements over the standardisation approach that has so far been mostly in use (Fischer, 2004), and are suitable as alternative methods in the analysis of cross-cultural research data in which response bias pose a problem.

4.2.2 The sub-problems

The first sub-problem is to examine whether total score simultaneous normalisation (SIMNORM) is better than standardisation to obtain equivalent scores across groups.

The second sub-problem extends the first sub-problem, and is to investigate how statistical results are affected when SIMNORM-scores, standardised scores and observed scores (based on the same set of empirical data) are used to analyse data, using level-orientated data analysis methods such as regression analysis and analysis of variance (ANOVA).

The third sub-problem is to examine whether simultaneous item normalisation (SIMITNORM) is better than standardisation at item level to obtain equivalent item scores across groups.

The fourth sub-problem extends the third, and investigates how the statistical results are affected when SIMITNORM scores, standardised scores and observed scores (based on the same set of empirical data) are subjected to structure-orientated data analysis methods, such as exploratory factor analysis (EFA) and multi-group confirmatory factor analysis (CFA) within in the structural equation modelling (SEM) framework.

4.2.3 Assumptions

The first key assumption is that the instrument being selected to demonstrate the use of SIMNORM and SIMITNORM is a valid and reliable instrument across the cultural groups. It is crucial that this assumption is tested before the research proceeds, since SIMNORM and SIMITNORM are not aimed at correcting problems arising from poorly designed survey instruments.

Construct equivalence is the second key assumption and pre-condition for the use of SIMNORM and SIMITNORM. Construct equivalence, or structural equivalence, implies that the theoretical constructs being measured have comparative meaning across cultures. The use of SIMNORM and SIMITNORM on instruments that do not posses construct equivalence is invalid. It is also crucial that this assumption be tested before the use of SIMINORM or SIMITNORM.

The third key assumption is that bias arising from physical conditions of administration, interviewer effects, differential familiarity with physical administration, and communication problems and other unknown sources of bias, is negligible and can be assumed to be equal across cultural groups.

The fourth key assumption is that item bias due to item wording, formulation and appropriateness of item content is negligible and can be assumed to be equal across cultural groups.

The fifth key assumption is that differences observed between cultural groups can be attributed mainly to cultural differences, or response style differences.

4.3 THE DATA USED IN THIS RESEARCH

4.3.1 The nature of this study

In this research a method is investigated for its effectiveness to deal with uniform and non-uniform bias in a research instrument. This study does not endeavour to explain sociological or anthropological reasons for bias in data. An existing method (SIMNORM) is proposed and a new method (SIMITNORM) is developed, based on a mathematical model, to be tested on a data set with several observations. Existing and commonly applied statistical methods are used to compare raw scores and standardised score results with three models of SIMNORM scores and seven models of SIMITNORM scores. This study therefore explores the applicability of two methods, namely SIMNORM and SIMITNORM to remove systematic bias from data.

This research approach is very commonly used in the development of statistical methods, psychometrics, econometrics and finance, where a unique problem occurs, and a method or model is developed to solve the specific problem.

The SIMNORM and SIMITNORM frameworks are quantitative methods and the analyses are quantitative and statistical, involving a large number of observations.

4.3.2 The COMMAP data

Impact, a marketing and advertising organisation founded by Erik du Plessis, was bought by Millward Brown International, which is regarded as one of the leading market research and organisations, internationally and in South Africa. Erik du Plessis heads the South African division. Since its founding years, Du Plessis has been actively involved in measuring and quantifying various aspects of marketing endeavours. In the production of television commercials more specifically, the response to the commercial was tested using a multi-cultural panel of viewers who watched the commercials and judged them using inter alia the Schlinger (1979) Viewer Response Profile. The commercials were produced for a wide variety of

products and services and for a diverse client group, from August 1989 to December 2001. The period spans twelve years, of which about half of the ads were produced prior to April 1994, and half after April 1994, a landmark in the South African history when Apartheid was officially abolished and all members of the South African nation were able to participate in elections and the Nelson Mandela became the first black president of South Africa, representing the ANC.

The Schlinger (1979) VRP was used by marketing practitioners who developed the COMMAP database. The development of the Schlinger (1979) VRP was described in more detail in Chapter 2.

Analyses presented in Chapter 5 demonstrated the ideal properties of the selected data for purposes of this research, and it showed that the data meets the conditions and assumptions stated for the purpose of this study. The COMMAP database is unique, since it has few missing observations and consists of several replications. Lastly, the data was collected in a multi-cultural context, and several biographic and advertising related variables were included in the database. Consequently, the data set provides an unusual access for academic research.

4.3.3 Background of the COMMAP database

The COMMAP data base consists of more than 40 000 individual responses of various cultural groups. Millward Brown, the advertising agency, produces television commercials for various clients such banks, life insurance institutions, various consumer products and other services. Every commercial is tested by exposing it to different panels of multi-cultural groups consisting of mainly black and white respondents, approximately equally divided between these racial groups. Within each of these two groups, several subgroups exist. The four main cultural groups presented in the study are the White Afrikaans speaking group; White English speaking group; the Black Nguni group (which includes the Zulu, Xhosa, Swazi and Ndebele groups) and the Black Sotho group (which includes the Northern Sotho, Southern Sotho and Tswana groups).

From the COMMAP database, a number of commercials were selected for purposes of this study. The researcher had no information about the commercials, only a single "Job Number" was available to identify different commercials.

4.3.4 The selection of commercials from COMMAP

The selection of commercials was done in the following way. Firstly the data was screened for missing items. It was found that for several commercials, a number of the 32 Schlinger (1979) items were not included in the survey. Only commercials using all 32 items were retained for this study. For some commercials, a relatively high occurrence of randomly missing observations was found. All these commercials were excluded. Lastly, the method of administration was also different for some commercials; some were face-to-face interviews, and some were a group exposure in a theatre room. In the theatre setting, the respondents were shown the commercial on a big cinema-like screen, and then the respondents were requested to complete a self-administered questionnaire using modern technology to capture the responses electronically straight away. Only the theatre method was used as a criterion to be selected from the database. Print and radio commercials were also excluded, so that the selected commercials consisted only of television commercials. After all these criteria of selection, 9737 usable observations remained, consisting of the responses to 50 different commercials.

On closer inspection, it became clear that for some of the 50 commercials, the balance between black or Afro-centric and white or Euro-centric respondents who completed the instrument varied considerably, whilst for others, there were groups of equal sizes. About half of the 50 commercials had exactly equal sizes of black and white respondents. From the data set of 9737 responses, only the ads that each had 100 black and 100 white respondents were used as the sample in this research.

This final selection therefore resulted in 24 independent commercials, each containing 200 independent responses, 100 from whites and 100 from blacks, a total of 4800 responses. The advantage of this approach is that each of the 24

commercials or stimuli was equally represented by the two major cultural groups in the sample, eliminating the need to use weighting in analyses.

4.3.5 Variables available on the COMMAP database

The COMMAP database contains, apart from the 32-item Schlinger (1979) VRP questions that were rated on a five-point Likert scale, also other variables such as biographical information, and a *Likeability* measurement, based on a 10-point scale. Likeability is a well-known and respected predictor for advertising effectiveness, (Biel and Bridgwater, 1990; Walker and Dubitsky, 1994). "...liking measures add substantial value to the assessment and optimization of advertising effectiveness" (Walker and Dubitsky, 1994:9) According to Du Plessis (1994), a number of relationships exist between advertising likeability and the seven dimensions of the Schlinger (1979) VRP. Du Plessis (1994) postulates a dynamic model: "Likeability is not an objective of the creative strategy but an intermediary variable that creates penetration. Likeability is the extent to which the objective of Entertainment, Empathy or Relevant News is achieved without Confusion, Alienation or Familiarity" (Du Plessis, 1994, RC-8).

One of the primary reasons why the COMMAP database is unique and useful for the purposes of this research is the fact that so many variables are available to examine the effect of SIMNORM and SIMITNORM. The composition of the cultural groups that will be used in this study is summarised in Table 4.1. When the responses to the Schlinger (1979) items are considered, the response patterns for English and Afrikaans groups (Euro-centric whites) are fairly similar, and so are the response patterns of the Nguni and Sotho groups (Afro-centric blacks).

Table 4.1: Distribution of cultural groups in the data set used in this study

Cultural group				
Euro-centric white groups		**Afro-centric black groups**		
English	**Afrikaans**	**Nguni**	**Sotho**	**Total**
1207	1193	1386	1014	**4800**
2400		2400		**4800**

4.4 TESTING PROPOSITIONS

4.4.1 Propositions related to SIMNORM

The first and second sub-problems are both related to the simultaneous normalisation (SIMNORM) of aggregated test scores. This study examined the effect of SIMNORM with the aim of obtaining equivalent scores across cultures. The methodologies followed for the assessment of each proposition are summarised in Tables 4.2 and 4.3.

Table 4.2: Assessment of propositions related to SIMNORM

Proposition 1: Since SIMNORM affects scale location, scale dispersion and scale symmetry, it is better than standardisation which only affects scale location and scale dispersion, to remove systematic bias which may present itself as uniform and non-uniform item bias in cross-cultural research.	This proposition will be evaluated by examining the score distributions for the raw scores, standardised scores and the different SIMNORM models.
Proposition 2: SIMNORM is a non-linear trans-formation method that has the largest effect on scale location and spread, and a smaller effect on scale symmetry. Standardisation is a linear trans-formation with the largest effect on scale location and spread, and no effect on scale symmetry. When the results of SIMNORM are compared with the results of standardisation, using level-oriented statistical methods such as ANOVA and regression, very similar results are expected, with minor differences due to the adjustments in the symmetry of the distributions when SIMNORM is used.	This proposition will be evaluated by examining the results of ANOVA and regression when performed on standardised and SIMNORM scores, and will be based on the ANOVA and regression tests of significance.

4.4.2 Propositions related to SIMITNORM

Sub-problems 3 and 4 are related to simultaneous item normalisation across different cultural groups. This study examines the effect of SIMITNORM with standardisation as alternative method of obtaining equivalent item scores across groups.

Table 4.3: Assessment of propositions related to SIMITNORM

Proposition	**Assessment of Proposition**
Proposition 3: Simultaneous normalisation of item scores (SIMITNORM) will affect item location, item spread and item symmetry. Item standardisation only affects item location and item spread. It is therefore expected that SIMITNORM will offer a better approach than item standardisation to equate scores over cultural groups, by removing systematic bias that my present itself as uniform and non-uniform bias.	The key assumption for this proposition is that the measurement instrument can be assumed to have structural equivalence. This proposition will be evaluated by examining the score distributions for the raw scores, standardised scores and the different SIMITNORM models.
Proposition 4: SIMITNORM will not have a significant effect on the reliability coefficient Cronbach Alpha when compared to the Cronbach Alpha of the raw items or to the Cronbach Alpha for standardised scores.	This proposition will be tested by comparing the Cronbach Alpha for raw scores, standardised scores and SIMITNORM scores.
Proposition 5: SIMITNORM will be better than standardisation to ameliorate uniform and non-uniform bias in the total scores and sub-scales of instruments where construct bias is not present, and construct equivalence can be assumed. Interpretation of the results subsequent to SIMITNORM should take the specific method used into account.	This proposition will be evaluated by identifying the type of bias present in the data, and by testing whether scores that have been transformed through the different SIMITNORM models still indicate the presence of bias.

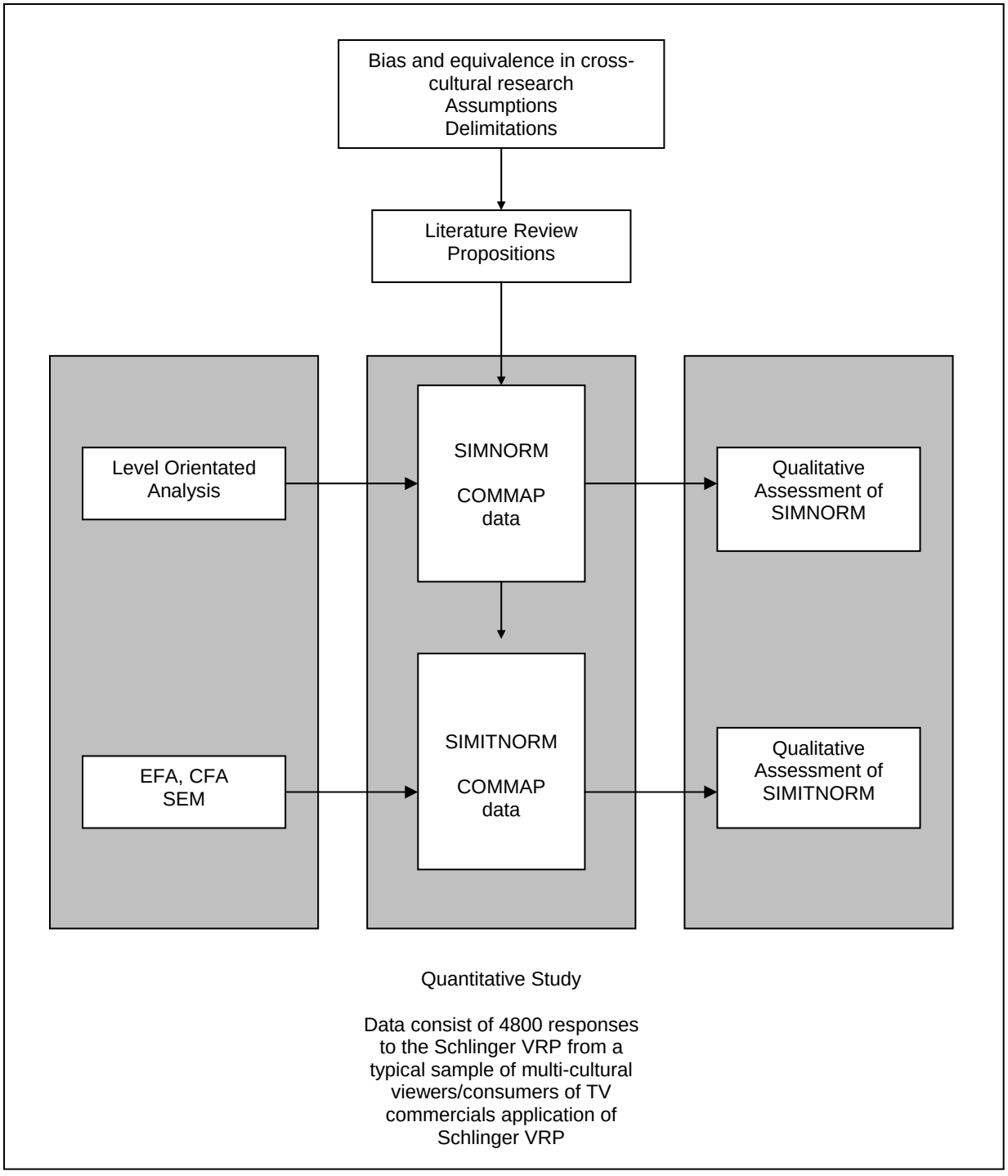
Table 4.3: (continued) Assessment of Propositions related to SIMITNORM

Proposition 6: SIMITNORM used in conjunction with EFA will not affect construct equivalence (the factor structure and factor loadings will remain the same).	This proposition will be evaluated by considering the changes in the factor structure when exploratory factor analysis using VARIMAX rotation is used on raw scores, standardised item scores and SIMITNORM scores.
Proposition 7: SIMITNORM used in conjunction with CFA will not affect construct equivalence, metric equivalence or scalar equivalence.	This proposition will be evaluated by comparing the results when multi-group confirmatory factor analysis is used. The method proposed by Steenkamp and Baumgartner (1998), and Cheung and Rensvold (2000) will be used on raw items, standardised items and SIMITNORM scores.
Proposition 8: SIMITNORM with CFA will result in improved model fit, since the assumption of normally distributed items can be made with more confidence.	Several CFA goodness-of-fit measures, using the guidelines proposed by Fabrigar et al., (1999); Hu and Bentler (1999) and Raykov and Marcoulides (2000).

4.5 METHODOLOGY ROADMAP

The proposed methodology is summarised on the roadmap depicted in Figure 4.1. The consistency matrix is presented in Section 4.6.

Figure 4.1: Methodology Roadmap



4.6 CONSISTENCY MATRIX

Table 4.4: Consistency Matrix

PURPOSE	Sub-Problems	Propositions	Data	Analysis	Assessment of Proposition
Primary Purpose The purpose of this study is to investigate whether two simultaneous normalisation approaches (SIMNORM and SIMITNORM) can offer improvements over the standardisation approach that has so far been mostly in use (Fischer, 2004), and are suitable as alternative methods in the analysis of cross-cultural research data in which response bias pose a problem.	Sub-Problem 1 To examine whether total score simultaneous normalisation (SIMNORM) is better than standardisation to obtain equivalent scores across groups. Sub-Problem 2 This sub-problem extends the first sub-problem, and is to investigate how statistical results are affected when SIMNORM-scores, standardised scores and observed scores (based on the same set of empirical data) are used to analyse data, using level-orientated data analysis methods such as regression analysis and analysis of variance (ANOVA).	Proposition 1: Since SIMNORM affects scale location, scale dispersion and scale symmetry, it will be better than standardisation which only affects scale location and scale dispersion, to remove systematic bias which may present itself as uniform and non-uniform item bias in cross-cultural research. Proposition 2: SIMNORM is a non-linear transformation method that has the largest effect on scale location and spread, and a smaller effect on scale symmetry. Standardisation is a linear transformation with the largest effect on scale location and spread, and no effect on scale symmetry. When the results of SIMNORM are compared with the results of standardisation, using level-oriented methods, very similar results are expected, with minor differences due to the adjustments in the symmetry of the distributions when SIMNORM is used.	Schlinger VRP, 4800 observations, four cultural groups	Application of SIMNORM over four cultural groups, by estimating parameters to obtain SIMNORM scores. Using Raw Scores, Standardised scores and SIMNORM scores for subscales and total scores. ANOVA and multiple regression analyses and comparison of statistical results across raw scores, standardised scores and SIMNORM scores. Fischer (2004); Van de Vijver and Leung (1997); Strasheim (1989)	Evaluation and comparison of outcomes of ANOVA and multiple regression on raw scores, standardised scores and SIMNORM scores, based on significance of results.

Table 4.4 (continued): Consistency Matrix

PURPOSE	Sub-Problems	Propositions	Data	Analysis	Assessment of Proposition
Primary Purpose The purpose of this study is to investigate whether two simultaneous normalisation approaches (SIMNORM and SIMITNORM) can offer improvements over the standardisation approach that has so far been mostly in use (Fischer, 2004), and are suitable as alternative methods in the analysis of cross-cultural research data in which response bias pose a problem.	The third sub-problem is to examine whether simultaneous item normalisation (SIMITNORM) is better than standardisation at item level to obtain equivalent items scores across groups. The fourth sub-problem extends the third, and investigates how the statistical results are affected when SIMITNORM scores, standardised scores and observed scores (based on the same set of empirical data) are subjected to structure-orientated data analysis methods, such as exploratory factor analysis (EFA) and multi-group confirmatory factor analysis (CFA) within in the structural equation modelling (SEM) framework.	Proposition 3: Simultaneous normalisation of item scores (SIMITNORM) will affect item location, item spread and item symmetry. Item standardisation only affects item location and item spread. It is therefore expected that SIMITNORM will offer a better approach than item standardisation to equate scores over cultural groups, by removing systematic bias that my present itself as uniform and non-uniform bias.	Schlinger VRP, 4800 observations, four cultural groups	ANOVA method for bias assessment as described in Chapter 2. Calculation of standardised scores, SIMITNORM scores and calculation of descriptive measures. Examination of score distributions and descriptive measures. Test the assumption of normally distributed scores. Fischer (2004); Van de Vijver and Leung (1997)	Interpretation of ANOVA method and comparison of descriptive measures for raw scores, standardised scores and SIMITNORM scores. Interpreting tests for normality.
		Proposition 4: SIMITNORM will not have a significant effect on the reliability coefficient Cronbach Alpha when compared to the Cronbach Alpha of the raw items or to the Cronbach Alpha for standardised scores.	Schlinger VRP, 4800 observations, four cultural groups	Calculation of SIMITNORM scores, and calculation of Cronbach Alpha for raw scores, standardised scores and SIMITNORM scores. Fischer (2004); Van de Vijver and Leung (1997)	Comparison of Cronbach Alpha coefficients for raw scores, standardised scores and SIMITNORM scores.

Table 4.4 (continued): Consistency Matrix

PURPOSE	Sub-Problems	Propositions	Data	Analysis	Assessment of Proposition
Primary Purpose The purpose of this study is to investigate whether two simultaneous normalisation approaches (SIMNORM and SIMITNORM) can offer improvements over the standardisation approach that has so far been mostly in use (Fischer, 2004), and are suitable as alternative methods in the analysis of cross-cultural research data in which response bias pose a problem.	Sub-problem 3: To examine whether simultaneous item normalisation (SIMITNORM) is better than standardisation at item level to obtain equivalent items scores across groups. Sub-problem 4: This sub-problem extends the third sub-problem, and investigates how the statistical results are affected when SIMITNORM scores, standardised scores and observed scores (based on the same set of empirical data) are subjected to structure-orientated data analysis methods, such as exploratory factor analysis (EFA) and multi-group confirmatory factor analysis (CFA) within in the structural equation modelling (SEM) framework.	Proposition 5: SIMITNORM will be better than standardisation to ameliorate uniform and non-uniform bias in the total scores and sub-scales of instruments where construct bias is not present, and construct equivalence can be assumed. Interpretation of the results subsequent to SIMITNORM should take the specific method used into account.	Schlinger VRP, 4800 observations, four cultural groups	Calculation of standardised scores, SIMITNORM scores and calculation of descriptive measures. Apply multi-group CFA over raw scores, standardised scores and SIMITNORM scores. Test hypotheses H1, H2a and H2b using multi-group CFA.	Comparison of descriptive measures for raw scores, standardised scores and SIMITNORM scores. If hypotheses H1 and H2a are tenable for raw scores, but H2b is not tenable, it will provide evidence for acquiescence bias in data. If H1, H2a and H2b are tenable for standardised scores and simitnorm scores are tenable, bias will be removed from the data. If the measures of fit are better for the SIMITNORM scores than for the standardised scores, it will provide evidence that the SIMITNORM scores were more effective in removing bias than the standardisation method. Cheung and Rensvold (2000)

Table 4.4 (continued): Consistency Matrix

PURPOSE	Sub-Problems	Propositions	Data	Analysis	Assessment of Proposition
Primary Purpose The purpose of this study is to investigate whether two simultaneous normalisation approaches (SIMNORM and SIMITNORM) can offer improvements over the standardisation approach that has so far been mostly in use (Fischer, 2004), and are suitable as alternative methods in the analysis of cross-cultural research data in which response bias pose a problem.	Sub-problem 3: To examine whether simultaneous item normalisation (SIMITNORM) is better than standardisation at item level to obtain equivalent items scores across groups. Sub-problem 4: This sub-problem extends the third sub-problem, and investigates how the statistical results are affected when SIMITNORM scores, standardised scores and observed scores (based on the same set of empirical data) are subjected to structure-orientated data analysis methods, such as exploratory factor analysis (EFA) and multi-group confirmatory factor analysis (CFA) within in the structural equation modelling (SEM) framework.	Proposition 6: When SIMITNORM scores and standardised scores are subjected to Exploratory Factor Analysis (EFA), it will not have a major effect on construct equivalence (the factor structure and factor loadings will remain intact).	Schlinger VRP, 4800 observations, four cultural groups	Calculation of SIMITNORM scores. Obtain Varimax rotated factor loadings for raw scores, standardised scores and SIMITNORM scores. Fischer (2004); Van de Vijver and Poortinga (1994); Strasheim (1989)	Comparison of factor loadings for raw scores, standardises scores and SIMITNORM scores. When these loadings are similar in magnitude and form similar patterns, proposition 6 will be supported.
		Proposition 7: SIMITNORM used in conjunction with CFA will not affect construct equivalence, metric equivalence or scalar equivalence.	Schlinger VRP, 4800 observations, four cultural groups	Calculation of SIMITNORM scores. Use LISREL to obtain estimates and measures of fit for raw scores, standardised scores and SIMITNORM scores in specified multi-group CFA model.	Use the approach described by Steenkamp and Baumgartner (1998) and Cheung and Rensvold (2000). It is expected that only H1 and H2 will be tenable for raw scores. Standardisation and SIMITNORM is proposed to 'cause' H2b to be tenable. If H1, H2 and H2a are tenable for standardised scores and SIMITNORM scores, proposition 7 will be supported.

Table 4.4 (continued): Consistency Matrix

PURPOSE	Sub-Problems	Propositions	Data	Analysis	Assessment of Proposition
Primary Purpose The purpose of this study is to investigate whether two simultaneous normalisation approaches (SIMNORM and SIMITNORM) can offer improvements over the standardisation approach that has so far been mostly in use (Fischer, 2004), and are suitable as alternative methods in the analysis of cross-cultural research data in which response bias pose a problem.	Sub-problem 3: To examine whether simultaneous item normalisation (SIMITNORM) is better than standardisation at item level to obtain equivalent items scores across groups. Sub-problem 4: This sub-problem extends the third sub-problem, and investigates how the statistical results are affected when SIMITNORM scores, standardised scores and observed scores (based on the same set of empirical data) are subjected to structure-orientated data analysis methods, such as exploratory factor analysis (EFA) and multi-group confirmatory factor analysis (CFA) within in the structural equation modelling (SEM) framework.	Proposition 8: SIMITNORM with SEM will result in improved model fit, since the assumption of normally distributed items can be made with more confidence.	Schlinger VRP, 4800 observations, four cultural groups	Calculation of SIMITNORM scores. Use LISREL to obtain estimates for the MG CGA model and measures of fit for raw scores, standardised scores and SIMITNORM scores.	Comparison of goodness-of-fit statistics of raw scores, standardised scores and SIMITNORM scores, across different models. Interpret the fit measure using the guidelines provided by Hu and Bentler (1999); Fabrigar *et al.*(1999); Raykov and Marcoulides (2000); Schwartz (1978) and Raftery (1995).
				Calculation of SIMITNORM scores. Use LISREL to obtain estimates and measures of fit for raw scores, standardised scores and SIMITNORM scores.	Comparison of goodness-of-fit statistics of raw scores, standardised scores and SIMITNORM scores, across different models. Use the multi-group CFA method described by Bollen, 1989; Byrne, 1998 and Steenkamp and Baumgartner 1998.

CHAPTER 5

SIMNORM RESEARCH RESULTS

5.1 THE COMMAP DATABASE

The focus in this chapter is to present the results for the SIMNORM method, where aggregated scores are normalised. Firstly, it is shown that the COMMAP database is very suitable for the purposes of this research, therefore in Section 5.1.1, the key assumptions made in this research are evaluated. Section 5.1.2 shows the biographic distribution of the selected data set, whilst the descriptive statistics of the research variables are presented in Section 5.1.3. In Section 5.1.4, the ANOVA method described by Van de Vijver and Leung (1997) is used to test for bias and identify whether the bias is uniform or non-uniform across cultural groups.

Sections 5.2 to 5.5 cover the preliminary analyses to test whether the Schlinger (1979) VRP is reliable and valid, and therefore tests the validity of the first key assumption stated in Chapter 1, namely that the instrument being used in cross-cultural research is valid and reliable in measuring the constructs for all the cultural groups studied. According to Leedy and Ormrod 2001, the validity of an instrument refers to how well it measures what it is supposed to be measuring, while reliability refers to the accuracy and consistency with which the instrument produces results.

In Sections 5.2 to 5.5 the psychometric properties of the 32-item Schlinger (1979) Viewer Response Profile (VRP) were evaluated, and it was contrasted with a more shorter 20-item VRP, where redundant or poor performing items were excluded.

5.1.1 The key assumptions

In order to test the suitability of the COMMAP database, it is necessary to evaluate the stated assumptions.

The first key assumption was that the instrument being used in cross-cultural research is valid and reliable in measuring the constructs for all the cultural groups in the database. This assumption is tested in Section 5.2, and the results indicate that the VRP is both reliable and valid, and is therefore suitable for the purposes of this research.

The second key assumption is that construct equivalence is a necessary pre-condition for the use of SIMNORM and SIMITNORM. Construct equivalence, or structural equivalence, implies that the theoretical constructs being measured have comparative meaning across cultures. The validity and reliability of the original 32-item VRP is covered in Sections 5.3 to 5.5. The construct equivalence of the VRP will be tested over the four cultural groups, using the CFA-approach as described by Byrne (1998); Steenkamp and Baumgartner (1998); Leung and Rensvold (2000); as well as Vandenberg and Lance (2000) in Chapter 6.

The third key assumption is that bias arising form physical conditions of administration, interviewer effects, varying familiarity with physical administration, and communication problems and other unknown sources of bias, is negligible and can be assumed to be equal across cultural groups. The fourth key assumption is that item bias due to translation, formulation, and appropriateness of item content is negligible and can be assumed to be equal across cultural groups.

These two assumptions will not be tested. It can be assumed that the physical conditions during the theatre exposure were similar for all cultural groups. The respondents were all seated within the same facility, and instructions were given in English, Afrikaans and the official indigenous languages. The instrument was presented in English only. The language used was simple, and, since all the cultural groups had a reasonable level of education, it can be assumed that translation issues did not pose a problem. The respondents were tested in

Johannesburg, Gauteng, the metropolitan capital of South Africa, and the respondents were all from an economic group that would normally have a television in the household.

The fifth key assumption is that differences observed between cultural groups can mainly be attributed to cultural differences. This assumption will be tested using the ANOVA-approach described by Van de Vijver and Leung, 1997, and the results are presented in Section 5.1.4.

5.1.2 The biographic distribution in the data set

The biographic variables included in the data set include the language group, age of respondents, gender, income group and the geographic area. Other variables that are more suitable as independent variables with regard to emotional response to advertising, are behavioural and attitudinal questions, namely: “Adseen” and “Usage”. Adseen is a nominal variable which indicates whether a person has seen the commercial before; and the variable Usage is a nominal variable which indicates whether or not the respondent uses the advertised product. The distributions of the respondents across cultural groups according to the additional categorical variables are given in Tables 5.1 to 5.9.

Table 5.1: The language distribution in the data set

RACE DISTRIBUTION ACROSS CULTURAL GROUP			
	White	Black	TOTAL
English	1 207	0	1 207
Afrikaans	1 193	0	1 193
Nguni	0	1 386	1 386
Sotho	0	1 104	1 014
TOTAL	2 400	2 400	4 800

From Table 5.1 it is clear that language groups coincided with cultural groups. Table 5.2 to 5.4 indicate that age groups, gender groups and income groups were more or less evenly spread over the four cultural groups. Table 5.5 shows that the distribution of respondents who saw the commercial before was similar across the four cultural groups.

Table 5.2: The age distribution in the data set

AGE DISTRIBUTION ACROSS CULTURAL GROUPS						
	Younger than 16 years	16-24 years	25-34 years	35-49 years	50+ years	TOTAL
English	0	137	401	449	220	1 207
Afrikaans	0	137	439	443	174	1 193
Nguni	0	159	564	506	157	1 386
Sotho	0	102	414	356	142	1 014
TOTAL	0	535	1 818	1 754	693	4 800

Table 5.3: The gender distribution in the data set

GENDER DISTRIBUTION ACROSS CULTURAL GROUPS			
	Male	Female	TOTAL
English	560	647	1 207
Afrikaans	560	633	1 193
Nguni	659	727	1 386
Sotho	500	514	1 014
TOTAL	2 279	2 521	4 800

Table 5.4: The income distribution in the data set

INCOME DISTRIBUTION ACROSS CULTURAL GROUPS							
	A	B	C	D	STUDENT	Not Applicable	TOTAL
English	424	344	352	87	0	0	1207
Afrikaans	389	317	408	79	0	0	1193
Nguni	410	415	386	175	0	0	1386
Sotho	322	308	274	110	0	0	1014
TOTAL	1 545	1 384	1 420	451	0	0	4800

Table 5.5: The distribution according to whether the commercial had been seen before

DISTRIBUTION of ADSEEN ACROSS CULTURAL GROUPS			
ADSEEN	Yes	No	TOTAL
English	657	402	1 207
Afrikaans	660	381	1 193
Nguni	669	545	1 386
Sotho	533	353	1 014
TOTAL	2 519	1 681	4 800

5.1.3 The descriptive statistics of the research variables in the data set

Table 5.6 presents the distribution of respondents who had used the advertised product, or service (the *Usage* variable) for the four cultural groups. There are a number of differences, but this is not a major cause for concern.

Table 5.6: The distribution according to whether the advertised product is being used

DISTRIBUTION of USAGE ACROSS CULTURAL GROUPS					
	USER	OTHER	NON-USER	NA	TOTAL
English	502	493	62	150	1 207
Afrikaans	524	469	50	150	1 193
Nguni	493	525	193	175	1 386
Sotho	363	370	156	125	1 014
TOTAL	1 882	1 857	461	600	4 800

The distribution on the perceived general enjoyment (*Genenjoy*) of the commercial is set out in Table 5.7. The tendency towards acquiescence bias can be seen in the higher response in the "Agree" and "Strongly Agree" categories for the Nguni and Sotho groups.

Table 5.7: The distribution according to the General Enjoyment of the commercials

DISTRIBUTION of *Genenjoy* ACROSS CULTURAL GROUPS						
	Strongly Disagree	Disagree	Neither Agree nor Disagree	Agree	Strongly Agree	TOTAL
English	104	179	303	338	86	1 207
Afrikaans	103	150	243	386	108	1 193
Nguni	10	35	90	621	390	1 386
Sotho	3	9	53	509	280	1 014
TOTAL	220	373	689	1 854	864	4 800

Descriptive statistics of the 32 VRP items and the *Likeability* and *General Enjoyment* (*Genenjoy*) variables for the sample of 4 800 responses are summarised in Table 5.8.

Table 5.8: Descriptive Statistics of the 32 Schlinger (1979) VRP items across four cultural groups

Item	Description of Item	Means					Standard Deviations				
		All	Eng	Afr	Ngu	Sot	All	Eng	Afr	Ngu	Sot
V1	The commercial was lots of fun to watch and listen to.	3.77	3.43	3.66	3.97	4.02	1.07	1.18	1.14	0.93	0.91
V2	I thought it was clever and entertaining.	3.70	3.43	3.48	3.96	3.94	1.07	1.19	1.15	0.90	0.89
V3	The enthusiasm of the commercial is catching – it picks you up.	3.67	3.33	3.37	3.99	4.00	1.08	1.19	1.18	0.89	0.78
V4	The commercial wasn't just selling the product, it was also entertaining me and I appreciated that.	3.53	3.30	3.43	3.70	3.71	1.11	1.19	1.17	1.03	0.98
V5	The characters (or persons) in the commercial capture your attention.	3.98	3.70	3.81	4.18	4.22	1.00	1.12	1.11	0.84	0.74
V6	It is the kind of commercial that keeps running through your mind after you have seen it.	3.49	3.03	3.05	3.89	4.02	1.16	1.19	1.23	0.98	0.80
V7	I just laughed at it – I thought it was very funny and good.	2.55	2.25	2.31	2.86	2.76	1.21	1.11	1.20	1.22	1.19
V8	It was distracting - trying to watch the screen and listen to the words at the same time.	2.00	1.92	1.99	2.10	1.98	0.97	0.93	1.04	1.01	0.88
V9	It required a lot of effort to follow the commercial.	2.01	1.85	1.91	2.17	2.09	1.07	0.97	1.02	1.14	1.08
V10	It was too complex. I was not sure what was going on.	1.84	1.77	1.77	1.91	1.89	0.90	0.89	0.87	0.93	0.89
V11	I was so busy watching the screen, I didn't listen to the words.	2.02	1.97	2.08	2.01	2.01	0.99	1.01	1.08	0.94	0.92
V12	The commercial gave me a new idea.	3.25	2.57	2.77	3.82	3.87	1.23	1.14	1.19	1.00	0.97
V13	The commercial reminded me that I'm dissatisfied with what I am using now and I'm looking for something better.	2.53	2.10	2.20	2.87	2.94	1.15	0.98	1.02	1.18	1.13
V14	I learned something from the commercial that I didn't know before.	3.12	2.55	2.62	3.65	3.67	1.23	1.15	1.15	1.07	1.03
V15	The commercial told me about a new product and I think I'd like to try it.	3.34	2.91	3.03	3.69	3.77	1.18	1.13	1.14	1.14	1.05
V16	During the commercial I thought how that product might be useful to me.	3.47	2.99	3.03	3.92	3.94	1.17	1.16	1.20	0.99	0.90
V17	That's a good brand and I wouldn't hesitate recommending it to others	3.98	3.79	3.93	4.11	4.07	0.89	0.96	0.95	0.81	0.78
V18	I know that the advertised brand is dependable and reliable.	3.94	3.89	3.99	3.97	3.92	0.88	0.89	0.91	0.87	0.82
V19	The commercial was very realistic – that is, true to life.	3.72	3.46	3.61	3.89	3.90	1.02	1.15	1.14	0.86	0.78
V20	I felt that the commercial was acting out what I feel at times.	3.30	2.93	3.03	3.63	3.62	1.12	1.16	1.25	0.94	0.88
V21	I felt as though I were right there in the commercial experiencing the same thing.	3.10	2.74	2.78	3.43	3.47	1.17	1.16	1.26	1.03	0.96
V22	That's my idea – the kind of life the commercial showed.	3.27	2.88	3.03	3.60	3.58	1.16	1.22	1.32	0.97	0.90
V23	I liked the commercial because it was personal and intimate.	3.33	2.99	3.06	3.62	3.65	1.15	1.19	1.26	1.03	0.92
V24	This kind of commercial has been done many times. It is the same old thing.	2.01	2.19	2.08	1.87	1.91	0.91	1.00	0.98	0.84	0.78
V25	I have seen this commercial so many times – I am tired of it.	1.77	1.96	1.83	1.65	1.66	0.78	0.91	0.87	0.65	0.58
V26	I think that this is an unusual commercial, I am not sure I've seen another one like it.	3.59	3.33	3.42	3.77	3.83	1.07	1.07	1.10	1.03	1.00
V27	What they showed didn't demonstrate the claims they were making about the product.	2.20	2.47	2.29	2.05	1.95	0.99	1.02	0.98	0.98	0.88
V28	The ad didn't have anything to do with me or my needs.	2.08	2.47	2.24	1.82	1.77	1.02	1.12	1.05	0.88	0.80
V29	The commercial did not show me anything that would make me want to use one of their products.	2.16	2.57	2.43	1.82	1.79	1.10	1.16	1.19	0.93	0.86
V30	The commercial made exaggerated claims. The product would not live up to what they said or implied.	2.02	2.22	2.11	1.90	1.84	0.93	0.92	0.94	0.92	0.89
V31	It was an unrealistic ad – very far-fetched.	1.91	2.11	2.04	1.77	1.73	0.97	1.02	1.06	0.88	0.85
V32	The commercial irritated me – it was annoying.	1.75	2.03	1.96	1.52	1.47	0.96	1.07	1.11	0.74	0.69
	Likeability	7.29	6.77	6.97	7.65	7.80	2.05	2.21	2.11	1.88	1.79
	General Enjoyment	3.69	3.12	3.25	4.17	4.23	1.08	1.12	1.15	0.77	0.64

5.2 ITEM BIAS ANALYSIS USING ANOVA

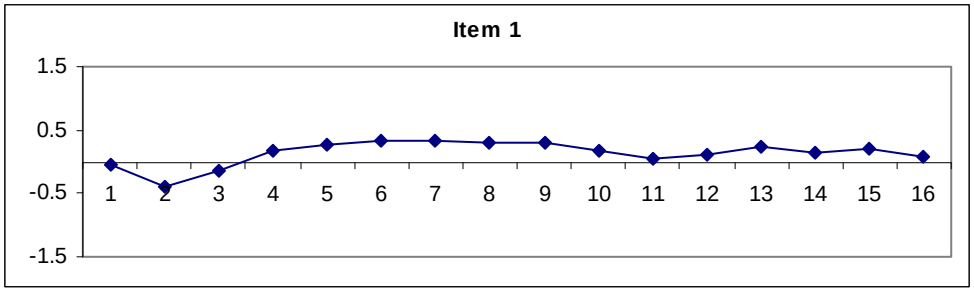
Using the approach given in Section 2.2.4 as described by Van de Vijver and Leung (1997), analyses of variance were conducted using 16 score levels, each approximately equally represented by the two main population groups, namely blacks and whites. This method becomes very cumbersome when four groups are involved, since six pair-wise comparisons should be done. Since the major differences are between the Euro-centric and the Afro-centric groups, it is more effective in the interpretation of the results to consider only two groups. For the results reported in Table 5.9, the total score was calculated across 32 items, with all negative items reversed. The dependent variables in the ANOVA are the item score, the predictors in the model are the score level (Level), the population group (Popgroup) and the interaction between the score level and the cultural group (Popgroup*Level). The results for each item are given in Table 5.9, with remarks about the significance of the tested effects, which shows that almost all the items show bias, and that the bias is mostly non-uniform in nature. This demonstrates once again to the suitability of the data set for the purposes of this study.

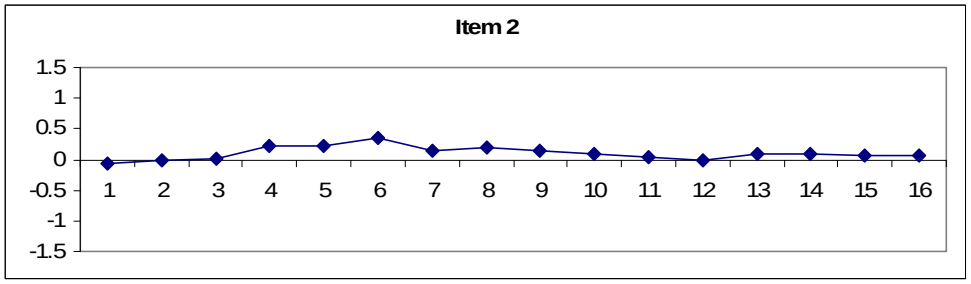
Figure 5.1 shows how the differences between the item means of the Afro-centric versus the Euro-centric groups across each score group plot against the score group. Most of these plots visually confirm the presence of bias, of which non-uniform bias is indicated for most items (27 of the 32 items show non-uniform bias).

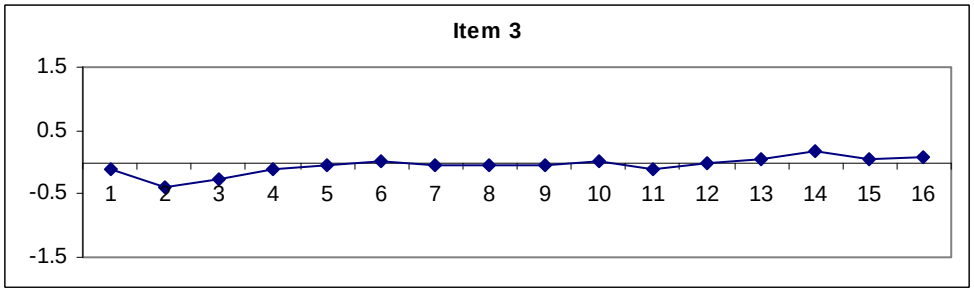
Table 5.9: Significance Levels of ANOVA analysis by score level for each of the 32 items across two population groups

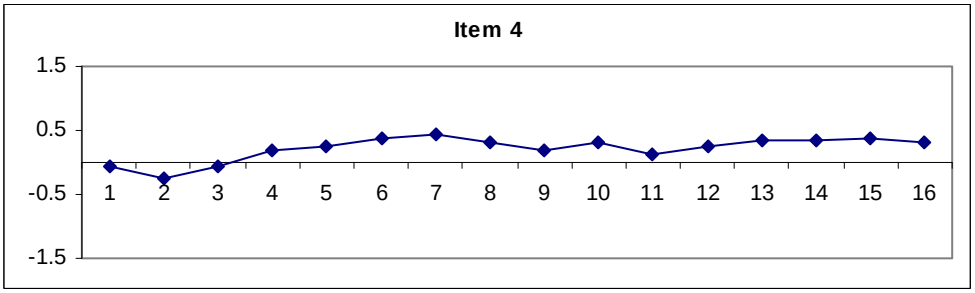
	LEVEL	POPGROUP	POPGROUP*LEVEL	Remark using (α = 0.10)
V1	<0.0001	<0.0001	<0.0001	Non-uniform bias
V2	<0.0001	<0.0001	<0.0001	Non-uniform bias
V3	<0.0001	<0.0001	<0.0001	Non-uniform bias
V4	<0.0001	<0.0001	<0.0001	Non-uniform bias
V5	<0.0001	<0.0001	<0.0001	Non-uniform bias
V6	<0.0001	<0.0001	<0.0001	Non-uniform bias
V7	<0.0001	<0.0001	<0.0001	Non-uniform bias
V8	<0.0001	<0.0001	0.0066	Non-uniform bias
V9	<0.0001	<0.0001	<0.0001	Non-uniform bias
V10	<0.0001	<0.0001	<0.0001	Non-uniform bias
V11	<0.0001	<0.0001	0.0284	Non-uniform bias
V12	<0.0001	<0.0001	<0.0001	Non-uniform bias
V13	0.1643	0.4645	<0.0001	Non-discriminating Item
V14	<0.0001	<0.0001	0.3576	Uniform bias
V15	<0.0001	<0.0001	0.0788	Uniform bias
V16	<0.0001	<0.0001	<0.0001	Non-uniform bias
V17	<0.0001	0.7303	0.0140	Non-uniform bias
V18	<0.0001	<0.0001	0.3603	Uniform bias
V19	<0.0001	0.0003	<0.0001	Non-uniform bias
V20	<0.0001	0.0010	<0.0001	Non-uniform bias
V21	<0.0001	<0.0001	<0.0001	Non-uniform bias
V22	<0.0001	<0.0001	<0.0001	Non-uniform bias
V23	<0.0001	<0.0001	<0.0001	Non-uniform bias
V24	<0.0001	<0.0001	<0.0001	Non-uniform bias
V25	<0.0001	<0.0001	<0.0001	Non-uniform bias
V26	<0.0001	<0.0001	<0.0001	Non-uniform bias
V27	<0.0001	<0.0001	<0.0001	Non-uniform bias
V28	<0.0001	0.0003	0.0037	Non-uniform bias
V29	<0.0001	<0.0001	<0.0001	Non-uniform bias
V30	<0.0001	<0.0001	0.0002	Non-uniform bias
V31	<0.0001	0.0010	0.8158	Uniform bias
V32	<0.0001	<0.0001	<0.0001	Non-uniform bias

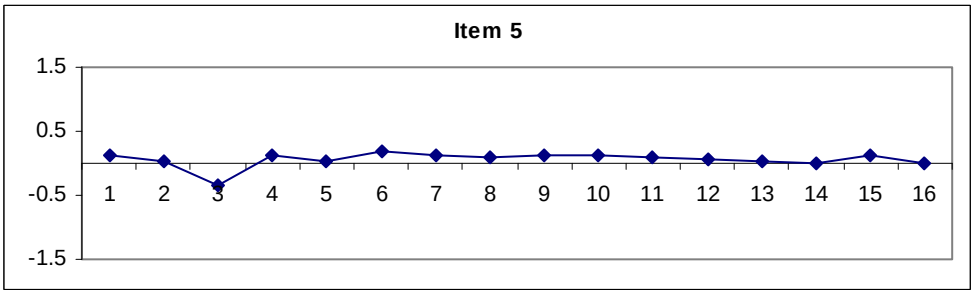
Figure 5.1: Nature of bias in the 32 items of the VRP











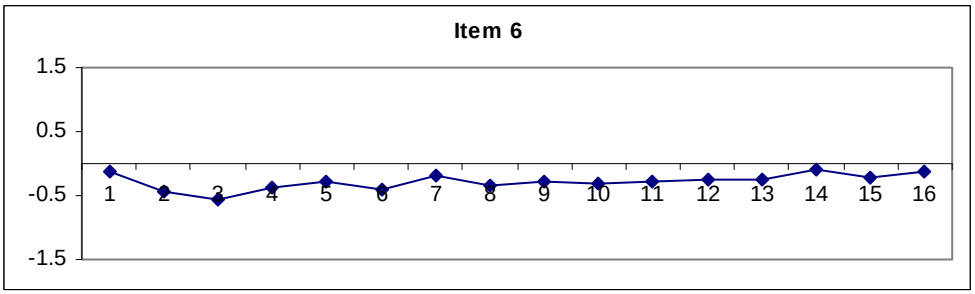


Figure 5.1 (continued): Nature of bias in the 32 items of the VRP

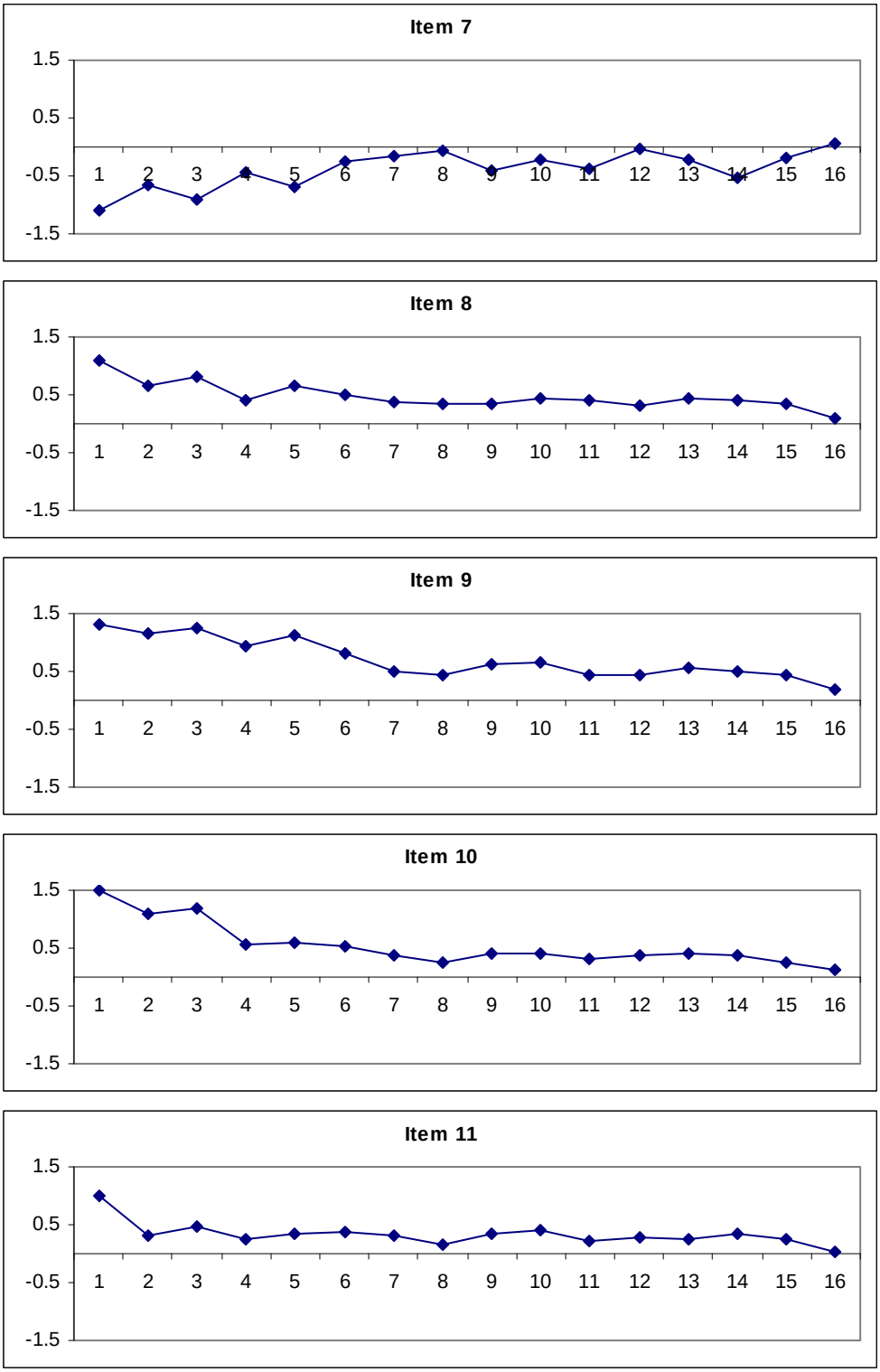


Figure 5.1 (continued): Nature of bias in the 32 items of the VRP

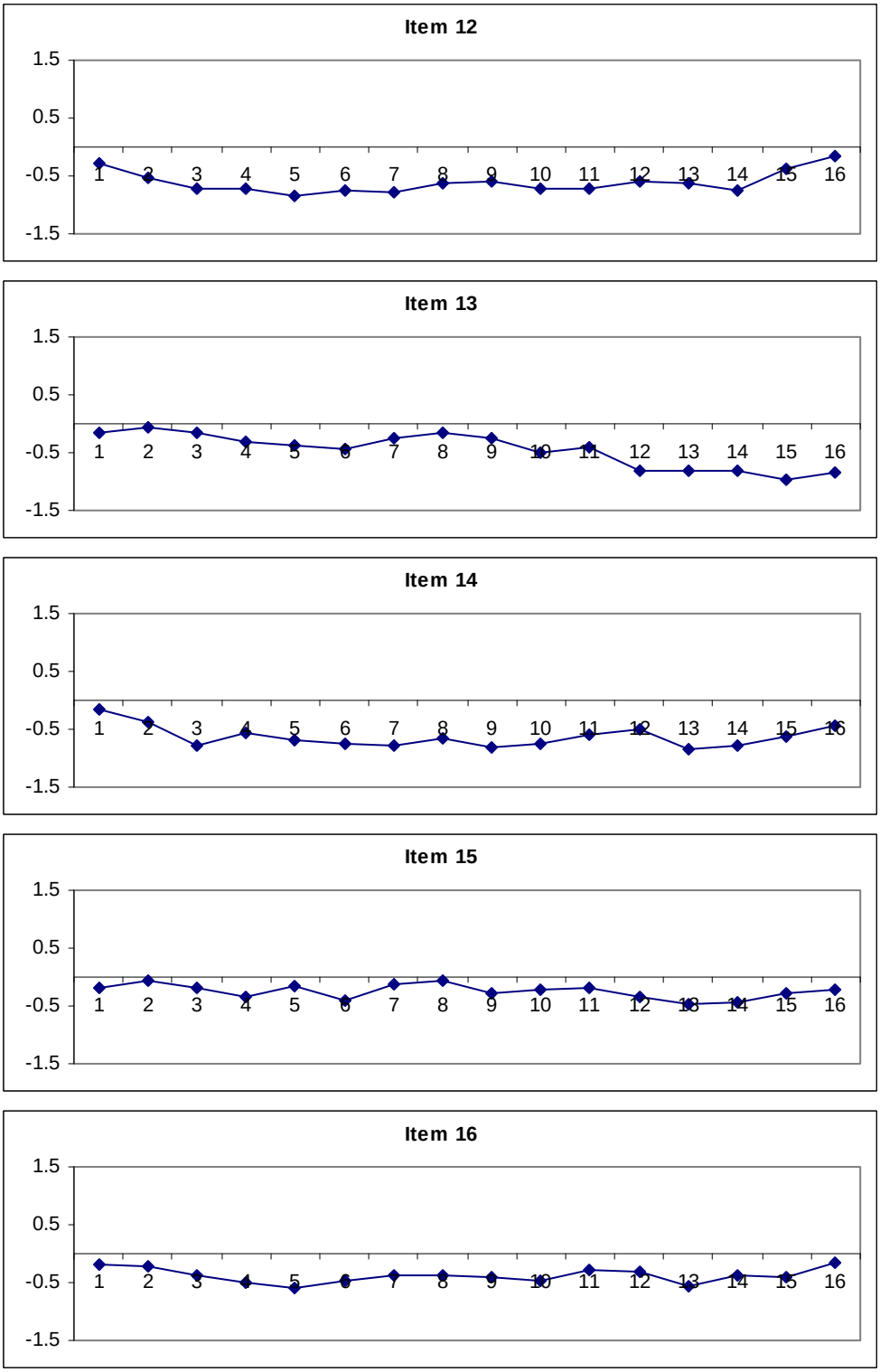
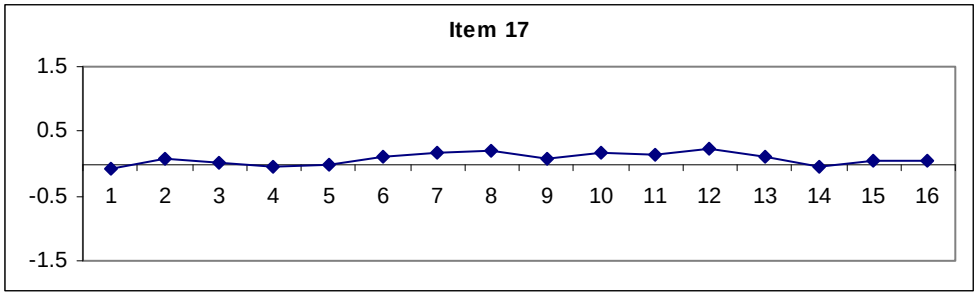
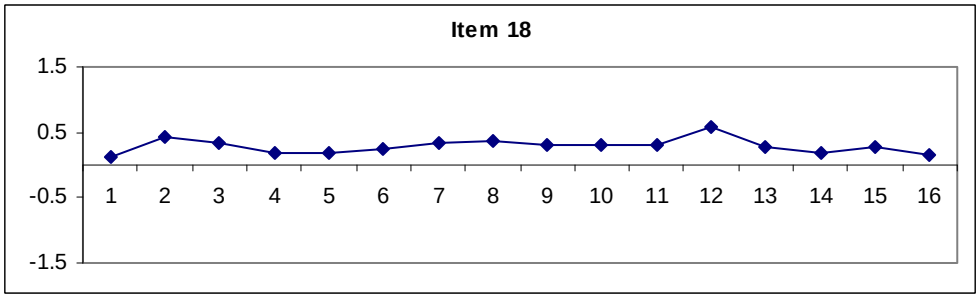
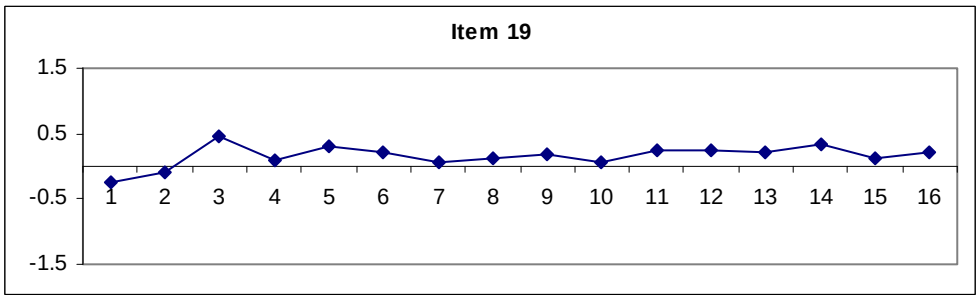
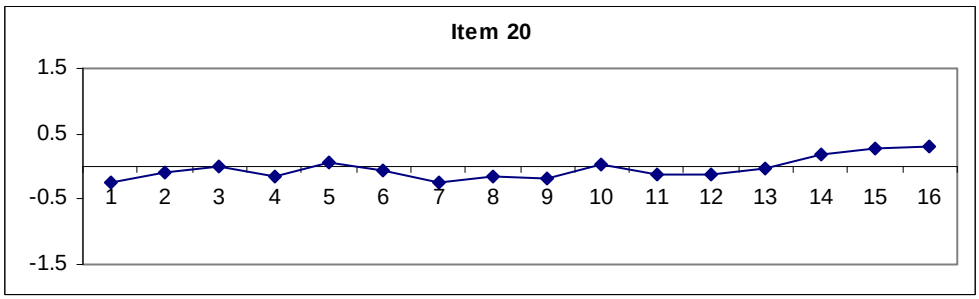


Figure 5.1 (continued): Nature of bias in the 32 items of the VRP









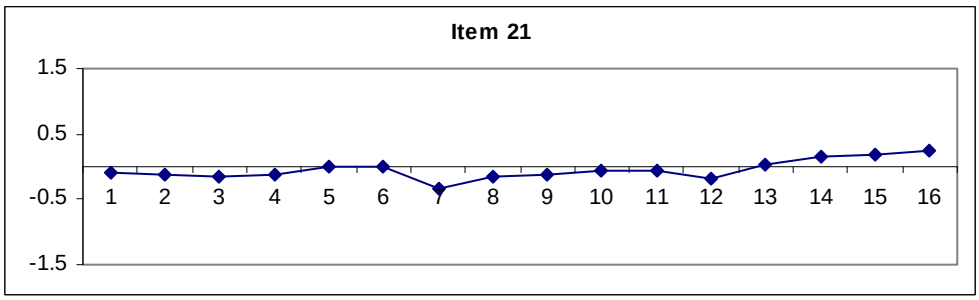


Figure 5.1 (continued): Nature of bias in the 32 items of the VRP

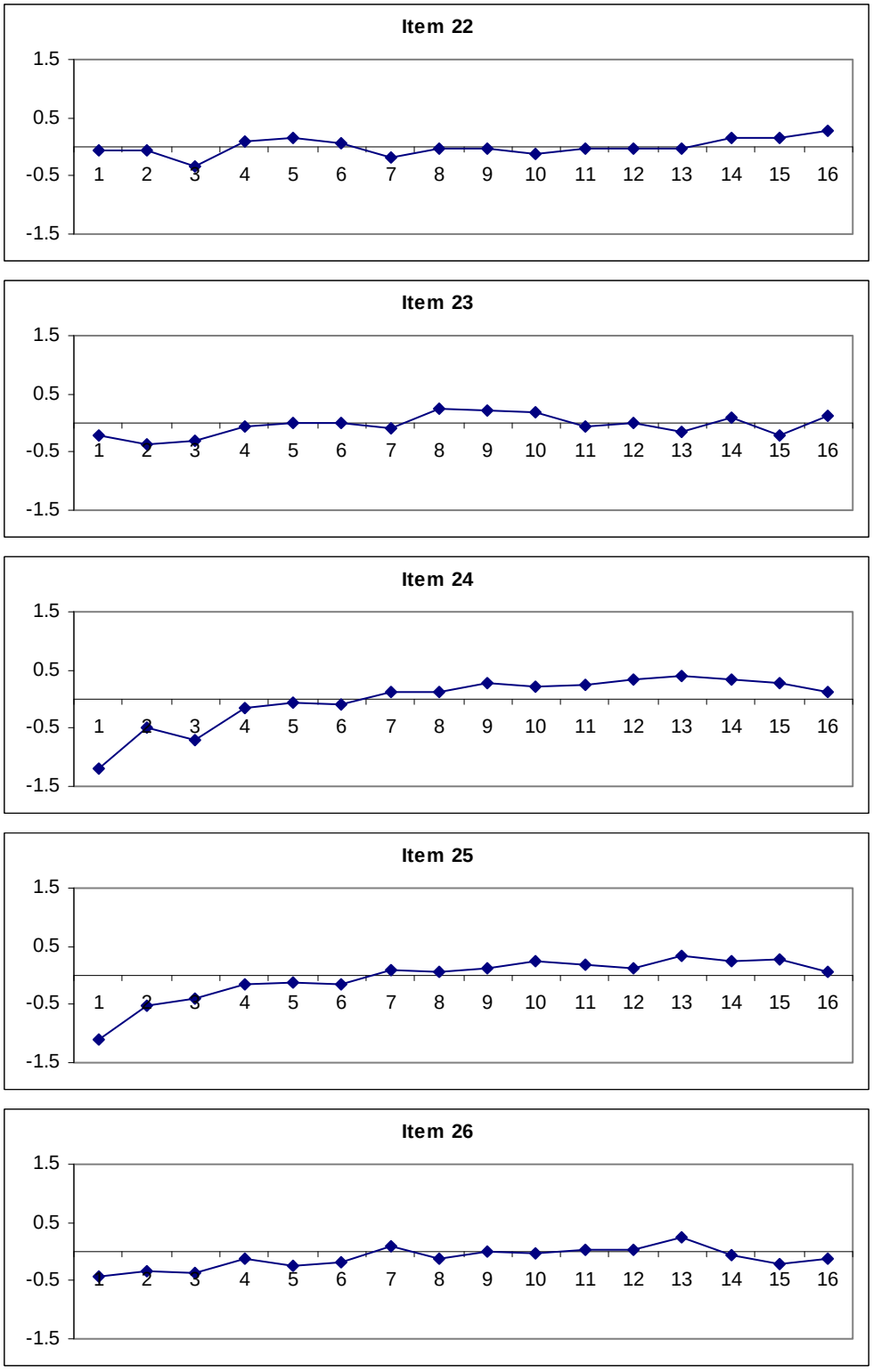


Figure 5.1 (continued): Nature of bias in the 32 items of the VRP

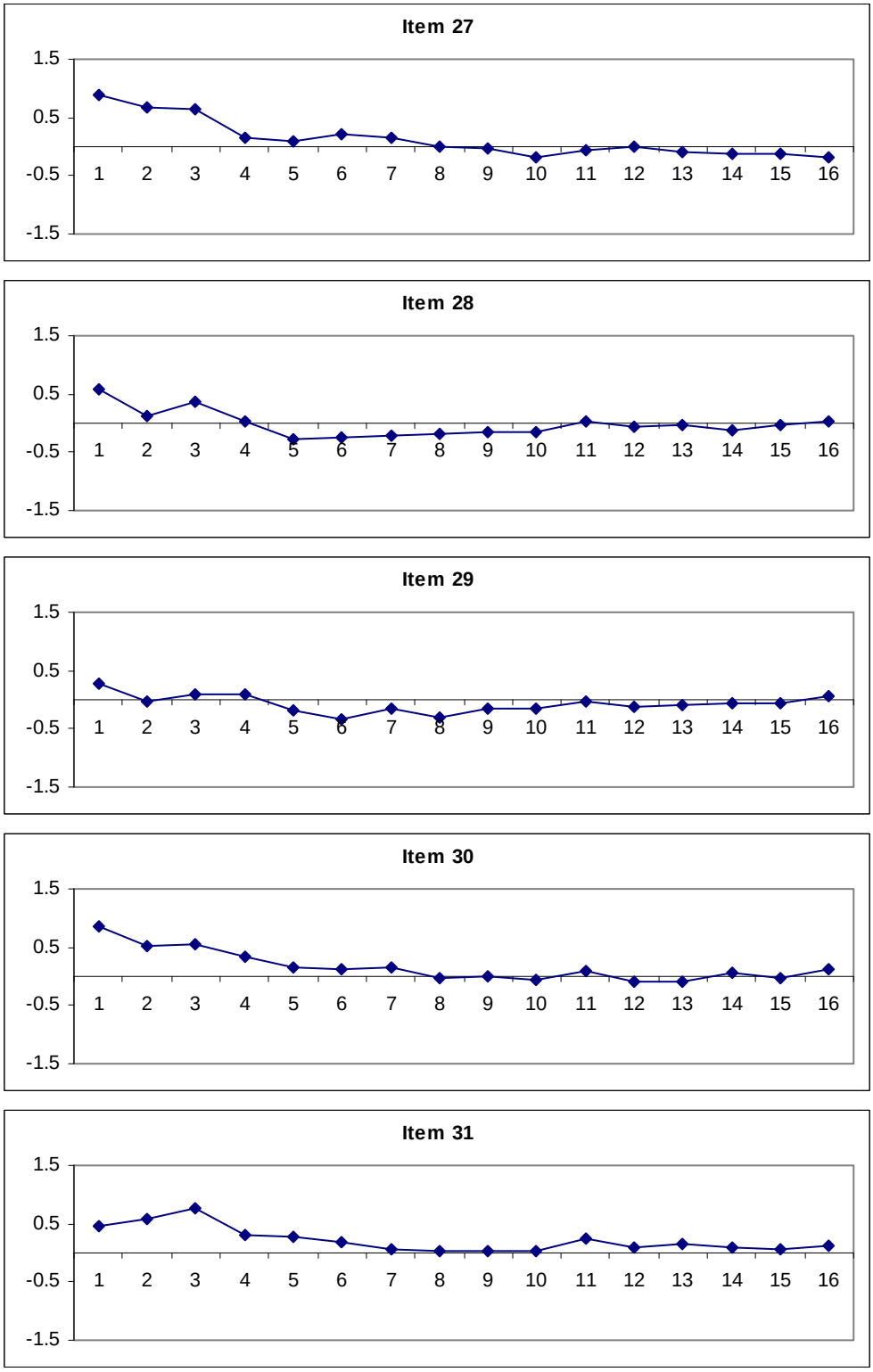
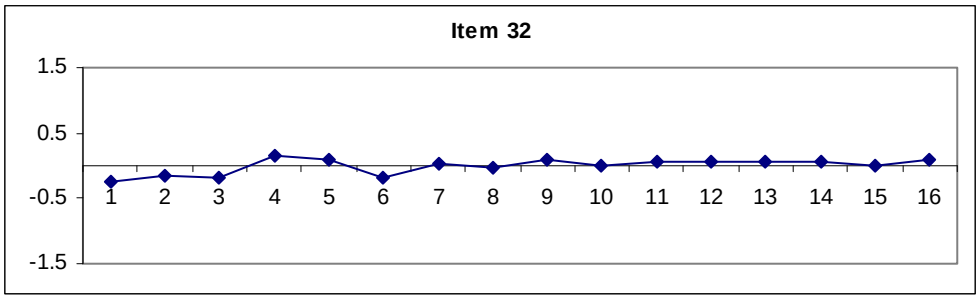


Figure 5.1 (continued): Nature of bias in the 32 items of the VRP



5.3 PSYCHOMETRIC PROPERTIES OF THE 32-ITEM SCHLINGER VRP[1]

In this section, the reliability and the validity of the VRP across cultures will be assessed. This is an important step in this research, since it tests whether the first assumption holds. If the Schlinger VRP is not found to be valid and reliable, the application of SIMNORM and SIMITNORM will be fruitless.

5.3.1 Comparison with similar advertising response research

If the approach used in this chapter to evaluate the psychometric properties of the Schlinger (1979) VRP is compared with other research aimed at scale development and testing, several similarities exist. In the development of the VRP scale, Schlinger (1979) used several independent individual responses to different commercials. Leavitt (1970) used 98 commercials and more than 3 160 respondents. Wells (1964a; 1964b) employed 171 print advertisements and more than 2 140 subjects. Nunnally and Bernstein (1994) point out the importance of using many subjects and repeated measures in scale validation. It is important to assess how the scale performs under a range of different conditions or stimuli.

[1] A more thorough exposition of the psychometric properties of the VRP has subsequent to this initial work been published in the Journal of Advertising. See Strasheim, Pitt and Caruana (2007).

5.3.2 Statements of VRP32

The item content and the overall descriptive measures of the 32 VRP items are summarised in Table 5.10, with all the subgroups combined.

Table 5.10: Descriptive statistics of the 32 VRP items (n = 4800)

Item	Description	Mean	Std Dev
V1	The commercial was lots of fun to watch and listen to.	3.769	1.073
V2	I thought it was clever and entertaining.	3.704	1.073
V3	The enthusiasm of the commercial is catching – it picks you up.	3.673	1.078
V4	The commercial wasn't just selling the product, it was also entertaining me and I appreciated that.	3.533	1.112
V5	The characters (or persons) in the commercial capture your attention.	3.978	0.995
V6	It is the kind of commercial that keeps running through your mind after you have seen it.	3.495	1.162
V7	I just laughed at it – I thought it was very funny and good.	2.550	1.212
V8	It was distracting - trying to watch the screen and listen to the words at the same time.	2.001	0.974
V9	It required a lot of effort to follow the commercial.	2.011	1.066
V10	It was too complex. I was not sure what was going on.	1.836	0.900
V11	I was so busy watching the screen, I didn't listen to the words.	2.018	0.991
V12	The commercial gave me a new idea.	3.254	1.229
V13	The commercial reminded me that I'm dissatisfied with what I am using now and I'm looking for something better.	2.526	1.147
V14	I learned something from the commercial that I didn't know before.	3.122	1.227
V15	The commercial told me about a new product and I think I'd like to try it.	3.345	1.184
V16	During the commercial I thought how that product might be useful to me.	3.471	1.169
V17	That's a good brand and I wouldn't hesitate recommending it to others	3.978	0.887
V18	I know that the advertised brand is dependable and reliable.	3.943	0.876
V19	The commercial was very realistic – that is, true to life.	3.717	1.017
V20	I felt that the commercial was acting out what I feel at times.	3.301	1.117
V21	I felt as though I were right there in the commercial experiencing the same thing.	3.104	1.166
V22	That's my idea – the kind of life the commercial showed.	3.273	1.161
V23	I liked the commercial because it was personal and intimate.	3.332	1.152
V24	This kind of commercial has been done many times. It is the same old thing.	2.008	0.915
V25	I have seen this commercial so many times – I am tired of it.	1.774	0.779
V26	I think that this is an unusual commercial, I am not sure I've seen another one like it.	3.587	1.073
V27	What they showed didn't demonstrate the claims they were making about the product.	2.196	0.990
V28	The ad didn't have anything to do with me or my needs.	2.078	1.016
V29	The commercial did not show me anything that would make me want to use on of their products.	2.155	1.104
V30	The commercial made exaggerated claims. The product would not live up to what they said or implied.	2.019	0.932
V31	It was an unrealistic ad – very far-fetched.	1.914	0.973
V32	The commercial irritated me – it was annoying.	1.747	0.957

5.3.3 The Structure of the VRP – Exploratory Factor Analysis of 32 items

The underlying structure of the VRP instrument was investigated in the following sequence. As a first stage, exploratory factor analysis (EFA) was used to establish whether the structure proposed by Schlinger (1979) was similar. Principal components exploratory factor analysis was used, using a Varimax rotation. For the number of factors extracted, the default criterion, namely the number of eigenvalues greater than one was used. Seven factors were extracted and jointly accounted for 67.8% of the variation. The results reported in Table 5.11 give the eigenvalues; the percentage of variation explained; the cumulative percentage variation explained; as well as the factor loadings. All factor loadings smaller than 0.3 were suppressed.

The results from Table 5.11 show an almost perfect resemblance to the factor structure proposed by Schlinger (1979). The factor loadings are all greater than 0.5, except item V6, which is very close to 0.5. Furthermore, almost all the items load only on the factor on which they were intended to load. The exceptions are items V2, V3, V5, V6, V12 and V23, but they still load noticeably higher on the factors they were supposed to, as indicated by Schlinger (1979).

Table 5.11: Results of Exploratory Factor Analysis on 32 item Schlinger Scale (n=4800)

Factor	1	2	3	4	5	6	7
Eigenvalue	11.82	2.70	1.81	1.73	1.39	1.16	1.09
% Variation	36.9	8.4	5.6	5.4	4.4	3.6	3.4
Cumulative	36.9	45.4	51.0	56.4	60.8	64.4	67.8
Varimax rotated factor loadings							
Item	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
V4	0.755						
V2	0.744	0.334					
V1	0.743						
V3	0.671	0.393					
V7	0.615						
V5	0.579	0.342					
V6	0.497	0.411					
V21		0.783					
V20		0.750					
V22		0.741					
V23	0.307	0.585					
V19		0.541					
V30			0.745				
V27			0.699				
V31			0.694				
V28			0.679				
V29			0.658				
V32			0.525				
V15				0.767			
V14				0.765			
V16				0.735			
V13				0.715			
V12		0.311		0.675			
V9					0.851		
V10					0.822		
V8					0.808		
V11					0.726		
V24						0.800	
V25						0.779	
V26						0.569	
V18							0.883
V17							0.855

* Factor loadings smaller than 0.3 were omitted

5.3.4 The Structure of VRP32 – Confirmatory Factor Analysis of 32 items

Exploratory factor analysis (EFA) is applied when the relationships between observed variables and latent factors are unknown. If the analysis is concerned with determining the number of factors needed to explain the relationships

between observed indicators, EFA is appropriate, and assists the researcher in reducing the dimensionality of the instrument. Confirmatory factor analysis (CFA) by contrast, is appropriate when the underlying latent structure is known beforehand, for example when the analysis is concerned with a pre-existing model. In CFA, the model comes first, and is tested for consistency with the observed data using a structural equation modelling (SEM) approach. Within the SEM framework, CFA is often referred to as the measurement model (Byrne, 1998; Raykov and Marcoulides, 2000).

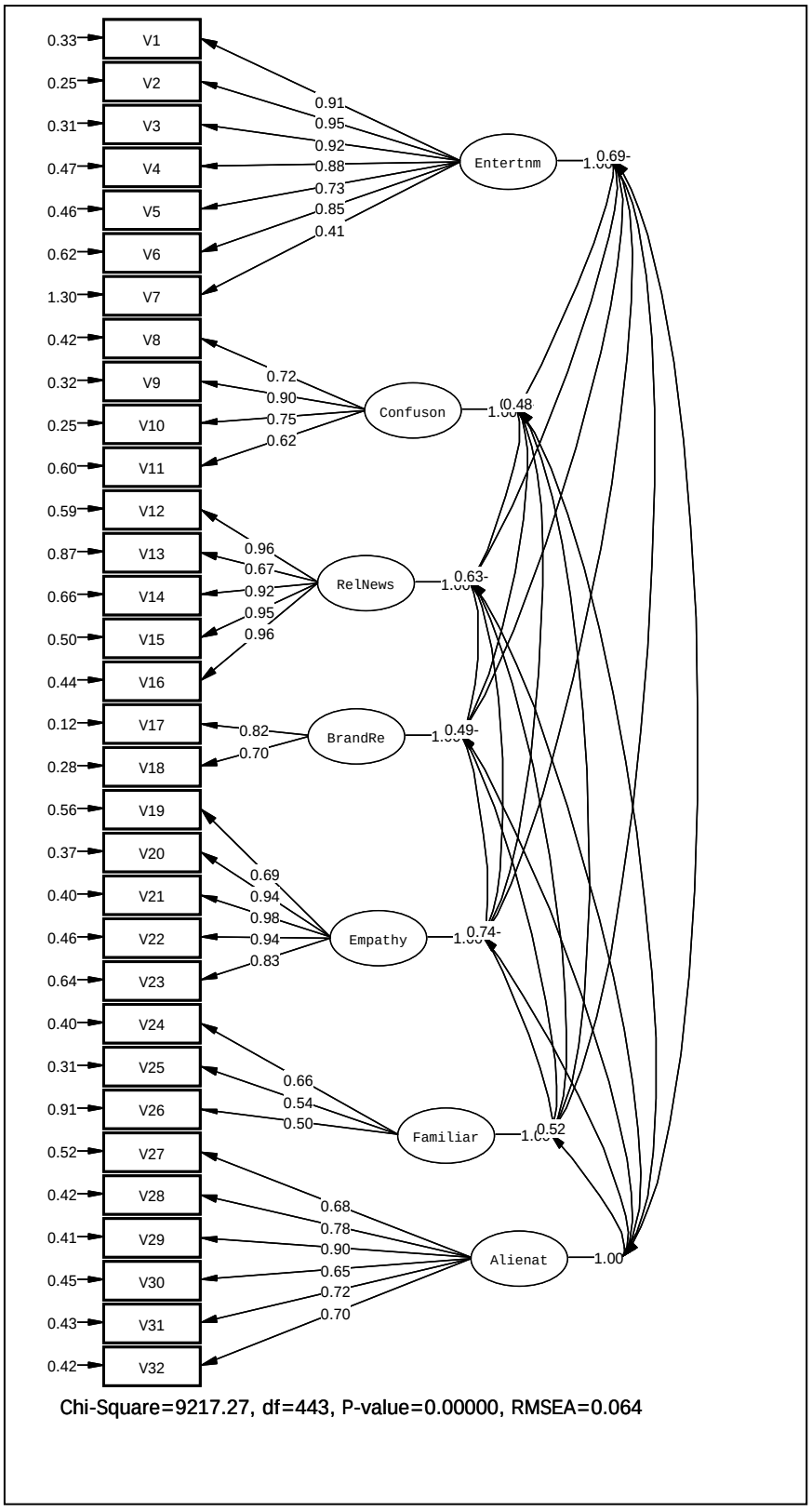
Confirmatory factor analysis using LISREL 8.8 (Jöreskog and Sörbom, 1993) was used to confirm the factor structure proposed by Schlinger (1979). Figure 5.2 shows the estimated first-order, 32-item seven factor VRP (VRP32) model as proposed by Schlinger (1979). In the model, the latent variable variances were constrained to be equal to 1. The confirmatory factor analysis can also be used to assess the reliability and construct validity of instruments across groups (Steenkamp and Van Trijp, 1991).

The estimated covariances between the latent variables are set out in Table 5.12.

Table 5.12: Estimated covariances between latent variables for VRP32

Estimated PHI matrix	Entertain-ment	Confusion	Relevant News	Brand reinforce-ment	Empathy	Familiarity	Alienation
Entertainment	1.00						
Confusion	-0.34	1.00					
Relevant News	0.62	-0.22	1.00				
Brand reinforcement	0.34	-0.19	0.33	1.00			
Empathy	0.76	-0.33	0.66	0.43	1.00		
Familiarity	-0.49	0.26	-0.34	-0.26	-0.41	1.00	
Alienation	-0.69	0.48	-0.63	-0.49	-0.74	0.52	1.00

Figure 5.2: Estimated First-Order, Seven-Factor Schlinger VRP32 Model



The Chi-Square value is a generally accepted goodness of fit index in SEM. Associated with the Chi-Square value is a p-value, which can be used to evaluate the model with a chosen significance level (e.g. 0.05), to test the hypothesis that the model is exactly capable of reproducing the relationships in the analysed covariance matrix. The model is rejected when its p-value is smaller than the significance level (which is the opposite of the approach generally used in inferential statistics). A problem encountered with the Chi-Square statistic is that it cannot be relied upon when large samples are used. With large samples, there will be an artificial tendency to reject models, even if they were only slightly inconsistent with the data. One proposed solution to the problem of large samples is to divide the Chi-Square by the degrees of freedom. If his number is smaller than 5, fit can be deemed appropriate. It is recommended in the literature that other fit indices should also be considered in order to obtain a better picture of model fit (Raykov and Marcoulides, 2000; Byrne, 1998).

Various descriptive fit indices have been developed, the first being the goodness-of-fit index (GFI), which can be loosely interpreted as the percentage of variance and covariance the model is able to explain. Alternative approaches to model assessment have also been developed. The comparative-fit index, (CFI) compares the proposed model with the null model that assumes no relationships. It measures the ratio of the improvement in the proposed model to that of the null model, and thus values of CFI close to 1 indicate that the model fits well. It is generally recommended that GFI and CFI should be approximately 0.90 or above, to indicate that the model gives a reasonable fit to the data. Another alternative fit index is the Root Mean Square Error of Approximation (RMSEA), which takes model complexity into account. It is generally accepted that RMSEA less than 0.05 indicates a good approximation; values between 0.05 and 0.08 acceptable fit; values between 0.08 and 0.10 marginal fit; and values above 0.10 poor fit. A very useful aspect of the RMSEA is that confidence intervals for the measure can be calculated. If the confidence interval is not too wide, and both limits are below 0.05, it can be argued that the model is a plausible means of describing the data (Fabrigar, Wegener, MacCallum and Strahan, 1999; Raykov and Marcoulides, 2000).

For the VRP32, where the estimated CFA-model is given in Figure 5.2, the Chi-Square (χ^2) associated with the 32-item seven-factor VRP was 9217.27 (df = 443, p < 0.01), GFI = 0.89, and CFI = 0.91. Although the χ^2 / df is 19.17, indicating poor fit, the other fit indices, namely GFI and CFI figures, suggest that the model fits the data reasonably well. The RMSEA = 0.061, which is an acceptable fit. All 32 item loadings were very highly significant (p < 0.001). The 32-item, seven-factor CFA model (VRP32) in Figure 5.2 can be regarded as a reasonable approximation to the data, but the results are somewhat disappointing, given the results obtained in the EFA and given the thorough methodology followed by Schlinger (1979) in the development of the scale.

5.4 THE 20-ITEM SCHLINGER VRP

Subsequently, various alternative models were investigated in an attempt to find a more satisfactory fit. All alternative models considered involved item deletion, and not factor structure adjustment. A more economical scale, but one with the same information richness and depth compared to the original 32-item scale, is advantageous. Other exogenous variables such as general attitude towards the brand, purchase intent, or other scales of interest may be included in the research instrument, which could enrich the scope of information.

As can be seen on Figure 5.3, the first factor, Entertainment is measured by seven items, whilst Brand Reinforcement is represent by two items. This may cause a loss of sensitivity for the Brand Reinforcement dimension. Several alternative models were considered, where items were eliminated based on the item-total correlations, and which resulted in an improved fit, based on GFI, CFI and RMSEA. One of the best alternative models shown in Figure 5.3, and consists of a 20 item seven-factor VRP (VRP20). The estimated covariances between the latent variables are listed in Table 5.13.

Table 5.13: Estimated covariances between latent variables for VRP20

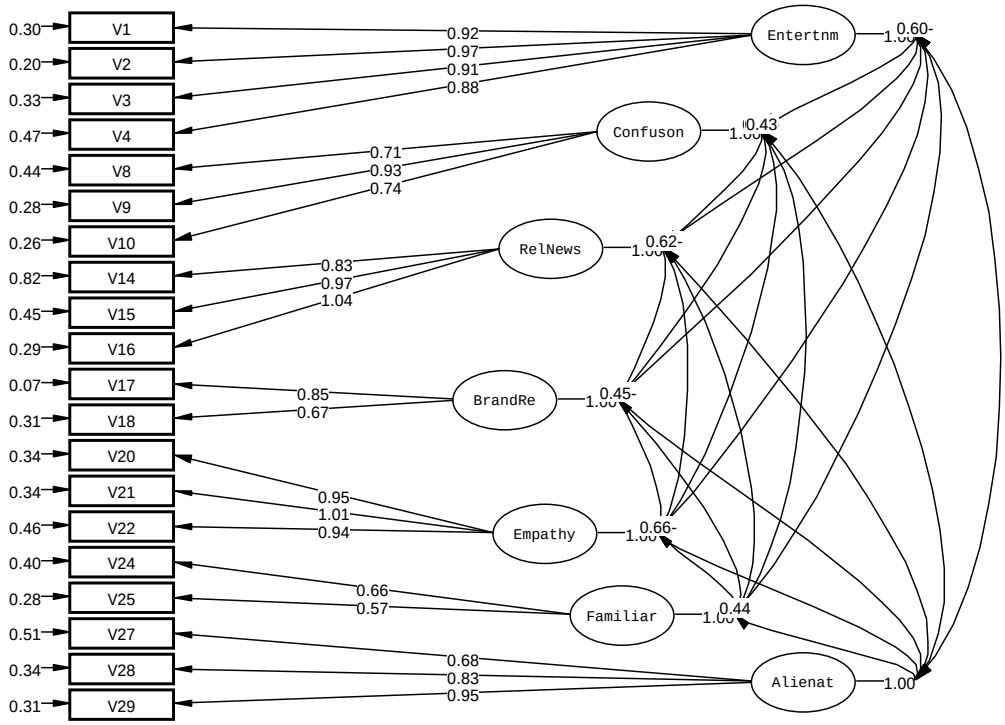
Estimated PHI matrix	Entertain-ment	Confusion	Relevant News	Brand reinforce-ment	Empathy	Familiarity	Alienation
Entertainment	1.00						
Confusion	-0.33	1.00					
Relevant News	0.55	-0.21	1.00				
Brand reinforcement	0.32	-0.16	0.36	1.00			
Empathy	0.69	-0.31	0.59	0.38	1.00		
Familiarity	-0.42	0.25	-0.27	-0.23	-0.35	1.00	
Alienation	-0.60	0.43	-0.62	-0.45	-0.66	0.44	1.00

A comparison of the fit measures for VRP32 and VRP20 appears in Table 5.14. It is clear from the all the fit measures that the VRP20 gives a significantly improved fit, with the difference in Chi-Square (8493 – 1278.9 = 7214.1) and corresponding df (443 – 149 = 294), which is very highly significant ($p < 0.0001$). Similary, all the goodness of `fit indices such as GFI and CFI improved remarkably. The most important improvement for the 20-item VRP over the 32-item VRP is the RMSEA which is 0.04, with the upper limit of the 90% confidence interval of RMSEA still below 0.05, and the width of the interval relatively small. All 20 item loadings were very highly significant ($p < 0.001$). The cross-validation of the CFA models for VRP32 and VRP20 was done on a holdout sample of 4937 observations. The results are in the lower half of Table 5.14. The VRP20-model was also fitted on the holdout sample, indicating that the findings could be extrapolated to other samples observed in similar conditions.

Table 5.14: Fit measures of VRP32 versus VRP20 for primary and secondary samples

Results for primary sample with n=4800			
Fit Statistic	32 Item VRP	20 Item VRP	Difference
Chi-Square	8493.0	1278.9	7214.1
Chi-Square DF	443	149	294
Pr > Chi-Square	0.000000	0.000000	0.000000
Chi-Square / DF	19.17	8.58	
Goodness of Fit Index (GFI)	0.8928	0.9740	
Bentler's Comparative Fit Index	0.9104	0.9794	
Bentler and Bonett's (1980) Non-normed Index	0.8997	0.9738	
Bentler and Bonett's (1980) NFI	0.9060	0.9768	
Root Mean Square Residual (RMR)	0.0613	0.0296	
RMSEA Estimate	0.0615	0.0398	
RMSEA 90% Lower Confidence Limit		0.0378	
RMSEA 90% Upper Confidence Limit		0.0418	
Results for secondary holdout sample with n=4937			
Fit Statistic	32 Item VRP	20 Item VRP	Difference
Chi-Square	9029.7	1202.4	7827.3
Chi-Square DF	443	149	294
Pr > Chi-Square	0.000000	0.000000	0.000000
Chi-Square / DF	20.38	8.07	
Goodness of Fit Index (GFI)	0.8892	0.9763	
Bentler's Comparative Fit Index	0.9056	0.9807	
Bentler and Bonett's (1980) Non-normed Index	0.8943	0.9754	
Bentler and Bonett's (1980) NFI	0.9012	0.9780	
Root Mean Square Residual (RMR)	0.0669	0.0370	
RMSEA Estimate	0.0627	0.0378	
RMSEA 90% Lower Confidence Limit		0.0359	
RMSEA 90% Upper Confidence Limit		0.0398	

Figure 5.3: Estimated First-Order, Seven-Factor Schlinger VRP20 Model



5.5 RELIABILITY AND VALIDITY OF VRP32 AND VRP20

For a scale to be used with confidence it must possess validity; in simple terms it must measure what it purports to measure. In this section, a number of aspects of the validity of the two versions of the VRP are considered, namely: the reliability, the nomological, convergent, discriminant, criterion, (concurrent and predictive) and constuct or face validity of both VRP32 and VRP20.

5.5.1 The Reliability of VRP32 and VRP20

The Cronbach Alpha (Cronbach and Meehl, 1955) measures the internal consistency or reliability of the items of a questionnaire. The reliability of an instrument relates to the ability of the instrument to produce consistent results, which is a necessary but not sufficient requirement for the validity of the instrument. Reliable instruments should be robust, in the sense that they work well under different times and conditions. Cronbach Alpha can be interpreted as the 'average' correlation among items in a test. However, it is partly a misnomer, because Cronbach Alpha both reflects the number of items and their average correlation. (Cooper and Schindler, 1998). The generally accepted norm for Cronbach Alpha is 0.7 or higher (Peterson, 1994; Fabrigar, Wegener, MacCallum and Strahan, 1999).

Cronbach Alpha was calculated for all advertisements combined (n=4800), and for VRP32, the estimated Alpha was equal to 0.940 for all 32 items. Alpha was also estimated for each of the 24 commercials individually, and the lowest Alpha was 0.917 and the highest 0.970. The median alpha for the 24 advertisements was 0.938. This demonstrates that the reliability of the original 32 items, even when considered over relatively small samples of 200 independent observations for each individual commercial, was very high (compared to a figure of 0.7).

For VRP20, the estimated Alpha was equal to 0.911 for the 20 remaining items. Alpha was again estimated for each of the 24 commercials individually, with the

lowest Alpha 0.880 and the highest 0.953. The median Alpha for the 24 advertisements was 0.910. The reliability of the remaining 20 items, considered over smaller samples of 200 independent samples from each commercial, was also more than adequate.

Reliabilities were also estimated for each of the seven dimensions for both VRP32 and VRP20. The estimated Alphas are given in Table 5.15. With a few exceptions, the reliabilities for the VRP32 and VRP20 on each of the seven dimensions are very similar. The only dimension for which Alpha is below 0.7 is Familiarity, but still acceptable, considering that it has only two items. Item 26, which was included in the Familiarity dimension of the 32-item VRP, was slightly confusing, and formulated in the opposite direction to the other statements used for the same latent variable. This researcher suspects that it could have resulted in respondent confusion.

Table 5.15: Reliability of the 32- and 20- item VRP

Cronbach Alpha	32-item VRP	20-item VRP
All items	0.940	0.911
Entertainment	0.889	0.912
Confusion	0.845	0.845
Relevant News	0.865	0.835
Brand reinforcement	0.847	0.847
Empathy	0.881	0.880
Familiarity	0.626	0.682
Alienation	0.879	0.833

5.5.2 Nomological validity

The confirmation of nomological validity requires a proof of the extent to which a measuring instrument "fit[s] lawfully into a network of expected relationships". (Nunnally and Bernstein, 1994: 91). This pattern of results is called a nomological network, and nomological validity can be established with factor analysis when items expected to load together actually do so (Cronbach and Meehl, 1955; Carman 1990).

For both VRP32 and VRP20, EFA solutions as shown in Tables 5.11 and 5.16 and CFA solutions as shown in Figures 5.2 and 5.3 and Table 5.14, provide sufficient evidence for nomological validity. For the VRP20, the results of EFA are shown in Table 5.16, with the number of factors extracted pre-specified to be equal to seven, using varimax rotation and suppressing factor loadings smaller than 0.3. The results show that the factor loadings load onto the dimensions as expected, and the seven factors account for 78.8% of the variation. Similar results were obtained for the hold-out sample.

Table 5.16: Results of exploratory factor analysis for VRP20

Factor	1	2	3	4	5	6	7
Eigenvalue	7.66	2.03	1.57	1.40	1.23	0.97	0.89
% Variation	38.3	10.2	7.8	7.0	6.2	4.9	4.4
Cumulative	38.3	48.5	56.3	63.3	69.5	74.4	78.8
Varimax rotated factor loadings							
Item	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
V2	0.834						
V1	0.828						
V4	0.821						
V3	0.758						
V15		0.835					
V16		0.800					
V14		0.779					
V9			0.885				
V10			0.845				
V8			0.827				
V21				0.821			
V20				0.782			
V22	0.345			0.744			
V27					0.777		
V28					0.773		
V29					0.709		
V18						0.910	
V17						0.880	
V25							0.845
V24							0.842

5.5.3 Convergent and discriminant validity

Discriminant validity is indicated when factors and their items are truly different from one another, in the sense that there are low correlations between variables of dissimilar constructs (Carman, 1990; Zikmund, 1997). Discriminant validity is necessary to ensure that the constructs are sufficiently different, and do not overlap the domain being tested. Convergent and discriminant validity were evaluated by calculating the average variance extracted (AVE) for each of the seven factors, using the method suggested by Fornell and Larcker (1981). Convergent validity is established if the shared variance accounts for 0.50 or more of the total variance. Discriminant validity is evident when the AVE for each construct is greater than the squared correlation between that construct and any other construct in the model. The results presented in Table 5.17 confirm both the convergent and discriminant validity of the seven-factor VRP32 and VRP20. The only exception is the Familiarity dimension in the VRP32, which does not possess convergent validity.

Table 5.17: Convergent and discriminant validity assessment for VRP32 and VRP20

VRP32	CR	AVE	$(Corr)^2$	Ave/ $(Corr)^2$	CV	DV
Entertainment	0.901	0.578	0.570	1.014	yes	yes
Confusion	0.849	0.588	0.234	2.510	yes	yes
Relevant News	0.866	0.567	0.433	1.309	yes	yes
Brand Reinforcement	0.851	0.742	0.236	3.143	yes	yes
Empathy	0.885	0.609	0.570	1.068	yes	yes
Familiarity	0.667	0.408	0.268	1.521	no	yes
Alienation	0.879	0.549	0.545	1.007	yes	yes
VRP20	CR	AVE	$(Corr)^2$	Ave/ $(Corr)^2$	CV	DV
Entertainment	0.913	0.726	0.475	1.528	yes	yes
Confusion	0.850	0.655	0.186	3.527	yes	yes
Relevant News	0.840	0.640	0.387	1.654	yes	yes
Brand Reinforcement	0.856	0.751	0.203	3.708	yes	yes
Empathy	0.881	0.713	0.475	1.501	yes	yes
Familiarity	0.688	0.524	0.195	2.684	yes	yes
Alienation	0.835	0.631	0.437	1.443	yes	yes

CR = Composite Reliability = $(\Sigma$ of std. Loading$)^2 / (\Sigma$ of std. Loading$)^2 + \Sigma$ of ε_j^2
AVE = Average Variance Extracted = Σ of (std. Loading$)^2 / \Sigma$ of (std. Loading$)^2 + \Sigma$ of ε_j^2
CV = Convergent Validity (AVE >0.50)
DV = Discriminant Validity = AVE/$(Corr)^2$ >1
$(Corr)^2$ = Highest $(Corr)^2$ between factor of interest and remaining factors

5.5.4 Criterion related validity

Criterion related validity offers an answer to the question of whether the instrument correlates with other measures of the same construct. The criterion validity of an instrument is indicated if a scale performs as expected in relation to other variables selected as meaningful criteria. To assess the criterion validity, at least concurrent or predictive validity should be assessed.

Concurrent validity concerns the degree to which dimensions in the research instrument can be used to predict the existence of a current behaviour, attitude or condition. In this research, the convergent validity of the instrument will be indicated, if the seven dimensions of the VRP correlate significantly and in ways that can be expected with a related, but different measurement, such as overall *Likeability*. Likeability has been one of the most successful predictors of advertising success (Cook, 1992).

Predictive validity is indicated when a model stating the relationships between the instrument and alternative measures (such as a model to establish concurrent validity) reasonably predicts the alternative criteria (such as *Likeability*).

5.5.5 Concurrent validity of VRP32 and VRP20

To assess the concurrent validity of the VRP in this research, two similar multiple regression models with *Likeability* as dependent variable and the seven dimensions of the VRP were performed with the seven dimensions of both VRP32 and VRP20 as independent variables. The results are given in Table 5.18. Both the R^2–values in the multiple regression models were 0.52, which is highly significant with $p<0.001$. The root mean square error (RMSE) is equal for the two models, which means that the expected error, should the model be used for prediction, will be the same for both models. From the multiple regression fit summaries, there is little choice between the VRP32 and VRP20.

In order to interpret the model, consider the estimated regression coefficients in Table 5.18. For VRP32, if all other dimensions of two persons were exactly equal, but there was a unit difference (on a 5-point scale) between them, the expected difference in *Likeability* for them would be 0.99, which is almost one unit on a scale from 1 to 10. Similarly, the higher the *Entertainment* score, the higher the *Relevance* score (though smaller in magnitude); and the higher the *Empathy* score, the higher the expected *Likeability*. In constrast, for higher scores of *Confusion*, *Familiarity* and *Alienation*, lower *Likeability* scores can be expected. From the model it seems that the *Brand Reinforcement* dimension is not a significant predictor of *Likeability* in a multiple regression model, given the other six dimensions.

Table 5.18: Regressions of *Likeability* (Dependent Variable) with the seven dimensions of VRP32 and VRP20 as independent variables

	VRP32		VRP20	
ANOVA	F-Value 733.858	Prob>F 0.0000	F-Value 732.680	Prob>F 0.0000
RMSE Dependent Mean R-Square Adjusted R-Square	1.427 7.292 0.5174 0.5167		1.427 7.292 0.5170 0.5163	
Estimates	Parameter Estimates	Prob>|T|	Parameter Estimates	Prob>|T|
Intercept	3.953	0.0000	3.414	0.0000
Entertainment	0.990	0.0000	0.987	0.0000
Confusion	-0.250	0.0000	-0.235	0.0000
Relevant News	0.078	0.0052	0.158	0.0000
Brand Reinforcement	-0.026	0.3569	0.043	0.1234
Empathy	0.351	0.0000	0.238	0.0000
Familiarity	-0.082	0.0139	-0.153	0.0000
Alienation	-0.402	0.0000	-0.211	0.0000

Although the estimated regression parameters for VRP20 and VRP32 are different, similar trends can be observed for VRP20. In summary, from the multiple regression analyses it is clear that both VRP32 and VRP20 are useful in predicting a very important external measure, namely *Likeability* of the commercial. This confirms the convergent validity of the VRP.

5.5.6 Construct validity

Construct validity refers to the extent to which an instrument covers the range of meanings included in the concept (Babbie, 1992). A simple but powerful means of confirming the construct validity of a measure is to compare scores achieved on the measure with independent measures of a similar construct.

In order to do this, the mean scores on each of the seven dimensions of both the VRP32 and VRP20 were computed. A table was created that enabled examination of the mean scores on the seven dimensions at each of the ten levels of the overall *Likeability* of the commercial. The *Likeability* score can be seen as an independent measure of the overall effectiveness of a commercial. In the table the seven dimensions were averaged by response to the overall *Likeability* scale. Table 5.19 provides the outcome of this examination of mean scores.

From Table 5.19 it is apparent that there is a strong positive relationship between the overall *Likeability* and the evaluation of the *Entertainment*, *Relevant News*, *Brand Reinforcement* and *Empathy* dimensions. In other words, the higher respondents rated the *Likeability* of the commercial, the higher their average rating was on these four latent variables of commercial effectiveness. On the other hand, the higher respondents rated the *Likeability* of the commercial, the lower their average rating was on *Confusion*, *Familiarity* and *Alienation*. It is also interesting to note from Table 5.19 that the majority of viewers regarded the commercials as very likeable, with more than half rating it as 8 or higher on a ten-point scale.

Table 5.19: Construct Validity of the VRP: Cross tabulation of the seven VRP dimensions against the 10-point *Likeability* scale

	Likeability									
VRP32	1	2	3	4	5	6	7	8	9	10
Entertainment	1.70	1.82	2.17	2.40	2.90	3.29	3.53	3.82	3.98	4.08
Confusion	2.50	2.38	2.59	2.58	2.33	2.20	1.94	1.85	1.72	1.66
Relevant News	1.72	1.86	2.00	2.23	2.73	2.99	3.12	3.33	3.51	3.62
Brand Reinforcement	3.31	3.36	3.47	3.51	3.68	3.89	3.93	4.08	4.13	4.21
Empathy	1.67	1.92	2.04	2.23	2.71	3.06	3.33	3.62	3.81	3.91
Familiarity	2.72	2.79	2.61	2.54	2.39	2.17	2.04	1.97	1.87	1.78
Alienation	3.29	3.29	3.28	2.97	2.56	2.19	1.97	1.77	1.66	1.62
VRP20	1	2	3	4	5	6	7	8	9	10
Entertainment	1.53	1.72	2.11	2.37	2.97	3.39	3.70	4.00	4.16	4.32
Confusion	2.54	2.40	2.60	2.60	2.33	2.19	1.92	1.81	1.70	1.64
Relevant News	1.81	1.88	2.14	2.33	2.91	3.17	3.30	3.51	3.70	3.77
Brand Reinforcement	3.31	3.36	3.47	3.51	3.68	3.89	3.93	4.08	4.13	4.21
Empathy	1.63	1.79	1.94	2.07	2.57	2.91	3.19	3.50	3.72	3.82
Familiarity	2.45	2.52	2.40	2.34	2.21	2.01	1.87	1.80	1.72	1.59
Alienation	3.26	3.46	3.35	3.12	2.72	2.33	2.12	1.88	1.77	1.75
Sample size	67	74	125	254	407	448	747	1287	763	628

It can thus be accepted that the VRP possesses construct validity for the measurement of effectiveness or likeability of the commercial.

5.5.7 Content validity

Probably the most challenging objective for the developers of research instruments is to provide evidence of content or face validity, which refers to the extent to which the measuring instruments provide sufficient, complete and adequate coverage of the study topic. This is often determined qualitatively by the subjective judgement or agreement between experts or professionals in the field of interest that the topic has been covered sufficiently.

When one considers the thorough methodology that Schlinger (1979) employed through as described in her 1979 article in The Journal of Advertising Research, one could argue that this provides sufficient evidence to prove the face validity of the Schlinger (1979) instrument. Another argument that may be used in favour of construct validity is that other attitudinal instruments developed so far (e.g. Wells, 1964; Leavitt, 1970; Wells, Leavitt and McConville, 1971 and Leavitt, 1983;

Lastovicka, 1983) have also found latent constructs that are very similar to those of Schlinger (1979).

5.5.8 Summary of reliability and validity assessment

Overall the results presented in the above sections support arguments for the psychometric soundness of both the VRP32 and the VRP20 as a first-order seven-factor instrument. In contrast to the findings of Zinkhan and Burton (1989), which were based on the reactions of 20 viewers, each exposed to 25 commercials, this research has found that in a large sample (n = 4 800), and given a large range of different stimuli (24 commercials), the VRP possesses the psychometric properties that are desirable in a research instrument. The results also support a more economical VRP20, which contains only a subset of 20 items of the original 32 items. The VRP20 was found to be in several aspects superior to the original VRP32.

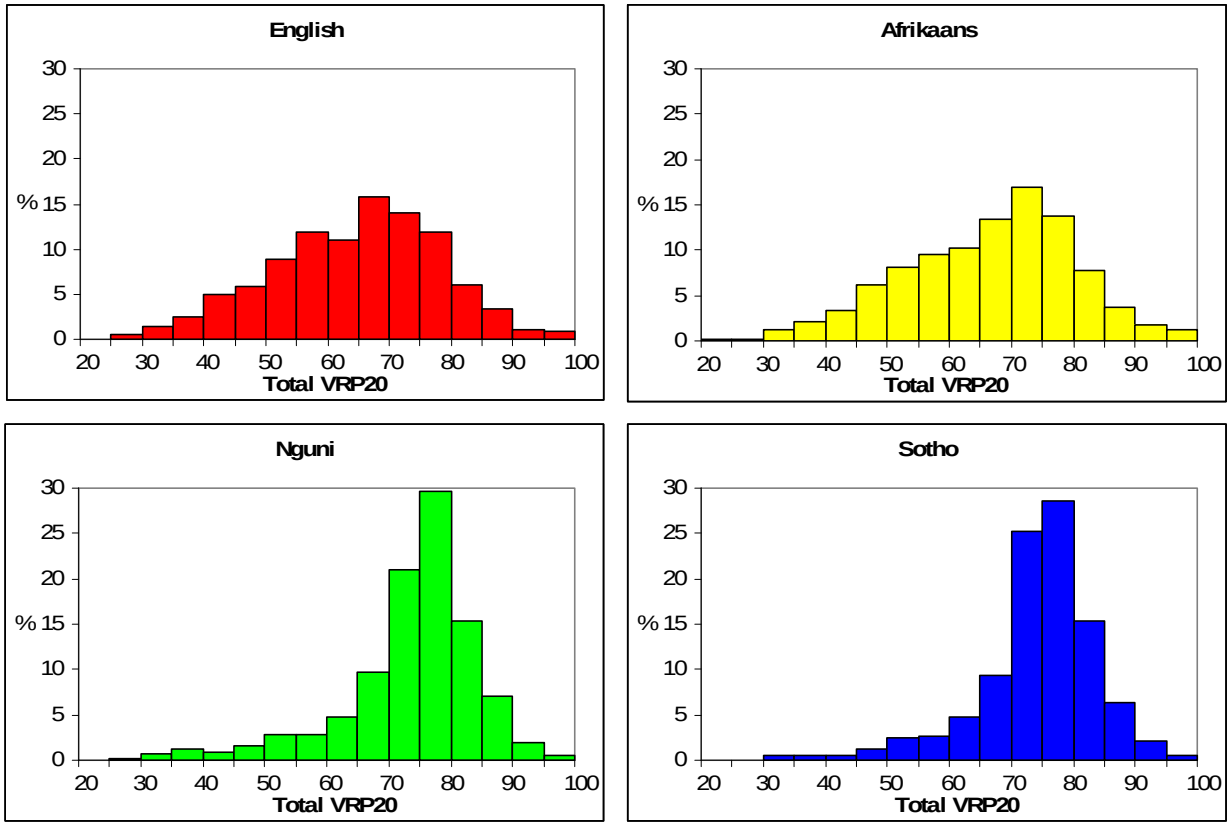
The seven dimensions resonate meaningfully with the range of responses that can be expected from viewers of television commercials. Advertisers can use the VRP on a continuous basis to gauge perceptions of specific commercials, and the results can be tracked over time. This application enables advertisers to establish whether or not a commercial is still achieving its purpose. The familiarity dimension is specifically useful since it is useful for tracking ad wearout.
In this research, the results obtained through the assessment of the reliability and validity of the VRP, imply that for the purposes of this research, it would be better to proceed with the VRP20 rather than the VRP32, since the VRP20 is more in line with the stated assumptions in this research. Therefore, all the results that follow this section, included only the twenty items and dimensions as found in this section.

5.6 RESULTS OF SIMNORM

5.6.1 SIMNORM of aggregate scores of the Schlinger (1979) VRP20 for four cultural groups

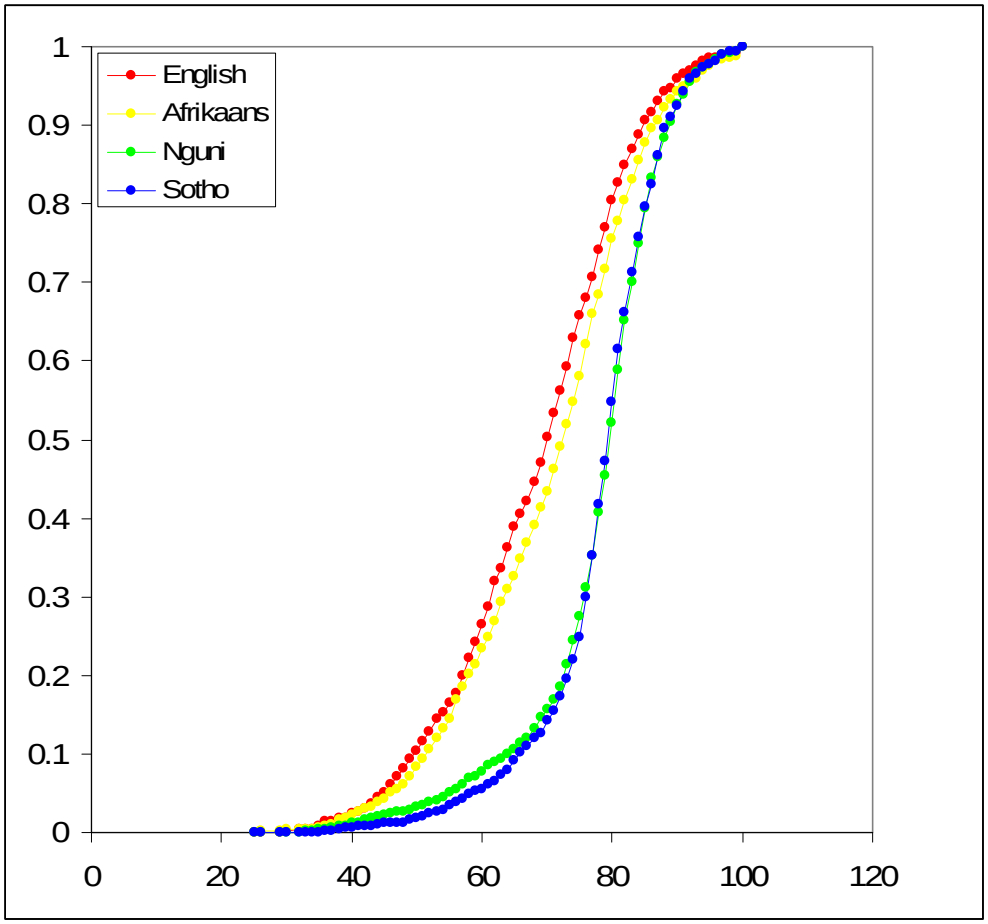
Figure 5.4 displays the total mean score of 20 items (with all negatively stated items reversed) of the Schlinger (1979) VRP20 scale. The results are from a sample of 4800 respondents from four cultural groups, who saw 24 different commercials. It is clear that the overall responses from the four cultural groups on Schlinger's (1979) VRP scale are very different across the four cultural groups.

Figure 5.4: Histograms of the sum of raw scores of the VRP20 Schlinger (1979) scale for four cultural groups



The cumulative densities are shown in Figure 5.5. The cumulative density functions clearly show that there is not only a locality difference between the four cultural groups, but also a difference in spread and shape.

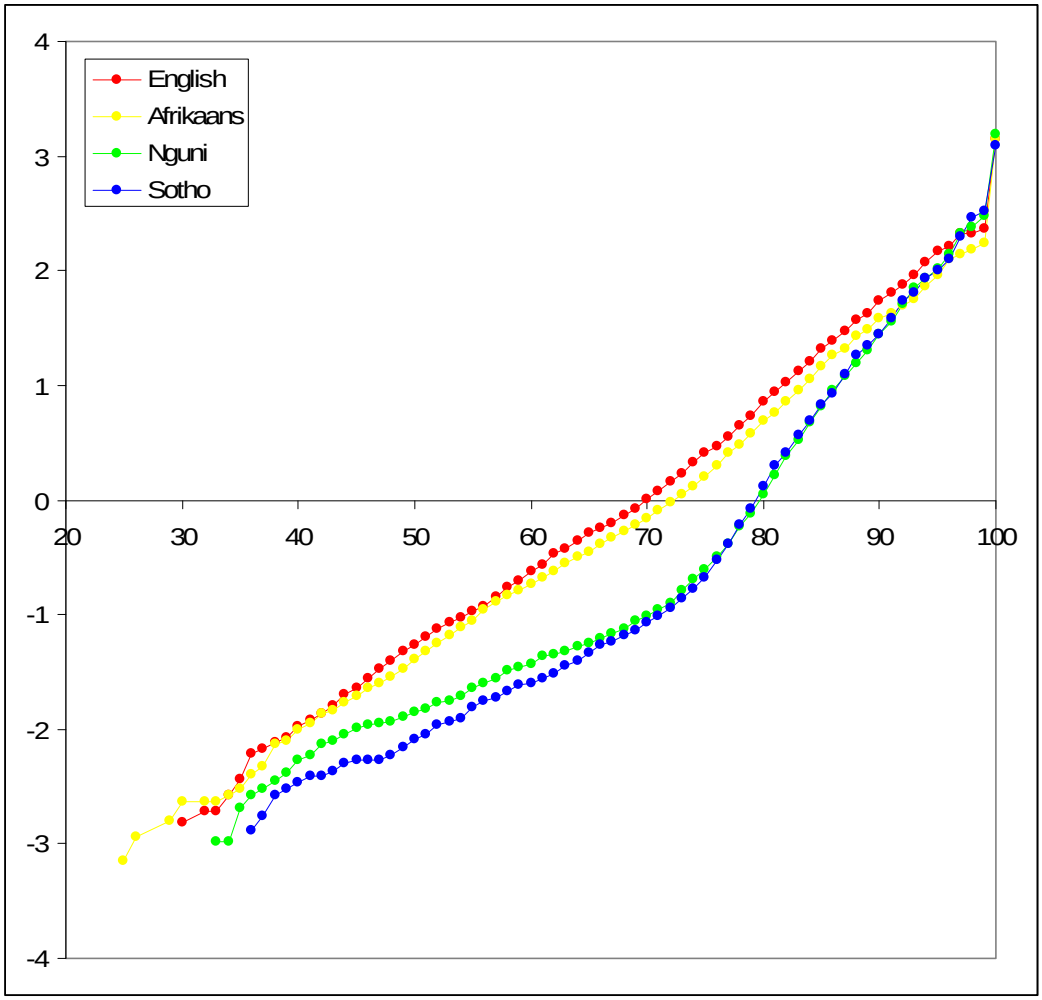
Figure 5.5: Cumulative density of the total score of the VRP20 scale for four cultural groups



The Q-Q plot in Figure 5.6 suggests the form of the set of transformation functions that will transform the scores to normally distributed distributions, thereby introducing equivalence to the data set.

SIMNORM was applied to the VRP20 data set, and the estimated parameters and fit statistics are summarised in Appendix B.

Figure 5.6: Q-Q plot for four cultural groups



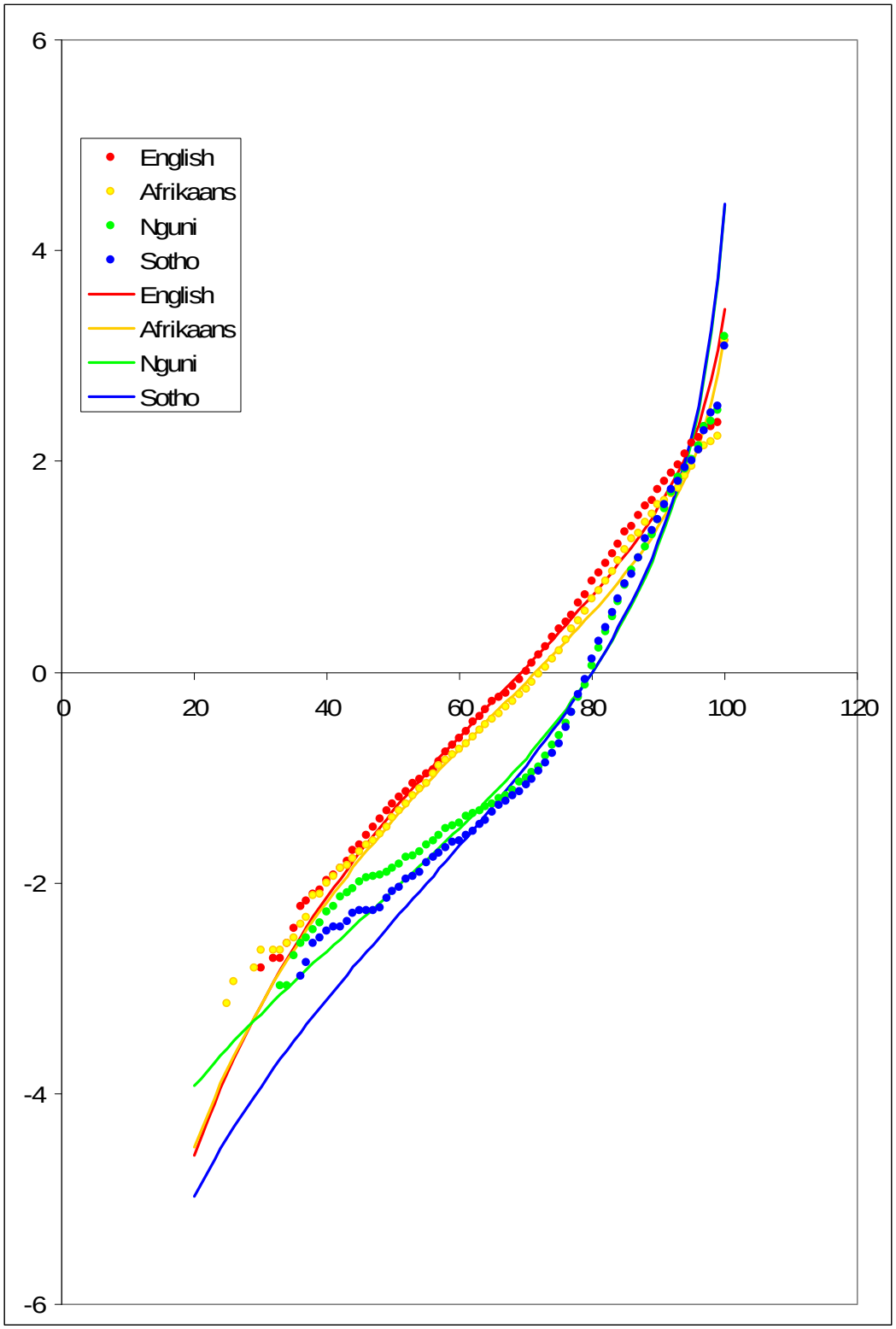
Model A

Model A is the least restrictive model and gives separate normalisations for each group, with U, the theoretical maximum of the total score, fixed for all the groups. The parameters ϕ_{1j} , ϕ_{2j} and ϕ_{3j} are estimated independently for each group, so that there are $p = 3g$ parameters, resulting in twelve parameters to be estimated. A graphical representation of the estimated normalisation functions is given in Figure 5.7.

It is interesting to note that the normalisation functions are not restricted by parameters that kept constant over groups, and therefore, each of the normalisation functions has been optimised or estimated to transform the aggregate scores to their “best ability” to obtain normal distributions for each

group. Therefore, the smooth functions in green and blue for the Nguni and Sotho groups, towards the lower part of the graphs, show how the normalisation functions attempt to “pull” the scores towards a more normal distribution.

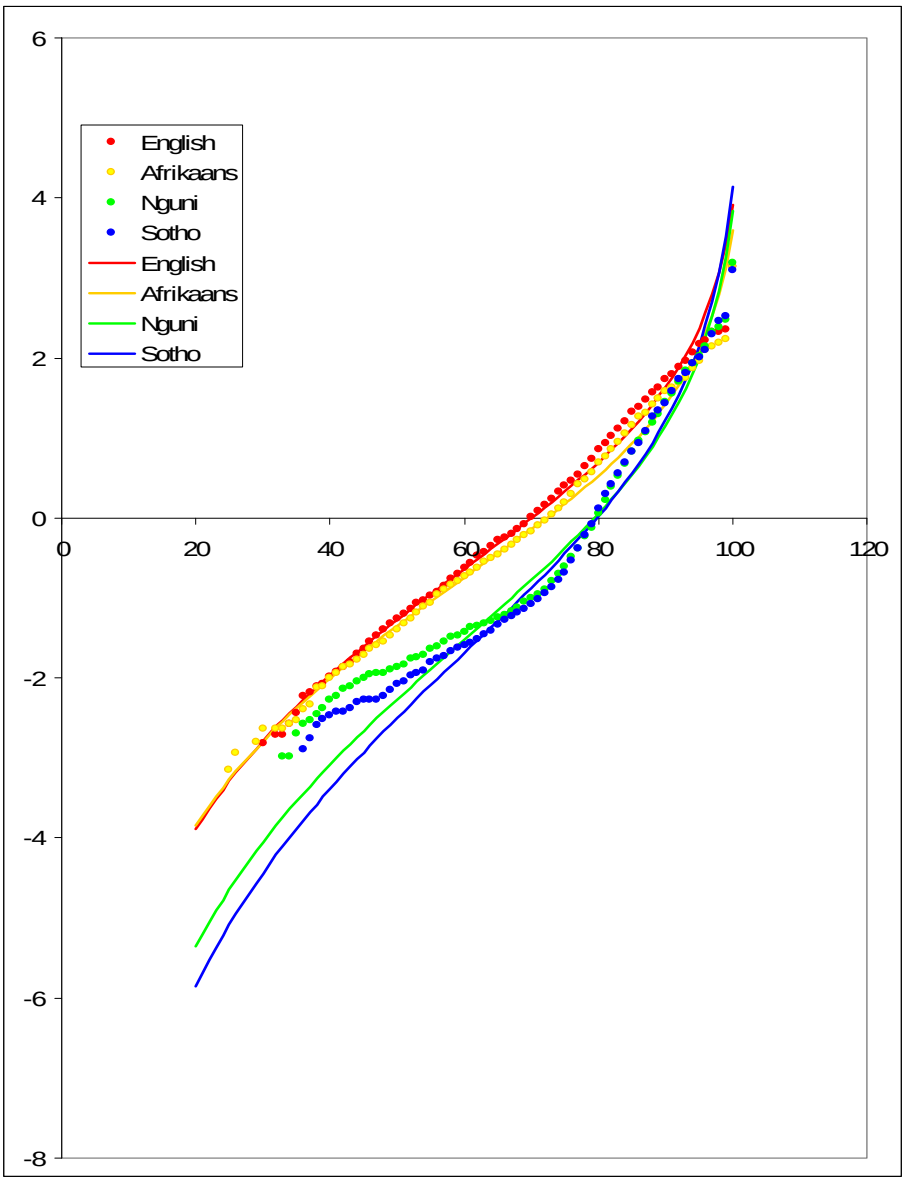
Figure 5.7: The SIMNORM Model A functions for four cultural groups



Model B

Model B is a special case of Model A, with γ equal for all groups. The number of parameters for Model B is $p = 2g + 1 = 2\times4+1$, resulting in nine parameters estimated when four groups are normalised. The estimated simultaneous normalisation functions are given in Figure 5.8. It is clear from this figure that the simultaneous normalisation functions are more restricted. The amiability or shape parameter, γ, has been restricted to remain constant across cultural groups. Although the fit is marginally worse than in Model A, the advantage is that there is more order (and predictability) in the transformations.

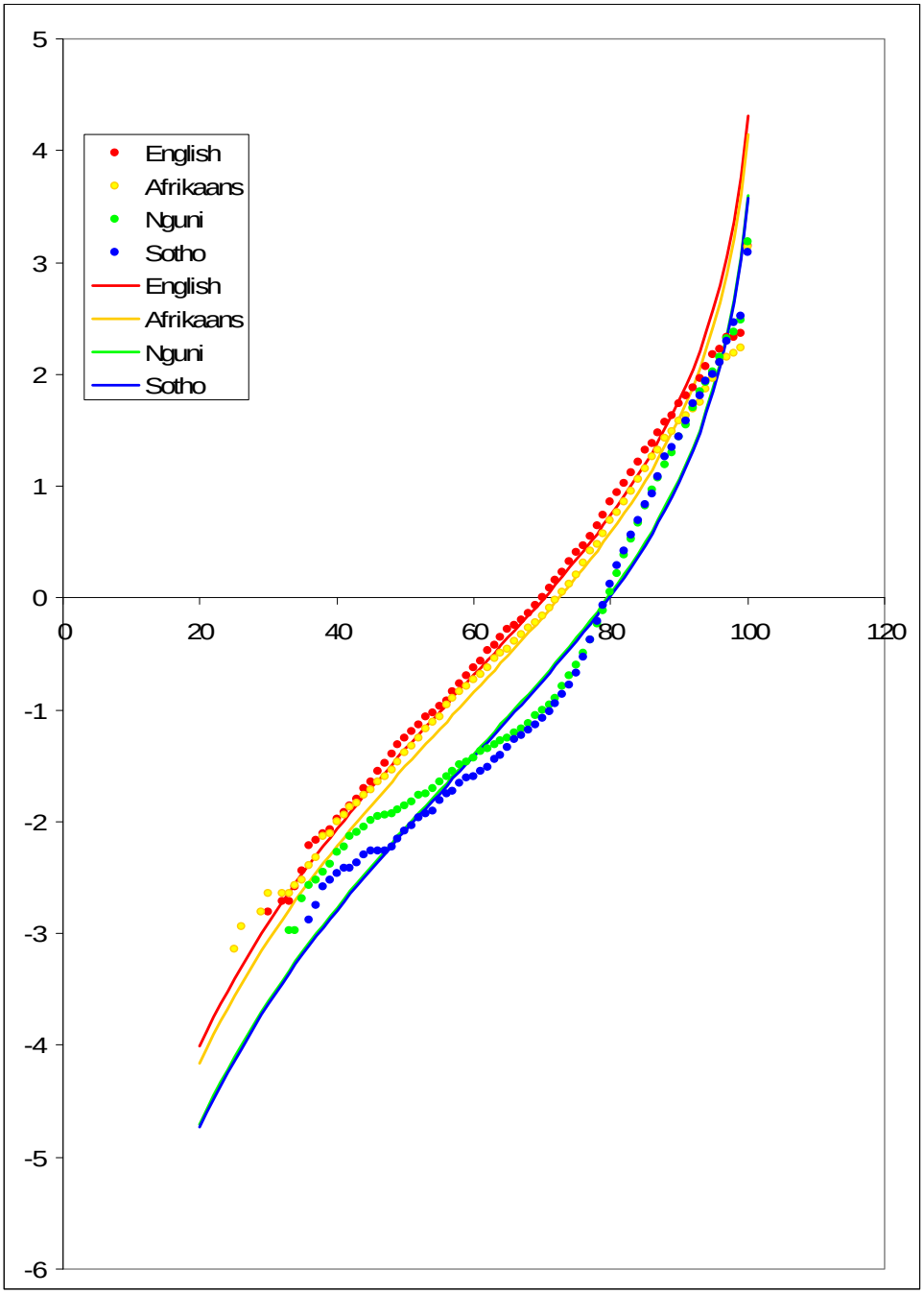
Figure 5.8: The SIMNORM Model B functions for four cultural groups



Model D

Model D is a special case of Model B, with γ and ϕ_2 restricted to be equal for all groups, so that the number of parameters is $p = g + 2 = 4 + 2$, which is six parameters. The estimated normalisation functions are plotted in Figure 5.9. In this plot, it is only the location parameter, ϕ_1 that is different across the functions. The result is that the SIMNORM functions are very orderly (and therefore far more predictable) than in the previous two models.

Figure 5.9: The SIMNORM Model D functions for four cultural groups



The estimated parameters were used to obtain normalised scores for each of the three models. The descriptive statistics for the normalised scores can be seen in Table 5.20. Note that the means and the standard deviations for the standardised scores and all three SIMNORM models are zero and one respectively. For the standardised scores, the skewness and kurtosis remained unchanged when compared to raw scores, but have improved considerably closer to zero for all three SIMNORM models.

The tests for normality in Table 5.20 all have as null hypothesis that the distribution is normal. In all the cases, even in the SIMNORM models, the null hypothesis is rejected. However, it should be noted that these tests are very sensitive, and in the case of large samples, the null hypothesis may be rejected too easily. It is therefore also appropriate to use visual representations to test whether the distribution can be assumed to be normal. The best methods are histograms, box-and-whisker plots and normal probability plots. The histograms are in Figures 5.10 to 5.13, while the box-and-whisker plots appear in Figures 5.14 to 5.18.

Figure 5.10 shows the distributions of standardised scores, and it is clear that the shapes of these distributions are identical to those in Figure 5.4. The distributions of the total scores that were normalised with each of the three models A, B and D, are presented in Figures 5.11 to 5.13 for the four cultural groups. Note that the SIMNORM distributions are more similar across the cultural groups, and that they indeed resemble the normal distribution better than the raw scores or standardised scores.

Figure 5.10: Distribution of standardised total scores for four cultural groups

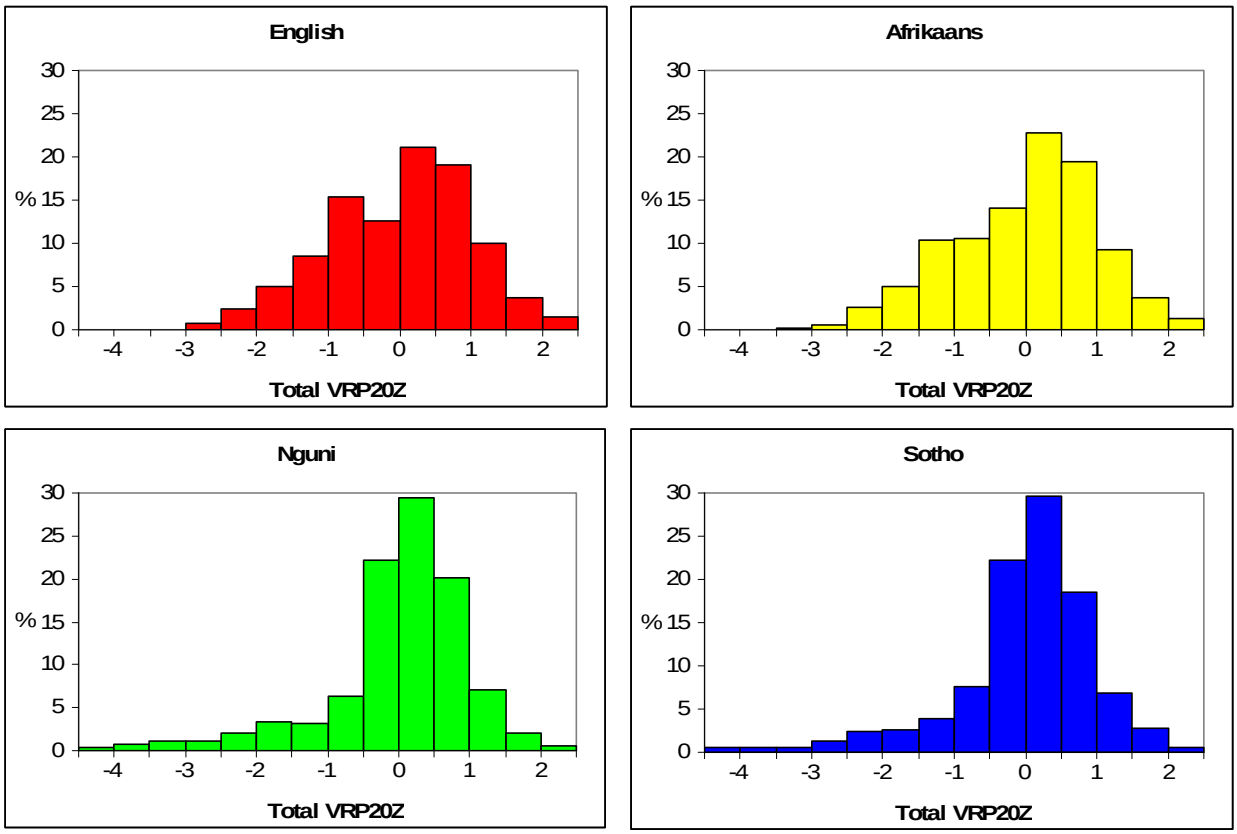


Table 5.20: Descriptive measures for raw scores, standardised scores and normalised scores over four cultural groups for *Overall Schlinger VRP20_10*

Descriptive	Raw Total Score				Standardised Scores				SIMITNORM Model A				SIMITNORM Model B				SIMITNORM Model D			
Measure	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
Mean	69.0	70.8	78.2	78.7	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Mode	74.0	76.0	80.0	80.0	0.37	0.38	0.16	0.14	0.31	0.29	0.00	0.00	0.26	0.25	0.05	0.02	0.27	0.27	0.04	0.01
Std Deviation	13.5	13.7	10.8	9.7	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.07	1.13	0.92	0.85
Skewness	-0.39	-1.40	-1.24	0.02	0.03	0.33	0.30	0.42	0.40	-0.32	-0.04	0.50	0.49	-0.23	0.06	-0.39	-1.40	-1.24	0.02	0.03
Kurtosis	-0.15	2.85	2.90	0.73	0.91	2.58	2.70	1.24	1.25	2.39	2.61	1.40	1.38	2.36	2.61	-0.15	2.85	2.90	0.73	0.91
Range	70.0	75.0	67.0	64.0	5.19	5.48	6.20	6.60	6.61	6.98	7.49	7.86	6.70	6.86	7.57	7.92	7.21	7.71	6.94	6.69
Max	100	100	100	100	2.30	2.13	2.01	2.20	3.44	3.22	4.44	4.44	3.92	3.59	3.84	4.14	4.31	4.15	3.60	3.58
99%	97	100	97	97	2.08	2.13	1.74	1.89	2.54	3.22	2.81	2.85	2.80	3.59	2.52	2.71	3.06	4.15	2.35	2.33
95%	90	92	92	92	1.56	1.55	1.27	1.37	1.56	1.59	1.54	1.59	1.63	1.66	1.45	1.55	1.77	1.89	1.35	1.32
90%	85	87	89	89	1.19	1.18	1.00	1.06	1.11	1.10	1.05	1.09	1.12	1.11	1.02	1.07	1.20	1.25	0.94	0.91
Q3	79	80	85	84	0.75	0.67	0.63	0.55	0.65	0.56	0.52	0.43	0.62	0.53	0.54	0.45	0.66	0.59	0.50	0.37
Median	70	73	80	80	0.08	0.16	0.16	0.14	0.05	0.10	0.00	0.00	0.00	0.05	0.05	0.02	-0.02	0.04	0.04	0.01
Q1	60	62	75	76	-0.66	-0.65	-0.30	-0.28	-0.61	-0.60	-0.44	-0.38	-0.64	-0.63	-0.38	-0.36	-0.69	-0.71	-0.36	-0.31
10%	50	52	64	66	-1.41	-1.38	-1.32	-1.31	-1.32	-1.26	-1.23	-1.20	-1.27	-1.23	-1.23	-1.21	-1.35	-1.37	-1.13	-1.02
5%	45	46	55	59	-1.78	-1.81	-2.15	-2.03	-1.71	-1.70	-1.78	-1.72	-1.61	-1.60	-1.89	-1.77	-1.70	-1.78	-1.72	-1.48
1%	36	37	40	44	-2.45	-2.47	-3.54	-3.57	-2.51	-2.45	-2.65	-2.80	-2.27	-2.23	-3.08	-3.02	-2.38	-2.45	-2.77	-2.50
Min	30	25	33	36	-2.89	-3.35	-4.19	-4.40	-3.16	-3.77	-3.06	-3.42	-2.79	-3.27	-3.74	-3.78	-2.90	-3.56	-3.33	-3.11
Kolmogorov-Smirnov D	0.06	0.07	0.14	0.14	0.06	0.07	0.14	0.14	0.05	0.05	0.10	0.10	0.05	0.06	0.11	0.11	0.05	0.06	0.11	0.11
Pr > D <	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Cramer-von Mises W-Sq	0.67	1.14	7.44	3.95	0.67	1.14	7.44	3.95	0.47	0.75	3.64	2.78	0.34	0.61	4.66	2.95	0.36	0.63	4.44	2.88
Pr > W-Sq <	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005
Anderson-Darling A-Sq	3.73	5.91	42.01	22.15	3.73	5.91	42.01	22.15	3.09	4.64	19.96	15.14	2.70	4.52	25.70	16.10	2.96	4.83	24.47	15.72
Pr > A-Sq <	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005	0.005

Figure 5.11: Distribution of SIMNORM Model A scores for four cultural groups

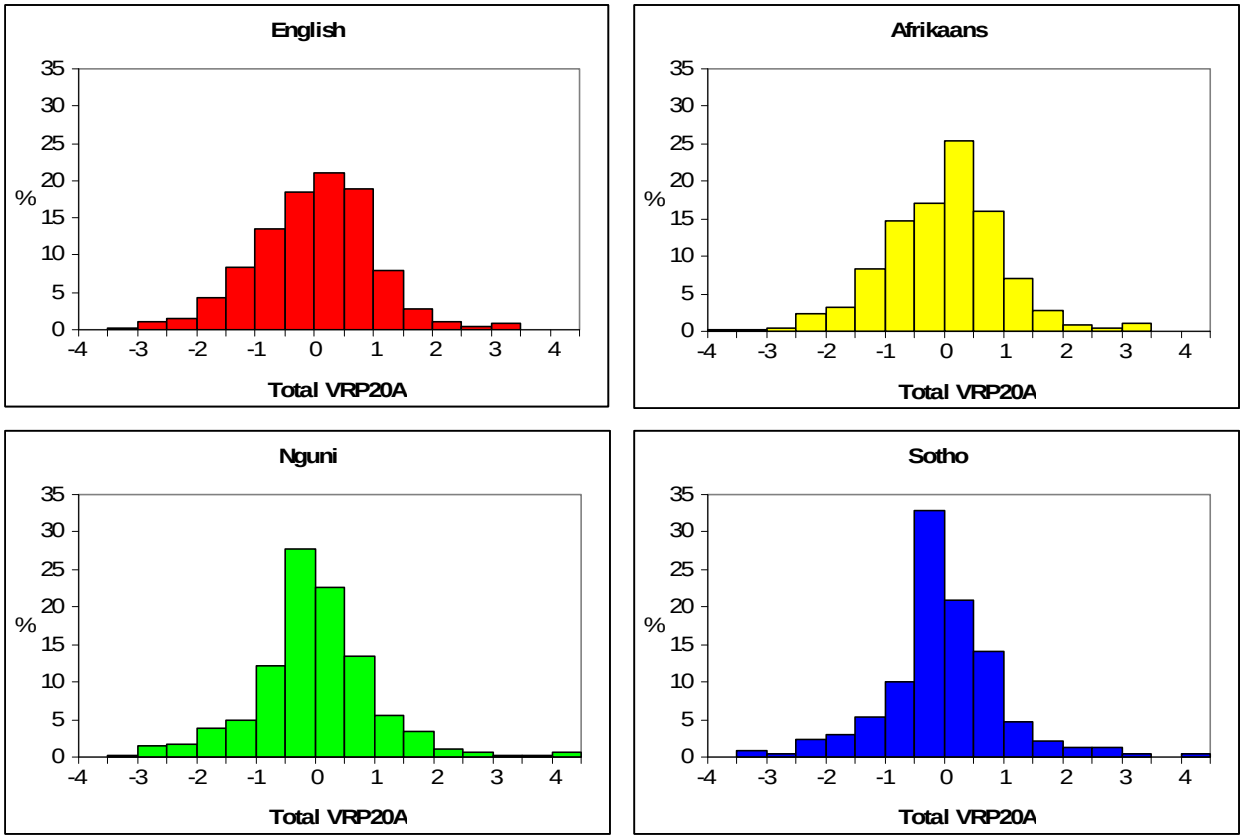


Figure 5.12: Distribution of SIMNORM Model B scores for four cultural groups

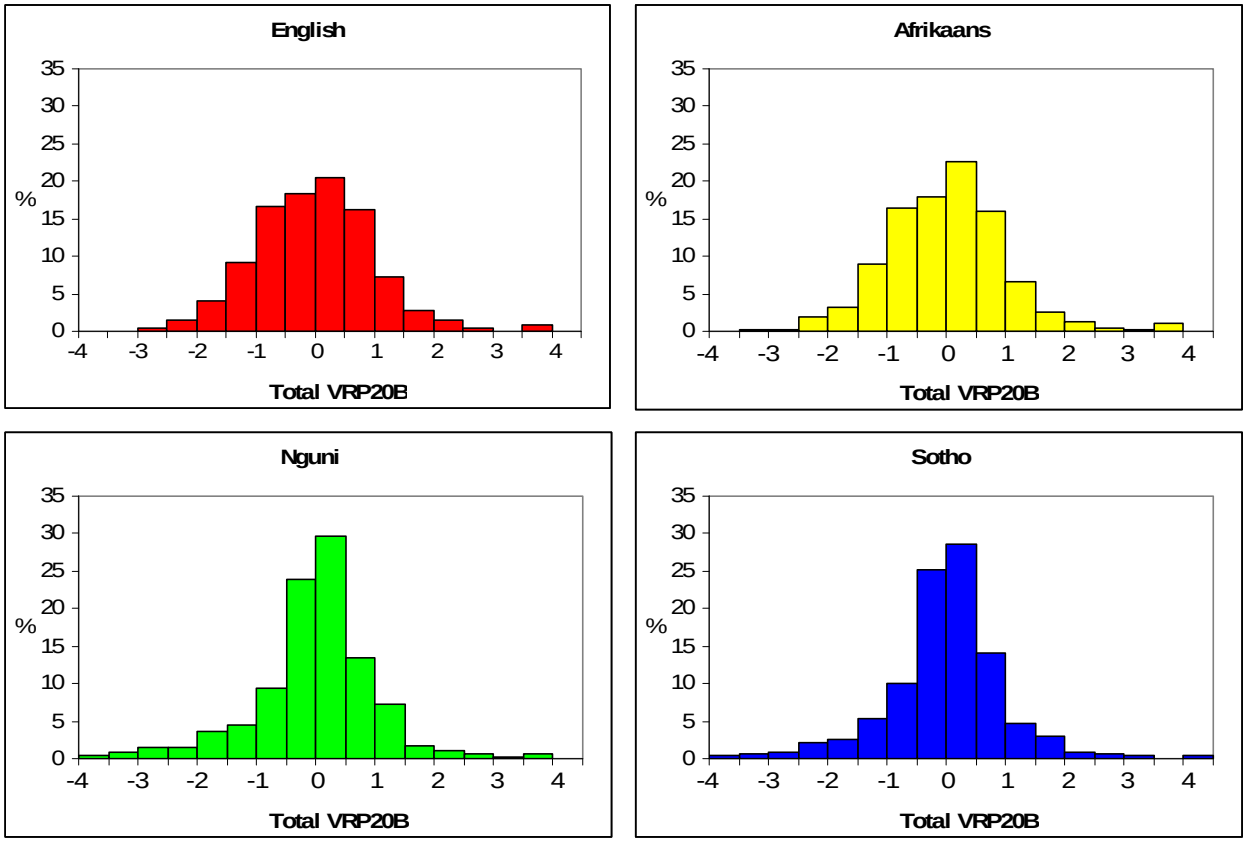


Figure 5.13: Distribution of SIMNORM Model D scores for four cultural groups

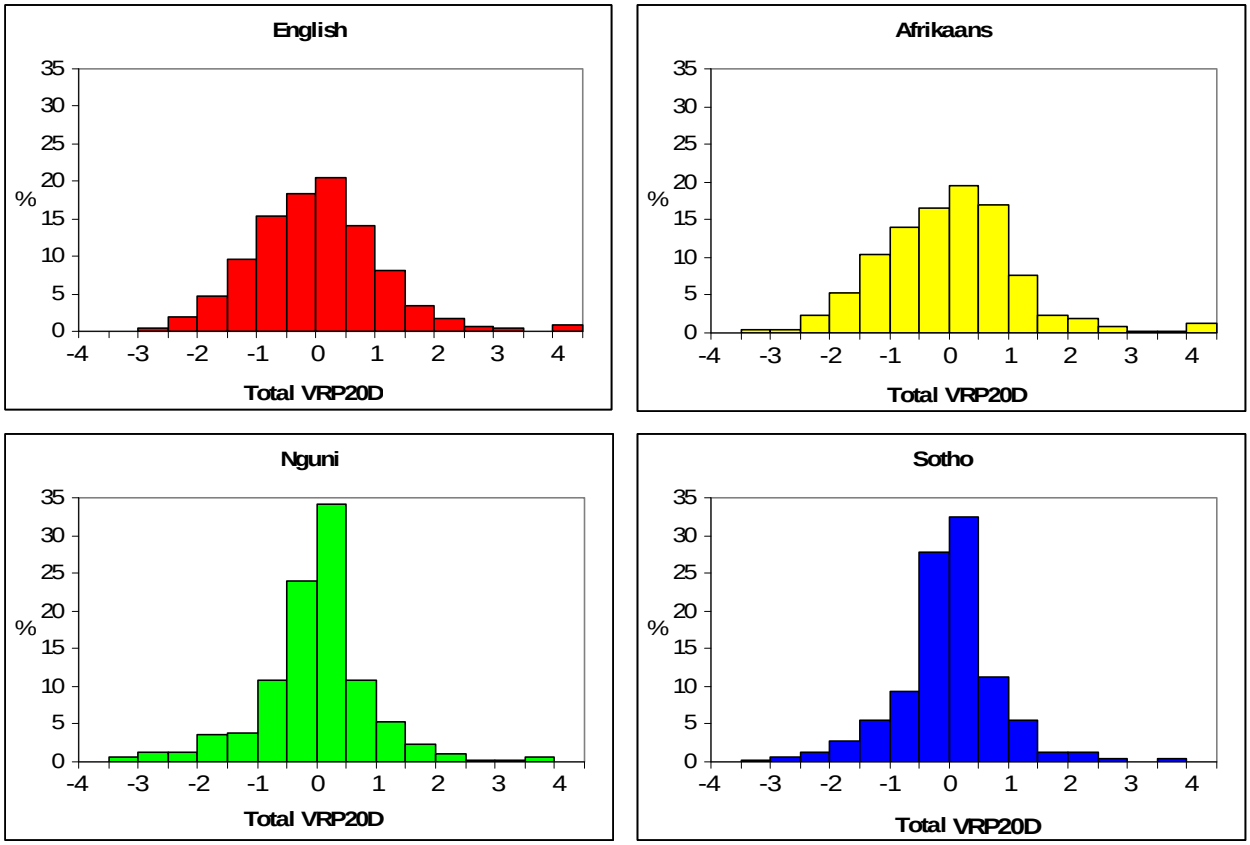


Figure 5.14: Box-and-Whisker Plot of Raw Total VRP20

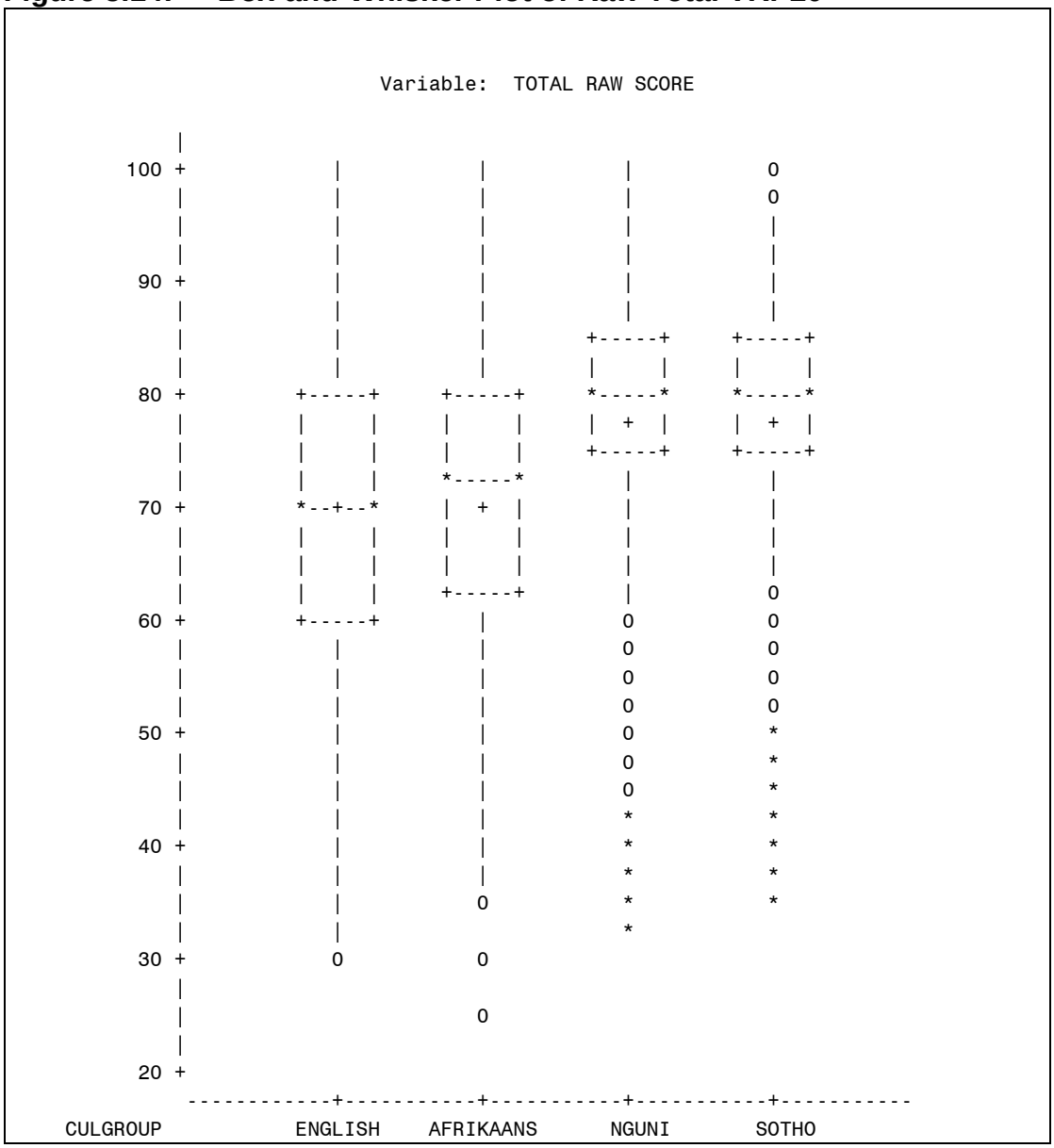


Figure 5.15: Box-and-Whisker Plot of Standardised Total VRP20

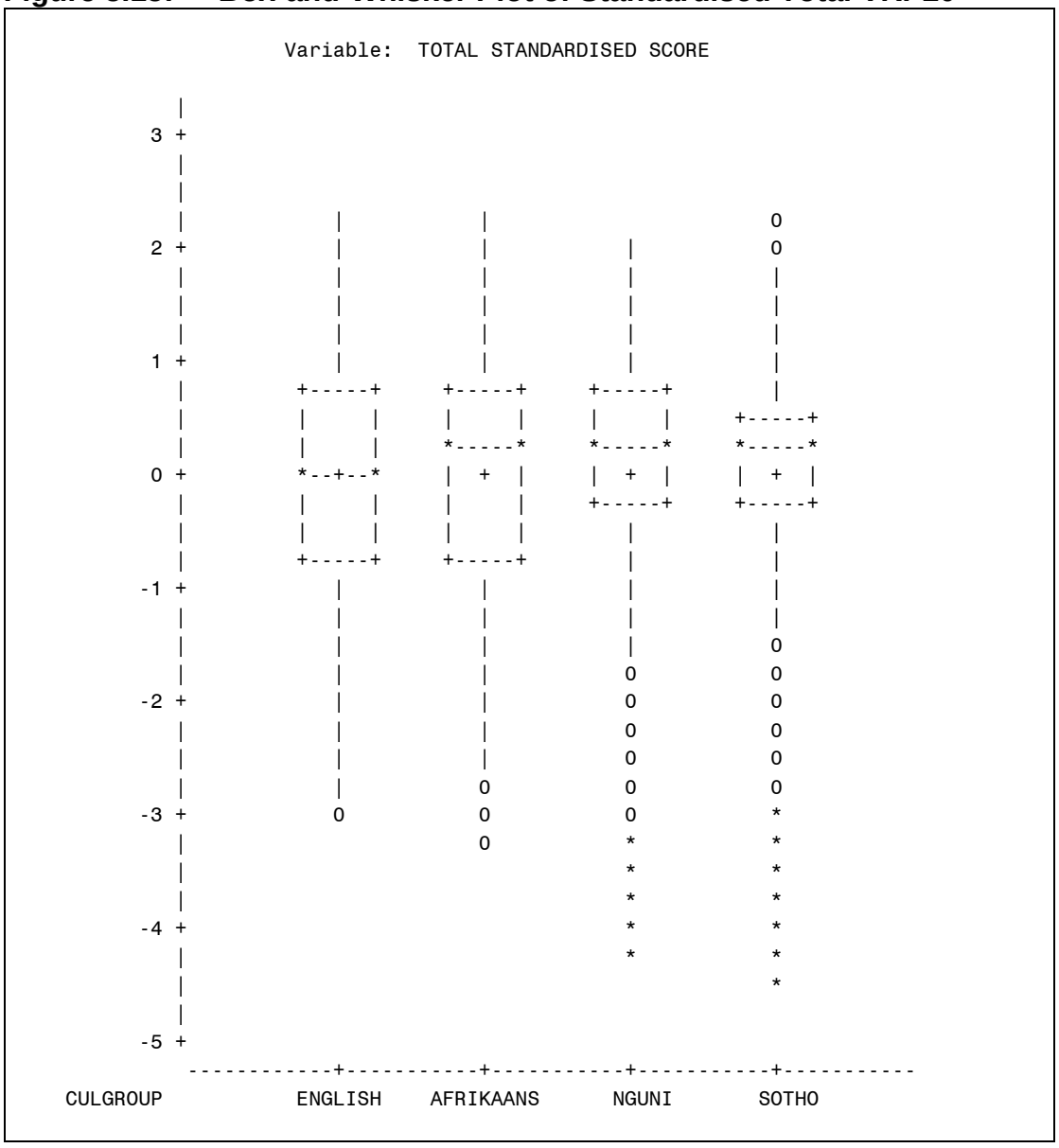


Figure 5.16: Box-and-Whisker Plot of SIMNORM Model A Total VRP20

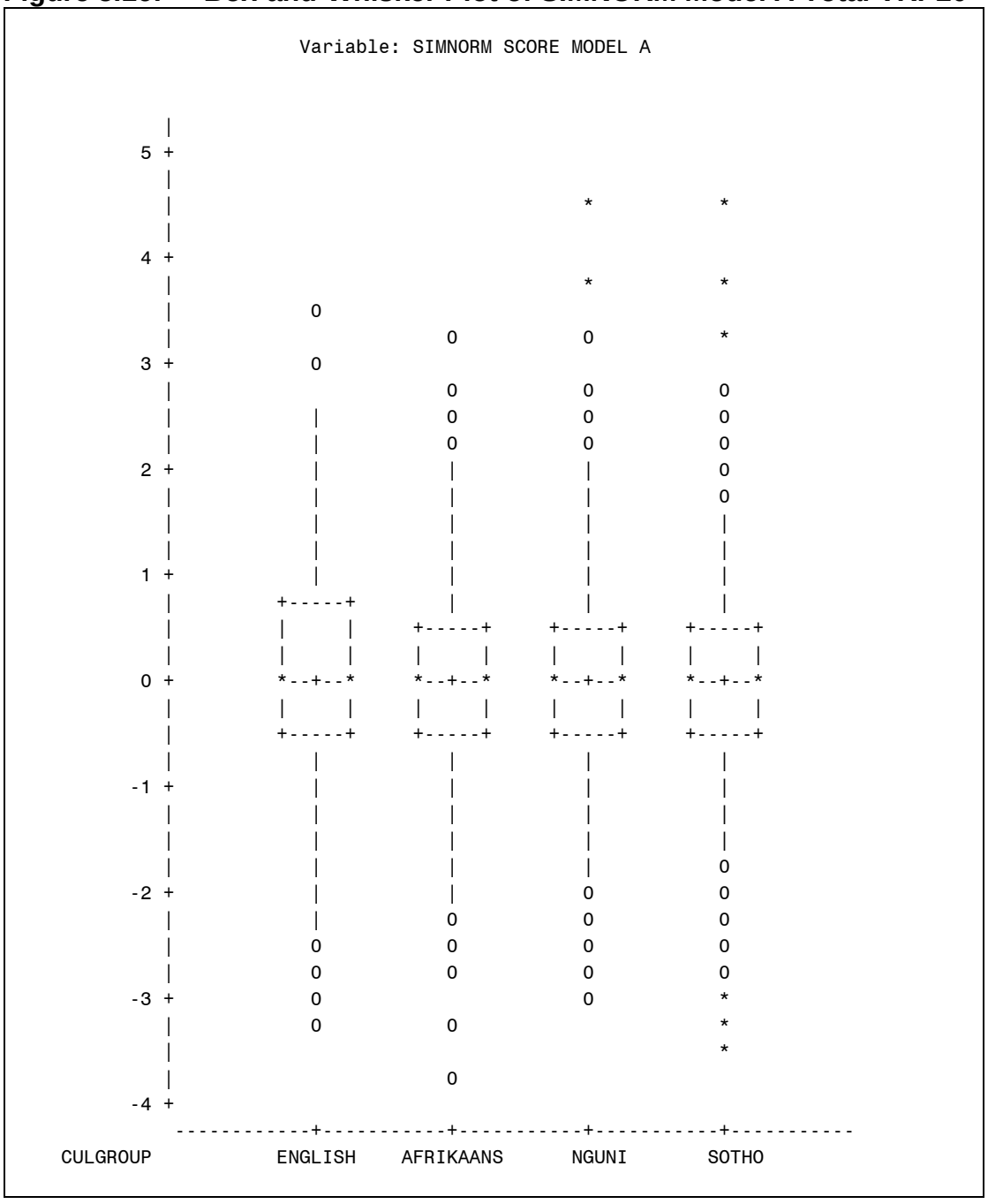


Figure 5.17: Box-and-Whisker Plot of SIMNORM Model B Total VRP20

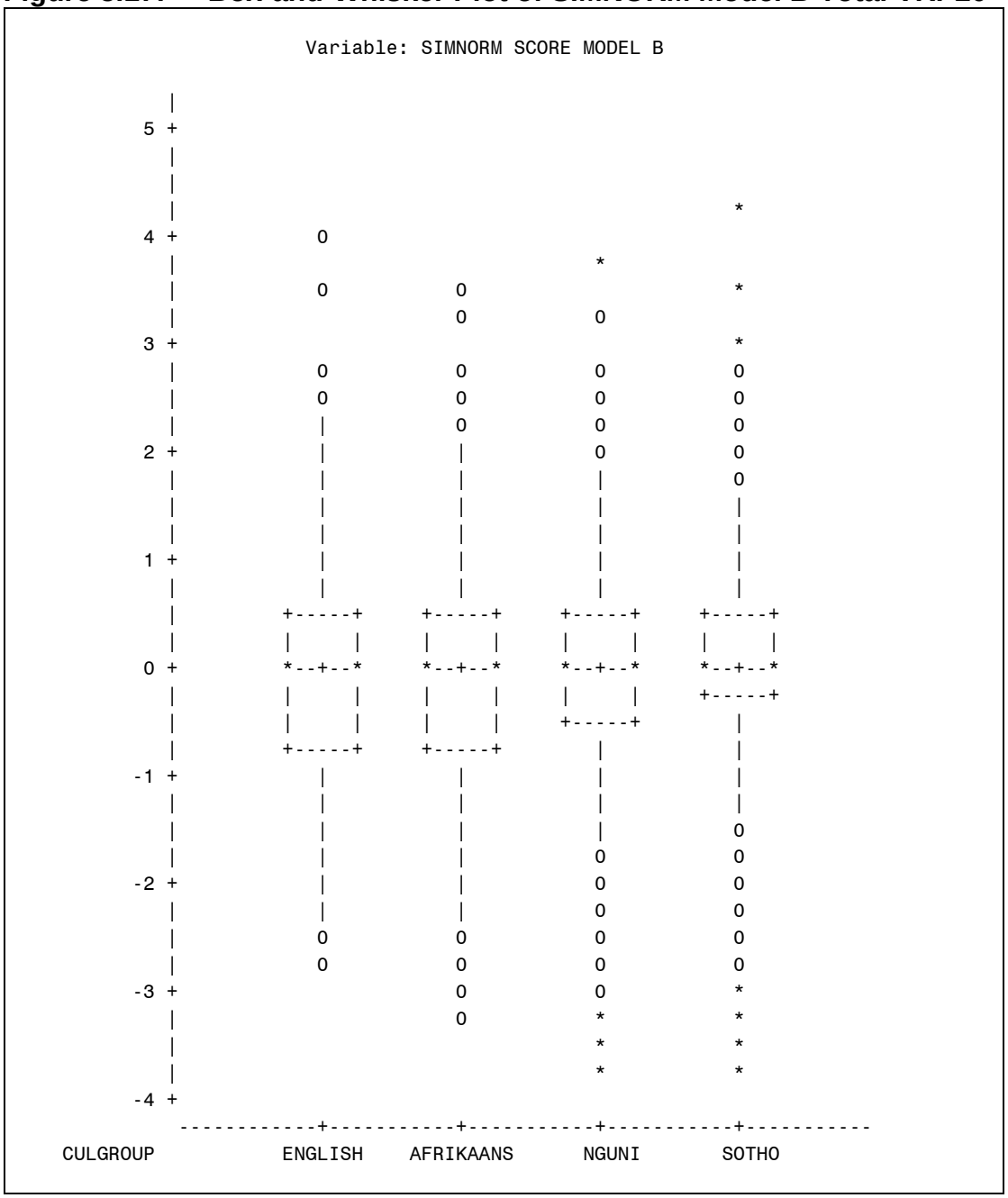
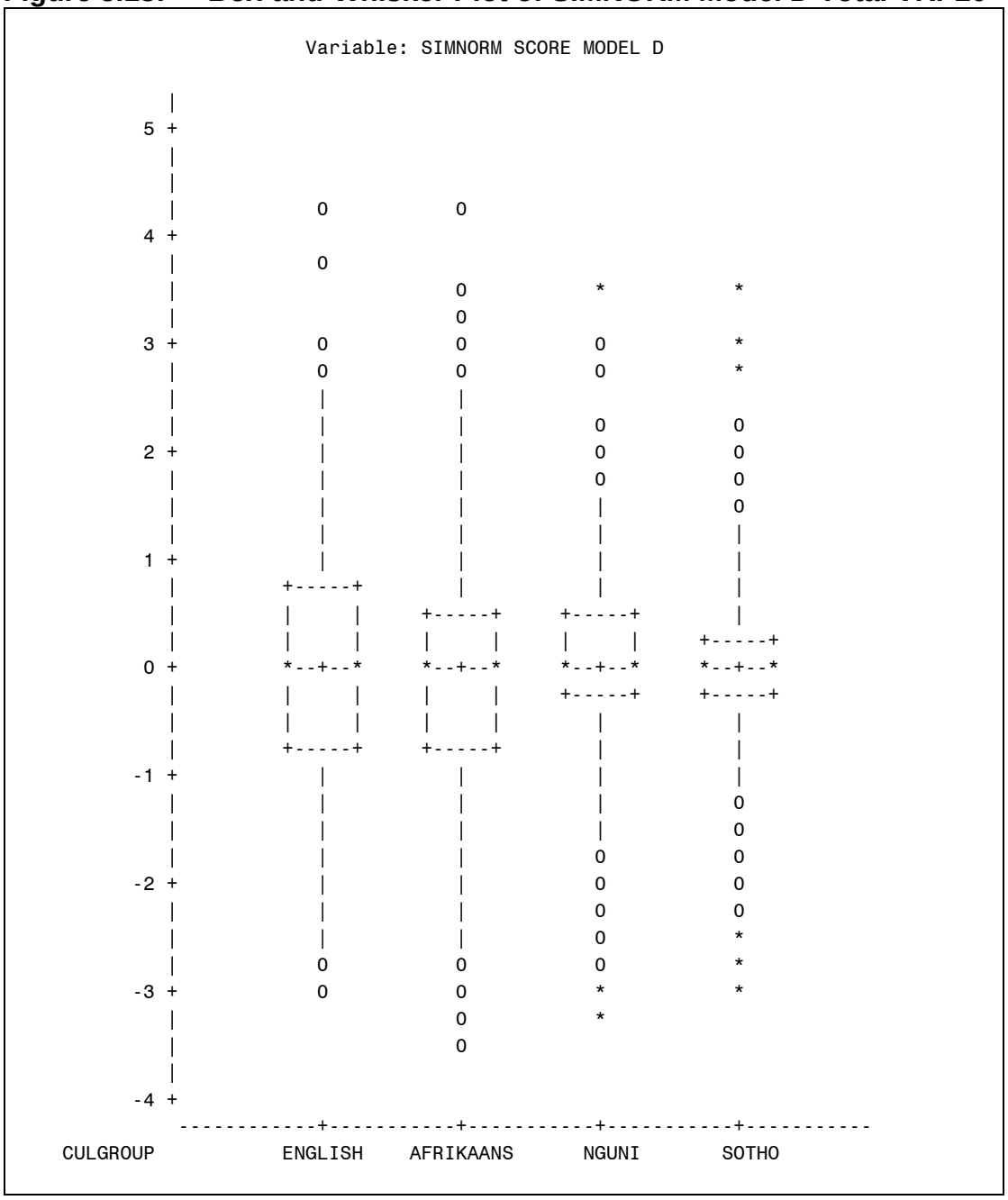


Figure 5.18: Box-and-Whisker Plot of SIMNORM Model D Total VRP20



5.6.2 The effect of SIMNORM on Level-Orientated Analyses

This section covers the results of total score normalisation (SIMNORM) when level-oriented results are involded. The effects of the three SIMNORM models are compared against raw scores and standardised Z-scores. SIMNORM was performed on the liking variable, *Likeability*, the seven subscales of Schlinger (1979), as well as three additional total scores. In total, eleven total score variables were analysed across four cultural groups, and these analyses were done using three normalisation models, as described in Table 5.21. Appendix B contains the detail of the estimated parameters.

Table 5.21: Variables used in SIMNORM

Variable	Description	Calculation
LIKE	*Likeability*	A score between 1 and 10, with 1, does not like the commercial, and 10, like it very much.
V20_1	*Entertainment*	Sum of items VV1, VV2, VV3 and VV4
R20_2	*Confusion*	Sum of items VV8, VV9 and VV10, where VV8, VV9 and VV10 were reverse scored
V20_3	*Relevant News*	Sum of items VV14, VV15 and VV16
V20_4	*Brand Reinforcement*	Sum of items VV17 and VV18
V20_5	*Empathy*	Sum of items VV20, VV21 and VV22
R20_6	*Familiarity*	Sum of reversed coded items VV24 and VV25
R20_7	*Alienation*	Sum of reversed coded items VV27, VV28 and VV29.
V20_8	*Positive Mood*	Sum of items VV1- VV4, VV14-VV16 VV17-VV18, and VV20-VV22
R20_9	*Negative Mood*	Sum of reversed coded items VV8-VV10, VV24-VV25 and VV27-VV29
V20_10	*Overall VRP20*	Sum of all positive items and all reversed negative items (V20_8 + R20_9)

Model A is the least restrictive model with three parameters for each of the four groups, which has in total twelve parameters. Model B has one fixed parameter for all groups, with two unique parameters per cultural group, which gives a total of

nine parameters to be estimated. Model D is the most restrictive (and most parsimonious) model, with two fixed parameters and one unique parameter per group, which results a total of six parameters. The estimated parameters, the fit measures, as well as the descriptive statistics of the normalised scores appear in Appendix B. All the normalisation results are compared to standardised scores, the required base for the purpose of this research.

It is important to determine how normalisation affects the results of level-orientated analyses, such as ANOVA, and to determine how it compares to standardisation. In order to obtain a broad perspective on how normalisation could affect the results when compared to standardisation, the analyses for the ANOVA are reported in four parts. Section 5.6.3 sets out the results with biographic variables as predictors, or as main effects and interaction effects, whereas Section 5.6.4 covers the results with advertising related variables as main effects and interaction effects. Section 5.6.5 reports the effects of normalisation on ANOVA from the perspective of pair-wise comparisons, using the biographic variables as predictors; and in Section 5.6.6 the the pair-wise comparisons are given where advertising variables were used as predictors.

The results of several ANOVAs, and different aspects of ANOVA are represented in each table. Tables 5.22 to 5.25 reveal the significance (p-values) for each main effect and interaction effects using the Type III sums of squares. The advantage of using the Type III sum of squares, is that the Type III sums of squares reflect the results, without any order effects. It reflects the results of the main effect or interaction effect, as if it had been entered into the model last. It therefore reflects the result of a specific effect, given that all the other main effects and interaction effects are already in the model, and therefore it is more suitable for comparison purposes (Searle, 1971).

5.6.3 Significance of main effects and interaction effects in ANOVA using the biographic variables as predictors

Table 5.22 displays the significance (p-values) for each main effect and interactions effect using the Type III sums of squares with *Likeability* as the dependent variable. As expected, the significance of the cultural group is eliminated with standardisation and normalisation, and therefore, the main *Culgroup* effect is significant in the Raw *Likeability* score, but not significant in the standardised *Z-score* or in any of the normalised scores, using Model A, Model B and Model D. It is interesting to note that *Gender* was the only other significant main effect before standardisation or normalisation, and that the standardisation and normalisation did not remove the gender effect.

Table 5.22: SIMNORM results on ANOVA, *Likeability* as dependent variable with biographic predictors

TYPE III SUM OF SQUARES Significance of Main Effects and Interaction Effects for Biographic Variables						
DEPENDENT VARIABLE *Likeability*		Raw score	Standardised Score	Normalised score Model A	Model B	Model D
Source	Df	Prob>F				
Gender	1	0.00	0.00	0.00	0.00	0.00
Age	3	0.13	0.19	0.39	0.40	0.38
Gender*Age	3	0.27	0.29	0.19	0.20	0.20
Income	3	0.62	0.52	0.57	0.56	0.58
Gender*Income	3	0.11	0.08	0.08	0.09	0.10
Age*Income	9	0.33	0.29	0.16	0.15	0.16
Gender*Age*Income	9	0.71	0.75	0.71	0.72	0.71
Culgroup	3	0.00	0.94	0.81	0.85	0.85
Gender*Culgroup	3	0.10	0.18	0.14	0.14	0.11
Age*Culgroup	9	0.12	0.13	0.18	0.19	0.18
Gender*Age*Culgroup	9	0.35	0.41	0.42	0.41	0.41
Income*Culgroup	9	0.25	0.24	0.13	0.13	0.14
Gender*Income*Culgroup	9	0.17	0.12	0.11	0.12	0.12
Age*Income*Culgroup	27	0.80	0.82	0.93	0.93	0.94
Gender*Age*Income*Culgroup	27	0.47	0.50	0.46	0.44	0.45
Age*Income*Culgroup	27	0.83	0.87	0.92	0.94	0.94
Gender*Age*Income*Culgroup	27	0.14	0.12	0.13	0.15	0.15

Overall, the results in Table 5.22 suggest that the results of the ANOVA have been affected by normalisation. The significance levels are all different across the raw scores, standardised scores and the normalised scores. In some cases, the significance diminished with normalisation, for example the main *Age* effect. In

other cases, the significance increased, for example the interaction effect *Gender*Age*. These changes are however not substantial, and from the results one could conclude that the effects of normalisation on the results are very similar to the results obtained by standardisation.

Table 5.23 displays the results of normalisation on ANOVA, using the V20_8, Positive Mood variable as the dependent variable. These results show some interesting patterns. As expected, standardisation and normalisation removed the effect of Cultural Group (*Culgroup*). The main *Gender* effect remained significant after standardisation and normalisation, but standardisation and normalisation removed the significant main *Age* effect from the raw score.

Table 5.23: SIMNORM results on ANOVA, V20_8 (*Positive Mood*) as dependent variable with biographic predictors

TYPE III SUM OF SQUARES Significance of Main Effects and Interaction Effects for Biographic Variables						
DEPENDENT VARIABLE *Positive Mood*	V20_8	Raw score	Standardised score	Normalised score Model A	Model B	Model D
Source	Df	Prob>F				
Gender	1	0.00	0.00	0.00	0.00	0.00
Age	3	0.03	0.05	0.08	0.08	0.06
Gender*Age	3	0.07	0.08	0.04	0.04	0.03
Income	3	0.45	0.51	0.44	0.43	0.41
Gender*Income	3	0.01	0.01	0.02	0.01	0.02
Age*Income	9	0.03	0.04	0.03	0.03	0.03
Gender*Age*Income	9	0.08	0.12	0.03	0.04	0.03
Culgroup	3	0.00	0.98	0.89	0.89	0.87
Gender*Culgroup	3	0.02	0.09	0.08	0.09	0.03
Age*Culgroup	9	0.05	0.08	0.07	0.07	0.05
Gender*Age*Culgroup	9	0.94	0.97	0.90	0.92	0.91
Income*Culgroup	9	0.37	0.41	0.31	0.33	0.36
Gender*Income*Culgroup	9	0.41	0.31	0.29	0.29	0.34
Age*Income*Culgroup	27	0.22	0.37	0.23	0.28	0.23
Gender*Age*Income*Culgroup	27	0.65	0.74	0.66	0.69	0.67

The non-significant *Gender*Age* and *Gender*Age*Income* effects in Table 5.24 (using α=0.05) were not significant in the raw scores or in the standardised scores, but became significant in the normalised scores. The significant *Age*Culgroup* effect became insignificant after standardisation and normalisation.

It is interesting to note that Table 5.23 reflects several differences between standardisation and normalisation. In some instances standardisation shows lower

significance than normalisation, for example in the *Gender*Income* interaction effect. Sometimes standardisation had a higher significance as can be seen with the *Gender*Age* interaction effect and the *Gender*Age*Income* interaction effects. In other instances similar significance between standardisation and normalisation resulted, as with the *Age*Income* effect. When V20_8 (Positive Mood) is used as the dependent variable, the results clearly show that normalisation results are remarkably different from standardisation.

Table 5.24 presents the normalisation results when R20_9 (the sum of the negatively phrased statements or the Negative Mood variable) are used as a dependent variable. The results are also mixed, showing that normalisation affects results in a slightly different manner than standardisation does.

Table 5.24: SIMNORM results on ANOVA, R20_9 (*Negative Mood*) as dependent variable with biographic predictors

TYPE III SUM OF SQUARES Significance of Main Effects and Interaction Effects for Biographic Variables						
DEPENDENT VARIABLE *Negative Mood*	R20_9	Raw score	Standardised Score	Normalised score Model A	Model B	Model D
Source	Df	Prob>F				
Gender	1	0.01	0.01	0.01	0.01	0.01
Age	3	0.04	0.04	0.03	0.03	0.03
Gender*Age	3	0.63	0.64	0.45	0.45	0.45
Income	3	0.02	0.02	0.01	0.01	0.01
Gender*Income	3	0.07	0.08	0.07	0.07	0.06
Age*Income	9	0.04	0.04	0.05	0.05	0.05
Gender*Age*Income	9	0.60	0.68	0.82	0.83	0.82
Culgroup	3	0.00	0.85	0.90	0.90	0.90
Gender*Culgroup	3	0.19	0.23	0.33	0.33	0.30
Age*Culgroup	9	0.13	0.14	0.20	0.21	0.19
Gender*Age*Culgroup	9	0.24	0.28	0.28	0.29	0.28
Income*Culgroup	9	0.55	0.49	0.56	0.57	0.61
Gender*Income*Culgroup	9	0.15	0.21	0.30	0.30	0.26
Age*Income*Culgroup	27	0.15	0.17	0.29	0.32	0.31
Gender*Age*Income*Culgroup	27	0.20	0.25	0.32	0.35	0.33

Table 5.25 shows the significance levels where V20_10 (overall Schlinger VRP20) was used as a dependent variable. This table also confirms that the effects of normalisation are very similar to those of standardisation, but with subtle differences.

Table 5.25: SIMNORM results on ANOVA, V20_10 (Overall Schlinger *VRP20*) as dependent variable with biographic predictors

TYPE III SUM OF SQUARES Significance of Main Effects and Interaction Effects for Biographic Variables						
DEPENDENT VARIABLE *Overall Schlinger VRP20*	V20_10	Raw score	Standardised Score	Normalised score Model A	Model B	Model D
Source	Df	Prob>F				
Gender	1	0.00	0.00	0.00	0.00	0.00
Age	3	0.04	0.06	0.12	0.10	0.06
Gender*Age	3	0.13	0.15	0.05	0.04	0.04
Income	3	0.28	0.20	0.15	0.15	0.20
Gender*Income	3	0.01	0.01	0.01	0.01	0.01
Age*Income	9	0.02	0.02	0.02	0.02	0.02
Gender*Age*Income	9	0.16	0.26	0.20	0.26	0.21
Culgroup	3	0.00	0.93	0.77	0.79	0.79
Gender*Culgroup	3	0.03	0.09	0.09	0.10	0.04
Age*Culgroup	9	0.04	0.07	0.03	0.03	0.02
Gender*Age*Culgroup	9	0.87	0.92	0.72	0.77	0.77
Income*Culgroup	9	0.52	0.46	0.35	0.39	0.48
Gender*Income*Culgroup	9	0.39	0.40	0.42	0.39	0.36
Age*Income*Culgroup	27	0.25	0.40	0.23	0.36	0.30
Gender*Age*Income*Culgroup	27	0.53	0.66	0.57	0.66	0.61

5.6.4 Significance of main effects and interaction effects in ANOVA, using the advertising variables as predictors

Tables 5.26 to 5.29 reflect the significance levels when advertising related variables are used as predictors, in addition to cultural group. These variables included *Genjoy, Adseen*, *Usage* and *Adno*. *Genjoy* is a variable measured on a five-point Likert type question, which indicates how much a respondent enjoys watching television commercials in general. The *Adseen* (Yes/No) variable indicates whether the respondent has seen the advertisement before, and the *Usage* variable refers to whether the respondent uses the advertised product ("Yes/No"). *Adno* is a variable that indicates which of the 24 variables are being evaluated. Table 5.26 sets out the standardisation and normalisation results using *Likeability* as the dependent variable, and using the advertising variables as predictors. The standardisation and normalisations were done across cultural groups, therefore the results all show a very highly significant cultural effect in the raw scores, but no significance after standardisation and normalisation, as expected. It is interesting to note that the strongly significant main effects *Genjoy*, *Adseen* and *Adno* remained significant after standardisation and normalisation. This seems to show that, although the cultural effect has been removed by

standardisation and normalisation, the transformations did not affect other important significant results in this instance. The interaction effect *Genjoy*Adno* was significant for the raw score (α=0.05), but it was not significant after normalisation.

Table 5.26: SIMNORM results on ANOVA, *Likeability* as dependent variable with advertising variables as predictors

TYPE III SUM OF SQUARES Significance of Main Effects and Interaction Effects for Advertising Related Variables						
DEPENDENT VARIABLE *Likeability*		Raw score	Standardised Score	Normalised score Model A	Model B	Model D
Source	Df	Prob>F				
Culgroup	3	0.00	0.61	0.88	0.88	0.89
Genjoy	4	0.00	0.00	0.00	0.00	0.00
Culgroup*Genjoy	12	0.08	0.07	0.11	0.10	0.11
Adseen	1	0.03	0.03	0.02	0.02	0.02
Culgroup*Adseen	3	0.36	0.29	0.44	0.44	0.45
Genjoy*Adseen	4	0.56	0.56	0.51	0.49	0.49
Culgroup*Adseen	10	0.13	0.12	0.11	0.11	0.11
Usage	2	0.23	0.23	0.37	0.37	0.37
Culgroup*Usage	6	0.05	0.08	0.15	0.14	0.13
Genjoy*Usage	8	0.23	0.25	0.30	0.31	0.30
Culgroup*Genjoy*Usage	20	0.23	0.21	0.19	0.19	0.20
Adseen*Usage	2	0.08	0.06	0.08	0.09	0.09
Culgroup*Adseen*Usage	6	0.21	0.20	0.17	0.16	0.16
Genjoy*Usage	7	0.54	0.54	0.57	0.57	0.59
Culgroup*Genjoy*Adseen*Usage	8	0.66	0.61	0.62	0.62	0.62
Adno	15	0.00	0.00	0.00	0.00	0.00
Culgroup*Adno	45	0.00	0.00	0.00	0.00	0.00
Genjoy*Adno	60	0.04	0.06	0.12	0.12	0.12
Culgroup*Genjoy*Adno	125	0.48	0.46	0.46	0.49	0.50
Adseen*Adno	15	0.23	0.17	0.10	0.09	0.10
Culgroup*Adseen*Adno	45	0.29	0.25	0.39	0.37	0.41
Genjoy*Adno	56	0.25	0.37	0.53	0.53	0.52
Culgroup*Genjoy*Adseen*Adno	59	0.55	0.63	0.77	0.76	0.75
Usage*Adno	21	0.34	0.35	0.51	0.51	0.52
Culgroup*Usage*Adno	51	0.31	0.20	0.17	0.18	0.20
Genjoy*Adno	69	0.52	0.53	0.64	0.65	0.65
Culgroup*Genjoy*Usage*Adno	78	0.16	0.16	0.22	0.23	0.22
Adseen*Usage*Adno	21	0.24	0.20	0.38	0.37	0.39
Culgroup*Adseen*Usage*Adno	41	0.08	0.05	0.09	0.09	0.09
Genjoy*Adseen*Usage*Adno	36	0.44	0.48	0.60	0.60	0.59
Culgroup*Genjoy*Adseen*Usage*Adno	14	0.24	0.14	0.10	0.10	0.11

Table 5.27 gives the results with V20_8, the sum of the positively stated items in the VRP20, with the advertising variables as predictors. In this case, the results are more mixed. The main effects *Genjoy* and *Adno*, as well as the interaction

effects *Culgroup*Adno*; *Culgroup*Genjoy*Adno* and *Culgroup*Adseen*Adno,* remained significant after standardisation and normalisation.

Table 5.27: SIMNORM results on ANOVA, V20_8 (*Positive Mood*) as dependent variable with advertising variables as predictors

TYPE III SUM OF SQUARES Significance of Main Effects and Interaction Effects for Advertising Related Variables						
DEPENDENT VARIABLE		RAW	Standardised	Normalised score		
Positive Mood	V20_8	Score	Score	Model A	Model B	Model D
Source	Df	Prob>F				
Culgroup	3	0.00	0.19	0.35	0.32	0.43
Genjoy	4	0.00	0.00	0.00	0.00	0.00
Culgroup*Genjoy	12	0.42	0.35	0.29	0.33	0.28
Adseen	1	0.50	0.35	0.54	0.48	0.59
Culgroup*Adseen	3	0.49	0.31	0.29	0.29	0.39
Genjoy*Adseen	4	0.76	0.70	0.88	0.86	0.88
Culgroup*Adseen	10	0.40	0.41	0.24	0.27	0.24
Usage	2	0.48	0.58	0.49	0.46	0.37
Culgroup*Usage	6	0.29	0.41	0.47	0.53	0.44
Genjoy*Usage	8	0.08	0.09	0.23	0.21	0.22
Culgroup*Genjoy*Usage	20	0.60	0.70	0.61	0.62	0.49
Adseen*Usage	2	0.32	0.28	0.13	0.15	0.15
Culgroup*Adseen*Usage	6	0.09	0.12	0.10	0.11	0.09
Genjoy*Usage	7	0.49	0.50	0.31	0.35	0.34
Culgroup*Genjoy*Adseen*Usage	8	0.59	0.72	0.63	0.66	0.58
Adno	15	0.00	0.00	0.00	0.00	0.00
Culgroup*Adno	45	0.00	0.00	0.00	0.00	0.00
Genjoy*Adno	60	0.05	0.08	0.07	0.05	0.02
Culgroup*Genjoy*Adno	125	0.01	0.00	0.00	0.00	0.00
Adseen*Adno	15	0.53	0.57	0.29	0.35	0.33
Culgroup*Adseen*Adno	45	0.03	0.02	0.03	0.03	0.03
Genjoy*Adno	56	0.06	0.16	0.09	0.10	0.03
Culgroup*Genjoy*Adseen*Adno	59	0.03	0.09	0.06	0.07	0.02
Usage*Adno	21	0.91	0.88	0.88	0.89	0.92
Culgroup*Usage*Adno	51	0.53	0.48	0.49	0.52	0.53
Genjoy*Adno	69	0.12	0.21	0.35	0.34	0.25
Culgroup*Genjoy*Usage*Adno	78	0.31	0.35	0.31	0.29	0.21
Adseen*Usage*Adno	21	0.61	0.56	0.36	0.45	0.51
Culgroup*Adseen*Usage*Adno	41	0.29	0.12	0.23	0.21	0.36
Genjoy*Adseen*Usage*Adno	36	0.46	0.58	0.26	0.37	0.33
Culgroup*Genjoy*Adseen*Usage*Adno	14	0.63	0.68	0.61	0.62	0.61

It is interesting to note that the interaction terms *Genjoy*Adno* and *Culgroup*Genjoy*Adseen*Adno* were significant in the raw scores, and were only significant again in the Model D normalised score.

In Table 5.28, the results of R20_9, (the sum of the reverse scored negatively stated items in the VRP20, or as phased by the researcher, the *Negative Mood* variable), are given, where the advertising variables were used as predictors. In Table 5.28, the results are also mixed. The main effects *Genjoy* and *Adno*, as well as the interaction effects *Genjoy*Usage*, *Culgroup*Adno*; and *Culgroup*Genjoy*Adseen*Adno*; remained significant after standardisation and normalisation.

It is interesting to note in Table 5.28, that the main effect *Adseen* and interaction terms *Culgroup*Adseen*Usage* and *Culgroup*Adseen*Adno* were significant in the raw scores and the standardised scores, but not in the normalised scores. The interaction term *Culgroup*Adseen*Usage*Adno* was not significant in the raw score, but became significant in the standardised and normalised scores.

Table 5.28: SIMNORM results on ANOVA, R20_9 (*Negative Mood*) as dependent variable with advertising variables as predictors

TYPE III SUM OF SQUARES Significance of Main Effects and Interaction Effects for Advertising Related Variables						
DEPENDENT VARIABLE		Raw	Standardised	Normalised score		
Negative Mood	R20_9	Score	Score	Model A	Model B	Model D
Source	DF	Prob>F				
Culgroup	3	0.16	0.01	0.03	0.03	0.04
Genjoy	4	0.00	0.00	0.00	0.00	0.00
Culgroup*Genjoy	12	0.28	0.26	0.16	0.15	0.16
Adseen	1	0.04	0.03	0.06	0.05	0.06
Culgroup*Adseen	3	0.33	0.35	0.42	0.41	0.40
Genjoy*Adseen	4	0.13	0.10	0.18	0.17	0.18
Culgroup*Adseen	10	0.61	0.58	0.66	0.64	0.64
Usage	2	0.43	0.47	0.35	0.34	0.32
Culgroup*Usage	6	0.48	0.52	0.25	0.24	0.21
Genjoy*Usage	8	0.01	0.01	0.01	0.01	0.01
Culgroup*Genjoy*Usage	20	0.15	0.14	0.16	0.16	0.16
Adseen*Usage	2	0.62	0.59	0.47	0.48	0.48
Culgroup*Adseen*Usage	6	0.02	0.03	0.07	0.07	0.06
Genjoy*Usage	7	0.44	0.47	0.45	0.44	0.42
Culgroup*Genjoy*Adseen*Usage	8	0.09	0.09	0.12	0.12	0.12
Adno	15	0.00	0.00	0.00	0.00	0.00
Culgroup*Adno	45	0.01	0.00	0.01	0.01	0.01
Genjoy*Adno	60	0.26	0.33	0.28	0.28	0.22
Culgroup*Genjoy*Adno	125	0.35	0.40	0.45	0.44	0.37
Adseen*Adno	15	0.23	0.24	0.30	0.31	0.32
Culgroup*Adseen*Adno	45	0.05	0.04	0.06	0.06	0.06
Genjoy*Adno	56	0.53	0.61	0.39	0.39	0.35
Culgroup*Genjoy*Adseen*Adno	59	0.00	0.00	0.00	0.00	0.00
Usage*Adno	21	0.27	0.25	0.50	0.49	0.51
Culgroup*Usage*Adno	51	0.70	0.70	0.80	0.80	0.79
Genjoy*Adno	69	0.13	0.12	0.38	0.38	0.38
Culgroup*Genjoy*Usage*Adno	78	0.12	0.10	0.13	0.13	0.14
Adseen*Usage*Adno	21	0.97	0.96	0.93	0.93	0.94
Culgroup*Adseen*Usage*Adno	41	0.05	0.03	0.03	0.03	0.04
Genjoy*Adseen*Usage*Adno	36	0.59	0.55	0.75	0.74	0.76
Culgroup*Genjoy*Adseen*Usage*Adno	14	0.37	0.26	0.27	0.26	0.32

Table 5.29 shows the results of V20_10, the overall VRP as the dependent variable.

Table 5.29: SIMNORM results on ANOVA, V20_10 (*Overall Schlinger VRP20*) as dependent variable with advertising variables as predictors

TYPE III SUM OF SQUARES Significance of Main Effects and Interaction Effects for Advertising Related Variables						
DEPENDENT VARIABLE		Raw	Standardised	Normalised score		
Overall Schlinger VRP20	V20_10	Score	Score	Model A	Model B	Model D
Source	Df	Prob>F				
Culgroup	3	0.00	0.03	0.10	0.08	0.16
Genjoy	4	0.00	0.00	0.00	0.00	0.00
Culgroup*Genjoy	12	0.49	0.42	0.54	0.47	0.42
Adseen	1	0.17	0.10	0.26	0.18	0.24
Culgroup*Adseen	3	0.52	0.38	0.43	0.41	0.46
Genjoy*Adseen	4	0.55	0.44	0.89	0.85	0.91
Culgroup*Adseen	10	0.60	0.57	0.59	0.55	0.51
Usage	2	0.38	0.47	0.42	0.35	0.26
Culgroup*Usage	6	0.37	0.49	0.31	0.37	0.27
Genjoy*Usage	8	0.02	0.02	0.04	0.04	0.02
Culgroup*Genjoy*Usage	20	0.35	0.41	0.32	0.36	0.27
Adseen*Usage	2	0.36	0.33	0.12	0.15	0.16
Culgroup*Adseen*Usage	6	0.02	0.04	0.03	0.04	0.03
Genjoy*Usage	7	0.49	0.53	0.35	0.40	0.37
Culgroup*Genjoy*Adseen*Usage	8	0.28	0.37	0.35	0.36	0.28
Adno	15	0.00	0.00	0.00	0.00	0.00
Culgroup*Adno	45	0.00	0.00	0.00	0.00	0.00
Genjoy*Adno	60	0.05	0.11	0.10	0.08	0.02
Culgroup*Genjoy*Adno	125	0.02	0.02	0.01	0.01	0.00
Adseen*Adno	15	0.49	0.47	0.22	0.33	0.35
Culgroup*Adseen*Adno	45	0.02	0.02	0.04	0.04	0.04
Genjoy*Adno	56	0.21	0.37	0.25	0.27	0.14
Culgroup*Genjoy*Adseen*Adno	59	0.00	0.00	0.00	0.00	0.00
Usage*Adno	21	0.85	0.84	0.89	0.90	0.90
Culgroup*Usage*Adno	51	0.64	0.60	0.62	0.66	0.66
Genjoy*Adno	69	0.08	0.11	0.35	0.29	0.24
Culgroup*Genjoy*Usage*Adno	78	0.10	0.10	0.05	0.04	0.02
Adseen*Usage*Adno	21	0.77	0.70	0.41	0.56	0.66
Culgroup*Adseen*Usage*Adno	41	0.09	0.02	0.02	0.02	0.06
Genjoy*Adseen*Usage*Adno	36	0.55	0.57	0.30	0.44	0.46
Culgroup*Genjoy*Adseen*Usage*Adno	14	0.55	0.48	0.28	0.29	0.37

Tables 5.26 to 5.29 all contained mixed results, indicating that there are broad similarities between standardisation and normalisation, with subtle differences.

5.6.5 The effect of SIMNORM on post-hoc comparisons in ANOVA using the biographic variables as predictors

The results of the post-hoc comparisons in ANOVA, when the biographic variables are used as predictors, are shown in Tables 5.30 to 5.33. Several conclusions can be drawn from the results.

The effect of culture has been completely removed with standardisation and normalisation, which is the reason why the four culture categories each have a mean equal to zero. These changes have altered the other effects in different ways. Consider Gender throughout these four tables. Females had higher scores than males for all four dependent variables (which could also be a matter of bias), and in both the standardised and normalised scores, the females still possessed significantly higher standardised and normalised scores than males. For the *Age* variable, the differences were significant between the age groups in the raw scores, and these differences were less significant or mostly insignificant after standardisation and normalisation. The significant differences between the income groups in the raw scores remained significant after standardisation and after normalisation.

This leads one to the conclusion that standardisation and normalisation have unpredictable effects on the outcome of level-oriented results. The effect of this is that, if this bias is ignored in analysis, other important differences may be disguised, or differences that appear significant may become insignificant after taking bias effects into account.

Table 5.30: Results of SIMNORM on Post-Hoc comparisons in ANOVA, with *Likeability* as dependent variable

RESULTS OF NORMALISATION ON POST-HOC COMPARISONS IN ANOVA Biographic Variables																		
			Raw			Standardised			Normalised Score									
DEPENDENT VARIABLE			Score			Score			Model A			Model B			Model D			
Likeability		N	Mean	Duncan		Mean	Duncan		Mean	Duncan		Mean	Duncan		Mean	Duncan		
GENDER	Female	2521	7.49	A		0.10	A		0.10	A		0.10	A		0.10	A		
	Male	2279	7.07	B		-0.11	B		-0.12	B		-0.11	B		-0.12	B		
AGE	25-34 yrs	1818	7.41		A	0.04	A		0.03	A		0.04	A		0.04	A		
	50+ yrs	693	7.27	B	A	0.01	A		0.02	A		0.02	A		0.02	A		
	16-24 yrs	535	7.30	B	A	0.01	A		0.00	A		0.00	A		0.00	A		
	35-49 yrs	1754	7.17	B		-0.05	A		-0.05	A		-0.05	A		-0.05	A		
INCOME	D	451	7.51		A	0.07		A	0.07		A	0.07		A	0.07		A	
	C	1420	7.29	B		0.01	B	A	0.01	B	A	0.02	B	A	0.02	B	A	
	B	1384	7.34	B	A	0.01	B	A	0.00	B	A	0.00	B	A	0.00	B	A	
	A	1545	7.19	B		-0.04	B		-0.04	B		-0.04	B		-0.04	B		
CULGROUP	Sotho	1014	7.80	A		0.00	A		0.00	A		0.00	A		0.00	A		
	Nguni	1386	7.65	A		0.00	A		0.00	A		0.00	A		0.00	A		
	Afrikaans	1193	6.97	B		0.00	A		-0.01	A		0.00	A		0.00	A		
	English	1207	6.77	C		0.00	A		0.00	A		0.00	A		0.00	A		

Table 5.31: Results of SIMNORM on Post-Hoc comparisons in ANOVA, with V20_8, *Positive Mood* as dependent variable

RESULTS OF NORMALISATION ON POST-HOC COMPARISONS IN ANOVA Biographic Variables															
DEPENDENT VARIABLE *Positive Mood*			Raw Score		Standardised Score			Normalised Score							
								Model A			Model B			Model D		
V20_8		N	Mean	Duncan	Mean	Duncan		Mean	Duncan		Mean	Duncan		Mean	Duncan	
GENDER	FEMALE	2521	42.91	A	0.08	A		0.08	A		0.08	A		0.08	A	
	MALE	2279	41.44	B	-0.09	B		-0.09	B		-0.08	B		-0.09	B	
AGE	25-34 yrs	1818	43.03	A	0.06		A	0.05		A	0.05		A	0.06		A
	16-24 yrs	535	41.83	B	-0.03	B	A	-0.04	B	A	-0.04	B	A	-0.04	B	
	50+ yrs	693	41.74	B	0.00	B	A	0.01	B	A	0.01	B	A	0.01	B	A
	35-49 yrs	1754	41.67	B	-0.05	B		-0.05	B		-0.05	B		-0.05	B	
INCOME	D	451	43.90	A	0.11		A	0.10		A	0.10		A	0.09		A
	C	1420	42.38	B	0.04	B	A	0.04		A	0.04	B	A	0.04		A
	B	1384	42.38	B	0.00	B	C	0.00	B	A	0.00	B	C	0.00	B	A
	A	1545	41.42	C	-0.07		C	-0.07	B		-0.07		C	-0.07	B	
CULGROUP	Sotho	1014	45.72	A	0.00	A		0.00	A		0.00	A		0.00	A	
	Nguni	1386	45.62	A	0.00	A		0.00	A		0.00	A		0.00	A	
	Afrikaans	1193	39.38	B	0.00	A		0.00	A		0.00	A		0.00	A	
	English	1207	38.17	C	0.00	A		0.00	A		0.00	A		0.00	A	

Table 5.32: Results of SIMNORM on Post-Hoc comparisons in ANOVA, with R20_9, *Negative Mood* as dependent variable

RESULTS OF NORMALISATION ON POST-HOC COMPARISONS IN ANOVA Biographic Variables												
DEPENDENT VARIABLE *Negative Mood*			Raw Score		Standardised Score		Normalised Score					
							Model A		Model B		Model D	
R20_9		N	Mean	Duncan	Mean	Duncan	Mean	Duncan	Mean	Duncan	Mean	Duncan
GENDER	FEMALE	2521	32.26	A	0.06	A	0.06	A	0.06	A	0.06	A
	MALE	2279	31.59	B	-0.07	B	-0.07	B	-0.07	B	-0.07	B
AGE	25-34 yrs	1818	32.16	A	0.03	A	0.02	A	0.02	A	0.02	A
	16-24 yrs	535	31.92	B A	0.00	A	-0.01	A	-0.01	A	-0.01	A
	35-49 yrs	1754	31.90	B A	0.00	A	0.01	A	0.01	A	0.01	A
	50+ yrs	693	31.48	B	-0.07	A	-0.07	A	-0.07	A	-0.07	A
INCOME	D	451	33.04	A	0.19	A	0.19	A	0.19	A	0.19	A
	B	1384	31.96	B	-0.01	B	0.00	B	0.00	B	0.00	B
	A	1545	31.79	B	-0.02	B	-0.02	B	-0.02	B	-0.02	B
	C	1420	31.74	B	-0.03	B	-0.04	B	-0.04	B	-0.04	B
CULGROUP	Sotho	1014	32.95	A	0.00	A	0.00	A	0.00	A	0.00	A
	Nguni	1386	32.62	A	0.00	A	0.00	A	0.00	A	0.00	A
	Afrikaans	1193	31.46	B	0.00	A	0.00	A	0.00	A	0.00	A
	English	1207	30.79	C	0.00	A	0.00	A	0.00	A	0.00	A

Table 5.33: Results of SIMNORM on Post-Hoc comparisons in ANOVA, with V20_10, *Overall Schlinger VRP20* as dependent variable

RESULTS OF NORMALISATION ON POST-HOC COMPARISONS IN ANOVA Biographic Variables												
DEPENDENT VARIABLE OVERALL SCHLINGER			Raw Score		Standardised Score		Normalised Score					
							Model A		Model B		Model D	
V20_10		N	Mean	Duncan	Mean	Duncan	Mean	Duncan	Mean	Duncan	Mean	Duncan
GENDER	FEMALE	2521	75.17	A	0.08	A	0.08	A	0.08	A	0.08	A
	MALE	2279	73.03	B	-0.09	B	-0.09	B	-0.09	B	-0.09	B
AGE	25-34 yrs	1818	75.19	A	0.05	A	0.04	A	0.04	A	0.05	A
	16-24 yrs	535	73.74	B	-0.02	A	-0.02	A	-0.02	A	-0.03	A
	35-49 yrs	1754	73.57	B	-0.04	A	-0.03	A	-0.03	A	-0.03	A
	50+ yrs	693	73.22	B	-0.03	A	-0.02	A	-0.02	A	-0.02	A
INCOME	D	451	76.94	A	0.15	A	0.15	A	0.16	A	0.15	A
	B	1384	74.34	B	-0.01	B	0.00	B	0.00	B	0.00	B
	C	1420	74.12	B	0.01	B	0.01	B	0.01	B	0.01	B
	A	1545	73.21	B	-0.05	B	-0.06	B	-0.05	B	-0.06	B
CULGROUP	Sotho	1014	78.67	A	0.00	A	0.00	A	0.00	A	0.00	A
	Nguni	1386	78.23	A	0.00	A	0.00	A	0.00	A	0.00	A
	Afrikaans	1193	70.84	B	0.00	A	0.00	A	0.00	A	0.00	A
	English	1207	68.96	C	0.00	A	0.00	A	0.00	A	0.00	A

5.6.6 The effect of SIMNORM on post-hoc comparisons in ANOVA, using the advertising variables as predictors

Tables 5.34 to 5.37 present the results of normalisation on the post-hoc comparisons in ANOVA when the advertising variables are used as predictors. It should be noted that the full sample was used in the normalisation, but, due to missing values among the advertising variables, the effect of normalisation has been eroded. There are, however, interesting observations to be made from the results. The very significant differences between the cultural groups that were present in the raw scores have been completely or largely removed after standardisation and normalisation, which is to be expected, since this was intended. The highly significant differences between the levels of *Genjoy* remained highly significant after standardisation and normalisation, and the order has not been affected. The significant differences between the *Adseen* categories remained significant in three of the four dependent variables investigated.

Significant differences between different levels became more significant after standardisation and normalisation in three of the four dependent variables analysed, and less significant in the other case. The significant differences in *Adno* remained significant, and mostly in the same order for all four dependent variables. Most of the normalisation results were very similar to the standardisation results.

The implications of these observations regarding the outcome of the different results, using the different models, is that the equating methods of standardisation and SIMNORM have their largest effect at removing cultural bias, but the greatest advantage is that the entire data set does not become distorted. The major advantage is that it gives the remaining effects, given that the effects of cultural differences were “removed” from the data. This is where the biggest advantage of data equation methods, such as standardisation and SIMNORM, lies. It is not a method to replace traditional analyses, but as a supplementary method, information in the data may offer deeper insight and reveal more subtleties than would otherwise be the case.

Table 5.34: Results of SIMNORM on Post-Hoc comparisons in ANOVA, *Likeability* as dependent variable

RESULTS OF NORMALISATION ON POST-HOC COMPARISONS IN ANOVA Advertising Variables												
DEPENDENT VARIABLE			Raw Scores		Standardised Scores		Normalised Scores					
							MODEL A		MODEL B		MODEL D	
Likeability		N	Mean	Duncan	Mean	Duncan	Mean	Duncan	Mean	Duncan	Mean	Duncan
CULTURAL	Nguni	974	7.69	A	0.02	A	0.02	A	0.02	A	0.02	A
GROUP	Afrikaans	838	6.99	B	0.01	A	-0.01	A	0.00	A	0.00	A
	Sotho	726	7.78	A	-0.01	A	-0.02	A	-0.02	A	-0.02	A
	English	862	6.67	C	-0.05	A	-0.05	A	-0.05	A	-0.05	A
GENENJOY	StrAgree	719	7.88	A	0.18	A	0.18	A	0.18	A	0.17	A
	Agree	1560	7.39	B	0.00	B	-0.01	B	-0.01	B	-0.01	B
	Neither	610	7.16	B	0.05	B A	0.03	B	0.04	B	0.04	B
	Disagree	327	6.35	C	-0.29	C	-0.29	C	-0.28	C	-0.28	C
	StrDisagree	184	6.01	D	-0.44	D	-0.45	D	-0.44	D	-0.45	D
ADSEEN	Yes	1940	7.44	A	0.08	A	0.07	A	0.07	A	0.07	A
	No	1460	7.06	B	-0.12	B	-0.13	B	-0.12	B	-0.12	B
USAGE	NA	600	7.57	A	0.13	A	0.14	A	0.14	A	0.14	A
	Non-User	310	7.32	B	0.02	B	0.01	B	0.01	B	0.01	B
	User	1339	7.28	B	-0.08	C B	-0.09	C B	-0.09	C B	-0.09	C B
	Other	1151	7.11	B	-0.11	C	-0.12	C	-0.12	C	-0.12	C

Table 5.34 (continued): Results of SIMNORM on Post-Hoc comparisons in ANOVA, *Likeability* as dependent variable

ADNO	23	200	8.15			A		0.41			A		0.44			A		0.45			A		0.45			A								
	16	200	8.05	B		A		0.34	B		A		0.39		B		A		0.39		B		A		0.40		B		A				
	19	200	7.84	B		A		0.29	B		A		0.28		B		A	C	0.28		B		A	C	0.28		B		A	C			
	17	200	7.80	B		A		0.27	B		A		0.23		B			C	0.23		B			C	0.22		B			C			
	20	200	7.64	B		C		0.18	B		C		0.13				D	C	0.13				D	C	0.13				D	C			
	12	200	7.40	D		C		0.05	D		C		-0.01		E		D		-0.01		E		D		-0.01		E		D				
	13	200	7.31	D		C	E	-0.01	D		C	E	-0.02		E		D	F	-0.02		E		D		-0.02		E		D				
	15	200	7.29	D		C	E	-0.02	D		C	E	-0.03		E		D	F	-0.02		E		D	F	-0.02		E		D	F			
	11	200	7.19	D		F	E	-0.04	D			E	-0.05		E	G	D	F	-0.05		E	G	D	F	-0.05		E	G	D	F			
	14	200	7.15	D		F	E	-0.07	D		F	E	-0.06		E	G	D	F	-0.06		E	G	D	F	-0.06		E	G	D	F			
	10	200	7.03	D	G	F	E	-0.11	D	G	F	E	-0.10		E	G	H	F	-0.10		E	G	H	F	-0.10		E	G	H	F			
	9	200	6.96		G	F	E	-0.14	D	G	F	E	-0.16	I	E	G	H	F	-0.16	I	E	G	H	F	-0.16	I	E	G	H	F			
	4	200	6.92		G	F	E	-0.20		G	F	E	-0.19	I	E	G	H	F	-0.18	I	E	G	H	F	-0.18	I	E	G	H	F			
	5	200	6.88		G	F	E	-0.20		G	F	E	-0.23	I		G	H	F	-0.23	I		G	H	F	-0.23	I		G	H	F			
	8	200	6.79		G	F		-0.26		G	F		-0.26	I		G	H		-0.26	I		G	H		-0.26	I		G	H				
	3	200	6.72		G			-0.27		G	F		-0.28	I			H		-0.28	I			H		-0.28	I			H				
	2	200	6.66		G			-0.31		G			-0.33	I					-0.32	I					-0.33	I							

Table 5.35: Results of SIMNORM on Post-Hoc comparisons in ANOVA, with V20_8, *Positive Mood* as dependent variable

RESULTS OF NORMALISATION ON POST-HOC COMPARISONS IN ANOVA Advertising Variables																	
			Raw Scores			Standardised Scores			Normalised Scores								
									MODEL A			MODEL B			MODEL D		
DEPENDENT VARIABLE																	
Positive Mood		N	Mean		Duncan	Mean		Duncan	Mean		Duncan	Mean		Duncan	Mean		Duncan
CULTURAL GROUP	Nguni	974	46.03	A		0.06		A	0.04		A	0.04		A	0.04		A
	Sotho	726	45.60	A		0.02	B	A	0.01	B	A	0.01	B	A	0.01	B	A
	Afrikaans	838	39.55	B		-0.02	B	A	-0.01	B	A	-0.01	B	A	-0.01	B	A
	English	862	37.59	C		-0.06	B		-0.07	B		-0.07	B		-0.08	B	
GENENJOY	StrAgree	719	46.72	A		0.31	A		0.33	A		0.33	A		0.32	A	
	Agree	1560	42.98	B		-0.01	B		-0.02	B		-0.03	B		-0.02	B	
	Neither	610	40.16	C		-0.02	B		-0.04	B		-0.04	B		-0.03	B	
	Disagree	327	36.98	D		-0.28	C		-0.26	C		-0.26	C		-0.28	C	
	StrDisagree	184	34.01	E		-0.56	D		-0.54	D		-0.54	D		-0.58	D	
ADSEEN	Yes	1940	42.74	A		0.08	A		0.07	A		0.07	A		0.06	A	
	No	1460	41.50	B		-0.11	B		-0.10	B		-0.11	B		-0.10	B	
USAGE	Non-User	310	43.78	A		0.13		A	0.15	A		0.14		A	0.12		A
	NA	600	43.11	A		0.03	B	A	0.02	B		0.02		B	0.02	B	A
	User	1339	42.16	B		-0.06	B	C	-0.08	C		-0.08	C	B	-0.07	B	C
	Other	1151	41.36	B		-0.08		C	-0.09	C		-0.09	C		-0.09		C

Table 5.35 (continued): Results of SIMNORM on Post-Hoc comparisons in ANOVA, with V20_8, *Positive Mood* as dependent variable

ADNO	16	200	45.94				A			0.41			A			0.42			A			0.41			A			0.43			A				
	19	200	44.46	B			A			0.29	B		A			0.26	B		A			0.26	B		A			0.24			B				
	15	200	44.30	B						0.24	B		A			0.23	B					0.23	B		A			0.23			B				
	23	200	44.23	B						0.21	B		A			0.16	B		C			0.17	B		C			0.18	C		B				
	17	200	43.69	B			C			0.18	B		C			0.14	B		C			0.15	B		C			0.14	C		B		D		
	12	200	43.55	B			C			0.14	B		C			0.08	B		C	D		0.09	B		C	D		0.10	C		E	B		D	
	8	200	42.15	E			C		D	-0.01	D		C			0.08	B		C	D		0.07	B		C	D		0.01	C		E	F		D	
	5	200	42.00	E					D	-0.01	D		C			-0.03	E		C	D		-0.03	E		C	D		-0.03			E	F		D	
	20	200	41.88	E					D	-0.01	D		C			-0.03	E		C	D		-0.04	E		C	D		-0.04			E	F		D	
	14	200	41.83	E						-0.02	D		C			-0.06	E			D		-0.06	E			D		-0.06	G		E	F		D	
	4	200	41.68	E			F			-0.08	D		E			-0.08	E		F	D		-0.08	E		F	D		-0.08	G		E	F			
	10	200	41.53	E			F			-0.09	D		E			-0.11	E		F	D		-0.11	E		F	D		-0.10	G		E	F		H	
	2	200	41.11	E			F			-0.12	D		E			-0.13	E		F	D		-0.11	E		F	D		-0.11	G			F		H	
	11	200	40.49	E		H	F		G	-0.17	D		E			-0.14	E		F			-0.14	E		F			-0.18	G		I	F		H	
	9	200	40.07			H	F		G	-0.22			E			-0.22	E		F	G		-0.22	E		F	G		-0.24	G		I			H	
	3	200	39.71			H			G	-0.27	F		E			-0.28			F	G		-0.27			F	G		-0.28			I			H	
	13	200	38.89			H			G	-0.44	F					-0.38				G		-0.40				G		-0.36			I				

Table 5.36: Results of SIMNORM on Post-Hoc comparisons in ANOVA, with R20_9, *Negative Mood* as dependent variable

RESULTS OF NORMALISATION ON POST-HOC COMPARISONS IN ANOVA Advertising Variables														
			Raw Scores			Standardised Scores		Normalised Scores						
								MODEL A		MODEL B		MODEL D		
DEPENDENT VARIABLE														
Negative Mood		N	Mean	Duncan		Mean	Duncan	Mean	Duncan	Mean	Duncan	Mean	Duncan	
CULTURAL GROUP	Nguni	974	32.66	A		0.01	A	-0.01	A	0.00	A	-0.01	A	
	Sotho	726	32.93	A		-0.01	A	-0.01	A	-0.01	A	-0.01	A	
	Afrikaans	838	31.30	B		-0.03	A	-0.03	A	-0.04	A	-0.04	A	
	English	862	30.34	C		-0.08	A	-0.09	A	-0.09	A	-0.10	A	
GENENJOY	StrAgree	719	33.04	A		0.12	A	0.11	A	0.11	A	0.11		A
	Agree	1560	32.17	B		0.01	A	0.00	A	0.00	A	-0.01	B	A
	Neither	610	31.40	C		-0.01	A	-0.02	A	-0.02	A	-0.02	B	
	Disagree	327	29.45	D		-0.35	B	-0.36	B	-0.36	B	-0.37		C
	StrDisagree	184	29.23	D		-0.38	B	-0.37	B	-0.37	B	-0.38		C
ADSEEN	Yes	1940	31.83	A		-0.02	A	-0.03	A	-0.03	A	-0.03	A	
	No	1460	31.74	A		-0.04	A	-0.04	A	-0.04	A	-0.05	A	
USAGE	User	1339	32.35		A	0.10	A	0.09	A	0.09	A	0.09	A	
	Other	1151	31.43	B	C	-0.09	B	-0.09	B	-0.09	B	-0.10	B	
	Non-User	310	31.85	B	A	-0.11	B	-0.14	B	-0.14	B	-0.14	B	
	NA	600	31.24		C	-0.15	B	-0.17	B	-0.17	B	-0.17	B	

Table 5.36 (continued): Results of SIMNORM on Post-Hoc comparisons in ANOVA, with R20_9, *Negative Mood* as dependent variable

| | | |

Table 5.37: Results of SIMNORM on Post-Hoc comparisons in ANOVA, with V20_10, *Overall Schlinger VRP20* as dependent variable

RESULTS OF NORMALISATION ON POST-HOC COMPARISONS IN ANOVA Advertising Variables												
			Raw Scores		Standardised Scores		Normalised Scores					
							MODEL A		MODEL B		MODEL D	
DEPENDENT VARIABLE												
Overall Schlinger		N	Mean	Duncan	Mean	Duncan	Mean	Duncan	Mean	Duncan	Mean	Duncan
CULTURAL	Nguni	974	78.70	A	0.04	A	0.02	A	0.03	A	0.03	A
GROUP	Afrikaans	838	70.85	B	0.00	B A	0.00	B A	-0.01	B A	-0.01	B A
	Sotho	726	78.53	A	-0.01	B A	-0.02	B A	-0.02	B A	-0.01	B A
	English	862	67.93	C	-0.08	B	-0.08	B	-0.09	B	-0.10	B
GENENJOY	StrAgree	719	79.76	A	0.26	A	0.27	A	0.28	A	0.28	A
	Agree	1560	75.15	B	0.00	B	-0.03	B	-0.03	B	-0.02	B
	Neither	610	71.56	C	-0.02	B	-0.03	B	-0.04	B	-0.04	B
	Disagree	327	66.44	D	-0.34	C	-0.32	C	-0.32	C	-0.35	C
	StrDisagree	184	63.24	E	-0.55	D	-0.52	D	-0.52	D	-0.56	D
ADSEEN	Yes	1940	74.57	A	0.05	A	0.03	A	0.03	A	0.03	A
	No	1460	73.24	B	-0.09	B	-0.08	B	-0.09	B	-0.09	B
USAGE	User	1339	74.50	A	0.06	A	0.05	A	0.05	A	0.05	A
	NA	600	74.35	A	0.02	A	0.02	A	0.01	A	0.00	B A
	Other	1151	72.79	B	-0.09	B	-0.09	B	-0.09	B	-0.10	B
	Non-User	310	75.63	A	-0.09	B	-0.11	B	-0.11	B	-0.10	B

Table 5.37 (continued): Results of SIMNORM on Post-Hoc comparisons in ANOVA, with V20_10, *Overall Schlinger VRP20* as dependent variable

ADNO	19	200	78.11		A			0.35		A			0.34		A			0.34		A			0.32		A
	17	200	77.24		A			0.27		A			0.24	B	A			0.24	B	A			0.23		A
	16	200	77.21		A			0.20		A			0.20	B	A			0.19	B	A			0.23		A
	12	200	76.76		A			0.20		A			0.16	B	A	C		0.17	B	A	C		0.18		A
	23	200	76.48		A			0.17		A			0.16	B	A	C		0.17	B	A	C		0.17		A
	15	200	76.02		A			0.15	B	A			0.11	B	D	C		0.12	B	D	C		0.13		A
	20	200	73.58		B			-0.02	B	C			-0.02	E	D	C		-0.02	E	D	C		-0.05		B
	11	200	73.43	C	B			-0.03	B	C	D		-0.06	E	D	F		-0.07	E	D	F		-0.07	C	B
	8	200	73.41	C	B			-0.07	E	C	D		-0.07	E	D	F		-0.07	E	D	F		-0.09	C	B
	3	200	72.67	C	B			-0.09	E	C	D		-0.08	E	D	F		-0.09	E		F		-0.10	C	B
	5	200	72.64	C	B			-0.11	E	C	D		-0.09	E		F		-0.09	E		F		-0.10	C	B
	2	200	72.63	C	B			-0.11	E	C	D		-0.11	E		F		-0.10	E		F		-0.13	C	B
	10	200	72.27	C	B			-0.13	E	C	D		-0.14	E		F		-0.14	E		F		-0.13	C	B
	14	200	71.94	C	B			-0.21	E	C	D		-0.21	E		F		-0.20	E		F		-0.18	C	B
	4	200	71.55	C	B			-0.24	E		D		-0.24			F		-0.24			F		-0.25	C	B
	13	200	71.17	C	B			-0.25	E				-0.25			F		-0.27			F		-0.26	C	B
	9	200	70.88	C				-0.27	E				-0.26			F		-0.27			F		-0.27	C	

5.6.7 Correlations between raw scores, standardised scores and SIMNORM scores

Table 5.38 contains the Pearson correlation coefficients between the raw scores, the standardised scores and the normalised scores. The correlation coefficients between the raw scores and standardised z-scores ranged between 0.93 and 0.99. These high coefficients are expected, since the z-score is a linear transformation of the raw score. The correlation coefficients between the raw scores and the normalised scores ranged between 0.88 and 0.98. The decline in this coefficient is expected, since the normalisation functions are non-linear transformations of the raw scores. The correlation coefficients between the z-scores and the normalised scores ranged between 0.96 and 0.99. This indicates that normalisation scores are in many cases very similar to z-scores. The differences are mainly in the extreme values. The correlation coefficients between the normalised scores ranged between 0.99 and 1.00, indicating that the differences between the three normalisation models are almost non-existent.

Table 5.38: Pearson correlation coefficients between raw scores, standardised scores and SIMNORM scores

Label		Variable	LIKE	LIKEZ	LIKEA	LIKEB	LIKED
LIKING on a 10-point scale	Raw Score	LIKE	1.00				
STANDARDISED LIKING	Z-score	LIKEZ	0.97	1.00			
	MODEL A	LIKEA	0.95	0.98	1.00		
NORMALISED LIKING	MODEL B	LIKEB	0.95	0.98	1.00	1.00	
	MODEL C	LIKED	0.95	0.98	1.00	1.00	1.00
Label		Variable	V20_1	V20_1Z	V20_1A	V20_1B	V20_1D
VRP20 Entertainment	Raw Score	V20_1	1.00				
STANDARDISED ENTERTAINMENT	Z-score	V20_1Z	0.95	1.00			
	MODEL A	V20_1A	0.93	0.97	1.00		
NORMALISED ENTERTAINMENT	MODEL B	V20_1B	0.93	0.97	1.00	1.00	
	MODEL C	V20_1D	0.94	0.96	0.99	0.99	1.00
Label		Variable	R20_2	R20_2Z	R20_2A	R20_2B	R20_2D
VRP20 Confusion	Raw Score	R20_2	1.00				
STANDARDISED CONFUSION	Z-score	R20_2Z	0.99	1.00			
	MODEL A	R20_2A	0.96	0.96	1.00		
NORMALISED CONFUSION	MODEL B	R20_2B	0.96	0.96	1.00	1.00	
	MODEL C	R20_2D	0.96	0.96	1.00	1.00	1.00
Label		Variable	V20_3	V20_3Z	V20_3A	V20_3B	V20_3D
VRP20 Relevant News	Raw Score	V20_3	1.00				
STANDARDISED RELEVANT NEWS	Z-score	V20_3Z	0.89	1.00			
	MODEL A	V20_3A	0.88	0.98	1.00		
NORMALISED RELEVANT NEWS	MODEL B	V20_3B	0.88	0.99	1.00	1.00	
	MODEL C	V20_3D	0.88	0.99	1.00	1.00	1.00

Table 5.38 (continued):

Label		Variable	V20_4	V20_4Z	V20_4A	V20_4B	V20_4D
VRP20 Brand Reinforcement	Raw Score	V20_4	1.00				
STANDARDISED BRAND REINFORCEMENT	Z-score	V20_4Z	0.99	1.00			
	MODEL A	V20_4A	0.98	0.99	1.00		
NORMALISED BRAND REINFORCEMENT	MODEL B	V20_4B	0.98	0.99	1.00	1.00	
	MODEL C	V20_4D	0.98	0.98	1.00	1.00	1.00
Label		Variable	V20_5	V20_5Z	V20_5A	V20_5B	V20_5D
VRP20 Empathy	Raw Score	V20_5	1.00				
STANDARDISED EMPATHY	Z-score	V20_5Z	0.94	1.00			
	MODEL A	V20_5A	0.92	0.98	1.00		
NORMALISED EMPATHY	MODEL B	V20_5B	0.93	0.99	1.00	1.00	
	MODEL C	V20_5D	0.93	0.97	0.99	0.99	1.00
Label		Variable	R20_6	R20_6Z	R20_6A	R20_6B	R20_6D
VRP20 REVERSED Familiarity	Raw Score	R20_6	1.00				
STANDARDISED FAMILIARITY	Z-score	R20_6Z	0.97	1.00			
	MODEL A	R20_6A	0.96	0.99	1.00		
NORMALISED FAMILIARITY	MODEL B	R20_6B	0.96	0.98	1.00	1.00	
	MODEL C	R20_6D	0.96	0.98	0.99	1.00	1.00
Label		Variable	R20_7	R20_7Z	R20_7A	R20_7B	R20_7D
VRP20 REVERSED Alienation	Raw Score	R20_7	1.00				
STANDARDISED ALIENATION	Z-score	R20_7Z	0.95	1.00			
	MODEL A	R20_7A	0.92	0.97	1.00		
NORMALISED ALIENATION	MODEL B	R20_7B	0.92	0.97	1.00	1.00	
	MODEL C	R20_7D	0.92	0.97	1.00	1.00	1.00
Label		Variable	V20_8	V20_8Z	V20_8A	V20_8B	V20_8D
VRP20 POSITIVE DIMENSIONS	Raw Score	V20_8	1.00				
STANDARDISED POSITIVE ITEMS	Z-score	V20_8Z	0.91	1.00			
	MODEL A	V20_8A	0.89	0.97	1.00		
NORMALISED POSITIVE ITEMS	MODEL B	V20_8B	0.89	0.98	1.00	1.00	
	MODEL C	V20_8D	0.90	0.96	0.99	0.99	1.00
Label		Variable	R20_9	R20_9Z	R20_9A	R20_9B	R20_9D
VRP20 NEGATIVE DIMENSIONS	Raw Score	R20_9	1.00				
STANDARDISED NEGATIVE ITEMS	Z-score	R20_9Z	0.98	1.00			
	MODEL A	R20_9A	0.96	0.98	1.00		
NORMALISED NEGATIVE ITEMS	MODEL B	R20_9B	0.96	0.98	1.00	1.00	
	MODEL C	R20_9D	0.97	0.98	1.00	1.00	1.00
Label		Variable	V20_10	V20_10Z	V20_10A	V20_10B	V20_10D
VRP20 OVERALL TOTAL SCORE NEGATIVES REVERSED	Raw Score	V20_10	1.00				
STANDARDISED TOTAL	Z-score	V20_10Z	0.93	1.00			
	MODEL A	V20_10A	0.91	0.97	1.00		
NORMALISED TOTAL	MODEL B	V20_10B	0.91	0.98	1.00	1.00	
	MODEL C	V20_10D	0.92	0.97	0.99	0.99	1.00

5.6.8 Results of SIMNORM on regression analyses

The results in Tables 5.39 to 5.41 reflect the results of normalisation in a regression analysis. The results of five regression models are presented. Model 1 has *Likeability* (LIKE) as the dependent variable in the raw form, with the raw scores of the seven subscales of the Schlinger VRP20 as predictors. Model 2 has the *Standardised Likeability* (LIKEZ) as the dependent variable, and the standardised seven subscales of the VRP20 as predictor variables. Models 3, 4 and 5 have the *Normalised Likeability* scores (LIKEA, LIKEB and LIKED) as predictors, using models A, B and D as dependent variables. In each of these models, the corresponding normalisation model was used to normalise the seven subscales used as predictors in the respective models.

Table 5.39 gives the fit statistics of the five regression models. All the models were highly significant, and the R-Square dropped from 0.52 to somewhere in the region of 0.4 as a result of normalisation. The R-Square for the standardised model is 0.47. It is interesting to note that the RMSE is very similar across the standarised model (LIKEZ) and the normalised models (LIKEA, LIKEB and LIKED).

Table 5.39: Fit statistics of five regression models

Dependent Variable	LIKE	LIKEZ	LIKEA	LIKEB	LIKED
F-value	732.700	613.700	463.100	456.800	465.700
Prob > F	0.000	0.000	0.000	0.000	0.000
Root MSE	1.430	0.726	0.773	0.775	0.772
Dependent Mean	7.290	0.000	0.000	0.000	0.000
R-Square	0.520	0.473	0.403	0.400	0.405
Adjusted R-Square	0.520	0.472	0.403	0.399	0.404

The estimated parameters and their corresponding levels of significance are presented in Tables 5.40 and 5.41. The standardised and normalised coefficients all have the same direction and magnitude. Predictors that were significant in the raw scores, remained significant after standardisation and after normalisation. The only insignificant predictor was Brand Reinforcement, and that remained insignificant after standardisation and normalisation. It seems that standardisation and normalisation did not have sufficient effect to change the overall covariance in the data set.

Table 5.40: Parameter estimates of five regression models

Dependent Variable	LIKE	LIKEZ	LIKEA	LIKEB	LIKED	
Predictor Variables	Regression Parameter Estimates					Label
Intercept	-0.182	0.000	-0.002	0.000	0.000	Intercept
V20_1	0.247	0.427	0.387	0.389	0.400	VRP20 Entertainment
R20_2	0.078	0.097	0.095	0.094	0.095	VRP20 REVERSED Confusion
V20_3	0.053	0.082	0.089	0.085	0.083	VRP20 Relevant News
V20_4	0.021	0.013	0.003	0.006	0.007	VRP20 Brand Reinforcement
V20_5	0.079	0.133	0.148	0.146	0.138	VRP20 Empathy
R20_6	0.077	0.064	0.058	0.062	0.051	VRP20 REVERSED Familiarity
R20_7	0.070	0.095	0.072	0.066	0.067	VRP20 REVERSED Alienation

Table 5.41: Significance of the estimates of five regression models

Dependent Variable	LIKE	LIKEZ	LIKEA	LIKEB	LIKED	
Predictor Variables	Significance of Regression Parameter Estimates					Label
Intercept	0.248	1.000	0.828	1.000	1.000	Intercept
V20_1	0.000	0.000	0.000	0.000	0.000	VRP20 Entertainment
R20_2	0.000	0.000	0.000	0.000	0.000	VRP20 REVERSED Confusion
V20_3	0.000	0.000	0.000	0.000	0.000	VRP20 Relevant News
V20_4	0.123	0.272	0.781	0.614	0.556	VRP20 Brand Reinforcement
V20_5	0.000	0.000	0.000	0.000	0.000	VRP20 Empathy
R20_6	0.000	0.000	0.000	0.000	0.000	VRP20 REVERSED Familiarity
R20_7	0.000	0.000	0.000	0.000	0.000	VRP20 REVERSED Alienation

5.6.9 Summary of SIMNORM results

The results presented in this section indicate that SIMNORM as a method of reducing bias in a data set can be seen to be very similar to standardisation, but with several added advantages. Standardisation is only a linear transformation, which implies that the shape of the distribution is not changed. SIMNORM removes the bias, but it also changes the shape of the distribution to a more equivalent shape. Although the results are in many respects very similar to standardisation in level-orientated analyses, there are subtle differences. It is not possible to predict how normalisation will affect the outcome of the significance of predictors in ANOVA, for example, but it could mean that, when the bias in one variable is removed, insignificant predictors may become significant after SIMNORM. Similarly, significant predictors may become insignificant.

It is recommended that when the researcher uses normalisation, it is always used as an additional analysis, to examine the effect of the removal of bias. It remains difficult (if not

impossible) to determine how much bias was true cultural bias, and how much resulted from actual differences in opinion.

In summary, the results in this chapter demonstrated that SIMNORM is more effective than standardisation as a method of adjust scores to a common scale when differences are introduced through bias. In this sense, SIMNORM can be seen to be an effective way of equating scores across cultures at aggregate score level. SIMNORM should be used as a supplementary method to the analysis of raw scores, in order to evaluate the research results by comparing the effects when cultural differences are ignored.

Of the three SIMNORM models presented, the research showed that the results between the three models were very similar. Therefore, it seems preferable to use the more parsimonious models B and D, rather than model A. Model D would be the best when variability in the raw scores is not too different across the cultural groups, and can be assumed to be equal. This model is also the most orderly model. In simple terms, Model D implies that, if the mean of cultural group 1 lies above the mean of cultural group 2, all transformed values of cultural group 1 will be above those of cultural group 2.

The results of SIMITNORM are presented in Chapter 6.

CHAPTER 6

SIMITNORM RESULTS

6.1 LAYOUT OF THIS CHAPTER

In this chapter the results of SIMITNORM are discussed. Section 6.2 briefly covers the SIMITNORM method, since the theoretical detail of SIMITNORM was covered in Section 3.3. The estimated parameters for the SIMITNORM procedure for the selected data set are set out in Appendix C, with the fitted transformation functions and item distributions in Appendix D for each of the 20 items and each of the seven SIMITNORM models. Section 6.2 also covers the resulting scoring levels of items 1 to 4, for each of the seven SIMITNORM models, and the implications of these scoring levels are briefly discussed. The scoring levels of all the other items appear in Appendix E.

Section 6.3 presents the descriptive measures of the *Raw Scores*, *Standardised Z-Scores* and *SIMITNORM Scores* by using the traditional descriptive measures, box-and-whisker plots, histograms, normal probability plots and tests of normality. In Section 6.4, the reliability of the *Raw Scores*, *Standardised Z-Scores* and *SIMITNORM Scores* are considered. This section also gives the resulting correlation coefficients between the aggregate scores for *Raw Scores*, *Standardised Z-Scores*, *SIMNORM Scores* and for the aggregated *SIMITNORM Scores*.

Since the SIMITNORM procedure involves non-linear transformations, the correlation coefficients between items will be affected. In Section 6.5, exploratory factor analysis was used to assess to what extent SIMITNORM has had an effect on the factorial structure within the data, and how it compares to standardisation. In Section 6.5, confirmatory factor analysis is used to consider the effect of SIMITNORM on the covariance structure within the data, and to compare the effect with standardisation.

Multi-group confirmatory factor analysis (MG CFA) is one of the most common and popular methods used to consider equivalence problems in data. Therefore, the *Raw Scores*, *Standardised Z-Scores* and *SIMITNORM Scores* are compared in Section 6.7. Lastly, the effects of SIMITNORM on three different types of bias observed in the items are tested.

6.2 THE SIMITNORM PROCEDURE

SIMITNORM is a simultaneous non-linear transformation of raw ordinal scores, to new ordinal scores with a mean of zero and standard deviation of one, which more closely resembles the normal distribution. Figure 6.1 shows the hierarchy of nested models in SIMITNORM. Model A, the most unrestricted model, has three parameters for each item and each group. In comparison, the standardisation method which was used as the benchmark, is also group independent and item independent, and is a linear transformation function with two parameters within each group and each item. Model B is a special case of Model A, with one parameter fixed across items. Model D is a special case of both Models A and B, with two parameters fixed across items. Models BB and DD are group independent models, where the normalisation happens independently for each cultural group. Model BB has one parameter fixed across all items within a group, and Model DD has two parameters fixed across all items within a group. Models BB and DD are not special cases of models B and D, but model DD is a special case of model BB. Models BBB and DDD are both group and item dependent. Model BBB has one parameter fixed across all groups and items, and Model DDD has two parameters fixed across all groups and items. Model BBB is a special case of both models B and BB; and similarly, model DDD is a special case of both models D and DD. Model DDD is also a special case of model BBB. Figure 6.1 shows the hierarchies and interdependencies graphically, and Table 6.1 further explains the implications of model assumptions.

As was discussed in Chapter 5, the basic SIMITNORM function is

$$z = t(x) = \frac{-\ln\left(\left(\frac{U}{x}\right)^{\gamma} - 1\right) - \phi_1}{\phi_2} \qquad (6.1)$$

with $\gamma = e^{\phi_3} > 0$ and U = the maximum score. The following parameter constraints apply:

$$-\infty < \phi_1 < \infty$$

$$\phi_2 > 0 \qquad (6.2)$$

and $-\infty < \phi_3 < \infty$.

The SIMITNORM function for the k-th item of the j-th group as perceived by the i-th respondent, is therefore

$$t(x_{ijk}) = \frac{y_{ijk} - \phi_{1jk}}{\phi_{2jk}} \qquad (6.3)$$

with $y_{ijk} = -\ln\left\{\left(\frac{U}{x_{ijk}}\right)^{\gamma_{jk}} - 1\right\}$ (6.4)

where $\gamma_{jk} = e^{\phi_{3jk}} > 0$, and U = the maximum score of the ordinal item. The constraints on the parameters are

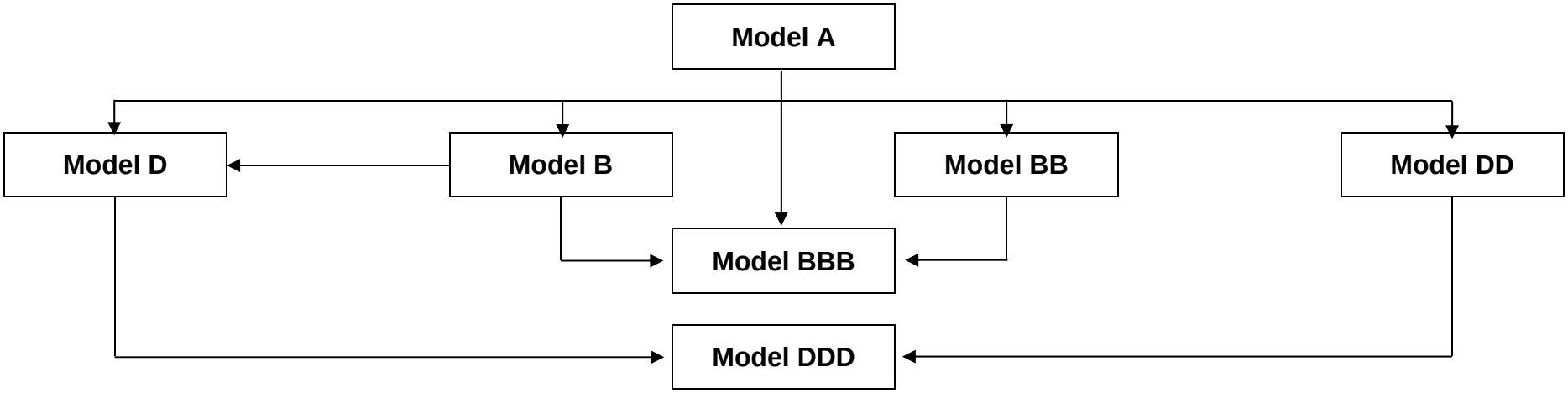
$$-\infty < \phi_{1jk} < \infty$$

$$\phi_{2jk} > 0 \qquad (6.2)$$

and $-\infty < \phi_{3jk} < \infty$.

The suitability of this mathematical expression for the purpose of this research lies in the fact that each of the parameters can be associated with a specific meaning within each item and each group. The parameter γ is associated with the amiability to the statement, almost similar to the item difficulty parameter in item response theory. The parameter ϕ_1 is associated with the item location and the parameter ϕ_2 is associated with the item spread. However, it should be noted that both ϕ_1 and ϕ_2 are functions of γ in the transformation function (6.3). This is a useful property, since it seems logic that the location and the spread of responses to an item are affected by the amiability to the statement.

Figure 6.1: Hierarchy of nested models for SIMITNORM in this research



Model A	Model independent of groups and items	ϕ_{1jk}, ϕ_{2jk} and γ_{jk} unique for each of 20 items and each of 4 cultural groups $p = m \times g \times 3 = 20 \times 4 \times 3 = 240$ parameters
Model B	Independent item model, one parameter fixed per item	γ_k fixed across groups for each of m items; ϕ_{1jk} and ϕ_{2jk} unique for each item and each group $p = m[2g + 1] = 20[2(4) +1]=180$ parameters
Model D	Independent item model, two parameters fixed per item	ϕ_{2k} and γ_k fixed across groups for each of m items; ϕ_{1jk} unique for each item and each group $p = m[g + 2] = 20[4+2] = 120$ parameters
Model BB	Independent group model, one parameter fixed per group	γ_j fixed across items for each of g groups; ϕ_{1jk} and ϕ_{2jk} unique for each item and each group $p = g[2m + 1]= 4[2(20)+1]=164$ parameters
Model BBB	Model both item and group dependent, one parameter fixed across groups and items	γ fixed across all items and groups, ϕ_{1jk} and ϕ_{2jk} unique for each item and each group $p = 2 \times m \times g + 1 = 2 \times 20 \times 4 +1 = 161$ parameters
Model DD	Independent group model, two parameters fixed per group	ϕ_{2j} and γ_j fixed across items for each of g groups; ϕ_{1jk} unique for each item and each group $p = g[m + 2] = 4 [20+2] = 88$ parameters
Model DDD	Model both item and group dependent, two parameters fixed across groups and items	ϕ_2 and γ fixed across all items and groups; ϕ_{1jk} unique for each item and each group $p = [m \times g] + 2 = [20 \times 4] +2 = 82$ parameters

p = number of parameters
m = number of items — In the Schlinger VRP20, 20 items were used
g = number of cultural groups — In the Schlinger VRP20, 4 cultural groups were used

Table 6.1 SIMITNORM models, assumptions, and practical implications

Model	Model Assumptions	Implications of assumptions
A	All item location, spread and amiability parameters unique across items and groups Model independent of groups and items	ϕ_{1jk}, ϕ_{2jk} and γ_{jk} unique for each item and each group $p = m \times g \times 3$ parameters Suitable baseline model, all parameters unique for each item and each group
B	Item location and spread parameters unique across items and groups Assumed that each item has the same amiability across groups Each item has its own unique amiability parameter that is fixed across groups Independent item model, one parameter fixed per item	γ_k fixed across groups for each of m items; ϕ_{1jk} and ϕ_{2jk} unique for each item and each group $p = m[2g + 1]$ parameters This model assumes that each item has a unique amiability that is equal across groups. Each group and each item have unique item location and item spread parameters
D	Item location parameters unique across items and groups Assumed that each item has the same amiability and the same spread across groups Each item has its own unique amiability parameter and own unique spread parameter that is fixed across groups Independent item model, two parameters fixed per item	ϕ_{2k} and γ_k fixed across groups for each of m items; ϕ_{1jk} unique for each item and each group $p = m[g + 2]$ parameters This model assumes that each item has unique amiability that is equal across groups and has unique spread that is equal across groups. Each group and each item has a unique location parameter
BB	Item location and item spread parameters unique across items and groups Assumed that all items within a group has the same amiability Each group has its own unique amiability parameter that is fixed across items Independent group model, one parameter fixed per group	γ_j fixed across items for each of g groups; ϕ_{1jk} and ϕ_{2jk} unique for each item and each group $p = g[2m + 1]$ parameters This model assumes that each group has a unique amiability parameter that is equal across items, but with item location and item spread unique across groups and items
BBB	Item location and item spread parameters unique across items and groups Assumed that all items and all groups has the same amiability parameter One unique amiability parameter that is fixed across groups and items Model both item and group dependent, one parameter fixed across groups and items	γ fixed across all items and groups, ϕ_{1jk} and ϕ_{2jk} unique for each item and each group $p = 2 \times m \times g + 1$ parameters This model assumes that there is one amiability parameter for all items and all groups, but with item location and item spread unique across groups and items
DD	Item location parameters unique across items and groups Assumed that all items within a group has the same amiability parameter and the same spread parameter One unique amiability parameter that is fixed across items for each group One unique spread parameter that is fixed across items for each group Independent group model, two parameters fixed per group	ϕ_{2j} and γ_j fixed across items for each of g groups; ϕ_{1jk} unique for each item and each group $p = g[m + 2]$ = parameters This model assumes that each group has a unique amiability parameter that is equal across items, and a unique spread parameter that is equal across items. Only item location parameters are unique across groups and items
DDD	Item location parameters unique across items and groups Assumed that all items and all groups has equal amiability parameters and equal spread parameters One unique amiability parameter that is fixed across items and groups One unique spread parameter that is fixed across items and group Model both item and group dependent, two parameters fixed across groups and items	ϕ_2 and γ fixed across all items and groups; ϕ_{1jk} unique for each item and each group $p = [m \times g] + 2$ parameters This model assumes that there is one amiability parameter that is equal for all groups and items, and one spread parameter that is equal for all groups and items. Only item location parameters are unique across groups and items

The arrows in Figure 6.1 indicate increasing restrictiveness. The restrictive models have fewer parameters, and therefore offer more parsimonious solutions. The benchmark model is the standardisation model, which is not shown on the figure. For the standardisation model, two parameters are required for each group and each item, which presents a total of 160 parameters in the example data set with 20 items and four cultural groups. Only models D, DD and DDD are more parsimonious than the standardisation method, with 120, 88 and 82 parameters respectively. Models B, BB and BBB are slightly less parsimonious, with 180, 164 and 161 parameters respectively, in comparison with 160 parameters used for standardisation.

Table 6.2 provides the frequencies as observed across all four cultural groups and for each item. The negatively phrased items were reverse coded as in SIMNORM (Chapter 5), for the sake of simplifying the calculations. The negatively phrased items are shaded in Table 6.2, and therefore present the observed frequencies after reverse coding was applied. Closer inspection of Table 6.2 clearly shows the shift in frequencies across the cultural groups from left to right across the table, towards a more positive stance in the categories 4, “agree” and 5, “strongly agree”, for almost all the items. This again reveals the cultural bias that is present in the data.

After the SIMITNORM estimation, with all the results given in Appendices C, D and E, a new scoring method is applied for each group and each item. Tables 6.3 to 6.9 set out the scoring levels as well as the distances between adjacent score levels for each of the SIMITNORM models, focusing only on the first four items. The complete results for the seven SIMITNORM models and all 20 items appear in Appendix E.

It is interesting to examine the distances between the scoring levels in Tables 6.3 to 6.9. It is clear that the distances between the scoring levels are all different for Models A, B, BB and BBB, which is expected, since the item spread parameters are not restricted in any of these models.

Table 6.2: Item Frequencies across cultural groups

GROUP	English					Afrikaans					Nguni					Sotho				
ITEM	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
VV1	93	191	242	463	218	76	141	163	543	270	24	106	166	686	404	10	82	98	509	315
VV2	96	199	216	477	219	85	184	199	525	200	29	92	147	758	360	13	87	94	569	251
VV3	103	209	276	423	196	101	205	226	477	184	29	72	158	746	381	8	52	108	610	236
VV4	113	219	245	454	176	94	199	187	526	187	54	151	235	668	278	21	130	158	520	185
VV8	20	79	127	544	437	33	111	104	508	437	46	137	91	742	370	19	75	44	600	276
VV9	21	99	69	512	506	30	110	64	511	478	75	172	71	664	404	39	127	33	507	308
VV10	21	57	75	526	528	14	61	79	518	521	36	87	71	719	473	21	65	30	564	334
VV14	229	450	212	263	53	208	430	221	280	54	60	193	172	703	258	26	164	118	520	186
VV15	154	294	342	341	76	135	258	341	357	102	73	196	160	618	339	26	149	106	488	245
VV16	139	309	278	387	94	150	282	250	399	112	39	129	123	708	387	16	81	103	557	257
VV17	38	59	307	521	282	33	42	251	512	355	7	25	268	595	491	4	13	210	463	324
VV18	20	51	274	559	303	22	50	224	523	374	12	42	350	558	424	10	19	270	457	258
VV20	150	312	314	339	92	167	276	241	377	132	43	149	262	755	177	23	110	187	604	90
VV21	197	351	309	274	76	224	327	231	305	106	66	210	330	619	161	22	172	231	483	106
VV22	200	278	293	339	97	216	209	246	366	156	53	140	299	712	182	24	120	213	560	97
VV24	30	112	219	539	307	27	101	155	562	348	16	53	151	679	487	8	41	96	571	298
VV25	23	58	163	566	397	13	61	103	548	468	6	19	45	723	593	2	7	24	595	386
VV27	48	146	315	516	182	26	117	287	512	251	42	93	176	659	416	12	63	113	504	322
VV28	64	192	213	521	217	29	166	167	534	297	26	69	64	693	534	11	46	35	531	391
VV29	72	228	234	456	217	75	199	169	473	277	34	78	56	656	562	14	57	38	500	405

Model A, with the score levels given in Table 6.3 for the first four items, is a completely unrestricted model, and can be viewed as a baseline model. This model has three unique parameters for each item and each group, allowing the item location, item spread and item amiability to be estimated freely for each item and each group. The differences between adjacent score categories are also unique across items and groups, as can be seen in the second column of Table 6.3, as would be expected from a completely unrestricted model.

Table 6.3: SIMITNORM score levels and distances between adjacent score categories: Model A

Model A									
SIMITNORM score Model A					Distance between adjacent scores Model A				
A	Eng	Afr	Nguni	Sotho	A	Eng	Afr	Nguni	Sotho
	Item 1					Item 1			
1	-1.583	-1.724	-2.185	-2.280	2-1	0.686	0.605	0.655	0.668
2	-0.898	-1.120	-1.530	-1.612	3-2	0.618	0.596	0.655	0.666
3	-0.280	-0.524	-0.875	-0.946	4-3	0.709	0.732	0.815	0.826
4	0.429	0.208	-0.060	-0.119	5-4	1.211	1.324	1.487	1.505
5	1.640	1.532	1.427	1.386					
Item 2					Item 2				
1	-1.568	-1.632	-2.198	-2.264	2-1	0.674	0.668	0.645	0.695
2	-0.894	-0.964	-1.553	-1.569	3-2	0.611	0.625	0.666	0.694
3	-0.284	-0.338	-0.888	-0.875	4-3	0.705	0.740	0.846	0.863
4	0.421	0.402	-0.041	-0.012	5-4	1.209	1.297	1.573	1.573
5	1.630	1.699	1.532	1.561					
Item 3					Item 3				
1	-1.522	-1.539	-2.240	-2.600	2-1	0.718	0.691	0.643	0.826
2	-0.805	-0.848	-1.597	-1.774	3-2	0.626	0.624	0.666	0.789
3	-0.179	-0.223	-0.931	-0.985	4-3	0.701	0.718	0.849	0.948
4	0.522	0.495	-0.083	-0.037	5-4	1.171	1.229	1.580	1.683
5	1.693	1.724	1.498	1.646					
Item 4					Item 4				
1	-1.485	-1.582	-1.904	-2.057	2-1	0.705	0.668	0.679	0.761
2	-0.780	-0.914	-1.225	-1.296	3-2	0.624	0.622	0.656	0.710
3	-0.156	-0.292	-0.569	-0.586	4-3	0.707	0.734	0.796	0.838
4	0.552	0.442	0.227	0.252	5-4	1.193	1.282	1.422	1.466
5	1.745	1.724	1.649	1.718					

Table 6.4 depicts the score levels after transformation for Model B, with distances between adjacent score categories in the second column of Table 6.4.

Model B has a restriction on the amiability parameter for each item. This model would be suitable where the items in the instrument represent several underlying constructs or ideas that are not necessarily closely related, thereby allowing the amiability parameter to be fixed per item, but with the item location and the item spread parameters to vary freely across groups and items.

Table 6.4: SIMITNORM score levels and distances between adjacent score categories: Model B

Model B									
SIMITNORM score Model B					Distance between adjacent scores Model B				
B	Eng	Afr	Nguni	Sotho	B	Eng	Afr	Nguni	Sotho
	Item 1					Item 1			
1	-1.536	-1.745	-2.249	-2.344	2-1	0.623	0.629	0.706	0.717
2	-0.913	-1.116	-1.543	-1.627	3-2	0.599	0.604	0.678	0.689
3	-0.314	-0.512	-0.865	-0.938	4-3	0.722	0.729	0.818	0.831
4	0.408	0.217	-0.046	-0.107	5-4	1.286	1.298	1.457	1.479
5	1.694	1.515	1.410	1.372					
Item 2					Item 2				
1	-1.531	-1.616	-2.317	-2.338	2-1	0.623	0.648	0.741	0.754
2	-0.908	-0.968	-1.576	-1.585	3-2	0.596	0.619	0.708	0.720
3	-0.312	-0.349	-0.868	-0.864	4-3	0.716	0.744	0.851	0.866
4	0.404	0.395	-0.017	0.001	5-4	1.271	1.322	1.511	1.537
5	1.675	1.717	1.494	1.539					
Item 3					Item 3				
1	-1.481	-1.522	-2.423	-2.669	2-1	0.659	0.667	0.788	0.876
2	-0.823	-0.855	-1.635	-1.793	3-2	0.610	0.618	0.730	0.812
3	-0.212	-0.238	-0.905	-0.981	4-3	0.716	0.725	0.857	0.953
4	0.504	0.488	-0.047	-0.028	5-4	1.247	1.262	1.492	1.659
5	1.751	1.750	1.445	1.631					
Item 4					Item 4				
1	-1.457	-1.586	-1.953	-2.068	2-1	0.662	0.674	0.728	0.770
2	-0.795	-0.913	-1.225	-1.297	3-2	0.614	0.624	0.674	0.714
3	-0.181	-0.289	-0.551	-0.584	4-3	0.720	0.733	0.792	0.838
4	0.539	0.444	0.240	0.254	5-4	1.254	1.275	1.377	1.458
5	1.793	1.719	1.618	1.712					

In Table 6.5, the score levels of Model BB are provided for the first four items. Model BB restricts the amiability parameter across items, allowing it to vary across groups. The model also allows the spread and location parameters to vary freely across items and groups. This model would be suitable where the items relate to a single underlying primary construct, and where there may be unique response styles across items and groups affecting both spread and location in the response pattern. If this model fits the data well, it implies that each group has its own amiability parameter.

Table 6.5: SIMITNORM score levels and distances between adjacent score categories: Model BB

Model BB									
SIMITNORM score Model BB					Distance between adjacent scores Model BB				
BB	Eng	Afr	Nguni	Sotho	BB	Eng	Afr	Nguni	Sotho
	Item 1					Item 1			
1	-1.594	-1.793	-2.249	-2.293	2-1	0.704	0.690	0.706	0.680
2	-0.890	-1.103	-1.543	-1.613	3-2	0.624	0.625	0.678	0.673
3	-0.267	-0.477	-0.865	-0.940	4-3	0.707	0.722	0.818	0.831
4	0.441	0.245	-0.046	-0.109	5-4	1.201	1.246	1.457	1.515
5	1.642	1.491	1.410	1.406					
	Item 2					Item 2			
1	-1.583	-1.655	-2.317	-2.280	2-1	0.698	0.702	0.741	0.710
2	-0.884	-0.953	-1.576	-1.570	3-2	0.618	0.636	0.708	0.702
3	-0.266	-0.317	-0.868	-0.868	4-3	0.701	0.735	0.851	0.867
4	0.435	0.418	-0.017	-0.001	5-4	1.191	1.268	1.511	1.581
5	1.626	1.686	1.494	1.580					
	Item 3					Item 3			
1	-1.511	-1.535	-2.423	-2.526	2-1	0.704	0.690	0.788	0.775
2	-0.807	-0.846	-1.635	-1.751	3-2	0.624	0.625	0.730	0.767
3	-0.183	-0.221	-0.905	-0.984	4-3	0.707	0.722	0.857	0.947
4	0.524	0.501	-0.047	-0.037	5-4	1.201	1.245	1.492	1.726
5	1.725	1.746	1.445	1.689					
	Item 4					Item 4			
1	-1.483	-1.600	-1.953	-1.976	2-1	0.706	0.697	0.728	0.691
2	-0.777	-0.903	-1.225	-1.285	3-2	0.625	0.631	0.674	0.684
3	-0.151	-0.272	-0.551	-0.602	4-3	0.709	0.729	0.792	0.844
4	0.558	0.457	0.240	0.242	5-4	1.205	1.258	1.377	1.539
5	1.762	1.715	1.618	1.782					

Model BBB, with the score levels given in Table 6.6, restricts the amiability parameter to be fixed across all items and groups. It therefore has a single amiability parameter, assumed to be equal across all groups and items. This model would be suitable when the items refer to a single underlying primary construct, and where item content can be assumed to carry similar substantive meaning across groups, thereby allowing a single amiability parameter. However, this model allows unique spread and location across items and groups, as in Model BB.

Table 6.6: SIMITNORM score levels and distances between adjacent score categories: Model BBB

Model BBB									
SIMITNORM score Model BBB					Distance between adjacent scores Model BBB				
BBB	Eng	Afr	Nguni	Sotho	BBB	Eng	Afr	Nguni	Sotho
	Item 1					Item 1			
1	-1.553	-1.765	-2.288	-2.388	2-1	0.648	0.655	0.741	0.754
2	-0.905	-1.109	-1.547	-1.634	3-2	0.608	0.614	0.694	0.706
3	-0.297	-0.495	-0.853	-0.928	4-3	0.720	0.728	0.823	0.837
4	0.423	0.233	-0.030	-0.091	5-4	1.270	1.284	1.451	1.476
5	1.693	1.517	1.421	1.384					
	Item 2					Item 2			
1	-1.544	-1.630	-2.349	-2.373	2-1	0.643	0.669	0.770	0.784
2	-0.900	-0.961	-1.579	-1.589	3-2	0.603	0.627	0.721	0.734
3	-0.298	-0.335	-0.858	-0.855	4-3	0.714	0.743	0.855	0.871
4	0.417	0.408	-0.003	0.016	5-4	1.260	1.310	1.508	1.535
5	1.677	1.718	1.506	1.551					
	Item 3					Item 3			
1	-1.472	-1.513	-2.403	-2.643	2-1	0.650	0.658	0.774	0.860
2	-0.823	-0.856	-1.629	-1.783	3-2	0.609	0.616	0.726	0.806
3	-0.214	-0.239	-0.9c03	-0.978	4-3	0.722	0.731	0.860	0.955
4	0.507	0.491	-0.043	-0.022	5-4	1.272	1.288	1.517	1.684
5	1.780	1.780	1.474	1.662					
	Item 4					Item 4			
1	-1.449	-1.577	-1.939	-2.051	2-1	0.654	0.664	0.716	0.758
2	-0.796	-0.913	-1.223	-1.293	3-2	0.612	0.622	0.671	0.710
3	-0.183	-0.291	-0.551	-0.583	4-3	0.726	0.738	0.796	0.842
4	0.543	0.447	0.244	0.258	5-4	1.280	1.301	1.403	1.484
5	1.823	1.749	1.647	1.743					

For Model D, in Table 6.7, the differences between scoring levels are constant over groups for each item as indicated by the cells with the same colour codes. Since this model restrains the spread parameter and amiability parameter to be fixed per item, it can be mathematically proven that the resulting distances between points would be equal across items. This result is shown in the second column of Table 6.7. Model D would be useful when there is sufficient reason to believe that amiability and spread is unique per item, but equal across groups.

In this model, bias is modelled as differences in location parameters alone. This model would be suitable in a context where firstly, the items in the questionnaire cover a broad range of issues, and are not relating to a single specific underlying construct; secondly, where the groups under consideration are assumed not to have an underlying bias in their response pattern such as extreme response style, therefore allowing each item to have a fixed item spread parameter; and thirdly, that the items have the same meaning (or amiability) across groups. The third condition implies that the items can be assumed not to have content that has alternative substantive meanings across groups.

Table 6.7: SIMITNORM score levels and distances between adjacent score categories: Model D

Model D									
SIMITNORM score Model D					Distance between adjacent scores Model D				
D	Eng	Afr	Nguni	Sotho	D	Eng	Afr	Nguni	Sotho
Item 1					Item 1				
1	-1.618	-1.823	-2.094	-2.150	2-1	0.644	0.644	0.644	0.644
2	-0.974	-1.179	-1.451	-1.507	3-2	0.630	0.630	0.630	0.630
3	-0.344	-0.549	-0.821	-0.877	4-3	0.771	0.771	0.771	0.771
4	0.427	0.222	-0.050	-0.106	5-4	1.389	1.389	1.389	1.389
5	1.816	1.610	1.339	1.283					
Item 2					Item 2				
1	-1.658	-1.683	-2.114	-2.097	2-1	0.659	0.659	0.659	0.659
2	-0.999	-1.024	-1.455	-1.438	3-2	0.645	0.645	0.645	0.645
3	-0.354	-0.379	-0.811	-0.793	4-3	0.788	0.788	0.788	0.788
4	0.434	0.409	-0.023	-0.005	5-4	1.418	1.418	1.418	1.418
5	1.852	1.827	1.396	1.413					
Item 3					Item 3				
1	-1.603	-1.627	-2.201	-2.177	2-1	0.686	0.686	0.686	0.686
2	-0.916	-0.941	-1.515	-1.491	3-2	0.661	0.661	0.661	0.661
3	-0.256	-0.280	-0.854	-0.831	4-3	0.798	0.798	0.798	0.798
4	0.542	0.518	-0.056	-0.033	5-4	1.423	1.423	1.423	1.423
5	1.965	1.941	1.367	1.390					
Item 4					Item 4				
1	-1.538	-1.647	-1.877	-1.877	2-1	0.692	0.692	0.692	0.692
2	-0.846	-0.955	-1.186	-1.185	3-2	0.648	0.648	0.648	0.648
3	-0.199	-0.307	-0.538	-0.537	4-3	0.767	0.767	0.767	0.767
4	0.568	0.460	0.229	0.229	5-4	1.344	1.344	1.344	1.344
5	1.912	1.804	1.573	1.573					

Equal distances between adjacent categories have the same colour coding

In Model DD, represented in Table 6.8, the differences are constant across items within a group. Equal distances have the same colour code in the table. Since Model DD restrains the amiability and the spread parameters to be fixed within a group, it can be mathematically proven that item transformations for this model would result in equal distances between categories across all items within a group. Model DD would be useful when the user views equivalence as having equal

distances between adjacent categories within a cultural group across all items. This model would therefore be suitable in a context where, firstly the items refer to a single underlying primary construct, secondly, where the group may have a unique response style (such as acquiescence or extreme response style) affecting the spread equally across all items; and thirdly, where the amiability to items may be unique within a group.

Table 6.8: SIMITNORM score levels and distances between adjacent score categories: Model DD

Model DD									
SIMITNORM score Model DD					Distance between adjacent scores Model DD				
DD	Eng	Afr	Nguni	Sotho	DD	Eng	Afr	Nguni	Sotho
	Item 1					Item 1			
1	-1.712	-1.853	-2.166	-2.389	2-1	0.746	0.701	0.635	0.703
2	-0.965	-1.152	-1.531	-1.687	3-2	0.670	0.646	0.650	0.700
3	-0.296	-0.507	-0.881	-0.986	4-3	0.767	0.755	0.822	0.869
4	0.472	0.249	-0.059	-0.117	5-4	1.314	1.318	1.528	1.593
5	1.786	1.567	1.469	1.476					
	Item 2					Item 2			
1	-1.714	-1.679	-2.136	-2.276	2-1	0.746	0.701	0.635	0.703
2	-0.967	-0.979	-1.501	-1.573	3-2	0.670	0.646	0.650	0.700
3	-0.298	-0.333	-0.851	-0.873	4-3	0.767	0.755	0.822	0.869
4	0.469	0.422	-0.029	-0.003	5-4	1.314	1.318	1.528	1.593
5	1.784	1.740	1.499	1.589					
	Item 3					Item 3			
1	-1.621	-1.586	-2.175	-2.309	2-1	0.746	0.701	0.635	0.703
2	-0.875	-0.885	-1.540	-1.606	3-2	0.670	0.646	0.650	0.700
3	-0.205	-0.240	-0.890	-0.905	4-3	0.767	0.755	0.822	0.869
4	0.562	0.516	-0.068	-0.036	5-4	1.314	1.318	1.528	1.593
5	1.876	1.834	1.460	1.556					
	Item 4					Item 4			
1	-1.587	-1.636	-1.884	-2.025	2-1	0.746	0.701	0.635	0.703
2	-0.840	-0.935	-1.249	-1.322	3-2	0.670	0.646	0.650	0.700
3	-0.171	-0.290	-0.600	-0.622	4-3	0.767	0.755	0.822	0.869
4	0.596	0.466	0.222	0.247	5-4	1.314	1.318	1.528	1.593

Equal distances between adjacent categories have the same colour coding

In Model DDD, as shown in Table 6.9, the distances between adjacent score categories are constant across items and groups. These constant differences are the results of the restrictions imposed on the parameters, which only allow the location parameters to be unique across items and groups. It can be mathematically proven that the quality restrictions of the item and group amiability and spread parameters result in equal distances between score levels across all groups and items. Model DDD is the most restrictive model of the seven SIMITNORM models, and has the greatest similarity with the method described

by Stacey (2005). Model DDD would be suitable within a context where firstly, all the items can be assumed to relate to a single primary construct; secondly, all the items and all groups are assumed to have similar spread parameters, and all the groups and items have similar amiability on the item content. This model assumes that all parameters across items and groups are fixed, except for a unique location parameter across items and groups.

Table 6.9: SIMITNORM score levels and distances between adjacent score categories: Model DDD

Model DDD									
SIMITNORM score Model DDD					Distance between adjacent scores Model DDD				
DDD	Eng	Afr	Nguni	Sotho	DDD	Eng	Afr	Nguni	Sotho
Item 1					Item 1				
1	-1.709	-1.924	-2.208	-2.267	2-1	0.699	0.699	0.699	0.699
2	-1.011	-1.225	-1.510	-1.568	3-2	0.668	0.668	0.668	0.668
3	-0.342	-0.556	-0.841	-0.899	4-3	0.805	0.805	0.805	0.805
4	0.462	0.248	-0.037	-0.095	5-4	1.438	1.438	1.438	1.438
5	1.900	1.686	1.401	1.343					
Item 2					Item 2				
1	-1.712	-1.739	-2.181	-2.163	2-1	0.699	0.699	0.699	0.699
2	-1.014	-1.040	-1.482	-1.464	3-2	0.668	0.668	0.668	0.668
3	-0.345	-0.372	-0.814	-0.796	4-3	0.805	0.805	0.805	0.805
4	0.459	0.433	-0.009	0.009	5-4	1.438	1.438	1.438	1.438
5	1.897	1.871	1.429	1.446					
Item 3					Item 3				
1	-1.616	-1.641	-2.219	-2.196	2-1	0.699	0.699	0.699	0.699
2	-0.918	-0.942	-1.520	-1.497	3-2	0.668	0.668	0.668	0.668
3	-0.249	-0.274	-0.852	-0.829	4-3	0.805	0.805	0.805	0.805
4	0.555	0.531	-0.047	-0.024	5-4	1.438	1.438	1.438	1.438
5	1.993	1.969	1.391	1.414					
Item 4					Item 4				
1	-1.581	-1.694	-1.932	-1.931	2-1	0.699	0.699	0.699	0.699
2	-0.882	-0.995	-1.233	-1.232	3-2	0.668	0.668	0.668	0.668
3	-0.214	-0.327	-0.565	-0.564	4-3	0.805	0.805	0.805	0.805
4	0.591	0.478	0.240	0.241	5-4	1.438	1.438	1.438	1.438
5	2.028	1.916	1.677	1.679					

Equal distances between adjacent categories have the same colour coding

The measures of fit are given in Table 6.10 for each of the SIMITNORM models. R-square is the highest for Model BBB, with a corresponding lowest RMSE. A very good fit is not expected, since the models transform an ordinal categorical variable to a continuous variable.

Table 6.10 Measures of fit for SIMITNORM models

Model	AIC	F	R-SQ	RMSE
A	2.514	0.335	0.745	0.518
B	2.521	0.340	0.742	0.522
D	2.530	0.345	0.735	0.529
BB	2.568	0.363	0.747	0.527
BBB	2.572	0.365	0.768	0.530
DD	2.576	0.368	0.741	0.533
DDD	2.584	0.372	0.760	0.539

6.3 DESCRIPTIVE MEASURES

In order to study the effect of the different SIMITNORM models, composite scores were calculated for the *Raw Scores*, and the *Standardised Z-Scores*, as well as for each of the seven SIMITNORM models, by averaging all the items. The descriptive statistics were calculated for the composite scores and are provided in Table 6.11, which shows that the means of the standardised and normalised functions are zero. In terms of locality transformation, standardisation and normalisation have achieved the same result, as was intended. The standard deviations for the *Standardised Z-Scores* are all equal to 1, but the standard deviations of the normalised scores are all less than 1, due to the effect of the SIMITNORM transformation. The descriptive statistics of the individual SIMITNORM items in Appendix C show that the individual items have standard deviations equal, or approximately equal to 1.

When the skewness and kurtosis of the scores are considered in Table 6.11, it clearly shows that standardisation has left skewness and kurtosis within each group unchanged, whilst all the SIMITNORM composite scores have more normal skewness and kurtosis. In the normal distribution, both skewness and kurtosis are zero. It is interesting to note that Model BB, which assumes that the amiability parameter is fixed for each group across all items, but which allows all the item location and spread parameters to be free across items, resulted in skewness and kurtosis measures that are the closest to zero of all the other models, and the skewness and kurtosis are also lower than for the *Z*-scores. This is an early indication that Model BB may be the most suitable model for the data used in this research. This model states that each group has a unique amiability parameter,

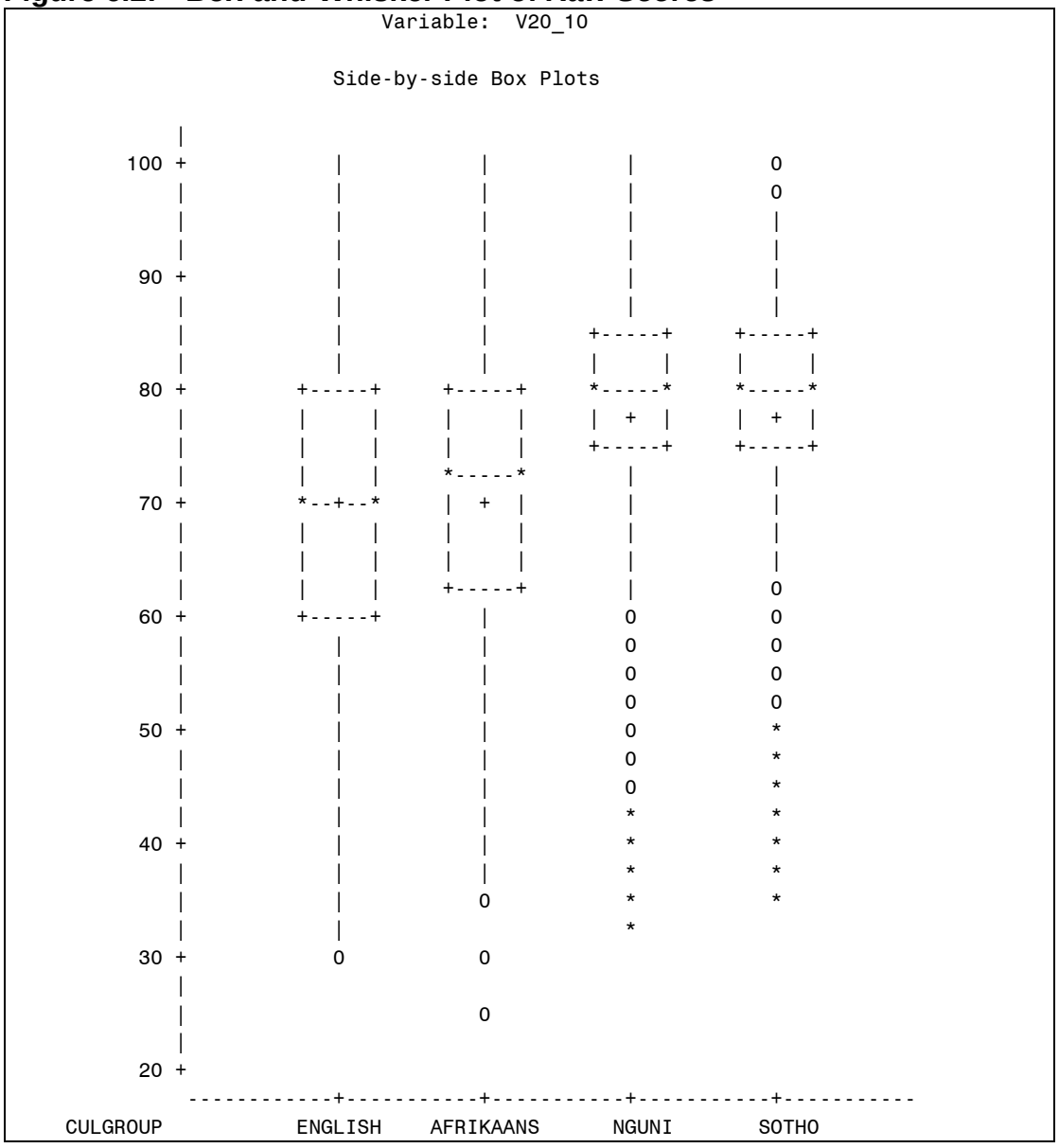
and each item and each group has two unique parameters for item location and item spread.

Table 6.11: Descriptive statistics of *Raw Scores*, *Standardised Z-scores* and *SIMITNORM Scores*

Sample Size	English 1207	Afrikaans 1193	Nguni 1386	Sotho 1014
Means				
Raw Score	69.0	70.8	78.2	78.7
Z-Score	0.000	0.000	0.000	0.000
A-Score	0.000	0.000	0.000	0.000
B-Score	0.000	0.000	0.000	0.000
D-Score	0.000	0.000	0.000	0.000
BB-Score	0.000	0.000	0.000	0.000
BBB-Score	0.000	0.000	0.000	0.000
DD-Score	0.000	0.000	0.000	0.000
DDD-Score	0.000	0.000	0.000	0.000
Standard Deviation				
Raw Score	13.478	13.693	10.808	9.699
Z-Score	1.000	1.000	1.000	1.000
A-Score	0.612	0.617	0.548	0.539
B-Score	0.613	0.618	0.551	0.540
D-Score	0.636	0.657	0.534	0.494
BB-Score	0.615	0.620	0.549	0.539
BBB-Score	0.616	0.620	0.552	0.541
DD-Score	0.619	0.623	0.549	0.538
DDD-Score	0.638	0.658	0.540	0.498
Skewness				
Raw Score	-0.287	-0.395	-1.403	-1.235
Z-Score	-0.287	-0.395	-1.403	-1.235
A-Score	0.054	-0.010	-0.475	-0.305
B-Score	0.147	0.038	-0.628	-0.386
D-Score	0.172	0.069	-0.594	-0.351
BB-Score	0.041	-0.023	-0.500	-0.318
BBB-Score	0.138	0.034	-0.655	-0.424
DD-Score	0.054	0.006	-0.509	-0.325
DDD-Score	0.161	0.066	-0.643	-0.405
Kurtosis				
Raw Score	-0.290	-0.146	2.850	2.898
Z-Score	-0.290	-0.146	2.850	2.898
A-Score	-0.142	-0.121	0.818	1.103
B-Score	-0.116	-0.112	1.136	1.200
D-Score	-0.120	-0.122	1.072	1.119
BB-Score	-0.104	-0.076	0.858	1.088
BBB-Score	-0.093	-0.089	1.178	1.278
DD-Score	-0.116	-0.101	0.881	1.063
DDD-Score	-0.103	-0.110	1.160	1.206

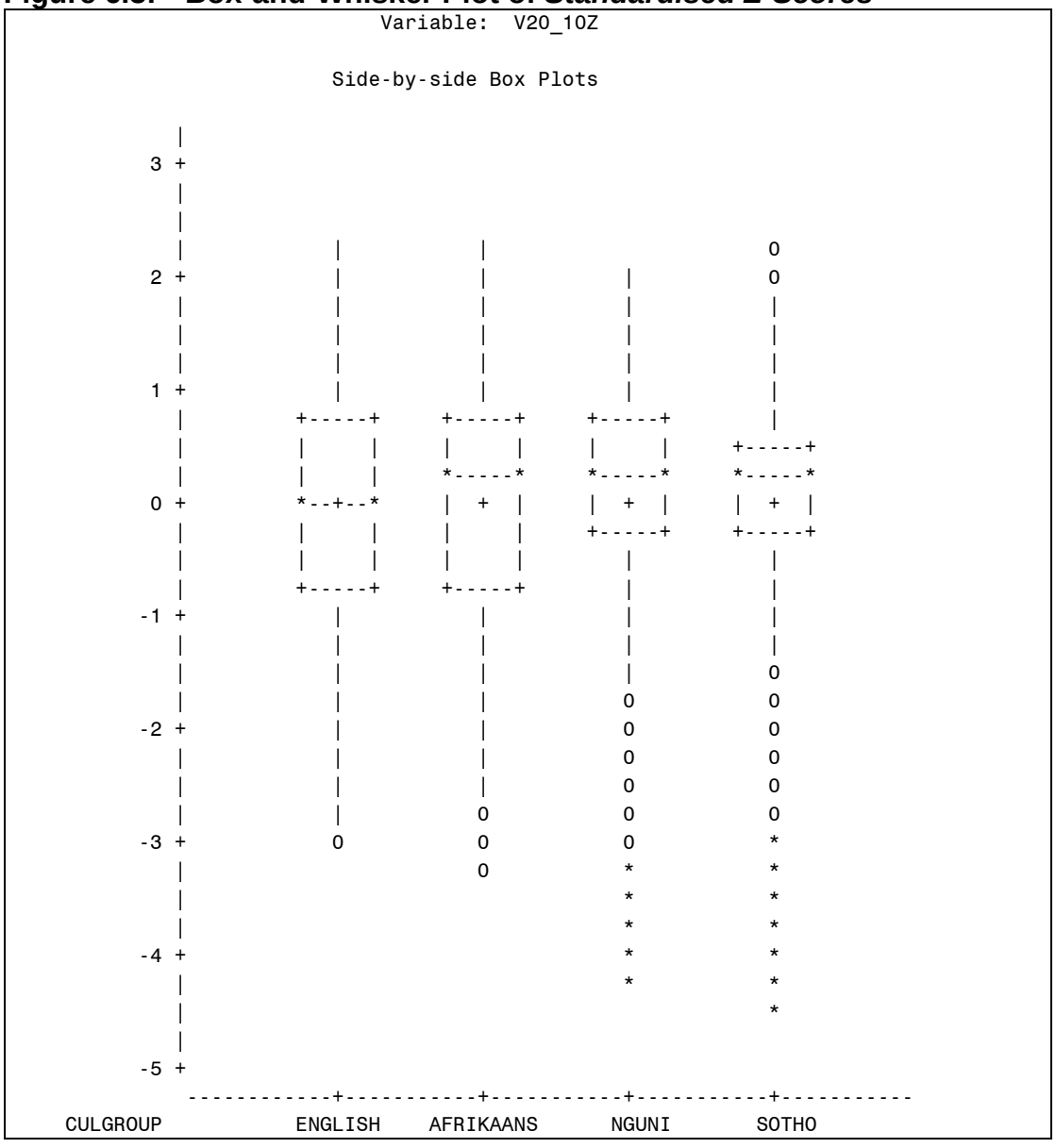
Figure 6.2 shows that the composite *Raw Scores* are all more positive for the Nguni and Sotho groups, and that the distributions of scores are more concentrated for these two groups.

Figure 6.2: Box-and-Whisker Plot of *Raw Scores*



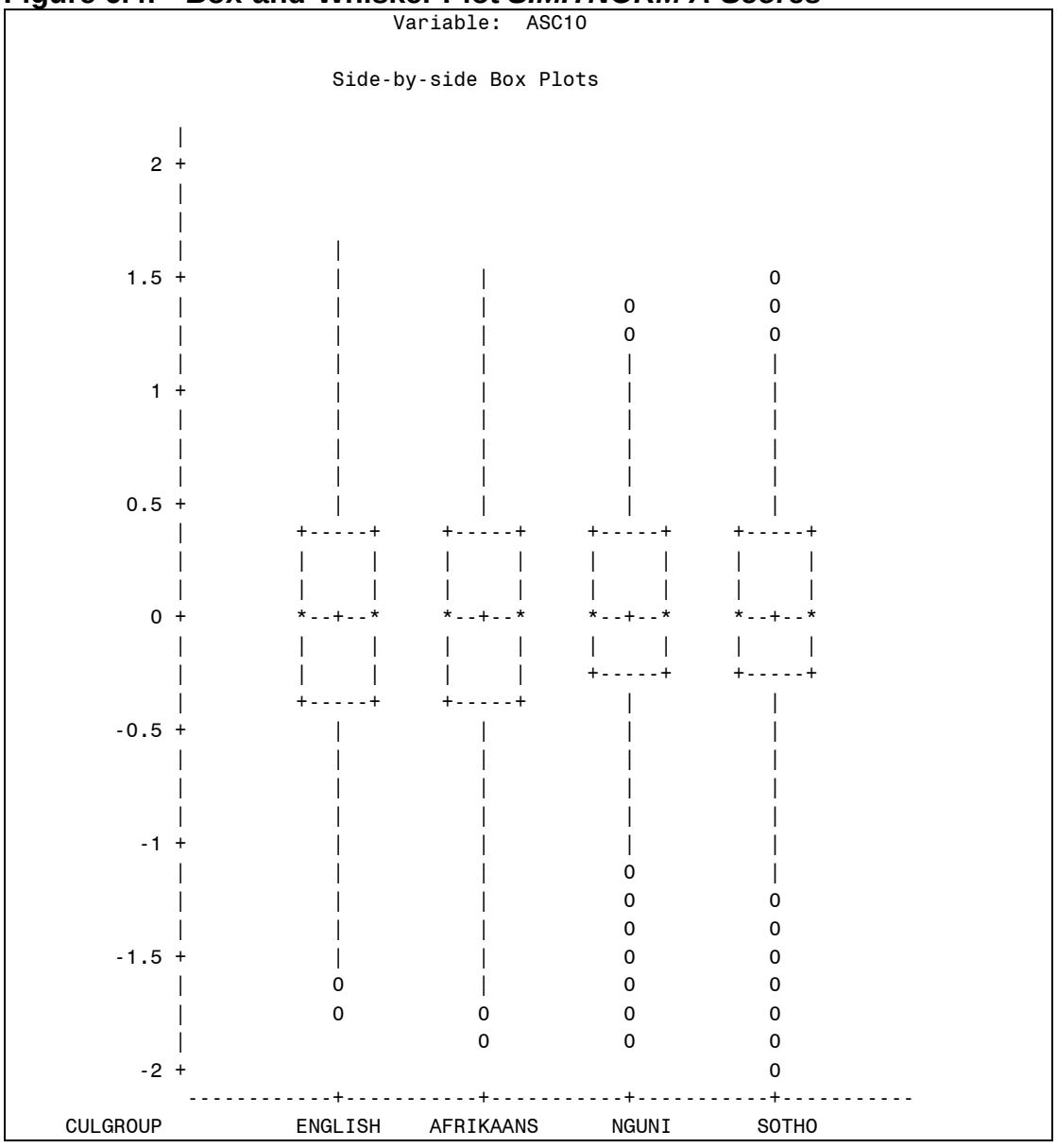
As can be seen in Figure 6.3, the composite score of the standardised items, resulted in a shift of location for the Nguni and Sotho groups, as well as a more comparable spread of scores, but the basic shape of the distribution was not affected.

Figure 6.3: Box-and-Whisker Plot of *Standardised Z-Scores*



Figures 6.4 to 6.10 show the box-and-whisker plots for the composite scores of all the SIMITNORM models. It is clear that there are hardly any noticeable differences across the models.

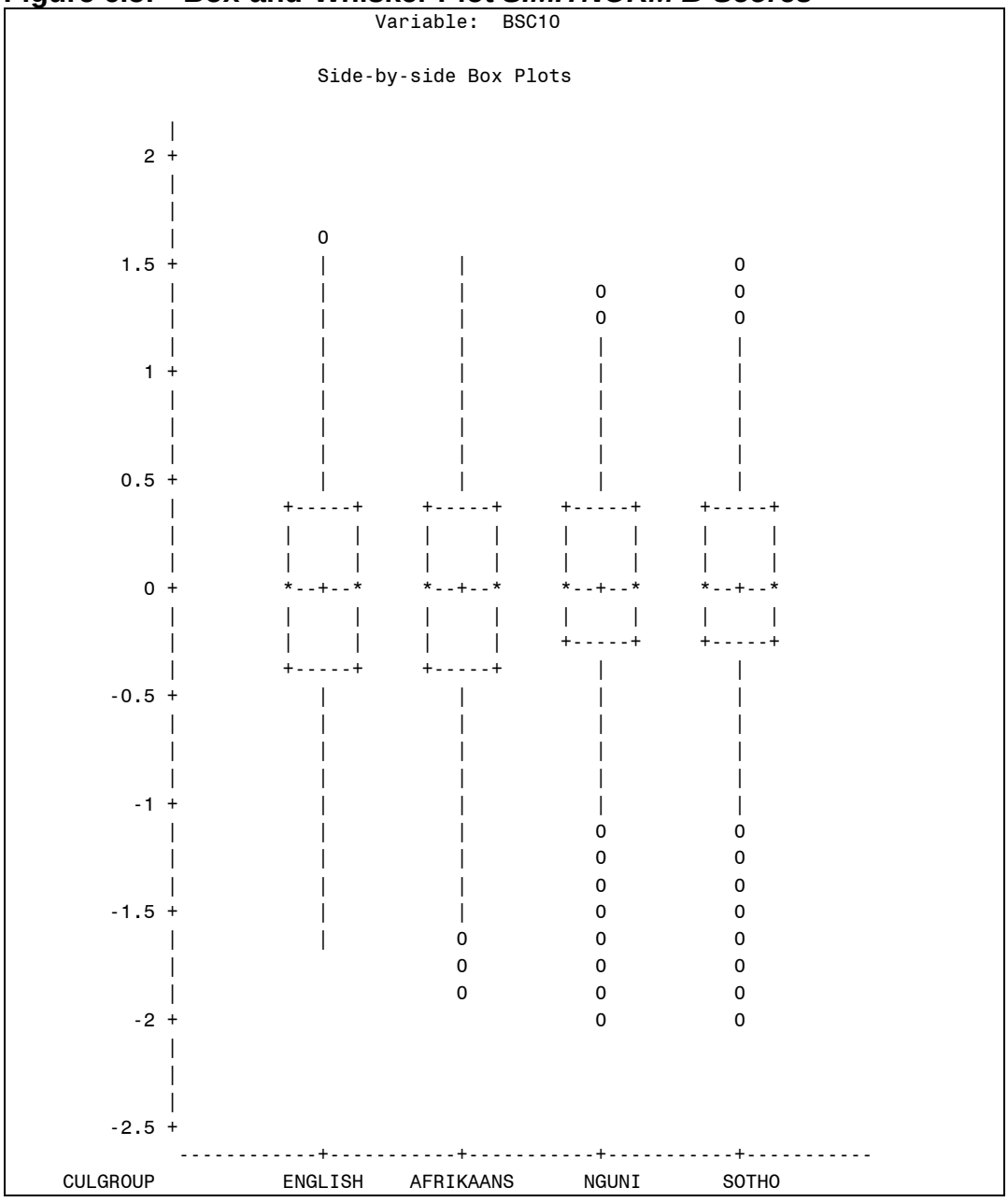
Figure 6.4: Box-and-Whisker Plot *SIMITNORM A-Scores*



In Figures 6.4 to 6.10, the location, spread, symmetry and general shape of the SIMITNORM total score distributions are more similar across groups, compared to the Z-score distribution in Figure 6.3. This can be seen from the position of the cross, which represents the position of the average, in the middle of the box, as well as the interquartile ranges. For the *Raw Scores* in Figure 6.2, and the *Standardised Z-Scores* in Figure 6.3, the means are not on the median (the line

that divides the box in two), whilst in all the *SIMITNORM Scores* (Figures 6.4 to 6.10), the means fall on the median. The boxes, representing the middle 50% of the scores, are also more equally divided in two similar halves for the *SIMITNORM Scores*, than for the *Standardised Z-Scores*, showing the improved symmetry after items were normalised.

Figure 6.5: Box-and-Whisker Plot *SIMITNORM B-Scores*



The least restricted models, A and B in Figures 6.4 and 6.5 were more appropriate to obtaining normal scores, than with the amiability restricted models D, DD and DDD in Figures 6.6, 6.9 and 6.10. In the more restricted models, the scores of the

Nguni and Sotho groups have less variability than the English and Afrikaans groups. It is understandable, since in these models, both the amiability and the spread parameters are held constant across groups, thereby not correcting sufficiently for the differences in spread in the original distributions.

Figure 6.6: Box-and-Whisker Plot *SIMITNORM D-Scores*

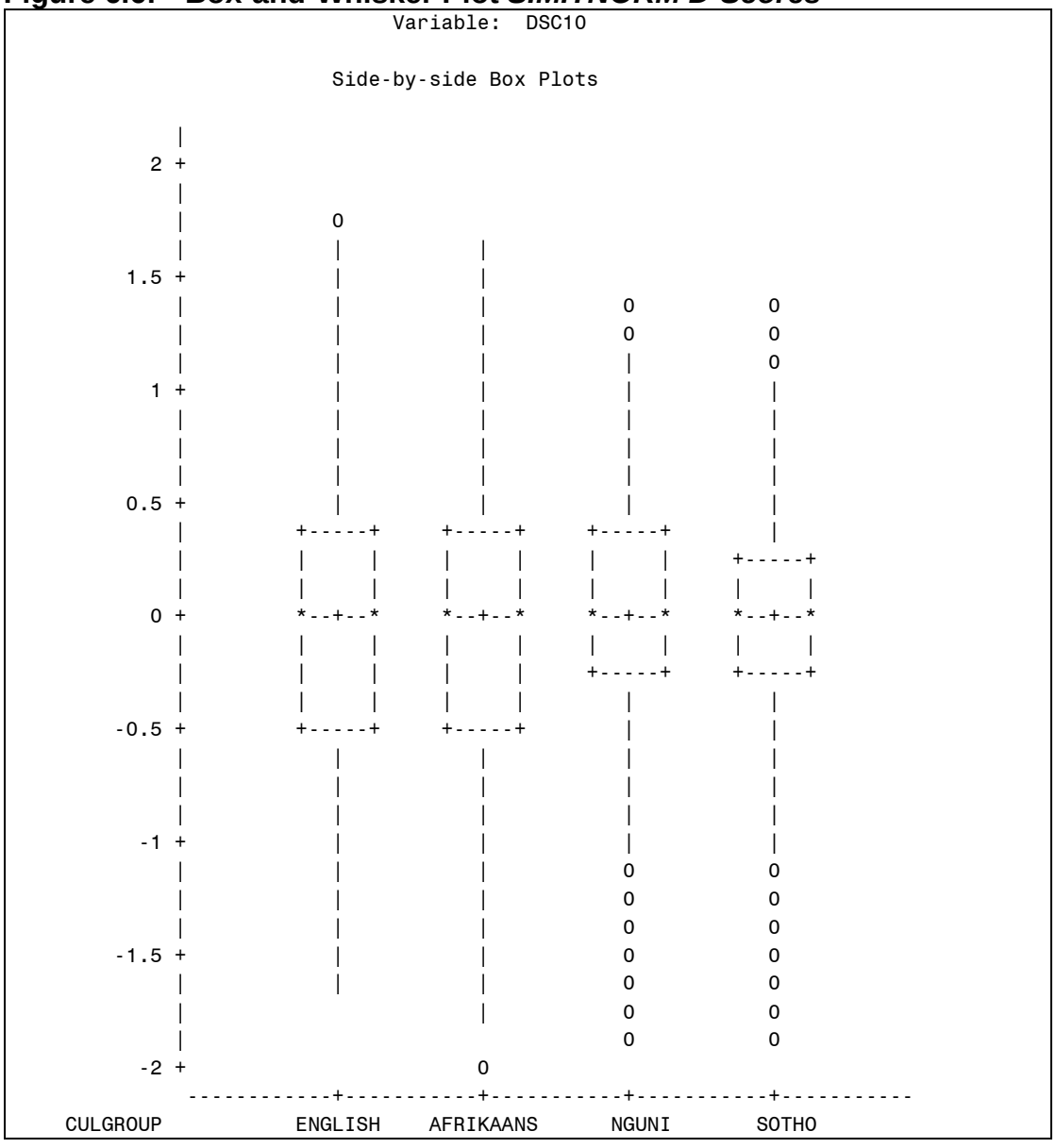


Figure 6.7: Box-and-Whisker Plot *SIMITNORM BB-Scores*

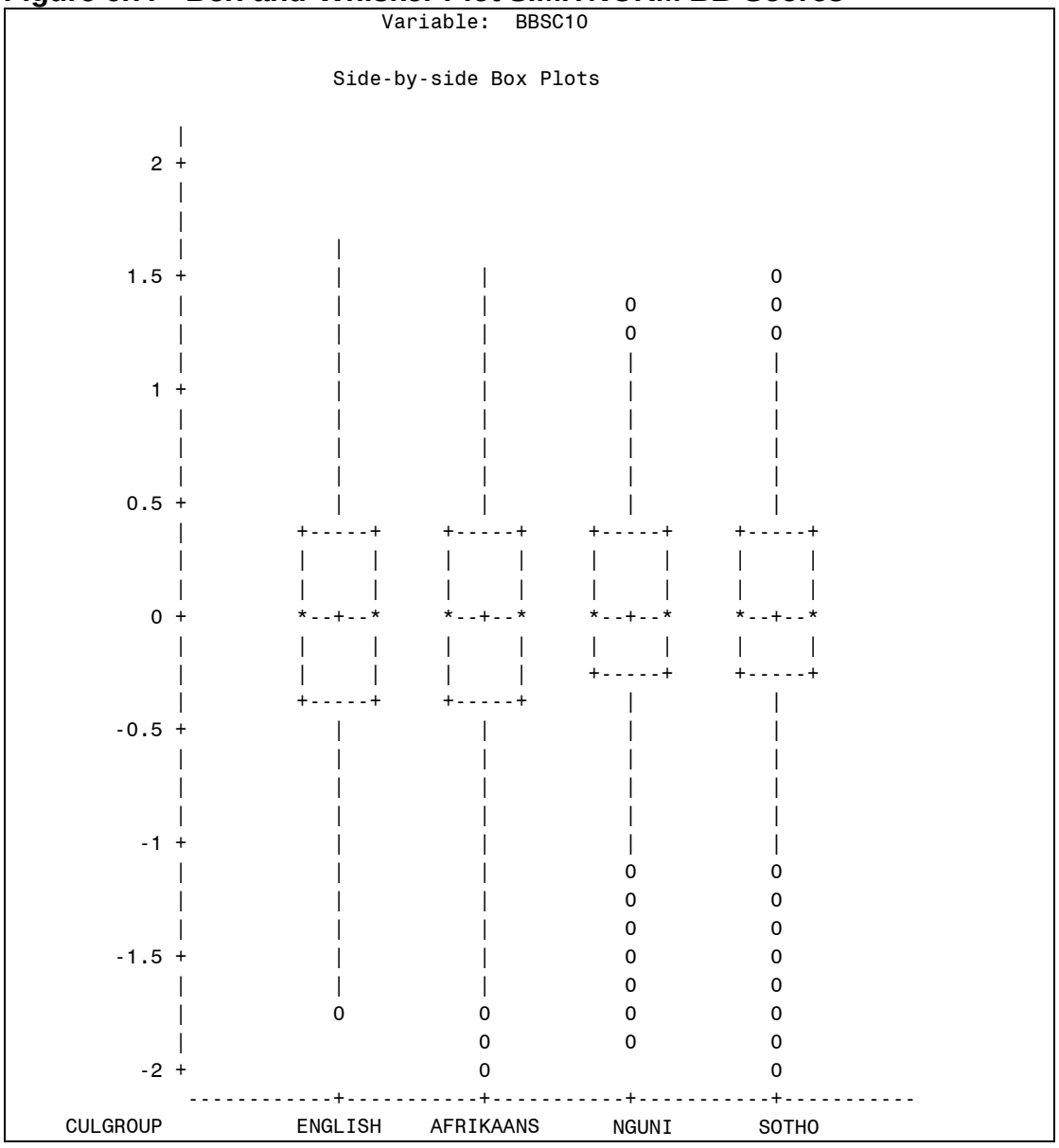


Figure 6.8: Box-and-Whisker Plot *SIMITNORM BBB-Scores*

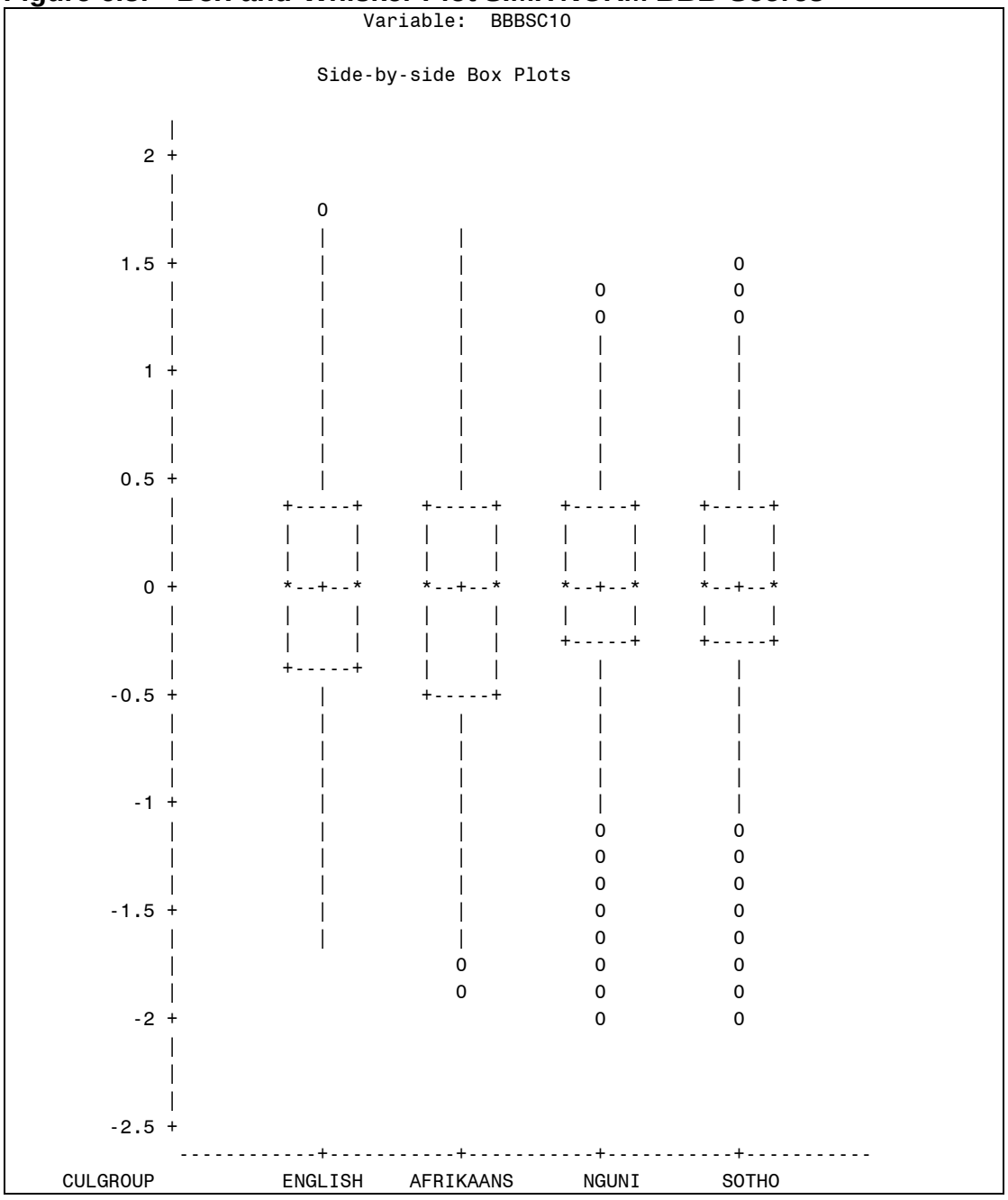


Figure 6.9: Box-and-Whisker Plot *SIMITNORM DD-Scores*

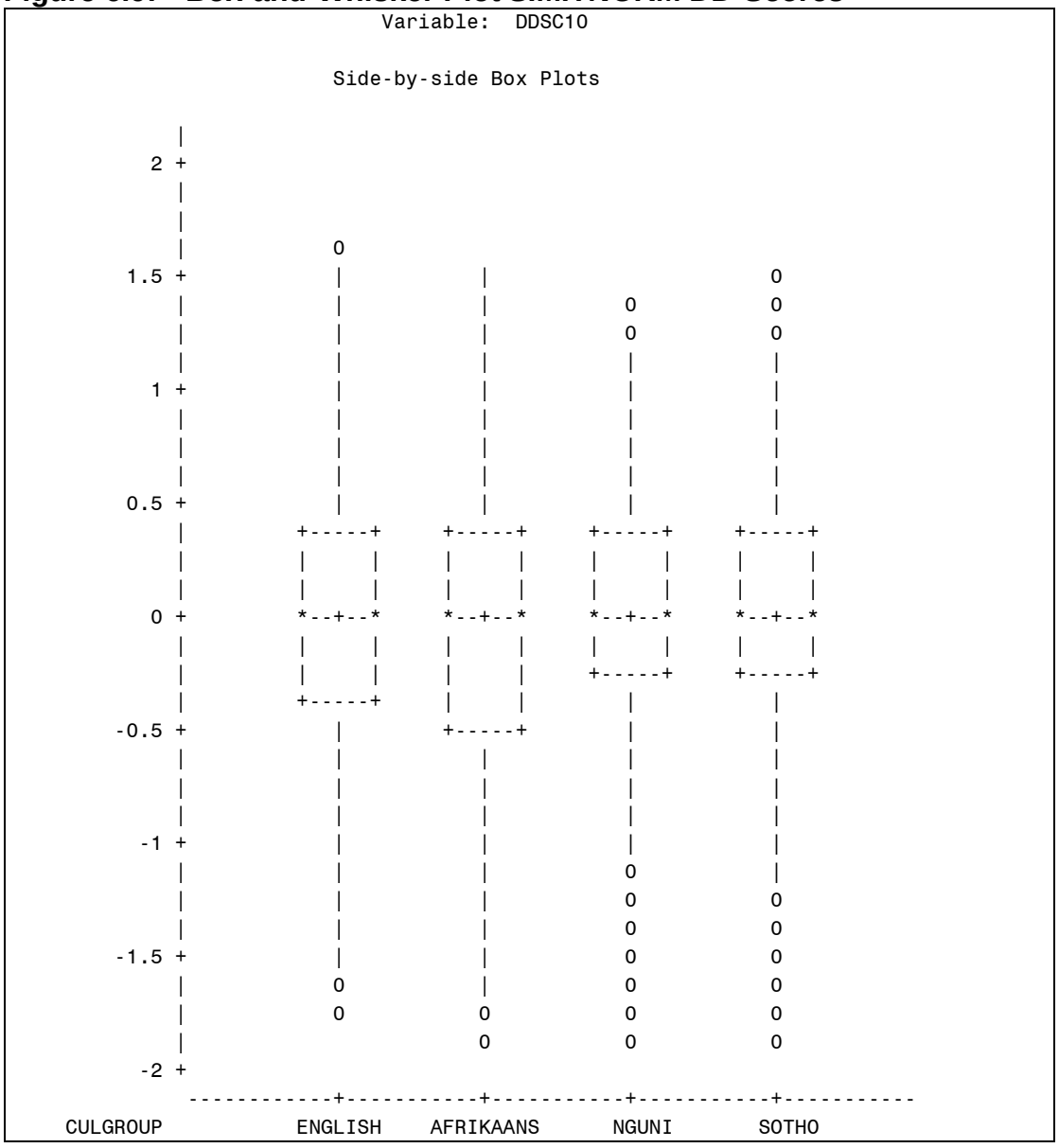
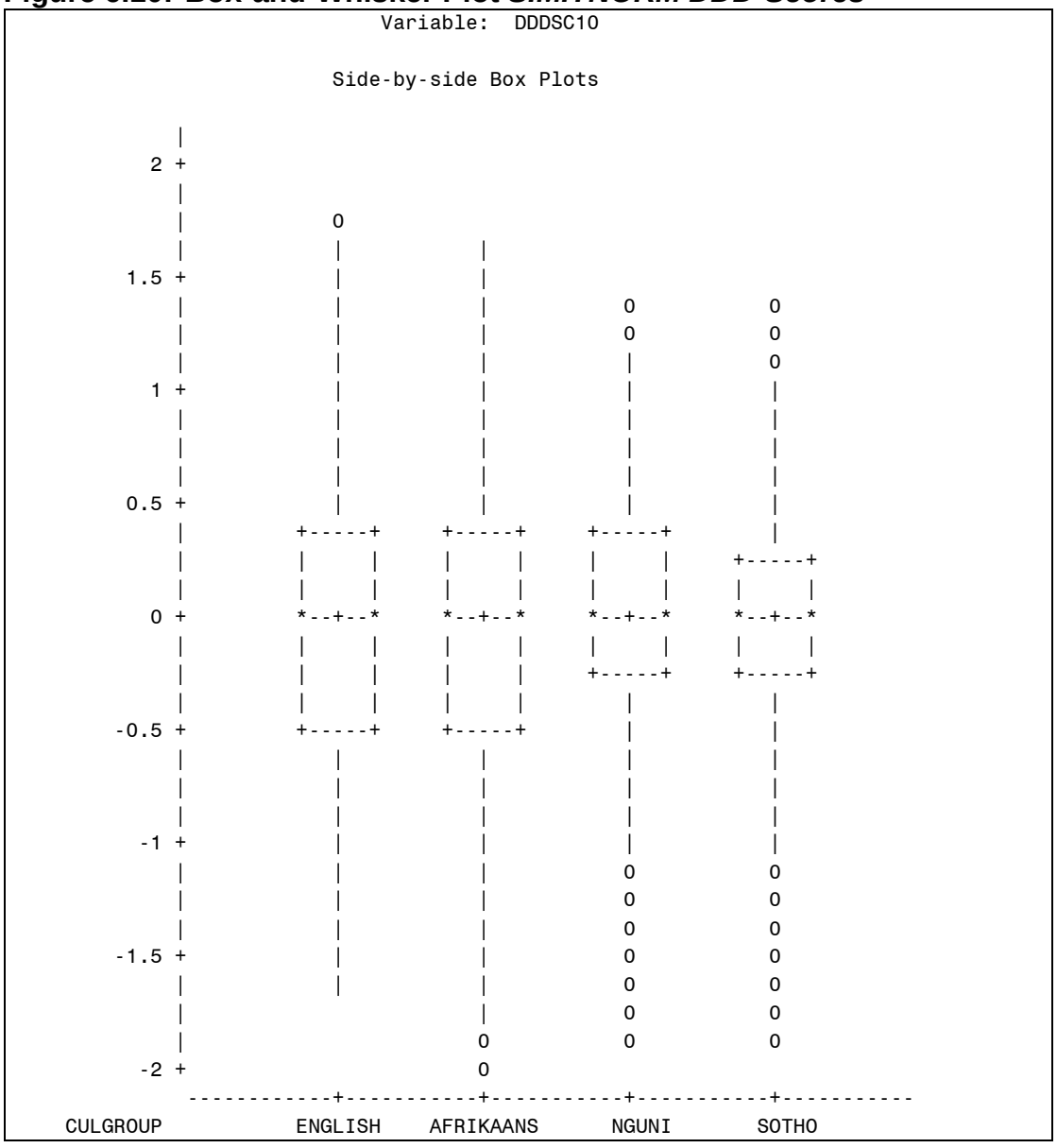


Figure 6.10: Box-and-Whisker Plot *SIMITNORM DDD-Scores*



In the histograms in Figure 6.11, one can clearly see how the locality, spread, symmetry and shape of the composite scores are affected by SIMITNORM. For the *Raw Scores* in the first row of the figure, the distributions are skewed to the left. The *Standardised Z-Scores* in the second row of the figure show that standardisation only affects the location and spread of the distributions, but the skewness and kurtosis have remained unchanged. All the distributions of the SIMITNORM models show largely equivalent means and spread across the cultural groups, and more equivalent spread and kurtosis, than were the case with *Standardised Z-Scores*. There is not much difference between the different SIMITNORM models based on the examining the histograms, which supports the supposition that Model A can be greatly simplified by choosing any of the more restricted models. The restricted models, Models B, BB, BBB have almost the same number of parameters that are necessary for standardisation, whereas the most restricted models D, DD or DDD have fewer parameters than standardisation, and should therefore be more desirable.

Figure 6.12 represents a three-dimensional view of the histograms in Figure 6.11. These figures show that SIMITNORM was not sufficiently successful in removing all the skewness and kurtosis in the Nguni and Sotho groups.

Figure 6.11: Histograms of *Raw Scores*, *Standardised Z-Scores* and *SIMITNORM Scores*

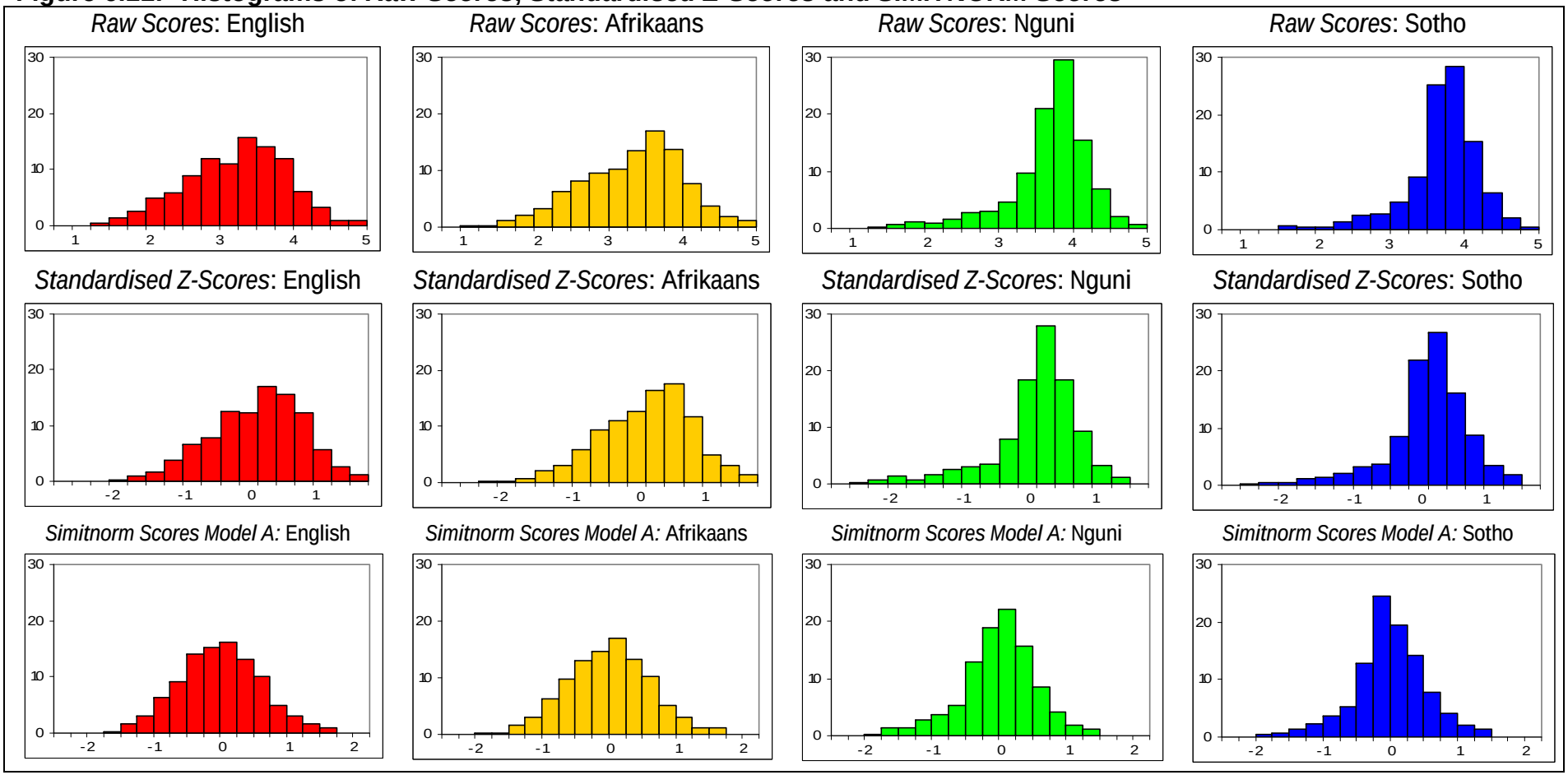


Figure 6.11(continued): Histograms of *Raw Scores*, *Standardised Z-Scores* and *SIMITNORM Scores*

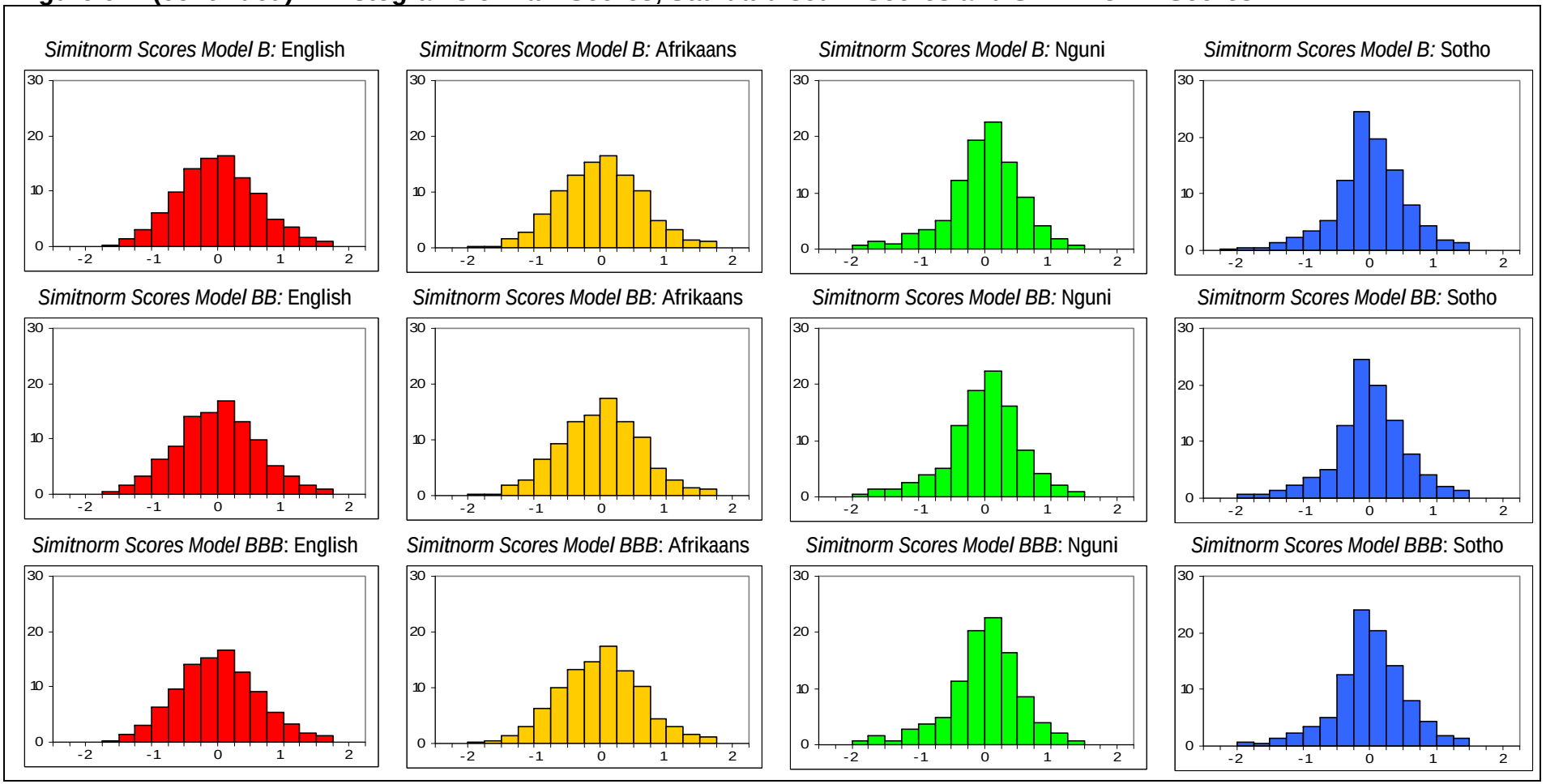


Figure 6.11(continued): Histograms of *Raw Scores*, *Standardised Z-Scores* and *SIMITNORM Scores*

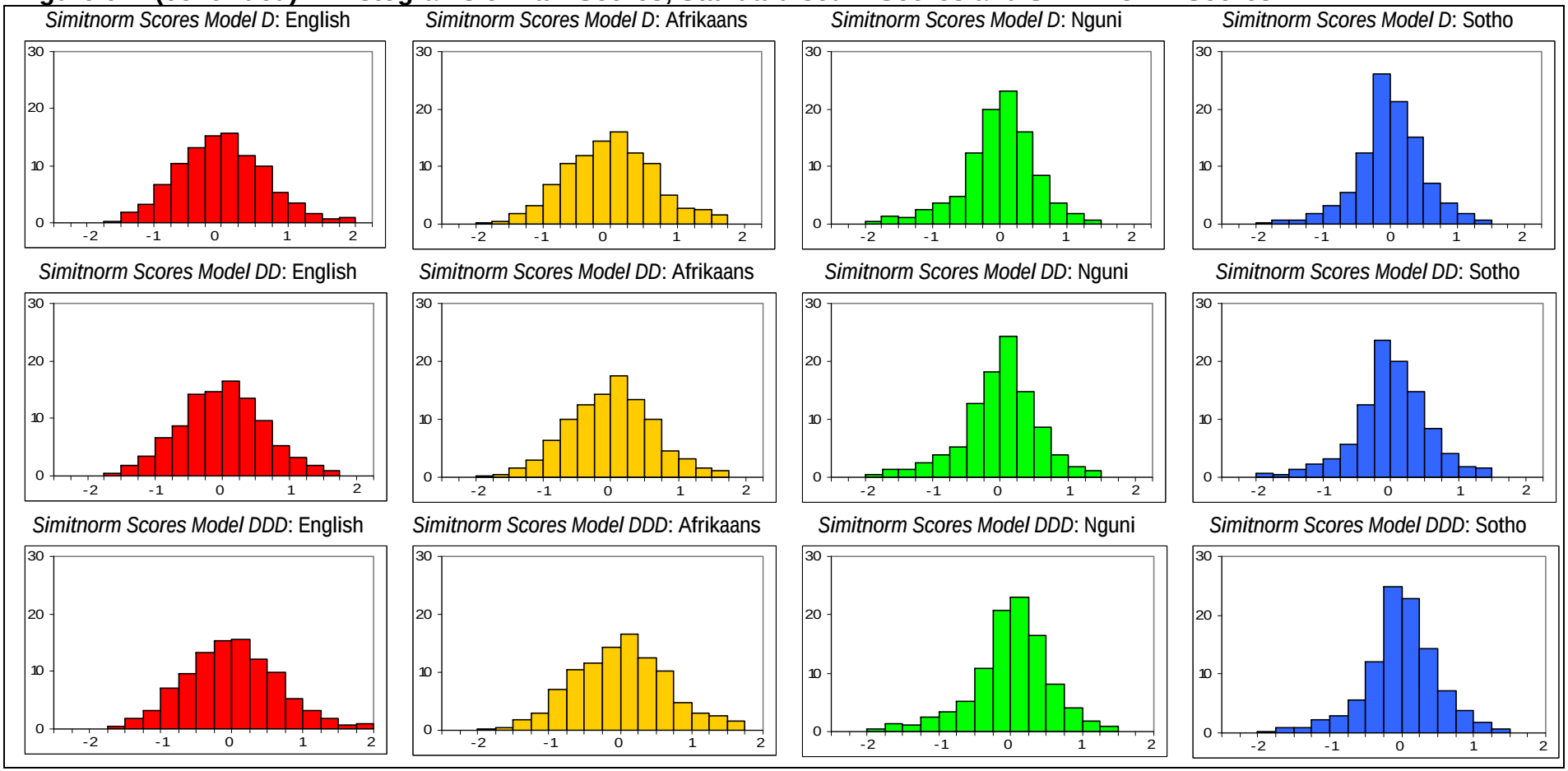


Figure 6.12:` Three-dimensional histograms of *Standardised Z-Scores* and *SIMITNORM Scores*

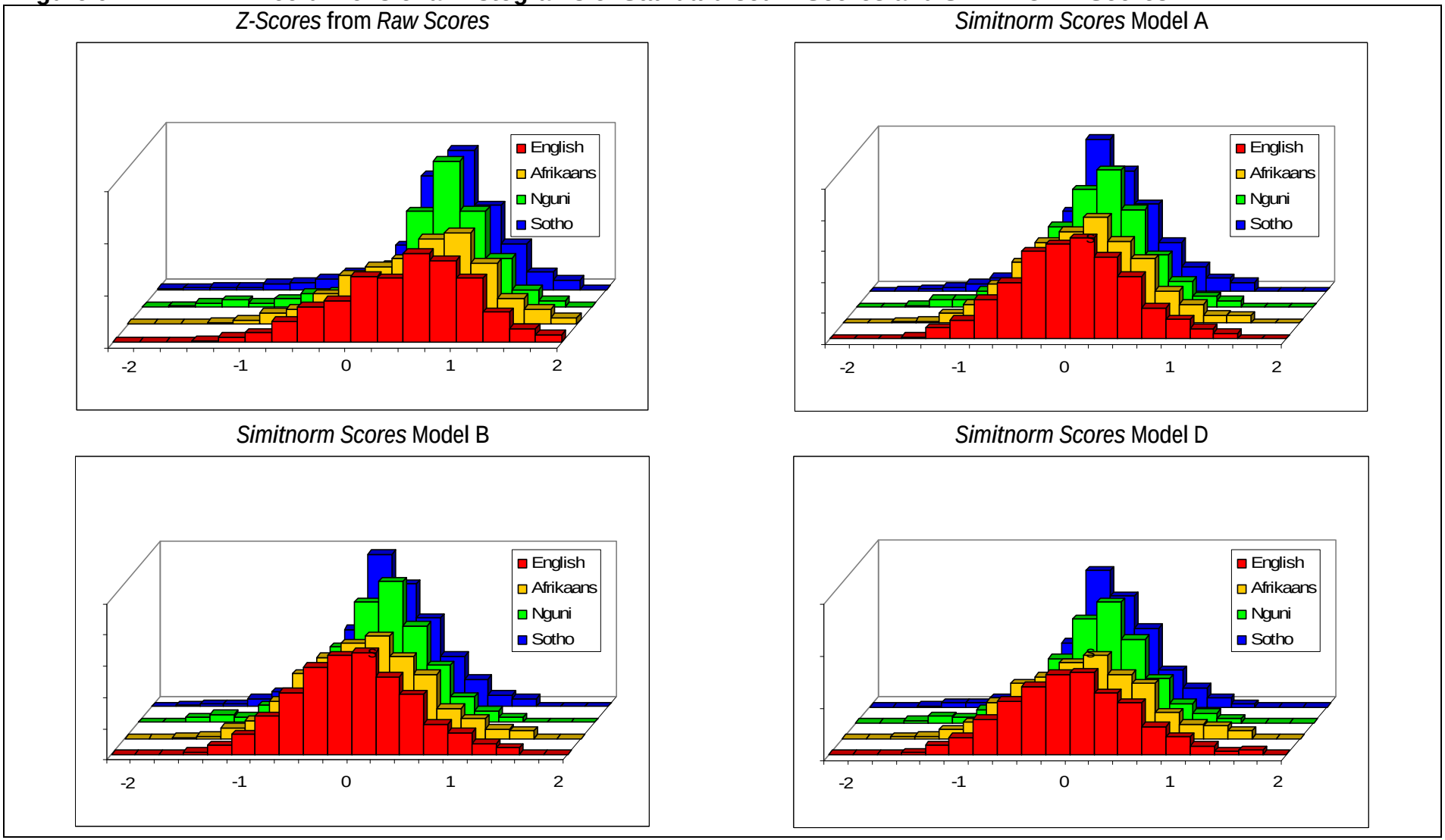
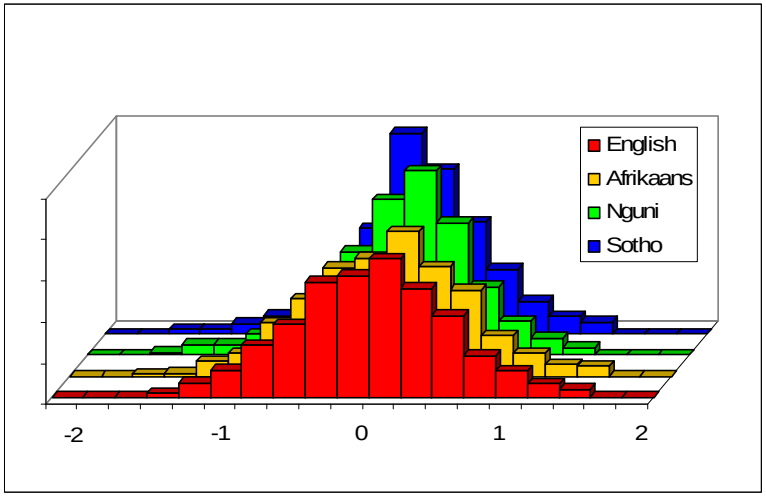
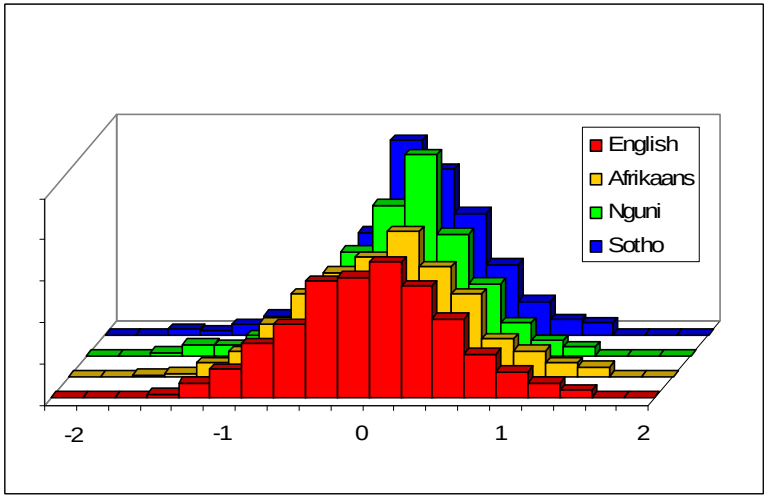


Figure 6.12(continued): Three-dimensional histograms of *Standardised Z-Scores* and *SIMITNORM Scores*

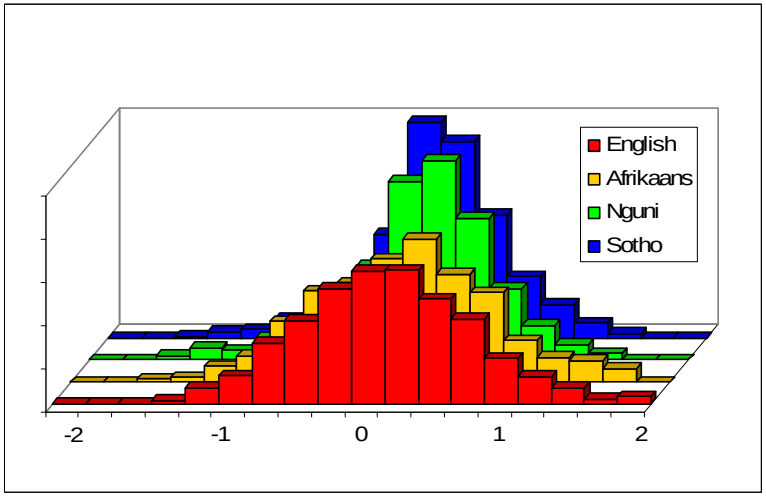
Simitnorm Scores Model BB



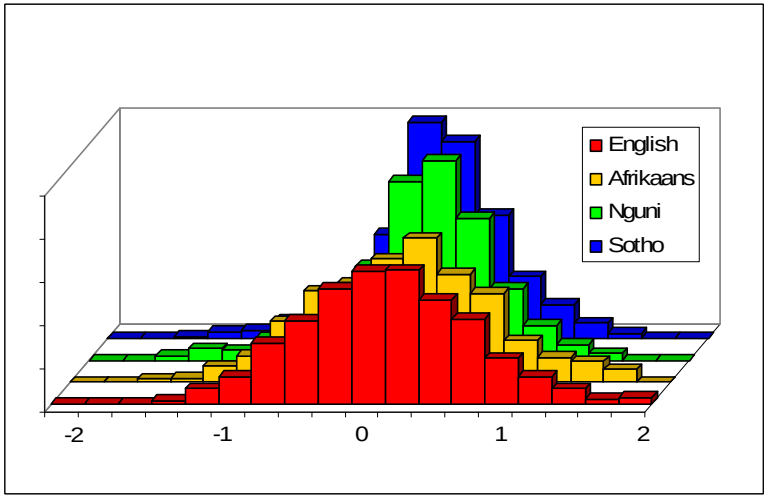
Simitnorm Scores Model DD



Simitnorm Scores Model BBB



Simitnorm Scores Model DDD



Tests for normality were applied to all the composite scores, and the results appear in Table 6.12. The Kolmogorov-Smirnov, Cramer-von-Mises and Anderson-Darling tests were applied using SAS version 9.1. For all four tests, the null hypothesis is that the data comes from a normal distribution. The null hypothesis is rejected for all four cultural groups when the *Raw Scores* and the *Standardised Z-Scores* are tested for normality. The box-and whisker plots in Figures 6.2 to 6.10, as well as the distribution functions in Figures 6.11 to 6.12 confirm the deviation from normality in the cases of the *Raw Scores* and *Standardised Z-Scores*.

In Table 6.12, it is clear that for the English and Afrikaans cultural groups, all the *SIMITNORM Scores* can be assumed to be normally distributed, but not for the Nguni and Sotho cultural groups. This may be mainly due to the fact there still remained excessive kurtosis after SIMITNORM. However, it should be noted that the normality tests are extremely sensitive to deviations from normality with increasing sample size. Therefore it would be appropriate also to do visual inspections by using normal probability plots.

The normal probability plots for the most restricted model, Model DDD, which is the “least successful” in normalisation, are given for the *Raw Scores* and for the *SIMITNORM DDD-Scores* in Figures 6.13 to 6.14. These figures indicate that the deviations from the straight line for the *DDD-Scores* are not too serious in the cases of the Nguni and Sotho groups, and that even the most restricted SIMITNORM model resulted in distributions that are closer to normal distributions.

Table 6.12: Normality tests for *Raw Scores*, *Standardised Z-Scores* and *SIMITNORM Scores*

			English		Afrikaans		Nguni		Sotho	
Variable	Test	p Type	Statistic	p-value	Statistic	p-Value	Statistic	p-Value	Statisitc	p-Value
Raw Score	Kolmogorov-Smirnov	Pr > D	0.059	0.0100	0.072	0.0100	0.144	0.0100	0.142	0.0100
Raw Score	Cramer-von Mises	Pr > W-Sq	0.671	0.0050	1.137	0.0050	7.440	0.0050	3.951	0.0050
Raw Score	Anderson-Darling	Pr > A-Sq	3.729	0.0050	5.914	0.0050	42.008	0.0050	22.149	0.0050
Z-Score	Kolmogorov-Smirnov	Pr > D	0.059	0.0100	0.072	0.0100	0.144	0.0100	0.142	0.0100
Z-Score	Cramer-von Mises	Pr > W-Sq	0.671	0.0050	1.137	0.0050	7.440	0.0050	3.951	0.0050
Z-Score	Anderson-Darling	Pr > A-Sq	3.729	0.0050	5.914	0.0050	42.008	0.0050	22.149	0.0050
A-Score	Kolmogorov-Smirnov	Pr > D	0.010	0.1500	0.016	0.1500	0.060	0.0100	0.072	0.0100
A-Score	Cramer-von Mises	Pr > W-Sq	0.014	0.2500	0.049	0.2500	1.287	0.0050	0.947	0.0050
A-Score	Anderson-Darling	Pr > A-Sq	0.138	0.2500	0.351	0.2500	8.048	0.0050	5.597	0.0050
B-Score	Kolmogorov-Smirnov	Pr > D	0.017	0.1500	0.015	0.1500	0.069	0.0100	0.077	0.0100
B-Score	Cramer-von Mises	Pr > W-Sq	0.056	0.2500	0.043	0.2500	1.836	0.0050	1.046	0.0050
B-Score	Anderson-Darling	Pr > A-Sq	0.420	0.2500	0.366	0.2500	11.297	0.0050	6.221	0.0050
D-Score	Kolmogorov-Smirnov	Pr > D	0.023	0.1290	0.017	0.1500	0.067	0.0100	0.074	0.0100
D-Score	Cramer-von Mises	Pr > W-Sq	0.078	0.2238	0.047	0.2500	1.706	0.0050	0.955	0.0050
D-Score	Anderson-Darling	Pr > A-Sq	0.565	0.1471	0.441	0.2500	10.557	0.0050	5.714	0.0050
BB-Score	Kolmogorov-Smirnov	Pr > D	0.010	0.1500	0.018	0.1500	0.063	0.0100	0.075	0.0100
BB-Score	Cramer-von Mises	Pr > W-Sq	0.018	0.2500	0.075	0.2439	1.447	0.0050	1.010	0.0050
BB-Score	Anderson-Darling	Pr > A-Sq	0.164	0.2500	0.471	0.2487	8.950	0.0050	5.997	0.0050
BBB-Score	Kolmogorov-Smirnov	Pr > D	0.016	0.1500	0.015	0.1500	0.072	0.0100	0.078	0.0100
BBB-Score	Cramer-von Mises	Pr > W-Sq	0.042	0.2500	0.054	0.2500	2.067	0.0050	1.174	0.0050
BBB-Score	Anderson-Darling	Pr > A-Sq	0.356	0.2500	0.422	0.2500	12.534	0.0050	6.967	0.0050
DD-Score	Kolmogorov-Smirnov	Pr > D	0.011	0.1500	0.018	0.1500	0.067	0.0100	0.077	0.0100
DD-Score	Cramer-von Mises	Pr > W-Sq	0.021	0.2500	0.079	0.2186	1.498	0.0050	0.985	0.0050
DD-Score	Anderson-Darling	Pr > A-Sq	0.196	0.2500	0.528	0.1863	9.263	0.0050	5.862	0.0050
DDD-Score	Kolmogorov-Smirnov	Pr > D	0.018	0.1500	0.016	0.1500	0.075	0.0100	0.081	0.0100
DDD-Score	Cramer-von Mises	Pr > W-Sq	0.056	0.2500	0.060	0.2500	2.040	0.0050	1.106	0.0050
DDD-Score	Anderson-Darling	Pr > A-Sq	0.464	0.2500	0.508	0.2084	12.385	0.0050	6.587	0.0050

Figure 6.13: Normal Probability Plot of *Raw Scores*

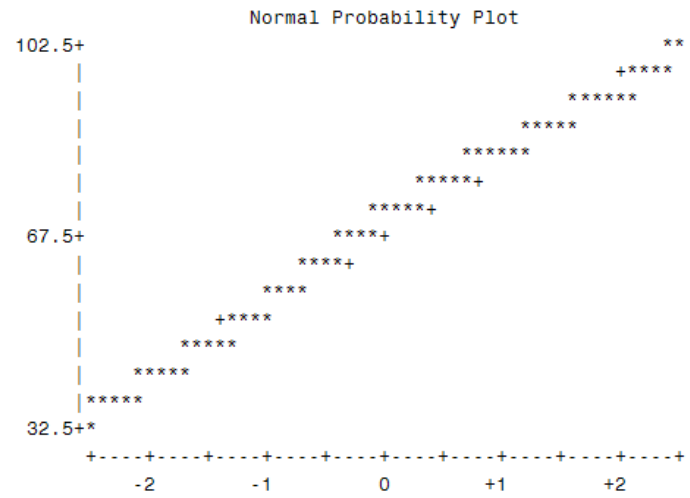
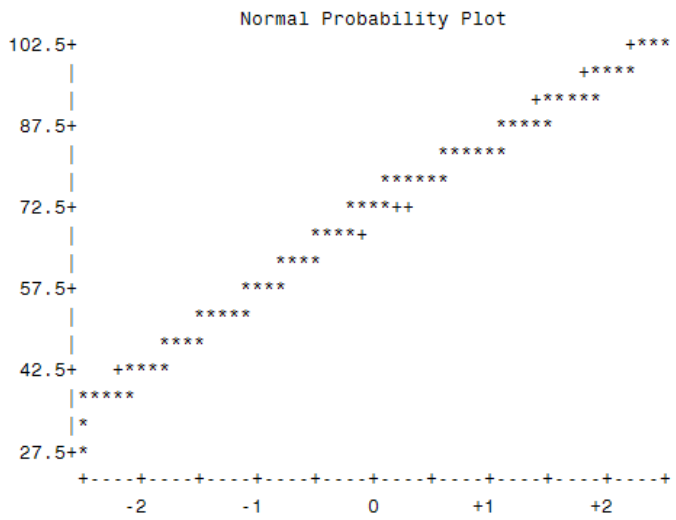
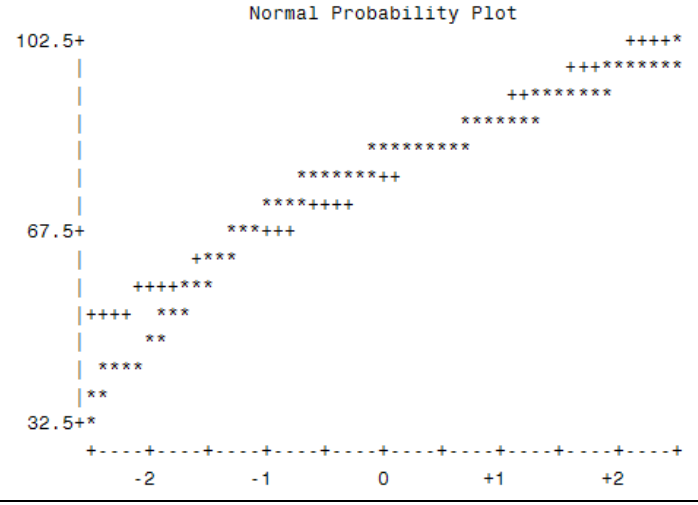
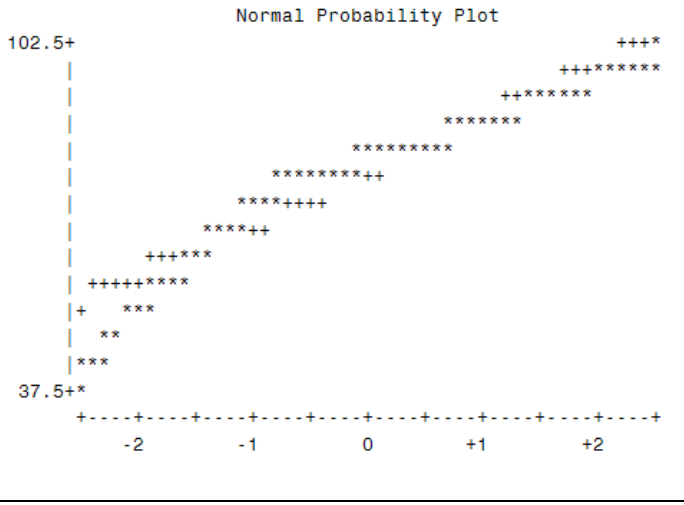
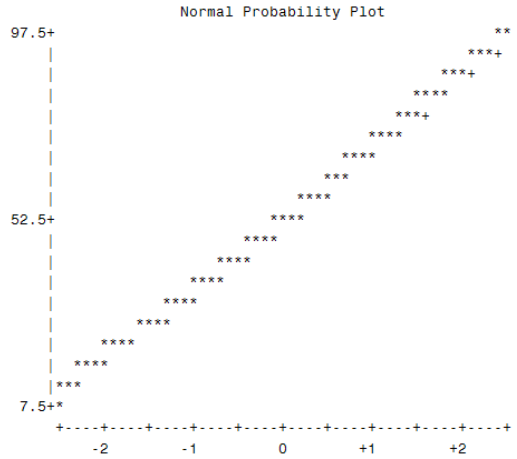
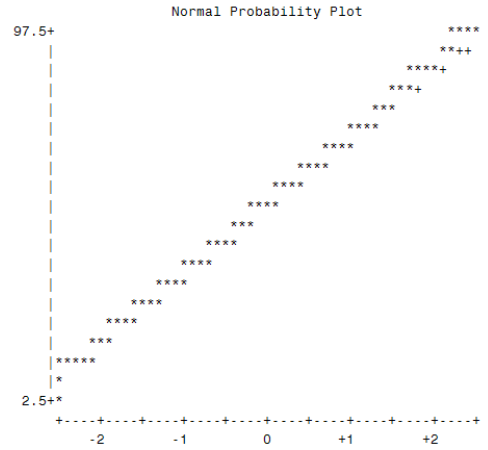
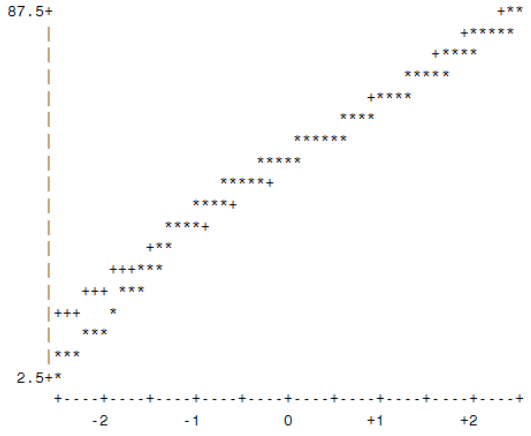
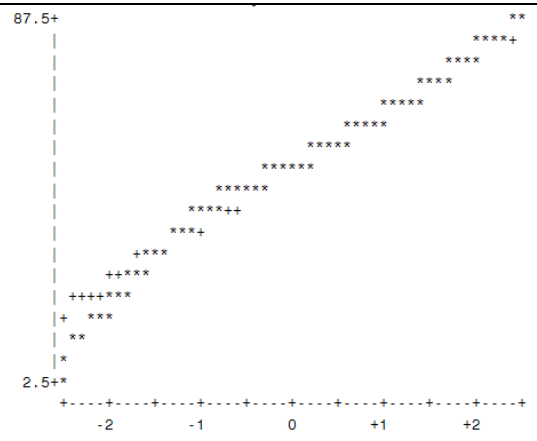
English: *Raw Scores*	Afrikaans: *Raw Scores*
	
Nguni: *Raw Scores*	**Sotho: *Raw Scores***
	

Figure 6.14: Normal Probability Plot of *SIMITNORM Scores*, Model DDD

English: *DDD-Score*	Afrikaans: *DDD-Score*
	
Nguni: *DDD-Score*	**Sotho: *DDD-Score***
	

6.4 RELIABILITY ANALYSES AND CORRELATIONS BETWEEN SCORES

Since the SIMITNORM transformation functions are all non-linear transformation functions, they will affect all methods that are based on correlations or covariances. It is therefore appropriate to examine the effect of SIMITNORM on Cronbach Alpha, as well as examining the Pearson correlation coefficients between all the composite scores.

Table 6.13 gives the Cronbach Alpha for each of the groups on the *Raw Scores*, *Z-Scores* and all the *SIMITNORM Scores*. The *Z-Score*, which is a linear transformation, left the Cronbach Alpha to remain unchanged, as is apparent when the raw score alphas and standardised score alphas are compared. As expected, SIMITNORM does affect Cronbach Alpha, due to the non-linear nature of the normalisation functions, but these differences are negligible.

Table 6.13: Cronbach Alpha Reliability Coefficient

	Cronbach Alpha			
	English	Afrikaans	Nguni	Sotho
Raw Scores	0.9135	0.9145	0.8865	0.8752
Standardised *Z-Scores*	0.9135	0.9145	0.8865	0.8752
Model A Scores	0.9120	0.9142	0.8774	0.8710
Model B Scores	0.9125	0.9147	0.8793	0.8721
Model D Scores	0.9126	0.9147	0.8788	0.8717
Model BB Scores	0.9134	0.9154	0.8781	0.8715
Model BBB Scores	0.9136	0.9155	0.8799	0.8724
Model DD Scores	0.9134	0.9155	0.8779	0.8714
Model DDD Scores	0.9137	0.9155	0.8794	0.8720

The Pearson correlation coefficients for the four cultural groups in Tables 6.14(a) to 6.14(d) are given for the *Raw Scores*, the *Standardised Z-Scores*, and the *SIMNORM Scores* as obtained in the results discussed in Chapter 5, as well as for the composite scores obtained after application of SIMITNORM. All the correlation coefficients are very significant, and they are all above 0.95. It is notable that several of the correlation coefficients are 0.999, and most are above 0.97. It is clear that the composite *SIMITNORM Scores* generally have a higher correlation with the *Raw Scores* than the *SIMNORM Scores*. The *SIMITNORM Scores* are very highly correlated with each other, once again showing that there are compelling reasons to settle for the more parsimonious models.

Table 6.14(a): Pearson Correlation Coefficients between *Raw Scores*, *Standardised Z-Scores*, *SIMNORM* and *SIMITNORM Scores* for English group

ENGLISH	RAW		SIMNORM			SIMITNORM						
Variable	Raw Score	Z-Score	SINORMA	SINORMB	SINORMD	A-Score	B-Score	BB-Score	BBB-Score	D-Score	DD-Score	DDD-Score
Raw Score	1.000											
Z-Score	1.000	1.000										
SINORMA	0.991	0.991	1.000									
SINORMB	0.982	0.982	0.997	1.000								
SINORMD	0.979	0.979	0.996	1.000	1.000							
A-Score	0.989	0.989	0.988	0.985	0.983	1.000						
B-Score	0.986	0.986	0.986	0.985	0.984	1.000	1.000					
BB-Score	0.991	0.991	0.990	0.987	0.985	1.000	0.999	1.000				
BBB-Score	0.987	0.987	0.988	0.986	0.985	0.999	1.000	0.999	1.000			
D-Score	0.985	0.985	0.986	0.984	0.983	0.999	1.000	0.999	1.000	1.000		
DD-Score	0.991	0.991	0.990	0.987	0.986	0.999	0.999	1.000	0.999	0.999	1.000	
DDD-Score	0.986	0.986	0.987	0.986	0.985	0.999	0.999	0.999	1.000	1.000	0.999	1.000

Table 6.14(b): Pearson Correlation Coefficients between *Raw Scores*, *Standardised Z-Scores*, SIMNORM and *SIMITNORM Scores* for Afrikaans group

AFRIKAANS	RAW		SIMNORM			SIMITNORM						
Variable	Raw Score	Z-Score	SINORMA	SINORMB	SINORMD	A-Score	B-Score	BB-Score	BBB-Score	D-Score	DD-Score	DDD-Score
Raw Score	1.000											
Z-Score	1.000	1.000										
SINORMA	0.988	0.988	1.000									
SINORMB	0.977	0.977	0.997	1.000								
SINORMD	0.974	0.974	0.996	1.000	1.000							
A-Score	0.986	0.986	0.985	0.981	0.979	1.000						
B-Score	0.985	0.985	0.985	0.982	0.980	1.000	1.000					
BB-Score	0.988	0.988	0.987	0.983	0.982	1.000	1.000	1.000				
BBB-Score	0.986	0.986	0.986	0.983	0.981	0.999	1.000	1.000	1.000			
D-Score	0.984	0.984	0.985	0.982	0.981	1.000	1.000	0.999	1.000	1.000		
DD-Score	0.989	0.989	0.988	0.984	0.983	0.999	0.999	1.000	1.000	0.999	1.000	
DDD-Score	0.986	0.986	0.986	0.983	0.982	0.999	0.999	0.999	1.000	1.000	1.000	1.000

Table 6.14(c): Pearson Correlation Coefficients between *Raw Scores*, *Standardised Z-Scores*, *SIMNORM* and *SIMITNORM Scores* for Nguni group

NGUNI	RAW		SIMNORM			SIMITNORM						
Variable	Raw Score	Z-Score	SINORMA	SINORMB	SINORMD	A-Score	B-Score	BB-Score	BBB-Score	D-Score	DD-Score	DDD-Score
Raw Score	1.000											
Z-Score	1.000	1.000										
SINORMA	0.949	0.949	1.000									
SINORMB	0.976	0.976	0.994	1.000								
SINORMD	0.972	0.972	0.996	1.000	1.000							
A-Score	0.968	0.968	0.967	0.975	0.975	1.000						
B-Score	0.976	0.976	0.966	0.978	0.977	0.999	1.000					
BB-Score	0.970	0.970	0.967	0.976	0.976	1.000	0.999	1.000				
BBB-Score	0.978	0.978	0.967	0.980	0.979	0.999	1.000	0.999	1.000			
D-Score	0.975	0.975	0.967	0.978	0.977	0.999	1.000	0.999	0.999	1.000		
DD-Score	0.971	0.971	0.968	0.978	0.977	0.999	0.999	1.000	0.999	0.999	1.000	
DDD-Score	0.979	0.979	0.969	0.981	0.980	0.998	0.999	0.999	1.000	0.999	0.999	1.000

Table 6.14(d): Pearson Correlation Coefficients between *Raw Scores*, *Standardised Z-Scores*, SIMNORM and *SIMITNORM Scores* for Nguni group

SOTHO	RAW		SIMNORM			SIMITNORM						
Variable	Raw Score	Z-Score	SINORMA	SINORMB	SINORMD	A-Score	B-Score	BB-Score	BBB-Score	D-Score	DD-Score	DDD-Score
Raw Score	1.000											
Z-Score	1.000	1.000										
SINORMA	0.962	0.962	1.000									
SINORMB	0.974	0.974	0.999	1.000								
SINORMD	0.971	0.971	0.999	1.000	1.000							
A-Score	0.971	0.971	0.975	0.978	0.978	1.000						
B-Score	0.975	0.975	0.974	0.979	0.978	1.000	1.000					
BB-Score	0.973	0.973	0.975	0.979	0.978	1.000	0.999	1.000				
BBB-Score	0.978	0.978	0.976	0.981	0.979	0.999	1.000	1.000	1.000			
D-Score	0.974	0.974	0.974	0.978	0.977	0.999	1.000	0.999	0.999	1.000		
DD-Score	0.974	0.974	0.976	0.980	0.979	0.999	0.999	0.999	0.999	0.999	1.000	
DDD-Score	0.978	0.978	0.977	0.981	0.980	0.999	0.999	0.999	0.999	0.999	1.000	1.000

6.5 EXPLORATORY FACTOR ANALYSIS

Exploratory factor analysis (EFA) is a very useful multivariate method to identify underlying factors or latent variables indicated by the items. EFA is based on the decomposition of the covariance matrix of the items in a questionnaire, and SIMITNORM may affect the factor pattern obtained. Tables 6.15 to 6.16 display the Varimax rotated factor pattern for the *Z-Scores* and for the most restrictive SIMITNORM Model DDD. The results of the other exploratory factor analyses appear in Appendix F. In each case, seven factors were extracted. The number of eigenvalues smaller than unity varied between 5 and 6 for the different scores.

Seven factors were extracted in an attempt obtain a factor pattern, as suggested by Schlinger (1979). It is interesting to note that all seven Schlinger (1979) dimensions were found as was originally postulated by Schlinger (1979) for all the exploratory factor analyses, as performed on the *Raw Scores*; the *Standardised Z-Scores*; as well as the *SIMITNORM Scores*. The seven factor pattern explained approximately 77% of the variation in the factor space. Note that all factor loadings smaller than 0.3 were suppressed in order to show the factor pattern more clearly.

Table 6.15: Varimax Rotated Factor Pattern: Standardised Items

Variable	Communalities	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
Z2	0.83	0.83						
Z1	0.79	0.83						
Z4	0.72	0.80						
Z3	0.75	0.77						
Z9	0.81		0.87					
Z10	0.77		0.83					
Z8	0.72		0.82					
Z21	0.83			0.83				
Z20	0.78			0.78				
Z22	0.74			0.74				
Z15	0.77				0.83			
Z16	0.77				0.79			
Z14	0.64				0.77			
Z28	0.78					0.79		
Z27	0.70					0.76		
Z29	0.76					0.75		
Z18	0.87						0.90	
Z17	0.86						0.89	
Z25	0.75							0.85
Z24	0.75							0.84
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
Eigenvalues		7.166	1.879	1.596	1.485	1.265	1.053	0.959
% Variation explained		35.8	9.4	8.0	7.4	6.3	5.3	4.8
Cum % Variation explained		35.8	45.2	53.2	60.6	67.0	72.2	77.0
Variation in Factor Space		3.124	2.365	2.265	2.200	2.156	1.775	1.516
% Variation explained		20.3	15.4	14.7	14.3	14.0	11.5	9.8
Cum % Variation explained		20.3	35.6	50.3	64.6	78.6	90.2	100.0

Table 6.16: Varimax Rotated Factor Pattern: SIMITNORM Model DDD

Variable	Communalities	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
DDD2	0.817	0.836						
DDD1	0.781	0.826						
DDD4	0.720	0.798						
DDD3	0.740	0.772						
DDD9	0.830		0.880					
DDD10	0.781		0.846					
DDD8	0.726		0.822					
DDD21	0.824			0.830				
DDD20	0.778			0.787				
DDD22	0.735			0.746				
DDD28	0.789				0.809			
DDD27	0.707				0.770			
DDD29	0.761				0.760			
DDD15	0.764					0.831		
DDD16	0.767					0.787		
DDD14	0.643					0.765		
DDD18	0.875						0.901	
DDD17	0.871						0.892	
DDD25	0.784							0.856
DDD24	0.780							0.850
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
Eigenvalues		7.093	2.033	1.593	1.464	1.231	1.072	0.988
% Variation explained		35.5	10.2	8.0	7.3	6.2	5.4	4.9
Cum % Variation explained		35.5	45.6	53.6	60.9	67.1	72.4	77.4
Variation in Factor Space		3.098	2.392	2.248	2.193	2.174	1.794	1.576
% Variation explained		20.0	15.5	14.5	14.2	14.0	11.6	10.2
Cum % Variation explained		20.0	35.5	50.0	64.2	78.2	89.8	100.0

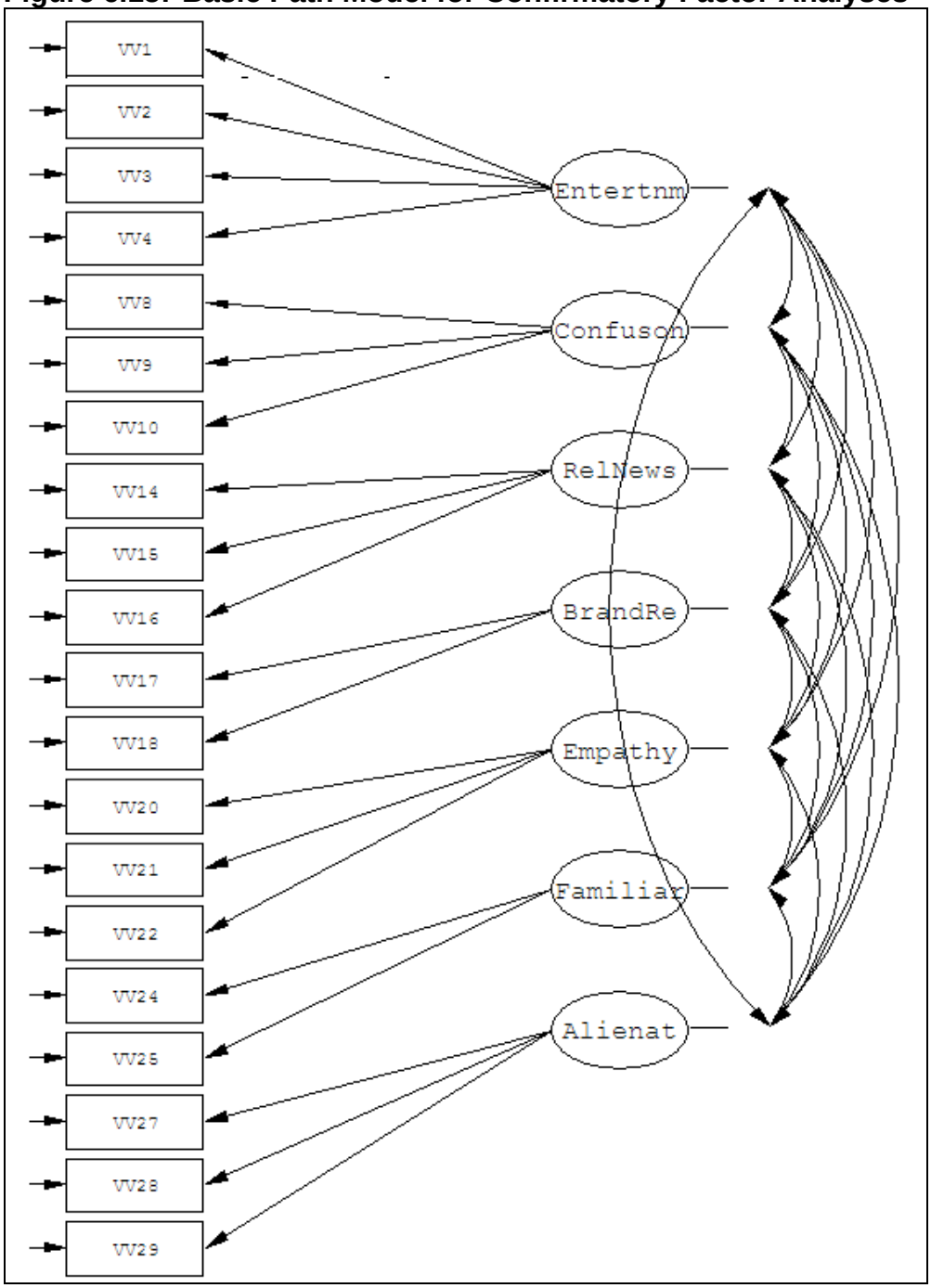
6.6 SINGLE GROUP CONFIRMATORY FACTOR ANALYSIS

In confirmatory factor analysis (CFA), the approach is the opposite of the one used in EFA. The researcher specifies the factor structure, and usually a structural equation modelling (SEM) approach is used to find estimates for the parameters and assess the fit of the specified model. A brief background to the theory in the SEM approach was covered in Chapter 2.

In this analysis, a strictly confirmatory approach was followed, in which the seven-factor model as suggested by Schlinger (1979) was specified and the fit of the model was assessed. All 4800 observations were used in the analysis reported in this section, and the cultural groups were not split into subgroups. The purpose was not to generate alternative models for the Schlinger (1979) VRP. In this analysis, the primary objective was to ascertain to what extent standardisation and SIMITNORM affect the fit of a specific CFA model. Since SIMITNORM involves non-linear transformations, it was expected that the covariance matrix would be affected, which in turn would affect the fit of the model. The purpose of this analysis was to determine how SIMITNORM affected the fit of a CFA model, and whether the results would show an improved fit to the model. It was expected that the *SIMITNORM Scores* would give an improved fit, since the ordinal scores had been transformed to normalised scores, one of the key assumptions in CFA.

LISREL 8.8 was used to estimate the parameters for the *Raw Scores*, the *Standardised Z-Scores* and all the *SIMITNORM Scores*. The basic path model that specifies the parameters to be estimated is shown in Figure 6.15. This model was fitted for all the scores. The estimated parameters can be seen in Appendix G and the goodness-of-fit measures are listed in Table 6.17. Each model gives the estimated parameters for *Raw Scores*, *Standardised Z-Scores*, and each of the seven SIMITNORM models. It is interesting to note the estimated parameters are all of the same order, in Appendix G, Tables G.1 to G.3, as one moves from left to right across the tables. The parameters have signs as expected, and are all highly significant ($p < 0.001$).

Figure 6.15: Basic Path Model for Confirmatory Factor Analyses



When the goodness-of-fit measures are considered in Table 6.17, all the models give satisfactory fit. Raykov and Marcoulides (2000) recommend that the Goodness-of-Fit Index (GFI) and Comparative Fit Index (CFI) should be in the mid 0.90 or above, to indicate that the model gives a reasonable fit to the data. According to Fabrigar *et al*, (1999), one of the most important measures to assess model fit is the Root Mean Square Error of Approximation (RMSEA) which takes

model complexity into account. RMSEA should be less than 0.05 for the model to indicate a good approximation. Values between 0.05 and 0.08 indicate acceptable fit, values between 0.08 and 0.10 marginal fit, and values above 0.10 poor fit. A very useful aspect of the RMSEA is that confidence intervals for the measure can be calculated. If the confidence interval is not too wide, and both limits are below 0.05, it can be argued that the model is a plausible means of describing the data (Hu and Bentler, 1999).

A comparison of the measures of fit in Table 6.17 shows that the models all provided a very good fit to the *Raw Scores* and *Standardised Z-Scores*. The CFI's are just above 0.99 for the *Raw Scores* and all the SIMITNORM models. The GFI's are 0.970 and 0.972 for the *Raw Scores* and *Standardised Z-Scores*, and 0.98 for all the SIMITNORM models.

The RMSEA's for the *Raw Scores* and the *Standardised Z-Scores* were 0.0398 and 0.0350 respectively. The upper limits of the 90% confidence intervals were 0.0418 and 0.0370. These were both below the recommended value of 0.05, which indicates that the models are very good approximations of the covariance structure in the data. It is very encouraging that all the SIMITNORM models had a better fit than the raw data, even the most restrictive model DDD. RMSEA's ranged between 0.0342 and 0.0345 across the models, and the upper limit of the 90% confidence interval ranged between 0.0362 and 0.0365.

It is very informative to study the Bayesian Information Criterion (BIC) (see Section 2.7.4). The smaller this number (larger negative numbers), the better the model fits Schwartz (1978) and Raftery (1995). When the BIC is studied in Table 6.17, it can be seen that Model BB provided the best fit, which had a unique item location parameter and a unique item spread parameter for all items and groups, and had four amiability parameters, one for each group.

In summary, when all the fit measures are considered, all the SIMITNORM models provided a better fit than the *Raw Scores* and the *Standardised Z-Scores*, when the data was treated in a single group analysis.

Table 6.17: Goodness-of-Fit Measures of Single Group Confirmatory Factor Analyses

	RAW		SIMITNORM						
GOODNESS OF FIT MEASURES	VV	Z	A	B	D	BB	BBB	DD	DDD
Degrees of Freedom	149	149	149	149	149	149	149	149	149
BIC	15.94	-238.60	-275.09	-275.22	-263.42	-278.04	-273.14	-264.94	-265.38
Minimum Fit Function Chi-Square	1278.92	1024.376	987.886	987.763	999.559	984.94	989.843	998.038	997.598
	(P=0.0)	(P=0.0)	(P=0.0)	(P=0.0)	(P=0.0)	(P=0.0)	(P=0.0)	(P=0.0)	(P=0.0)
Root Mean Square Error of Approximation (RMSEA)	0.0398	0.0350	0.0343	0.0342	0.0345	0.0342	0.0343	0.0344	0.0344
90 Percent Confidence Interval for RMSEA: Lower Limit	0.0378	0.0330	0.0322	0.0322	0.0324	0.0322	0.0323	0.0324	0.0324
90 Percent Confidence Interval for RMSEA: Upper Limit	0.0418	0.0370	0.0363	0.0363	0.0365	0.0362	0.0363	0.0365	0.0365
P-Value for Test of Close Fit (RMSEA < 0.05)	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
Expected Cross-Validation Index (ECVI)	0.2930	0.2390	0.2310	0.2310	0.2330	0.2300	0.2310	0.2330	0.2330
90 Percent Confidence Interval for ECVI: Lower Limit	0.2690	0.2180	0.2530	0.2110	0.2130	0.2110	0.2110	0.2130	0.2130
90 Percent Confidence Interval for ECVI: Upper Limit	0.3170	0.2180	0.2530	0.2520	0.2550	0.2520	0.2530	0.2550	0.2550
ECVI for Saturated Model	0.0875	0.0875	0.0875	0.0875	0.0875	0.0875	0.0875	0.0875	0.0875
ECVI for Independence Model	26.027	22.514	21.341	21.626	22.153	21.520	21.725	21.614	22.235
Normed Fit Index (NFI)	0.9900	0.9910	0.9900	0.9900	0.9910	0.9900	0.9910	0.9900	0.9910
Non-Normed Fit Index (NNFI)	0.9880	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900	0.9900
Parsimony Normed Fit Index (PNFI)	0.7760	0.7770	0.7770	0.7770	0.7770	0.7770	0.7770	0.7770	0.7770
Comparative Fit Index (CFI)	0.9910	0.9920	0.9920	0.9920	0.9920	0.9920	0.9920	0.9920	0.9920
Standardized RMR	0.0261	0.0228	0.0227	0.0226	0.0229	0.0226	0.0226	0.0229	0.0229
Goodness of Fit Index (GFI)	0.9740	0.9790	0.9800	0.9800	0.9800	0.9800	0.9800	0.9800	0.9800
Adjusted Goodness of Fit Index (AGFI)	0.9630	0.9710	0.9720	0.9720	0.9710	0.9720	0.9720	0.9710	0.9710
Parsimony Goodness of Fit Index (PGFI)	0.6910	0.6950	0.6950	0.6950	0.6950	0.6950	0.6950	0.6950	0.6950

6.7 MULTI-GROUP CONFIRMATORY FACTOR ANALYSES

6.7.1 Presentation of Hypotheses 1 to 6

The literature that covers the multi-group method of testing invariance across groups was presented in Chapter 2. In this section, the results of these hypotheses are reported for various sets of variables. The results reported in this section include a selection of goodness-of-fit measures, and they involved a total of 63 multi-group confirmatory factor analyses. The goodness-of-fit measures are reported in Tables 6.18 to 6.20. The group goodness-of-fit measures are reported in Appendix H.

6.7.2 Hypothesis 1: Configural Invariance

This hypothesis tests whether the basic factor structure is equivalent across groups. The results are given in the first column (H1) across Tables 6.18 to 6.20. When the results are considered, it can be inferred that the *Raw Scores*, the *Standardised Z-Scores*, and all seven *SIMITNORM Scores* possess configural invariance across the four cultural groups tested. Overall the model fitted very well across all tables. The RMSEA and both confidence limits are in the region of 0.04 or lower, NNFI, IFI and RFI (in column H1) are all between 0.979 to 0.986. It is interesting to note that these fit measures are all at least the same, or slightly better for the SIMITNORM models.

It is also necessary to consider the group goodness-of-fit measures as presented in Appendix H, Tables H.1 to H.9. For Hypothesis 1, the RMR and the Standardised RMR are all in the region of 0.026 to 0.030 across all cultural groups. The GFI ranged between 0.958 and 0.971 for the *Raw Scores* and the *Standardised Z-Scores* across the cultural groups, and dropped slightly to the region of 0.959 and 0.969 for the various SIMITNORM models. It is interesting to note that, within a specific cultural group, the GFI remains fairly stable across the different SIMITNORM models.

Overall, it can be concluded that the VRP20 possesses configural invariance across the four cultural groups tested, which means that the hypothesis that there is a similar structure across these groups is tenable. This finding is supported for *Raw Scores, Standardised Z-Scores* and all seven *SIMITNORM Scores*.

6.7.3 Hypothesis 2a: Metric Invariance

This hypothesis not only keeps the factor structure similar, but it further constrains the factor loadings to be equal across the different groups. If this hypothesis is tenable, it can be inferred that the instrument can be assumed to measure on the same scale across the different cultural groups. An inspection of the column H2a over Tables 6.18 to 6.20 is appropriate to assess the outcome of this hypothesis.

Across all tables it is clear that the model imposed in Hypothesis 2a fits the data very well. The RMSEA is 0.0432 for the *Raw Scores,* dropped to 0.0427 for *Standardised Z-Scores*, and drops slightly further to range between 0.0408 and 0.0413 for the different SIMITNORM models. The lower confidence limit for the RMSEA is 0.0411 for the *Raw Scores (VV)*; 0.0406 for the *Standardised Z-Scores*; and has dropped to a range of between 0.0387 and 0.0392 for the *SIMITNORM Scores*. The upper confidence limit for the RMSEA is 0.0453 for *Raw Scores*, 0.0448 for *Standardised Z-Scores*, and ranges between 0.0429 and 0.0434 for the seven different *SIMITNORM Scores*. Since the model is more restricted, one would expect RMSEA to be higher in Hypothesis 2a than in Hypothesis 1 for all variables, but the increases are not large.

These results indicate that the hypothesis of equal factor loadings (H2a) is tenable for *Raw Scores* and *Standardised Z-Scores*, and more tenable for *SIMITNORM Scores*. It can therefore be concluded that the VRP20 possesses metric invariance for all the scores reported in this section, and more so for *SIMITNORM Scores*.

6.7.4 Hypothesis 2b: Equal intercepts

This hypothesis keeps the factor structure similar (H1), and constrains the factor loadings to be equal across the different groups (H2a) and further constrains the

factor intercepts to be equal across groups (H2b). An inspection of the column H2b over Tables 6.18 to 6.20 is appropriate to assess the outcome of this hypothesis. If this hypothesis is not tenable, it would mean that there is bias present in the data (Cheung and Rensvold, 2000). When the results are studied for the RMSEA in Table 6.18, it is clear that the hypothesis is rejected for the *Raw Scores (VV)*, but not for the *Standardised Z-Scores* or any of the SIMITNORM models, when the RMSEA is considered. This provides strong support that there is bias present in the data, and that Standardisation and SIMITNORM were suitable to remove the bias. It further show that all the SIMITNORM models were better than standardisation.

6.7.5 Hypothesis 3: Factor Covariance and Variance Invariance

This hypothesis is even more restricted than Hypotheses 1, 2a and 2b. Not only are the factor loadings restricted to be equal, but the factor variances and covariances are restricted to be equal across cultural groups. If this hypothesis is tenable, it means that the relationships between the latent variables; and their respective variances, are very similar across the cultural groups.

The H3 column in Tables 6.18 to 6.20 reveals that the RMSEA is 0.0580 for the *Raw Scores*, which is just above the accepted 0.05 level. The RMSEA is 0.0444 for the *Standardised Z-Scores*, and range between 0.0424 (Models B and BB) and 0.0458 (Model DDD). Both limits of the 90% confidence interval for the *Raw Scores* fall above 0.05, which is higher than the generally accepted norm. For the *Standardised Z-Scores* and all the *SIMITNORM Scores*, both these limits are below 0.05, and the corresponding p-values still indicate that the model is tenable for the *Z-Scores* and the *SIMITNORM Scores*, but not for the *Raw Scores*. When all the other fit indices are considered, strong support for H3 is found for the *Z-Scores* and the *SIMITNORM Scores*, and H3 is also slightly supported for the *Raw Scores*.

When the group goodness-of-fit measures are considered in Appendix H, Tables H.1 to H.9, further support for H3 is found in the cases of the *Standardised Z-Scores* and *SIMITNORM Scores*. The RMR shows a meaningful increase across

all four cultural groups for the *Raw Scores* when compared to the value in H2, and similarly the GFI shows a meaningful decrease in Table 6.14. For the *Standardised Z-Scores* and the *SIMITNORM Scores*, there is also an increase in the RMR across all groups and all scores, but the increase is generally smaller than is the case with the *Raw Scores*. Similarly, as expected, the GFI decreases for this more restrictive model, but the decrease is less profound in the cases of the *Z-Scores* and the *SIMITNORM Scores*, than is the case in the *Raw Scores*.

Overall, one can conclude that the Schlinger (1979) VRP20 can be assumed to possess Factor Covariance and Variance Invariance for the *Standardised Z-Scores* and the *SIMITNORM Scores*, but not for the *Raw Scores*. It seems, therefore, that both standardisation and SIMITNORM procedures have succeeded in preserving a similar Factor Covariance and Variance across cultural groups.

6.7.6 Hypothesis 4: Error Variance Invariance

This hypothesis is the most restricted hypothesis, and is the last hypothesis that concerns the covariance structure. Over and above the restrictions imposed in Hypothesis 1, Hypothesis 2a and 2b; as well as Hypothesis 3, additional constraints are added to the model, namely that the errors in the observed variables are equal across the cultural groups. When the global fit results of Hypothesis 4 are considered across the *Raw Scores*, *Z-Scores* and *SIMITNORM Scores* in Tables 6.18 to 6.20, it can be observed that:

- the RMSEA and both confidence limits are above 0.06 for the *Raw Scores*;
- the RMSEA and both confidence limits are between 0.05 and 0.06 across the *Z-Scores* and all *SIMITNORM Scores*;
- the NNFI and RFI remain above 0.95 for all models, including the *Raw Scores*;
- the IFI remain above 0.97 for all models, except for the *Raw Scores*, for which IFI = 0.946; and

When the group goodness-of-fit measures are considered for Hypothesis 4 in Appendix H, Tables H.1 to H.9, the following points are observed:

- the magnitude of RMR is considerably smaller for the*Standardised Z-Scores* and the *SIMITNORM Scores*, than for the *Raw Scores* (almost half the size);
- there is a small increase in RMR across all groups from H3 to H4 for the *Raw Scores*, but there is a small decrease in magnitude for the RMR across all groups for the *Z-Scores* and the *SIMITNORM Scores*;
- there is a small decrease in the GFI for the *Raw Scores* across all four groups when H4 is compared to H3; and
- there is a surprising larger decrease in the GFI across all four groups for the *Z-Scores* and the *SIMITNORM Scores*. These results are in contrast with the results when the RMR are considered.

When all these observations are considered, the conclusions listed below may be drawn.

- The fourth hypothesis is not tenable for *Raw Scores*, *Z-Scores* or *SIMITNORM Scores*.
- The fourth hypothesis is least tenable for *Raw Scores*, and slightly less tenable for *Z-Scores* and *SIMITNORM Scores*.
- This implies that standardisation and SIMITNORM were slightly better to improve error variance invariance, but not sufficiently to obtain an acceptable model fit.
- Since this level of invariance is overly restrictive, the lack of fit in the model is not regarded as serious evidence against the merits of standardisation or SIMITNORM.

6.7.7 Hypothesis 5: Latent Mean Invariance

The fifth hypothesis of interest in this research is to test whether the latent means are equal. This should not be confused with Hypothesis 2. If Hypothesis 2 is tenable, it only means that the measurement instrument can be assumed to measure in the same units of measurement across cultural groups, but the groups may not necessarily have the same latent means across groups. Therefore, even

if the covariance structures are similar, the average levels on the latent variables across the groups may still be different. In order to test Hypothesis 5, the model is specified so that all the latent means are constrained equal across the groups. If this model fits poorly, it can be assumed that the latent means are significantly different across groups.

From the observed means of the observed items, as set out in Table 5.9, it is expected that Hypothesis 5 will be rejected for the *Raw Scores*. It is also expected, that Hypothesis 5 will be preferable to Hypothesis 4 for the *Z-Scores* and the *SIMITNORM Scores*, since these methods are transformations of the raw data that had deliberately set all the item means equal to zero. The results, appearing in Tables 6.18 to 6.20, confirm the expectations. Hypothesis 5 is totally rejected in Table 6.18. However, slightly better fit measures were obtained for the *Standardised Z-Scores* for Hypothesis when compared to Hypothesis 4; and similarly for the *SIMITNORM Scores*. This is to be expected, since the procedures were intended to set the latent means equal across the groups.

6.7.8 Hypothesis 6: Latent Mean Invariance, scalar invariance, equal intercepts and unequal error variances

The sixth hypothesis that was tested was a nested model that entailed H1, H2, H2a, H3 and H5. Steenkamp and Baumgartner (1998) have indicated that H4 may be an overly strict hypothesis, and it would be more realistic to test for H1: similar structure; H2: equal factor loadings; H2a: equal intercepts; H3: equal variances and covariances between latent variables; and H5: equal latent means. When the outcome of this hypothesis is compared across the different models, it is clear that the hypothesis H6 is rejected for raw scores, and tenable for Standardised Z-scores and all SIMITNORM models. It is also clear that all the SIMITNORM models performed better than standardisation, with model BB performing the best. This clearly indicates that a model with different amiability parameters for each group is preferable to a model with this parameter constrained across groups.

6.7.9 Implications of Hypotheses 1 to 6 in terms of Equivalence

The results presented above lead to the following conclusions: SIMITNORM helped to achieve improvements over all the hypotheses when compared to *Raw Scores* in terms of equivalence. Configural invariance (H1), metric invariance (H2) and factor covariance and variance invariance (H3), improved remarkably after standardisation and SIMITNORM. Error variance invariance (H4) was slightly improved by standardisation and SIMITNORM, but it was still better than using *Raw Scores*. Complete scalar or latent mean invariance (H5) was achieved through standardisation and SIMITNORM.

When these results are considered, the results between standardisation and SIMITNORM seem very similar. However, one of the models, Model BB, was superior in most respects, at least in the MG CFA analyses. This model had one parameter fixed within a cultural group across all items. This implies that the amiability parameter is kept constant across all items within the cultural group (the γ-parameter), with the level parameters (ϕ_1's) and the spread parameters (ϕ_2's) free to vary across items and groups. There seems to some inherent logic in this model for the data set used in this thesis.

In order to evaluate the effectiveness of SIMNORM and SIMITNORM on item bias, and to compare the results with the standardisation method, the ANOVA method, as described in Chapter 2, is followed to assess the results of SIMNORM. These results are given and discussed in Section 6.8.

Table 6.18 Root Mean Square Error of Approximation (RMSEA)

Root Mean Square Error of Approximation (RMSEA)							
Model	H1	H2a	H2b	H3	H4	H5	H6
VV	0.0412	0.0432	0.0515	0.0580	0.0679	0.0760	0.0687
Z	0.0412	0.0427	0.0409	0.0444	0.0552	0.0543	0.0436
A	0.0404	0.0412	0.0393	0.0428	0.0555	0.0546	0.0420
B	0.0399	0.0408	0.0390	0.0424	0.0551	0.0543	0.0418
D	0.0401	0.0412	0.0393	0.0460	0.0536	0.0528	0.0451
BB	0.0399	0.0408	0.0390	0.0424	0.0551	0.0543	0.0416
BBB	0.0400	0.0409	0.0391	0.0426	0.0547	0.0538	0.0418
DD	0.0399	0.0413	0.0395	0.0448	0.0570	0.0561	0.0440
DDD	0.0399	0.0412	0.0394	0.0458	0.0546	0.0537	0.0450
90 Percent Confidence Interval for RMSEA (Lower Limit)							
Model	H1	H2a	H2b	H3	H4	H5	H6
VV	0.0390	0.0411	0.0495	0.0562	0.0661	0.0743	0.0669
Z	0.0390	0.0406	0.0388	0.0425	0.0534	0.0525	0.0417
A	0.0382	0.0391	0.0373	0.0409	0.0538	0.0529	0.0401
B	0.0378	0.0387	0.0369	0.0405	0.0534	0.0525	0.0399
D	0.0379	0.0391	0.0373	0.0441	0.0519	0.0510	0.0432
BB	0.0378	0.0387	0.0369	0.0405	0.0534	0.0525	0.0397
BBB	0.0378	0.0388	0.0371	0.0407	0.0529	0.0521	0.0399
DD	0.0377	0.0392	0.0375	0.0430	0.0553	0.0544	0.0421
DDD	0.0377	0.0391	0.0374	0.0440	0.0528	0.0520	0.0431
90 Percent Confidence Interval for RMSEA (Upper Limit)							
Model	H1	H2a	H2b	H3	H4	H5	H6
VV	0.0433	0.0453	0.0534	0.0599	0.0696	0.0777	0.0705
Z	0.0433	0.0448	0.0429	0.0463	0.0569	0.0560	0.0455
A	0.0426	0.0433	0.0414	0.0447	0.0573	0.0564	0.0439
B	0.0421	0.0429	0.0411	0.0443	0.0569	0.0560	0.0437
D	0.0423	0.0433	0.0414	0.0479	0.0554	0.0545	0.0470
BB	0.0421	0.0429	0.0411	0.0443	0.0569	0.0560	0.0435
BBB	0.0422	0.0430	0.0412	0.0445	0.0565	0.0556	0.0437
DD	0.0421	0.0434	0.0416	0.0468	0.0588	0.0579	0.0459
DDD	0.0421	0.0434	0.0415	0.0477	0.0564	0.0555	0.0468
P-Value for Test of Close Fit (RMSEA < 0.05)							
Model	H1	H2a	H2b	H3	H4	H5	H6
VV	1.000	1.000	0.110	0.000	0.000	0.000	0.000
Z	1.000	1.000	1.000	1.000	0.000	0.000	1.000
A	1.000	1.000	1.000	1.000	0.000	0.000	1.000
B	1.000	1.000	1.000	1.000	0.000	0.000	1.000
D	1.000	1.000	1.000	1.000	0.000	0.005	1.000
BB	1.000	1.000	1.000	1.000	0.000	0.000	1.000
BBB	1.000	1.000	1.000	1.000	0.000	0.000	1.000
DD	1.000	1.000	1.000	1.000	0.000	0.000	1.000
DDD	1.000	1.000	1.000	1.000	0.000	0.000	1.000

Table 6.19 Other fit measures

Non-Normed Fit Index (NNFI) (Tucker and Lewis)							
Model	H1	H2a	H2b	H3	H4	H5	H6
VV	0.986	0.985	0.979	0.973	0.962	0.951	0.960
Z	0.986	0.985	0.987	0.984	0.975	0.976	0.985
A	0.986	0.986	0.987	0.984	0.973	0.974	0.985
B	0.986	0.986	0.987	0.985	0.974	0.975	0.985
D	0.986	0.986	0.987	0.982	0.976	0.976	0.983
BB	0.986	0.986	0.987	0.985	0.974	0.975	0.985
BBB	0.986	0.986	0.987	0.985	0.974	0.975	0.985
DD	0.986	0.986	0.987	0.983	0.972	0.973	0.983
DDD	0.986	0.986	0.987	0.982	0.975	0.976	0.983
Incremental Fit Index (IFI) (Bollen)							
Model	H1	H2a	H2b	H3	H4	H5	H6
VV	0.989	0.987	0.981	0.973	0.959	0.946	0.959
Z	0.989	0.988	0.988	0.984	0.973	0.973	0.984
A	0.989	0.988	0.988	0.984	0.971	0.971	0.984
B	0.989	0.988	0.989	0.985	0.972	0.972	0.985
D	0.989	0.988	0.988	0.982	0.974	0.974	0.982
BB	0.989	0.988	0.989	0.985	0.972	0.972	0.985
BBB	0.989	0.988	0.989	0.985	0.972	0.973	0.985
DD	0.989	0.988	0.988	0.983	0.970	0.970	0.983
DDD	0.989	0.988	0.988	0.982	0.973	0.973	0.982
Relative Fit Index (RFI)							
Model	H1	H2a	H2b	H3	H4	H5	H6
VV	0.980	0.978	0.972	0.966	0.955	0.944	0.954
Z	0.980	0.979	0.980	0.977	0.968	0.969	0.978
A	0.979	0.979	0.980	0.977	0.966	0.967	0.978
B	0.979	0.979	0.980	0.978	0.967	0.968	0.978
D	0.979	0.979	0.980	0.975	0.969	0.970	0.976
BB	0.979	0.979	0.980	0.978	0.967	0.968	0.978
BBB	0.980	0.979	0.980	0.978	0.968	0.968	0.978
DD	0.979	0.979	0.980	0.976	0.965	0.966	0.976
DDD	0.980	0.979	0.980	0.975	0.968	0.969	0.976

Table 6.20 Fit measures based on Chi-Square value

Bayesian Information Criterion							
Model	H1	H2a	H2b	H3	H4	H5	H6
VV	-3246.6	-3335.8	-2939.2	-2609.6	-1538.0	-207.1	-1269.7
Z	-3246.6	-3383.7	-3714.3	-3875.9	-3124.5	-3302.5	-4053.9
A	-3295.2	-3478.9	-3809.5	-4005.4	-3064.1	-3242.1	-4183.4
B	-3316.4	-3497.3	-3827.9	-4028.2	-3092.2	-3270.2	-4194.8
D	-3310.9	-3471.2	-3801.8	-3743.2	-3319.9	-3497.9	-3921.2
BB	-3316.4	-3497.3	-3827.9	-4028.2	-3092.2	-3270.2	-4206.2
BBB	-3312.0	-3490.5	-3821.1	-4015.3	-3150.5	-3328.5	-4193.3
DD	-3320.6	-3458.9	-3789.5	-3838.8	-2928.9	-3106.9	-4016.8
DDD	-3318.5	-3463.1	-3793.7	-3752.0	-3197.2	-3375.2	-3930.0
Minimum Fit Function Chi-Square							
Df	596	635	674	758	818	839	779
Model	H1	H2a	H2b	H3	H4	H5	H6
VV	1805.3	2046.7	2773.9	3815.5	5395.6	6904.6	5333.406
Z	1805.3	1998.8	1998.8	2549.2	3809.2	3809.2	2549.237
A	1756.8	1903.6	1903.6	2419.7	3869.5	3869.5	2419.653
B	1735.5	1885.2	1885.2	2396.9	3841.5	3841.5	2408.302
D	1741.1	1911.3	1911.3	2681.9	3613.7	3613.7	2681.926
BB	1735.5	1885.2	1885.2	2396.9	3841.5	3841.5	2396.875
BBB	1739.9	1892.0	1892.0	2409.8	3783.2	3783.2	2409.828
DD	1731.3	1923.6	1923.6	2586.3	4004.8	4004.8	2586.333
DDD	1733.4	1919.4	1919.4	2673.1	3736.5	3736.5	2673.057

6.8 THE EFFECT OF SIMITNORM ON ITEM BIAS

The primary purpose of SIMITNORM is to remove bias in order to obtain equivalence in marketing scales. It is therefore important to use the scoring group method described in Section 2.3.2, in order to ascertain which of the seven SIMITNORM models were successful in “removing” bias from the data. In Section 5.2, the initial item bias analysis was based on all 32 items. Subsequent analyses in Chapter 5 have led to the exclusion of a number of items, so that this section only presents the item bias analyses of the selected VRP20 items, using the same two subgroups used in Section 5.1.4, namely the Euro-centric Afrikaans and English groups in one group, (white group), and the Afro-centric Nguni and Sotho group (black group). This was done since the graphs require that the differences between the group means are plotted against score group levels. When four groups are involved, it would require a calculation of all the differences between each possible pair, which implies in total six plots per item. It is obvious that this

becomes too cumbersome, and therefore only the two major cultural groups as mentioned earlier are used in this analysis.

The results of the analyses presented here required a number of steps, and the details are described below.

- All negatively phrased items were reverse coded.
- Total scores were calculated for *Raw Scores*, *Standardised Z-Scores*, and each of the seven SIMITNORM models.
- The total scores were then categorised into *Score Levels*, with approximately equal numbers of persons per score level, and approximately equal interval lengths, ignoring the highest and the lowest score groups, but also so that there were not less than 50 persons in a *Score Level*.
- For the *Raw Scores*, 20 ANOVA's were done, one for each item.
- The dependent variable in the ANOVA was the item score, and the independent variables were the *Score Level*, the *Population Group* variable (white/black), and also the interaction term between *Score Level* and *Population Group*.
- *Score Level* was always a significant predictor, and it should be, otherwise the item does not discriminate across score levels, and since these were all highly significant (<0.0001), they were not reported here.
- If the Popgroup variable was significant, but the interaction term was not significant ($\alpha = 0.05$), then the item was assumed to possess uniform bias.
- If the interaction term was significant, it was assumed that the item possessed non-uniform bias.
- The steps described above were repeated for the *Standardised Z-Scores*, and for each of the seven SIMITNORM models.

Overall, the results presented in Table 6.21 and 6.22 give the results for *Raw Scores*, *Standardised Z-Scores* and each of the seven SIMITNORM models, for 20 items each, in total 180 ANOVA analyses. The interest lies primarily with the significance levels of the *Population Group (Popgroup)* term and the interaction term between *Popgroup* and *Score Level*. The results are presented separately for items with uniform bias and those with non-uniform bias, to ease interpretation.

The purpose of the analysis was to determine the extent to which SIMITNORM could remove bias from the observed items, and how it compared to the *Standardised Z-Score*.

Uniform bias is shown when only the grouping variable (in this case *Popgroup*) is significant, but not the interaction term between *Popgroup* and *Score Level*. Uniform bias was only present for five items, namely items 2, 15, 16, 21 and 22. The rest of the items showed non-uniform bias and are listed in Table 6.22, which is indicated when both *Popgroup* and the interaction term between *Popgroup* and *Score Level* are significant.

When Table 6.21 is studied, it seems that both standardisation and all seven SIMITNORM models were successful in removing uniform bias from the data, since the significant main effect of *Popgroup* became insignificant for the *Standardised Z-Score* and for all seven SIMITNORM models.

In the second part of Table 6.21, where the significance of the interaction terms is given, the results are somewhat surprising. It seems that the *Standardised Z-Score*, and also all the SIMITNORM models have introduced non-uniform bias into the data, especially with the more restrictive SIMITNORM models.

When Table 6.22 is considered, it is clear that both standardisation and all seven SIMITNORM models were successful in removing bias at the mean level from the data. This can be seen in the first part of the table, where the significance of the main effect of *Popgroup* is reported. All the significant *Popgroup* effects became insignificant after standardisation and SIMITNORM, and the single insignificant *Popgroup* effect for Item 4 also remained insignificant after standardisation and normalisation. When the interaction terms are considered in the second part of Table 6.22, it is clear that standardisation and SIMITNORM were not as successful in removing non-uniform bias. Several items still showed significant interaction terms.

Table 6.21: Significance of *Popgroup* main effect and *Popgroup*Score Level* interaction effect in ANOVA bias detection for items with uniform bias

ITEMS WITH UNIFORM BIAS				SIMITNORM Models						
Dependent Variable	Main Effect	Raw Score	Z-score	Model A	Model B	Model D	Model BB	Model DD	Model BBB	Model DDD
Item 2	Popgroup	0.000	0.493	0.865	0.853	0.585	0.804	0.689	0.780	0.546
Item 15	Popgroup	0.000	0.994	0.999	0.997	0.997	0.999	0.994	0.995	0.988
Item 16	Popgroup	0.000	0.981	0.996	0.956	0.988	0.991	0.994	0.967	0.968
Item 21	Popgroup	0.000	0.985	0.966	0.976	0.984	0.991	0.992	0.995	0.993
Item 22	Popgroup	0.000	0.899	0.930	0.945	0.906	0.958	0.917	0.965	0.973
Dependent Variable	Interaction Effect	Raw Score	Z-score	Model A	Model B	Model D	Model BB	Model DD	Model BBB	Model DDD
Item 2	Popgroup* Score Level	0.097	0.002	0.009	0.010	0.003	0.008	0.001	0.024	0.002
Item 15	Popgroup* Score Level	0.707	0.655	0.283	0.235	0.092	0.334	0.346	0.077	0.113
Item 16	Popgroup* Score Level	0.432	0.864	0.591	0.421	0.511	0.492	0.601	0.276	0.413
Item 21	Popgroup* Score Level	0.079	0.000	0.187	0.022	0.000	0.011	0.001	0.034	0.000
Item 22	Popgroup* Score Level	0.260	0.013	0.024	0.003	0.000	0.002	0.000	0.003	0.001

Significance level $\alpha = 0.01$

Table 6.22: Significance of *Popgroup* main effect and *Popgroup*Score Level* interaction effect in ANOVA bias detection for items with non-uniform bias

ITEMS WITH NON-UNIFORM BIAS				SIMITNORM Models						
Dependent Variable	Main Effect	Raw Score	Z-score	Model A	Model B	Model D	Model BB	Model DD	Model BBB	Model DDD
ITEM 1	Popgroup	0.000	0.579	0.868	0.633	0.822	0.686	0.774	0.750	0.596
ITEM 3	Popgroup	0.000	0.951	0.991	0.934	0.996	0.962	0.993	0.989	0.934
ITEM 4	Popgroup	0.315	0.769	0.941	0.809	0.935	0.888	0.927	0.926	0.855
ITEM 8	Popgroup	0.000	0.970	0.978	0.968	0.982	0.997	0.999	0.998	0.994
ITEM 9	Popgroup	0.000	0.950	0.999	0.997	0.998	0.970	0.956	0.948	0.908
ITEM 10	Popgroup	0.000	0.975	0.979	0.990	0.988	0.996	0.999	0.989	0.991
ITEM 14	Popgroup	0.000	0.656	0.905	0.829	0.781	0.826	0.797	0.661	0.635
ITEM 17	Popgroup	0.001	0.854	0.891	0.835	0.925	0.932	0.908	0.903	0.776
ITEM 18	Popgroup	0.000	0.709	0.891	0.786	0.921	0.887	0.867	0.889	0.716
ITEM 20	Popgroup	0.000	0.957	0.962	0.990	0.981	0.985	0.978	0.988	0.999
ITEM 24	Popgroup	0.000	0.828	0.930	0.734	0.880	0.949	0.950	0.904	0.798
ITEM 25	Popgroup	0.000	0.751	0.833	0.624	0.808	0.944	0.902	0.877	0.656
ITEM 27	Popgroup	0.000	0.965	0.987	0.965	0.942	0.998	0.998	0.991	0.997
ITEM 28	Popgroup	0.000	0.905	0.994	0.983	0.971	0.998	0.999	0.991	0.997
ITEM 29	Popgroup	0.000	0.944	0.987	0.947	0.998	0.906	0.954	0.969	0.847

Table 6.22 (continued): Significance of *Popgroup* main effect and *Popgroup*Score Level* interaction effect in ANOVA bias detection for items with non-uniform bias

ITEMS WITH NON-UNIFORM BIAS				SIMITNORM Models						
Dependent Variable	Interaction Effect	Raw Score	Z-score	Model A	Model B	Model D	Model BB	Model DD	Model BBB	Model DDD
ITEM 1	Popgroup* Score Level	0.003	0.000	0.016	0.003	0.020	0.002	0.006	0.008	0.001
ITEM 3	Popgroup* Score Level	0.005	0.069	0.045	0.000	0.058	0.002	0.076	0.055	0.000
ITEM 4	Popgroup* Score Level	0.023	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ITEM 8	Popgroup* Score Level	0.000	0.315	0.029	0.010	0.138	0.006	0.032	0.035	0.015
ITEM 9	Popgroup* Score Level	0.000	0.074	0.189	0.006	0.395	0.000	0.146	0.620	0.002
ITEM 10	Popgroup* Score Level	0.000	0.000	0.001	0.000	0.002	0.000	0.012	0.003	0.000
ITEM 14	Popgroup* Score Level	0.033	0.474	0.011	0.014	0.032	0.000	0.002	0.001	0.011
ITEM 17	Popgroup* Score Level	0.013	0.085	0.775	0.486	0.337	0.678	0.494	0.539	0.756
ITEM 18	Popgroup* Score Level	0.001	0.487	0.573	0.318	0.573	0.415	0.603	0.408	0.671
ITEM 20	Popgroup* Score Level	0.027	0.000	0.024	0.000	0.001	0.000	0.019	0.001	0.000
ITEM 24	Popgroup* Score Level	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ITEM 25	Popgroup* Score Level	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ITEM 27	Popgroup* Score Level	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
ITEM 28	Popgroup* Score Level	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
ITEM 29	Popgroup* Score Level	0.000	0.000	0.004	0.002	0.002	0.001	0.000	0.002	0.000

Significance level $\alpha = 0.05$

The researcher adopted a strategy to help interpret the results of Tables 6.21 and 6.22 further. Table 6.23 was used to classify the items into three groups, namely five items with uniform bias, three borderline items (4, 14 and 20), and the remaining twelve items with non-uniform bias. According to Table 6.23, it seems that SIMITNORM was completely effective in removing the bias as reflected in the main effect of *Popgroup*. When the interaction term is considered, however, the results were rather disappointing. For seven of the 20 items, namely items 4, 20, 24, 25, 27, 28 and 29, the interaction terms were not insignificant after SIMITNORM, indicating that it was not possible to remove non-uniform bias with SIMITNORM in seven of the 20 items.

Table 6.23: Classification of items in terms of the success of SIMITNORM in removing bias, considering the interaction term Popgroup by Score Level, using the ANOVA approach

Success to transform the MAIN EFFECT to an insignificant effect			
	Insignificant after SIMITNORM	**In-between**	**Still significant after SIMITNORM**
Items with uniform bias	2, 15, 16, 21, 22		
Borderline items	4, 14, 20		
Items with non-uniform bias	Rest of items		
Success to transform the INTERACTION TERM to an insignificant effect			
	Insignificant after SIMITNORM	**In-between**	**Still significant after SIMITNORM**
Items with uniform bias	15, 16	2, 21, 22	
Borderline items		14	4, 20
Items with non-uniform bias	17, 18	1, 3, 8, 9, 10	24, 25, 27, 28, 29

In order to investigate these results further, item bias plots were done for each of the 20 items. This involved the calculation of the average item score for each population group (white/black), and then calculating the differences between the mean levels, and then plotting the differences against the score level for the entire

scale, using the same approach as in the ANOVA analysis. In these plots, the averages of whites were subtracted from the averages of blacks, to obtain a difference score. If the difference score lies above the zero line on the plot, it indicates that at that score level of the entire scale, blacks had a higher mean item value than whites.

Consider the item bias plot of item 16, an item with uniform bias, shown in Figure 6.16, and ignore the first and last points plotted for the *Raw Scores*. Since all the other points lie above the zero line, it means that across all score levels of the entire scale, blacks consistently rated higher item mean levels for item 16 across all score levels, which was phrased: “During the commercial I thought how that product might be useful to me”.

It is also worth noting the standardisation plots and each of the seven SIMITNORM models of item 16 in Figure 6.16. These plots all represent a more-or-less random pattern around the zero line, which indicates the absence of bias after standardisation and SIMITNORM for item 16.

Item 2 as shown in Appendix I, Table I.2, is another item for which uniform bias was reported, but when the plots are observed, it is not clear why the ANOVA result indicated uniform bias. For this item, there seems to be a trend across score levels. Similar patterns are observed for items 15, 21 and 22 that are set out in Appendix I.

Also notable is an item with non-uniform bias, for example item 18, with the item bias plot given in Figure 6.18. In the *Raw Score* plot, when the first two points are ignored, there is a consistent downward trend across score levels. It means that at the lower score levels, blacks on average gave higher average item scores than whites, but at the highest two levels, blacks gave lower average item scores than whites. When the standardisation and the SIMITNORM plots are considered, it seems that these methods were all successful in removing the bias from this item.

Figure 6.16: Item Bias Analysis - Item 16

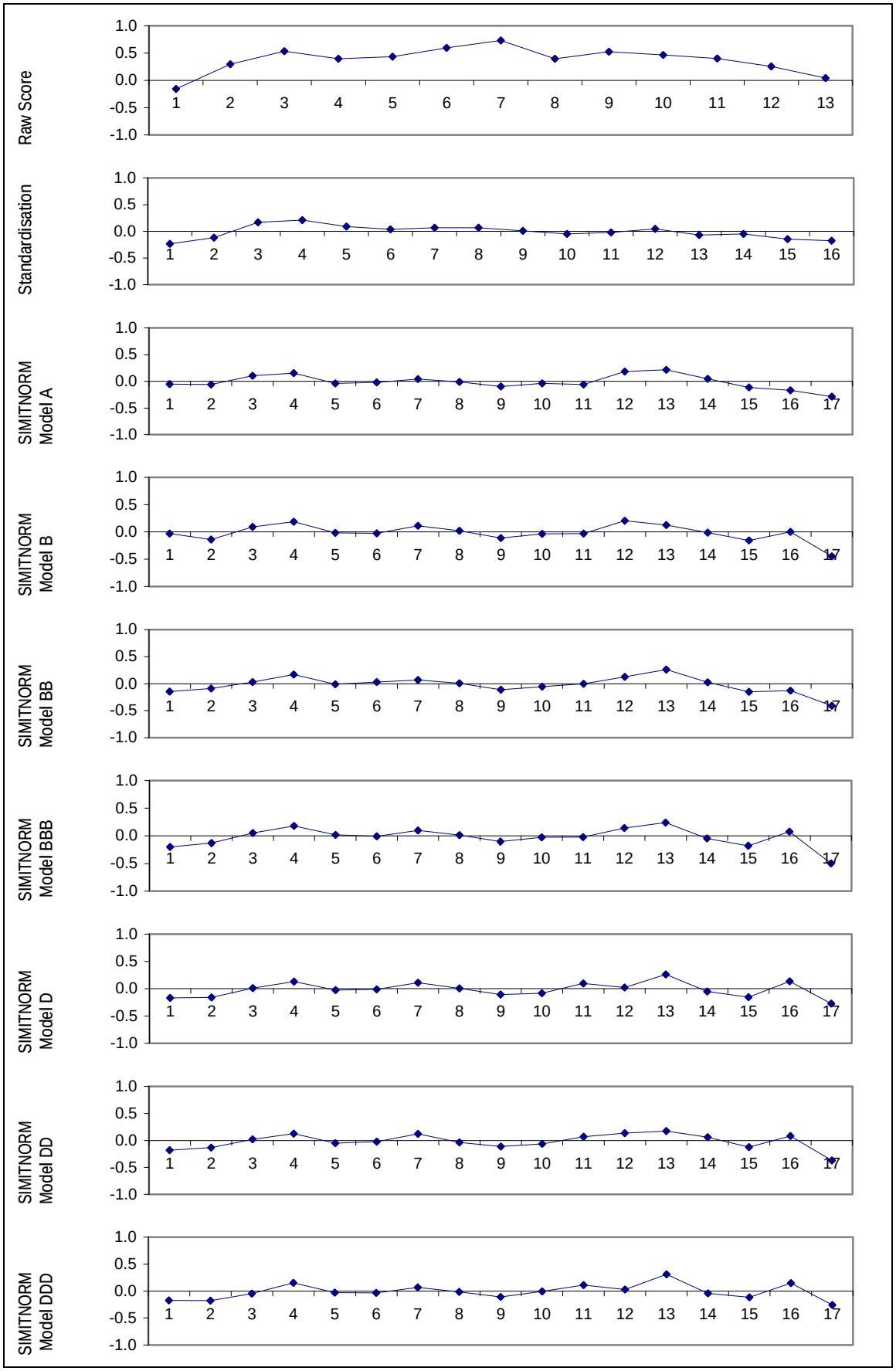


Figure 6.17: Item Bias Analysis - Item 18

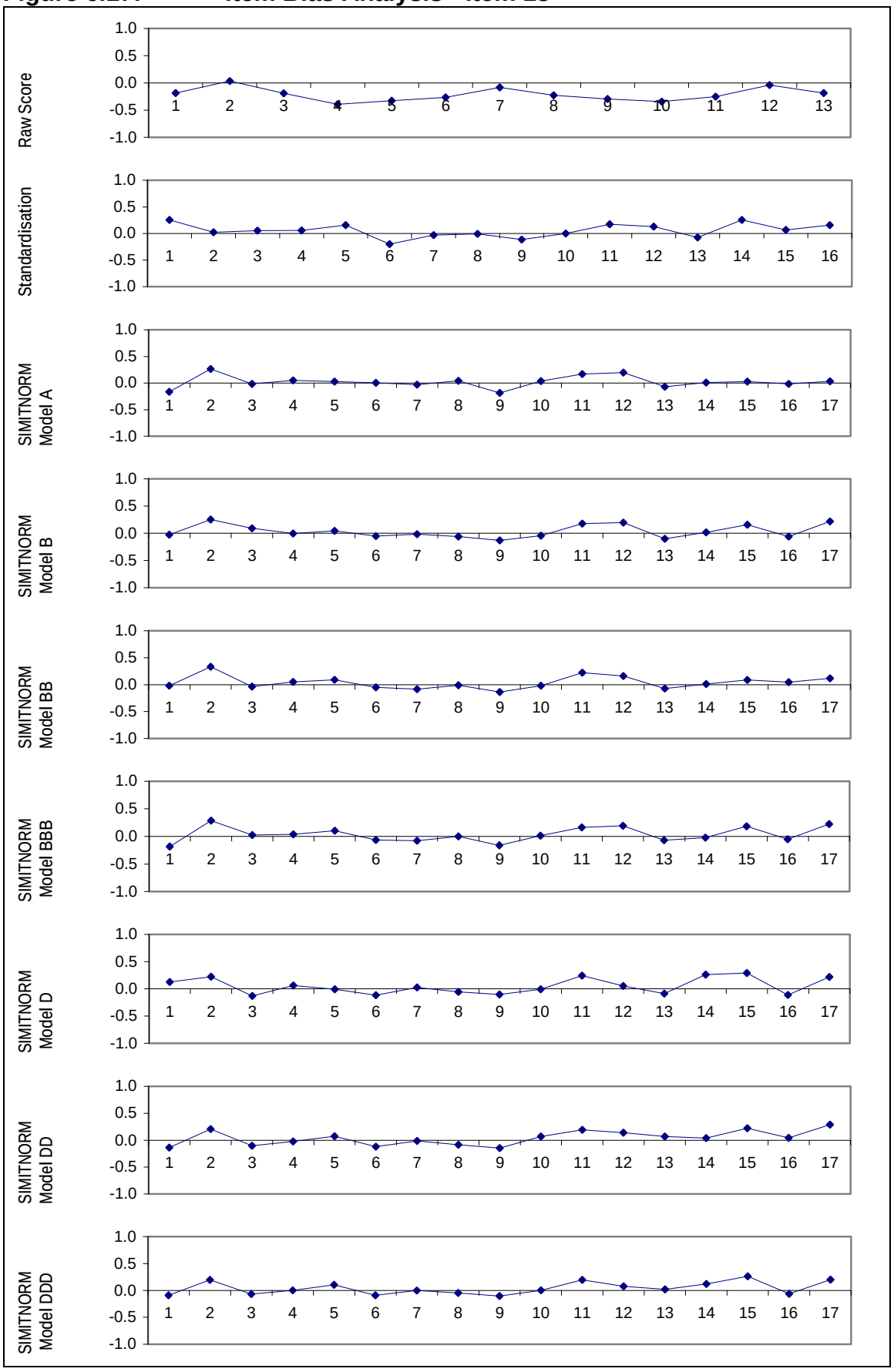
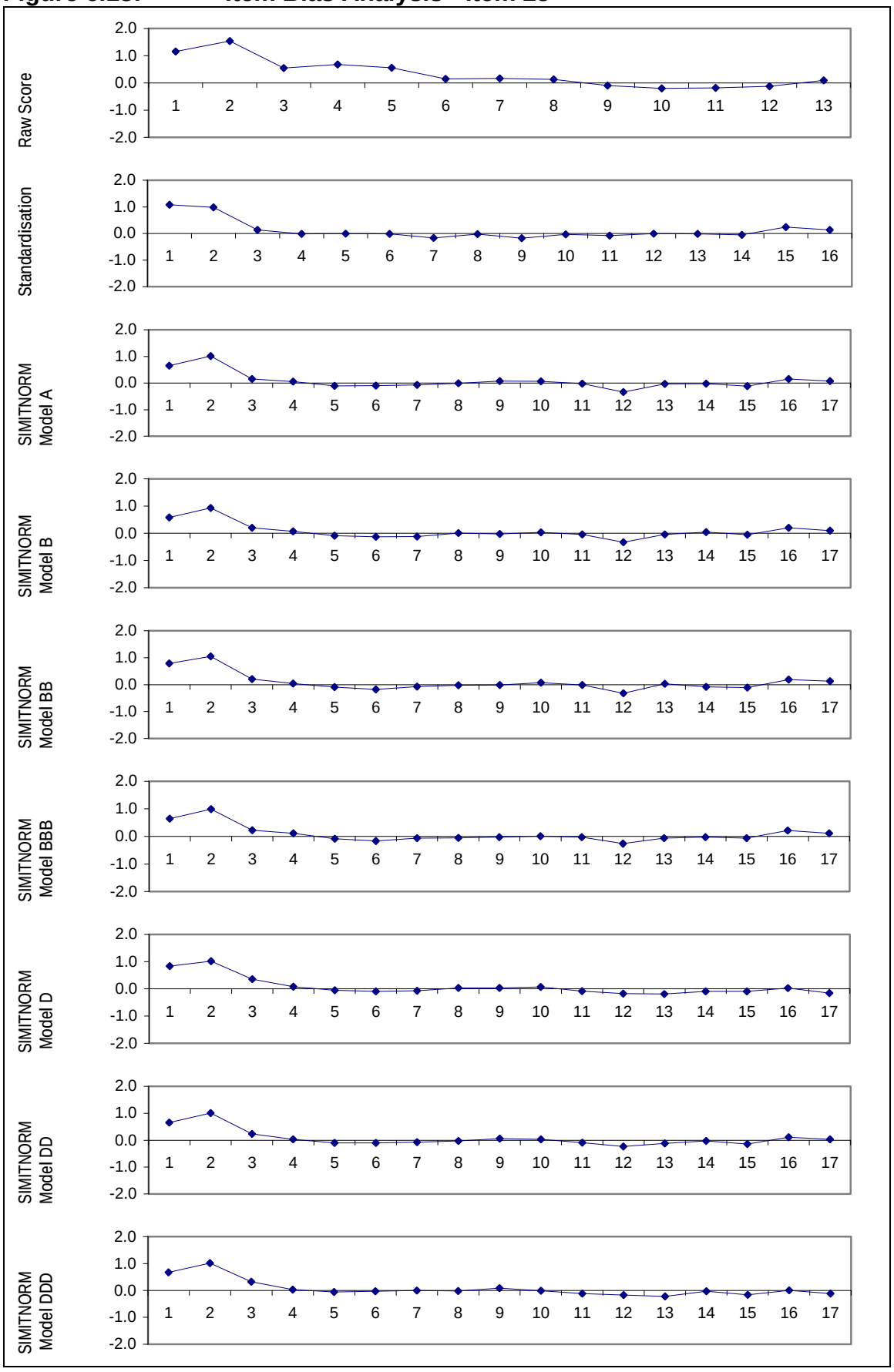


Figure 6.18: Item Bias Analysis - Item 25



Note that items 8, 9, 10 (*Confusion*), items 24 and 25 (*Familiarity*) and items 27, 28 and 29 (*Alienation*) were reverse coded for the analyses, and were plotted in their reversed form. For reverse coded items, a point above the zero line would indicate that at that score level, blacks had a higher reversed mean item level than whites. An alternative equivalent interpretation is that a point above the zero line would indicate that whites had a higher average item value than blacks at that score level for the non-reversed item. For example, consider item 25, which is shown in Figure 6.18. It reads: “I have seen this commercial so many times – I am tired of it.” At the lower score levels (for the entire scale), whites had a higher item mean than blacks, and at the higher score levels there was not much difference between the mean item values.

Item 25 was one of the items for which the ANOVA results indicated that Standardisation and SIMITNORM were not successful in removing non-uniform bias. Similar examples for items 26, 27, 28 and 29 are presented in Appendix I. If these are considered, it seems that most of the problems that arise from “removing” bias, are rooted in the higher and lower ends of the scale. In the middle *Score Levels* of most of the plots in Appendix I, the SIMITNORM results indicate a random pattern around the zero line.

Note that the rest of the figures for the standardised *Z-Scores* and *SIMITNORM Scores* reveal that, except for the first two categories, the remaining score levels indicate approximately a random pattern across score levels, except perhaps for items 4, 9, 10 and 27.

In summary, it should be noted that this method of bias detection is quite sensitive to how the score groups are calculated. Another problem that needs to be considered in the interpretation of these results is that large samples have a tendency to show significant results, even if small differences are observed. Therefore, in an attempt to judge the merits and demerits of SIMITNORM and standardisation, one should take several factors into account. These are discussed in Chapter 7, which concludes this research.

CHAPTER 7

DISCUSSION OF FINDINGS

7.1 LAYOUT OF THIS CHAPTER

The purpose of this study was to compare various simultaneous normalisation approaches with standardisation in the analysis of cross-cultural research data. The purpose of the normalisation methods and standardisation was to eliminate bias from the data in order to obtain equivalent measures across cultures. This chapter evaluates the research results and revisits the stated propositions in order to assess their merits. Most of these evaluations are based on an assessment across repeated statistical methods. The propositions related to SIMNORM are evaluated in Section 7.2, and those related to SIMITNORM in Section 7.3. The implications of the research findings are covered in Section 7.4, with suggestions for further research in Section 7.5.

7.2 PROPOSITIONS RELATED TO SIMNORM

The first and second sub-problems both related to the simultaneous normalisation (SIMNORM) of aggregated perception scores. In Section 5.6, SIMNORM, the simultaneous normalisation of composite scores (where several items are combined to obtain a total score) was compared to standardised scores and raw scores using level-orientated analyses, such as regression analysis and analysis of variance. Propositions 1 and 2 briefly stated SIMNORM scores will be better than standardised to equate composite scores and that SIMNORM results will be very similar to standardised scores in level-orientated analyses with subtle differences, but that they will be different from the results when raw observed scores are used.

Proposition 1:	Since SIMNORM affects scale location, scale dispersion and scale symmetry, it is better than standardisation which only affects scale location and scale dispersion, to remove systematic bias which may present itself as uniform and non-uniform item bias in cross-cultural research.
Proposition 2:	SIMNORM is a non-linear transformation method that has the largest effect on scale location and spread, and a smaller effect on scale symmetry. Standardisation is a linear transformation with the largest effect on scale location and spread, and no effect on scale symmetry. When the results of SIMNORM are compared with the results of standardisation, using level-oriented statistical methods such as ANOVA and regression, very similar results are expected, with minor differences due to the adjustments in the symmetry of the distributions when SIMNORM is used.

The results presented in Chapter 5, all indicated that there were almost no differences between the three models of SIMNORM scores and that the differences between the SIMNORM models and standardised scores were not profound. However, when the differences in the results of raw scores were compared to standardised or normalised scores, the results were remarkably different. The results in Figures 5.10, 5.11, 5.12 and 5.13 clearly showed that SIMNORM was better than standardisation in obtaining equivalence across groups, since the resulting shapes of the SIMNORM score distributions were normally distributed. If the descriptive statistics in Table 5.20 were considered, the means were equal to zero for the standardised scores, and for all three SIMNORM models. The standard deviations of the distributions were equal to 1 for the *Standardised Z-Scores*, and for the SIMNORM Model A and Model B scores. The SIMNORM Model D model was more parsimonious, which means that it had fewer parameters. However, this also meant that the model was more restricted. The result was that Model D was less flexible for individual cultural differences, and therefore was not able also to produce a standard deviation of unity within the

given set of restrictions. Therefore, in the results, the SIMNORM Model D did not have a standard deviation of 1.

When the results of the ANOVA analyses are considered, as given in Table 5.22 to 5.37, it is apparent that there were more similarities than differences between the *Standardised Score* results and the SIMNORM model results. It is also clear that the main difference between the *Raw Scores* and the *Standardised* and SIMNORM results, were in the *Culgroup* variable. For all the *Raw Scores*, *Culgroup* has a very significant effect, whilst for the *Standardised Scores* and for all three SIMNORM scores, *Culgroup* has no effect, which is the result of the standardisation/normalisation.

Overall, it can be concluded that SIMNORM produced scores that are more equivalent across cultural groups than when standardised scores are used. The correlations in Table 5.38 between *Raw Scores*, *Standardised Scores* and *Normalised Scores* are all very high, significant and positive. In the regression analyses, the overall results were very similar, all insignificant predictors remained insignificant, and all significant predictors remained significant, as shown in Table 5.41.

It may be concluded that SIMNORM achieved very similar results compared to standardisation, but had the added benefit that the overall shape of the resulting distribution more closely resembles the normal distribution. The SIMNORM scores resulted in more symmetrical distributions, and a more acceptable kurtosis, than when standardisation is used. All these observations and analyses lead to the non-rejection of Propositions 1 and 2.

Overall the results indicate that SIMNORM could be a very useful alternative to standardisation in the search for equivalence of scores across cultures.

7.3 PROPOSITIONS RELATED TO SIMITNORM

The purpose of SIMITNORM was to obtain equivalence of item scores across different cultures. The purpose stated in the sub-problems was to compare the results of simultaneous item normalisation (SIMITNORM) with standardisation, and to examine the effect of SIMITNORM in structure-orientated data analysis techniques such as exploratory factor analysis (EFA), confirmatory factor analysis (CFA) and structural equation modelling (SEM). Five propositions were stated with respect to SIMITNORM and are repeated below.

Proposition 3:	Simultaneous normalisation of item scores (SIMITNORM) will affect item location, item spread and item symmetry. Item standardisation only affects item location and item spread. It is therefore expected that SIMITNORM will offer a better approach than item standardisation to equate scores over cultural groups, by removing systematic bias that may present itself as uniform and non-uniform bias.
Proposition 4:	SIMITNORM will not have a significant effect on the reliability coefficient Cronbach Alpha when compared to the Cronbach Alpha of the raw items or to the Cronbach Alpha for standardised scores.
Proposition 5:	SIMITNORM will be better than standardisation to ameliorate uniform and non-uniform bias in the total scores and sub-scales of instruments where construct bias is not present, and construct equivalence can be assumed. Interpretation of the results subsequent to SIMITNORM should take the specific method used into account.

Proposition 6:	SIMITNORM used in conjunction with EFA will not affect construct equivalence (the factor structure and factor loadings will remain the same).
Proposition 7:	SIMITNORM used in conjunction with CFA will not affect construct equivalence, metric equivalence or scalar equivalence.
Proposition 8:	SIMITNORM with SEM will result in improved model fit, since the assumption of normally distributed items can be made with more confidence.

In order to test the above propositions, it was necessary to ascertain that the key assumption of structural equivalence of the VRP20 be met. The results obtained in Sections 5.3 to 5.5, where the reliability and validity of the VRP32 and the more parsimonious VRP20 were assessed, provided sufficient evidence for the construct validity and reliability for the VRP20. The results presented in Section 6.7 provided sufficient evidence for the configural invariance and metric invariance of the VRP20 across the four cultural groups, implying that the VRP20 was reliable and valid across cultural groups. Therefore, sufficient evidence does exist that the key assumptions necessary to assess Propositions 3 and 8 were met.

Proposition 3 was supported through the results presented in Section 6.7. H2a was rejected for raw scores, indicating the presence of bias in the data. H2a was not rejected for *Z-scores* or SIMITNORM scores, indicating that both standardisation and SIMITNORM were successful in removing bias. The model fit was in each case better for the SIMITNORM scores than for the Z-scores, indicating that SIMITNORM was better than standardisation to remove bias from the data.

Proposition 4 stated that SIMITNORM will not lower the reliability coefficient, the Cronbach Alpha, when compared to the Cronbach Alpha of the raw item scores or the Cronbach Alpha of standardised scores. The results presented in Table 6.13 show that the Cronbach Alpha is exactly equal between *Raw Scores* and *Standardised Scores*, which is to be expected because standardisation is a linear transformation. With the non-linear transformation in SIMITNORM, in most cases,

as can be seen in Table 6.13, the Cronbach Alpha mostly dropped very slightly, but the change seems insignificant. Overall, it can be concluded that SIMITNORM did not lower the Cronbach Alpha.

The results in Section 6.8 suggest that SIMITNORM was as effective as standardisation to remove uniform bias (see Table 6.21). In the presence of non-uniform bias, the effects of SIMITNORM are very similar to standardisation in removing the cultural effect from the data, but it seems that neither method was sufficiently capable of removing non-uniform bias from the data (Table 6.22). SIMITNORM provided very similar and sometimes slightly better results than standardisation. These results therefore provided some support for Proposition 5, but the results are somewhat disappointing.

Proposition 6 stated that SIMITNORM used in conjunction with EFA will not affect construct equivalence (the factor structure will remain intact). An assessment of the results provided in Section 6.5 clearly indicates that this proposition is supported.

The last two propositions concerned the use of confirmatory factor analyses.

When the results in Tables 6.18 to 6.20 in the column headings “H1” and “H2a” are considered for the overall fit measures, it is clear that SIMITNORM, in conjunction with CFA produced very good results. All the fit measures across these hypotheses were within an acceptable range for all the SIMITNORM models. This implies that SIMITNORM was successful in removing the bias, a clear support for Proposition 7. In addition to this, the fit measures improved if they are compared to standardisation. This could be attributed to the fact that SIMITNORM scores are normally distributed, one of the critical assumptions in CFA, which then in turn is a support for Proposition 8.

It can therefore be concluded that SIMITNORM, in conjunction with CFA produced very good results. With all seven SIMITNORM models, similar and better results to standardised scores were obtained. All seven SIMITNORM models fitted similarly

and for the most part better than the *Standardised Z-scores*, when a variety of fit measures were considered.

Overall, the research results indicated that SIMITNORM has promising potential to be useful in the equating of scores across cultural groups and to remove bias from data.

7.4 IMPLICATIONS OF RESEARCH RESULTS

The SIMITNORM model allows parameter restrictions, so that the user can select a model that has a fixed parameter per item, or a fixed parameter per group. For example, it may sometimes be more appropriate to have one fixed parameter across the scale for a specific cultural group, because the entire group has a certain tendency to respond to all items, irrespective of their content, in a similar way. The plausibility of this can be tested by constraining the amiability parameter to be equal across items. Or, in another application, it may be that a particular item has the tendency to elicit certain responses that are unique to the item, and then the amiability parameter of this item can be restricted to be equal across groups. Since the parameters in SIMITNORM are associated with different aspects of the transformation function, which in turn all affect the final shape, through model choice the user can decide which model is most appropriate in a specific application.

In SIMITNORM, it is would be possible to proceed in a similar manner as has been done in the step-by-step CFA approach. Model A could be seen as H1, where all parameters are freely estimated, and therefore should provide a good baseline model for comparison purposes. Subsequently, constraints on the parameters can be imposed in step-wise procedure, allowing for comparison between nested models.

7.5 SUGGESTIONS FOR FURTHER RESEARCH REGARDING SIMITNORM

This research did not investigate the plausibility of applying SIMNORM on SIMITNORM aggregate scores. This may be useful in smoothing out further irregularities.

The function that was used in this research could be replaced by other mathematical functions, or by alternative parameterisations. It was found that the choice of the upper bound, U, used in the analyses, has a significant impact on the outcome of the parameter estimates. It may be necessary to parameterise U, instead of leaving it for the user to specify.

The issue of normalisation and deviations from normal scores had been addressed by several previous researchers. In LISREL, a different normalisation procedure is available to normalise scores. Further research could compare SIMITNORM with the normalisation procedure available in LISREL. Researchers have also addressed the issue of non-normality by suggesting the use of polychoric correlations and using Robust Maximum Likelihood (RML) estimation procedures (Jöreskog, 1994). Further research could also examine the effectiveness of SIMITNORM when compared to these procedures.

The literature review covered a very robust and efficient normalisation approach recently developed by Stacey (2005). Although Stacey (A.G. Stacey, personal communication, 20 February 2007) tested his method across a variety of demographic variables and found promising results, the research has not yet been published. The Stacey (2005) method could easily be applied across cultures to equate scores. It may be very interesting to compare SIMITNORM with the approach suggested by Stacey (2005) as a method to obtain equivalence.

7.6 SUGGESTIONS FOR FURTHER RESEARCH REGARDING EQUIVALENCE

Salzberger, Sinkovics and Schlegelmilch (2000) have compared Classical Test Theory (CTT) approaches, which include the CFA approach, with Latent Trait Theory / Item Response Theory (LTT/IRT) approaches, in order to obtain data equivalence in cross-cultural research. Their research has shown that LTT approaches overcome several of the problems associated with the CTT approach. The advantage of the LTT approach is that person parameters and item parameters are independent, and the Rasch Model, which is often used in LTT, has less stringent assumptions than CFA.

Some of the shortcomings of the LTT approach are that it has not yet been widely adopted, the computing procedures are not widely accessible nor widely understood, and some of the LTT models are built upon very strict assumptions. Salzberger *et al* (2000) suggested that LTT approaches are used in conjunction with CFA methods. They further argued that it is important to consider data equivalence at the scale development phase of a study, and that researchers should be much more aware of data equivalence issues throughout the design of a study.

It would be very interesting to compare the results of this study by following an LTT approach to obtain data equivalence. Recent studies by Meade and Lautenschlager (2004) and Ewing, Salzberger and Sinkovics (2005) have certainly proved that this could indeed be a fruitful avenue. Meade and Lautenschlager (2004) used simulated data with known properties to assess the appropriateness, similarities and differences between CFA methods and LTT approaches. They found that neither approach is flawless, but that the item response theory (IRT), based approach seemed to be better suited for certain types of measurement equivalence problems. The graded response model was used, compared to the Rasch model used by Salzberger *et al* (2000), which is better suited for ordinal items. Meade and Lautenschlager (2004) indicated that much more work needs to

be done to address issues such as the effect of sample size and violations of key assumptions, such as normality on the use of CFA versus IRT methods.

Ewing *et al* (2005) provided a comparison between CFA/CTT (confirmatory factor analysis / classical test theory) and IRT/LTT (item response theory/latent trait theory) approaches. They analysed a subset of the data used in this research, and examined the *Confusion* subscale using the responses to four commercials, each watched by 100 “non-indigenous” (in this research the white Eurocentric group) and 100 indigenous (in this research the black Afrocentric group) of respondents. Ewing *et al* (2005) found that the IRT approach and CFA approach led to similar conclusions, but that the IRT method was more helpful in identifying specific problematic items. Ewing *et al* (2005) quoted a lucid summary from Molenaar (1997:492), which is also very relevant in this research: “Careful modelling and careful measurement bear some resemblance to airbags and safety belts in a car: they are somewhat costly, and one rarely needs them, but when one does, their presence matters an awful lot”.

Another possibility is to further refine SIMITNORM to a method that can be used to not only diagnose equivalence problems, but also as a method to correct for these problems. In order to do so, further refinements are necessary. Table 7.1 gives a summary of how SIMITNORM can be used to supplement the multi-group confirmatory factor analysis approach, and where the method could be further refined.

Table 7.1: Potential of SIMITNORM as an alternative approach in measurement equivalence applications

Classical Test Theory and CFA in an SEM framework	Latent Trait Theory and IRT	SIMITNORM
The modelled relationship between the latent variable and the items are linear in nature Raju *et al.,* (2002).	The modelled relationship between the items and the latent traits are non-linear Raju *et al.,* (2002).	The modelled relationship between the target distribution function and the items are non-linear in nature.
The methodology is thoroughly developed and software and procedures are more available to handle multiple constructs and multiple groups simultaneously in an analysis Raju *et al.,* (2002).	The differential item functioning (DIF) methodology can only be applied on uni-dimensional constructs Raju *et al.,* (2002).	SIMITNORM can handle multiple groups and is therefore suitable to test parameter constraints across groups.
CFA allows the user to test whether item error variances are equal across groups although it is generally regarded as an overly strict requirement in most practical applications Raju *et al.,* (2002).	In the IRT framework, testing for item level error variances has not been fully explored by statisticians, and it is only possible to test for group-based aggregate standard errors Raju *et al.,* (2002).	An investigation is required to establish whether item error variances can be incorporated into the SIMITNORM model.
CFA are more geared towards studying smaller sets of continuous observed variables assumed to be normally distributed (Embretson and Reise, 2000).	IRT methods are more useful to detect item bias when a large pool (more than 40) dichotomous and polytomous items that are uni-dimensional (relate to a single latent variable) (Embretson and Reise, 2000).	SIMITNORM is suitable to use on a small pool of items. Uni-dimensionality has been assumed in this research, however, it will be possible to allow multiple dimensions in future applications.
Hypothesis by hypothesis invariance testing in CFA has been well developed and is widely applied by researchers (Reise *et al.*, 1993; Embretson and Reise, 2000).	In the IRT-framework, investigation of item bias has developed into a branch called differential item functioning (DIF) (Embretson and Reise, 2000).	It is possible to test hierarchical models. This requires further refinement and investigation.

Table 7.1 (cont): Potential of SIMITNORM as an alternative approach in measurement equivalence applications

Classical Test Theory and CFA in an SEM framework	Latent Trait Theory and IRT	SIMITNORM
Can be used to test for two major response bias patterns, namely extreme response style and acquiescence response style (Cheung and Rensvold, 2000).	A particular problem that arises is that DIF is often indicated when the items relating to the latent construct are marginally not uni-dimensional. Very few constructs are so narrowly defined, and DIF may be indicated falsely, since DIF could be a function of the combination of items, and not a result of item bias. (Embretson and Reise, 2000).	Can be used to remove several types of bias from a set of data.
It is required that the researcher use a strategy to specify certain parameters as fixed in order to identify the model, and the usual way is to either set one of the path coefficients equal to one for each latent variable, or set the error variances equal to one (Reise *et al.*, 1993).	As in CFA, the researcher needs to specify a distribution of the latent trait variable θ (often specified as standard normal), in order for the model to be identified. In the multiple group setting, the researcher has to specify an anchor item in which the parameters are constrained to be equal across groups, in order to allow the researcher to compare the values of the other parameters across groups (Reise *et al.*, 1993).	It is not necessary to constrain parameters in order for the model to be identified, if the three-parameter transformation function is used.
Items paths need to be set equal to 1 in order to assign a scale to the latent variable and in order for the CFA model to be identified (Embretson and Reise, 2000).	Items that have no DIF (no bias) need to be identified to provide the "anchor" items that are assumed to be equal across all groups, thereafter the item parameters across all groups are freely estimated.	It is not necessary to constrain parameters in order to assign a scale to the items. In this research, the scale was set to the standard normal distribution.

Table 7.1 (cont): Potential of SIMITNORM as an alternative approach in measurement equivalence applications

Classical Test Theory and CFA in an SEM framework	Latent Trait Theory and IRT	SIMITNORM
Successive increasingly restrictive hypotheses can be tested, by restricting parameters to be equal across groups (Steenkamp and Baumgartner, 1998).	Items are held constant one by one, and if the restriction is found tenable, those items can be viewed as having no DIF, which could be a very tedious process (Embretson and Reise, 2000).	Successive increasingly restrictive hypotheses can be tested, by restricting parameters to be equal across groups or items.
It is possible to identify partial measurement invariance (Byrne, *et al.,* 1989)	It is possible to identify partial measurement invariance using differential item functioning (Reise *et al.*, 1993; Embretson and Reise, 2000; Raju *et al.*, 2002).	It should be possible to allow certain items / groups to have certain parameters fixed and others free. There is a clear meaning attached to the parameters.
Statistical fit indices are available for CFA measures, namely the chi-square test. The procedure allows for chi-square difference tests when the models are nested by restraining parameters to be equal across groups, thereby allowing the testing of increasingly restrictive hypotheses (Reise *et al.*, 1993).	As in CFA, it is possible to use the chi-square difference tests for nested models to assess the plausibility of increasingly restrictive models (Reise *et al.*, 1993).	The statistical fit indices are AIC, R-square and RMSE. It should be possible in future research to extend these fit indices so that nested models can be compared.
Practical fit indices are available over and above the chi-square values, which allow the researcher to assess model fit, supplementing the chi-square measures of fit that may indicate poor fit in the case of large samples, even if the model deviates slightly from the covariance matrix.	Practical fit indices have not yet been developed, and according to Reise *et al.* (1993), it is a matter that should be the focus for future research.	It should be possible in future research to extend these fit indices so that nested models can be compared.
Modification indices are calculated that helps the researcher to identify the source of poor fit in models.	The development of modification indices in IRT is also identified as a potential area for future research (Reise *et al.*, 1993).	It should be possible to calculate modification indices in help researchers identify the source of poor fit in the models.

Table 7.1 (cont): Potential of SIMITNORM as an alternative approach in measurement equivalence applications

Classical Test Theory and CFA in an SEM framework	Latent Trait Theory and IRT	SIMITNORM
The multi-group CFA approach can be refined by using polychoric correlations, in cases where the assumption of normality is violated.	IRT does not require normally distributed items.	SIMITNORM scores can be calculated and then be used as input into a CFA model. It is not possible with currently available software, to use SIMITNORM scores first, and then apply IRT.
CFA models include the estimation of variances and covariances between latent constructs	Only uni-dimensional constructs can be used, a very limiting aspect of IRT.	It should be possible in future research to estimate variances and covariances between latent constructs.

7.7 CONCLUSION

Van Herk, Poortinga and Verhallen (2004) covered several issues concerning equivalence, which in essence summarised and enhanced their view of available literature. They offer more clarity on several confusing matters, and made the following important remark: “Equivalence cannot be assumed: construct, measurement unit, or scalar equivalence has to be established by means of explicit procedures.” (Van Herk *et al*, 2000:361). This research proposed a possible explicit procedure, which may in support of existing procedures, provide deeper insights which were previously not possible. It may be worthwhile to further test and refine SIMNORM and SIMITNORM in a variety of ways as noted in these suggestions for further research, and on more sets of data.

According to Reise *et al* (1993), since the CFA and IRT approaches have been proposed as methods to deal with the invariance problem, frequent application of these methods are sometimes ignored in situations where it could be fruitful. Also, since these methods have been proposed, the topic of invariance and methods dealing with invariance has been largely ignored in the psychometric literature, and new approaches have not been forwarded recently. SIMITNORM, if further refined, could be a very promising alternative and supplementary approach for dealing with equivalence problems in the analysis of data where bias is a problem.

REFERENCES

Aaker, D.A. & Bruzzone, D.E. (1985): Causes of Irritation in Advertising, *Journal of Marketing*, 49(2), Spring, 47-57.

affable. (n.d.): Dictionary.com. *Dictionary.com Unabridged (v 1.1).* Random House Inc.
http://dictionary.reference.com/browse/affable (accessed: April 18, 2008).

agreeability. (n.d.): Dictionary.com. *Dictionary.com Unabridged (v 1.1).* Random House Inc.
http://dictionary.reference.com/browse/agreeability (accessed: April 18, 2008).

amiability. (n.d.): Dictionary.com. *Dictionary.com Unabridged (v 1.1).* Random House Inc.
http://dictionary.reference.com/browse/amiability (accessed: April 18, 2008).

amicable. (n.d.): Dictionary.com. *Dictionary.com Unabridged (v 1.1).* Random House Inc.
http://dictionary.reference.com/browse/amicable (accessed: April 18, 2008).

Anonymous (2008): Research Bias
http://tigger.uic.edu/~lwbenn/jacswcourses/socw360/week14.htm
(accessed March 14, 2008).

Akaike, H. (1987): Factor analysis and AIC, *Psychometrika,* 52, 317:332.

Angoff, W.H. (1971): Scales, norms and equivalent scores, in *Educational Measurement,* R.L. Thorndika (ed.), second edition, Washington, D.C.: American Council on Education.

Angoff, W.H. & Robertson G.J. (1987): A procedure for standardizing individually administered tests, normed by age or grade level, *Applied Psychological Measurement*, 1(1), March, 33-46.

Appiah, O. Ethnic Identification of Adolescents' Evaluations of Advertisements, *Journal of Advertising Research,* September-October, 7-22.

Babbie, E. (1992): *The practice of social research,* sixth edition, Belmont, CA: Wadsworth Publishing.

Baumgartner. H. & Steenkamp, J.E.M. (2001): Response styles in Marketing Research: A Cross-National Investigation, *Journal of Marketing Research*, 38(2) May 2001, 143-156.

Baumgartner. H. & Steenkamp, J.E.M. (2006): An Extended Paradigm for Measurement Analysis of Marketing Constructs Applicable to Panel Data, *Journal of Marketing Research,* XLIII (August), 431-442.

Bearden, W.O., Netemeyer, R.G. & Mobley, M.F. (1993): *Handbook of Marketing Scales: Multi-Item Measures for Marketing and Consumer Behaviour Research,* London: Sage.

Bendixen, M.T. & Sandler, M. (1995): Converting Verbal Scales to Interval Scales Using Correspondence Analysis, *Management Dynamics: Contemporary Research*, 4(1), 32-50.

Bensaou, M., Coyne, M. & Venkatraman N. (1999): Testing Metric Equivalence in Cross-National Strategy Research: An Empirical Test across the United States and Japan, *Strategic Management Journal,* 20(7), 671-689.

Bentler, P.M. (1990): Comparative Fit Indices in Structural Equation Models, *Psychological Bulletin,* 107, 238-246.

Bentler, P.M. & Bonnet, D. (1980): Significance Tests and Goodness of Fit in the Analysis of Covariance Structures, *Psychological Bulletin,* 88, 588-606.

Bernardi, R.A. (2006): Associations between Hofstede's Cultural Constructs and Social Desirability Response Bias, *Journal of Business Ethics,* 65: 43-53.

Biel, A.L. & Bridgewater, C.A. (2001): Attributes of Likable Commercials, *Journal of Advertising Research,* 30(3), 38-44.

Bollen, K.A. (1989): *Structural Equations with Latent Variables,* New York: John Wiley & Sons.

Box, M.J., Davies, D. & Swann, W.H. (1969): *Non-linear Optimization Techniques,* Edinburgh: Oliver and Boyd Ltd.

Browne, M.W. (1984): Asymptotically distribution-free methods for the analysis of covariance structures, *British Journal of Mathematical and Statistical Psychology,* 37, 62-83.

Browne, M.W. & Cudeck, R. (1993): Alternative ways of assessing model fit. In: K.A. Bollen & J.S. Long (Eds), *Testing Structural Equation Models* (136-162)*,* Newbury Park, CA: Sage.

Bullinger, M., Anderson, R., Cella, D. and Aaronson, N. (1993): Developing and evaluating cross-cultural instruments from minimum requirements to optimal models, *Quality of Life Research,* 2, 451:459.

Byrne, B.M. (1998): *Structural Equation Modeling with LISREL, PRELIS and SIMPLIS: Basic Concepts, Applications and Programming*, London: Lawrence Erlbaum Associates, Publishers.

Byrne, B.M. & Campbell T.L. (1999): Cross-Cultural Comparisons and the Presumption of Equivalent Measurement and Theoretical Structure: A Look beneath the surface, *Journal of Cross-Cultural Psychology,* 30(5), 555-574.

Byrne, B.M. & Shavelson, R.J. (1987). Adolescent Self-Concept: Testing the Assumption of Equivalent Structure Across Gender, *American Educational Research Journal,* 24(3), 365-385.

Byrne, B.M., Shavelson, R.J. and Muthén, B. (1989). Testing for the Equivalence of Factor Covariance and Mean Structures: The Issue of Partial Measurement Invariance, *Psychological Bulletin,* 105(May), 456-466.

Campbell, D.T. & Fiske, D.W. (1959): Convergent and discriminant validation by the multi-trait-multimethod matrix, *Psychological Bulletin, 56*, 81-105.

Carman, J.M. (1990): Consumer perceptions of service quality: an assessment of the SERVQUAL dimensions, *Journal of Retailing*, 66(Spring), 33-56.

Cheung G.W. & Rensvold R.B. (2000): Assessing Extreme and Acquiescence Response Sets in Cross-Cultural Reseach Using Structural Equation Modelling, *Journal of Cross-Cultural Research Psychology,* 31(2), 187-212.

Churchill, G.A. (1979): A Paradigm for Developing Better Measures of Marketing Constructs, *Journal of Marketing Research*, 16(Feb), 64-73.

Clauser, B.E., Mazor, K.M. & Hambleton, R.K. (1994): The effects of score group width on the Mantel-Haensel Procedure, *Journal of Educational Measurement,* 31, 67-78.

Cleary, T.A. & Hilton, T.L. (1968): An investigation of item bias, *Educational and Psychological Measurement,* 28, 61-75.

Cole, M.S., Bedeian, A.G. and Feild, H.S. (2006): The Measurement Equivalence of Web-Based and Paper-and-Pencil Measures of Transformational Leadership, *Organisational Research Methods,* 9(3), 339-368.

Cook, W.A. (1992): Love, Hate and Likability, *Journal of Advertising Research,* 32(2), 7-9.

Cooper, D.R. & Schindler, P.S. (1998): *Business Research Methods*, sixth edition, Boston: Irwin/McGraw-Hill.

Cronbach, L.J. & Meehl, P.E. (1955): Construct validity in psychological tests, *Psychological Bulletin,* 52*,* 317-327.

Cudeck, R. (1989): Analysis of Correlation Matrices Using Covariance Structure Models, *Psychological Bulletin,* 105 (March), 317-327.

Day, G.S. & Montgomery, D.B. (1999): Charting New Directions for Marketing, *Journal of Marketing,* 63(Special Issue)*,* 3-13.

Draper N.R. & Smith, H. (1981): *Applied Regression Analysis, second edition,* New York: John Wiley & Sons.

Du Plessis, E. (1994): Understanding and using likability, *Journal of Advertising Research,* 34(5), RC3-RC10.

Du Toit, M. (Ed.) (2003): *IRT from SSI: BILOG-MG MULTILOG PARSCALE TESTFACT,* Chicago: Scientific Software International.

Du Toit, S.H.C. (1979): The analysis of growth curves*.* Unpublished Ph.D. Thesis, Pretoria: Unisa.

Du Toit, M. & Du Toit, S.H.C. (2001): *Interactive LISREL, User's Guide,* Chicago: Scientific Software International.

Du Toit, S.H.C., Steyn, A.G.W. & Stumpf R.H. (1986): *Graphical exploratory data analysis,* New York: Springer Verlag.

Elphinstone, C.D. (1983): A target distribution model for nonparametric density estimation, *Communications in Statistics, A:Theory and Methods,* 161-198.

Elphinstone, C.D. (1985): *A method of distribution and density estimation,* Pretoria: CSIR Technical Report, TWISK 421.

Embretson, S.E. & Reise, S.P. (2000): *Item Response Theory for Psychologists,* Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Ewing, M.T. Salzberger, T. & Sinkovics, R.R. (2005): An alternative approach to assessing cross-cultural measurement equivalence in advertising research, *Journal of Advertising,* 34(1), 17-36.

Fabrigar, L.R., Wegener, D.T., MacCallum, R.C. & Strahan, E.J. (1999): Evaluating the Use of Exploratory Factor Analysis in Psychological Research, *Psychological Methods,* 4(3), 272-299.

Fan, X. (1998): Item Response Theory and Classical Test Theory: An empirical comparison of their item/person statistics, *Educational and Psychological Measurement,* 58(3), 457-381.

Fischer, R. (2004): Standardisation to Account for Cross-Cultural Response Bias: A Classification of Score Adjustment Procedures and Review of Research in *JCCP*, *Journal of Cross-Cultural Psychology,* 35(3), May, 263-282.

Fornell, C. & Larcker D.F. (1981): Evaluating Structural Equation Models with Unobservable Variables and Measurement Error, *Journal of Marketing Research*, 18 (February), 39–50.

Hair, J.F. Jr., Anderson, R.E., Tatham, R.L. & Black W.C. (1998): *Multivariate Data Analysis,* fifth edition, London: Prentice-Hall International, Inc.

Hicks, L.E. (1970): Some properties of ipsative, normative and forced-choice normative measures. *Psychological Bulletin*, 74, 167-184.

Hofstede G. (1980): *Culture's consequences: International differences in work-related values,* Beverly Hills, CA: Sage.

Horn, J.L. & McArdle, J.J. (1992): A Practical and Theoretical Guide to Measurement Invariance in Aging Research, *Experimental Aging Research,* 18 (Fall-Winter), 117-144.

Horn, J.L., McArdle, J.J. & Mason, R. (1983): When is Invariance not Invariant: A Practical Scientist's Look at the Ethereal Concept of Factor Invariance, *The Southern Psychologist,* 1 (Summer-Fall), 179-188.

Hu, L.-T. & Bentler, P.M. (1999): Cut-off Criteria for Fit Indices in Covariance Structure Analysis: Conventional Criteria versus New Alternatives, *Structure Equation Modeling,* 6(1), 1-55.

Hui, C.H. & Triandis, H.C. (1985): Measurement in Cross-Cultural Psychology, *Journal of Cross-Cultural Psychology,* 16(2), 131-152.

Hui, C.H. & Triandis, H.C. (1989): Effects of Culture and Response Format on Extreme Response Style, *Journal of Cross-Cultural Psychology,* 20(3), 296-309.

Johnson, N.L. & Kotz, S. (1970): *Distributions in Statistics: Continuous Univariate Distributions-1.* New York: John Wiley & Sons.

Jones, P.B. & Sabers, D. (1984): *An Investigation of Two Procedures for Smoothing Test Norms.* Paper presented to the 1984 Annual Meeting of the American Educational Research Association. (58th, New Orleans, Louisiana, April 23-27).

Jöreskog, K. (1994): On the estimation of polychoric correlations and their asymptotic covariance Matrix, *Psychometrika,* 59, 381-389.

Jöreskog, K. & Sörbom, D. (1993): *LISREL 8: Structural Equation Modelling with the SIMPLIS Command Language,* Chicago: Scientific Software International.

Jöreskog, K. & Sörbom, D. (1996a): *LISREL 8: User's Reference Guide,* Chicago: Scientific Software International.

Jöreskog, K. & Sörbom, D. (1996b): *PRELIS 2: User's Reference Guide,* Chicago: Scientific Software International.

Juster, F.T. (1964): *Anticipants and Purchases: An Analysis of Consumer Behavior.* Princeton, NJ: National Bureau of Economic Research.

Kalleberg, A.L. & Kluegel, J.R. (1975): Analysis of the Multitrait-Multimethod Matrix: Some Limitations and An Alternative, *Journal of Applied Psychology,* 60(Feb), 1:9.

Kendall, M. & Stuart, A. (1977): *The Advanced Theory Of Statistics Vol.1: Distribution Theory,* fourth edition, London: Charles Griffin & Company Limited.

King, G., Murray, C.J.L., Salomon, J.A. and Tandon, A. (2004): Enhancing the Validity and Cross-Cultural Comparability of Measurement in Survey Research, *American Political Science Review*, 98(1), 191-207).

Kolen, M.J. (1981): Comparison of Traditional and Item Response Theory Methods for Equating Tests, *Journal of Educational Measurement,* 18(1) 1-11.

Kolen, M.J. (1983): Effectiveness of Analytic Smoothing in Equipercentile Equating, *ACT Technical Bulletin,* No. 41. American College Testing Program, Iowa City, Iowa.

Krugman, H.E. (1967): The Measurement of Advertising Involvement, *Public Opinion Quarterly,* 30, 583-596.

Lastovicka, J.L. (1983): Convergent and Discriminant Validity of Television Commercial Rating Scales, *Journal of Advertising*, 12(2), 14-23.

Lawley, D.M. & Maxwell, A.E. (1971): *Factor Analysis as a Statistical Method,* London: Butterworth & Co (Publishers) Ltd.

Leavitt, C. (1970): A Multi-Dimensional Set of Rating Scales for Television Commercials, *Journal of Applied Psychology,* 54, 427-429.

Leedy, P.D. & Ormrod, J.E. (2001): *Practical Research: Planning and Design*, seventh edition, Upper Saddle River: Merrill Prentice Hall.

Leung, K. & Bond, M.H. (1989): On the Empirical Identification of Dimensions for Cross-Cultural Comparison. *Journal of Cross-Cultural Psychology,* 20, 133-151.

Lindsay, C.A. & Prichard, M.A. (1971): An Analytic Procedure for the Equipercentile Method of Equating Tests, *Journal of Educational Measurement,* 8, 203-207.

Little, T.D. (1997): Means and Covariance Structures (MACS) Analyses of Cross-Cultural Data: Practical and Theoretical Issues, *Multivariate Behavioral Research,* 32(1), 53-76.

Little, T.D. (2000): On the Comparability of Constructs in Cross-Cultural Research: A critique of Cheung and Rensvold, *Journal of Cross-Cultural Psychology*, 31(2), 213-219.

Lord, F.M. (1977): Practical Application of Item Characteristic Theory*, Journal of Educational Measurement,* 14, 117-138.

Mackenzie, S.B. (2001): Opportunities for Improving Consumer Research through Latent Variable Structural Equation Modeling, *Journal of Consumer Research,* 28(1), 159-166.

Marco, G.L. (1977): Use of the logistic model as an alternative to linear interpolation for computing percentile ranks, *Journal of Educational Measurement*, 14, 271-275.

Marsh, H.W. and Hovecar, D. (1985): Application of Confirmatory Factor Analysis to the Study of Self-Concept: First and Higher Order Factor Models and Their Invariance Across Groups, *Psychological Bulletin,* 97(3), 562-582.

McGaw B. & Jöreskog, K.G. (1971): Factorial Invariance of Ability Measures in Groups Differing in Intelligence and Socioeconomic Status, *British Journal of Mathematical and Statistical Psychology,* 24, 154-168.

Meade, A.W. & Lautenschlager, G.J. (2004): A Comparison of Item Response Theory and Confirmatory Factor Analytic Methodologies for Establishing Measurement Equivalence/Invariance, *Organizational Research Methods,* 7(4), October, 361-388.

Moldovan, S.E. (1985): Copy Factors Related to Persuasion Scores, *Journal of Advertising Research,* 24(6), 16-22.

Molenaar, I.W. (1997): Lenient or Strict Application of IRT with an Eye on the Practical Consequences, in *Applications of Latent Trait and Latent Class Models in the Social Sciences,* J. Rost & T. Langenheine R. (eds.), Münster: Waxmann, 38-49.

Myers, M.B., Calantone, R.J., Page Jr., T.J. & Taylor, C.R. (2000): Academic Insights: An Application of Multiple-Group Causal Models in Assessing Cross-Cultural Measurement Equivalence, *Journal of International Marketing*, 8(4), 108-121.

Nunnally, J. C. & Bernstein, I. H. (1994): *Psychometric Theory*, third edition, New York: McGraw-Hill Series in Psychology.

Peterson, R.A. (1994): A Meta-Analysis of Cronbach's Coefficient Alpha, *Journal of Consumer Research,* 21(September), 381-391.

Petty, R.E., Cacioppo, J.T. & Schumann, D. (1983): Central and Peripheral Routes to Advertising Effectiveness: The Moderation Role of Involvement, *Journal of Consumer Research,* 10, (September), 135-147.

Poortinga, Y.H. (1989): Equivalence of Cross-Cultural Data: An Overview of Basic Issues, *International Journal of Psychology,* 24(1), 737-756.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P. (1996): *Numerical recipes in Fortran90,* Cambridge: Cambridge University Press.

Raftery, A. (1995): Bayesian model selection in social research. *Sociological Methodology*, 25: 111-163.

Raju, N.S., Laffitte, L.J. & Byrne, B.M. (2002): Measurement Equivalence: A Comparison of Methods Based on Confirmatory Factor Analysis and Item Response Theory, *Journal of Applied Psychology* 87(3), 517-529.

Raykov, T. & Marcoulides, G.A. (2000): *A First Course in Structural Equation Modeling,* Mahwah, New Jersey: Lawrence Erlbaum Associates, Publishers.

Reise, S.P., Widaman, K.F. & Pugh, R.H. (1993): Confirmatory Factor Analysis and Item Response Theory Approaches for Exploring Measurement Invariance, *Psychological Bulletin,* 114(3), 552-566.

Rensvold, R.B. & Cheung, G.W. (1998): Testing Measurement Models for Factorial Invariance: A Systematic Approach, *Educational and Psychological Measurement,* 58(6), 1017-1034.

Richards, F.J. (1959): A Flexible Growth Function for Empirical Use, *Journal of Experimental Botany,* 10(29), 290-300.

Ross, G.J.S. (1970): The Efficient Use of Function Minimisation in Non-Linear Maximum-Likelihood Estimation, *Applied Statistics*, Vol.19, pp205-221.

Salzberger, T., Sinkovics, R. & Schlegelmilch, B. (2000): Data Equivalence in Cross-Cultural Research: A comparison of Classical Test Theory and Latent Trait Theory Based Approaches, *Australasian Marketing Journal*, 7(2), 23-38.

Schlinger, M.J. (1979): A Profile of Responses to Commercials, *Journal of Advertising Research*, 19(2), (April), 37-46.

Schlinger, M.J. (1982): Respondent Characteristics that Affect Copy-Test Attitude Scales, *Journal of Advertising Research,* 22(1), 29-35.

Scientific Software International (n.d.): http://www.ssicentral.com (accessed: April 10, 2008).

Schwarz, G. (1987): Estimating the dimension of a model. *Annals of Statistics,* 6, 461-464.

Searle, S.R. (1971): *Linear Models,* New York: John Wiley & Sons.

Sin, L.Y.M., Cheung, G.W.H. and Lee, R. (1999): Methodology in Cross-Cultural Research: A Review and Critical Assessment, *Journal of International Consumer Marketing,* 11(4), 75-96.

Smith, A.M. & Reynolds, N.L. (2001): Measuring cross-cultural service quality, *International Marketing Review,* 19(5), 450-481.

Smith, P.B. (2004): Acquiescent and Response Bias as an Aspect of Cultural Communication Style, *Journal of Cross-Cultural Psychology*, 35, 50-61.

Sörbom, D. (1974): A General Method for Studying Differences in Factor Means and Factor Structure between Groups, *British Journal of Mathematical and Statistical Psychology,* 27, 229-239.

Stacey, A.G. (2005): The Reliability and Validity of the Item Means and Standard Deviations of Ordinal Level Response Data, *Management Dynamics,* 14(3), 2-25.

Steenkamp, J.E.M. & Baumgartner, H.B. (1998): Assessing Measurement Invariance in Cross-National Consumer Research, *Journal of Consumer Research,* 25(June), 78-90.

Steenkamp, J.E.M. & Van Trijp, H.C.M. (1991): The use of LISREL in Validating Marketing Constructs, *International Journal of Research in Marketing,* 8, 283-299.

Strasheim, C. (1989): Simultaneous Normalisation of Test Scores over Various Age Levels, Unpubished MSc Thesis, Pretoria: University of South Africa.

Strasheim, C., Pitt, L.F. & Caruana, A. (2007): Psychometric Properties of the Schlinger Viewer Response Profile (VRP): Evidence from a Large Sample, *Journal of Advertising,* 36(4), 101-114.

Trochim, W.M.K. (2006): *Social Research Methods,* Web Centre for Social Research Methods, http://www.socialresearchmethods.net/kb/index.php (accessed: February 18, 2008).

Van de Vijver, F.J.R. & Leung, K. (1997): *Methods and Data Analysis for Cross-Cultural Research*, London: Sage.

Van de Vijver, F.J.R. & Leung, K. (2000): Methodological Issues in Psychological Research on Culture, *Journal of Cross-Cultural Psychology,* 31(1), 33-51.

Van de Vijver, F. & Poortinga, Y.H. (1994): Methodological Issues in Cross-Cultural Studies on Parental Rearing Behavior and Psychopathology. In C. Perris, W.A. Arrindell & M. Eisenman (Eds.), *Parental Rearing and Psychopathology* (173-197). Chicester, UK: John Wiley.

Van Hemert, D.A., Van de Vijver, F.J.R., Poortinga, Y.H. and Georgas, J. (2002): Structural and functional equivalence of the Eysenck Personality Questionnaire within and between countries. *Personality and Individual,* 33, 1229-1249.

Van Herk, H., Poortinga, Y.H. & Verhallen, M.M. (2005): Equivalence of Survey Data: Relevance for International Marketing, *European Journal of Marketing,* 39(3/4), 351-364.

Vandenberg, R.J. (2002). Toward a Further Understanding of and Improvement in Measurement Invariance Methods and Procedures, *Organizational Research Methods,* 5(2), 139-158.

Vandenberg, R.J. & Lance, C.E. (2000): A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices and Recommendations for Organisational Research, *Organisational Research Methods,* 3(1), 4-69.

Walker, D. & Dubitsky, T.M. (1994): Why Liking Matters, *Journal of Advertising Research,* (May/June), 9-17.

Wells, W.D. (1964): Recognition, Recall and Rating Scales, *Journal of Advertising Research,* 4, 2-8.

Wells W.D., Leavitt C. & McConville, M. (1971): A Reaction Profile for TV Commercials, *Journal of Advertising Research*, 11(December), 11-17.

Zikmund, W.G. (1997): *Business Research Methods,* fifth edition, New York: The Dryden Press.

Zinkhan, G.M. & Burton, S. (1989): An Examination of Three Multidimensional Profiles for Assessing Consumer Reactions to Advertisements, *Journal of Advertising,* 18(4), 6-14.

Zinkhan, G.M. & Fornell, C. (1985): A Test of Two Consumer Response Scales in Advertising, *Journal of Marketing Research,* 22(November), 447-452.

APPENDIX A: The COMMAP Variables

Nr	Questionnaire Number
Likability	Give the commercial points out of 10 in terms of how much you like it.

V1	The commercial was lots of fun to watch and listen to.
V2	I thought it was clever and quite entertaining
V3	The enthusiasm of the commercial is catching – it picks you up
V4	The commercial wasn't just selling ….. it was also entertaining me. I appreciated that.
V5	The characters (or persons) in the commercial capture your attention.
V6	It is the kind of commercial that keeps running through your mind after you have seen it.
V7	I just laughed at it – I thought it was very funny and good.
V8	It was distracting trying to watch the screen and listen to the words at the same time.
V9	It required a lot of effort to follow the commercial.
V10	It was too complex. I was not sure what was going on.
V11	I was so busy watching the screen, I didn't listen to the words.
V12	The commercial gave me a new idea.
V13	The commercial reminded me that I'm dissatisfied with the ……. I have now and Im looking for something better.
V14	I learned something from the commercial that I did not know before.
V15	The commercial told me about the benefits of ….. and I think I would like to try it.
V16	During the commercial I thought how …. Might be useful to me.
V17	…… is a good financial institution and I wouldn't hesitate recommending it to others
V18	I know that the advertised …. is dependable and reliable.
V19	The commercial was very realistic – that is true to life.
V20	I felt that the commercial was acting out what I feel at times.
V21	I felt as though I were right there in the commercial experiencing the same old thing.
V22	That's my idea – the kind of life the commercial showed.
V23	I liked the commercial because it was personal and intimate.
V24	This kind of commercial has been done many times ……. it is the same old thing.
V25	I have seen this commercial so many times – I am tired of it.
V26	I think that this is an unusual commercial, I am not sure I've seen another one like it.
V27	What they showed did not demonstrate the claims they were making about …..
V28	The ad didn't have anything to do with me or my needs.
V29	The commercial did not show me anything that would make me want to have a ……..
V30	The commercial made exaggerated claims. …… would not live up to what they said or implied.
V31	It was an unrealistic ad – very far-fetched.
V32	That commercial irritated me – it was annoying

V33	The commercial makes me consider trying a ……..
V34	The commercial makes me feel that ……. cares about me.
V35	The commercial is well suited to the image I have of ……..
V36	The commercial is convincing.
V37	The commercial makes me feel that ……. treats their clients as stupid people.
V38	I generally watch and enjoy television commercials.
V39	This type of commercial is appealing to me.
V40	The commercial makes me think that …… treats their clients as intelligent people.
V41	I do not like television commercials in general.
V42	This is a good …… and I wouldnt hesitate recommending it to others.
V43	I know that the advertised account is dependable and reliable.

Biographic Variables

V44 Age
V45 Population Group
V46 Gender
V47 Income
V48 Language
V49 Ad Done
V50 Type of ad
V51 Area
V52 Type
V53 Area
V54 Date
V55 Ad Seen
V56 Usage
V57 Classification
V58 Job Number

Variable Labels

Value labels Likablity 1 Do not like at all 10 Like very much

V1 to V43
1 Strongly disagree
2 Disagree
3 Neither agree nor disagree
4 Agree
5 Strongly agree

V44
1 Less than 16 Years
2 16 - 24 Years
3 25 - 34 Years
4 35 - 49 Years
5 50+ Years
6 Not applicable

V45
1 White 2 Black

V46
1 Male 2 Female

V47
1 A 2 B 3 C 4 D 5 Student 6 Not applicable

V48
1 English / Other
2 Afrikaans / Both
3 Nguni
4 Sotho
5 Not applicable

V49
1 Pre-Test ad 2 Post-Test ad

V50
4 TV

V51
1 Gauteng 2 Kwazulu-Natal 3 Free State 4 Cape Town

V52
1 Public

V53
1 Urban 2 Rural

V54

1 Aug-89	31 Feb-92	61 Aug-94	91 Feb-97	121 Aug-99
2 Sep-89	32 Mar-92	62 Sep-94	92 Mar-97	122 Sep-99
3 Oct-89	33 Apr-92	63 Oct-94	93 Apr-97	123 Oct-99
4 Nov-89	34 May-92	64 Nov-94	94 May-97	124 Nov-99
5 Dec-89	35 Jun-92	65 Dec-94	95 Jun-97	125 Dec-99
6 Jan-90	36 Jul-92	66 Jan-95	96 Jul-97	126 Jan-00
7 Feb-90	37 Aug-92	67 Feb-95	97 Aug-97	127 Feb-00
8 Mar-90	38 Sep-92	68 Mar-95	98 Sep-97	128 Mar-00
9 Apr-90	39 Oct-92	69 Apr-95	99 Oct-97	129 Apr-00
10 May-90	40 Nov-92	70 May-95	100 Nov-97	130 May-00
11 Jun-90	41 Dec-92	71 Jun-95	101 Dec-97	131 Jun-00
12 Jul-90	42 Jan-93	72 Jul-95	102 Jan-98	132 Jul-00
13 Aug-90	43 Feb-93	73 Aug-95	103 Feb-98	133 Aug-00
14 Sep-90	44 Mar-93	74 Sep-95	104 Mar-98	134 Sep-00
15 Oct-90	45 Apr-93	75 Oct-95	105 Apr-98	135 Oct-00
16 Nov-90	46 May-93	76 Nov-95	106 May-98	136 Nov-00
17 Dec-90	47 Jun-93	77 Dec-95	107 Jun-98	137 Dec-00
18 Jan-91	48 Jul-93	78 Jan-96	108 Jul-98	138 Jan-01
19 Feb-91	49 Aug-93	79 Feb-96	109 Aug-98	139 Feb-01
20 Mar-91	50 Sep-93	80 Mar-96	110 Sep-98	140 Mar-01
21 Apr-91	51 Oct-93	81 Apr-96	111 Oct-98	141 Apr-01
22 May-91	52 Nov-93	82 May-96	112 Nov-98	142 May-01
23 Jun-91	53 Dec-93	83 Jun-96	113 Dec-98	143 Jun-01
24 Jul-91	54 Jan-94	84 Jul-96	114 Jan-99	144 Jul-01
25 Aug-91	55 Feb-94	85 Aug-96	115 Feb-99	145 Aug-01
26 Sep-91	56 Mar-94	86 Sep-96	116 Mar-99	146 Sep-01
27 Oct-91	57 Apr-94	87 Oct-96	117 Apr-99	147 Oct-01
28 Nov-91	58 May-94	88 Nov-96	118 May-99	148 Nov-01
29 Dec-91	59 Jun-94	89 Dec-96	119 Jun-99	149 Dec-01
30 Jan-92	60 Jul-94	90 Jan-97	120 Jul-99	

V59
1 Yes 2 No

V60
1 User 2 Other 3 Non Users 4 Not applicable

V62
1 Face to Face 2 Theatre

APPENDIX B: Estimated Parameters of SIMNORM

This appendix provides the results from the SIMNORM program. The total scores that were subjected to normalisation are given in Table B.1. Due to space limitations, the SIMNORM results of the subscales V20_1, R20_2, V20_3, V20_4, V20_5, R20_6 and R20_7 are not given.

Table B.1: Descriptions of all total scores used for the SIMNORM results

Variable	Description	Calculation
LIKE	Likeability	A score between 1 and 10, with 1, does not like the commercial, and 10, like it very much.
V20_1	Entertainment	Sum of items VV1, VV2, VV3 and VV4
R20_2	Confusion	Sum of items VV8, VV9 and VV10, where VV8, VV9 and VV10 were reverse scored
V20_3	Relevant News	Sum of items VV14, VV15 and VV16
V20_4	Brand Reinforcement	Sum of items VV17 and VV18
V20_5	Empathy	Sum of items VV20, VV21 and VV22
R20_6	Familiarity	Sum of reversed coded items VV24 and VV25
R20_7	Alienation	Sum of reversed coded items VV27, VV28 and VV29.
V20_8	Positive Moods	Sum of items VV1- VV4, VV14-VV16 VV17-VV18, and VV20-VV22
R20_9	Negative Moods	Sum of reversed coded items VV8-VV10, VV24-VV25 and VV27-VV29
V20_10	Overall VRP20	Sum of all positive items and all reversed negative items (V20_8 + R20_9)
LIKEZ	Likeability	LIKE standardised over cultural groups
V20_1Z	Entertainment	V20_1 standardised over cultural groups
R20_2Z	Confusion	R20_2 standardised over cultural groups
V20_3Z	Relevant News	V20_3 standardised over cultural groups
V20_4Z	Brand Reinforcement	V20_4 standardised over cultural groups
V20_5Z	Empathy	V20_5 standardised over cultural groups
R20_6Z	Familiarity	R20_6 standardised over cultural groups
R20_7Z	Alienation	R20_7 standardised over cultural groups
V20_8Z	Positive Moods	V20_8 standardised over cultural groups
R20_9Z	Negative Moods	R20_9 standardised over cultural groups
V20_10Z	Overall VRP20	V20_10 normalised over cultural groups using Model A

Table B.1 (continued):

LIKEA	Likeability	LIKE normalised over cultural groups using Model A
V20_1A	Entertainment	V20_1 normalised over cultural groups using Model A
R20_2A	Confusion	R20_2 normalised over cultural groups using Model A
V20_3A	Relevant News	V20_3 normalised over cultural groups using Model A
V20_4A	Brand Reinforcement	V20_4 normalised over cultural groups using Model A
V20_5A	Empathy	V20_5 normalised over cultural groups using Model A
R20_6A	Familiarity	R20_6 normalised over cultural groups using Model A
R20_7A	Alienation	R20_7 normalised over cultural groups using Model A
V20_8A	Positive Moods	V20_8 normalised over cultural groups using Model A
R20_9A	Negative Moods	R20_9 normalised over cultural groups using Model A
V20_10A	Overall VRP20	V20_10 normalised over cultural groups using Model A
LIKEB	Likeability	LIKE normalised over cultural groups using Model B
V20_1B	Entertainment	V20_1 normalised over cultural groups using Model B
R20_2B	Confusion	R20_2 normalised over cultural groups using Model B
V20_3B	Relevant News	V20_3 normalised over cultural groups using Model B
V20_4B	Brand Reinforcement	V20_4 normalised over cultural groups using Model B
V20_5B	Empathy	V20_5 normalised over cultural groups using Model B
R20_6B	Familiarity	R20_6 normalised over cultural groups using Model B
R20_7B	Alienation	R20_7 normalised over cultural groups using Model B
V20_8B	Positive Moods	V20_8 normalised over cultural groups using Model B
R20_9B	Negative Moods	R20_9 normalised over cultural groups using Model B
V20_10B	Overall VRP20	V20_10 normalised over cultural groups using Model B
LIKED	Likeability	LIKE normalised over cultural groups using Model D
V20_1D	Entertainment	V20_1 normalised over cultural groups using Model D
R20_2D	Confusion	R20_2 normalised over cultural groups using Model D
V20_3D	Relevant News	V20_3 normalised over cultural groups using Model D
V20_4D	Brand Reinforcement	V20_4 normalised over cultural groups using Model D
V20_5D	Empathy	V20_5 normalised over cultural groups using Model D
R20_6D	Familiarity	R20_6 normalised over cultural groups using Model D
R20_7D	Alienation	R20_7 normalised over cultural groups using Model D
V20_8D	Positive Moods	V20_8 normalised over cultural groups using Model D
R20_9D	Negative Moods	R20_9 normalised over cultural groups using Model D
V20_10D	Overall VRP20	V20_10 normalised over cultural groups using Model D

Table B.2 gives the means and standard deviations of all the raw scores before they were transformed.

Table B.2: Means and standard deviations of all raw total scores

	Means				Standard Deviation			
	English	Afrikaans	Nguni	Sotho	English	Afrikaans	Nguni	Sotho
LIKEZ	6.773	6.972	7.647	7.801	2.209	2.110	1.876	1.792
V20_1Z	13.497	13.938	15.615	15.676	4.348	4.240	3.148	2.962
R20_2Z	12.451	12.330	11.820	12.039	2.440	2.569	2.699	2.512
V20_3Z	8.453	8.678	11.262	11.378	2.941	2.942	2.708	2.487
V20_4Z	7.677	7.920	8.076	7.996	1.752	1.759	1.540	1.460
V20_5Z	8.542	8.841	10.662	10.670	3.168	3.478	2.535	2.379
R20_6Z	7.853	8.096	8.486	8.432	1.701	1.621	1.253	1.171
R20_7Z	10.484	11.034	12.310	12.482	2.784	2.684	2.441	2.283
V20_8Z	38.169	39.377	45.615	45.719	9.468	9.721	7.452	6.512
R20_9Z	30.788	31.460	32.617	32.954	5.358	5.229	4.774	4.517
V20_10Z	68.957	70.837	78.232	78.673	13.478	13.693	10.808	9.699

Likeability could be considered to be a continuous variable. This score reflects the opinion of a viewer on how much he/she like the commercial viewed, with 1 meaning nor at all, and 10, liked it very much. This score was also subjected to normalisation. The frequency distribution of the Liking variable is given in Table B.3. The results of the normalisation of the Liking variable are given in Tables B.4, B.5 and B.6.

Table B.3: Frequency distribution of *Likeability* across cultural groups

	Culgroup				
Likeability	Afrikaans	English	Nguni	Sotho	Grand Total
1	29	28	8	2	67
2	33	27	10	4	74
3	48	37	28	12	125
4	94	73	49	38	254
5	123	99	106	79	407
6	119	122	126	81	448
7	224	208	184	131	747
8	278	336	385	288	1287
9	151	169	263	180	763
10	108	94	227	199	628
Grand Total	1207	1193	1386	1014	4800

Table B.4: Estimated Parameters using SIMNORM for *Likeability*

U=11.01 Number of parameters	MODEL A 12			MODEL B 9			MODEL D 6		
GROUP	PHI1	PHI2	GAMMA	PHI1	PHI2	GAMMA	PHI1	PHI2	GAMMA
English	2.4349	0.7967	0.1931	3.1328	0.7764	0.0988	3.4285	0.7426	0.0740
Afrikaans	19.7282	0.7198	0.0000	3.1887	0.7396		3.4839		
Nguni	3.2858	0.7399	0.1163	3.4532	0.7373		3.7468		
Sotho	1.9024	0.7824	0.4626	3.5174	0.7329		3.8107		

Table B.5: Fit statistics of SIMNORM for *Likeability*

	MODEL A	MODEL B	MODEL D
F	1.08225	1.08294	1.08345
AIC	4.00738	4.00750	4.00728
R-SQ	0.92614	0.92565	0.92535
RMSE	0.55695	0.55898	0.56030
TERMINATION CRITERION		1E-06	

Table B.6: Descriptive statistics of raw socres, standardised scores and normalised scores using SIMNORM for *Likeability*

Descriptive Statistics of Raw Scores					
Group	Mean	Std dev	Skewness	Kurtosis	N
English	6.273	2.209	-0.668	-0.145	1207
Afrikaans	6.472	2.110	-0.893	0.378	1193
Nguni	7.147	1.876	-0.907	0.586	1386
Sotho	7.301	1.792	-0.802	0.252	1014
Descriptive Statistics of Standardised Scores: LIKEZ					
Group	Mean	Std dev	Skewness	Kurtosis	N
English	0.000	1.000	-0.025	-0.130	1207
Afrikaans	0.000	1.000	-0.096	0.061	1193
Nguni	0.000	1.000	-0.052	-0.346	1386
Sotho	0.000	1.000	-0.076	-0.482	1014
Descriptive Statistics of Normalised Scores using Model A: LIKEA					
Group	Mean	Std dev	Skewness	Kurtosis	N
English	0.000	1.000	-0.025	-0.130	1207
Afrikaans	0.000	1.000	-0.096	0.061	1193
Nguni	0.000	1.000	-0.052	-0.346	1386
Sotho	0.000	1.000	-0.076	-0.482	1014
Descriptive Statistics of Normalised Scores using Model B: LIKEB					
Group	Mean	Std dev	Skewness	Kurtosis	N
English	0.000	1.000	0.026	-0.162	1207
Afrikaans	0.000	1.000	-0.153	0.110	1193
Nguni	0.000	1.000	-0.046	-0.355	1386
Sotho	0.000	1.000	0.031	-0.622	1014
Descriptive Statistics of Normalised Scores using Model D: LIKED					
Group	Mean	Std dev	Skewness	Kurtosis	N
English	0.000	1.039	0.039	-0.170	1207
Afrikaans	0.000	0.990	-0.139	0.097	1193
Nguni	0.000	0.988	-0.036	-0.368	1386
Sotho	0.000	0.983	0.038	-0.629	1014

The V20_8 *Positive Mood* variable was calculated by combining all the positively stated items into a summary score, by obtaining the sum of VV1- VV4, VV14-VV16 VV17-VV18, and VV20-VV22. The results of the normalisation are given in tables B.7 to B.9.

Table B.7: Estimated Parameters using SIMNORM for V20_8: *Positive Mood* VRP20 variable

U=61.01 Number of parameters	MODEL A 12			MODEL B 9			MODEL D 6		
GROUP	PHI1	PHI2	GAMMA	PHI1	PHI2	GAMMA	PHI1	PHI2	GAMMA
English	-1.0178	1.1321	2.8513	-0.6139	1.0200	2.2824	-0.2903	0.8809	1.8728
Afrikaans	-1.0548	1.2474	3.1396	-0.4786	1.0788		-0.1634		
Nguni	0.9034	0.7536	1.2812	0.1528	0.8671		0.4226		
Sotho	0.5044	0.7015	1.7464	0.1458	0.7535		0.4133		

Table B.8: Fit statistics SIMNORM: V20_8: *Positive Mood* VRP20 variable

	MODEL A	MODEL B	MODEL D
F	2.60081	2.60401	2.62062
AIC	7.04450	7.04965	7.08161
R-SQ	0.97894	0.97819	0.97017
RMSE	1.32077	1.34450	1.57281
TERMINATION CRITERION		1E-06	

Table B.9: Descriptive statistics of raw socres, standardised scores and normalised scores using SIMNORM for V20_8: *Positive Mood* VRP20 variable

Descriptive Statistics of Raw Scores					
Group	Mean	Std dev	Skewness	Kurtosis	N
English	37.669	9.468	-0.313	-0.291	1207
Afrikaans	38.877	9.721	-0.363	-0.268	1193
Nguni	45.115	7.452	-1.197	2.354	1386
Sotho	45.219	6.512	-1.058	2.414	1014
Descriptive Statistics of Standardised Scores: V20_8Z					
Group	Mean	Std dev	Skewness	Kurtosis	N
English	0.000	1.000	-0.314	-0.288	1207
Afrikaans	0.000	1.000	-0.364	-0.264	1193
Nguni	0.000	1.000	-1.198	2.367	1386
Sotho	0.000	1.000	-1.059	2.432	1014
Descriptive Statistics of Normalised Scores using Model A: V20_8A					
Group	Mean	Std dev	Skewness	Kurtosis	N
English	0.000	1.000	0.065	1.364	1207
Afrikaans	0.000	1.000	0.044	1.241	1193
Nguni	0.000	1.000	0.327	2.633	1386
Sotho	0.000	1.000	0.297	3.257	1014
Descriptive Statistics of Normalised Scores using Model B: V20_8B					
Group	Mean	Std dev	Skewness	Kurtosis	N
English	0.000	1.000	0.274	1.695	1207
Afrikaans	0.000	1.000	0.353	1.628	1193
Nguni	0.000	1.000	-0.115	2.615	1386
Sotho	0.000	1.000	0.071	3.253	1014
Descriptive Statistics of Normalised Scores using Model D: V20_8D					
Group	Mean	Std dev	Skewness	Kurtosis	N
English	0.000	1.072	0.448	2.043	1207
Afrikaans	0.000	1.141	0.526	1.927	1193
Nguni	0.000	0.930	0.061	2.570	1386
Sotho	0.000	0.810	0.242	3.244	1014

The V20_9 NEGATIVE MOOD variable was calculated by combining all the reverse scored negatively stated items into a summary score, by obtaining the sum of VV8-VV10, VV24-VV25 and VV27-VV29. The results of the normalisation are given in tables B.10 to B.12.

Table B.10: Estimated Parameters using SIMNORM for R20_9: *Negative Mood* VRP20 variable

U=42.01 Number of parameters	MODEL A 12			MODEL B 9			MODEL D 6		
GROUP	PHI1	PHI2	GAMMA	PHI1	PHI2	GAMMA	PHI1	PHI2	GAMMA
English	-0.1207	0.9239	2.5958	-0.0423	0.9076	2.4627	0.0134	0.8689	2.3708
Afrikaans	-0.1745	0.9651	2.9144	0.0751	0.9126	2.4627	0.1294	0.8689	2.3708
Nguni	0.3707	0.8441	2.2758	0.2635	0.8624	2.4627	0.3155	0.8689	2.3708
Sotho	0.6090	0.7790	1.9668	0.3134	0.8237	2.4627	0.3647	0.8689	2.3708

Table B.11: SIMNORM Fit statistics for R20_9: *Negative Mood* VRP20 variable

	MODEL A	MODEL B	MODEL D
F	2.05630	2.05665	2.05820
AIC	5.95547	5.95493	5.95678
R-SQ	0.97046	0.97021	0.96920
RMSE	0.86907	0.87294	0.88790
TERMINATION CRITERION	1E-06		

Table B.12: Descriptive statistics of raw socres, standardised scores and normalised scores using SIMNORM for R20_9: *Negative Mood* VRP20 variable

Descriptive Statistics of Raw Scores					
Group	Mean	Std dev	Skewness	Kurtosis	N
English	30.288	5.358	-0.567	0.286	1207
Afrikaans	30.960	5.229	-0.598	0.356	1193
Nguni	32.117	4.774	-0.885	1.137	1386
Sotho	32.454	4.517	-0.987	1.748	1014
Descriptive Statistics of Standardised Scores: R20_9Z					
Group	Mean	Std dev	Skewness	Kurtosis	N
English	0.000	1.000	-0.287	-0.290	1207
Afrikaans	0.000	1.000	-0.395	-0.146	1193
Nguni	0.000	1.000	-1.403	2.850	1386
Sotho	0.000	1.000	-1.235	2.898	1014
Descriptive Statistics of Normalised Scores using Model A: R20_9A					
Group	Mean	Std dev	Skewness	Kurtosis	N
English	0.000	1.000	0.010	0.560	1207
Afrikaans	0.000	1.000	-0.003	0.389	1193
Nguni	0.000	1.000	0.047	0.459	1386
Sotho	0.000	1.000	0.070	0.615	1014
Descriptive Statistics of Normalised Scores using Model B: R20_9B					
Group	Mean	Std dev	Skewness	Kurtosis	N
English	0.000	1.000	0.043	0.532	1207
Afrikaans	0.000	1.000	0.096	0.308	1193
Nguni	0.000	1.000	0.005	0.490	1386
Sotho	0.000	1.000	-0.048	0.743	1014
Descriptive Statistics of Normalised Scores using Model D: R20_9D					
Group	Mean	Std dev	Skewness	Kurtosis	N
English	0.000	1.032	0.065	0.513	1207
Afrikaans	0.000	1.039	0.117	0.294	1193
Nguni	0.000	0.982	0.025	0.475	1386
Sotho	0.000	0.939	-0.026	0.717	1014

The V20_10 *Overall Schlinger VRP20* variable was calculated by combining all the items into a summary score, by obtaining the sum of V20_8 + R20_9. The results of the normalisation are given in tables B.13 to B.15.

Table B.13: Estimated Parameters using SIMNORM for V20_10: *Overall Schlinger VRP20*

U=102.01 Number of parameters	MODEL A 12			MODEL B 9			MODEL D 6		
GROUP	PHI1	PHI2	GAMMA	PHI1	PHI2	GAMMA	PHI1	PHI2	GAMMA
English	-1.2185	1.0908	3.8213	-0.2762	0.8593	2.2380	-0.0681	0.7658	1.9493
Afrikaans	-0.9954	1.1100	3.6856	-0.1523	0.9022		0.0509		
Nguni	2.0769	0.5764	0.4860	0.2869	0.7303		0.4700		
Sotho	1.0015	0.5965	1.2882	0.3090	0.6713		0.4904		

Table B.14: SIMNORM fit statistics for V20_10: *Overall Schlinger VRP20*

	MODEL A	MODEL B	MODEL D
F	2.94267	2.95102	2.96396
AIC	7.72821	7.74366	7.76830
R-SQ	0.98329	0.97879	0.97250
RMSE	1.65533	1.86600	2.12511
TERMINATION CRITERION	1E-06		

Table B.15: Descriptive statistics of raw socres, standardised scores and normalised scores using SIMNORM for V20_10: *Overall Schlinger VRP20* variable

Descriptive Statistics of Raw Scores					
Group	Mean	Std dev	Skewness	Kurtosis	N
English	68.457	13.478	-0.286	-0.294	1207
Afrikaans	70.337	13.693	-0.395	-0.150	1193
Nguni	77.732	10.808	-1.401	2.835	1386
Sotho	78.173	9.699	-1.234	2.878	1014
Descriptive Statistics of Standardised Scores: V20_10Z					
Group	Mean	Std dev	Skewness	Kurtosis	N
English	0.000	1.000	-0.568	0.292	1207
Afrikaans	0.000	1.000	-0.599	0.362	1193
Nguni	0.000	1.000	-0.886	1.146	1386
Sotho	0.000	1.000	-0.988	1.762	1014
Descriptive Statistics of Normalised Scores using Model A: V20_10A					
Group	Mean	Std dev	Skewness	Kurtosis	N
English	0.000	1.000	0.017	0.814	1207
Afrikaans	0.000	1.000	0.018	1.016	1193
Nguni	0.000	1.000	0.357	2.860	1386
Sotho	0.000	1.000	0.312	2.937	1014
Descriptive Statistics of Normalised Scores using Model B: V20_10B					
Group	Mean	Std dev	Skewness	Kurtosis	N
English	0.000	1.000	0.419	1.382	1207
Afrikaans	0.000	1.000	0.413	1.404	1193
Nguni	0.000	1.000	-0.315	2.573	1386
Sotho	0.000	1.000	-0.023	2.808	1014
Descriptive Statistics of Normalised Scores using Model D: V20_10D					
Group	Mean	Std dev	Skewness	Kurtosis	N
English	0.000	1.073	0.508	1.561	1207
Afrikaans	0.000	1.129	0.505	1.546	1193
Nguni	0.000	0.918	-0.212	2.551	1386
Sotho	0.000	0.846	0.076	2.818	1014

APPENDIX C: Results of SIMITNORM

Table C.1: Item Frequencies

Group	English					Afrikaans					Nguni					Sotho				
Item	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	93	191	242	463	218	76	141	163	543	270	24	106	166	686	404	10	82	98	509	315
2	96	199	216	477	219	85	184	199	525	200	29	92	147	758	360	13	87	94	569	251
3	103	209	276	423	196	101	205	226	477	184	29	72	158	746	381	8	52	108	610	236
4	113	219	245	454	176	94	199	187	526	187	54	151	235	668	278	21	130	158	520	185
8	20	79	127	544	437	33	111	104	508	437	46	137	91	742	370	19	75	44	600	276
9	21	99	69	512	506	30	110	64	511	478	75	172	71	664	404	39	127	33	507	308
10	21	57	75	526	528	14	61	79	518	521	36	87	71	719	473	21	65	30	564	334
14	229	450	212	263	53	208	430	221	280	54	60	193	172	703	258	26	164	118	520	186
15	154	294	342	341	76	135	258	341	357	102	73	196	160	618	339	26	149	106	488	245
16	139	309	278	387	94	150	282	250	399	112	39	129	123	708	387	16	81	103	557	257
17	38	59	307	521	282	33	42	251	512	355	7	25	268	595	491	4	13	210	463	324
18	20	51	274	559	303	22	50	224	523	374	12	42	350	558	424	10	19	270	457	258
20	150	312	314	339	92	167	276	241	377	132	43	149	262	755	177	23	110	187	604	90
21	197	351	309	274	76	224	327	231	305	106	66	210	330	619	161	22	172	231	483	106
22	200	278	293	339	97	216	209	246	366	156	53	140	299	712	182	24	120	213	560	97
24	30	112	219	539	307	27	101	155	562	348	16	53	151	679	487	8	41	96	571	298
25	23	58	163	566	397	13	61	103	548	468	6	19	45	723	593	2	7	24	595	386
27	48	146	315	516	182	26	117	287	512	251	42	93	176	659	416	12	63	113	504	322
28	64	192	213	521	217	29	166	167	534	297	26	69	64	693	534	11	46	35	531	391
29	72	228	234	456	217	75	199	169	473	277	34	78	56	656	562	14	57	38	500	405

Table C.2: Item Percentages

	English					Afrikaans					Nguni					Sotho				
Item	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
1	8	16	20	38	18	6	12	14	46	23	2	8	12	49	29	1	8	10	50	31
2	8	16	18	40	18	7	15	17	44	17	2	7	11	55	26	1	9	9	56	25
3	9	17	23	35	16	8	17	19	40	15	2	5	11	54	27	1	5	11	60	23
4	9	18	20	38	15	8	17	16	44	16	4	11	17	48	20	2	13	16	51	18
8	2	7	11	45	36	3	9	9	43	37	3	10	7	54	27	2	7	4	59	27
9	2	8	6	42	42	3	9	5	43	40	5	12	5	48	29	4	13	3	50	30
10	2	5	6	44	44	1	5	7	43	44	3	6	5	52	34	2	6	3	56	33
14	19	37	18	22	4	17	36	19	23	5	4	14	12	51	19	3	16	12	51	18
15	13	24	28	28	6	11	22	29	30	9	5	14	12	45	24	3	15	10	48	24
16	12	26	23	32	8	13	24	21	33	9	3	9	9	51	28	2	8	10	55	25
17	3	5	25	43	23	3	4	21	43	30	1	2	19	43	35	0	1	21	46	32
18	2	4	23	46	25	2	4	19	44	31	1	3	25	40	31	1	2	27	45	25
20	12	26	26	28	8	14	23	20	32	11	3	11	19	54	13	2	11	18	60	9
21	16	29	26	23	6	19	27	19	26	9	5	15	24	45	12	2	17	23	48	10
22	17	23	24	28	8	18	18	21	31	13	4	10	22	51	13	2	12	21	55	10
24	2	9	18	45	25	2	8	13	47	29	1	4	11	49	35	1	4	9	56	29
25	2	5	14	47	33	1	5	9	46	39	0	1	3	52	43	0	1	2	59	38
27	4	12	26	43	15	2	10	24	43	21	3	7	13	48	30	1	6	11	50	32
28	5	16	18	43	18	2	14	14	45	25	2	5	5	50	39	1	5	3	52	39
29	6	19	19	38	18	6	17	14	40	23	2	6	4	47	41	1	6	4	49	40

Table C.3: Descriptive Statistics of Raw Scores

Raw	Mean				Standard Deviation				Skewness				Kurtosis			
Item	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
1	2.933	3.162	3.467	3.523	1.178	1.138	0.934	0.905	-0.506	-0.831	-1.008	-1.019	-0.655	-0.102	0.845	0.825
2	2.934	2.979	3.458	3.445	1.190	1.150	0.905	0.893	-0.522	-0.625	-1.135	-1.051	-0.691	-0.497	1.445	1.062
3	2.831	2.867	3.494	3.500	1.186	1.181	0.886	0.785	-0.384	-0.486	-1.159	-1.030	-0.772	-0.719	1.700	1.750
4	2.799	2.930	3.196	3.208	1.195	1.169	1.032	0.976	-0.408	-0.591	-0.815	-0.753	-0.814	-0.612	0.164	0.040
8	3.576	3.510	3.404	3.525	0.934	1.037	1.009	0.883	-1.122	-1.117	-1.161	-1.346	1.080	0.666	0.949	2.111
9	3.646	3.587	3.330	3.405	0.970	1.021	1.139	1.084	-1.293	-1.263	-1.047	-1.127	1.263	1.042	0.228	0.515
10	3.729	3.733	3.587	3.610	0.891	0.871	0.933	0.889	-1.456	-1.359	-1.405	-1.472	2.335	1.966	2.097	2.543
14	2.053	2.116	3.154	3.167	1.152	1.152	1.067	1.033	0.356	0.279	-0.823	-0.741	-0.910	-0.971	-0.037	-0.227
15	2.410	2.528	3.188	3.266	1.132	1.145	1.142	1.054	-0.090	-0.165	-0.791	-0.803	-0.886	-0.833	-0.257	-0.136
16	2.490	2.534	3.420	3.445	1.162	1.204	0.994	0.902	-0.124	-0.176	-1.097	-1.061	-0.990	-1.037	0.852	1.117
17	3.287	3.434	3.610	3.575	0.960	0.946	0.809	0.780	-0.750	-0.904	-0.636	-0.493	0.536	0.867	0.153	-0.020
18	3.390	3.487	3.467	3.421	0.886	0.914	0.870	0.824	-0.722	-0.884	-0.528	-0.478	0.609	0.773	-0.041	0.282
20	2.426	2.526	3.131	3.119	1.156	1.245	0.944	0.876	-0.043	-0.141	-0.862	-0.962	-0.936	-1.092	0.456	0.705
21	2.236	2.284	2.932	2.972	1.164	1.261	1.033	0.964	0.143	0.116	-0.563	-0.504	-0.922	-1.137	-0.309	-0.414
22	2.380	2.531	3.099	3.078	1.217	1.315	0.967	0.903	-0.054	-0.190	-0.815	-0.809	-1.052	-1.154	0.381	0.338
24	3.313	3.425	3.631	3.595	0.998	0.979	0.837	0.783	-0.780	-0.969	-1.115	-1.069	0.171	0.596	1.649	1.895
25	3.541	3.671	3.855	3.837	0.909	0.868	0.655	0.580	-1.073	-1.209	-1.185	-0.667	1.249	1.573	3.340	2.382
27	3.029	3.208	3.448	3.546	1.016	0.977	0.983	0.885	-0.562	-0.583	-1.087	-1.040	-0.147	-0.082	0.996	1.098
28	3.026	3.258	3.683	3.728	1.116	1.052	0.875	0.805	-0.593	-0.725	-1.481	-1.457	-0.481	-0.239	2.660	2.995
29	2.929	3.068	3.679	3.708	1.158	1.192	0.928	0.862	-0.422	-0.603	-1.524	-1.458	-0.770	-0.647	2.461	2.505

Table C.4: Estimated Parameters: Model A

Model A		240 Parameters											
		English			Afrikaans			Nguni			Sotho		
Item	U	Phi1	Phi2	Gamma	Phi1	Phi2	Gamma	Phi1	Phi2	Gamma	Phi1	Phi2	Gamma
VV1	5.51	0.814	1.039	0.859	2.082	0.909	0.337	2.618	0.804	0.253	2.590	0.796	0.273
VV2	5.51	0.878	1.037	0.817	1.238	0.951	0.620	3.577	0.750	0.098	2.542	0.761	0.263
VV3	5.51	0.436	1.094	1.069	0.778	1.022	0.840	3.811	0.745	0.080	1.873	0.725	0.503
VV4	5.51	0.535	1.064	0.969	1.147	0.964	0.647	1.798	0.854	0.440	1.374	0.843	0.640
VV8	5.21	1.387	1.296	1.265	1.579	1.337	1.030	1.354	1.255	1.050	1.418	1.165	1.092
VV9	5.21	2.098	1.257	0.761	2.043	1.284	0.752	1.392	1.363	0.987	1.508	1.317	0.957
VV10	5.21	2.471	1.182	0.584	2.147	1.199	0.789	1.931	1.198	0.782	1.887	1.163	0.817
VV14	5.51	-1.869	1.494	2.438	-1.561	1.400	2.216	1.953	0.851	0.370	1.443	0.869	0.589
VV15	5.51	-0.312	1.080	1.349	-0.144	1.089	1.304	1.866	0.937	0.422	1.727	0.895	0.501
VV16	5.51	-0.258	1.117	1.371	-0.123	1.142	1.292	4.299	0.795	0.047	2.714	0.764	0.224
VV17	5.51	1.449	0.869	0.642	2.160	0.841	0.382	1.033	0.879	1.194	0.451	0.927	1.822
VV18	5.51	1.221	0.856	0.842	2.012	0.837	0.457	0.671	0.950	1.420	0.450	0.923	1.599
VV20	5.51	-0.519	1.172	1.549	-0.266	1.231	1.413	2.069	0.734	0.317	3.540	0.626	0.075
VV21	5.51	-1.120	1.315	1.924	-1.105	1.446	1.953	0.987	0.863	0.726	0.628	0.852	0.990
VV22	5.51	-0.496	1.221	1.483	-0.137	1.278	1.306	1.870	0.762	0.374	1.919	0.690	0.346
VV24	5.31	0.467	1.421	1.928	0.929	1.345	1.534	1.161	1.257	1.575	0.707	1.245	2.086
VV25	5.31	1.626	1.057	0.837	2.169	1.011	0.591	1.775	0.921	0.993	0.311	1.047	3.154
VV27	5.26	0.100	1.273	1.796	0.079	1.347	2.146	1.474	1.172	0.936	1.172	1.163	1.315
VV28	5.26	0.308	1.344	1.582	0.561	1.332	1.629	2.608	1.049	0.427	2.230	1.030	0.628
VV29	5.26	-0.108	1.488	1.947	0.450	1.443	1.521	3.229	1.058	0.240	2.565	1.049	0.459

Table C.5: Descriptive Statistics of SIMITNORM Scores: Model A

Model	Mean				Standard Deviation				Skewness				Kurtosis			
A	English	Afr	Nguni	Sotho	English	Afr	Nguni	Sotho	English	Afr	Nguni	Sotho	English	Afr	Nguni	Sotho
1	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.079	-0.129	-0.102	-0.121	-0.578	-0.580	-0.583	-0.660
2	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.089	-0.080	-0.038	-0.036	-0.616	-0.484	-0.339	-0.313
3	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.057	-0.069	-0.039	0.047	-0.552	-0.536	-0.381	0.022
4	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.060	-0.083	-0.051	-0.028	-0.556	-0.511	-0.333	-0.243
8	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.168	-0.208	-0.055	0.000	-0.798	-0.896	-0.450	-0.264
9	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.275	-0.262	-0.142	-0.142	-1.013	-0.974	-0.725	-0.714
10	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.269	-0.264	-0.117	-0.077	-1.024	-1.032	-0.679	-0.580
14	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.076	0.056	-0.072	-0.065	-0.687	-0.660	-0.366	-0.395
15	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.006	-0.002	-0.150	-0.119	-0.352	-0.336	-0.689	-0.616
16	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.009	-0.023	-0.123	-0.042	-0.481	-0.589	-0.557	-0.344
17	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.024	-0.086	-0.129	-0.101	-0.284	-0.580	-0.756	-0.454
18	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.027	-0.102	-0.098	-0.050	-0.273	-0.648	-0.529	-0.020
20	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.004	-0.024	0.050	0.111	-0.438	-0.700	0.235	0.711
21	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.032	0.037	0.003	0.010	-0.537	-0.818	-0.058	-0.004
22	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.003	-0.041	0.043	0.077	-0.639	-0.855	0.173	0.464
24	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.072	-0.093	-0.125	-0.059	-0.367	-0.537	-0.601	-0.106
25	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.109	-0.192	-0.146	-0.175	-0.648	-0.905	-0.863	0.464
27	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.009	-0.051	-0.093	-0.097	0.105	-0.117	-0.578	-0.568
28	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.029	-0.083	-0.168	-0.138	-0.354	-0.561	-0.833	-0.799
29	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.041	-0.090	-0.232	-0.193	-0.468	-0.690	-0.912	-0.891

Table C.6: Estimated Parameters: Model B

Model B 180 Parameters			English		Afrikaans		Nguni		Sotho	
Item	U	Gamma	Phi1	Phi2	Phi1	Phi2	Phi1	Phi2	Phi1	Phi2
1	5.51	0.475	1.523	0.947	1.705	0.938	1.948	0.836	1.997	0.823
2	5.51	0.502	1.461	0.961	1.484	0.924	1.863	0.808	1.848	0.794
3	5.51	0.682	1.014	0.994	1.036	0.982	1.555	0.831	1.536	0.747
4	5.51	0.683	0.981	0.989	1.083	0.973	1.298	0.900	1.298	0.850
8	5.21	1.098	1.555	1.269	1.503	1.350	1.300	1.264	1.412	1.166
9	5.21	0.884	1.931	1.276	1.861	1.307	1.524	1.340	1.602	1.302
10	5.21	0.745	2.208	1.204	2.210	1.193	1.985	1.192	1.991	1.152
14	5.51	1.372	-0.668	1.103	-0.610	1.100	0.381	1.066	0.395	1.027
15	5.51	0.922	0.255	0.954	0.358	0.980	0.957	1.047	1.018	0.975
16	5.51	0.824	0.480	0.958	0.520	1.002	1.288	0.926	1.293	0.853
17	5.51	0.795	1.201	0.895	1.345	0.907	1.508	0.830	1.463	0.803
18	5.51	0.995	1.022	0.879	1.131	0.919	1.112	0.889	1.048	0.839
20	5.51	0.929	0.266	0.985	0.359	1.078	0.841	0.842	0.813	0.762
21	5.51	1.429	-0.566	1.143	-0.519	1.251	0.096	1.021	0.133	0.939
22	5.51	0.961	0.178	1.050	0.319	1.160	0.774	0.871	0.742	0.794
24	5.31	1.095	1.059	1.132	1.186	1.132	1.411	1.047	1.333	0.980
25	5.31	0.999	1.423	1.081	1.586	1.066	1.768	0.922	1.714	0.864
27	5.26	1.513	0.349	1.210	0.578	1.222	0.885	1.278	0.999	1.193
28	5.26	1.133	0.765	1.236	1.036	1.231	1.534	1.148	1.575	1.093
29	5.26	1.132	0.682	1.276	0.852	1.343	1.550	1.194	1.571	1.141

Table C.7: Descriptive Statistics of SIMITNORM Scores: Model B

Model B	Mean				Standard Deviation				Skewness				Kurtosis			
Item	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
1	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.061	-0.181	-0.172	-0.178	-0.637	-0.536	-0.471	-0.568
2	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.024	-0.034	-0.192	-0.119	-0.663	-0.503	-0.099	-0.203
3	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.083	-0.010	-0.269	-0.018	-0.595	-0.553	0.039	0.126
4	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.045	-0.096	-0.147	-0.044	-0.574	-0.507	-0.243	-0.231
8	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.128	-0.225	-0.071	-0.002	-0.877	-0.866	-0.429	-0.261
9	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.301	-0.293	-0.110	-0.120	-0.957	-0.914	-0.761	-0.741
10	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.304	-0.256	-0.106	-0.055	-0.937	-1.053	-0.699	-0.621
14	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.349	0.282	-0.451	-0.330	-0.440	-0.482	-0.007	-0.175
15	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.171	0.147	-0.319	-0.252	-0.265	-0.291	-0.513	-0.475
16	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.190	0.142	-0.403	-0.259	-0.403	-0.538	-0.125	-0.014
17	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.083	-0.230	-0.050	0.100	-0.195	-0.299	-0.944	-0.928
18	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.079	-0.270	0.004	0.126	-0.178	-0.293	-0.748	-0.402
20	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.229	0.136	-0.232	-0.312	-0.336	-0.659	0.417	0.856
21	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.183	0.177	-0.282	-0.149	-0.437	-0.739	0.096	0.056
22	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.180	0.066	-0.232	-0.207	-0.538	-0.826	0.364	0.569
24	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.002	-0.142	-0.166	0.026	-0.534	-0.522	-0.561	-0.361
25	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.155	-0.288	-0.147	0.197	-0.554	-0.681	-0.859	-1.133
27	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.118	0.145	-0.281	-0.151	0.040	-0.333	-0.246	-0.463
28	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.127	0.057	-0.369	-0.269	-0.426	-0.697	-0.348	-0.465
29	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.218	0.027	-0.486	-0.366	-0.558	-0.770	-0.301	-0.468

Table C.8: Estimated Parameters: Model D

Model D		120 Parameters					
		Common Parameters		English	Afrikaans	Nguni	Sotho
Item	U	Phi2	Gamma	Phi1	Phi1	Phi1	Phi1
1	5.510	0.870	0.372	1.799	1.977	2.213	2.262
2	5.510	0.852	0.377	1.785	1.807	2.174	2.159
3	5.510	0.855	0.463	1.472	1.492	1.983	1.963
4	5.510	0.917	0.618	1.103	1.203	1.414	1.414
8	5.210	1.266	1.090	1.563	1.511	1.308	1.420
9	5.210	1.305	0.866	1.953	1.884	1.548	1.626
10	5.210	1.188	0.749	2.201	2.204	1.978	1.984
14	5.510	1.058	1.308	-0.585	-0.528	0.446	0.460
15	5.510	1.003	0.963	0.193	0.298	0.902	0.964
16	5.510	0.915	0.717	0.660	0.700	1.448	1.452
17	5.510	0.846	0.690	1.366	1.508	1.667	1.622
18	5.510	0.882	0.981	1.040	1.148	1.129	1.065
20	5.510	0.888	0.763	0.531	0.622	1.079	1.050
21	5.510	1.046	1.248	-0.340	-0.292	0.289	0.322
22	5.510	0.932	0.766	0.490	0.627	1.052	1.018
24	5.310	1.065	1.017	1.150	1.275	1.497	1.419
25	5.310	0.961	0.748	1.755	1.912	2.087	2.032
27	5.260	1.229	1.511	0.351	0.579	0.887	1.001
28	5.260	1.165	1.044	0.871	1.137	1.629	1.668
29	5.260	1.222	1.031	0.805	0.973	1.659	1.679

Table C.9: Descriptive Statistics of SIMITNORM Scores: Model D

	Means				Standard Deviation				Skewness				Kurtosis			
Item	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
1	0.000	0.000	0.000	0.000	1.063	1.054	0.944	0.931	0.100	-0.142	-0.139	-0.148	-0.648	-0.570	-0.525	-0.616
2	0.000	0.000	0.000	0.000	1.095	1.053	0.927	0.912	0.070	0.015	-0.143	-0.075	-0.676	-0.519	-0.179	-0.262
3	0.000	0.000	0.000	0.000	1.102	1.087	0.934	0.843	0.166	0.073	-0.182	0.061	-0.605	-0.567	-0.130	0.000
4	0.000	0.000	0.000	0.000	1.061	1.043	0.968	0.915	0.070	-0.072	-0.121	-0.021	-0.576	-0.515	-0.269	-0.249
8	0.000	0.000	0.000	0.000	1.002	1.065	0.997	0.920	-0.126	-0.223	-0.069	0.001	-0.881	-0.869	-0.433	-0.265
9	0.000	0.000	0.000	0.000	0.976	1.000	1.025	0.996	-0.298	-0.289	-0.105	-0.115	-0.965	-0.922	-0.767	-0.748
10	0.000	0.000	0.000	0.000	1.015	1.005	1.005	0.971	-0.305	-0.257	-0.108	-0.056	-0.935	-1.051	-0.697	-0.618
14	0.000	0.000	0.000	0.000	1.023	1.020	0.994	0.958	0.370	0.303	-0.428	-0.309	-0.416	-0.461	-0.036	-0.197
15	0.000	0.000	0.000	0.000	0.963	0.989	1.054	0.981	0.154	0.130	-0.333	-0.266	-0.276	-0.298	-0.496	-0.460
16	0.000	0.000	0.000	0.000	1.016	1.063	0.992	0.914	0.232	0.182	-0.364	-0.220	-0.377	-0.518	-0.194	-0.080
17	0.000	0.000	0.000	0.000	1.037	1.052	0.967	0.936	-0.042	-0.192	-0.030	0.119	-0.257	-0.377	-0.988	-0.963
18	0.000	0.000	0.000	0.000	0.995	1.040	1.007	0.950	-0.074	-0.265	0.008	0.130	-0.187	-0.303	-0.755	-0.410
20	0.000	0.000	0.000	0.000	1.058	1.159	0.914	0.826	0.296	0.195	-0.157	-0.233	-0.288	-0.633	0.354	0.801
21	0.000	0.000	0.000	0.000	1.037	1.136	0.935	0.863	0.247	0.232	-0.212	-0.085	-0.384	-0.699	0.045	0.025
22	0.000	0.000	0.000	0.000	1.063	1.178	0.894	0.815	0.254	0.132	-0.142	-0.119	-0.480	-0.800	0.287	0.519
24	0.000	0.000	0.000	0.000	1.050	1.050	0.974	0.912	0.022	-0.118	-0.146	0.048	-0.563	-0.559	-0.610	-0.405
25	0.000	0.000	0.000	0.000	1.087	1.075	0.936	0.880	-0.083	-0.228	-0.102	0.228	-0.698	-0.824	-1.015	-1.232
27	0.000	0.000	0.000	0.000	0.984	0.994	1.040	0.971	0.119	0.146	-0.280	-0.150	0.040	-0.334	-0.247	-0.464
28	0.000	0.000	0.000	0.000	1.044	1.043	0.975	0.929	0.158	0.082	-0.342	-0.245	-0.436	-0.718	-0.418	-0.530
29	0.000	0.000	0.000	0.000	1.025	1.079	0.964	0.922	0.252	0.058	-0.455	-0.338	-0.562	-0.788	-0.382	-0.539

Table C.10: Estimated Parameters: Model BB

Model		84 Parameters							
Bb		English		Afrikaans		Nguni		Sotho	
	Gamma	0.973		0.825		0.193		0.319	
Item	U	Phi1	Phi2	Phi1	Phi2	Phi1	Phi2	Phi1	Phi2
1	5.51	0.653	1.068	1.061	1.017	2.902	0.795	2.422	0.802
2	5.51	0.655	1.077	0.892	1.000	2.879	0.763	2.335	0.768
3	5.51	0.564	1.068	0.801	1.018	2.909	0.760	2.361	0.704
4	5.51	0.530	1.065	0.850	1.008	2.680	0.810	2.143	0.789
8	5.51	1.255	0.934	1.400	0.983	2.853	0.818	2.403	0.767
9	5.51	1.341	0.967	1.477	0.975	2.823	0.900	2.347	0.895
10	5.51	1.414	0.913	1.605	0.877	3.010	0.800	2.492	0.788
14	5.51	-0.110	0.975	0.179	0.930	2.648	0.818	2.120	0.820
15	5.51	0.180	0.968	0.507	0.954	2.705	0.891	2.218	0.862
16	5.51	0.253	0.999	0.518	1.002	2.868	0.818	2.338	0.777
17	5.51	0.958	0.927	1.302	0.912	3.023	0.762	2.455	0.751
18	5.51	1.049	0.875	1.351	0.892	2.903	0.785	2.315	0.754
20	5.51	0.201	0.997	0.519	1.047	2.594	0.714	2.030	0.662
21	5.51	0.042	0.999	0.324	1.050	2.464	0.759	1.944	0.732
22	5.51	0.161	1.053	0.529	1.116	2.576	0.731	2.006	0.686
24	5.51	0.992	0.955	1.295	0.924	3.039	0.762	2.458	0.728
25	5.51	1.209	0.911	1.534	0.869	3.226	0.676	2.667	0.642
27	5.51	0.715	0.925	1.088	0.898	2.896	0.824	2.440	0.794
28	5.51	0.729	1.015	1.149	0.961	3.095	0.780	2.600	0.757
29	5.51	0.653	1.050	0.996	1.061	3.105	0.810	2.596	0.791

Table C.11: Descriptive Statistics of SIMITNORM Scores: Model BB

Model BB	Means				Standard Deviation				Skewness				Kurtosis			
Item	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
1	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.120	-0.312	-0.083	-0.134	-0.555	-0.404	-0.612	-0.639
2	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.144	-0.157	-0.073	-0.055	-0.587	-0.444	-0.288	-0.288
3	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.023	-0.064	-0.080	0.112	-0.565	-0.538	-0.313	-0.074
4	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.061	-0.150	0.045	0.088	-0.556	-0.485	-0.408	-0.315
8	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.480	-0.523	-0.217	-0.229	-0.231	-0.391	-0.371	-0.013
9	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.685	-0.645	-0.290	-0.356	-0.130	-0.240	-0.659	-0.530
10	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.747	-0.620	-0.354	-0.370	0.242	-0.149	-0.307	-0.077
14	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.487	0.475	-0.003	0.030	-0.263	-0.255	-0.405	-0.441
15	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.150	0.187	-0.073	-0.061	-0.279	-0.273	-0.753	-0.666
16	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.133	0.142	-0.174	-0.076	-0.432	-0.538	-0.488	-0.298
17	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.153	-0.241	0.057	0.183	-0.081	-0.276	-1.161	-1.071
18	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.072	-0.215	0.178	0.304	-0.192	-0.415	-1.041	-0.689
20	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.212	0.172	0.108	-0.014	-0.347	-0.644	0.216	0.716
21	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.351	0.374	0.236	0.268	-0.283	-0.575	-0.072	0.016
22	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.176	0.112	0.128	0.090	-0.541	-0.809	0.141	0.463
24	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.210	-0.305	-0.142	0.002	-0.353	-0.339	-0.695	-0.401
25	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.397	-0.464	-0.155	0.168	-0.104	-0.297	-0.880	-1.075
27	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.047	0.030	-0.171	-0.132	-0.065	-0.387	-0.557	-0.609
28	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.151	-0.161	-0.400	-0.354	-0.431	-0.601	-0.349	-0.313
29	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.032	-0.165	-0.510	-0.446	-0.616	-0.744	-0.315	-0.335

Table C.12: Estimated Parameters: Model BBB

Model	81 Parameters							
BBB	Gamma		0.622					
	English		Afrikaans		Nguni		Sotho	
Item	Phi1	Phi2	Phi1	Phi2	Phi1	Phi2	Phi1	Phi2
1	1.209	0.981	1.398	0.971	1.650	0.859	1.701	0.844
2	1.211	0.989	1.235	0.951	1.626	0.826	1.611	0.812
3	1.127	0.979	1.149	0.967	1.659	0.822	1.640	0.740
4	1.095	0.974	1.195	0.958	1.407	0.888	1.407	0.840
8	1.767	0.879	1.723	0.945	1.594	0.893	1.682	0.811
9	1.849	0.908	1.798	0.938	1.555	0.989	1.617	0.953
10	1.917	0.860	1.920	0.848	1.763	0.866	1.774	0.832
14	0.516	0.872	0.560	0.872	1.371	0.900	1.380	0.875
15	0.773	0.873	0.868	0.901	1.428	0.980	1.483	0.917
16	0.840	0.904	0.880	0.946	1.611	0.890	1.614	0.822
17	1.486	0.865	1.626	0.878	1.782	0.810	1.738	0.784
18	1.569	0.823	1.673	0.861	1.653	0.839	1.591	0.791
20	0.793	0.902	0.883	0.989	1.320	0.786	1.290	0.710
21	0.651	0.898	0.700	0.989	1.172	0.841	1.196	0.784
22	0.761	0.949	0.895	1.054	1.298	0.806	1.264	0.735
24	1.519	0.892	1.620	0.889	1.799	0.815	1.741	0.763
25	1.721	0.857	1.851	0.841	2.001	0.713	1.962	0.664
27	1.258	0.856	1.419	0.863	1.641	0.894	1.720	0.835
28	1.276	0.936	1.480	0.922	1.856	0.839	1.888	0.794
29	1.208	0.967	1.336	1.013	1.865	0.875	1.882	0.832

Table C.13: Descriptive Statistics of SIMITNORM Scores: Model BBB

BBB	Means				Standard Deviation				Skewness				Kurtosis			
Item	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
1	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.007	-0.236	-0.220	-0.220	-0.618	-0.483	-0.389	-0.495
2	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.019	-0.080	-0.239	-0.162	-0.648	-0.484	-0.017	-0.142
3	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.106	0.013	-0.245	0.004	-0.599	-0.558	-0.010	0.090
4	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.069	-0.073	-0.123	-0.022	-0.576	-0.514	-0.267	-0.248
8	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.373	-0.462	-0.382	-0.353	-0.483	-0.521	-0.129	0.218
9	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.584	-0.584	-0.437	-0.459	-0.391	-0.384	-0.464	-0.375
10	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.633	-0.562	-0.517	-0.490	-0.116	-0.322	0.057	0.199
14	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.625	0.556	-0.170	-0.076	-0.050	-0.137	-0.297	-0.388
15	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.299	0.275	-0.218	-0.157	-0.164	-0.225	-0.624	-0.578
16	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.270	0.219	-0.329	-0.185	-0.350	-0.498	-0.253	-0.137
17	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.016	-0.168	-0.018	0.131	-0.295	-0.425	-1.015	-0.985
18	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.046	-0.152	0.089	0.227	-0.396	-0.549	-0.904	-0.579
20	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.355	0.247	-0.092	-0.164	-0.240	-0.606	0.308	0.764
21	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.497	0.448	0.048	0.150	-0.110	-0.497	-0.069	-0.011
22	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.312	0.183	-0.074	-0.052	-0.430	-0.776	0.238	0.491
24	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.092	-0.238	-0.265	-0.092	-0.520	-0.456	-0.407	-0.208
25	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.279	-0.406	-0.249	0.121	-0.381	-0.451	-0.545	-0.912
27	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.097	0.099	-0.324	-0.222	-0.155	-0.453	-0.289	-0.438
28	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	-0.024	-0.099	-0.550	-0.453	-0.512	-0.667	0.051	-0.035
29	0.000	0.000	0.000	0.000	1.000	1.000	1.000	1.000	0.086	-0.100	-0.660	-0.543	-0.666	-0.795	0.088	-0.070

Table C.14: Estimated Parameters: Model DD

Model	88 Parameters			
DD	Eng	Afr	Nguni	Sotho
Gamma	0.890	0.721	0.147	0.280
Phi2	0.970	0.953	0.784	0.760
Item	Phi1	Phi1	Phi1	Phi1
1	0.768	1.223	3.181	2.560
2	0.770	1.058	3.157	2.473
3	0.681	0.969	3.188	2.498
4	0.647	1.017	2.961	2.283
8	1.360	1.555	3.132	2.541
9	1.444	1.631	3.104	2.486
10	1.516	1.756	3.288	2.629
14	0.022	0.364	2.929	2.259
15	0.304	0.682	2.986	2.357
16	0.376	0.693	3.148	2.476
17	1.067	1.457	3.300	2.591
18	1.156	1.506	3.182	2.453
20	0.325	0.695	2.875	2.169
21	0.170	0.506	2.746	2.084
22	0.287	0.706	2.857	2.146
24	1.100	1.451	3.317	2.595
25	1.314	1.686	3.502	2.802
27	0.827	1.247	3.175	2.578
28	0.842	1.308	3.373	2.736
29	0.768	1.160	3.383	2.732

Table C.15: Descriptive Statistics of SIMITNORM Scores: Model DD

Model DD	Means				Standard Deviation				Skewness				Kurtosis			
Item	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
1	0.000	0.000	0.000	0.000	1.080	1.042	1.007	1.048	-0.090	-0.273	-0.069	-0.123	-0.572	-0.446	-0.632	-0.657
2	0.000	0.000	0.000	0.000	1.089	1.023	0.966	1.004	-0.115	-0.118	-0.056	-0.042	-0.603	-0.466	-0.313	-0.306
3	0.000	0.000	0.000	0.000	1.080	1.041	0.963	0.920	0.007	-0.025	-0.063	0.125	-0.575	-0.549	-0.341	-0.093
4	0.000	0.000	0.000	0.000	1.076	1.030	1.024	1.030	-0.031	-0.111	0.063	0.102	-0.563	-0.501	-0.420	-0.322
8	0.000	0.000	0.000	0.000	0.950	1.011	1.035	1.002	-0.454	-0.492	-0.200	-0.213	-0.293	-0.459	-0.394	-0.041
9	0.000	0.000	0.000	0.000	0.983	1.003	1.138	1.168	-0.661	-0.614	-0.274	-0.343	-0.194	-0.315	-0.677	-0.549
10	0.000	0.000	0.000	0.000	0.929	0.904	1.012	1.029	-0.719	-0.590	-0.338	-0.355	0.154	-0.239	-0.341	-0.110
14	0.000	0.000	0.000	0.000	0.980	0.944	1.033	1.071	0.518	0.516	0.015	0.044	-0.218	-0.197	-0.414	-0.447
15	0.000	0.000	0.000	0.000	0.975	0.972	1.126	1.125	0.184	0.231	-0.057	-0.049	-0.256	-0.250	-0.765	-0.676
16	0.000	0.000	0.000	0.000	1.007	1.021	1.035	1.016	0.164	0.181	-0.158	-0.062	-0.417	-0.519	-0.511	-0.317
17	0.000	0.000	0.000	0.000	0.941	0.938	0.966	0.983	-0.120	-0.203	0.065	0.189	-0.135	-0.354	-1.174	-1.081
18	0.000	0.000	0.000	0.000	0.890	0.919	0.995	0.986	-0.043	-0.183	0.187	0.313	-0.244	-0.485	-1.054	-0.701
20	0.000	0.000	0.000	0.000	1.005	1.067	0.902	0.863	0.245	0.210	0.130	0.006	-0.325	-0.626	0.210	0.713
21	0.000	0.000	0.000	0.000	1.005	1.068	0.959	0.955	0.384	0.411	0.256	0.284	-0.247	-0.537	-0.068	0.021
22	0.000	0.000	0.000	0.000	1.060	1.137	0.924	0.895	0.207	0.148	0.150	0.109	-0.518	-0.793	0.135	0.463
24	0.000	0.000	0.000	0.000	0.970	0.951	0.965	0.952	-0.182	-0.271	-0.129	0.014	-0.395	-0.400	-0.722	-0.423
25	0.000	0.000	0.000	0.000	0.927	0.897	0.858	0.841	-0.369	-0.434	-0.146	0.174	-0.173	-0.377	-0.911	-1.093
27	0.000	0.000	0.000	0.000	0.937	0.924	1.043	1.038	-0.013	0.065	-0.156	-0.120	-0.090	-0.422	-0.583	-0.629
28	0.000	0.000	0.000	0.000	1.027	0.987	0.988	0.990	-0.121	-0.130	-0.385	-0.341	-0.453	-0.636	-0.388	-0.346
29	0.000	0.000	0.000	0.000	1.062	1.087	1.026	1.035	-0.005	-0.132	-0.494	-0.433	-0.630	-0.771	-0.354	-0.367

Table C.16: Estimated Parameters: Model DDD

Model DDD				
82 Parameters				
Gamma	0.504			
Phi2	0.858			
Item	English Phi1	Afrikaans Phi1	Nguni Phi1	Sotho Phi1
1	1.456	1.639	1.884	1.934
2	1.458	1.481	1.860	1.845
3	1.376	1.397	1.893	1.873
4	1.345	1.442	1.647	1.646
8	1.999	1.957	1.830	1.915
9	2.081	2.031	1.794	1.853
10	2.147	2.149	1.996	2.005
14	0.785	0.828	1.612	1.620
15	1.031	1.124	1.669	1.721
16	1.097	1.136	1.846	1.848
17	1.723	1.861	2.013	1.969
18	1.804	1.907	1.886	1.825
20	1.052	1.140	1.560	1.530
21	0.915	0.964	1.417	1.439
22	1.022	1.153	1.539	1.505
24	1.756	1.855	2.029	1.972
25	1.954	2.081	2.227	2.189
27	1.501	1.658	1.875	1.953
28	1.520	1.718	2.087	2.117
29	1.454	1.580	2.095	2.112

Table C.17: Descriptive Statistics of SIMITNORM Scores: Model DDD

DDD	Means				Standard Deviation				Skewness				Kurtosis			
Item	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
1	0.000	0.000	0.000	0.000	1.112	1.101	0.980	0.965	0.051	-0.192	-0.182	-0.186	-0.634	-0.526	-0.456	-0.554
2	0.000	0.000	0.000	0.000	1.120	1.077	0.942	0.926	0.023	-0.034	-0.192	-0.120	-0.663	-0.503	-0.098	-0.202
3	0.000	0.000	0.000	0.000	1.109	1.094	0.937	0.846	0.150	0.058	-0.198	0.047	-0.604	-0.565	-0.101	0.022
4	0.000	0.000	0.000	0.000	1.101	1.084	1.009	0.955	0.114	-0.028	-0.076	0.021	-0.577	-0.527	-0.311	-0.277
8	0.000	0.000	0.000	0.000	1.003	1.077	1.016	0.925	-0.338	-0.426	-0.336	-0.304	-0.560	-0.592	-0.200	0.124
9	0.000	0.000	0.000	0.000	1.036	1.069	1.123	1.084	-0.551	-0.549	-0.396	-0.418	-0.472	-0.464	-0.521	-0.438
10	0.000	0.000	0.000	0.000	0.982	0.970	0.987	0.949	-0.596	-0.529	-0.471	-0.443	-0.227	-0.417	-0.050	0.087
14	0.000	0.000	0.000	0.000	0.978	0.979	1.022	0.995	0.675	0.606	-0.124	-0.035	0.037	-0.059	-0.332	-0.412
15	0.000	0.000	0.000	0.000	0.982	1.016	1.112	1.043	0.352	0.327	-0.178	-0.120	-0.113	-0.189	-0.663	-0.615
16	0.000	0.000	0.000	0.000	1.019	1.067	1.013	0.937	0.319	0.266	-0.285	-0.142	-0.313	-0.468	-0.323	-0.203
17	0.000	0.000	0.000	0.000	0.986	1.002	0.928	0.899	0.029	-0.127	0.004	0.152	-0.358	-0.504	-1.059	-1.021
18	0.000	0.000	0.000	0.000	0.940	0.984	0.960	0.905	0.085	-0.116	0.114	0.258	-0.456	-0.621	-0.946	-0.625
20	0.000	0.000	0.000	0.000	1.016	1.115	0.892	0.805	0.406	0.292	-0.037	-0.106	-0.193	-0.579	0.275	0.739
21	0.000	0.000	0.000	0.000	1.010	1.113	0.953	0.889	0.550	0.494	0.099	0.196	-0.039	-0.446	-0.076	-0.004
22	0.000	0.000	0.000	0.000	1.068	1.187	0.915	0.834	0.361	0.226	-0.019	0.003	-0.382	-0.753	0.204	0.475
24	0.000	0.000	0.000	0.000	1.017	1.014	0.932	0.873	-0.053	-0.199	-0.230	-0.055	-0.570	-0.520	-0.493	-0.287
25	0.000	0.000	0.000	0.000	0.979	0.962	0.819	0.764	-0.240	-0.373	-0.221	0.140	-0.465	-0.535	-0.646	-0.980
27	0.000	0.000	0.000	0.000	0.971	0.983	1.019	0.954	0.146	0.138	-0.281	-0.186	-0.176	-0.487	-0.369	-0.508
28	0.000	0.000	0.000	0.000	1.061	1.050	0.958	0.908	0.020	-0.063	-0.507	-0.413	-0.533	-0.702	-0.067	-0.148
29	0.000	0.000	0.000	0.000	1.097	1.149	0.999	0.951	0.127	-0.062	-0.618	-0.504	-0.677	-0.822	-0.030	-0.177

Table C.18: Fit Measures of SIMITNORM Functions, Models A, B, D

	AIC			F			RMSE			R-SQ		
Item	Model A	Model B	Model D	Model A	Model B	Model D	Model A	Model B	Model D	Model A	Model B	Model D
1	2.624	2.625	2.632	0.391	0.392	0.396	0.517	0.518	0.523	0.767	0.766	0.763
2	2.648	2.649	2.661	0.402	0.404	0.410	0.531	0.534	0.542	0.754	0.752	0.744
3	2.629	2.632	2.655	0.393	0.395	0.407	0.522	0.526	0.541	0.765	0.762	0.747
4	2.809	2.809	2.814	0.483	0.484	0.487	0.530	0.531	0.535	0.772	0.771	0.768
8	2.258	2.257	2.261	0.208	0.208	0.210	0.523	0.523	0.529	0.711	0.711	0.705
9	2.266	2.265	2.264	0.212	0.212	0.212	0.526	0.526	0.526	0.756	0.756	0.756
10	1.977	1.975	1.975	0.067	0.067	0.067	0.503	0.503	0.503	0.688	0.688	0.687
14	2.830	2.852	2.852	0.494	0.505	0.506	0.554	0.565	0.564	0.796	0.788	0.789
15	2.863	2.869	2.870	0.510	0.514	0.515	0.527	0.526	0.530	0.802	0.802	0.799
16	2.737	2.748	2.753	0.447	0.453	0.456	0.532	0.537	0.538	0.793	0.789	0.788
17	2.357	2.360	2.363	0.257	0.259	0.261	0.500	0.502	0.503	0.682	0.680	0.678
18	2.392	2.394	2.395	0.275	0.276	0.277	0.503	0.505	0.506	0.670	0.667	0.666
20	2.787	2.799	2.831	0.472	0.479	0.495	0.537	0.550	0.573	0.768	0.757	0.737
21	2.896	2.903	2.923	0.527	0.531	0.541	0.522	0.531	0.543	0.799	0.792	0.783
22	2.851	2.859	2.898	0.504	0.509	0.529	0.531	0.542	0.571	0.790	0.782	0.758
24	2.240	2.295	2.300	0.199	0.226	0.230	0.501	0.497	0.502	0.699	0.704	0.699
25	1.886	1.891	1.906	0.022	0.025	0.033	0.493	0.495	0.507	0.598	0.595	0.575
27	2.486	2.489	2.489	0.321	0.324	0.324	0.501	0.502	0.504	0.743	0.742	0.741
28	2.327	2.332	2.335	0.242	0.245	0.248	0.511	0.514	0.516	0.746	0.743	0.742
29	2.412	2.424	2.430	0.284	0.291	0.295	0.504	0.510	0.513	0.791	0.786	0.784

Table C.19: Fit Measures of SIMITNORM Functions, Models BB, DD, BBB, DDD

	English		Afrikaans		Nguni		Sotho			All Items and All Groups		
Model	BB	DD	BB	DD	BB	DD	BB	DD		MODEL	BBB	DDD
F	0.462	0.465	0.448	0.453	0.290	0.295	0.246	0.252		F	0.365	0.372
AIC	2.765	2.770	2.738	2.745	2.422	2.429	2.334	2.345		AIC	2.572	2.584
R-SQ	0.811	0.810	0.809	0.806	0.707	0.699	0.651	0.639		R-SQ	0.768	0.760
RMSE	0.516	0.518	0.523	0.527	0.530	0.537	0.542	0.552		RMSE	0.530	0.539

APPENDIX D : SIMITNORM Figures

Figure D1: Cumulative Density Functions of 20 VRP items

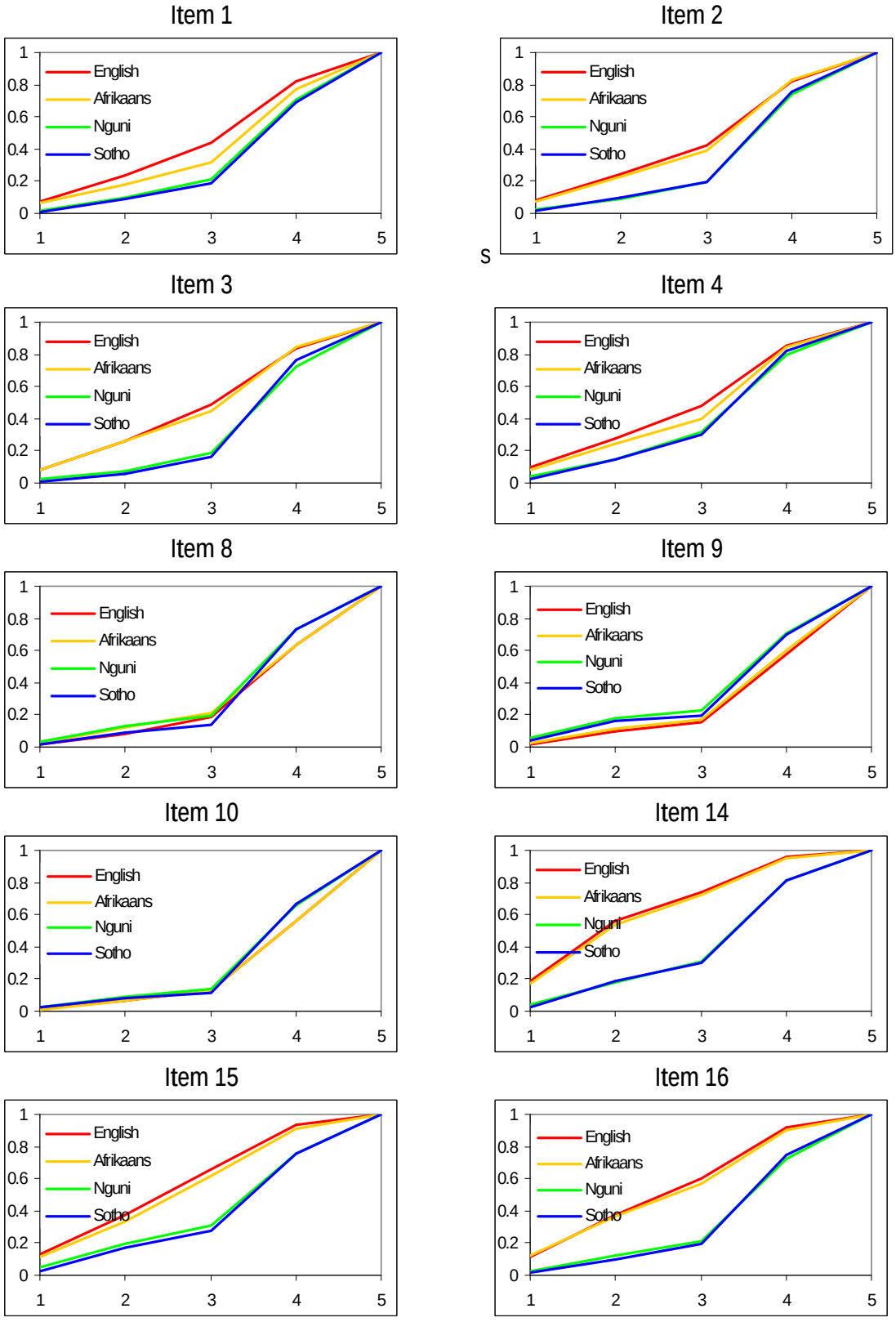
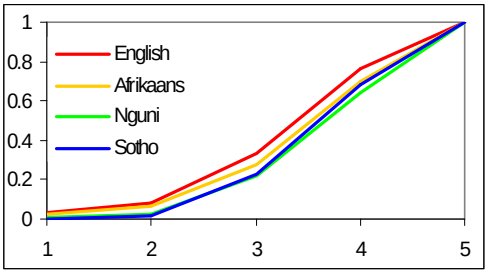
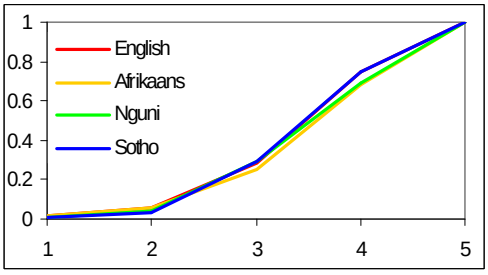


Figure D1 (continued): Cumulative Density Functions of 20 VRP items

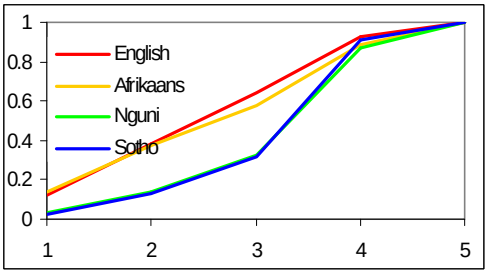
Item 17



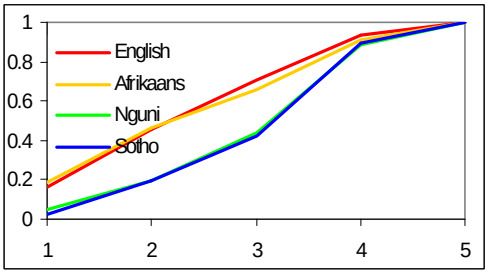
Item 18



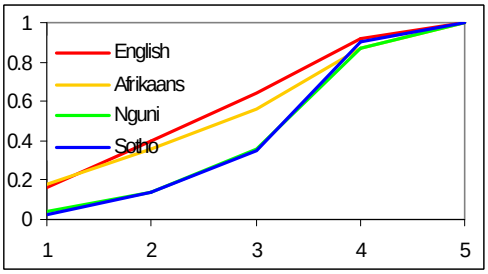
Item 20



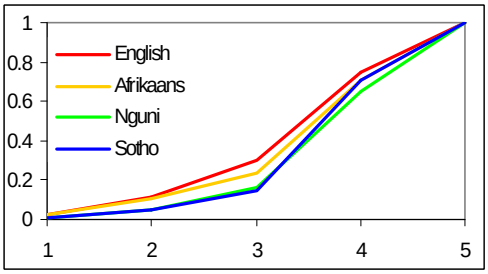
Item 21



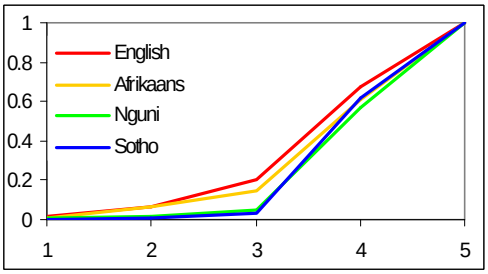
Item 22



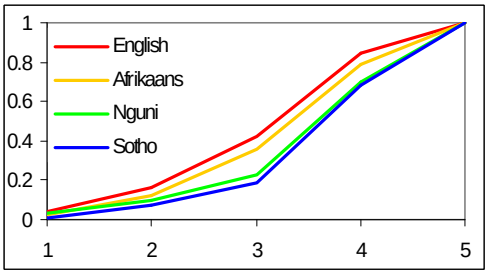
Item 24



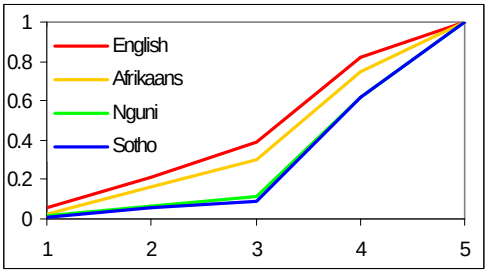
Item 25



Item 27



Item 28



Item 29

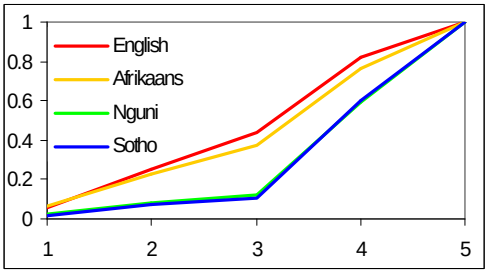


Figure D2: Q-Q plot of 20 VRP items over four cultural groups

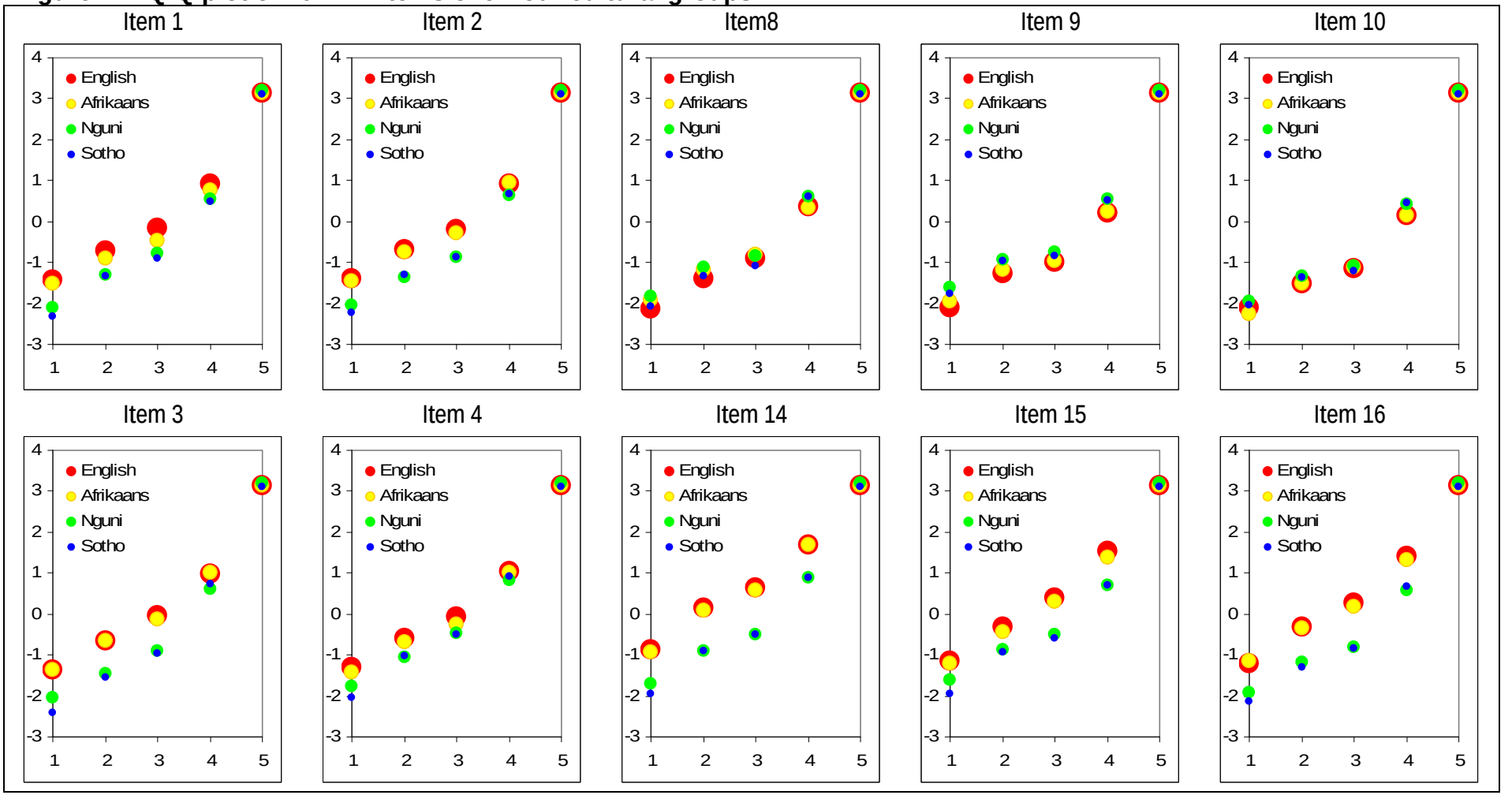


Figure D2 (continued): Q-Q plot of 20 VRP items over four cultural groups

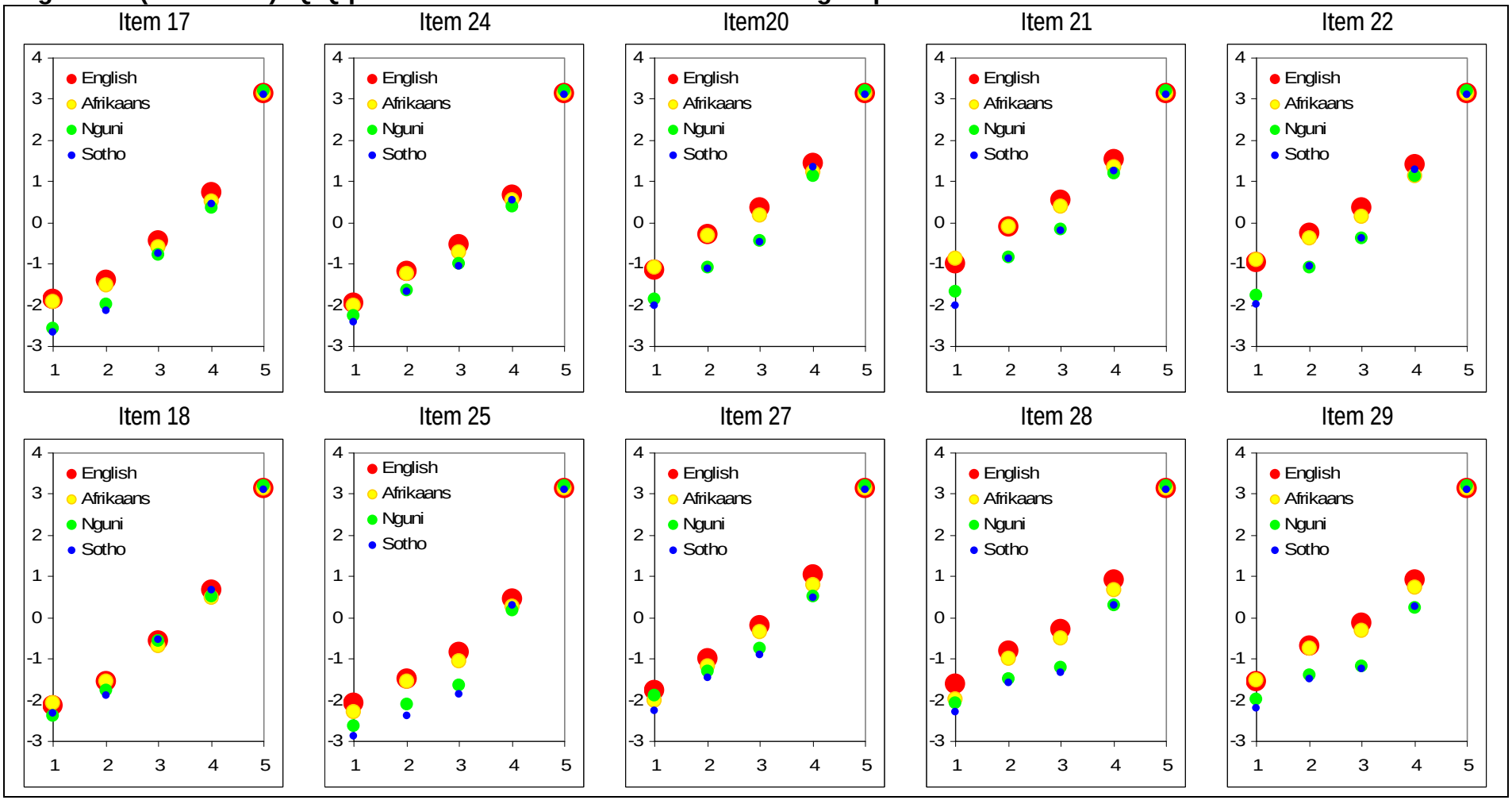


Figure D3.1: SIMITNORM functions for Item 1 over four cultural groups

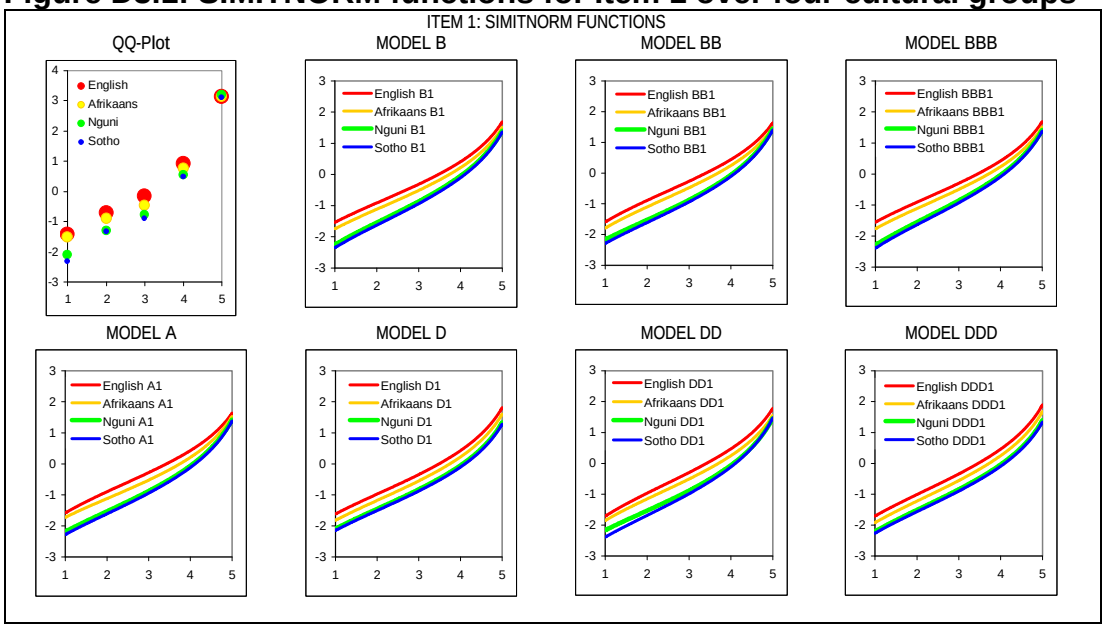


Figure D3.2: SIMITNORM functions for Item 2 over four cultural groups

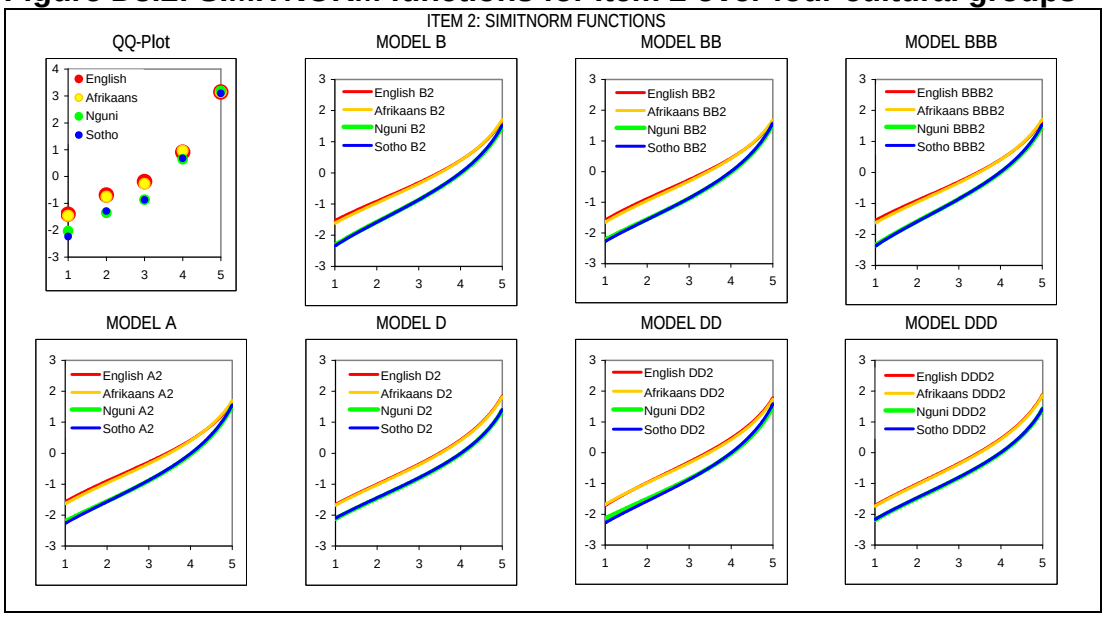


Figure D3.3: SIMITNORM functions for Item 3 over four cultural groups

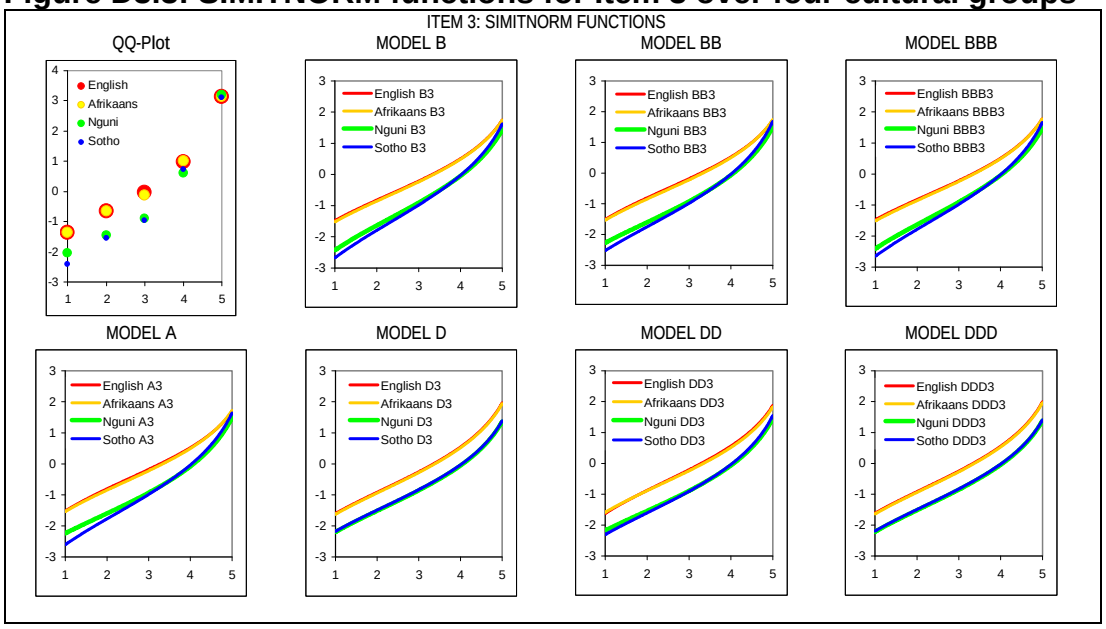


Figure D3.4: SIMITNORM functions for Item 4 over four cultural groups

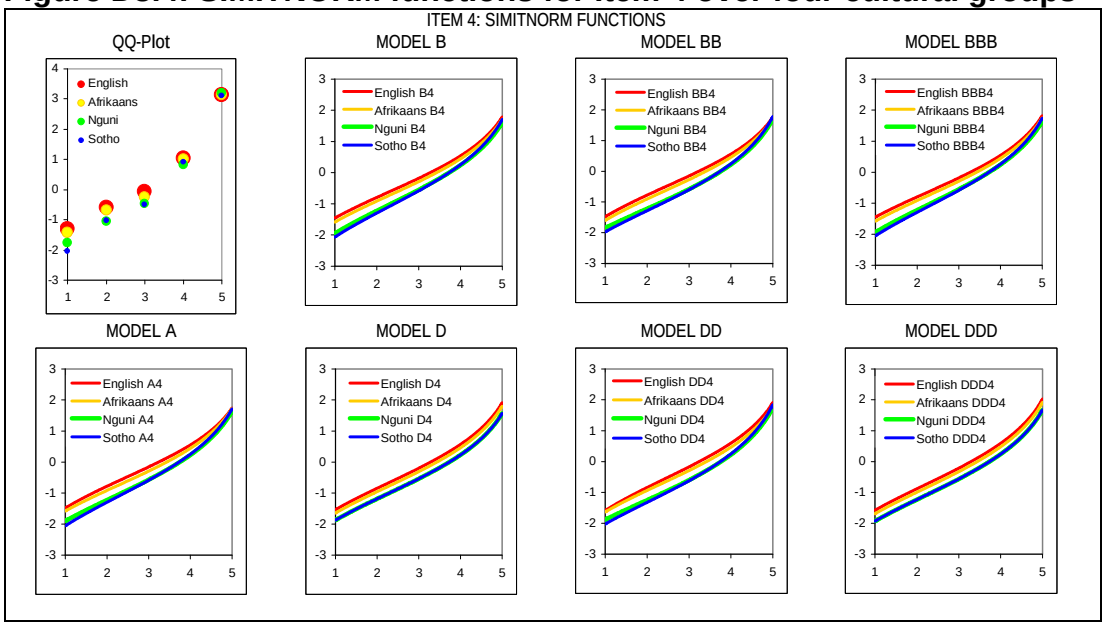


Figure D3.5: SIMITNORM functions for Item 8 over four cultural groups

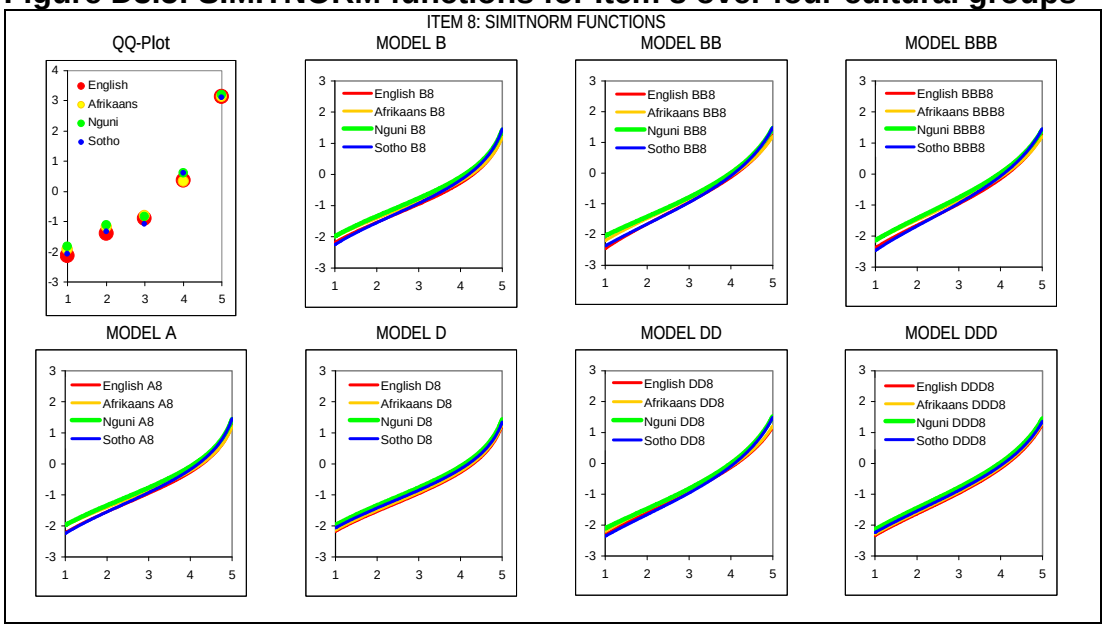


Figure D3.6: SIMITNORM functions for Item 9 over four cultural groups

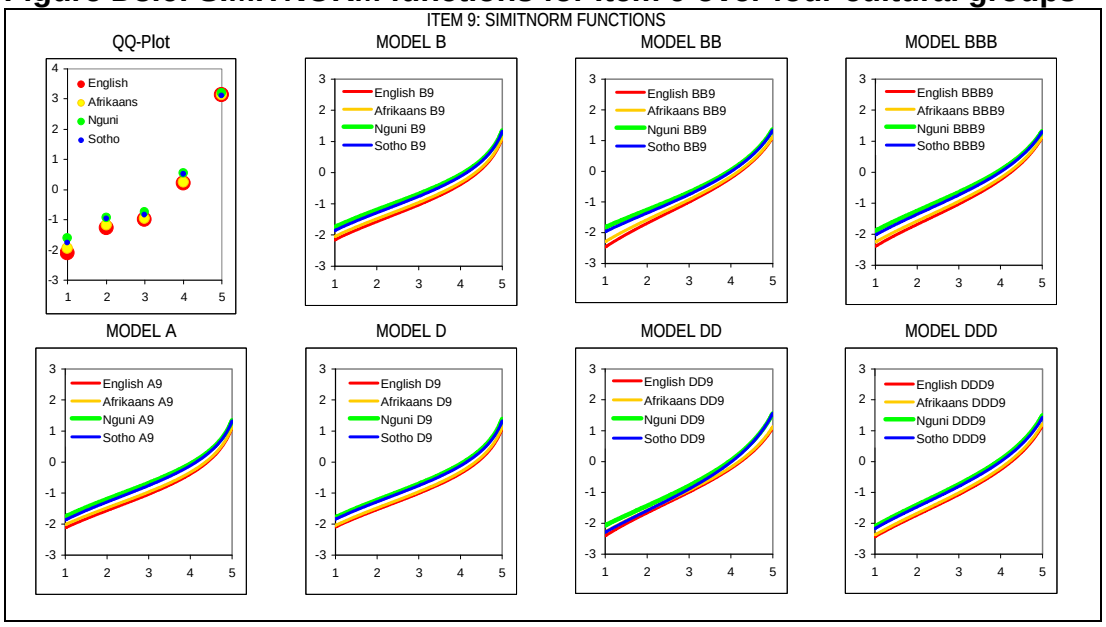


Figure D3.7: SIMITNORM functions for Item 10 over four cultural groups

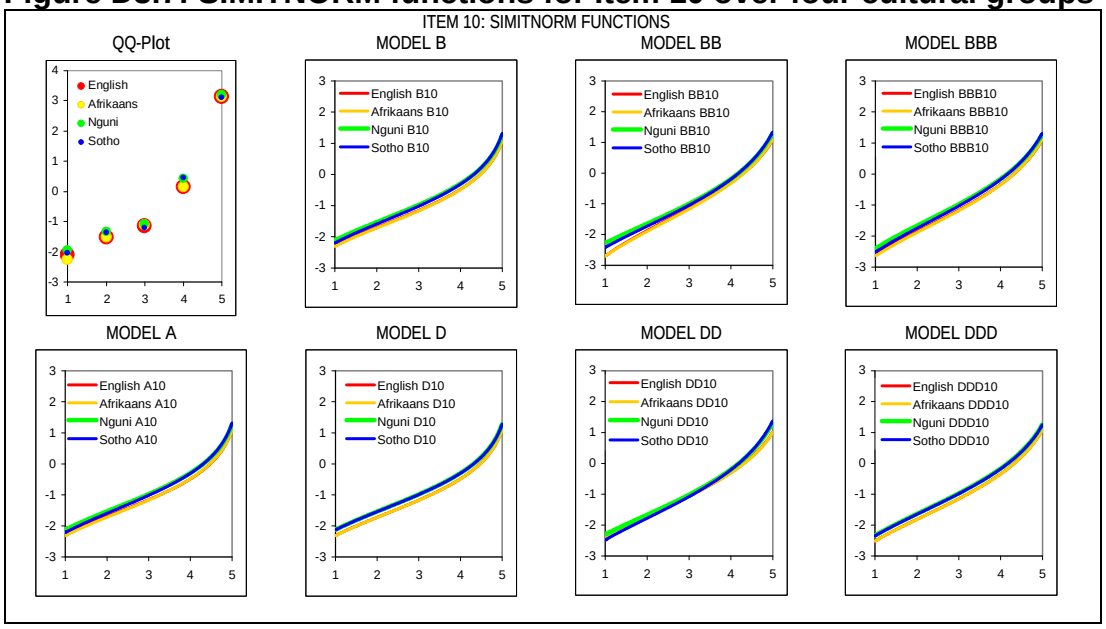


Figure D3.8: SIMITNORM functions for Item 14 over four cultural groups

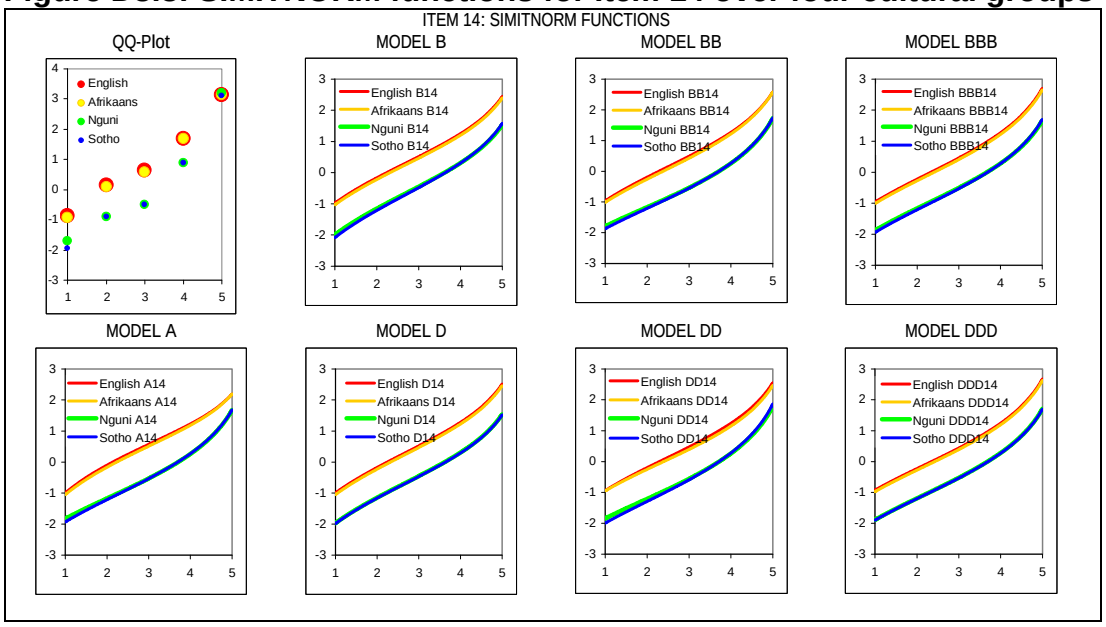


Figure D3.9: SIMITNORM functions for Item 15 over four cultural groups

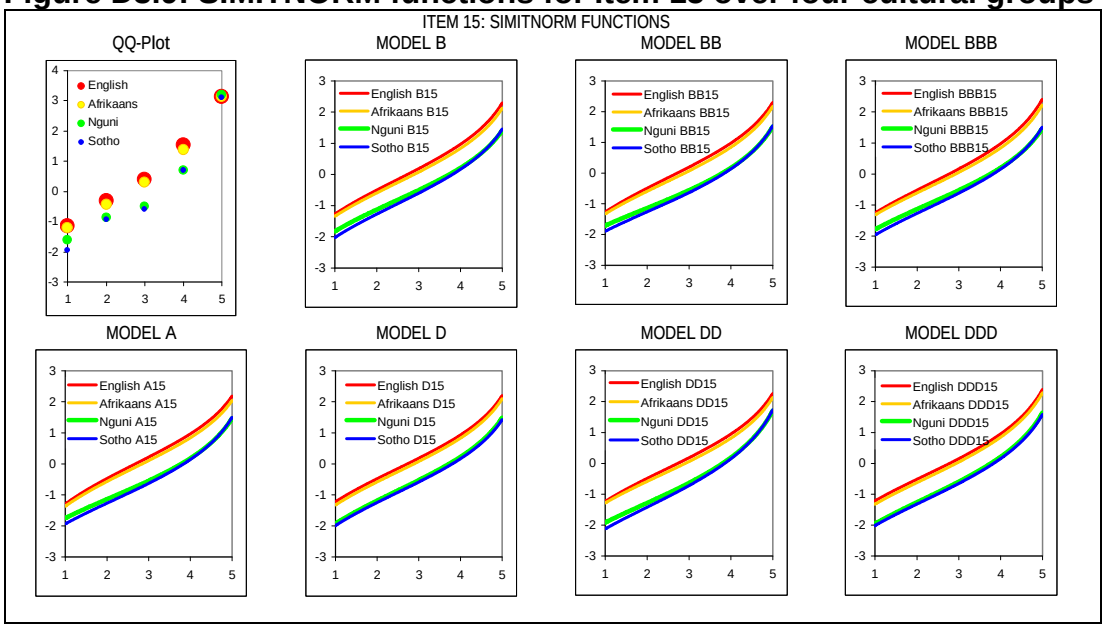


Figure D3.10: SIMITNORM functions for Item 16 over four cultural groups

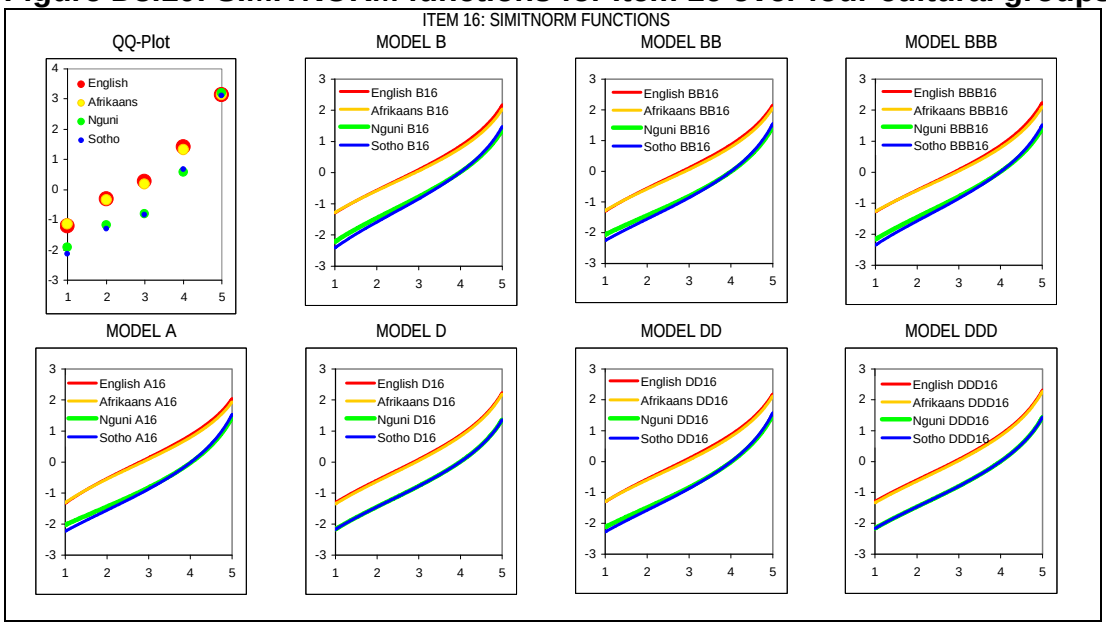


Figure D3.11: SIMITNORM functions for Item 17 over four cultural groups

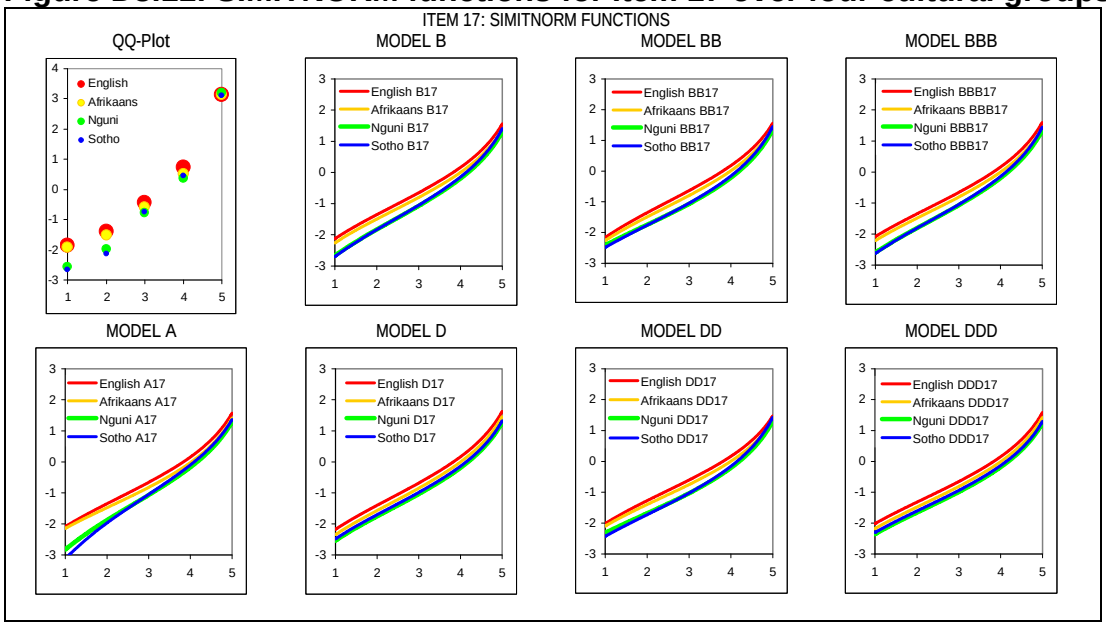


Figure D3.12: SIMITNORM functions for Item 18 over four cultural groups

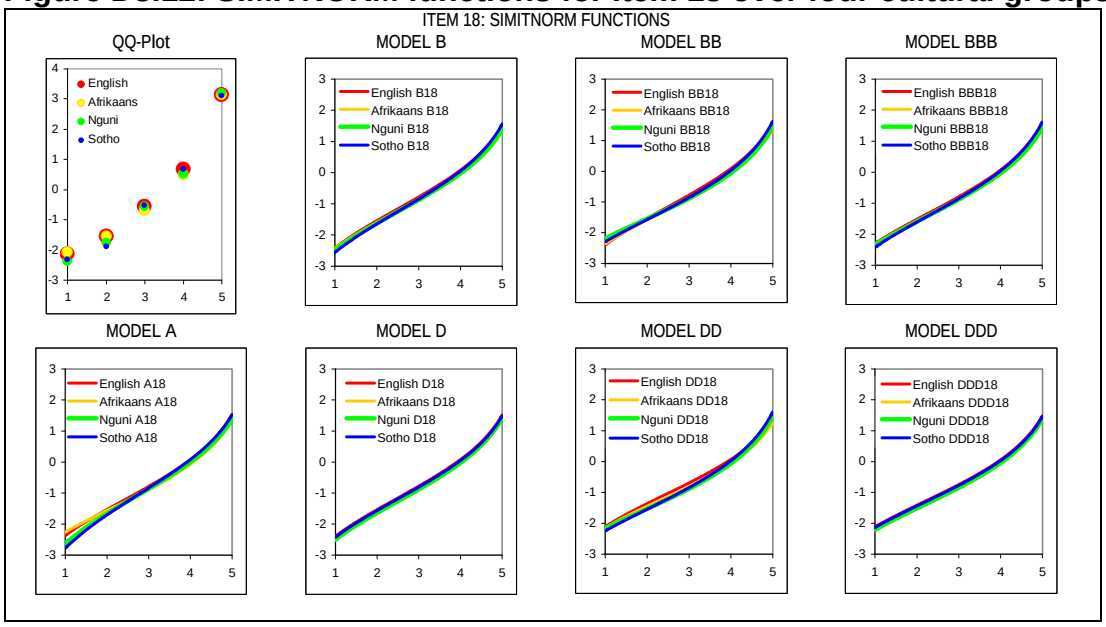


Figure D3.13: SIMITNORM functions for Item 20 over four cultural groups

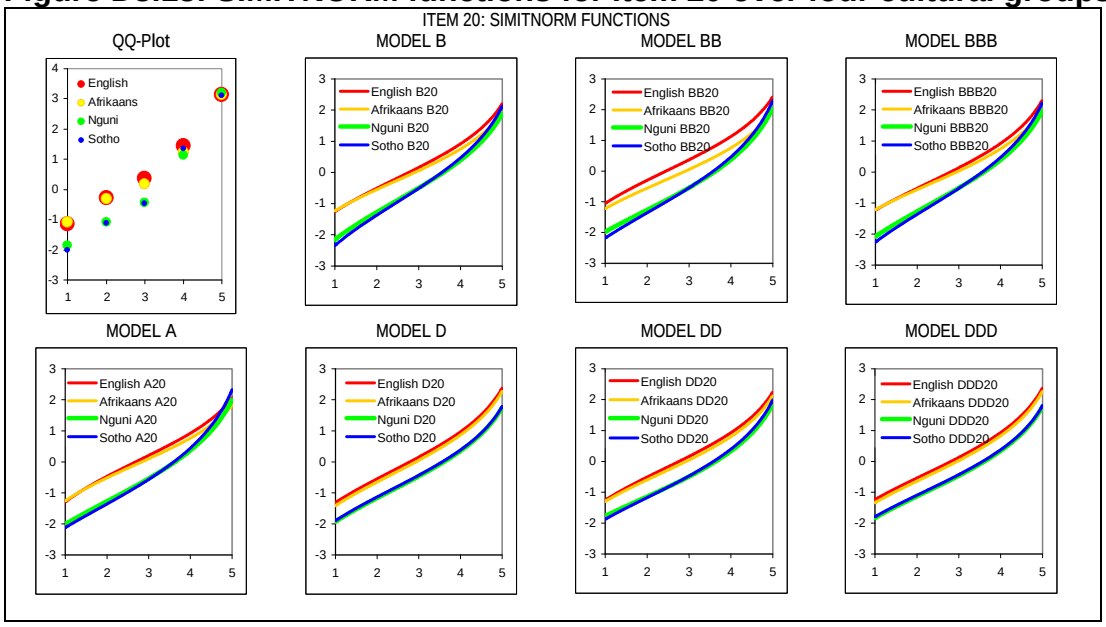


Figure D3.14: SIMITNORM functions for Item 21 over four cultural groups

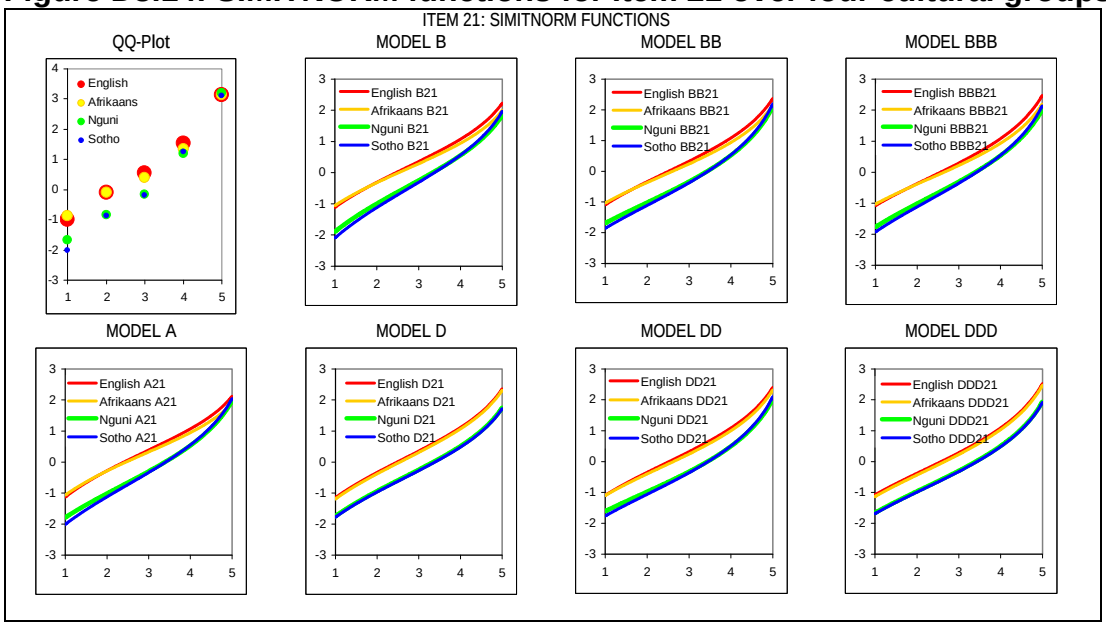


Figure D3.15: SIMITNORM functions for Item 22 over four cultural groups

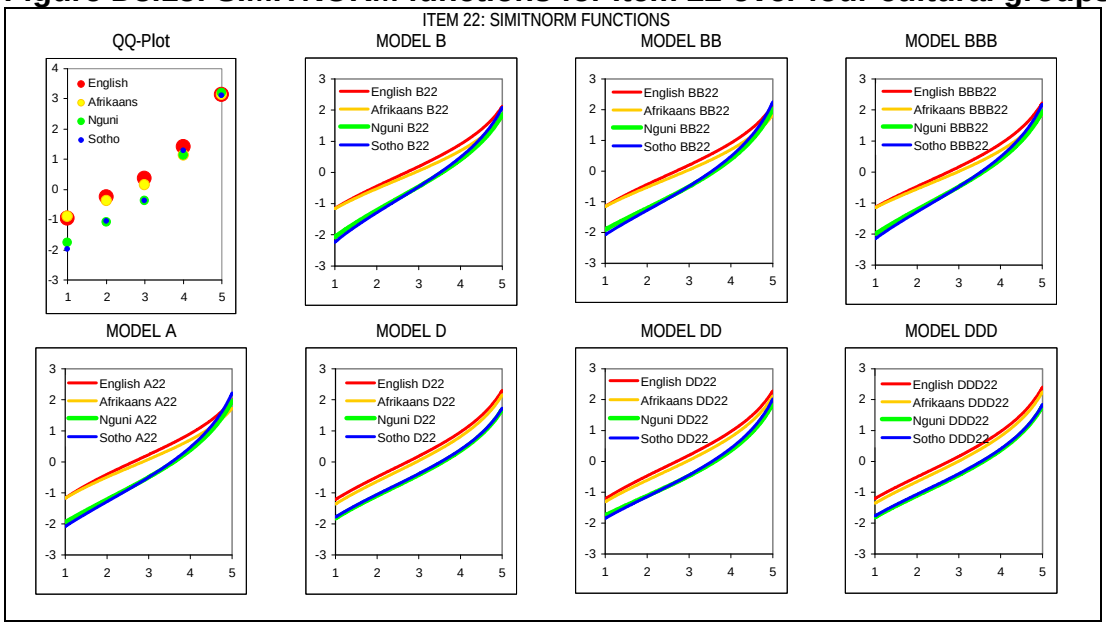


Figure D3.16: SIMITNORM functions for Item 24 over four cultural groups

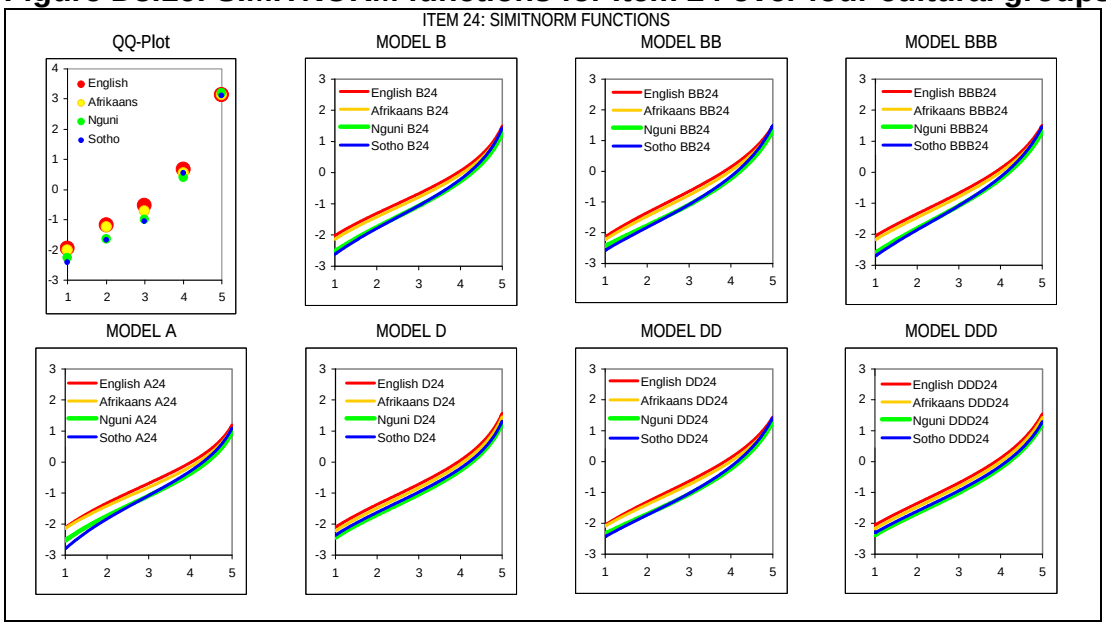


Figure D3.17: SIMITNORM functions for Item 25 over four cultural groups

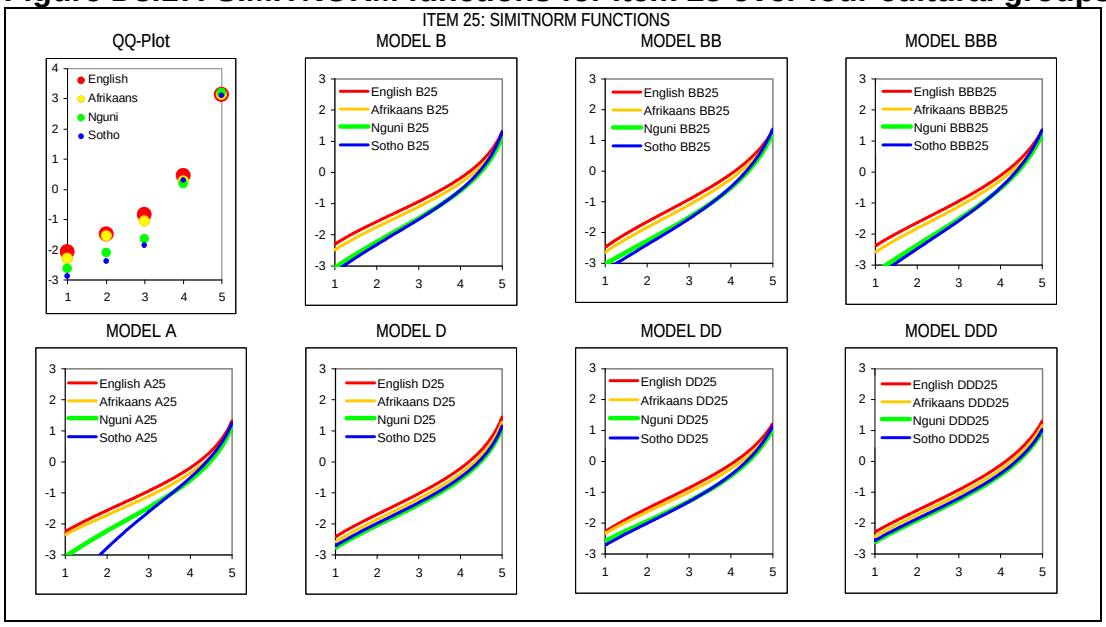


Figure D3.18: SIMITNORM functions for Item 27 over four cultural groups

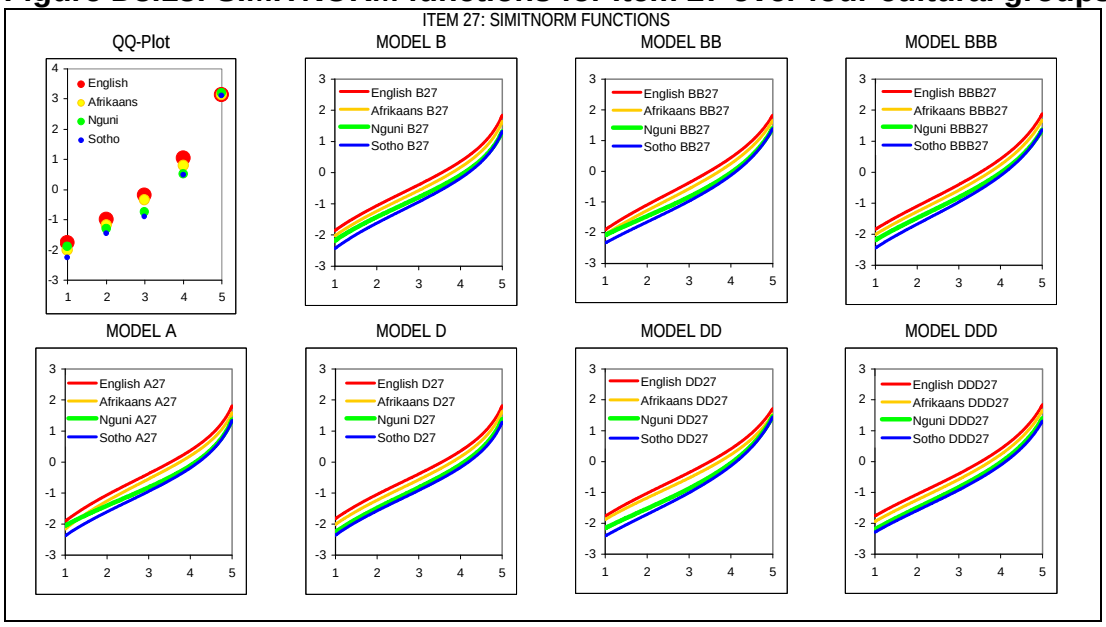


Figure D3.19: SIMITNORM functions for Item 28 over four cultural groups

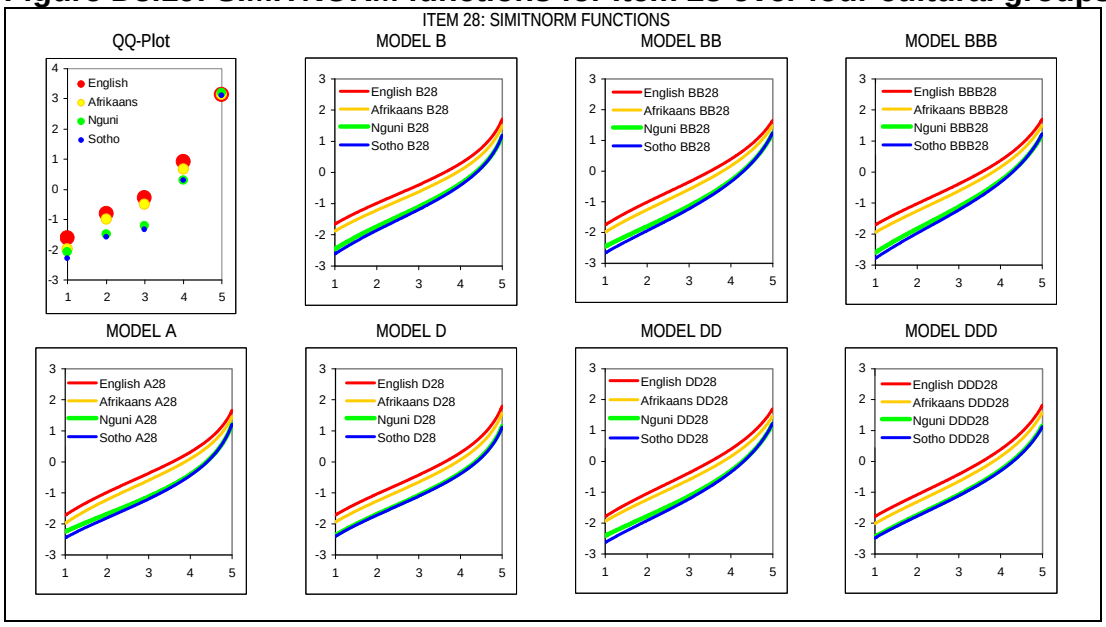


Figure D3.20: SIMITNORM functions for Item 29 over four cultural groups

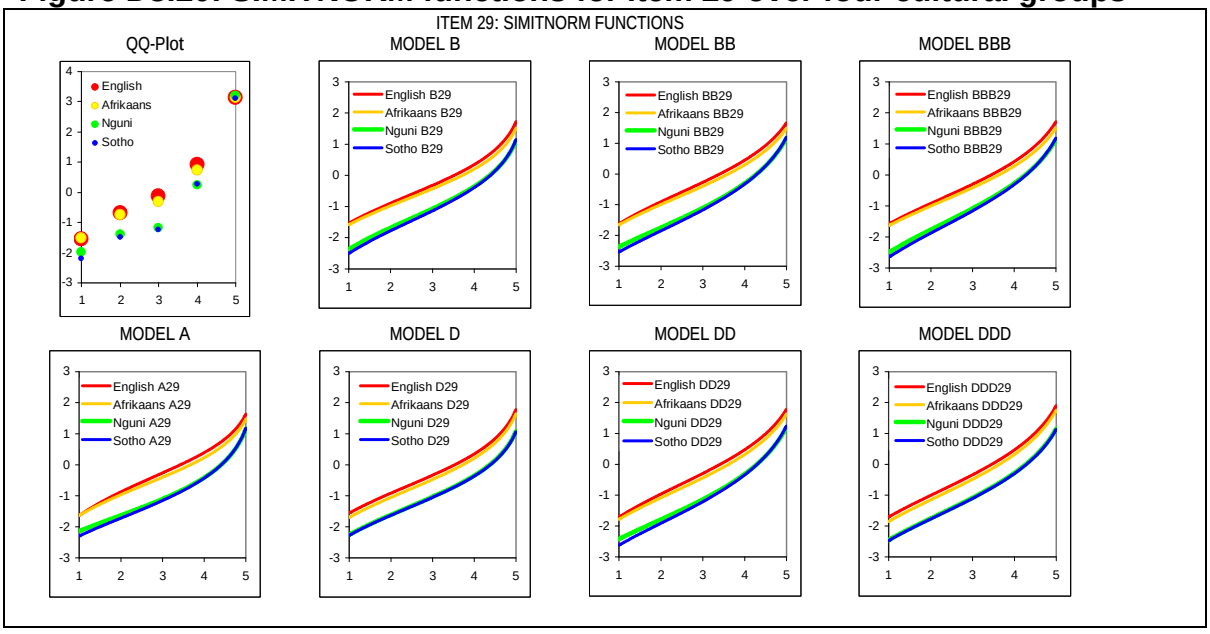


Figure D4.1: Seven SIMITNORM model score distributions for Item 1 by cultural group

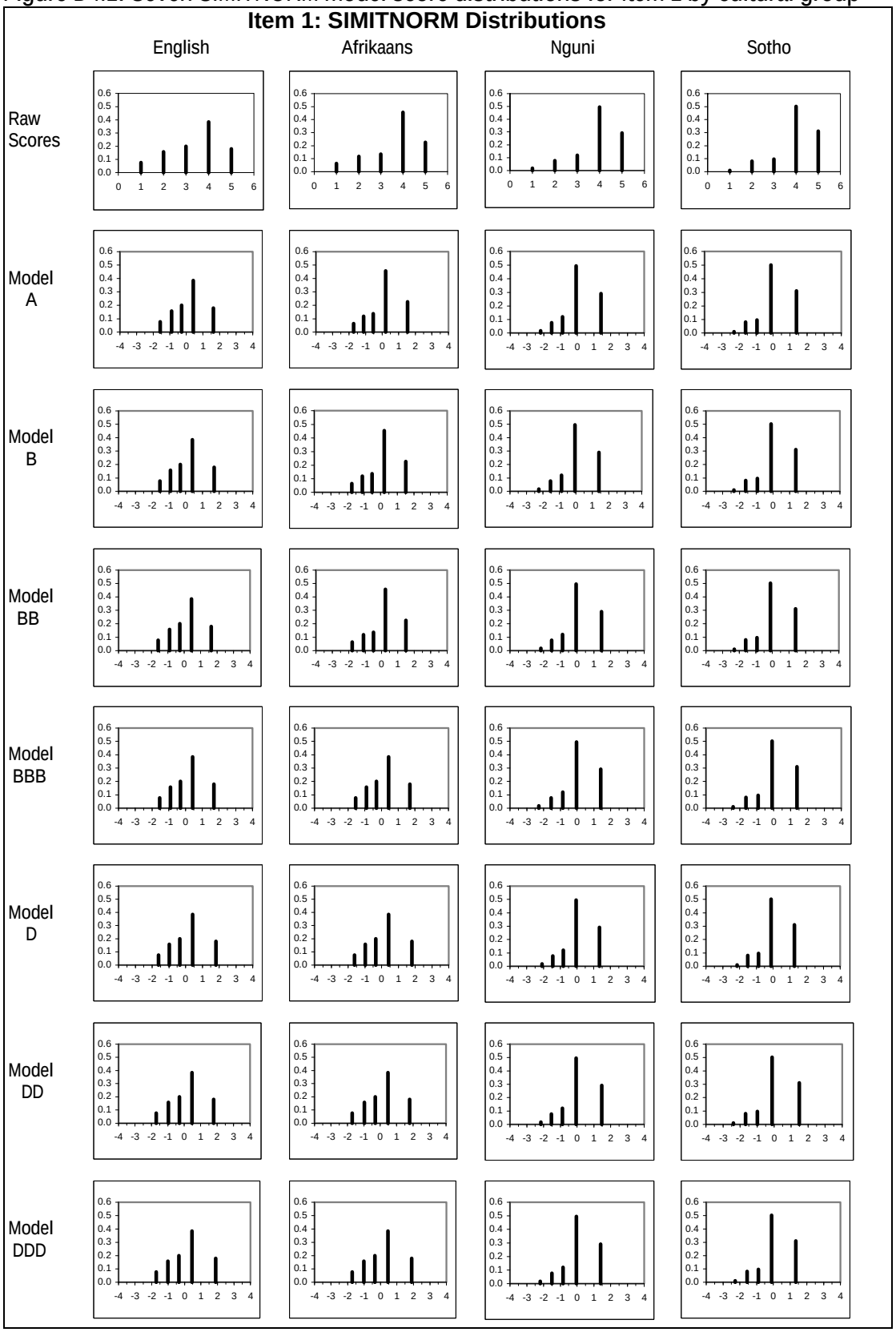


Figure D4.2: Seven SIMITNORM model score distributions for Item 2 by cultural group

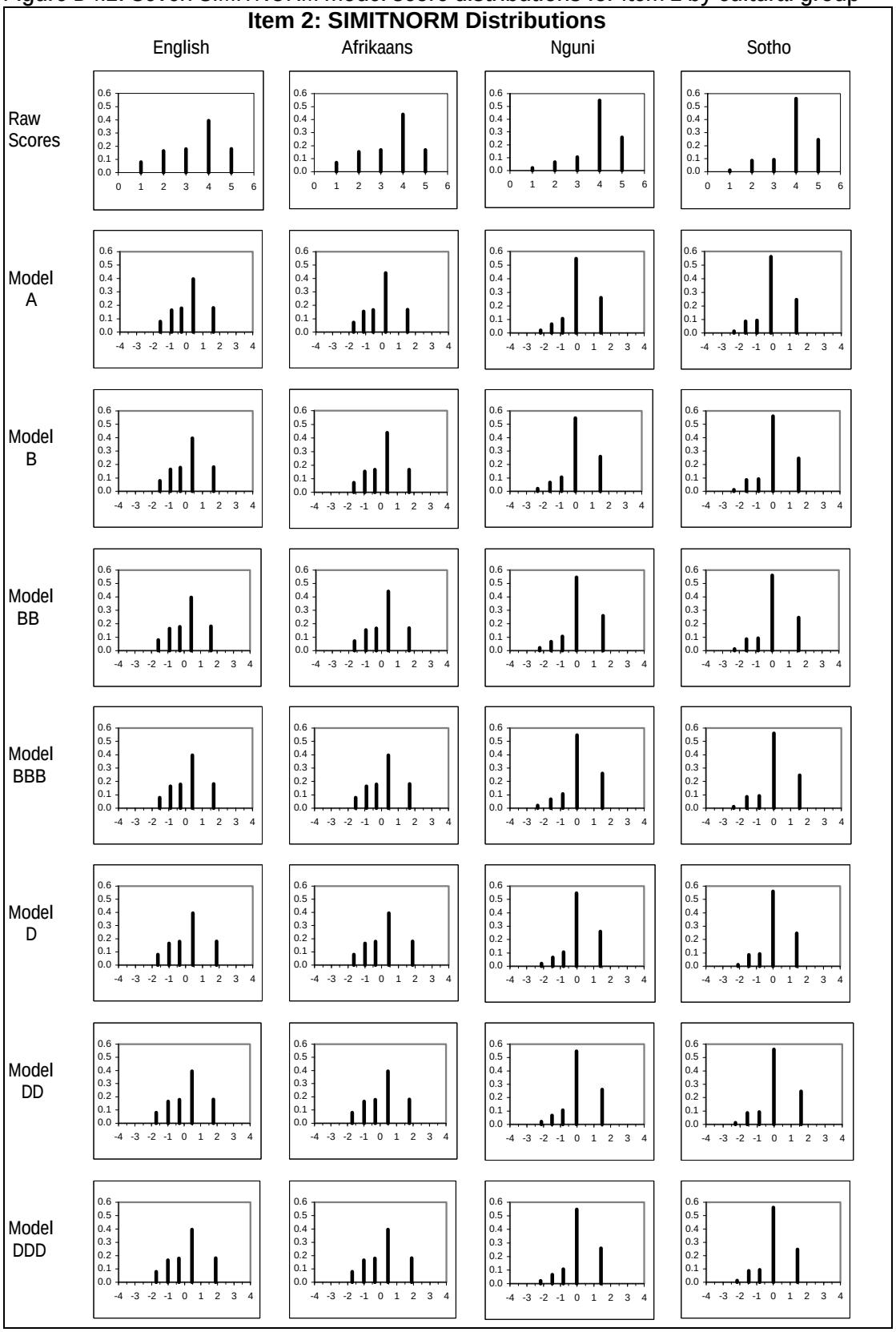


Figure D4.3: Seven SIMITNORM model score distributions for Item 3 by cultural group

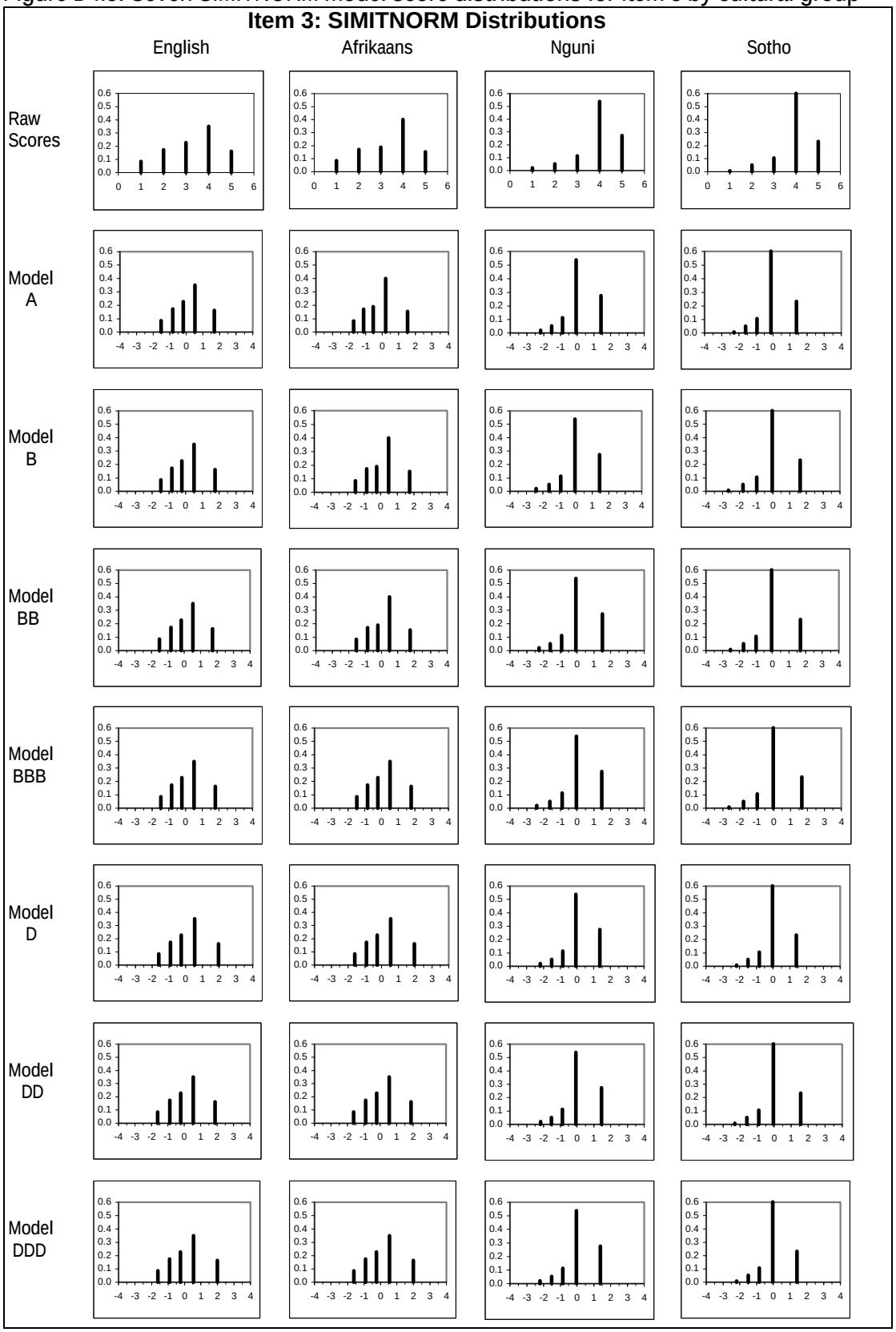


Figure D4.4: Seven SIMITNORM model score distributions for Item 4 by cultural group

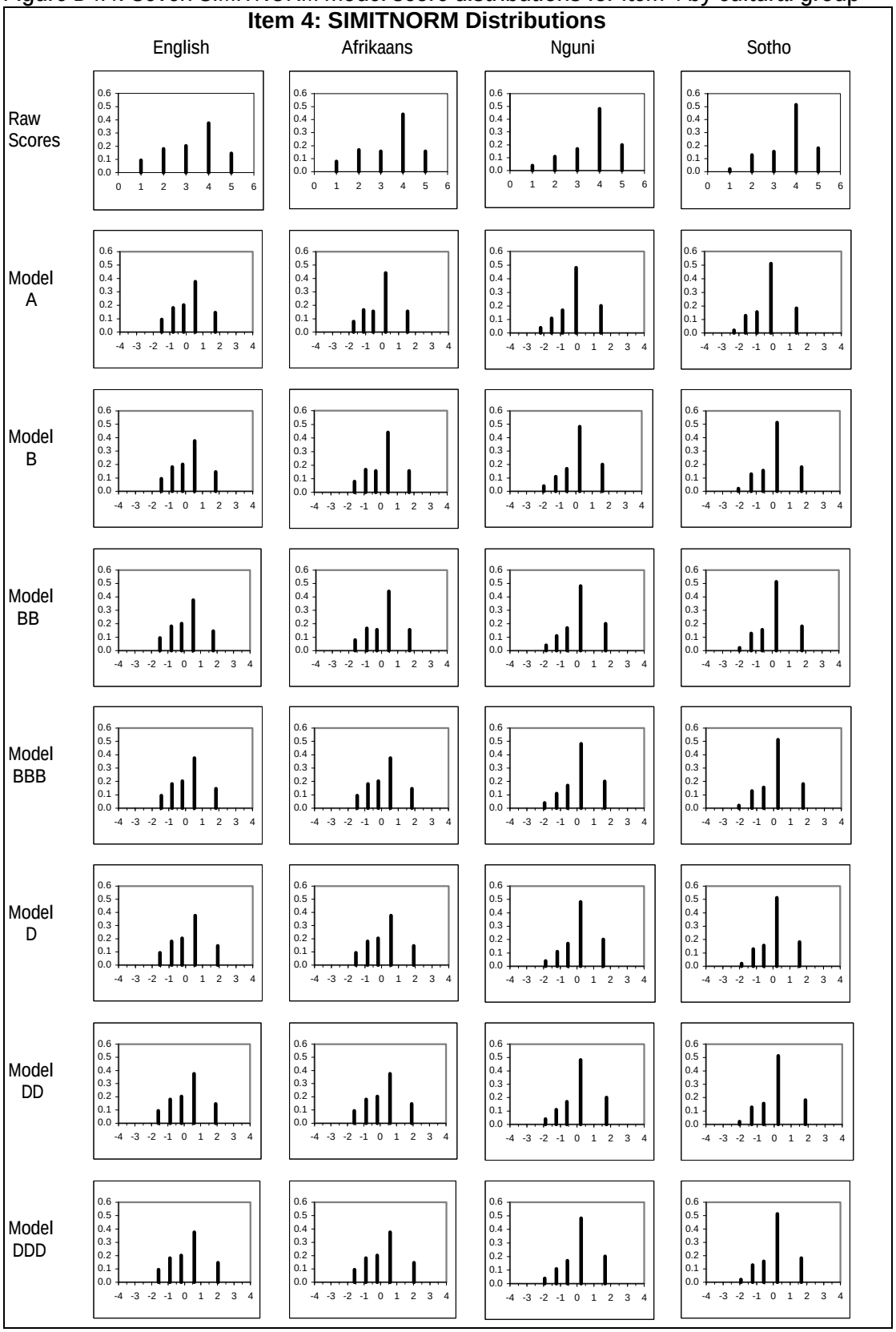


Figure D4.5: Seven SIMITNORM model score distributions for Item 8 by cultural group

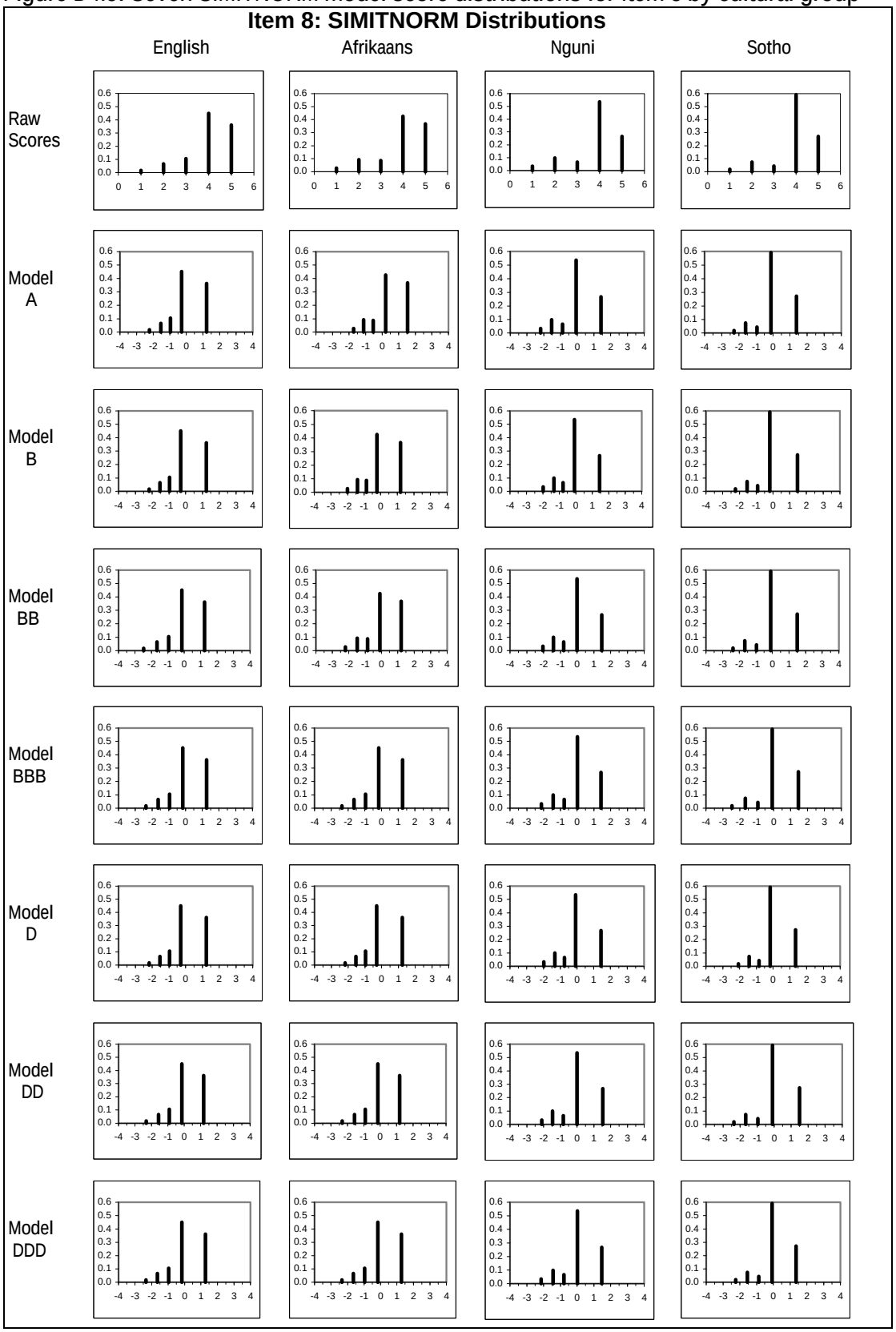


Figure D4.6: Seven SIMITNORM model score distributions for Item 9 by cultural group

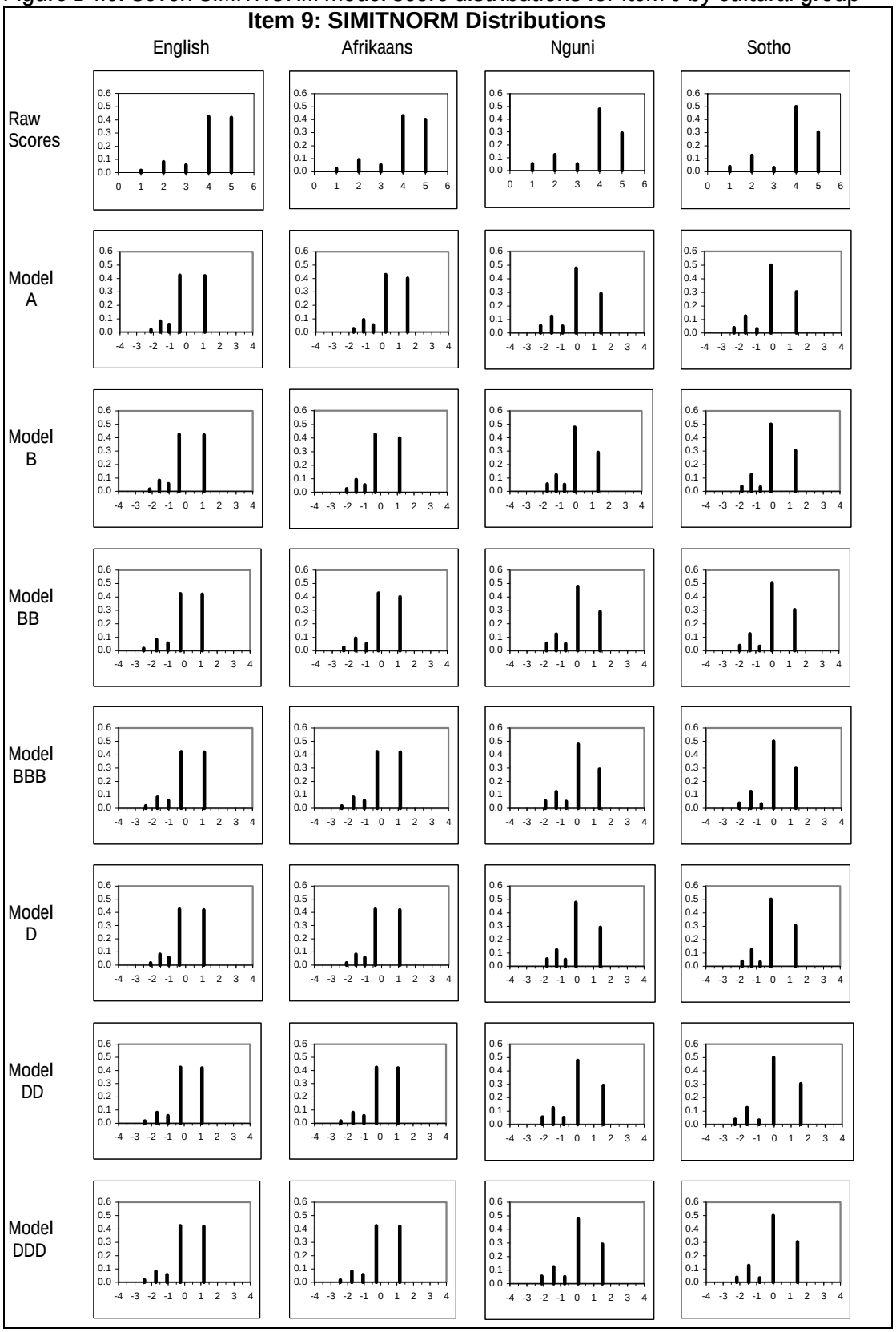


Figure D4.7: Seven SIMITNORM model score distributions for Item 10 by cultural group

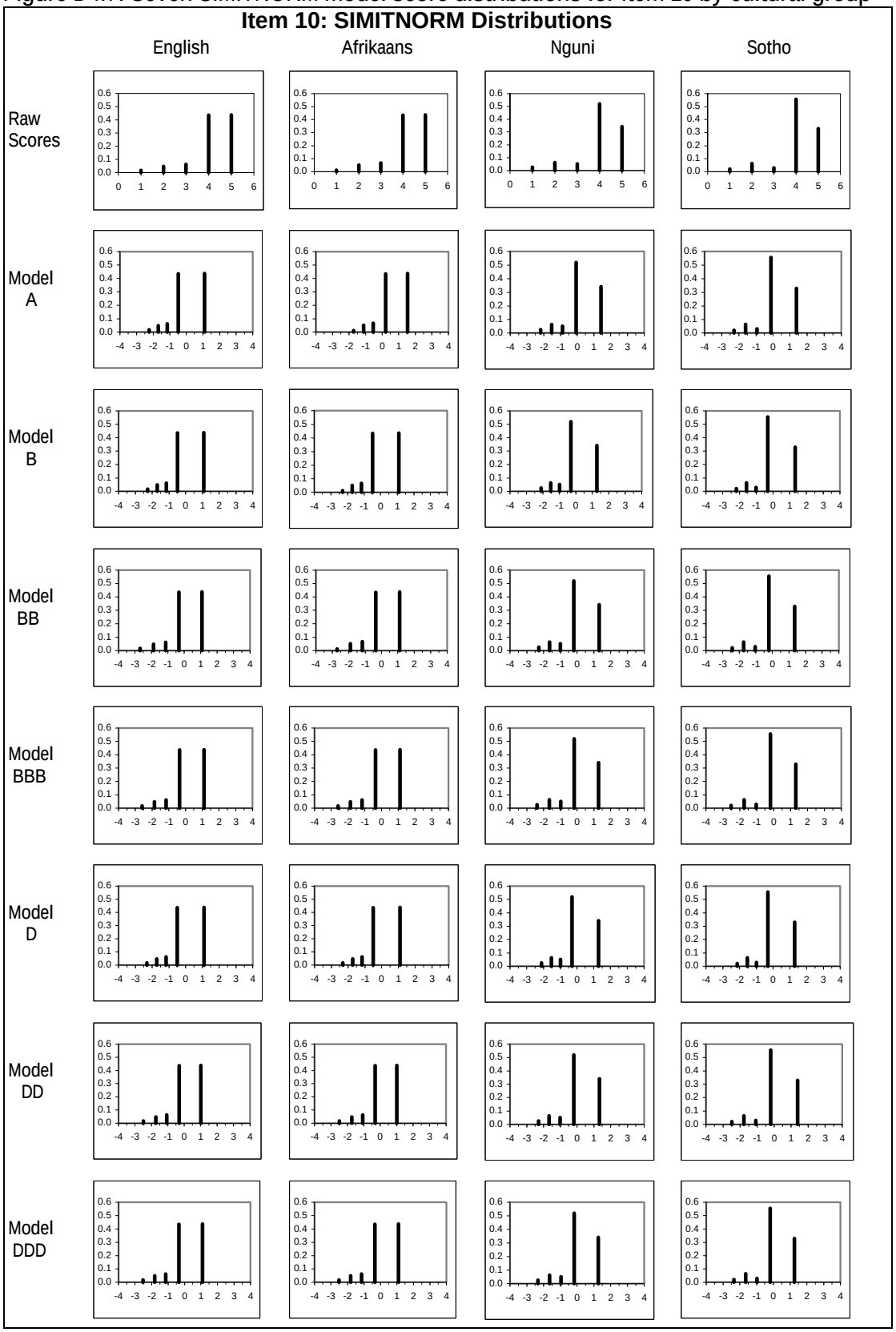


Figure D4.8: Seven SIMITNORM model score distributions for Item 14 by cultural group

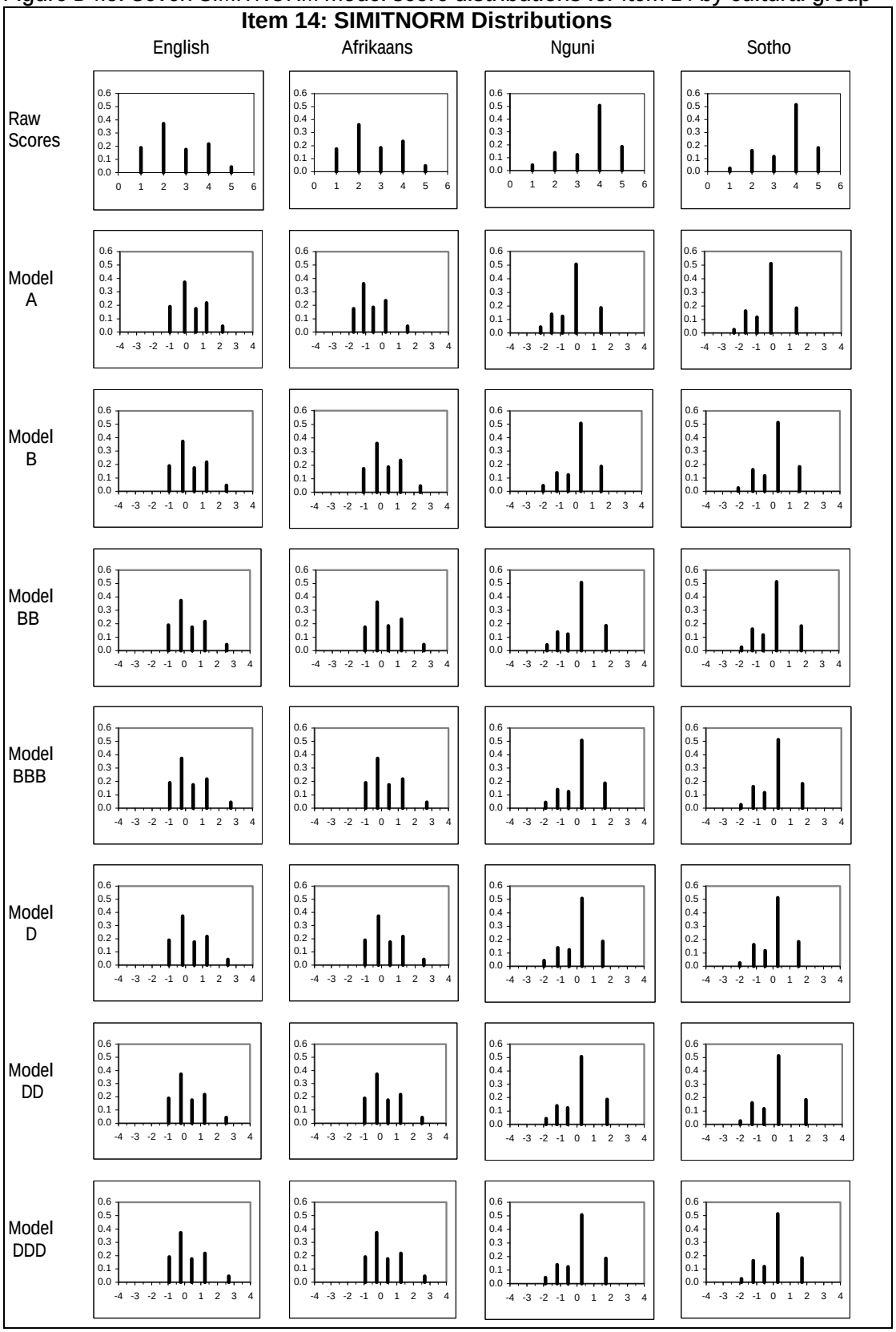


Figure D4.9: Seven SIMITNORM model score distributions for Item 15 by CULTURAL GROUP

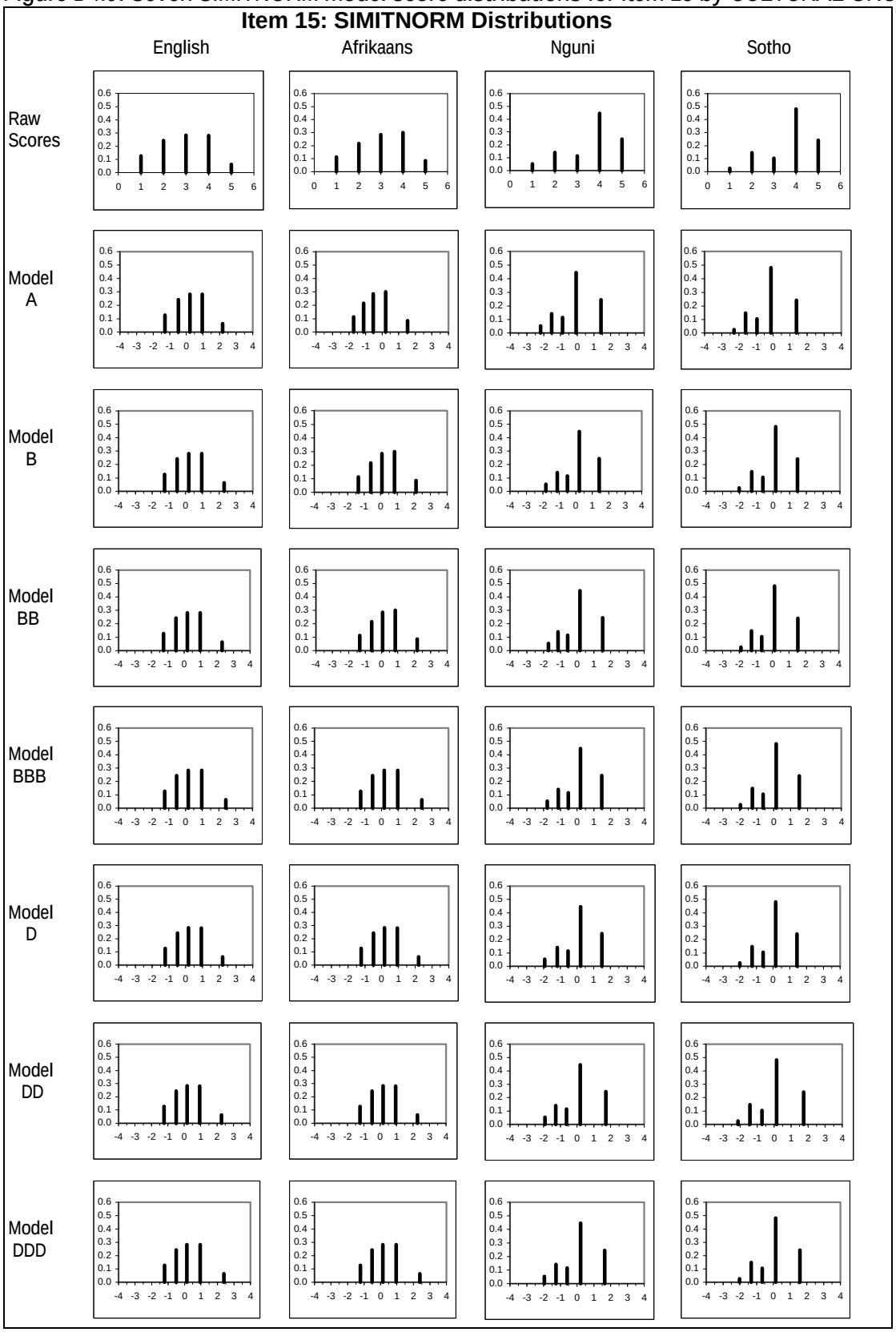


Figure D4.10: Seven SIMITNORM model score distributions for Item 16 by cultural group

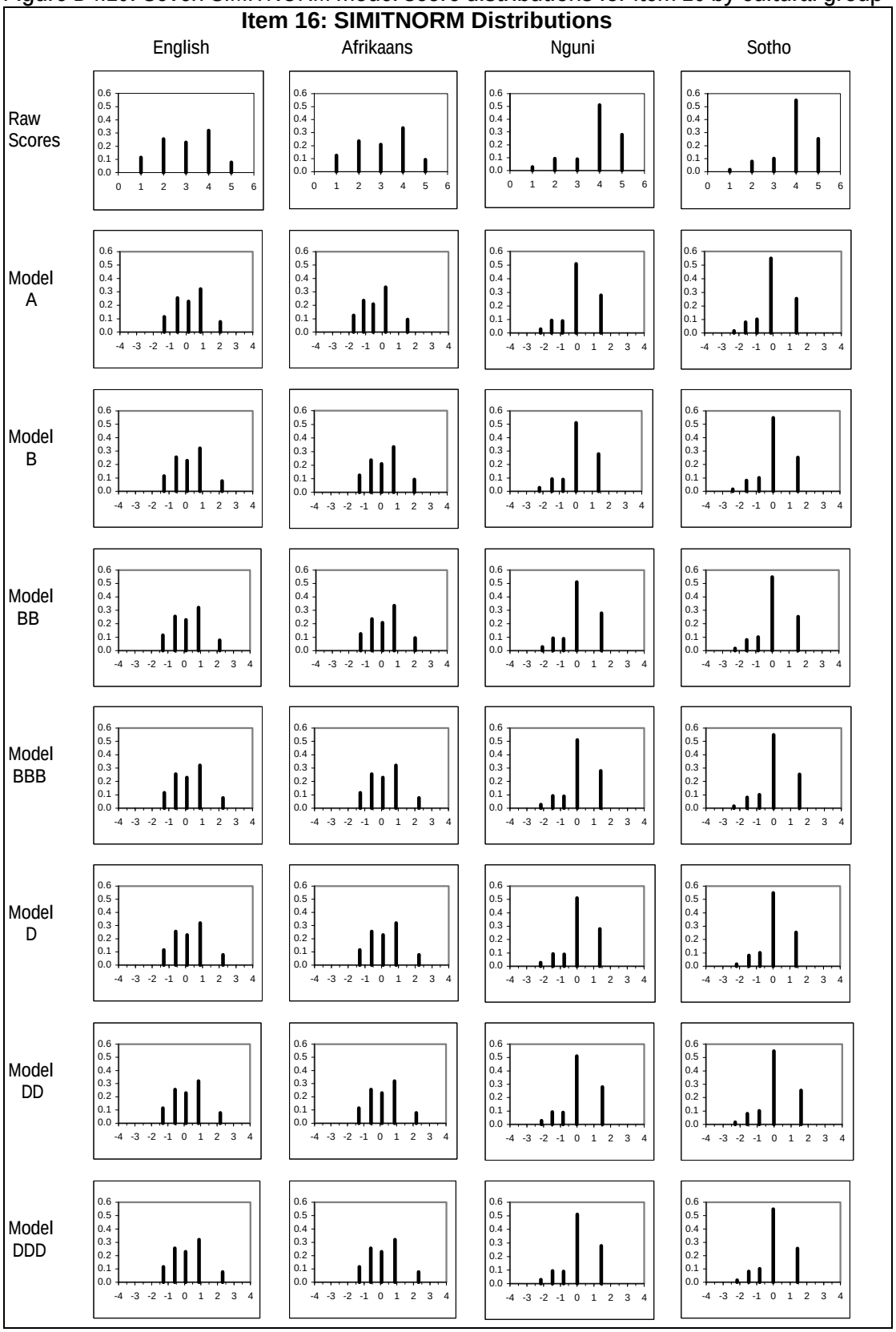


Figure D4.11: Seven SIMITNORM model score distributions for Item 17 by cultural group

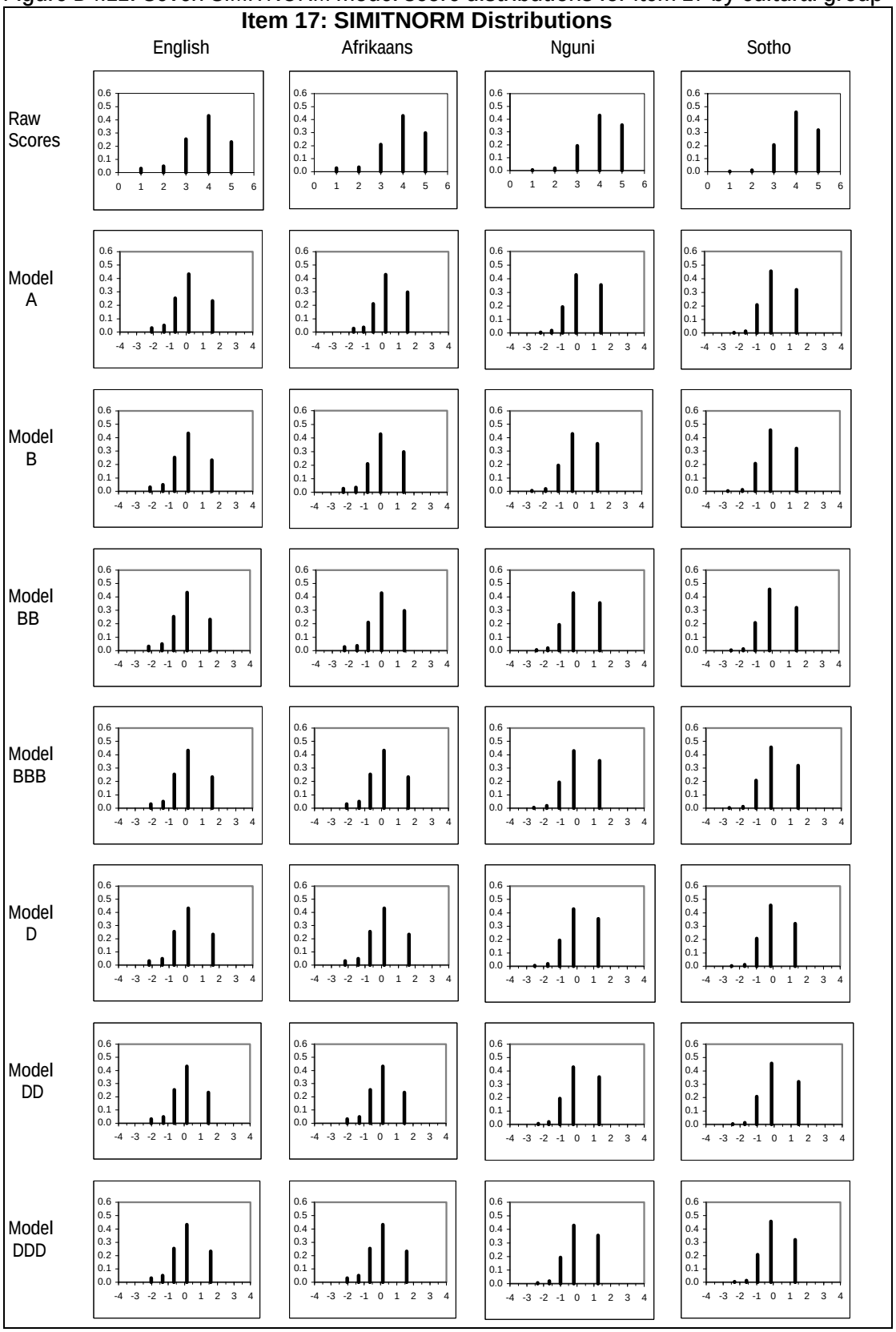


Figure D4.12: Seven SIMITNORM model score distributions for Item 18 by cultural group

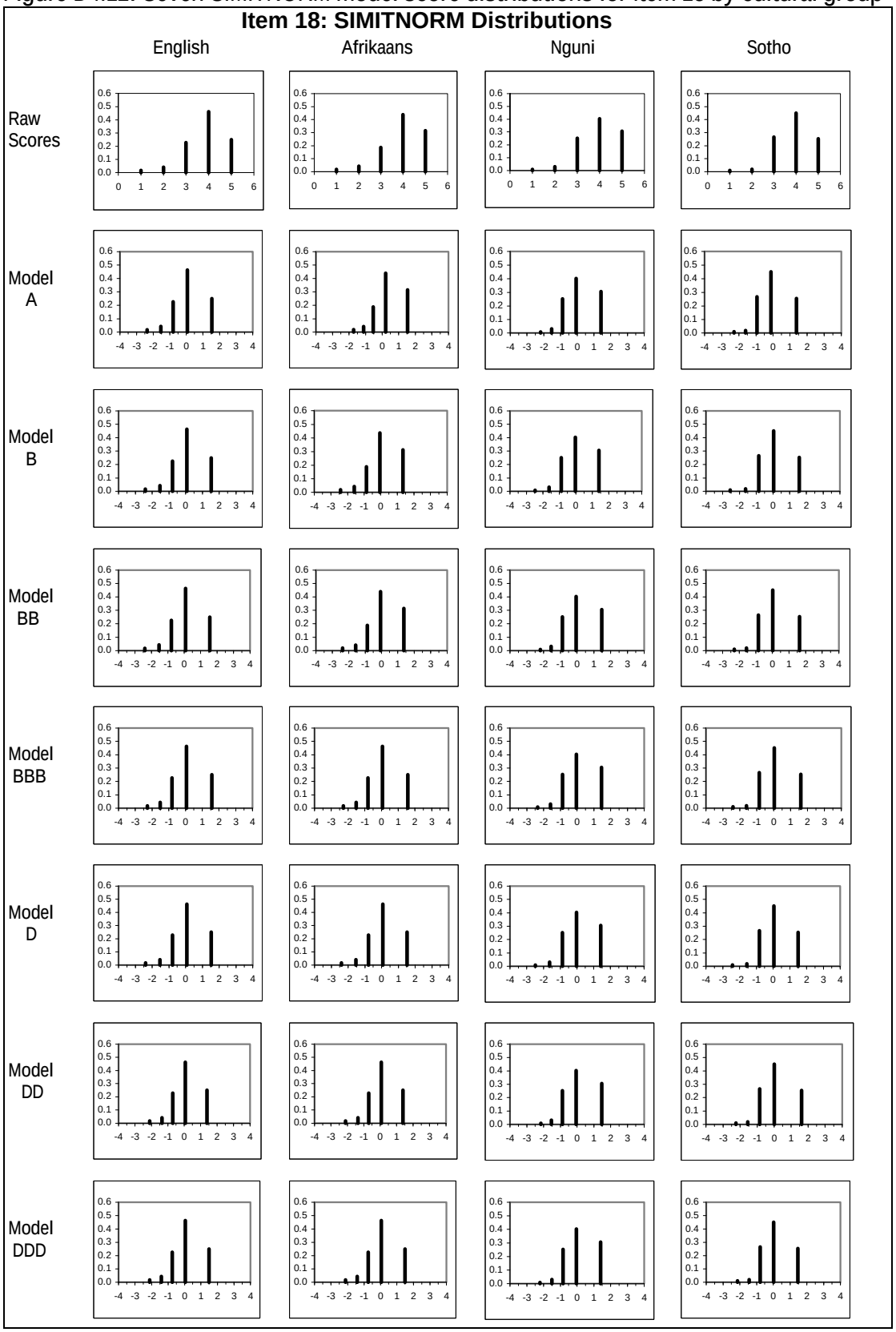


Figure D4.13: Seven SIMITNORM model score distributions for Item 20 by cultural group

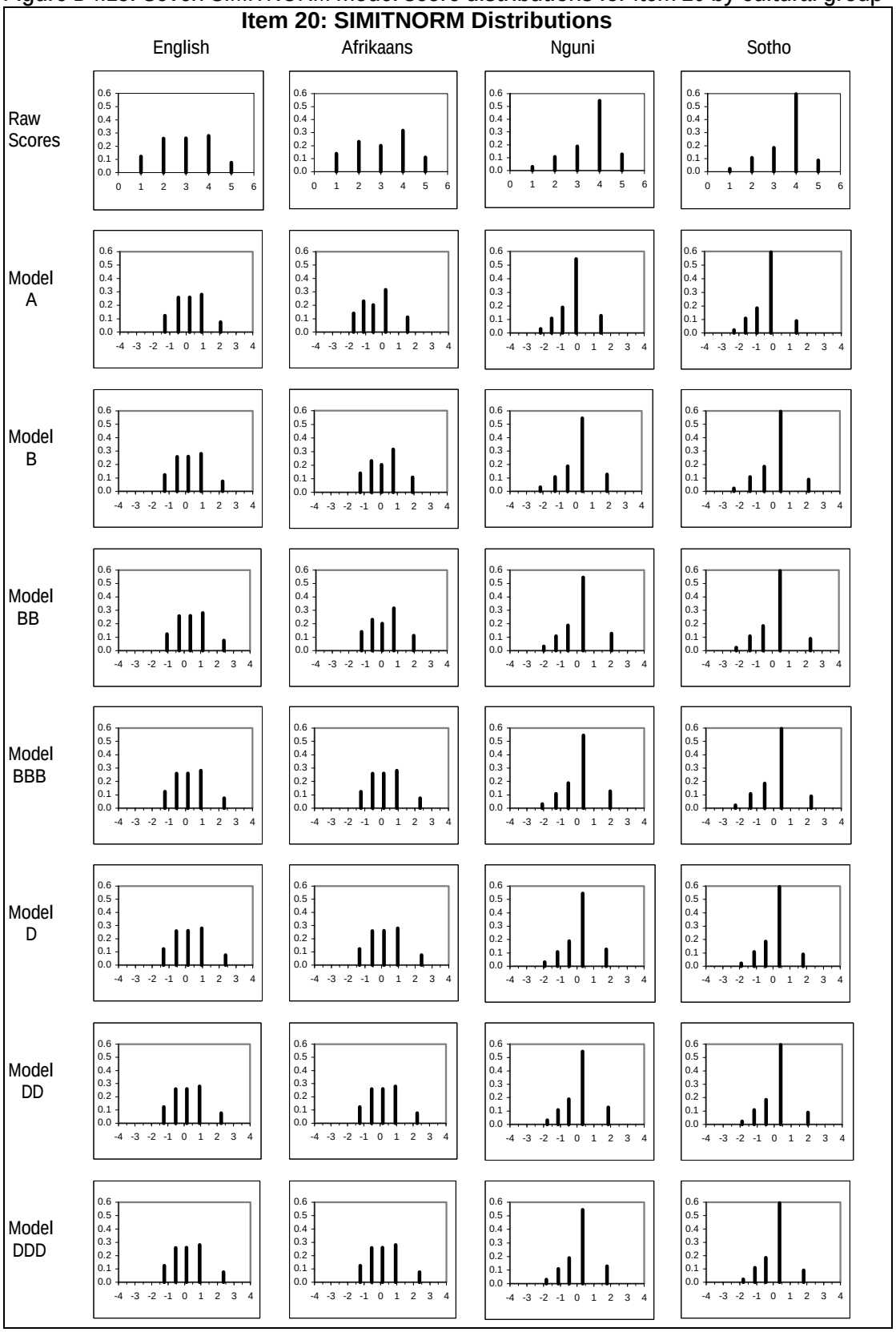


Figure D4.14: Seven SIMITNORM model score distributions for Item 21 by cultural group

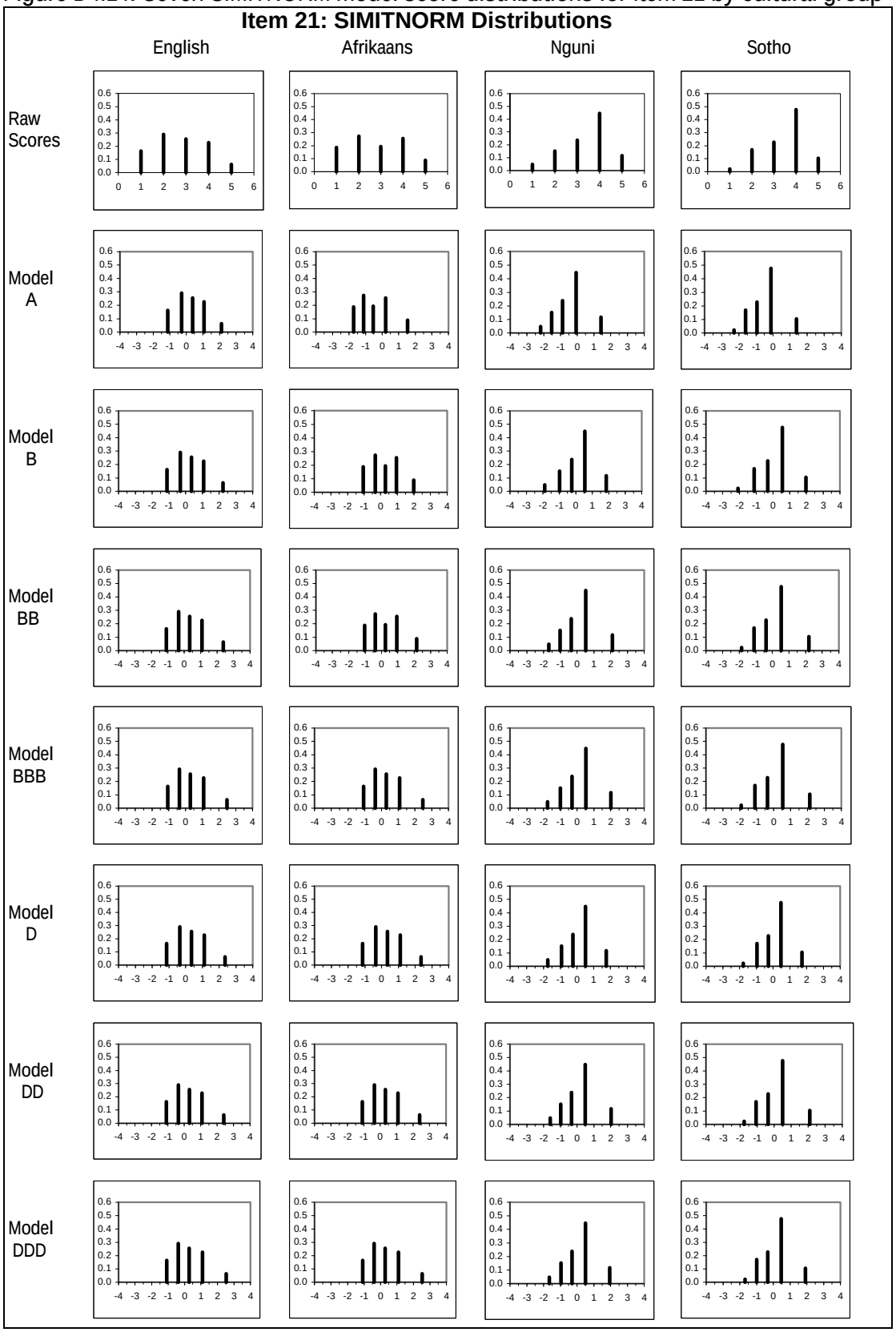


Figure D4.15: Seven SIMITNORM model score distributions for Item 22 by cultural group

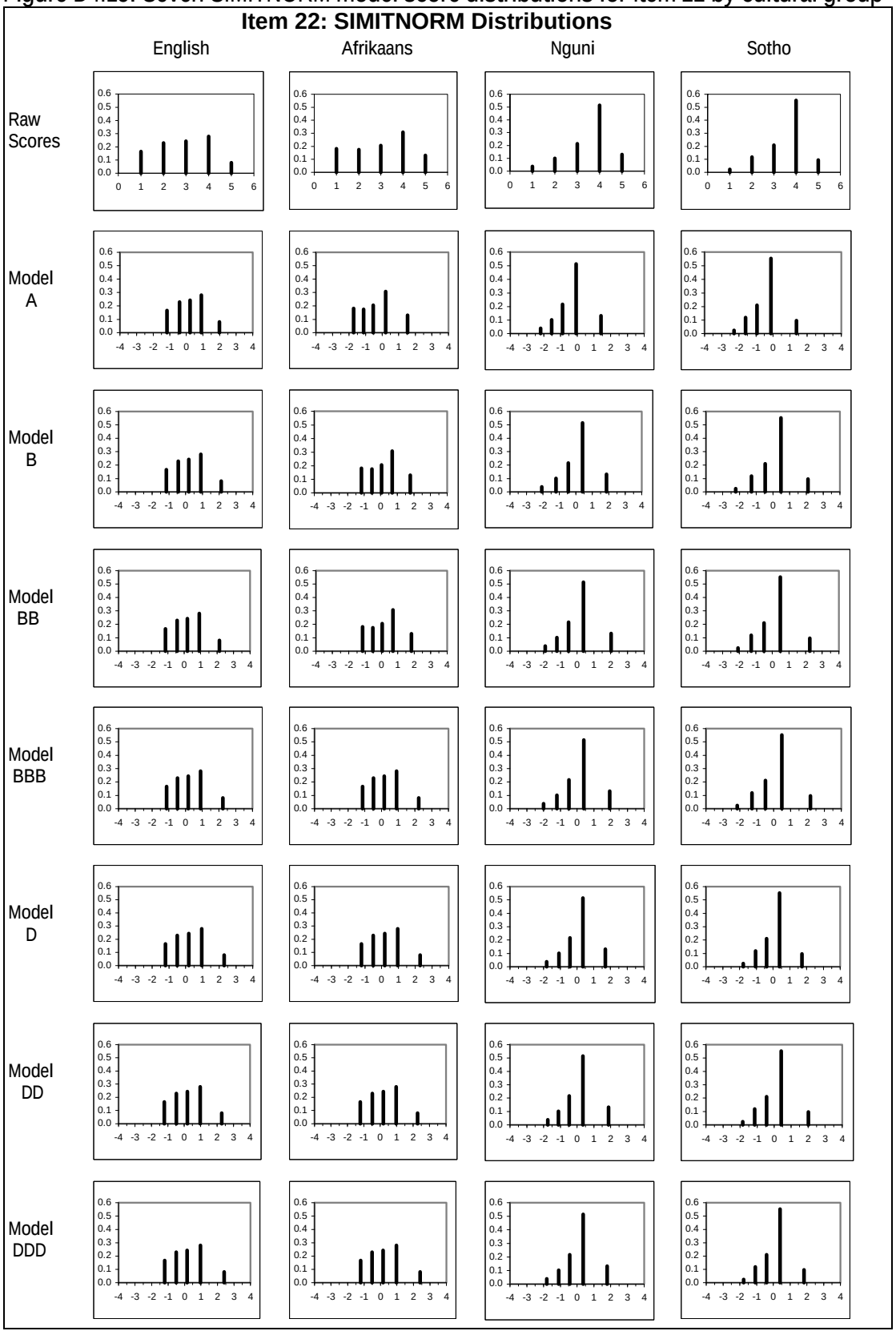


Figure D4.16: Seven SIMITNORM model score distributions for Item 24 by cultural group

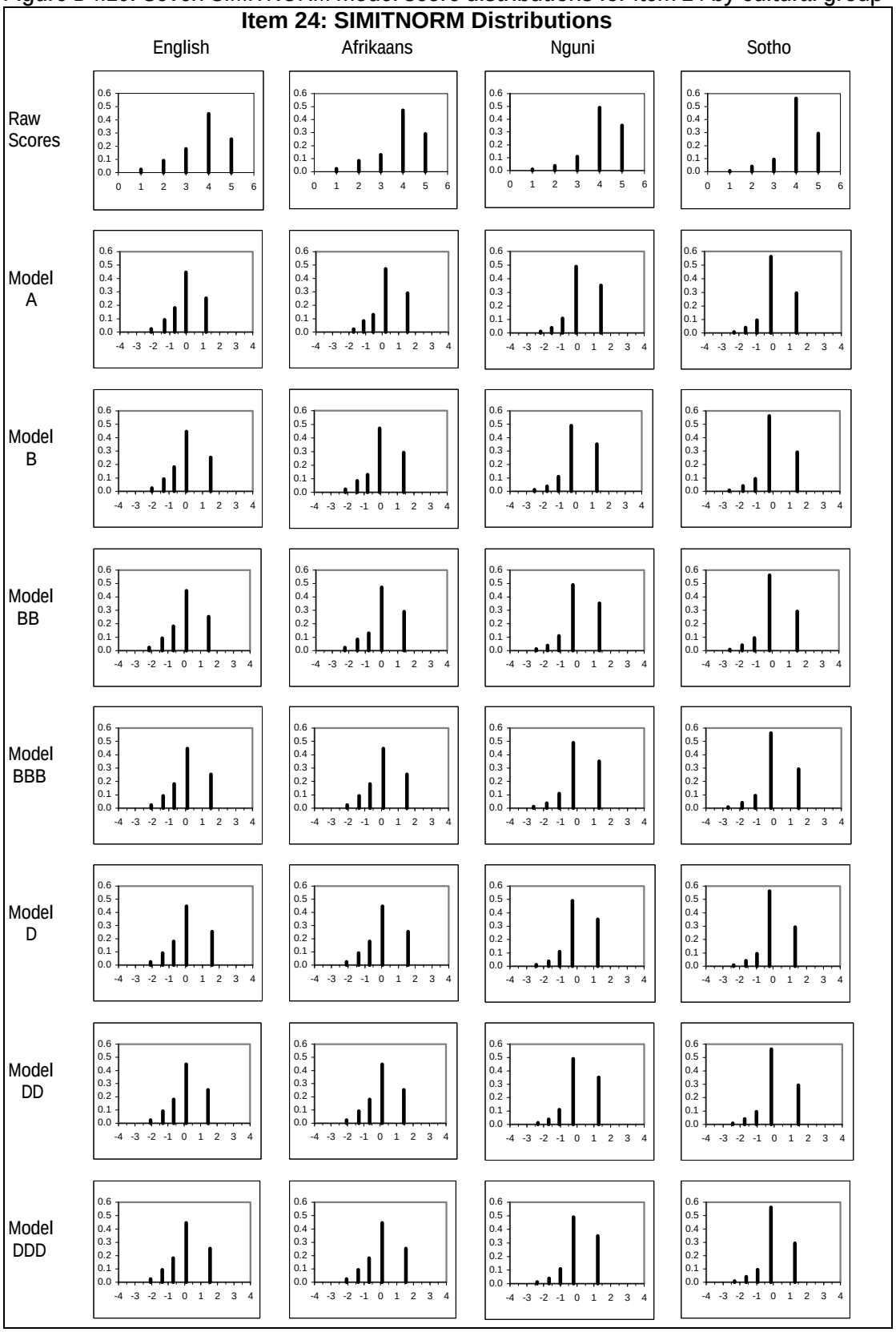


Figure D4.17: Seven SIMITNORM model score distributions for Item 25 by cultural group

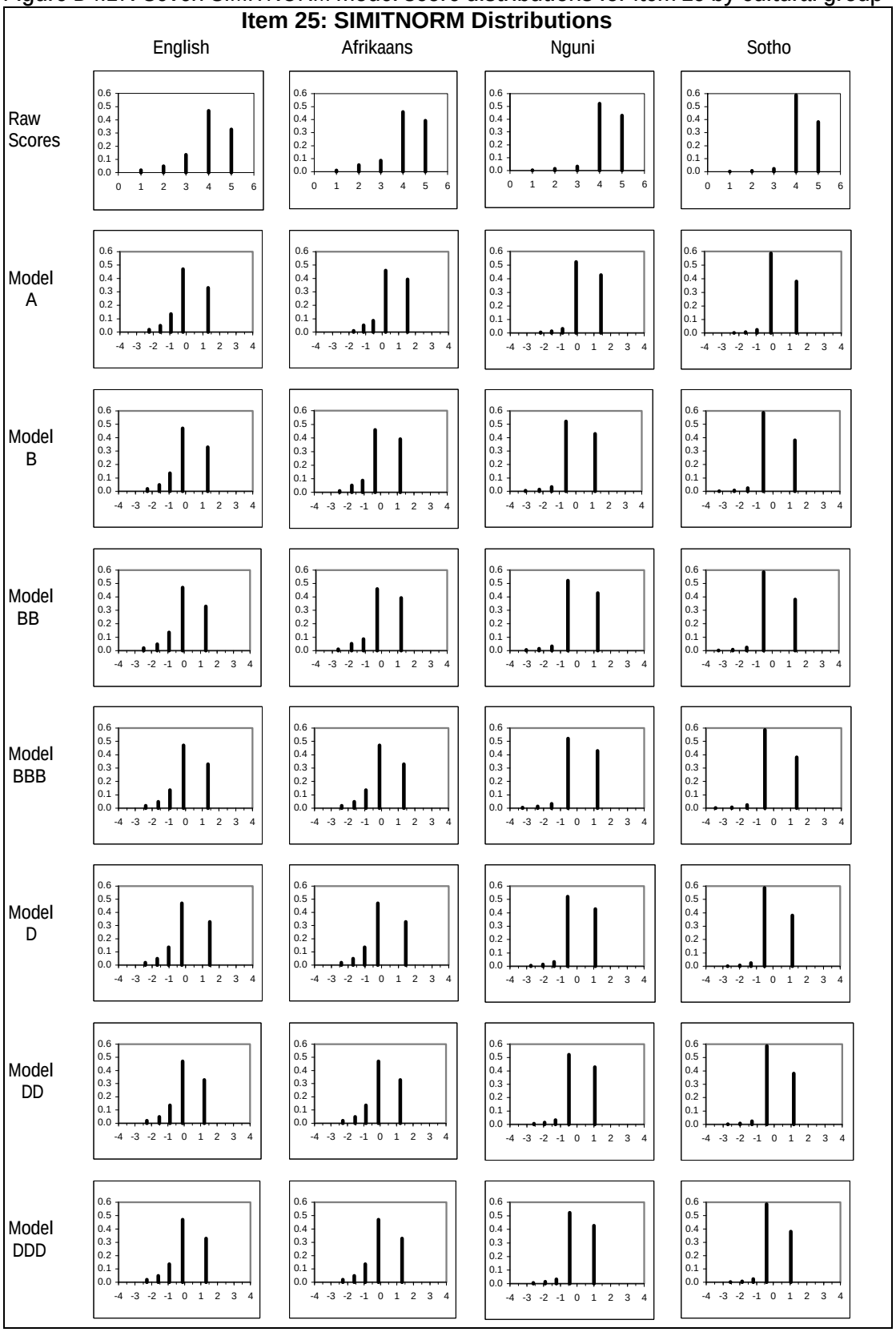


Figure D4.18: Seven SIMITNORM model score distributions for Item 27 by cultural group

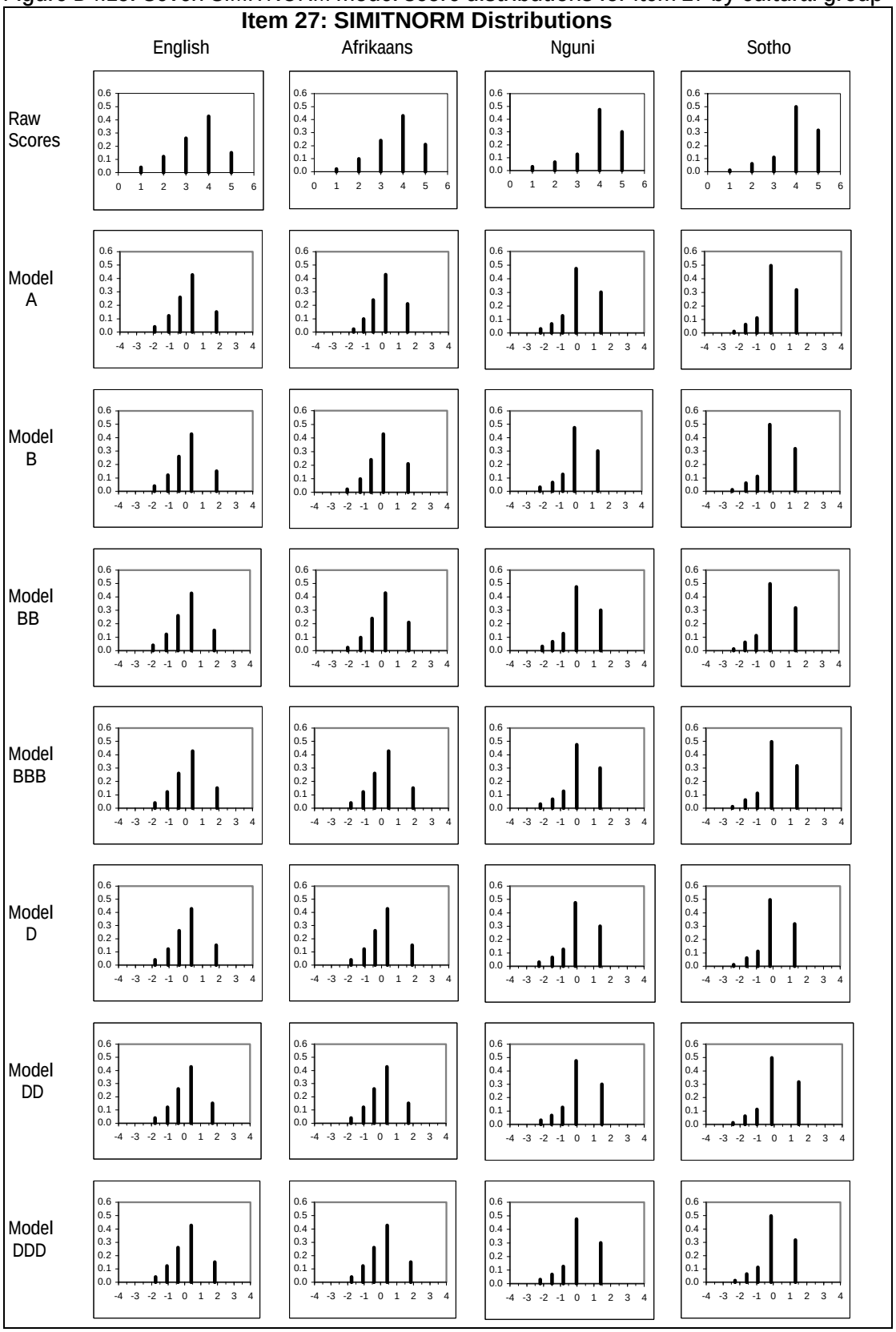


Figure D4.19: Seven SIMITNORM model score distributions for Item 28 by cultural group

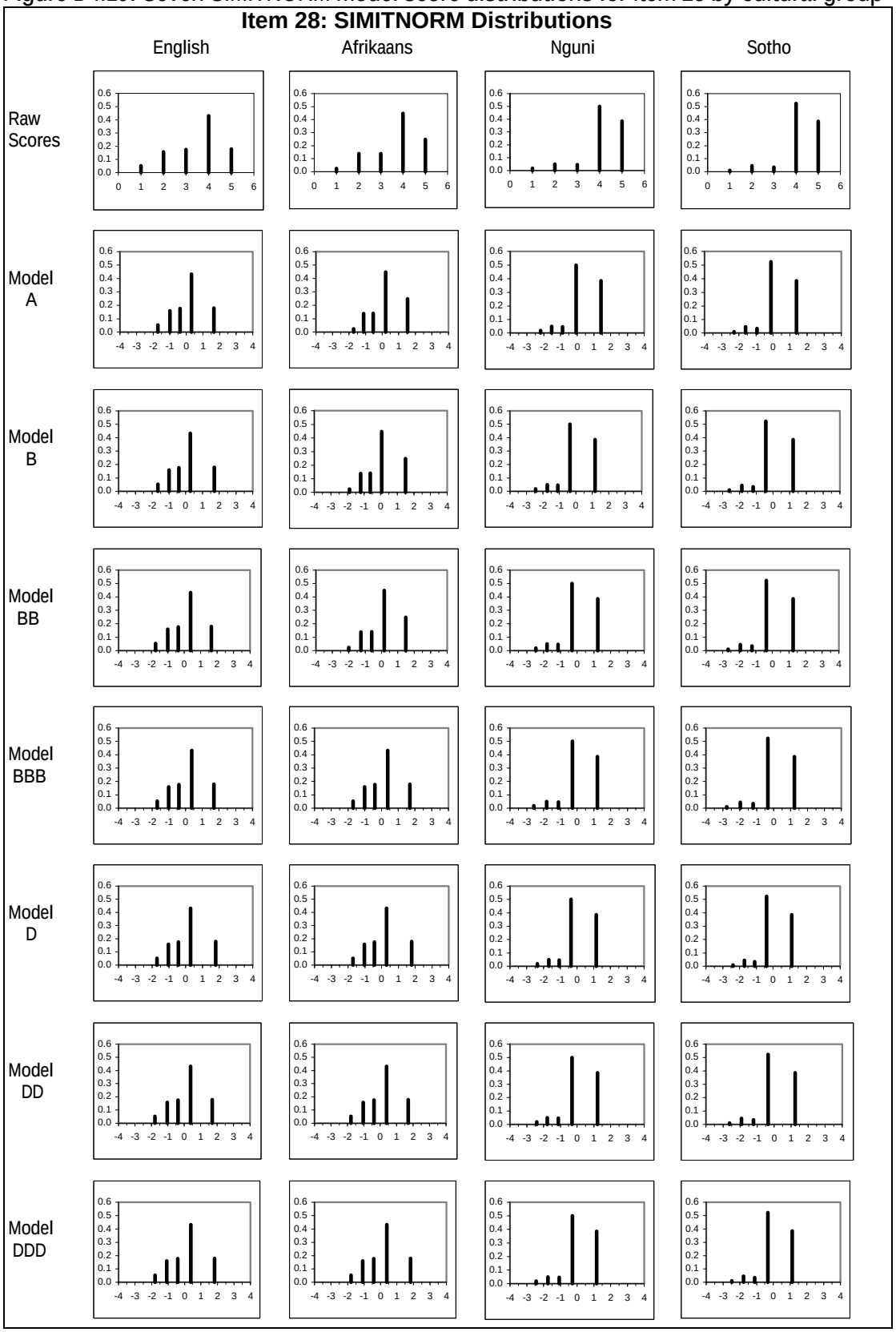
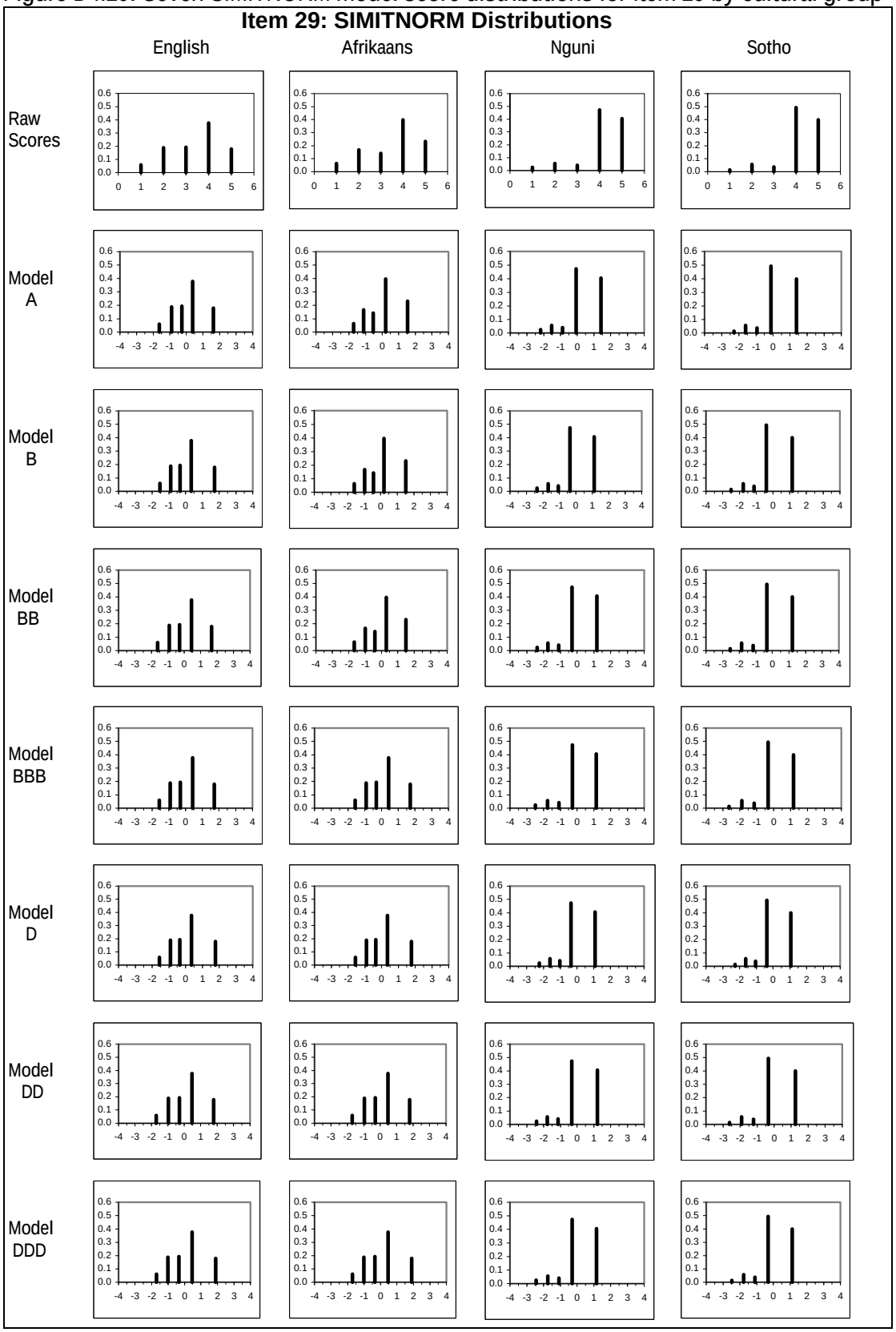


Figure D4.20: Seven SIMITNORM model score distributions for Item 29 by cultural group



APPENDIX E: Resulting Scoring Levels of SIMITNORM

Table E.1(a): SIMITNORM Scores: Model A

A	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
	Item 1				Item 9				Item 17				Item 24			
1	-1.583	-1.724	-2.185	-2.280	-2.124	-2.021	-1.763	-1.871	-2.089	-2.147	-2.824	-3.129	-2.115	-2.137	-2.523	-2.791
2	-0.898	-1.120	-1.530	-1.612	-1.565	-1.477	-1.193	-1.289	-1.349	-1.477	-1.879	-1.971	-1.322	-1.416	-1.739	-1.833
3	-0.280	-0.524	-0.875	-0.946	-1.040	-0.964	-0.675	-0.758	-0.660	-0.823	-1.069	-1.050	-0.688	-0.814	-1.087	-1.079
4	0.429	0.208	-0.060	-0.119	-0.390	-0.329	-0.053	-0.117	0.154	-0.024	-0.176	-0.101	-0.017	-0.150	-0.372	-0.293
5	1.640	1.532	1.427	1.386	1.113	1.141	1.350	1.333	1.576	1.414	1.297	1.370	1.192	1.095	0.964	1.101
	Item 2				Item 10				Item 18				Item 25			
1	-1.568	-1.632	-2.198	-2.264	-2.234	-2.315	-2.124	-2.213	-2.360	-2.261	-2.627	-2.771	-2.246	-2.349	-3.061	-4.369
2	-0.894	-0.964	-1.553	-1.569	-1.688	-1.722	-1.533	-1.594	-1.534	-1.562	-1.664	-1.705	-1.566	-1.716	-2.224	-2.774
3	-0.284	-0.338	-0.888	-0.875	-1.158	-1.167	-0.978	-1.016	-0.788	-0.889	-0.863	-0.837	-0.940	-1.109	-1.471	-1.615
4	0.421	0.402	-0.041	-0.012	-0.488	-0.482	-0.294	-0.307	0.07v1	-0.074	-0.002	0.079	-0.190	-0.359	-0.587	-0.511
5	1.630	1.699	1.532	1.561	1.094	1.095	1.285	1.322	1.539	1.379	1.386	1.529	1.325	1.199	1.168	1.263
	Item 3				Item 14				Item 20				Item 27			
1	-1.522	-1.539	-2.240	-2.600	-0.994	-1.050	-1.829	-1.932	-1.289	-1.256	-2.012	-2.120	-1.906	-2.163	-2.050	-2.371
2	-0.805	-0.848	-1.597	-1.774	-0.100	-0.166	-1.172	-1.213	-0.467	-0.515	-1.271	-1.358	-1.061	-1.256	-1.405	-1.606
3	-0.179	-0.223	-0.931	-0.985	0.573	0.512	-0.528	-0.535	0.208	0.102	-0.538	-0.566	-0.373	-0.545	-0.817	-0.945
4	0.522	0.495	-0.083	-0.037	1.229	1.186	0.260	0.270	0.923	0.765	0.366	0.443	0.373	0.201	-0.115	-0.191
5	1.693	1.724	1.498	1.646	2.189	2.191	1.679	1.685	2.060	1.836	2.004	2.324	1.814	1.591	1.372	1.342
	Item 4				Item 15				Item 21				Item 28			
1	-1.485	-1.582	-1.904	-2.057	-1.293	-1.367	-1.750	-1.937	-1.127	-1.065	-1.788	-2.011	-1.721	-1.980	-2.249	-2.442
2	-0.780	-0.914	-1.225	-1.296	-0.471	-0.568	-1.137	-1.268	-0.279	-0.286	-1.013	-1.122	-0.983	-1.222	-1.682	-1.806
3	-0.156	-0.292	-0.569	-0.586	0.219	0.107	-0.542	-0.629	0.389	0.326	-0.300	-0.337	-0.368	-0.593	-1.119	-1.197
4	0.552	0.442	0.227	0.252	0.967	0.842	0.180	0.140	1.070	0.949	0.533	0.551	0.314	0.100	-0.400	-0.439
5	1.745	1.724	1.649	1.718	2.181	2.041	1.475	1.504	2.116	1.902	1.975	2.045	1.662	1.464	1.211	1.221
	Item 8				Item 16				Item 22				Item 29			
1	-2.221	-1.994	-1.975	-2.246	-1.331	-1.307	-2.028	-2.230	-1.169	-1.173	-1.950	-2.083	-1.640	-1.637	-2.142	-2.297
2	-1.545	-1.402	-1.338	-1.548	-0.529	-0.548	-1.437	-1.553	-0.401	-0.492	-1.214	-1.283	-0.875	-0.966	-1.631	-1.722
3	-0.956	-0.868	-0.766	-0.924	0.143	0.094	-0.820	-0.870	0.234	0.084	-0.495	-0.495	-0.263	-0.402	-1.105	-1.154
4	-0.271	-0.229	-0.083	-0.183	0.869	0.794	-0.030	-0.018	0.912	0.711	0.386	0.471	0.391	0.227	-0.416	-0.431
5	1.228	1.204	1.447	1.468	2.045	1.937	1.447	1.542	1.997	1.733	1.971	2.217	1.635	1.478	1.162	1.182

Table E.1(b): Differences between adjacent SIMITNORM Score categories: Model A

A	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
			Item 1				Item 9				Item 17				Item 24	
2-1	0.686	0.605	0.655	0.668	0.558	0.544	0.570	0.582	0.739	0.670	0.945	1.158	0.793	0.721	0.784	0.958
3-2	0.618	0.596	0.655	0.666	0.526	0.513	0.518	0.531	0.689	0.654	0.810	0.921	0.634	0.602	0.652	0.754
4-3	0.709	0.732	0.815	0.826	0.650	0.635	0.622	0.641	0.814	0.799	0.893	0.949	0.671	0.664	0.715	0.786
5-4	1.211	1.324	1.487	1.505	1.503	1.470	1.403	1.450	1.422	1.438	1.473	1.471	1.209	1.245	1.336	1.394
			Item 2				Item 10				Item 18				Item 25	
2-1	0.674	0.668	0.645	0.695	0.546	0.593	0.591	0.619	0.826	0.699	0.962	1.066	0.680	0.633	0.838	1.595
3-2	0.611	0.625	0.666	0.694	0.530	0.555	0.555	0.578	0.746	0.673	0.801	0.868	0.626	0.607	0.753	1.159
4-3	0.705	0.740	0.846	0.863	0.670	0.684	0.684	0.709	0.859	0.814	0.861	0.916	0.750	0.750	0.884	1.104
5-4	1.209	1.297	1.573	1.573	1.582	1.578	1.579	1.629	1.469	1.453	1.388	1.450	1.515	1.557	1.755	1.774
			Item 3				Item 14				Item 20				Item 27	
2-1	0.718	0.691	0.643	0.826	0.894	0.884	0.658	0.719	0.823	0.740	0.741	0.763	0.845	0.906	0.645	0.766
3-2	0.626	0.624	0.666	0.789	0.673	0.678	0.644	0.677	0.674	0.617	0.733	0.791	0.688	0.712	0.588	0.660
4-3	0.701	0.718	0.849	0.948	0.656	0.673	0.788	0.805	0.715	0.664	0.904	1.009	0.746	0.745	0.702	0.755
5-4	1.171	1.229	1.580	1.683	0.960	1.005	1.419	1.415	1.137	1.071	1.638	1.880	1.441	1.390	1.487	1.533
			Item 4				Item 15				Item 21				Item 28	
2-1	0.705	0.668	0.679	0.761	0.822	0.799	0.613	0.669	0.848	0.779	0.775	0.889	0.738	0.758	0.567	0.636
3-2	0.624	0.622	0.656	0.710	0.690	0.675	0.595	0.639	0.668	0.612	0.713	0.785	0.615	0.629	0.563	0.609
4-3	0.707	0.734	0.796	0.838	0.748	0.735	0.723	0.769	0.682	0.623	0.832	0.887	0.682	0.694	0.719	0.758
5-4	1.193	1.282	1.422	1.466	1.214	1.199	1.294	1.365	1.046	0.953	1.442	1.494	1.348	1.364	1.611	1.660
			Item 8				Item 16				Item 22				Item 29	
2-1	0.676	0.592	0.636	0.698	0.802	0.759	0.591	0.678	0.768	0.682	0.736	0.800	0.765	0.671	0.511	0.576
3-2	0.590	0.534	0.573	0.624	0.672	0.642	0.617	0.682	0.635	0.576	0.720	0.787	0.612	0.564	0.526	0.568
4-3	0.685	0.639	0.683	0.741	0.726	0.700	0.790	0.852	0.678	0.627	0.880	0.967	0.654	0.629	0.689	0.723
5-4	1.499	1.433	1.529	1.651	1.176	1.143	1.477	1.561	1.086	1.022	1.585	1.746	1.244	1.251	1.579	1.614

Table E.2(a): SIMITNORM Scores: Model B

B	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
	Item 1				Item 9				Item 17				Item 24			
1	-1.536	-1.745	-2.249	-2.344	-2.154	-2.049	-1.747	-1.857	-2.127	-2.259	-2.663	-2.697	-2.020	-2.133	-2.520	-2.613
2	-0.913	-1.116	-1.543	-1.627	-1.572	-1.482	-1.193	-1.288	-1.354	-1.496	-1.830	-1.836	-1.307	-1.420	-1.749	-1.789
3	-0.314	-0.512	-0.865	-0.938	-1.035	-0.958	-0.682	-0.762	-0.651	-0.802	-1.072	-1.052	-0.676	-0.788	-1.067	-1.060
4	0.408	0.217	-0.046	-0.107	-0.382	-0.320	-0.060	-0.122	0.162	0.001	-0.195	-0.146	0.056	-0.056	-0.275	-0.215
5	1.694	1.515	1.410	1.372	1.109	1.135	1.358	1.338	1.561	1.382	1.313	1.413	1.494	1.382	1.279	1.446
	Item 2				Item 10				Item 18				Item 25			
1	-1.531	-1.616	-2.317	-2.338	-2.278	-2.302	-2.115	-2.193	-2.408	-2.421	-2.481	-2.553	-2.292	-2.477	-3.064	-3.207
2	-0.908	-0.968	-1.576	-1.585	-1.700	-1.718	-1.531	-1.589	-1.544	-1.595	-1.627	-1.648	-1.577	-1.752	-2.225	-2.311
3	-0.312	-0.349	-0.868	-0.864	-1.155	-1.168	-0.980	-1.018	-0.782	-0.866	-0.874	-0.850	-0.934	-1.100	-1.471	-1.507
4	0.404	0.395	-0.017	0.001	-0.478	-0.485	-0.296	-0.312	0.079	-0.043	-0.023	0.051	-0.180	-0.336	-0.587	-0.564
5	1.675	1.717	1.494	1.539	1.089	1.097	1.287	1.326	1.527	1.342	1.409	1.569	1.316	1.181	1.168	1.309
	Item 3				Item 14				Item 20				Item 27			
1	-1.481	-1.522	-2.423	-2.669	-0.977	-1.033	-1.995	-2.085	-1.254	-1.232	-2.149	-2.338	-1.858	-2.027	-2.178	-2.429
2	-0.823	-0.855	-1.635	-1.793	-0.165	-0.218	-1.154	-1.212	-0.506	-0.549	-1.275	-1.371	-1.060	-1.237	-1.423	-1.620
3	-0.212	-0.238	-0.905	-0.981	0.516	0.464	-0.450	-0.482	0.160	0.060	-0.496	-0.511	-0.390	-0.573	-0.788	-0.940
4	0.504	0.488	-0.047	-0.028	1.251	1.201	0.311	0.308	0.919	0.753	0.391	0.469	0.360	0.169	-0.079	-0.180
5	1.751	1.750	1.445	1.631	2.442	2.395	1.543	1.587	2.204	1.927	1.894	2.130	1.851	1.645	1.332	1.332
	Item 4				Item 15				Item 21				Item 28			
1	-1.457	-1.586	-1.953	-2.068	-1.269	-1.340	-1.826	-2.024	-1.113	-1.054	-1.894	-2.098	-1.656	-1.882	-2.452	-2.612
2	-0.795	-0.913	-1.225	-1.297	-0.500	-0.592	-1.125	-1.272	-0.310	-0.321	-0.996	-1.122	-0.989	-1.212	-1.734	-1.858
3	-0.181	-0.289	-0.551	-0.584	0.187	0.076	-0.500	-0.600	0.358	0.289	-0.248	-0.309	-0.399	-0.620	-1.099	-1.191
4	0.539	0.444	0.240	0.254	0.969	0.837	0.212	0.165	1.075	0.944	0.554	0.563	0.289	0.071	-0.358	-0.413
5	1.793	1.719	1.618	1.712	2.295	2.128	1.420	1.462	2.229	1.999	1.847	1.968	1.716	1.503	1.177	1.200
	Item 8				Item 16				Item 22				Item 29			
1	-2.177	-2.008	-1.985	-2.248	-1.295	-1.279	-2.212	-2.409	-1.152	-1.163	-2.071	-2.231	-1.537	-1.588	-2.370	-2.500
2	-1.535	-1.404	-1.340	-1.548	-0.564	-0.579	-1.455	-1.587	-0.439	-0.519	-1.213	-1.290	-0.892	-0.974	-1.680	-1.777
3	-0.961	-0.865	-0.763	-0.923	0.099	0.054	-0.770	-0.843	0.192	0.052	-0.453	-0.456	-0.320	-0.432	-1.070	-1.138
4	-0.280	-0.225	-0.079	-0.183	0.863	0.785	0.020	0.016	0.908	0.700	0.409	0.490	0.346	0.202	-0.357	-0.392
5	1.236	1.201	1.443	1.468	2.173	2.037	1.374	1.487	2.117	1.794	1.866	2.088	1.728	1.515	1.119	1.153

Table E.2(b): Differences between adjacent SIMITNORM Score categories: Model B

B	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
	Item 1				Item 9				Item 17				Item 24			
2-1	0.623	0.629	0.706	0.717	0.581	0.568	0.554	0.570	0.773	0.763	0.833	0.861	0.713	0.713	0.771	0.824
3-2	0.599	0.604	0.678	0.689	0.537	0.524	0.511	0.526	0.703	0.694	0.758	0.783	0.631	0.631	0.682	0.729
4-3	0.722	0.729	0.818	0.831	0.653	0.638	0.622	0.640	0.813	0.803	0.877	0.907	0.732	0.732	0.791	0.846
5-4	1.286	1.298	1.457	1.479	1.490	1.455	1.419	1.460	1.398	1.381	1.508	1.559	1.438	1.438	1.554	1.661
	Item 2				Item 10				Item 18				Item 25			
2-1	0.623	0.648	0.741	0.754	0.578	0.583	0.584	0.604	0.864	0.826	0.854	0.905	0.716	0.726	0.839	0.896
3-2	0.596	0.619	0.708	0.720	0.546	0.551	0.551	0.570	0.762	0.729	0.753	0.798	0.642	0.652	0.754	0.804
4-3	0.716	0.744	0.851	0.866	0.676	0.683	0.683	0.707	0.860	0.823	0.851	0.901	0.754	0.764	0.884	0.943
5-4	1.271	1.322	1.511	1.537	1.567	1.582	1.583	1.638	1.448	1.385	1.432	1.517	1.496	1.517	1.755	1.872
	Item 3				Item 14				Item 20				Item 27			
2-1	0.659	0.667	0.788	0.876	0.812	0.814	0.840	0.872	0.748	0.683	0.874	0.966	0.797	0.790	0.755	0.809
3-2	0.610	0.618	0.730	0.812	0.680	0.682	0.704	0.731	0.666	0.609	0.779	0.861	0.671	0.664	0.635	0.680
4-3	0.716	0.725	0.857	0.953	0.736	0.737	0.761	0.790	0.758	0.693	0.887	0.980	0.749	0.742	0.709	0.760
5-4	1.247	1.262	1.492	1.659	1.191	1.194	1.232	1.279	1.286	1.174	1.503	1.661	1.491	1.476	1.411	1.512
	Item 4				Item 15				Item 21				Item 28			
2-1	0.662	0.674	0.728	0.770	0.770	0.749	0.701	0.753	0.803	0.733	0.899	0.977	0.667	0.669	0.718	0.754
3-2	0.614	0.624	0.674	0.714	0.686	0.668	0.625	0.671	0.668	0.610	0.747	0.812	0.590	0.592	0.635	0.667
4-3	0.720	0.733	0.792	0.838	0.782	0.761	0.712	0.765	0.717	0.655	0.803	0.873	0.688	0.691	0.741	0.778
5-4	1.254	1.275	1.377	1.458	1.327	1.291	1.208	1.297	1.154	1.055	1.292	1.405	1.426	1.432	1.535	1.613
	Item 8				Item 16				Item 22				Item 29			
2-1	0.642	0.604	0.645	0.699	0.732	0.700	0.757	0.822	0.712	0.644	0.858	0.941	0.646	0.614	0.690	0.722
3-2	0.574	0.540	0.576	0.625	0.663	0.633	0.685	0.744	0.631	0.571	0.761	0.834	0.571	0.543	0.610	0.639
4-3	0.681	0.640	0.684	0.741	0.764	0.731	0.790	0.858	0.716	0.648	0.862	0.946	0.667	0.634	0.713	0.746
5-4	1.516	1.425	1.522	1.650	1.310	1.252	1.354	1.471	1.209	1.094	1.457	1.598	1.382	1.313	1.476	1.546

Table E.3(a): SIMITNORM Scores: Model D

D	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
	Item 1				Item 9				Item 17				Item 24			
1	-1.618	-1.823	-2.094	-2.150	-2.097	-2.044	-1.787	-1.846	-2.178	-2.346	-2.534	-2.481	-2.102	-2.220	-2.428	-2.355
2	-0.974	-1.179	-1.451	-1.507	-1.533	-1.480	-1.222	-1.282	-1.401	-1.569	-1.756	-1.704	-1.369	-1.487	-1.695	-1.622
3	-0.344	-0.549	-0.821	-0.877	-1.011	-0.958	-0.700	-0.760	-0.682	-0.849	-1.037	-0.984	-0.714	-0.831	-1.040	-0.967
4	0.427	0.222	-0.050	-0.106	-0.374	-0.321	-0.063	-0.123	0.162	-0.006	-0.194	-0.141	0.054	-0.064	-0.272	-0.199
5	1.816	1.610	1.339	1.283	1.082	1.135	1.393	1.333	1.629	1.461	1.274	1.326	1.575	1.457	1.249	1.322
	Item 2				Item 10				Item 18				Item 25			
1	-1.658	-1.683	-2.114	-2.097	-2.312	-2.315	-2.125	-2.130	-2.390	-2.513	-2.491	-2.419	-2.412	-2.577	-2.758	-2.702
2	-0.999	-1.024	-1.455	-1.438	-1.725	-1.727	-1.538	-1.543	-1.534	-1.658	-1.636	-1.564	-1.694	-1.859	-2.040	-1.984
3	-0.354	-0.379	-0.811	-0.793	-1.171	-1.173	-0.983	-0.988	-0.778	-0.901	-0.879	-0.807	-1.024	-1.188	-1.369	-1.313
4	0.434	0.409	-0.023	-0.005	-0.485	-0.487	-0.297	-0.302	0.077	-0.046	-0.024	0.048	-0.212	-0.376	-0.557	-0.501
5	1.852	1.827	1.396	1.413	1.104	1.102	1.292	1.287	1.520	1.396	1.418	1.490	1.445	1.281	1.099	1.156
	Item 3				Item 14				Item 20				Item 27			
1	-1.603	-1.627	-2.201	-2.177	-0.997	-1.051	-1.972	-1.986	-1.313	-1.416	-1.931	-1.899	-1.828	-2.014	-2.264	-2.356
2	-0.916	-0.941	-1.515	-1.491	-0.173	-0.227	-1.148	-1.161	-0.547	-0.650	-1.164	-1.132	-1.043	-1.229	-1.480	-1.572
3	-0.256	-0.280	-0.854	-0.831	0.523	0.469	-0.452	-0.465	0.155	0.051	-0.463	-0.431	-0.383	-0.570	-0.820	-0.912
4	0.542	0.518	-0.056	-0.033	1.280	1.227	0.306	0.292	0.969	0.866	0.351	0.384	0.354	0.168	-0.082	-0.175
5	1.965	1.941	1.367	1.390	2.515	2.461	1.541	1.527	2.375	2.272	1.757	1.789	1.821	1.635	1.385	1.292
	Item 4				Item 15				Item 21				Item 28			
1	-1.538	-1.647	-1.877	-1.877	-1.225	-1.329	-1.931	-1.993	-1.146	-1.192	-1.747	-1.780	-1.713	-1.942	-2.364	-2.398
2	-0.846	-0.955	-1.186	-1.185	-0.478	-0.582	-1.185	-1.247	-0.334	-0.380	-0.935	-0.968	-1.033	-1.261	-1.684	-1.717
3	-0.199	-0.307	-0.538	-0.537	0.183	0.079	-0.523	-0.586	0.357	0.311	-0.244	-0.276	-0.423	-0.651	-1.074	-1.107
4	0.568	0.460	0.229	0.229	0.932	0.828	0.226	0.164	1.115	1.069	0.514	0.482	0.297	0.068	-0.354	-0.388
5	1.912	1.804	1.573	1.573	2.198	2.094	1.492	1.430	2.357	2.312	1.756	1.724	1.802	1.573	1.151	1.118
	Item 8				Item 16				Item 22				Item 29			
1	-2.179	-2.137	-1.977	-2.066	-1.307	-1.351	-2.169	-2.173	-1.215	-1.362	-1.818	-1.782	-1.560	-1.698	-2.259	-2.275
2	-1.537	-1.495	-1.335	-1.424	-0.579	-0.623	-1.441	-1.445	-0.483	-0.630	-1.086	-1.050	-0.915	-1.053	-1.614	-1.630
3	-0.963	-0.921	-0.761	-0.850	0.091	0.047	-0.770	-0.775	0.185	0.039	-0.417	-0.382	-0.336	-0.474	-1.035	-1.051
4	-0.281	-0.240	-0.080	-0.168	0.875	0.831	0.013	0.009	0.962	0.815	0.359	0.395	0.348	0.210	-0.350	-0.367
5	1.238	1.279	1.439	1.351	2.234	2.189	1.372	1.368	2.301	2.155	1.699	1.735	1.782	1.644	1.083	1.067

Table E.3(b): Differences between adjacent SIMITNORM Score categories: Model D

D	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
	Item 1				Item 9				Item 17				Item 24			
2-1	0.644	0.644	0.644	0.644	0.564	0.564	0.564	0.564	0.777	0.777	0.777	0.777	0.732	0.732	0.732	0.732
3-2	0.630	0.630	0.630	0.630	0.522	0.522	0.522	0.522	0.719	0.719	0.719	0.719	0.656	0.656	0.656	0.656
4-3	0.771	0.771	0.771	0.771	0.637	0.637	0.637	0.637	0.844	0.844	0.844	0.844	0.768	0.768	0.768	0.768
5-4	1.389	1.389	1.389	1.389	1.456	1.456	1.456	1.456	1.467	1.467	1.467	1.467	1.521	1.521	1.521	1.521
	Item 2				Item 10				Item 18				Item 25			
2-1	0.659	0.659	0.659	0.659	0.587	0.587	0.587	0.587	0.856	0.856	0.856	0.856	0.718	0.718	0.718	0.718
3-2	0.645	0.645	0.645	0.645	0.554	0.554	0.554	0.554	0.756	0.756	0.756	0.756	0.671	0.671	0.671	0.671
4-3	0.788	0.788	0.788	0.788	0.686	0.686	0.686	0.686	0.855	0.855	0.855	0.855	0.812	0.812	0.812	0.812
5-4	1.418	1.418	1.418	1.418	1.589	1.589	1.589	1.589	1.442	1.442	1.442	1.442	1.657	1.657	1.657	1.657
	Item 3				Item 14				Item 20				Item 27			
2-1	0.686	0.686	0.686	0.686	0.824	0.824	0.824	0.824	0.767	0.767	0.767	0.767	0.784	0.784	0.784	0.784
3-2	0.661	0.661	0.661	0.661	0.696	0.696	0.696	0.696	0.701	0.701	0.701	0.701	0.660	0.660	0.660	0.660
4-3	0.798	0.798	0.798	0.798	0.758	0.758	0.758	0.758	0.815	0.815	0.815	0.815	0.737	0.737	0.737	0.737
5-4	1.423	1.423	1.423	1.423	1.235	1.235	1.235	1.235	1.406	1.406	1.406	1.406	1.467	1.467	1.467	1.467
	Item 4				Item 15				Item 21				Item 28			
2-1	0.692	0.692	0.692	0.692	0.746	0.746	0.746	0.746	0.812	0.812	0.812	0.812	0.681	0.681	0.681	0.681
3-2	0.648	0.648	0.648	0.648	0.661	0.661	0.661	0.661	0.691	0.691	0.691	0.691	0.610	0.610	0.610	0.610
4-3	0.767	0.767	0.767	0.767	0.750	0.750	0.750	0.750	0.758	0.758	0.758	0.758	0.720	0.720	0.720	0.720
5-4	1.344	1.344	1.344	1.344	1.266	1.266	1.266	1.266	1.243	1.243	1.243	1.243	1.505	1.505	1.505	1.505
	Item 8				Item 16				Item 22				Item 29			
2-1	0.642	0.642	0.642	0.642	0.728	0.728	0.728	0.728	0.732	0.732	0.732	0.732	0.645	0.645	0.645	0.645
3-2	0.574	0.574	0.574	0.574	0.671	0.671	0.671	0.671	0.669	0.669	0.669	0.669	0.579	0.579	0.579	0.579
4-3	0.681	0.681	0.681	0.681	0.784	0.784	0.784	0.784	0.777	0.777	0.777	0.777	0.684	0.684	0.684	0.684
5-4	1.519	1.519	1.519	1.519	1.359	1.359	1.359	1.359	1.340	1.340	1.340	1.340	1.434	1.434	1.434	1.434

Table E.4(a): SIMITNORM Scores: Model BB

BB	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
		Item 1				Item 9				Item 17				Item 24		
1	-1.594	-1.793	-2.249	-2.293	-2.473	-2.297	-1.747	-1.971	-2.167	-2.264	-2.663	-2.491	-2.138	-2.227	-2.520	-2.576
2	-0.890	-1.103	-1.543	-1.613	-1.695	-1.577	-1.193	-1.361	-1.356	-1.494	-1.830	-1.765	-1.351	-1.467	-1.749	-1.826
3	-0.267	-0.477	-0.865	-0.940	-1.006	-0.925	-0.682	-0.758	-0.637	-0.796	-1.072	-1.046	-0.653	-0.778	-1.067	-1.085
4	0.441	0.245	-0.046	-0.109	-0.224	-0.171	-0.060	-0.014	0.179	0.010	-0.195	-0.160	0.138	0.017	-0.275	-0.169
5	1.642	1.491	1.410	1.406	1.103	1.129	1.358	1.344	1.563	1.400	1.313	1.458	1.481	1.388	1.279	1.500
		Item 2				Item 10				Item 18				Item 25		
1	-1.583	-1.655	-2.317	-2.280	-2.699	-2.701	-2.115	-2.424	-2.398	-2.370	-2.481	-2.296	-2.478	-2.643	-3.064	-3.247
2	-0.884	-0.953	-1.576	-1.570	-1.875	-1.900	-1.531	-1.731	-1.539	-1.582	-1.627	-1.573	-1.653	-1.835	-2.225	-2.397
3	-0.266	-0.317	-0.868	-0.868	-1.145	-1.174	-0.980	-1.046	-0.778	-0.869	-0.874	-0.857	-0.922	-1.103	-1.471	-1.556
4	0.435	0.418	-0.017	-0.001	-0.317	-0.336	-0.296	-0.200	0.085	-0.046	-0.023	0.026	-0.093	-0.258	-0.587	-0.518
5	1.626	1.686	1.494	1.580	1.088	1.110	1.287	1.343	1.551	1.375	1.409	1.637	1.315	1.201	1.168	1.375
		Item 3				Item 14				Item 20				Item 27		
1	-1.511	-1.535	-2.423	-2.526	-0.965	-1.013	-1.995	-1.873	-1.053	-1.224	-2.149	-2.184	-1.907	-2.060	-2.178	-2.340
2	-0.807	-0.846	-1.635	-1.751	-0.193	-0.258	-1.154	-1.208	-0.299	-0.554	-1.275	-1.360	-1.094	-1.278	-1.423	-1.652
3	-0.183	-0.221	-0.905	-0.984	0.490	0.427	-0.450	-0.550	0.369	0.054	-0.496	-0.545	-0.374	-0.570	-0.788	-0.973
4	0.524	0.501	-0.047	-0.037	1.265	1.217	0.311	0.262	1.127	0.756	0.391	0.461	0.442	0.248	-0.079	-0.133
5	1.725	1.746	1.445	1.689	2.582	2.580	1.543	1.743	2.414	1.966	1.894	2.295	1.829	1.659	1.332	1.397
		Item 4				Item 15				Item 21				Item 28		
1	-1.483	-1.600	-1.953	-1.976	-1.270	-1.331	-1.826	-1.896	-1.093	-1.035	-1.894	-1.858	-1.753	-1.990	-2.452	-2.665
2	-0.777	-0.903	-1.225	-1.285	-0.493	-0.595	-1.125	-1.263	-0.340	-0.366	-0.996	-1.113	-1.012	-1.259	-1.734	-1.944
3	-0.151	-0.272	-0.551	-0.602	0.195	0.072	-0.500	-0.638	0.327	0.240	-0.248	-0.376	-0.355	-0.597	-1.099	-1.231
4	0.558	0.457	0.240	0.242	0.975	0.842	0.212	0.135	1.083	0.939	0.554	0.534	0.389	0.168	-0.358	-0.351
5	1.762	1.715	1.618	1.782	2.301	2.171	1.420	1.544	2.368	2.147	1.847	2.194	1.654	1.488	1.177	1.255
		Item 8				Item 16				Item 22				Item 29		
1	-2.467	-2.200	-1.985	-2.374	-1.304	-1.277	-2.212	-2.258	-1.150	-1.157	-2.071	-2.074	-1.622	-1.657	-2.370	-2.543
2	-1.662	-1.486	-1.340	-1.662	-0.551	-0.577	-1.455	-1.556	-0.436	-0.528	-1.213	-1.279	-0.906	-0.996	-1.680	-1.853
3	-0.949	-0.839	-0.763	-0.959	0.116	0.058	-0.770	-0.862	0.197	0.042	-0.453	-0.492	-0.271	-0.396	-1.070	-1.172
4	-0.141	-0.091	-0.079	-0.090	0.872	0.791	0.020	-0.005	0.914	0.700	0.409	0.479	0.448	0.297	-0.357	-0.330
5	1.233	1.199	1.443	1.495	2.156	2.056	1.374	1.558	2.132	1.836	1.866	2.250	1.671	1.492	1.119	1.205

Table E.4(b): Differences between adjacent SIMITNORM Score categories: Model BB

BB	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
			Item 1				Item 9				Item 17				Item 24	
2-1	0.704	0.690	0.706	0.680	0.778	0.720	0.554	0.609	0.812	0.770	0.833	0.726	0.787	0.760	0.771	0.750
3-2	0.624	0.625	0.678	0.673	0.689	0.653	0.511	0.603	0.719	0.698	0.758	0.718	0.697	0.688	0.682	0.742
4-3	0.707	0.722	0.818	0.831	0.782	0.754	0.622	0.744	0.815	0.806	0.877	0.887	0.791	0.795	0.791	0.915
5-4	1.201	1.246	1.457	1.515	1.328	1.300	1.419	1.358	1.385	1.391	1.508	1.617	1.343	1.372	1.554	1.670
			Item 2				Item 10				Item 18				Item 25	
2-1	0.698	0.702	0.741	0.710	0.824	0.801	0.584	0.693	0.859	0.787	0.854	0.723	0.825	0.808	0.839	0.850
3-2	0.618	0.636	0.708	0.702	0.730	0.726	0.551	0.685	0.761	0.713	0.753	0.716	0.731	0.732	0.754	0.841
4-3	0.701	0.735	0.851	0.867	0.828	0.838	0.683	0.846	0.863	0.824	0.851	0.883	0.829	0.846	0.884	1.038
5-4	1.191	1.268	1.511	1.581	1.406	1.446	1.583	1.543	1.466	1.421	1.432	1.611	1.408	1.459	1.755	1.893
			Item 3				Item 14				Item 20				Item 27	
2-1	0.704	0.690	0.788	0.775	0.772	0.755	0.840	0.665	0.754	0.670	0.874	0.824	0.813	0.782	0.755	0.687
3-2	0.624	0.625	0.730	0.767	0.683	0.684	0.704	0.658	0.668	0.608	0.779	0.815	0.720	0.708	0.635	0.680
4-3	0.707	0.722	0.857	0.947	0.775	0.790	0.761	0.812	0.758	0.702	0.887	1.006	0.817	0.818	0.709	0.839
5-4	1.201	1.245	1.492	1.726	1.316	1.363	1.232	1.481	1.287	1.211	1.503	1.835	1.387	1.411	1.411	1.531
			Item 4				Item 15				Item 21				Item 28	
2-1	0.706	0.697	0.728	0.691	0.777	0.736	0.701	0.633	0.753	0.669	0.899	0.745	0.741	0.731	0.718	0.721
3-2	0.625	0.631	0.674	0.684	0.688	0.667	0.625	0.626	0.667	0.606	0.747	0.737	0.657	0.662	0.635	0.713
4-3	0.709	0.729	0.792	0.844	0.781	0.770	0.712	0.773	0.757	0.700	0.803	0.910	0.745	0.765	0.741	0.880
5-4	1.205	1.258	1.377	1.539	1.326	1.329	1.208	1.409	1.285	1.207	1.292	1.660	1.265	1.320	1.535	1.606
			Item 8				Item 16				Item 22				Item 29	
2-1	0.805	0.714	0.645	0.711	0.753	0.700	0.757	0.702	0.714	0.629	0.858	0.795	0.716	0.662	0.690	0.689
3-2	0.713	0.647	0.576	0.704	0.667	0.635	0.685	0.694	0.633	0.570	0.761	0.787	0.635	0.600	0.610	0.682
4-3	0.809	0.748	0.684	0.869	0.756	0.733	0.790	0.857	0.717	0.658	0.862	0.971	0.720	0.693	0.713	0.842
5-4	1.373	1.290	1.522	1.585	1.284	1.265	1.354	1.563	1.218	1.135	1.457	1.771	1.222	1.195	1.476	1.535

Table E.5(a): SIMITNORM Scores: Model BBB

BBB	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
		Item 1				Item 9				Item 17				Item 24		
1	-1.553	-1.765	-2.288	-2.388	-2.385	-2.253	-1.891	-2.026	-2.081	-2.210	-2.591	-2.620	-2.055	-2.176	-2.593	-2.696
2	-0.905	-1.109	-1.547	-1.634	-1.684	-1.574	-1.248	-1.359	-1.346	-1.486	-1.805	-1.808	-1.342	-1.461	-1.813	-1.862
3	-0.297	-0.495	-0.853	-0.928	-1.027	-0.939	-0.645	-0.734	-0.657	-0.807	-1.069	-1.047	-0.674	-0.790	-1.082	-1.081
4	0.423	0.233	-0.030	-0.091	-0.249	-0.186	0.069	0.008	0.160	-0.002	-0.196	-0.146	0.118	0.004	-0.215	-0.154
5	1.693	1.517	1.421	1.384	1.124	1.143	1.329	1.315	1.600	1.417	1.344	1.445	1.514	1.406	1.314	1.480
		Item 2				Item 10				Item 18				Item 25		
1	-1.544	-1.630	-2.349	-2.373	-2.596	-2.635	-2.401	-2.511	-2.289	-2.309	-2.347	-2.411	-2.377	-2.576	-3.248	-3.431
2	-0.900	-0.961	-1.579	-1.589	-1.856	-1.885	-1.666	-1.747	-1.516	-1.570	-1.588	-1.606	-1.634	-1.819	-2.356	-2.472
3	-0.298	-0.335	-0.858	-0.855	-1.163	-1.182	-0.977	-1.030	-0.792	-0.878	-0.878	-0.852	-0.939	-1.110	-1.520	-1.574
4	0.417	0.408	-0.003	0.016	-0.341	-0.349	-0.161	-0.180	0.067	-0.057	-0.035	0.041	-0.114	-0.270	-0.529	-0.510
5	1.677	1.718	1.506	1.551	1.108	1.120	1.278	1.318	1.581	1.390	1.451	1.617	1.340	1.212	1.219	1.367
		Item 3				Item 14				Item 20				Item 27		
1	-1.472	-1.513	-2.403	-2.643	-0.953	-1.004	-1.873	-1.937	-1.229	-1.212	-2.080	-2.261	-1.839	-2.009	-2.187	-2.438
2	-0.823	-0.856	-1.629	-1.783	-0.223	-0.275	-1.166	-1.210	-0.524	-0.568	-1.271	-1.365	-1.096	-1.272	-1.476	-1.676
3	-0.214	-0.239	-0.903	-0.978	0.460	0.409	-0.504	-0.529	0.137	0.034	-0.512	-0.526	-0.399	-0.581	-0.809	-0.962
4	0.507	0.491	-0.043	-0.022	1.271	1.219	0.281	0.279	0.921	0.749	0.387	0.470	0.427	0.237	-0.019	-0.116
5	1.780	1.780	1.474	1.662	2.700	2.649	1.665	1.703	2.304	2.009	1.973	2.225	1.883	1.681	1.375	1.376
		Item 4				Item 15				Item 21				Item 28		
1	-1.449	-1.577	-1.939	-2.051	-1.247	-1.314	-1.780	-1.961	-1.076	-1.027	-1.768	-1.928	-1.701	-1.946	-2.588	-2.776
2	-0.796	-0.913	-1.223	-1.293	-0.518	-0.607	-1.130	-1.267	-0.368	-0.383	-1.012	-1.117	-1.021	-1.256	-1.830	-1.975
3	-0.183	-0.291	-0.551	-0.583	0.165	0.054	-0.522	-0.617	0.296	0.220	-0.303	-0.356	-0.384	-0.610	-1.119	-1.224
4	0.543	0.447	0.244	0.258	0.975	0.838	0.200	0.154	1.083	0.934	0.536	0.546	0.372	0.156	-0.277	-0.333
5	1.823	1.749	1.647	1.743	2.403	2.222	1.472	1.513	2.471	2.195	2.017	2.136	1.703	1.508	1.208	1.237
		Item 8				Item 16				Item 22				Item 29		
1	-2.369	-2.157	-2.138	-2.463	-1.278	-1.263	-2.164	-2.348	-1.134	-1.149	-2.001	-2.149	-1.575	-1.631	-2.490	-2.640
2	-1.645	-1.484	-1.425	-1.679	-0.574	-0.591	-1.449	-1.574	-0.464	-0.545	-1.212	-1.284	-0.917	-1.002	-1.763	-1.876
3	-0.967	-0.853	-0.757	-0.943	0.085	0.039	-0.780	-0.848	0.164	0.021	-0.472	-0.473	-0.301	-0.414	-1.082	-1.160
4	-0.163	-0.105	0.034	-0.072	0.866	0.786	0.014	0.012	0.909	0.692	0.404	0.489	0.430	0.284	-0.275	-0.311
5	1.256	1.213	1.429	1.465	2.244	2.103	1.415	1.529	2.222	1.875	1.950	2.184	1.718	1.515	1.148	1.187

Table E.5(b): Differences between adjacent SIMITNORM Score categories: Model BBB

BBB	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
	Item 1				Item 9				Item 17				Item 24			
2-1	0.648	0.655	0.741	0.754	0.701	0.678	0.643	0.667	0.735	0.725	0.786	0.812	0.713	0.715	0.780	0.834
3-2	0.608	0.614	0.694	0.706	0.657	0.636	0.603	0.625	0.689	0.679	0.736	0.761	0.668	0.670	0.731	0.782
4-3	0.720	0.728	0.823	0.837	0.778	0.753	0.714	0.741	0.817	0.805	0.873	0.902	0.792	0.795	0.867	0.926
5-4	1.270	1.284	1.451	1.476	1.373	1.329	1.260	1.307	1.440	1.419	1.540	1.590	1.397	1.401	1.529	1.634
	Item 2				Item 10				Item 18				Item 25			
2-1	0.643	0.669	0.770	0.784	0.740	0.750	0.735	0.765	0.773	0.739	0.759	0.805	0.742	0.757	0.892	0.958
3-2	0.603	0.627	0.721	0.734	0.693	0.703	0.689	0.717	0.724	0.692	0.711	0.754	0.696	0.709	0.836	0.898
4-3	0.714	0.743	0.855	0.871	0.822	0.833	0.816	0.849	0.858	0.821	0.843	0.894	0.825	0.840	0.991	1.064
5-4	1.260	1.310	1.508	1.535	1.449	1.469	1.439	1.498	1.514	1.447	1.486	1.576	1.454	1.482	1.748	1.877
	Item 3				Item 14				Item 20				Item 27			
2-1	0.650	0.658	0.774	0.860	0.730	0.730	0.707	0.727	0.706	0.643	0.810	0.896	0.744	0.737	0.711	0.762
3-2	0.609	0.616	0.726	0.806	0.684	0.684	0.662	0.681	0.661	0.603	0.759	0.840	0.697	0.691	0.667	0.714
4-3	0.722	0.731	0.860	0.955	0.810	0.810	0.785	0.808	0.784	0.714	0.899	0.995	0.826	0.819	0.790	0.846
5-4	1.272	1.288	1.517	1.684	1.429	1.429	1.384	1.424	1.382	1.260	1.586	1.755	1.456	1.444	1.394	1.492
	Item 4				Item 15				Item 21				Item 28			
2-1	0.654	0.664	0.716	0.758	0.729	0.706	0.650	0.694	0.708	0.644	0.756	0.812	0.680	0.690	0.758	0.802
3-2	0.612	0.622	0.671	0.710	0.683	0.662	0.609	0.650	0.664	0.603	0.708	0.761	0.637	0.646	0.711	0.751
4-3	0.726	0.738	0.796	0.842	0.810	0.784	0.721	0.771	0.787	0.715	0.840	0.902	0.755	0.766	0.842	0.890
5-4	1.280	1.301	1.403	1.484	1.428	1.383	1.272	1.359	1.388	1.261	1.481	1.590	1.332	1.351	1.485	1.570
	Item 8				Item 16				Item 22				Item 29			
2-1	0.724	0.673	0.712	0.785	0.703	0.672	0.715	0.774	0.671	0.604	0.789	0.866	0.658	0.628	0.727	0.764
3-2	0.678	0.631	0.668	0.735	0.659	0.630	0.670	0.726	0.628	0.566	0.739	0.811	0.616	0.589	0.681	0.716
4-3	0.804	0.748	0.791	0.872	0.781	0.747	0.794	0.860	0.745	0.671	0.877	0.961	0.731	0.698	0.807	0.849
5-4	1.418	1.319	1.396	1.537	1.378	1.317	1.400	1.517	1.313	1.183	1.546	1.696	1.288	1.231	1.424	1.497

Table E.6(a): SIMITNORM Scores: Model DD

DD	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
	Item 1				Item 9				Item 17				Item 24			
1	-1.712	-1.853	-2.166	-2.389	-2.409	-2.280	-2.067	-2.292	-2.019	-2.098	-2.318	-2.431	-2.054	-2.092	-2.339	-2.435
2	-0.965	-1.152	-1.531	-1.687	-1.663	-1.580	-1.432	-1.589	-1.273	-1.398	-1.683	-1.728	-1.308	-1.391	-1.704	-1.733
3	-0.296	-0.507	-0.881	-0.986	-0.993	-0.934	-0.782	-0.889	-0.603	-0.752	-1.033	-1.028	-0.638	-0.746	-1.055	-1.032
4	0.472	0.249	-0.059	-0.117	-0.226	-0.179	0.039	-0.020	0.164	0.003	-0.212	-0.159	0.129	0.010	-0.233	-0.163
5	1.786	1.567	1.469	1.476	1.089	1.139	1.567	1.573	1.478	1.321	1.316	1.434	1.443	1.328	1.295	1.430
	Item 2				Item 10				Item 18				Item 25			
1	-1.714	-1.679	-2.136	-2.276	-2.483	-2.411	-2.302	-2.480	-2.111	-2.149	-2.166	-2.249	-2.274	-2.338	-2.576	-2.709
2	-0.967	-0.979	-1.501	-1.573	-1.737	-1.711	-1.667	-1.778	-1.365	-1.448	-1.531	-1.546	-1.528	-1.638	-1.941	-2.006
3	-0.298	-0.333	-0.851	-0.873	-1.067	-1.065	-1.017	-1.077	-0.695	-0.803	-0.882	-0.846	-0.858	-0.992	-1.291	-1.306
4	0.469	0.422	-0.029	-0.003	-0.300	-0.310	-0.196	-0.208	0.072	-0.047	-0.060	0.024	-0.091	-0.236	-0.470	-0.436
5	1.784	1.740	1.499	1.589	1.015	1.008	1.333	1.385	1.386	1.271	1.468	1.616	1.223	1.081	1.059	1.156
	Item 3				Item 14				Item 20				Item 27			
1	-1.621	-1.586	-2.175	-2.309	-0.942	-0.952	-1.844	-1.995	-1.254	-1.299	-1.776	-1.876	-1.772	-1.878	-2.158	-2.413
2	-0.875	-0.885	-1.540	-1.606	-0.196	-0.251	-1.209	-1.292	-0.508	-0.599	-1.141	-1.173	-1.026	-1.177	-1.523	-1.710
3	-0.205	-0.240	-0.890	-0.905	0.474	0.394	-0.560	-0.592	0.162	0.047	-0.491	-0.473	-0.356	-0.532	-0.873	-1.010
4	0.562	0.516	-0.068	-0.036	1.241	1.150	0.262	0.278	0.929	0.802	0.331	0.396	0.411	0.224	-0.052	-0.141
5	1.876	1.834	1.460	1.556	2.555	2.468	1.790	1.870	2.243	2.120	1.859	1.989	1.725	1.542	1.477	1.452
	Item 4				Item 15				Item 21				Item 28			
1	-1.587	-1.636	-1.884	-2.025	-1.233	-1.285	-1.917	-2.123	-1.095	-1.101	-1.611	-1.764	-1.788	-1.941	-2.411	-2.622
2	-0.840	-0.935	-1.249	-1.322	-0.486	-0.584	-1.282	-1.421	-0.348	-0.400	-0.976	-1.062	-1.041	-1.241	-1.776	-1.919
3	-0.171	-0.290	-0.600	-0.622	0.183	0.061	-0.632	-0.720	0.321	0.245	-0.326	-0.361	-0.372	-0.595	-1.126	-1.219
4	0.596	0.466	0.222	0.247	0.950	0.817	0.189	0.149	1.089	1.001	0.495	0.508	0.396	0.160	-0.304	-0.349
5	1.911	1.784	1.750	1.840	2.265	2.135	1.717	1.742	2.403	2.319	2.023	2.101	1.710	1.478	1.224	1.243
	Item 8				Item 16				Item 22				Item 29			
1	-2.322	-2.201	-2.103	-2.365	-1.307	-1.297	-2.123	-2.280	-1.215	-1.310	-1.752	-1.846	-1.711	-1.786	-2.424	-2.616
2	-1.575	-1.500	-1.468	-1.662	-0.560	-0.596	-1.488	-1.577	-0.469	-0.610	-1.117	-1.143	-0.965	-1.086	-1.789	-1.913
3	-0.906	-0.855	-0.819	-0.961	0.109	0.049	-0.838	-0.877	0.201	0.036	-0.468	-0.442	-0.295	-0.440	-1.139	-1.213
4	-0.139	-0.099	0.003	-0.092	0.877	0.805	-0.017	-0.007	0.968	0.791	0.354	0.427	0.472	0.315	-0.317	-0.344
5	1.176	1.219	1.531	1.500	2.191	2.123	1.512	1.585	2.282	2.109	1.882	2.019	1.786	1.633	1.211	1.249

Table E.6(b): Differences between adjacent SIMITNORM Score categories: Model DD

DD	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
	Item 1				Item 9				Item 17				Item 24			
2-1	0.746	0.701	0.635	0.703	0.746	0.701	0.635	0.703	0.746	0.701	0.635	0.703	0.746	0.701	0.635	0.703
3-2	0.670	0.646	0.650	0.700	0.670	0.646	0.650	0.700	0.670	0.646	0.650	0.700	0.670	0.646	0.650	0.700
4-3	0.767	0.755	0.822	0.869	0.767	0.755	0.822	0.869	0.767	0.755	0.822	0.869	0.767	0.755	0.822	0.869
5-4	1.314	1.318	1.528	1.593	1.314	1.318	1.528	1.593	1.314	1.318	1.528	1.593	1.314	1.318	1.528	1.593
	Item 2				Item 10				Item 18				Item 25			
2-1	0.746	0.701	0.635	0.703	0.746	0.701	0.635	0.703	0.746	0.701	0.635	0.703	0.746	0.701	0.635	0.703
3-2	0.670	0.646	0.650	0.700	0.670	0.646	0.650	0.700	0.670	0.646	0.650	0.700	0.670	0.646	0.650	0.700
4-3	0.767	0.755	0.822	0.869	0.767	0.755	0.822	0.869	0.767	0.755	0.822	0.869	0.767	0.755	0.822	0.869
5-4	1.314	1.318	1.528	1.593	1.314	1.318	1.528	1.593	1.314	1.318	1.528	1.593	1.314	1.318	1.528	1.593
	Item 3				Item 15				Item 20				Item 27			
2-1	0.746	0.701	0.635	0.703	0.746	0.701	0.635	0.703	0.746	0.701	0.635	0.703	0.746	0.701	0.635	0.703
3-2	0.670	0.646	0.650	0.700	0.670	0.646	0.650	0.700	0.670	0.646	0.650	0.700	0.670	0.646	0.650	0.700
4-3	0.767	0.755	0.822	0.869	0.767	0.755	0.822	0.869	0.767	0.755	0.822	0.869	0.767	0.755	0.822	0.869
5-4	1.314	1.318	1.528	1.593	1.314	1.318	1.528	1.593	1.314	1.318	1.528	1.593	1.314	1.318	1.528	1.593
	Item 4				Item 15				Item 21				Item 28			
2-1	0.746	0.701	0.635	0.703	0.746	0.701	0.635	0.703	0.746	0.701	0.635	0.703	0.746	0.701	0.635	0.703
3-2	0.670	0.646	0.650	0.700	0.670	0.646	0.650	0.700	0.670	0.646	0.650	0.700	0.670	0.646	0.650	0.700
4-3	0.767	0.755	0.822	0.869	0.767	0.755	0.822	0.869	0.767	0.755	0.822	0.869	0.767	0.755	0.822	0.869
5-4	1.314	1.318	1.528	1.593	1.314	1.318	1.528	1.593	1.314	1.318	1.528	1.593	1.314	1.318	1.528	1.593
	Item 8				Item 16				Item 22				Item 29			
2-1	0.746	0.701	0.635	0.703	0.746	0.701	0.635	0.703	0.746	0.701	0.635	0.703	0.746	0.701	0.635	0.703
3-2	0.670	0.646	0.650	0.700	0.670	0.646	0.650	0.700	0.670	0.646	0.650	0.700	0.670	0.646	0.650	0.700
4-3	0.767	0.755	0.822	0.869	0.767	0.755	0.822	0.869	0.767	0.755	0.822	0.869	0.767	0.755	0.822	0.869
5-4	1.314	1.318	1.528	1.593	1.314	1.318	1.528	1.593	1.314	1.318	1.528	1.593	1.314	1.318	1.528	1.593

Table E.7(a): SIMITNORM Scores: Model DDD

DDD	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
	Item 1				Item 9				Item 17				Item 24			
1	-1.709	-1.924	-2.208	-2.267	-2.438	-2.379	-2.103	-2.173	-2.021	-2.181	-2.359	-2.307	-2.059	-2.175	-2.378	-2.311
2	-1.011	-1.225	-1.510	-1.568	-1.739	-1.681	-1.405	-1.474	-1.322	-1.482	-1.660	-1.609	-1.361	-1.476	-1.679	-1.613
3	-0.342	-0.556	-0.841	-0.899	-1.071	-1.012	-0.736	-0.805	-0.654	-0.814	-0.992	-0.940	-0.692	-0.808	-1.011	-0.944
4	0.462	0.248	-0.037	-0.095	-0.266	-0.208	0.068	-0.001	0.151	-0.009	-0.187	-0.136	0.112	-0.003	-0.206	-0.140
5	1.900	1.686	1.401	1.343	1.172	1.230	1.506	1.437	1.589	1.428	1.251	1.302	1.550	1.435	1.231	1.298
	Item 2				Item 10				Item 18				Item 25			
1	-1.712	-1.739	-2.181	-2.163	-2.514	-2.517	-2.339	-2.350	-2.115	-2.235	-2.211	-2.140	-2.290	-2.438	-2.609	-2.563
2	-1.014	-1.040	-1.482	-1.464	-1.816	-1.819	-1.640	-1.651	-1.416	-1.536	-1.512	-1.441	-1.591	-1.739	-1.910	-1.865
3	-0.345	-0.372	-0.814	-0.796	-1.147	-1.150	-0.972	-0.983	-0.748	-0.868	-0.844	-0.773	-0.923	-1.071	-1.242	-1.196
4	0.459	0.433	-0.009	0.009	-0.343	-0.346	-0.167	-0.178	0.057	-0.063	-0.039	0.032	-0.118	-0.266	-0.437	-0.392
5	1.897	1.871	1.429	1.446	1.095	1.092	1.271	1.259	1.494	1.375	1.398	1.470	1.319	1.172	1.001	1.046
	Item 3				Item 14				Item 20				Item 27			
1	-1.616	-1.641	-2.219	-2.196	-0.928	-0.978	-1.891	-1.901	-1.239	-1.342	-1.831	-1.796	-1.763	-1.945	-2.198	-2.289
2	-0.918	-0.942	-1.520	-1.497	-0.229	-0.279	-1.193	-1.202	-0.540	-0.643	-1.132	-1.098	-1.064	-1.246	-1.500	-1.590
3	-0.249	-0.274	-0.852	-0.829	0.439	0.389	-0.524	-0.534	0.128	0.025	-0.464	-0.429	-0.396	-0.578	-0.831	-0.922
4	0.555	0.531	-0.047	-0.024	1.244	1.194	0.280	0.271	0.933	0.830	0.341	0.375	0.409	0.227	-0.027	-0.117
5	1.993	1.969	1.391	1.414	2.682	2.632	1.718	1.708	2.371	2.267	1.779	1.813	1.847	1.664	1.411	1.321
	Item 4				Item 15				Item 21				Item 28			
1	-1.581	-1.694	-1.932	-1.931	-1.215	-1.323	-1.958	-2.019	-1.079	-1.136	-1.665	-1.690	-1.784	-2.015	-2.444	-2.480
2	-0.882	-0.995	-1.233	-1.232	-0.516	-0.624	-1.259	-1.320	-0.381	-0.438	-0.966	-0.992	-1.086	-1.317	-1.746	-1.782
3	-0.214	-0.327	-0.565	-0.564	0.152	0.044	-0.591	-0.652	0.288	0.231	-0.298	-0.323	-0.417	-0.648	-1.077	-1.113
4	0.591	0.478	0.240	0.241	0.956	0.849	0.214	0.153	1.092	1.035	0.507	0.481	0.387	0.156	-0.273	-0.309
5	2.028	1.916	1.677	1.679	2.394	2.286	1.651	1.591	2.530	2.473	1.944	1.919	1.825	1.594	1.165	1.129
	Item 8				Item 16				Item 22				Item 29			
1	-2.342	-2.294	-2.145	-2.244	-1.292	-1.337	-2.165	-2.166	-1.204	-1.357	-1.807	-1.767	-1.708	-1.854	-2.455	-2.474
2	-1.644	-1.595	-1.446	-1.546	-0.593	-0.639	-1.466	-1.468	-0.505	-0.658	-1.108	-1.069	-1.009	-1.155	-1.756	-1.775
3	-0.975	-0.927	-0.778	-0.877	0.075	0.030	-0.798	-0.799	0.163	0.010	-0.440	-0.400	-0.341	-0.487	-1.088	-1.107
4	-0.171	-0.122	0.027	-0.073	0.880	0.834	0.007	0.005	0.968	0.815	0.365	0.404	0.464	0.318	-0.283	-0.302
5	1.267	1.315	1.464	1.365	2.318	2.272	1.445	1.443	2.405	2.252	1.803	1.842	1.902	1.755	1.155	1.135

Table E.7(b): Differences between adjacent SIMITNORM Score categories: Model DDD

DDD	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho	Eng	Afr	Nguni	Sotho
	Item 1				Item 9				Item 17				Item 24			
2-1	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699
3-2	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668
4-3	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805
5-4	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438
	Item 2				Item 10				Item 18				Item 25			
2-1	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699
3-2	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668
4-3	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805
5-4	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438
	Item 3				Item 14				Item 20				Item 27			
2-1	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699
3-2	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668
4-3	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805
5-4	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438
	Item 4				Item 15				Item 21				Item 28			
2-1	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699
3-2	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668
4-3	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805
5-4	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438
	Item 8				Item 16				Item 22				Item 29			
2-1	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699	0.699
3-2	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668	0.668
4-3	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805	0.805
5-4	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438	1.438

APPENDIX F: SIMITNORM Results Exploratory Factor Analysis

Table F.1: Varimax Rotated Factor Pattern: Raw Items

Variable	Communalities	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
VV2	0.85	0.83						
VV1	0.81	0.83						
VV4	0.76	0.82						
VV3	0.77	0.76						
VV15	0.79		0.83					
VV16	0.81		0.80					
VV14	0.70		0.78					
VV9	0.82			0.89				
VV10	0.77			0.85				
VV8	0.72			0.83				
VV21	0.85				0.82			
VV20	0.81				0.78			
VV22	0.77				0.74			
VV27	0.73					0.78		
VV28	0.78					0.77		
VV29	0.77					0.71		
VV18	0.88						0.91	
VV17	0.86						0.88	
VV25	0.77							0.84
VV24	0.76							0.84
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
Eigenvalues		7.664	2.033	1.567	1.400	1.232	0.975	0.889
% Variation explained		38.3	10.2	7.8	7.0	6.2	4.9	4.4
Cum % Variation explained		38.3	48.5	56.3	63.3	69.5	74.4	78.8
Variation in Factor Space		3.215	2.383	2.375	2.323	2.117	1.785	1.561
% Variation explained		20.4	15.1	15.1	14.7	13.4	11.3	9.9
Cum % Variation explained		20.4	35.5	50.6	65.3	78.8	90.1	100.0

Table F.2: Varimax Rotated Factor Pattern: SIMITNORM Model A

Variable	Communalities	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
A2	0.807	0.832						
A1	0.774	0.825						
A4	0.705	0.790						
A3	0.728	0.768						
A9	0.835		0.883					
A10	0.786		0.852					
A8	0.726		0.822					
A21	0.818			0.829				
A20	0.766			0.784				
A22	0.730			0.746				
A28	0.794				0.818			
A29	0.766				0.774			
A27	0.704				0.769			
A15	0.765					0.831		
A16	0.767					0.791		
A14	0.644					0.767		
A18	0.872						0.900	
A17	0.867						0.890	
A25	0.784							0.857
A24	0.780							0.851
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
Eigenvalues		6.935	2.126	1.584	1.446	1.246	1.097	0.986
% Variation explained		34.7	10.6	7.9	7.2	6.2	5.5	4.9
Cum % Variation explained		34.7	45.3	53.2	60.5	66.7	72.2	77.1
Variation in Factor Space		3.054	2.398	2.226	2.214	2.169	1.785	1.572
% Variation explained		19.8	15.6	14.4	14.4	14.1	11.6	10.2
Cum % Variation explained		19.8	35.4	49.8	64.2	78.2	89.8	100.0

Table F.3: Varimax Rotated Factor Pattern: SIMITNORM Model B

Variable	Communalities	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
B2	0.809	0.832						
B1	0.775	0.826						
B4	0.706	0.791						
B3	0.730	0.768						
B9	0.835		0.882					
B10	0.786		0.851					
B8	0.727		0.822					
B21	0.822			0.829				
B20	0.770			0.784				
B22	0.734			0.747				
B28	0.794				0.815			
B27	0.708				0.771			
B29	0.767				0.771			
B15	0.766					0.832		
B16	0.770					0.792		
B14	0.646					0.768		
B18	0.872						0.901	
B17	0.867						0.891	
B25	0.785							0.858
B24	0.781							0.852
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
Eigenvalues		6.982	2.113	1.582	1.452	1.247	1.094	0.978
% Variation explained		34.9	10.6	7.9	7.3	6.2	5.5	4.9
Cum % Variation explained		34.9	45.5	53.4	60.6	66.9	72.3	77.2
Variation in Factor Space		3.062	2.400	2.233	2.213	2.179	1.786	1.574
% Variation explained		19.8	15.5	14.5	14.3	14.1	11.6	10.2
Cum % Variation explained		19.8	35.4	49.8	64.1	78.2	89.8	100.0

Table F.4: Varimax Rotated Factor Pattern: SIMITNORM Model D

Variable	Communalities	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
D2	0.815	0.836						
D1	0.780	0.827						
D4	0.720	0.796						
D3	0.739	0.772						
D9	0.835		0.883					
D10	0.785		0.851					
D8	0.727		0.822					
D21	0.828			0.830				
D20	0.783			0.790				
D22	0.739			0.749				
D28	0.791				0.812			
D27	0.709				0.773			
D29	0.762				0.765			
D15	0.766					0.833		
D16	0.769					0.789		
D14	0.646					0.768		
D18	0.874						0.901	
D17	0.870						0.892	
D25	0.787							0.856
D24	0.784							0.852
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
Eigenvalues		7.063	2.112	1.592	1.455	1.222	1.082	0.983
% Variation explained		35.3	10.6	8.0	7.3	6.1	5.4	4.9
Cum % Variation explained		35.3	45.9	53.8	61.1	67.2	72.6	77.5
Variation in Factor Space		3.096	2.401	2.256	2.206	2.179	1.792	1.580
% Variation explained		20.0	15.5	14.5	14.2	14.1	11.6	10.2
Cum % Variation explained		20.0	35.4	50.0	64.2	78.3	89.8	100.0

Table F.5: Varimax Rotated Factor Pattern: SIMITNORM Model BB

Variable	Communalities	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
BB2	0.808	0.832						
BB1	0.774	0.825						
BB4	0.706	0.791						
BB3	0.727	0.768						
BB9	0.830		0.879					
BB10	0.781		0.846					
BB8	0.725		0.822					
BB21	0.816			0.828				
BB20	0.764			0.781				
BB22	0.727			0.744				
BB28	0.793				0.814			
BB27	0.705				0.769			
BB29	0.766				0.769			
BB15	0.765					0.831		
BB16	0.767					0.789		
BB14	0.643					0.765		
BB18	0.873						0.900	
BB17	0.868						0.890	
BB25	0.783							0.858
BB24	0.779							0.852
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
Eigenvalues		6.981	2.046	1.585	1.452	1.252	1.088	0.994
% Variation explained		34.9	10.2	7.9	7.3	6.3	5.4	5.0
Cum % Variation explained		34.9	45.1	53.1	60.3	66.6	72.0	77.0
Variation in Factor Space		3.057	2.387	2.221	2.205	2.170	1.787	1.572
% Variation explained		19.9	15.5	14.4	14.3	14.1	11.6	10.2
Cum % Variation explained		19.9	35.4	49.8	64.1	78.2	89.8	100.0

Table F.6: Varimax Rotated Factor Pattern: SIMITNORM Model BBB

Variable		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
BBB2	0.810	0.832						
BBB1	0.775	0.825						
BBB4	0.706	0.791						
BBB3	0.730	0.768						
BBB9	0.830		0.879					
BBB10	0.782		0.845					
BBB8	0.726		0.823					
BBB21	0.818			0.829				
BBB20	0.766			0.781				
BBB22	0.730			0.744				
BBB28	0.792				0.812			
BBB27	0.707				0.769			
BBB29	0.766				0.768			
BBB15	0.766					0.832		
BBB16	0.769					0.791		
BBB14	0.644					0.765		
BBB18	0.873						0.901	
BBB17	0.868						0.891	
BBB25	0.781							0.858
BBB24	0.777							0.851
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
Eigenvalues		7.016	2.032	1.585	1.457	1.254	1.083	0.989
% Variation explained		35.1	10.2	7.9	7.3	6.3	5.4	4.9
Cum % Variation explained		35.1	45.2	53.2	60.4	66.7	72.1	77.1
Variation in Factor Space		3.063	2.391	2.226	2.203	2.175	1.788	1.569
% Variation explained		19.9	15.5	14.4	14.3	14.1	11.6	10.2
Cum % Variation explained		19.9	35.4	49.8	64.1	78.2	89.8	100.0

Table F.7: Varimax Rotated Factor Pattern: SIMITNORM Model DD

Variable	Communalities	Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
DD2	0.812	0.835						
DD1	0.776	0.826						
DD4	0.711	0.793						
DD3	0.733	0.770						
DD9	0.829		0.879					
DD10	0.782		0.846					
DD8	0.724		0.821					
DD21	0.820			0.830				
DD20	0.773			0.786				
DD22	0.732			0.746				
DD28	0.793				0.814			
DD27	0.708				0.772			
DD29	0.764				0.767			
DD15	0.762					0.831		
DD16	0.765					0.788		
DD14	0.642					0.765		
DD18	0.871						0.900	
DD17	0.866						0.890	
DD25	0.783							0.857
DD24	0.780							0.853
		Factor1	Factor2	Factor3	Factor4	Factor5	Factor6	Factor7
Eigenvalues		6.993	2.051	1.582	1.472	1.242	1.087	0.999
% Variation explained		35.0	10.3	7.9	7.4	6.2	5.4	5.0
Cum % Variation explained		35.0	45.2	53.1	60.5	66.7	72.1	77.1
Variation in Factor Space		3.072	2.387	2.238	2.204	2.165	1.784	1.575
% Variation explained		19.9	15.5	14.5	14.3	14.0	11.6	10.2
Cum % Variation explained		19.9	35.4	49.9	64.2	78.2	89.8	100.0

APPENDIX G : Single Group Confirmatory Factor Analysis: Estimated Parameters

Table G.1: Lambda-X estimates Completely Standardised Solutions for *Raw Scores*, *Standardised Z-Scores* and SIMITNORM Models

		Raw scores VV1 – VV29	Z-scores Z1 - Z29	Model A A1 - A29	Model B B1 - B29	Model D D1 - D29	Model BB BB1 - BB29	Model BBB BBB1 - BBB29	Model DD DD1 - DD29	Model DDD DDD1 - DDD29
Entertainment	VV1	0.86	0.84	0.83	0.83	0.84	0.83	0.83	0.83	0.84
	VV2	0.91	0.90	0.88	0.88	0.88	0.88	0.88	0.88	0.88
	VV3	0.85	0.82	0.81	0.81	0.82	0.81	0.81	0.81	0.82
	VV4	0.79	0.76	0.75	0.75	0.77	0.76	0.75	0.76	0.77
Confusion	VV8	0.73	0.73	0.74	0.74	0.74	0.74	0.74	0.74	0.74
	VV9	0.87	0.87	0.89	0.89	0.89	0.89	0.89	0.89	0.89
	VV10	0.82	0.83	0.83	0.83	0.83	0.83	0.83	0.83	0.83
Relevant	VV14	0.68	0.60	0.60	0.60	0.60	0.60	0.60	0.60	0.60
News	VV15	0.82	0.81	0.81	0.81	0.80	0.81	0.81	0.80	0.80
	VV16	0.89	0.86	0.85	0.85	0.86	0.85	0.85	0.85	0.85
Brand	VV17	0.95	0.89	0.90	0.90	0.90	0.90	0.90	0.90	0.90
Reinforcement	VV18	0.77	0.82	0.83	0.83	0.84	0.83	0.83	0.83	0.84
Empathy	VV20	0.85	0.82	0.81	0.81	0.83	0.81	0.81	0.82	0.82
	VV21	0.87	0.85	0.84	0.84	0.85	0.84	0.84	0.84	0.85
	VV22	0.81	0.80	0.78	0.79	0.79	0.78	0.78	0.78	0.78
Familiarity	VV24	0.72	0.72	0.76	0.76	0.76	0.76	0.75	0.75	0.76
	VV25	0.73	0.70	0.75	0.75	0.75	0.74	0.74	0.75	0.75
Alienation	VV27	0.69	0.69	0.71	0.71	0.71	0.71	0.71	0.71	0.70
	VV28	0.82	0.81	0.83	0.83	0.83	0.83	0.83	0.83	0.82
	VV29	0.86	0.85	0.84	0.84	0.84	0.84	0.84	0.84	0.84

Table G.2: Estimated PHI Matrix Parameters for *Raw Scores*, *Standardised Z-Scores* and SIMITNORM Models

		VV1-VV29	Z1-Z29	A1-A29	B1-B29	D1-D29	BB1-BB29	BBB1-BBB29	DD1-DD29	DDD1-DDD29
Entertainment	Entertainment	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Confusion		0.33	0.38	0.34	0.34	0.34	0.35	0.35	0.35	0.35
Relevant News		0.55	0.49	0.50	0.50	0.50	0.50	0.50	0.50	0.50
Brand Reinforcement		0.32	0.31	0.32	0.32	0.32	0.32	0.32	0.32	0.32
Empathy		0.69	0.66	0.65	0.65	0.65	0.65	0.65	0.65	0.65
Familiarity		0.42	0.34	0.34	0.34	0.36	0.34	0.34	0.35	0.37
Alienation		0.61	0.56	0.50	0.50	0.51	0.51	0.51	0.51	0.52
Confusion	Confusion	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Relevant News		0.21	0.30	0.25	0.25	0.25	0.27	0.27	0.27	0.27
Brand Reinforcement		0.16	0.19	0.19	0.19	0.19	0.19	0.19	0.19	0.19
Empathy		0.31	0.39	0.33	0.34	0.33	0.35	0.36	0.35	0.35
Familiarity		0.25	0.28	0.35	0.35	0.35	0.33	0.33	0.32	0.33
Alienation		0.43	0.51	0.48	0.48	0.48	0.49	0.49	0.49	0.49
Relevant News	Relevant News	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Brand Reinforcement		0.36	0.34	0.35	0.35	0.35	0.36	0.35	0.35	0.35
Empathy		0.59	0.51	0.52	0.52	0.52	0.51	0.51	0.51	0.52
Familiarity		0.27	0.17	0.17	0.17	0.18	0.18	0.18	0.18	0.19
Alienation		0.62	0.54	0.47	0.47	0.48	0.48	0.49	0.48	0.49
Brand Reinforcement	Brand Reinforcement	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Empathy		0.38	0.38	0.39	0.39	0.39	0.39	0.39	0.39	0.39
Familiarity		0.23	0.21	0.22	0.22	0.23	0.22	0.22	0.22	0.23
Alienation		0.45	0.45	0.43	0.43	0.43	0.43	0.43	0.43	0.44
Empathy	Empathy	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Familiarity		0.35	0.26	0.26	0.25	0.28	0.27	0.27	0.28	0.29
Alienation		0.66	0.61	0.56	0.57	0.57	0.57	0.57	0.57	0.58
Familiarity	Familiarity	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Alienation		0.44	0.37	0.43	0.42	0.44	0.42	0.41	0.42	0.43
Alienation	Alienation	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table G.3: THETA-EPS Estimated Parameters for *Raw Scores*, *Standardised Z-Scores* and SIMITNORM Models

	VV1 - VV29	Z1 - Z20	A1 - A29	B1 - B29	D1 - D29	BB1 - BB29	BBB1 - BBB29	DD1 - DD29	DDD1 - DDD29
Item 1	0.264	0.289	0.312	0.311	0.303	0.311	0.309	0.309	0.300
Item 2	0.175	0.193	0.229	0.227	0.222	0.228	0.225	0.224	0.219
Item 3	0.281	0.320	0.347	0.343	0.332	0.348	0.345	0.339	0.332
Item 4	0.378	0.416	0.432	0.431	0.409	0.431	0.431	0.425	0.410
Item 8	0.466	0.467	0.448	0.448	0.447	0.454	0.452	0.454	0.451
Item 9	0.246	0.248	0.202	0.202	0.202	0.214	0.215	0.216	0.213
Item 10	0.322	0.319	0.311	0.311	0.313	0.313	0.311	0.312	0.314
Item 14	0.544	0.639	0.640	0.635	0.637	0.640	0.638	0.642	0.641
Item 15	0.324	0.352	0.348	0.350	0.354	0.350	0.351	0.362	0.358
Item 16	0.212	0.255	0.278	0.271	0.269	0.276	0.273	0.275	0.271
Item 17	0.090	0.201	0.194	0.195	0.191	0.193	0.195	0.193	0.190
Item 18	0.408	0.326	0.308	0.307	0.302	0.307	0.304	0.314	0.300
Item 20	0.276	0.327	0.347	0.341	0.318	0.347	0.344	0.331	0.322
Item 21	0.247	0.286	0.292	0.287	0.271	0.301	0.297	0.293	0.285
Item 22	0.339	0.367	0.391	0.384	0.382	0.394	0.389	0.389	0.385
Item 24	0.484	0.489	0.429	0.427	0.425	0.427	0.431	0.432	0.428
Item 25	0.468	0.517	0.445	0.444	0.435	0.451	0.456	0.442	0.444
Item 27	0.525	0.520	0.495	0.493	0.498	0.499	0.497	0.498	0.504
Item 28	0.328	0.338	0.311	0.311	0.316	0.314	0.314	0.314	0.321
Item 29	0.255	0.281	0.300	0.298	0.300	0.298	0.297	0.299	0.300

APPENDIX H : Multi-Group Confirmatory Factor Analysis Group Fit Statistics

Table H.1: Group Fit Measures: Multi-group CFA on *Raw Scores*

RAW SCORES VV	Hypothesis	English	Afrikaans	Nguni	Sotho
Contribution to Chi-Square	Hypothesis 1	367.7	523.0	549.9	364.8
	Hypothesis 2	412.5	605.3	607.5	421.4
	Hypothesis 3	631.3	814.7	961.6	675.9
	Hypothesis 4	885.0	1168.9	1296.8	1332.2
	Hypothesis 5	1723.9	1871.1	2117.0	2003.3
Percentage Contribution to Chi-Square	Hypothesis 1	20.4	29.0	30.5	20.2
	Hypothesis 2	20.2	29.6	29.7	20.6
	Hypothesis 3	20.5	26.4	31.2	21.9
	Hypothesis 4	18.9	25.0	27.7	28.4
	Hypothesis 5	22.3	24.3	27.4	26.0
Root Mean Square Residual (RMR)	Hypothesis 1	0.031	0.038	0.030	0.024
	Hypothesis 2	0.041	0.053	0.038	0.031
	Hypothesis 3	0.128	0.148	0.114	0.157
	Hypothesis 4	0.130	0.151	0.117	0.166
	Hypothesis 5	0.162	0.147	0.123	0.168
Standardized RMR	Hypothesis 1	0.026	0.031	0.032	0.030
	Hypothesis 2	0.033	0.043	0.041	0.037
	Hypothesis 3	0.122	0.134	0.111	0.153
	Hypothesis 4	0.128	0.140	0.116	0.155
	Hypothesis 5	0.149	0.128	0.113	0.147
Goodness of Fit Index (GFI)	Hypothesis 1	0.971	0.958	0.961	0.965
	Hypothesis 2	0.967	0.951	0.957	0.960
	Hypothesis 3	0.948	0.932	0.939	0.940
	Hypothesis 4	0.932	0.908	0.918	0.875
	Hypothesis 5	0.901	0.878	0.904	0.857

Table H.2: Group Fit Measures: Multi-group CFA on *Standardised Z-Scores*

Standardised Scores Z	Hypothesis	English	Afrikaans	Nguni	Sotho
Contribution to Chi-Square	Hypothesis 1	367.7	523.0	549.9	364.8
	Hypothesis 2	397.5	570.5	613.6	417.1
	Hypothesis 3	505.5	705.1	805.6	533.1
	Hypothesis 4	871.0	1142.9	1020.7	774.6
	Hypothesis 5	871.0	1142.9	1020.7	774.6
Percentage Contribution to Chi-Square	Hypothesis 1	20.4	29.0	30.5	20.2
	Hypothesis 2	19.9	28.5	30.7	20.9
	Hypothesis 3	19.8	27.7	31.6	20.9
	Hypothesis 4	22.9	30.0	26.8	20.3
	Hypothesis 5	22.9	30.0	26.8	20.3
Root Mean Square Residual (RMR)	Hypothesis 1	0.026	0.031	0.032	0.030
	Hypothesis 2	0.028	0.035	0.043	0.041
	Hypothesis 3	0.062	0.068	0.080	0.083
	Hypothesis 4	0.064	0.069	0.072	0.075
	Hypothesis 5	0.064	0.069	0.072	0.075
Standardized RMR	Hypothesis 1	0.026	0.031	0.032	0.030
	Hypothesis 2	0.028	0.034	0.043	0.039
	Hypothesis 3	0.066	0.071	0.075	0.078
	Hypothesis 4	0.064	0.069	0.072	0.075
	Hypothesis 5	0.064	0.069	0.072	0.075
Goodness of Fit Index (GFI)	Hypothesis 1	0.971	0.958	0.961	0.965
	Hypothesis 2	0.968	0.954	0.956	0.960
	Hypothesis 3	0.960	0.943	0.945	0.951
	Hypothesis 4	0.932	0.912	0.931	0.928
	Hypothesis 5	0.932	0.912	0.931	0.928

Table H.3: Group Fit Measures: Multi-group CFA on *SIMITNORM A-Scores*

SIMITNORM Model A	Hypothesis	English	Afrikaans	Nguni	Sotho
Contribution to Chi-Square	Hypothesis 1	405.1	496.4	481.4	373.9
	Hypothesis 2	428.6	526.6	527.7	420.7
	Hypothesis 3	527.2	658.4	698.2	535.9
	Hypothesis 4	971.4	1125.2	969.1	803.8
	Hypothesis 5	971.4	1125.2	969.1	803.8
Percentage Contribution to Chi-Square	Hypothesis 1	23.1	28.3	27.4	21.3
	Hypothesis 2	22.5	27.7	27.7	22.1
	Hypothesis 3	21.8	27.2	28.9	22.1
	Hypothesis 4	25.1	29.1	25.0	20.8
	Hypothesis 5	25.1	29.1	25.0	20.8
Root Mean Square Residual (RMR)	Hypothesis 1	0.027	0.032	0.030	0.030
	Hypothesis 2	0.029	0.036	0.038	0.038
	Hypothesis 3	0.064	0.071	0.075	0.081
	Hypothesis 4	0.064	0.071	0.068	0.074
	Hypothesis 5	0.064	0.071	0.068	0.074
Standardized RMR	Hypothesis 1	0.027	0.032	0.030	0.030
	Hypothesis 2	0.029	0.035	0.038	0.037
	Hypothesis 3	0.067	0.074	0.070	0.076
	Hypothesis 4	0.064	0.071	0.067	0.074
	Hypothesis 5	0.064	0.071	0.067	0.074
Goodness of Fit Index (GFI)	Hypothesis 1	0.968	0.959	0.966	0.964
	Hypothesis 2	0.966	0.957	0.962	0.960
	Hypothesis 3	0.959	0.946	0.952	0.950
	Hypothesis 4	0.924	0.911	0.936	0.926
	Hypothesis 5	0.924	0.911	0.936	0.926

Table H.4: Group Fit Measures: Multi-group CFA on *SIMITNORM B-Scores*

SIMITNORM Model B	Hypothesis	English	Afrikaans	Nguni	Sotho
Contribution to Chi-Square	Hypothesis 1	396.5	498.0	483.9	366.9
	Hypothesis 2	419.4	529.5	533.2	413.6
	Hypothesis 3	519.9	654.9	709.1	524.4
	Hypothesis 4	921.9	1096.6	951.9	782.0
	Hypothesis 5	921.9	1096.6	951.9	782.0
Percentage Contribution to Chi-Square	Hypothesis 1	22.7	28.5	27.7	21.0
	Hypothesis 2	22.1	27.9	28.1	21.8
	Hypothesis 3	21.6	27.2	29.4	21.8
	Hypothesis 4	24.6	29.2	25.4	20.8
	Hypothesis 5	24.6	29.2	25.4	20.8
Root Mean Square Residual (RMR)	Hypothesis 1	0.026	0.032	0.029	0.030
	Hypothesis 2	0.028	0.035	0.038	0.038
	Hypothesis 3	0.063	0.070	0.075	0.080
	Hypothesis 4	0.063	0.070	0.068	0.073
	Hypothesis 5	0.063	0.070	0.068	0.073
Standardized RMR	Hypothesis 1	0.026	0.032	0.029	0.030
	Hypothesis 2	0.028	0.035	0.038	0.037
	Hypothesis 3	0.066	0.073	0.070	0.075
	Hypothesis 4	0.063	0.070	0.068	0.073
	Hypothesis 5	0.063	0.070	0.068	0.073
Goodness of Fit Index (GFI)	Hypothesis 1	0.969	0.959	0.966	0.965
	Hypothesis 2	0.967	0.957	0.962	0.960
	Hypothesis 3	0.959	0.947	0.951	0.951
	Hypothesis 4	0.928	0.914	0.937	0.928
	Hypothesis 5	0.928	0.914	0.937	0.928

Table H.5: Group Fit Measures: Multi-group CFA on *SIMITNORM D-Scores*

SIMITNORM Model D	Hypothesis	English	Afrikaans	Nguni	Sotho
Contribution to Chi-Square	Hypothesis 1	395.5	494.0	483.9	367.6
	Hypothesis 2	424.5	542.3	523.7	420.7
	Hypothesis 3	555.3	732.3	780.1	614.2
	Hypothesis 4	762.5	979.5	980.9	890.9
	Hypothesis 5	762.5	979.5	980.9	890.9
Percentage Contribution to Chi-Square	Hypothesis 1	22.7	28.4	27.8	21.1
	Hypothesis 2	22.2	28.4	27.4	22.0
	Hypothesis 3	20.7	27.3	29.1	22.9
	Hypothesis 4	21.1	27.1	27.1	24.7
	Hypothesis 5	21.1	27.1	27.1	24.7
Root Mean Square Residual (RMR)	Hypothesis 1	0.028	0.036	0.028	0.025
	Hypothesis 2	0.033	0.045	0.034	0.031
	Hypothesis 3	0.092	0.121	0.093	0.124
	Hypothesis 4	0.092	0.123	0.090	0.125
	Hypothesis 5	0.092	0.123	0.090	0.125
Standardized RMR	Hypothesis 1	0.026	0.032	0.029	0.030
	Hypothesis 2	0.031	0.040	0.036	0.036
	Hypothesis 3	0.094	0.121	0.089	0.125
	Hypothesis 4	0.092	0.123	0.090	0.125
	Hypothesis 5	0.092	0.123	0.090	0.125
Goodness of Fit Index (GFI)	Hypothesis 1	0.969	0.960	0.966	0.965
	Hypothesis 2	0.967	0.956	0.963	0.960
	Hypothesis 3	0.956	0.938	0.949	0.945
	Hypothesis 4	0.941	0.920	0.937	0.916
	Hypothesis 5	0.941	0.920	0.937	0.916

Table H.6: Group Fit Measures: Multi-group CFA on *SIMITNORM BB-Scores*

SIMITNORM Model BB	Hypothesis	English	Afrikaans	Nguni	Sotho
Contribution to Chi-Square	Hypothesis 1	394.1	497.0	476.7	367.8
	Hypothesis 2	418.0	528.9	523.4	414.9
	Hypothesis 3	514.3	662.5	695.0	525.1
	Hypothesis 4	944.3	1140.6	961.6	794.9
	Hypothesis 5	944.3	1140.6	961.6	794.9
Percentage Contribution to Chi-Square	Hypothesis 1	22.7	28.6	27.5	21.2
	Hypothesis 2	22.2	28.1	27.8	22.0
	Hypothesis 3	21.5	27.6	29.0	21.9
	Hypothesis 4	24.6	29.7	25.0	20.7
	Hypothesis 5	24.6	29.7	25.0	20.7
Root Mean Square Residual (RMR)	Hypothesis 1	0.026	0.031	0.030	0.030
	Hypothesis 2	0.028	0.034	0.038	0.038
	Hypothesis 3	0.063	0.071	0.075	0.081
	Hypothesis 4	0.063	0.071	0.068	0.074
	Hypothesis 5	0.063	0.071	0.068	0.074
Standardized RMR	Hypothesis 1	0.026	0.031	0.030	0.030
	Hypothesis 2	0.028	0.034	0.038	0.037
	Hypothesis 3	0.067	0.075	0.070	0.076
	Hypothesis 4	0.063	0.071	0.068	0.074
	Hypothesis 5	0.063	0.071	0.068	0.074
Goodness of Fit Index (GFI)	Hypothesis 1	0.969	0.960	0.966	0.965
	Hypothesis 2	0.967	0.957	0.963	0.960
	Hypothesis 3	0.960	0.946	0.952	0.951
	Hypothesis 4	0.927	0.911	0.936	0.927
	Hypothesis 5	0.927	0.911	0.936	0.927

Table H.7: Group Fit Measures: Multi-group CFA on *SIMITNORM BBB-Scores*

SIMITNORM Model BBB	Hypothesis	English	Afrikaans	Nguni	Sotho
Contribution to Chi-Square	Hypothesis 1	392.6	491.7	489.0	366.6
	Hypothesis 2	416.3	524.3	537.8	413.6
	Hypothesis 3	515.3	655.1	716.2	523.1
	Hypothesis 4	923.7	1109.8	962.1	787.5
	Hypothesis 5	923.7	1109.8	962.1	787.5
Percentage Contribution to Chi-Square	Hypothesis 1	22.6	28.3	28.1	21.1
	Hypothesis 2	22.0	27.7	28.4	21.9
	Hypothesis 3	21.4	27.2	29.7	21.7
	Hypothesis 4	24.4	29.3	25.4	20.8
	Hypothesis 5	24.4	29.3	25.4	20.8
Root Mean Square Residual (RMR)	Hypothesis 1	0.026	0.031	0.030	0.030
	Hypothesis 2	0.028	0.034	0.039	0.038
	Hypothesis 3	0.063	0.070	0.076	0.081
	Hypothesis 4	0.063	0.070	0.069	0.074
	Hypothesis 5	0.063	0.070	0.069	0.074
Standardized RMR	Hypothesis 1	0.026	0.031	0.030	0.030
	Hypothesis 2	0.028	0.034	0.039	0.037
	Hypothesis 3	0.067	0.074	0.071	0.076
	Hypothesis 4	0.063	0.070	0.069	0.074
	Hypothesis 5	0.063	0.070	0.069	0.074
Goodness of Fit Index (GFI)	Hypothesis 1	0.969	0.960	0.966	0.965
	Hypothesis 2	0.967	0.957	0.962	0.960
	Hypothesis 3	0.960	0.947	0.951	0.951
	Hypothesis 4	0.928	0.913	0.936	0.927
	Hypothesis 5	0.928	0.913	0.936	0.927

Table H.8: Group Fit Measures: Multi-group CFA on *SIMITNORM DD-Scores*

SIMITNORM Model DD	Hypothesis	English	Afrikaans	Nguni	Sotho
Contribution to Chi-Square	Hypothesis 1	393.7	494.3	475.3	367.9
	Hypothesis 2	428.5	551.9	516.1	427.1
	Hypothesis 3	548.5	729.4	739.9	568.5
	Hypothesis 4	967.5	1165.2	1003.7	868.4
	Hypothesis 5	967.5	1165.2	1003.7	868.4
Percentage Contribution to Chi-Square	Hypothesis 1	22.7	28.6	27.5	21.3
	Hypothesis 2	22.3	28.7	26.8	22.2
	Hypothesis 3	21.2	28.2	28.6	22.0
	Hypothesis 4	24.2	29.1	25.1	21.7
	Hypothesis 5	24.2	29.1	25.1	21.7
Root Mean Square Residual (RMR)	Hypothesis 1	0.026	0.031	0.030	0.030
	Hypothesis 2	0.032	0.041	0.037	0.038
	Hypothesis 3	0.080	0.089	0.084	0.095
	Hypothesis 4	0.081	0.091	0.080	0.092
	Hypothesis 5	0.081	0.091	0.080	0.092
Standardized RMR	Hypothesis 1	0.026	0.031	0.030	0.030
	Hypothesis 2	0.032	0.040	0.037	0.036
	Hypothesis 3	0.083	0.092	0.080	0.089
	Hypothesis 4	0.081	0.091	0.081	0.091
	Hypothesis 5	0.081	0.091	0.081	0.091
Goodness of Fit Index (GFI)	Hypothesis 1	0.969	0.960	0.967	0.965
	Hypothesis 2	0.966	0.955	0.964	0.959
	Hypothesis 3	0.957	0.940	0.951	0.947
	Hypothesis 4	0.924	0.905	0.936	0.920
	Hypothesis 5	0.924	0.905	0.936	0.920

Table H.9: Group Fit Measures: Multi-group CFA on *SIMITNORM DDD-Scores*

SIMITNORM Model DDD	Hypothesis	English	Afrikaans	Nguni	Sotho
Contribution to Chi-Square	Hypothesis 1	392.1	488.6	485.6	367.1
	Hypothesis 2	425.1	543.7	526.9	423.7
	Hypothesis 3	556.2	729.7	780.2	607.0
	Hypothesis 4	803.6	1000.7	1012.9	919.4
	Hypothesis 5	803.6	1000.7	1012.9	919.4
Percentage Contribution to Chi-Square	Hypothesis 1	22.6	28.2	28.0	21.2
	Hypothesis 2	22.1	28.3	27.5	22.1
	Hypothesis 3	20.8	27.3	29.2	22.7
	Hypothesis 4	21.5	26.8	27.1	24.6
	Hypothesis 5	21.5	26.8	27.1	24.6
Root Mean Square Residual (RMR)	Hypothesis 1	0.027	0.035	0.029	0.026
	Hypothesis 2	0.033	0.045	0.035	0.032
	Hypothesis 3	0.092	0.120	0.092	0.124
	Hypothesis 4	0.092	0.121	0.090	0.125
	Hypothesis 5	0.092	0.121	0.090	0.125
Standardized RMR	Hypothesis 1	0.026	0.031	0.030	0.030
	Hypothesis 2	0.031	0.040	0.037	0.036
	Hypothesis 3	0.094	0.120	0.089	0.123
	Hypothesis 4	0.092	0.121	0.091	0.124
	Hypothesis 5	0.092	0.121	0.091	0.124
Goodness of Fit Index (GFI)	Hypothesis 1	0.969	0.960	0.966	0.965
	Hypothesis 2	0.967	0.956	0.963	0.959
	Hypothesis 3	0.956	0.939	0.949	0.945
	Hypothesis 4	0.938	0.919	0.935	0.915
	Hypothesis 5	0.938	0.919	0.935	0.915

APPENDIX I: Item Bias Analysis

Figure I.1: Item Bias Analysis - Item 1

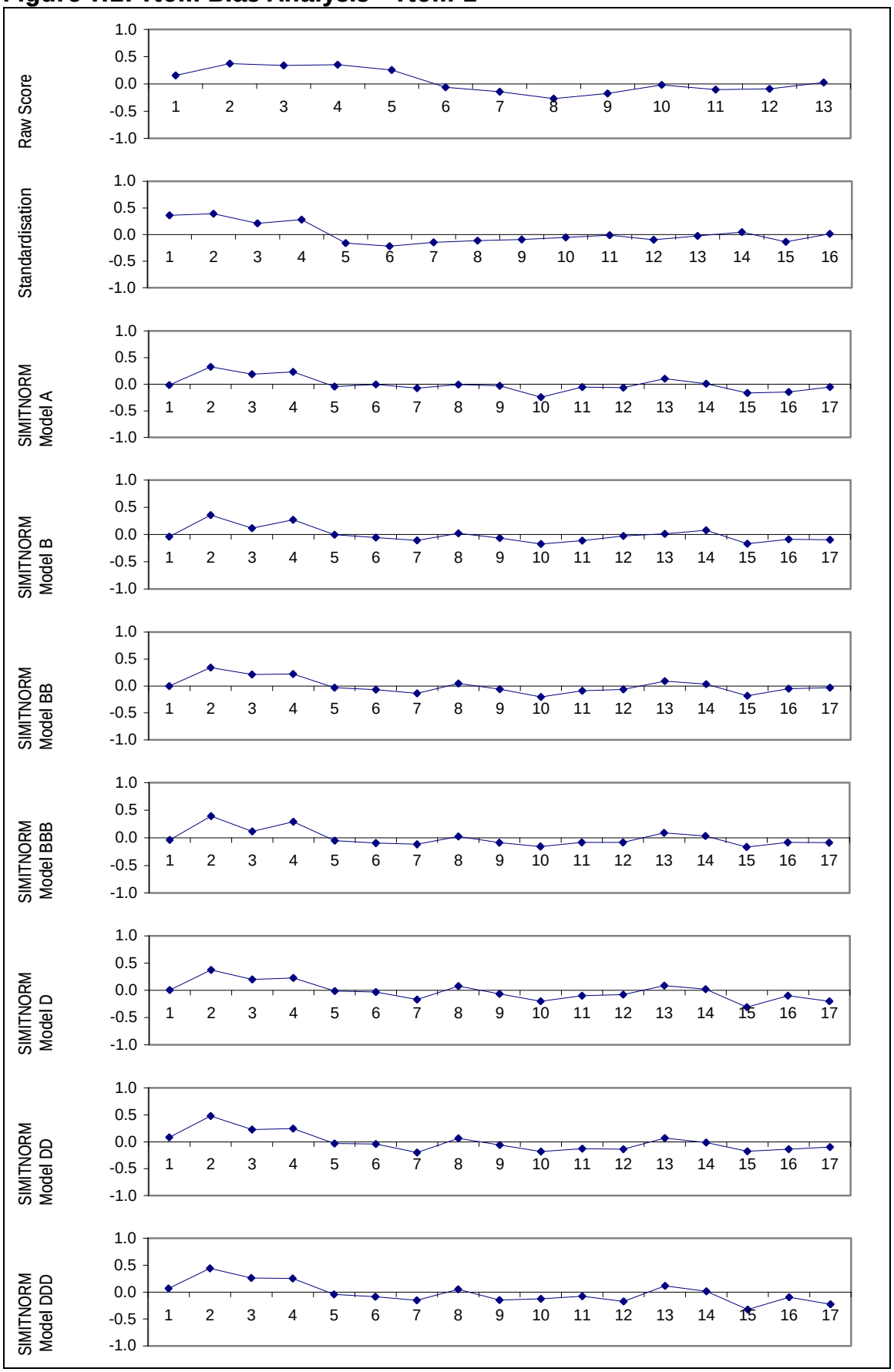


Figure I.2: Item Bias Analysis - Item 2

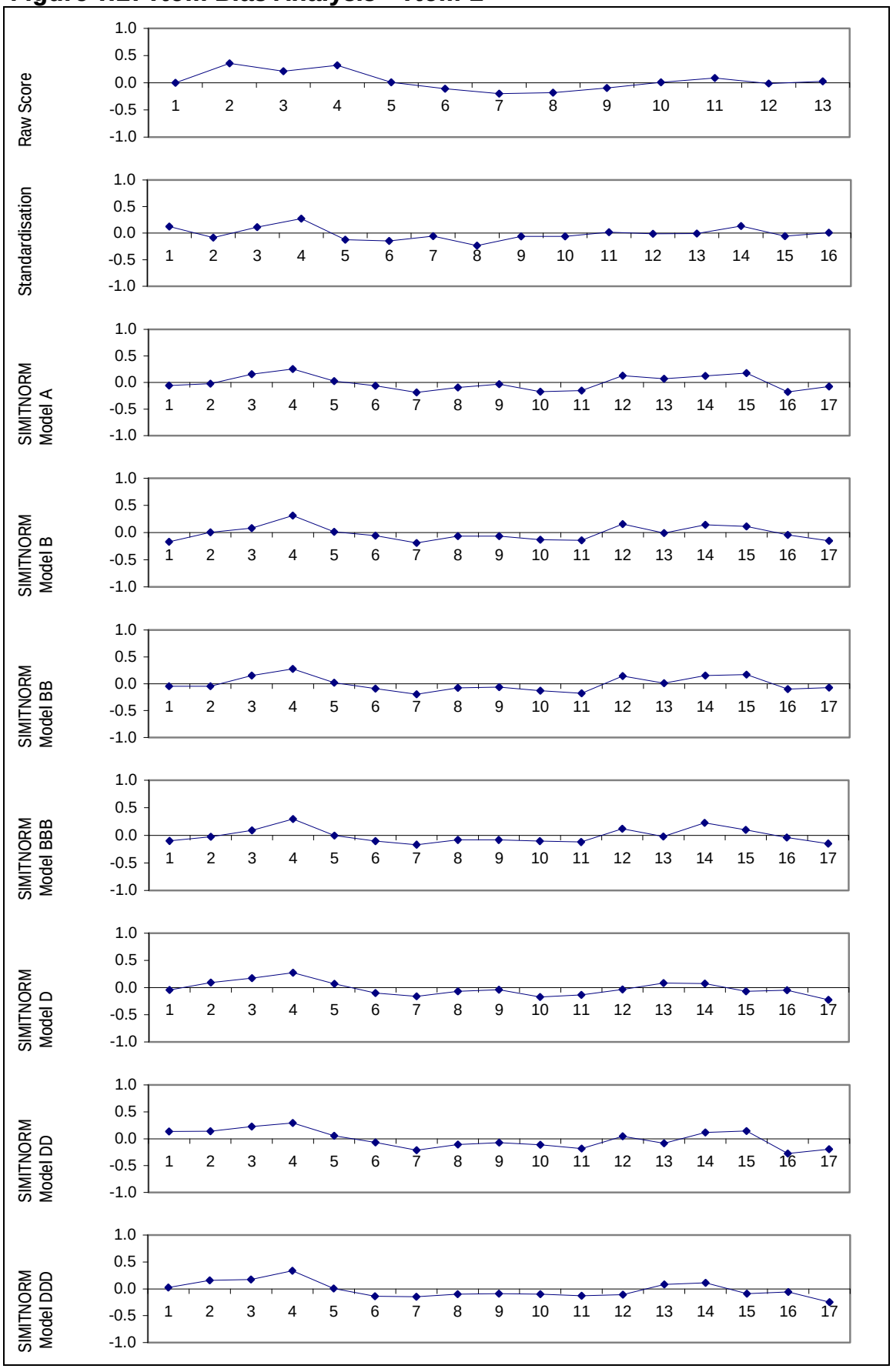


Figure I.3: Item Bias Analysis - Item 3

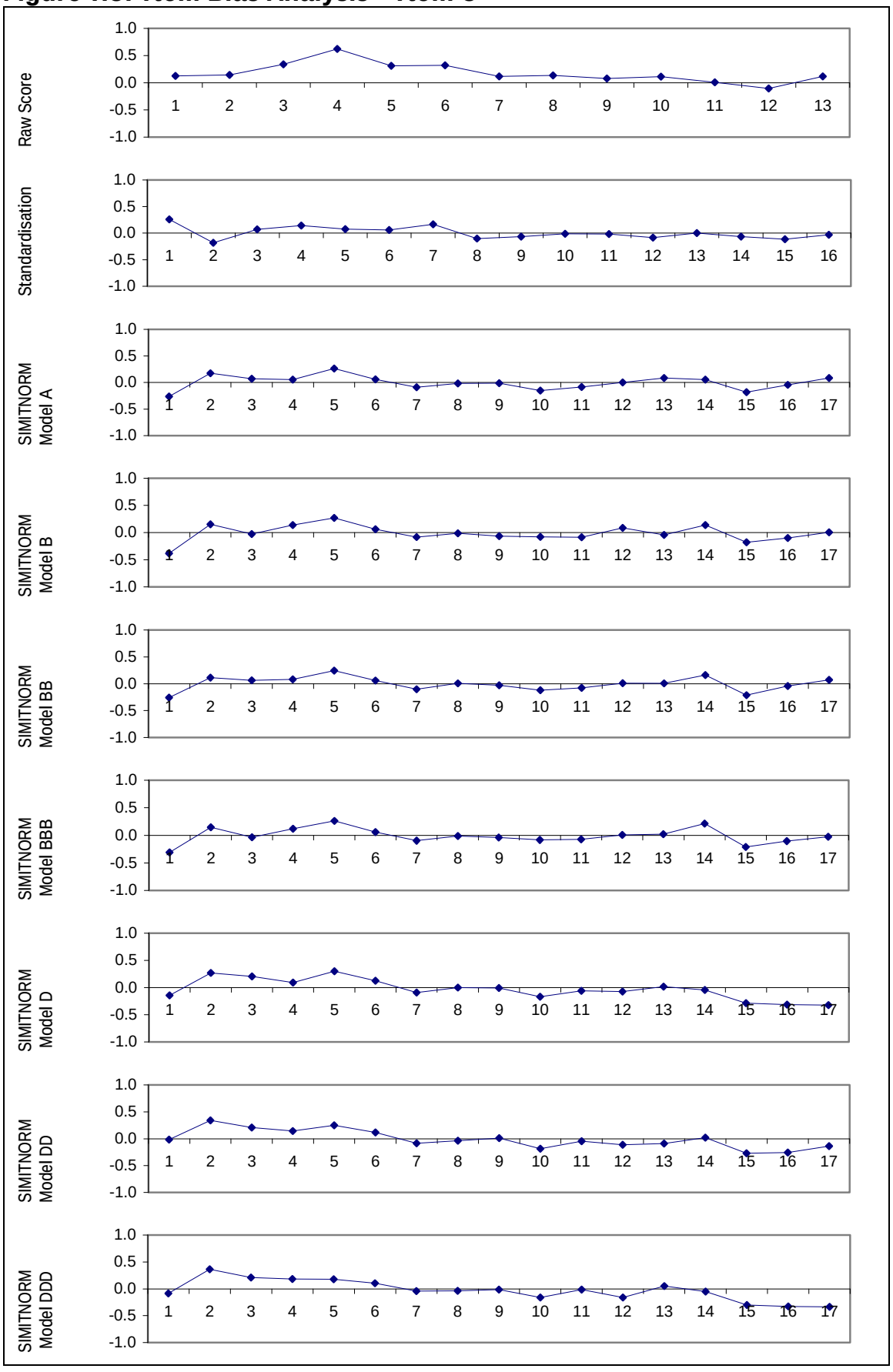


Figure I.4: Item Bias Analysis - Item 4

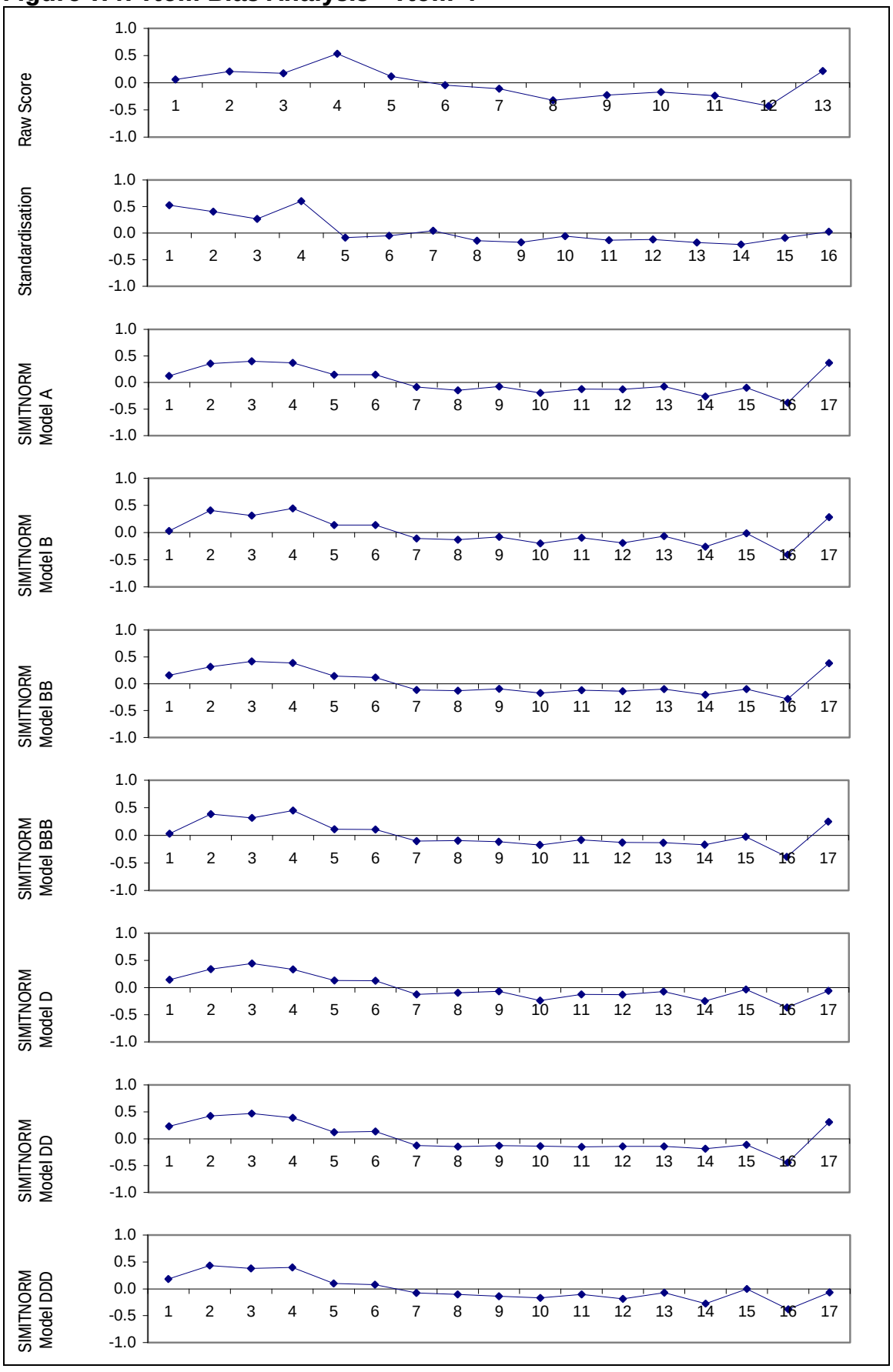


Figure I.5: Item Bias Analysis - Item 8

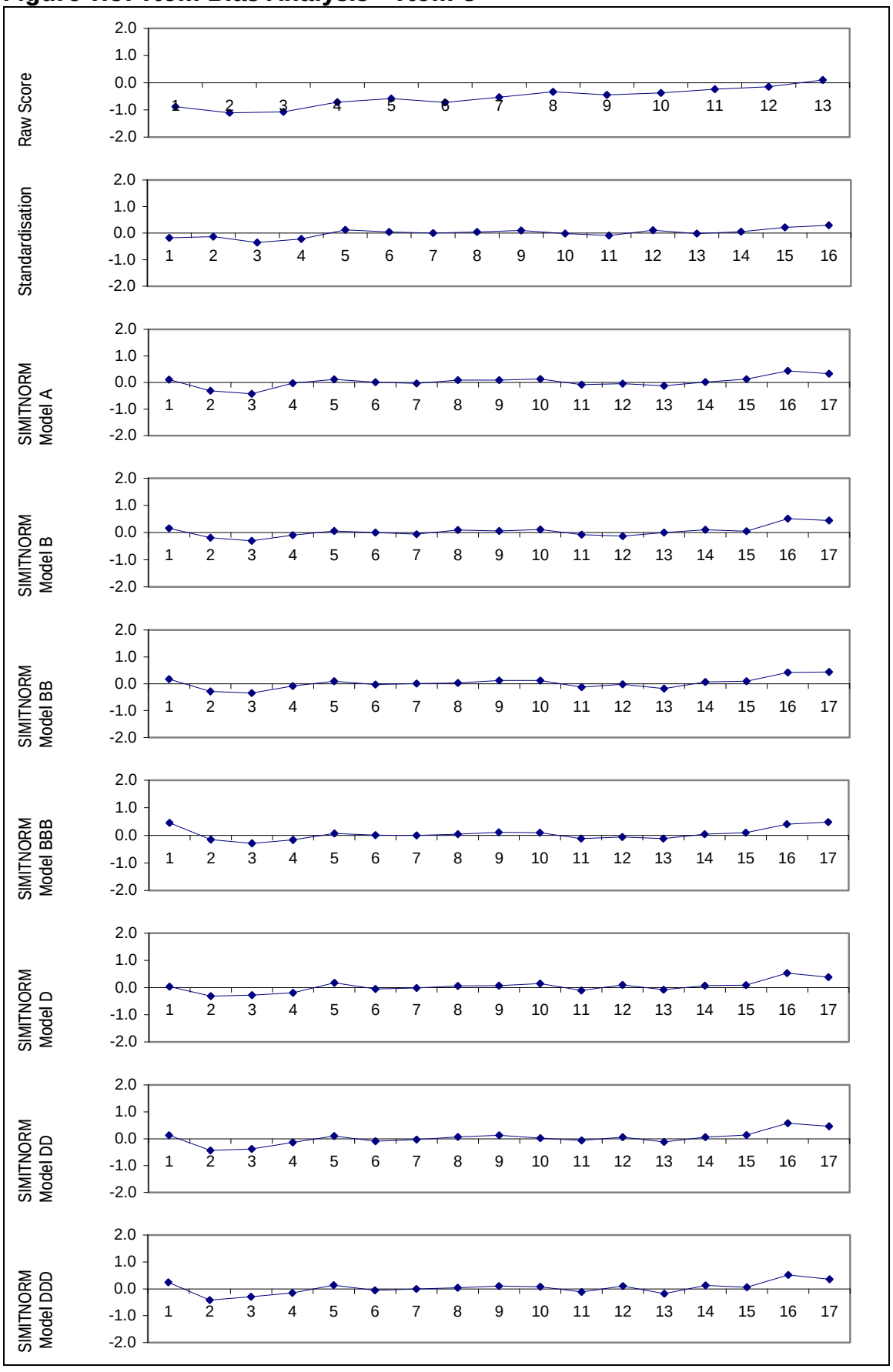


Figure I.6: Item Bias Analysis - Item 9

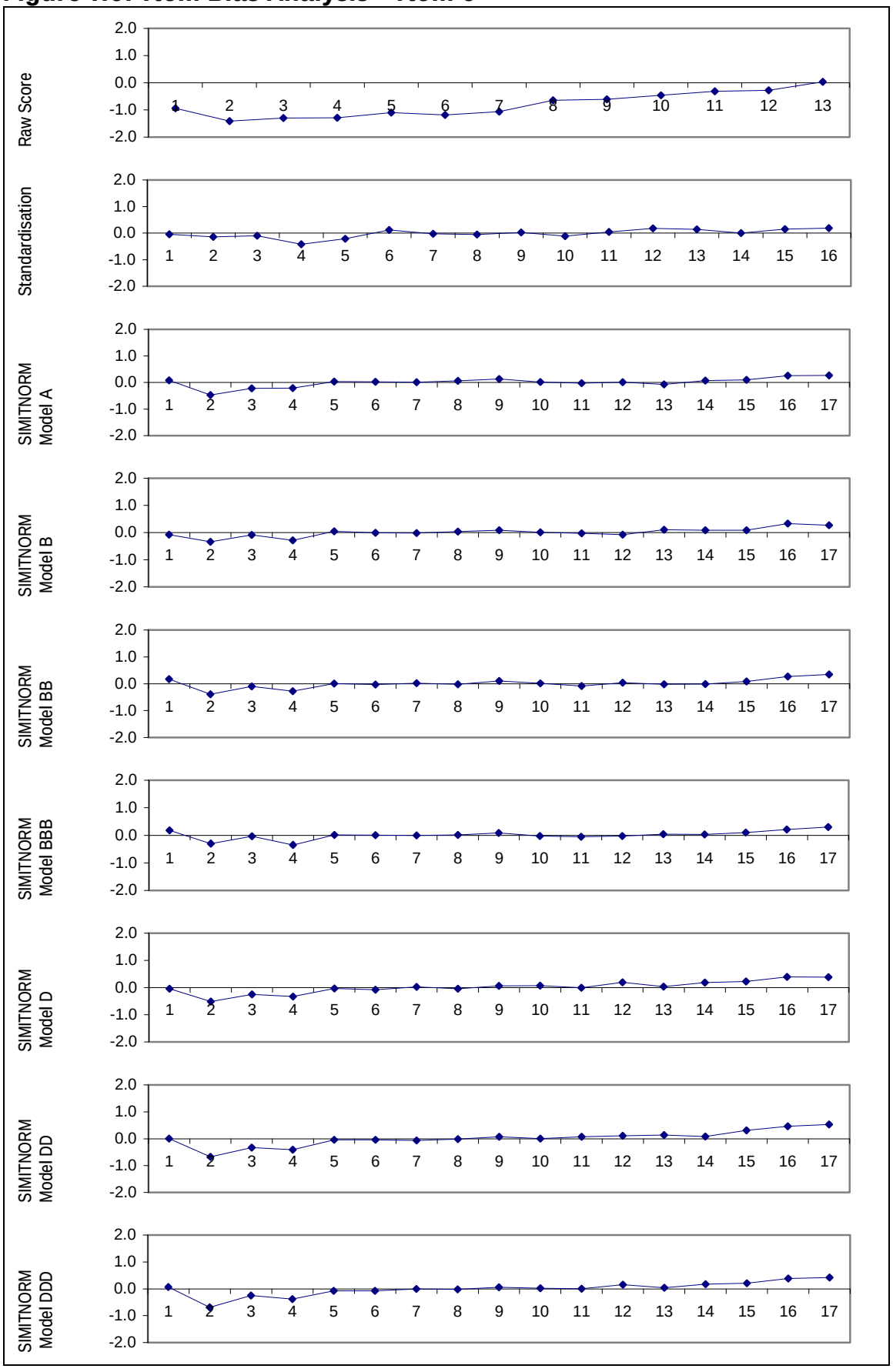


Figure I.7: Item Bias Analysis - Item 10

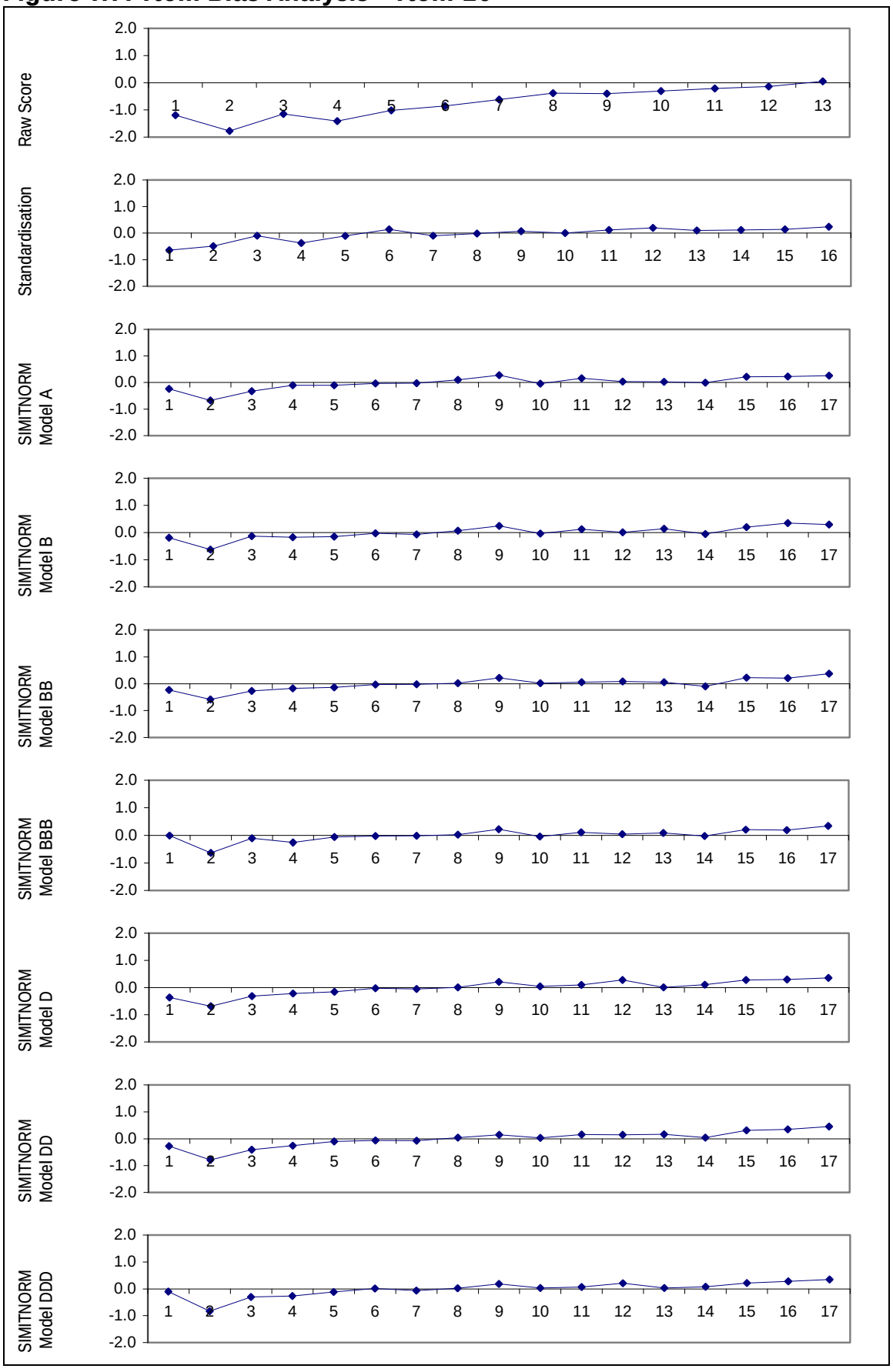


Figure I.8: Item Bias Analysis - Item 14

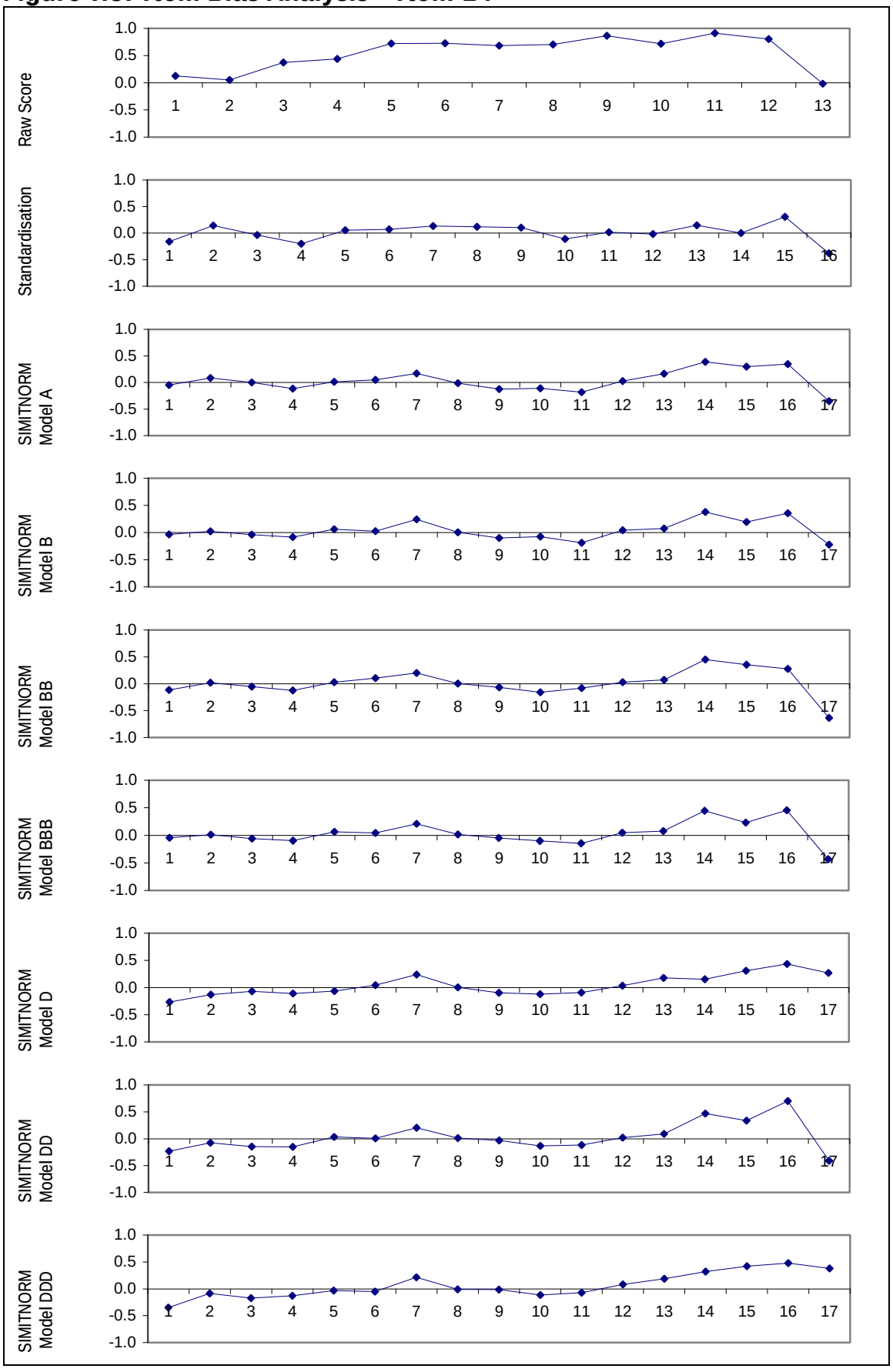


Figure I.9: Item Bias Analysis - Item 15

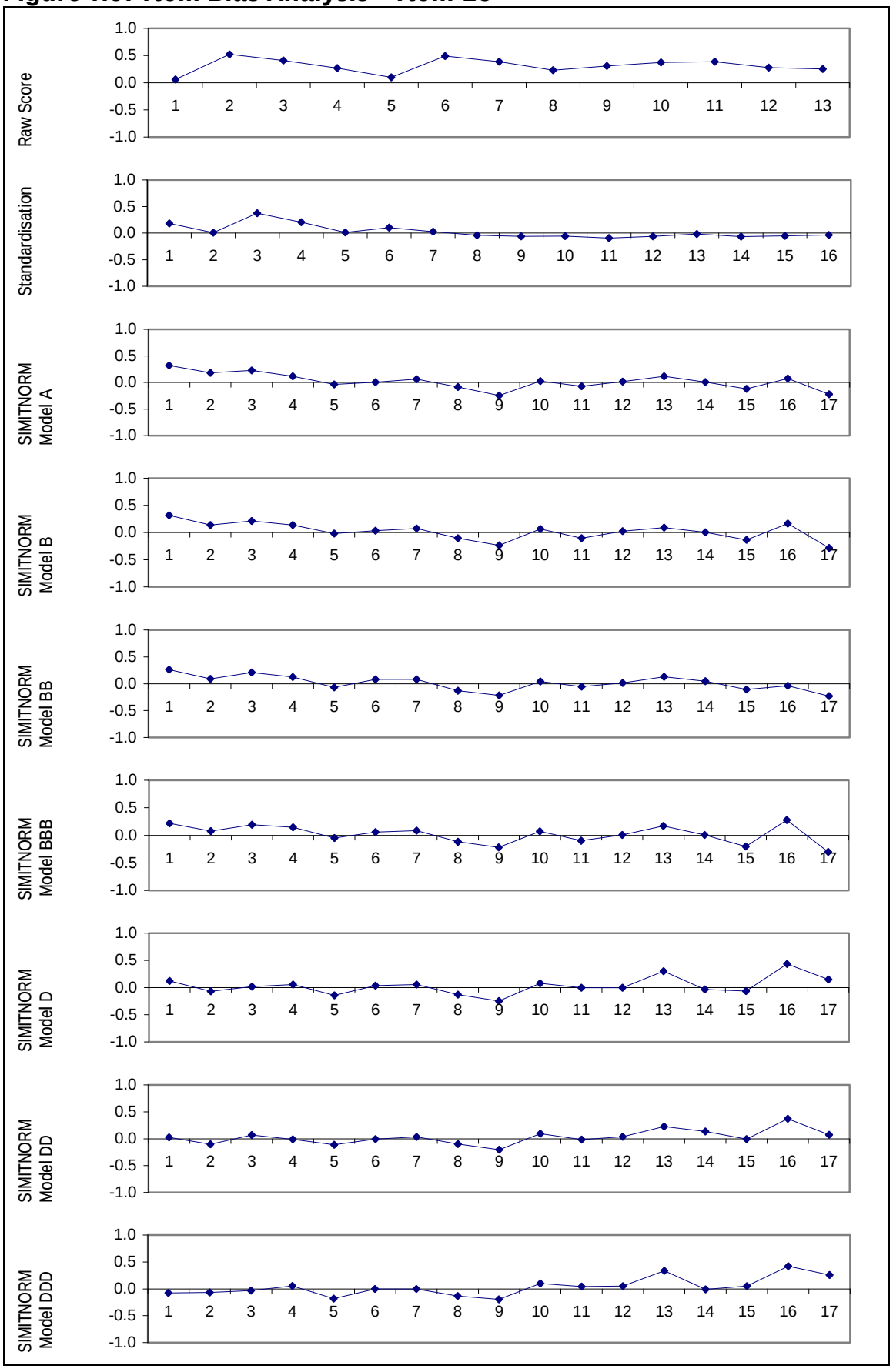


Figure I.10: Item Bias Analysis - Item 16

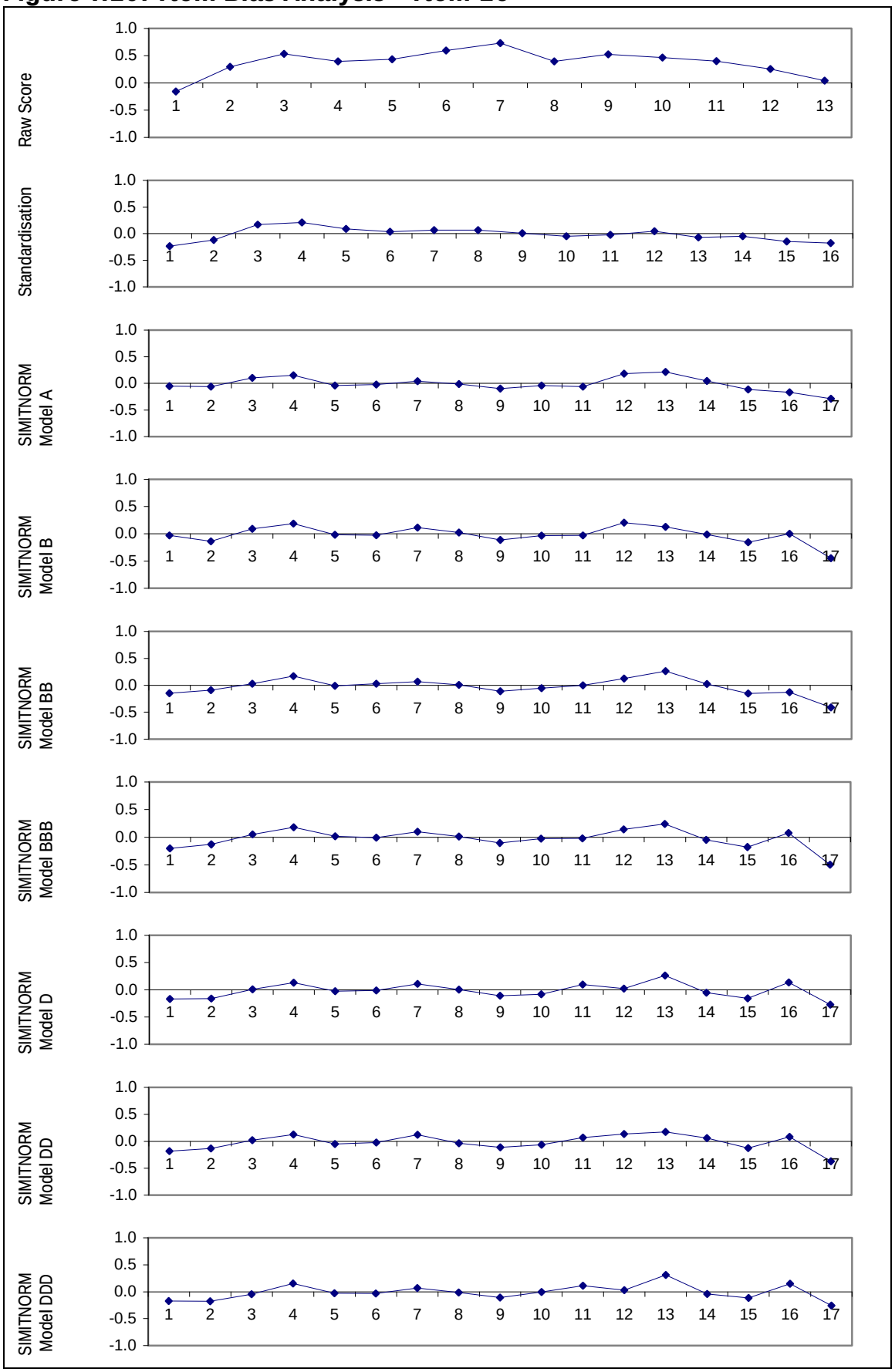


Figure I.11: Item Bias Analysis - Item 17

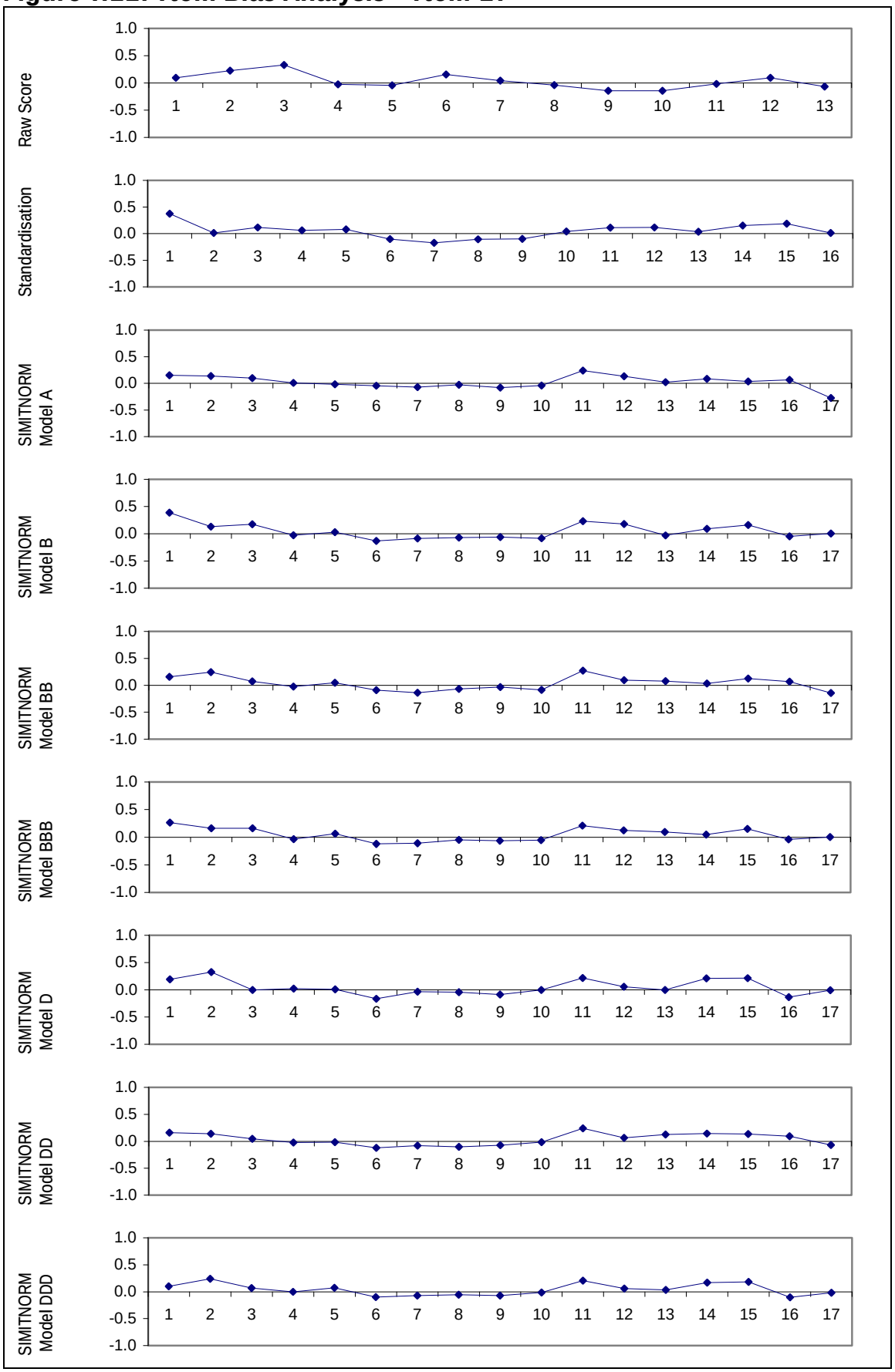


Figure I.12: Item Bias Analysis - Item 18

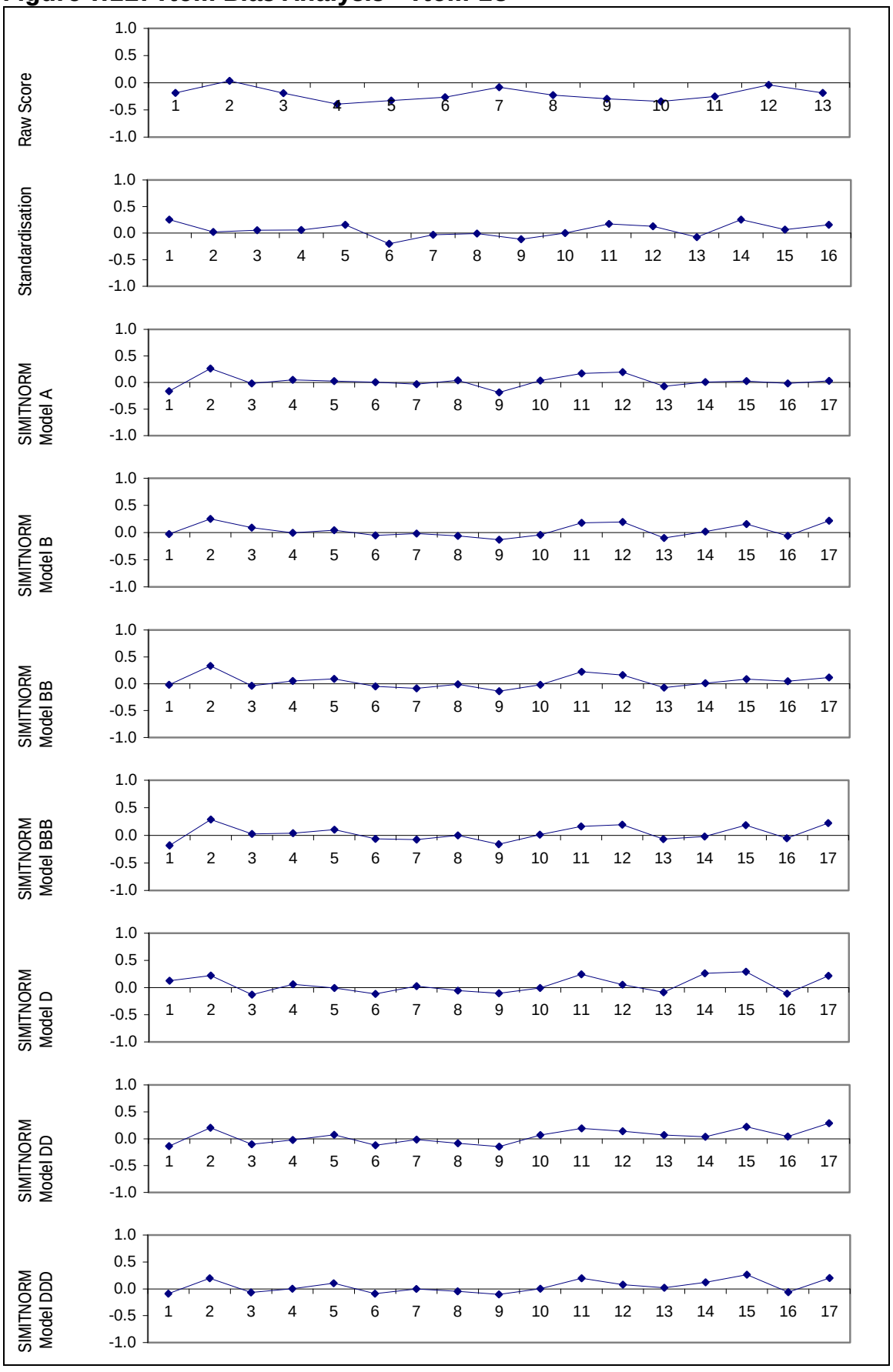


Figure I.13: Item Bias Analysis - Item 20

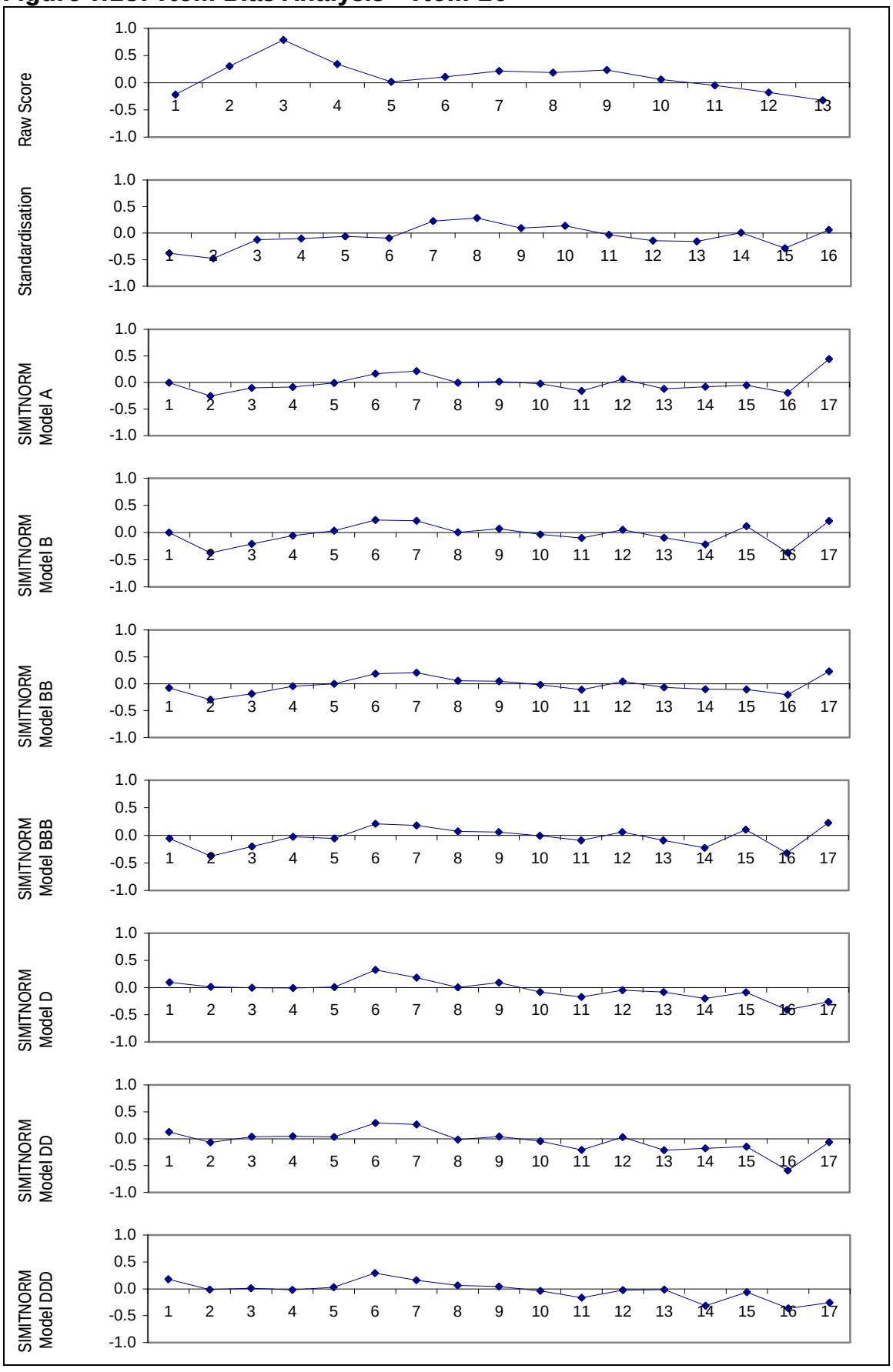


Figure I.14: Item Bias Analysis - Item 21

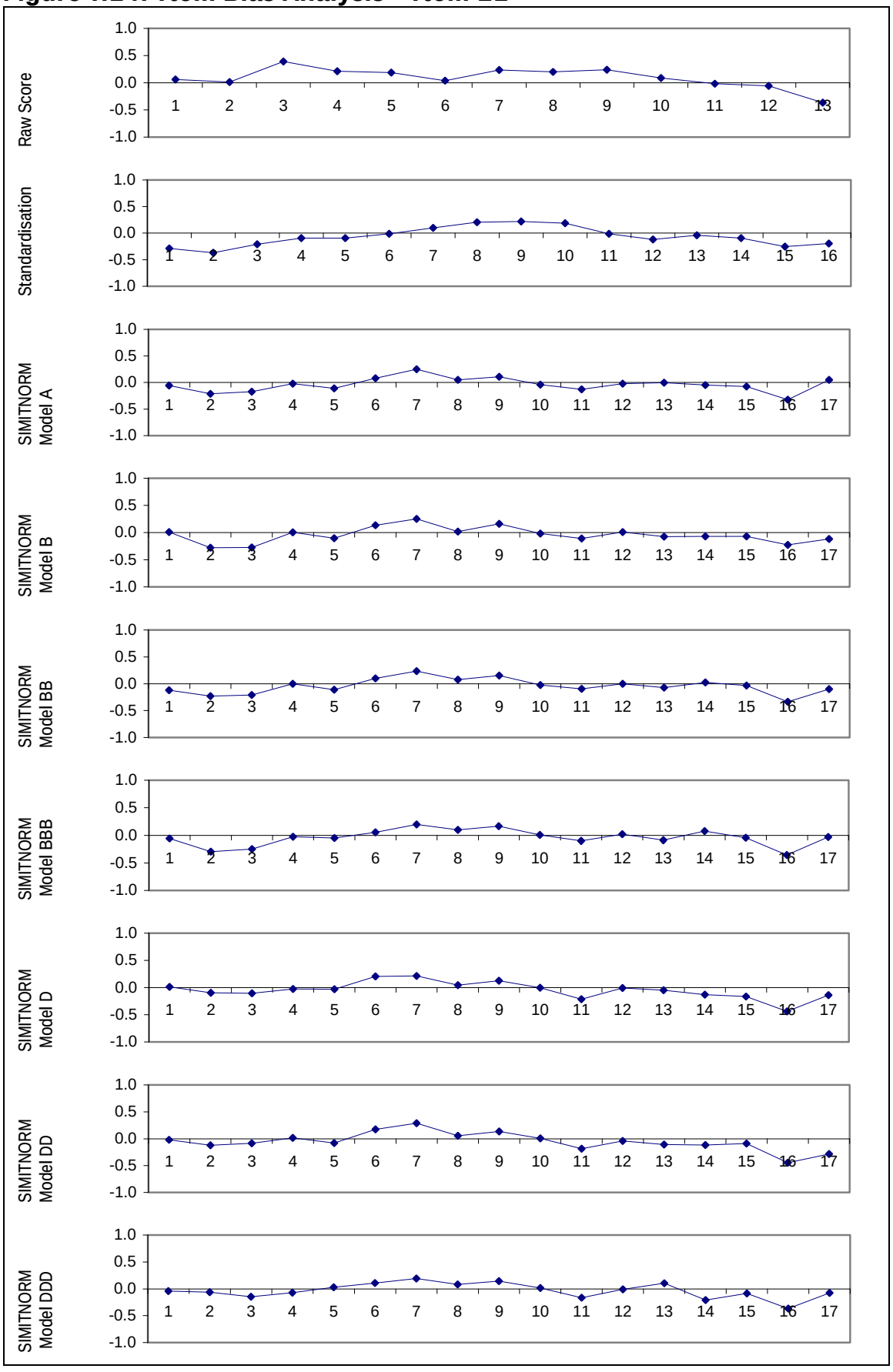


Figure I.15: Item Bias Analysis - Item 22

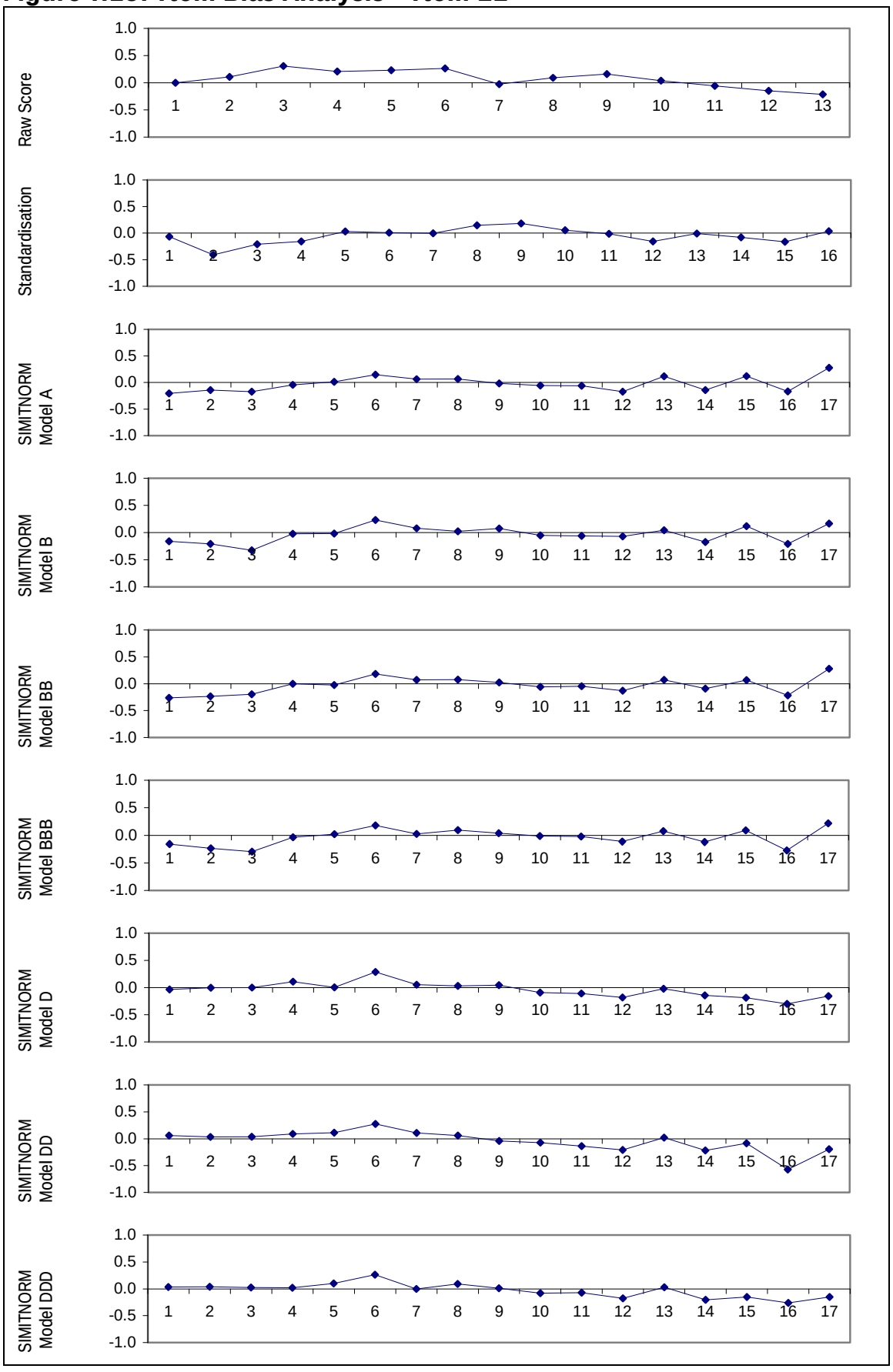


Figure I.16: Item Bias Analysis - Item 24

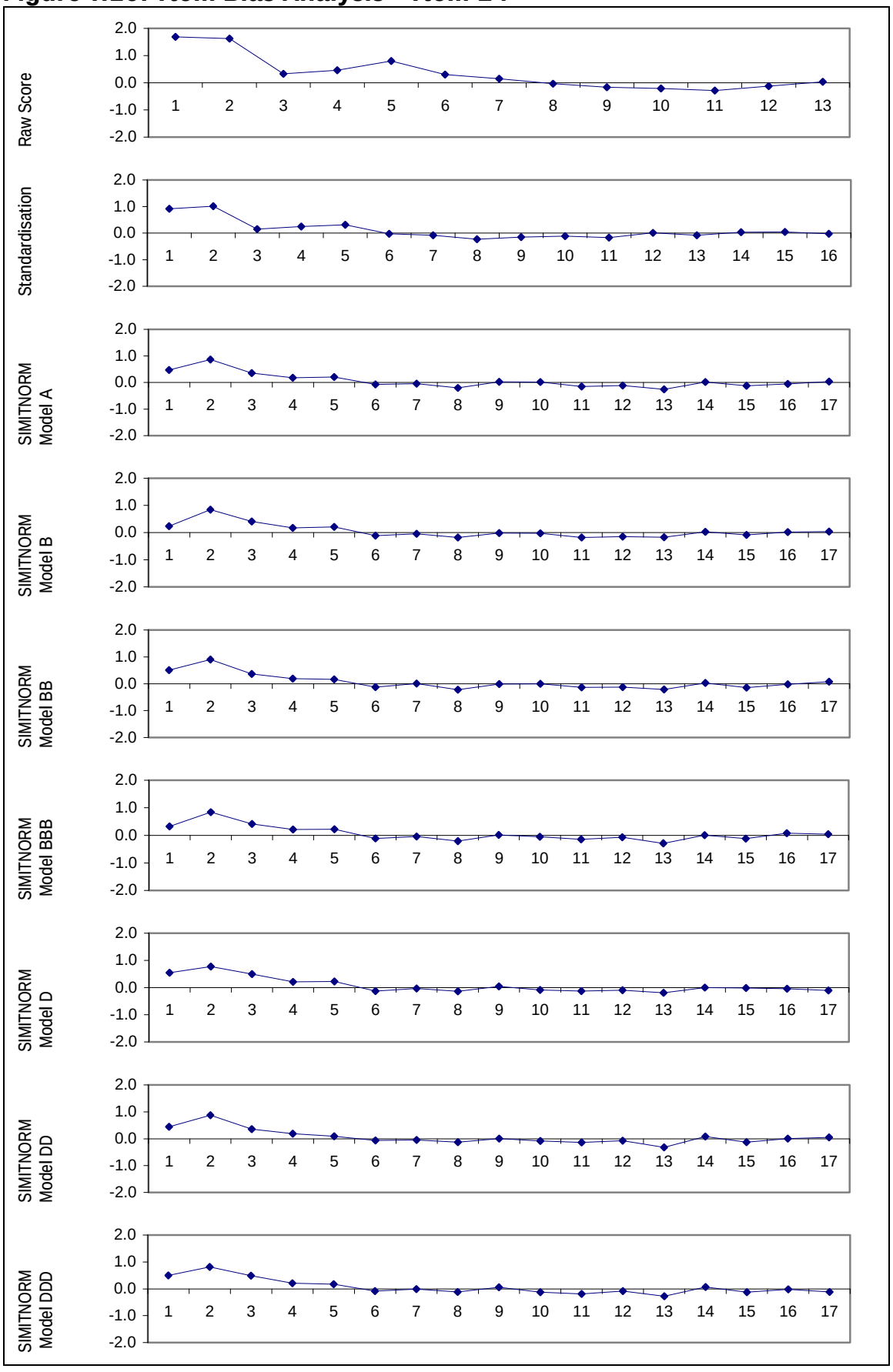


Figure I.17: Item Bias Analysis - Item 25

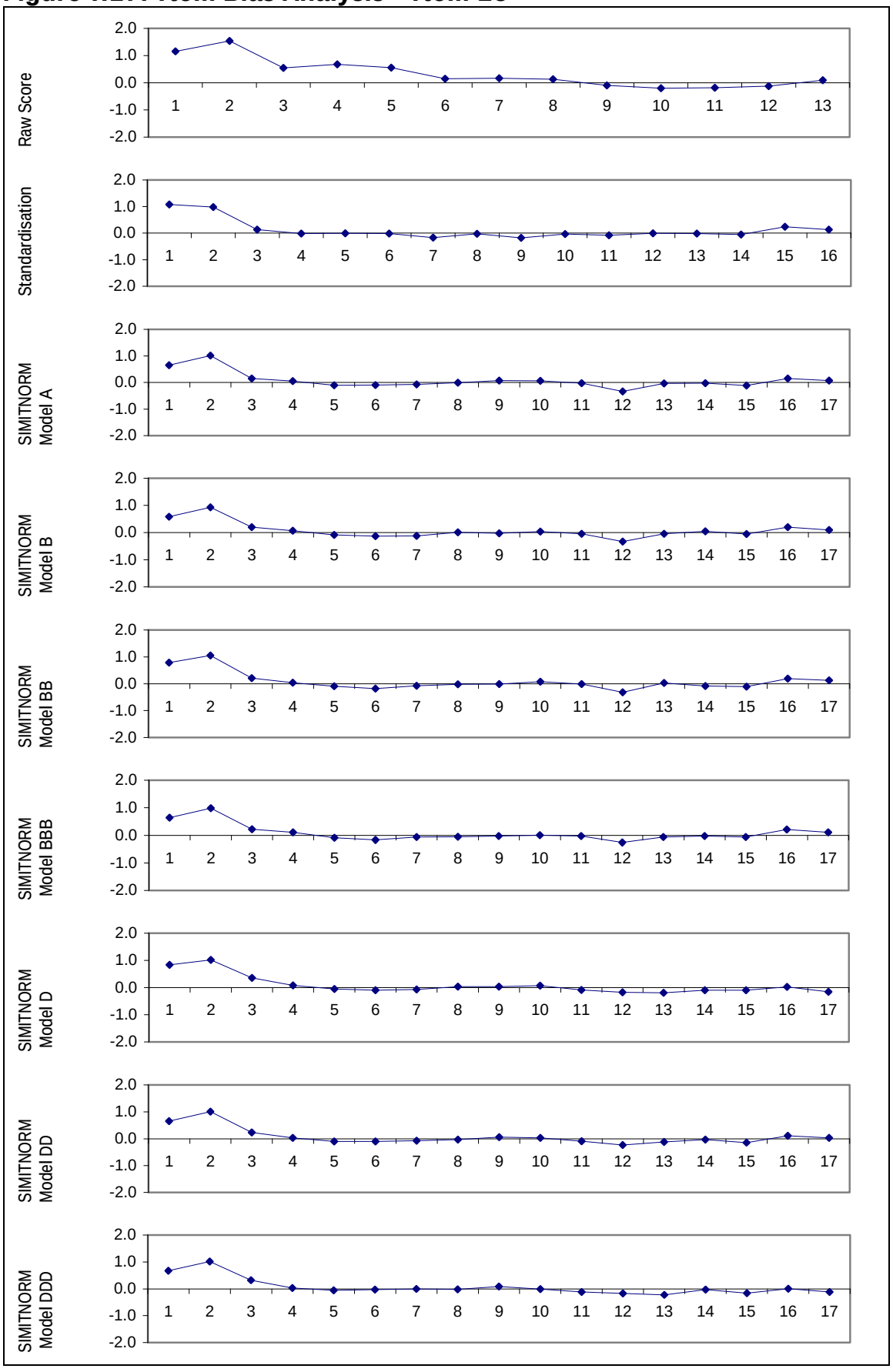


Figure I.18: Item Bias Analysis - Item 27

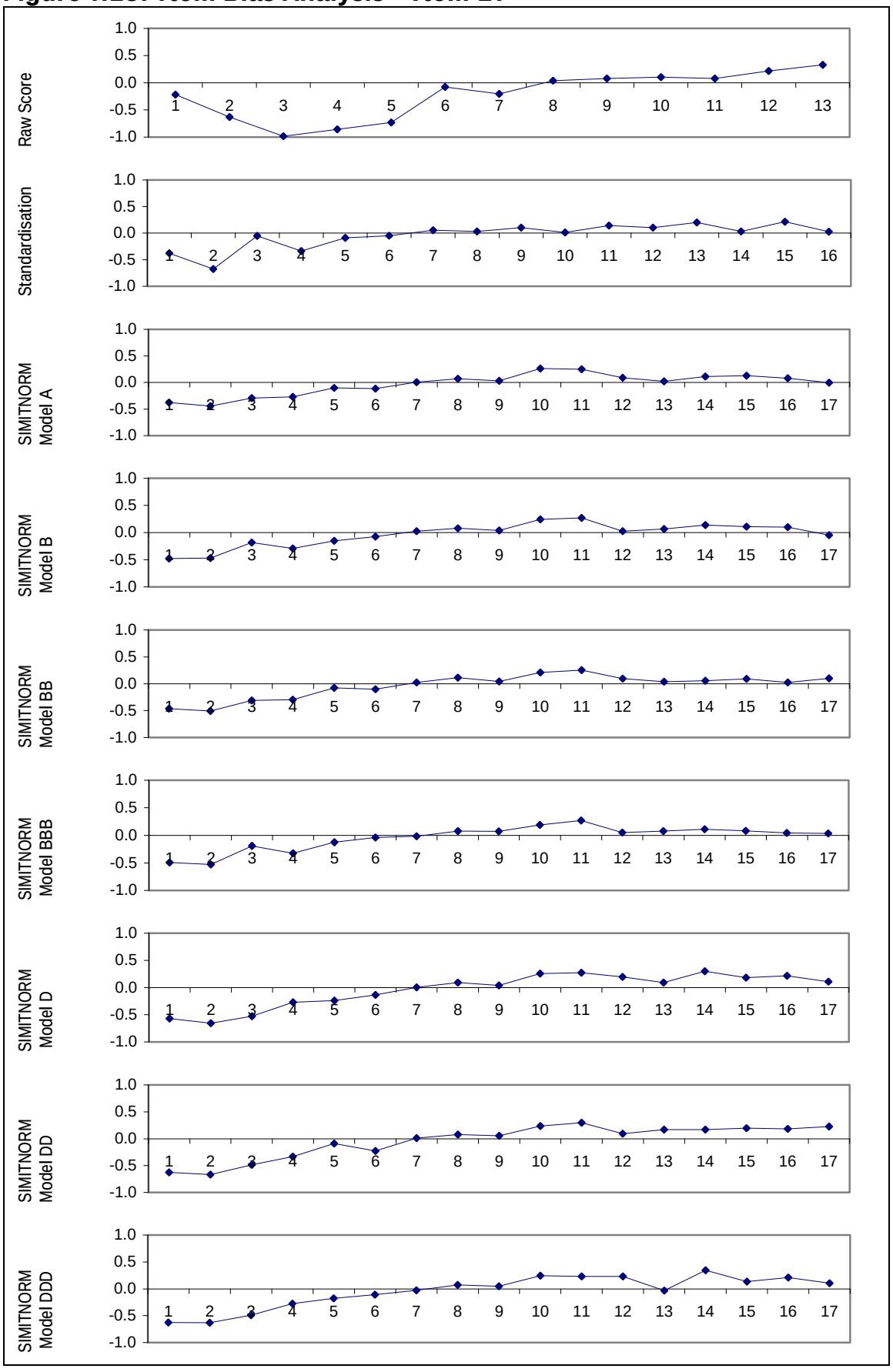


Figure I.19: Item Bias Analysis - Item 28

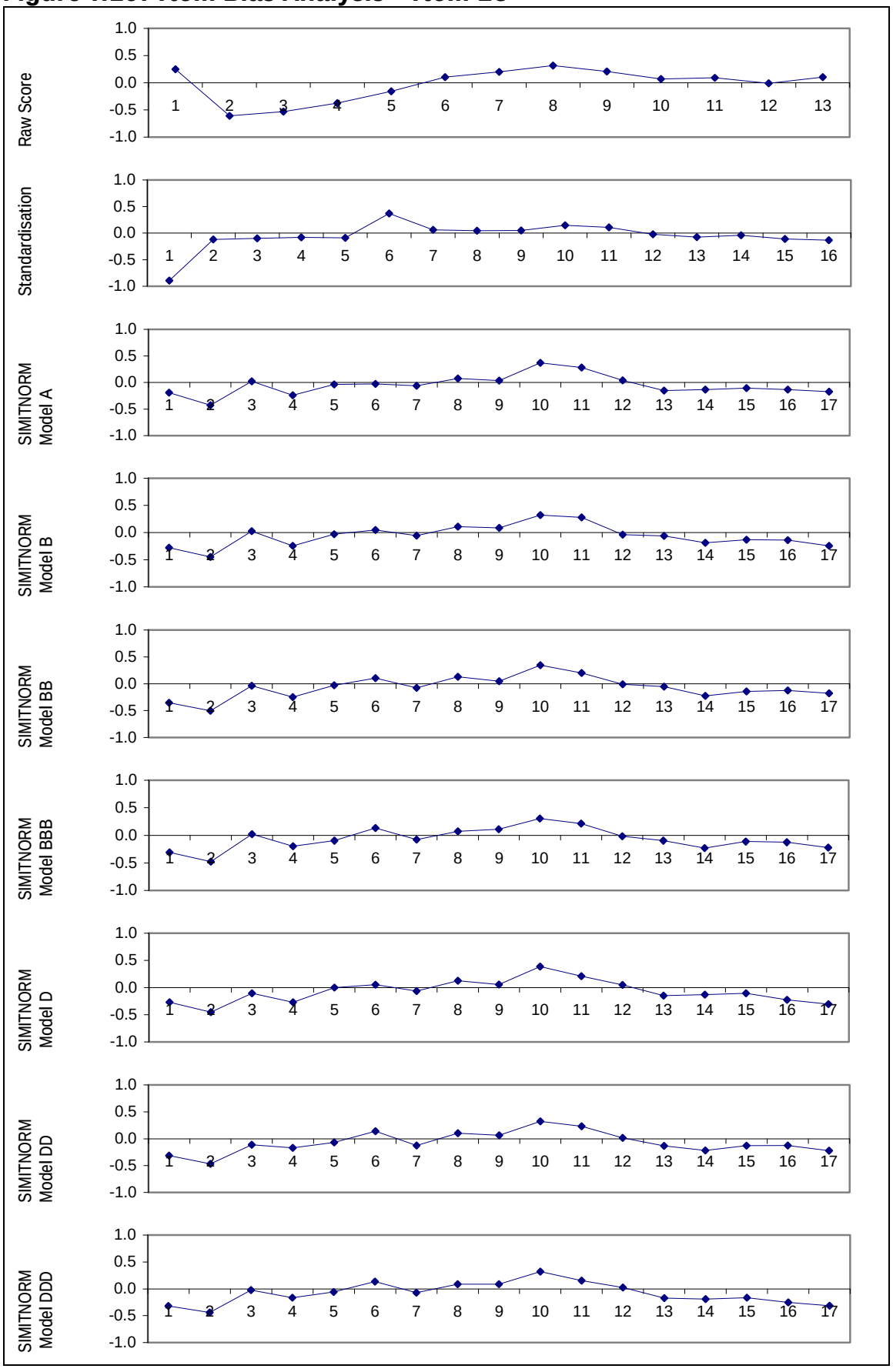


Figure I.20: Item Bias Analysis - Item 29

