

Table of Contents

1 DATA PREPARATION	16
2 BILOG-MG.....	24
2.1 NEW FEATURES IN BILOG-MG	24
2.2 PHASES OF THE ANALYSIS: INPUT, CALIBRATION AND SCORING	26
2.3 THE BILOG-MG INTERFACE	37
2.3.1 File menu.....	38
2.3.2 Edit menu	40
2.3.3 Setup menu.....	40
2.3.4 Data menu	54
2.3.5 Technical menu	64
2.3.6 Save menu	80
2.3.7 Run menu.....	81
2.3.8 Output menu	82
2.3.9 View menu.....	82
2.3.10 Options menu	83
2.3.11 Window menu.....	85
2.3.12 Help menu	85
2.3.13 Location of keywords in interface	85
2.4 GETTING STARTED WITH BILOG-MG.....	91
2.4.1 A first model: 2PL model for spelling data	92
2.4.2 A second model: DIF model for spelling data	100
2.5 SYNTAX.....	108
2.5.1 Data structures: ITEMS, TEST, GROUP and FORM commands	108
2.6 USING THE COMMAND LANGUAGE.....	113

2.6.1 Overview of syntax.....	113
2.6.2 Order of commands	114
2.6.3 CALIB command	116
2.6.4 COMMENT command.....	141
2.6.5 DRIFT command.....	142
2.6.6 FORM command.....	144
2.6.7 GLOBAL command	147
2.6.8 GROUP command	159
2.6.9 INPUT command.....	163
2.6.10 ITEMS command	182
2.6.11 LENGTH command.....	185
2.6.12 PRIORS command	187
2.6.13 QUAD command.....	193
2.6.14 QUADS command	196
2.6.15 SAVE command.....	199
2.6.16 SCORE command.....	208
2.6.17 TEST command	224
2.6.18 TITLE command	234
2.6.19 Variable format statement	235
2.6.20 Input and output files.....	241
3 PARSCALE	257
3.1 THE PARSCALE INTERFACE.....	258
3.1.1 Main menu.....	258
3.1.2 Workspace	259
3.1.3 Run menu.....	259
3.1.4 Output menu	260

3.1.5 Font option.....	260
3.1.6 Window menu.....	261
3.2 COMMAND SYNTAX.....	261
3.2.1 Order of commands.....	262
3.2.2 BLOCK command.....	265
3.2.3 CALIB command.....	274
3.2.4 COMBINE command.....	285
3.2.5 COMMENT command.....	287
3.2.6 FILES command.....	288
3.2.7 INPUT command.....	292
3.2.8 MGROUP command.....	300
3.2.9 MRATER command.....	303
3.2.10 PRIORS command.....	305
3.2.11 QUADP command.....	308
3.2.12 QUADS command.....	310
3.2.13 SAVE command.....	312
3.2.14 SCORE command.....	316
3.2.15 TEST/SCALE command.....	325
3.2.16 TITLE command.....	330
3.2.17 Variable format statements.....	331
3.3 INPUT FILES.....	333
3.3.1 Specification of input files.....	333
3.3.2 Individual level data.....	333
3.3.3 Group-level data.....	335
3.3.4 Key files.....	336

3.4 OUTPUT FILES.....	337
3.4.1 Format of output files	337
3.4.2 Combined score file.....	337
3.4.3 Fit statistics file	338
3.4.4 Item parameter file	340
3.4.5 Item information file.....	342
3.4.6 Subject scores file.....	343
4 MULTILOG	345
4.1 THE MULTILOG USER'S INTERFACE	345
4.1.1 Main menu.....	346
4.1.2 Run menu.....	346
4.1.3 Output menu	347
4.1.4 Window menu	347
4.1.5 Font option.....	348
4.2 CREATING SYNTAX USING THE MULTILOG SYNTAX WIZARD	349
4.2.1 New Analysis dialog box	349
4.2.2 Fixed Theta dialog box	350
4.2.3 Input Data dialog box.....	351
4.2.4 Input Parameters dialog box.....	352
4.2.5 Test Model dialog box	354
4.2.6 Response Codes (Binary Data) dialog box.....	355
4.2.7 Response Codes (Non-Binary Data) dialog box	356
4.3 GETTING STARTED WITH MULTILOG	357
4.3.1 Two-parameter model for the skeletal maturity data.....	357
4.3.2 Three-parameter (and guessing) model for the LSAT6 data.....	364

4.3.3	Generating syntax for a fixed- θ model.....	370
4.4	COMMAND SYNTAX.....	375
4.4.1	Overview of syntax.....	375
4.4.2	END command.....	378
4.4.3	EQUAL command.....	379
4.4.4	ESTIMATE command.....	382
4.4.5	FIX command.....	385
4.4.6	LABELS command.....	387
4.4.7	PROBLEM command.....	388
4.4.8	PRIORS command.....	393
4.4.9	SAVE command.....	395
4.4.10	START command.....	396
4.4.11	TEST command.....	398
4.4.12	TGROUPS command.....	401
4.4.13	TMATRIX command.....	403
4.4.14	Variable format statement.....	405
5	TESTFACT.....	410
5.1	INTRODUCTION.....	410
5.2	THE TESTFACT INTERFACE.....	411
5.2.1	Main menu.....	411
5.2.2	Run menu.....	412
5.2.3	Output menu.....	412
5.2.4	Window menu.....	412
5.2.5	Font option.....	412
5.3	COMMAND SYNTAX.....	413
5.3.1	Order of commands.....	414

5.3.2 Overview of syntax.....	415
5.3.3 BIFACTOR command.....	418
5.3.4 CLASS command	424
5.3.5 COMMENT command.....	426
5.3.6 CONTINUE command.....	427
5.3.7 CRITERION command.....	428
5.3.8 EXTERNAL command.....	430
5.3.9 FACTOR command	431
5.3.10 FRACTILES command	435
5.3.11 FULL command	437
5.3.12 INPUT command.....	441
5.3.13 KEY command	448
5.3.14 NAMES command	449
5.3.15 PLOT command.....	450
5.3.16 PRIOR command.....	452
5.3.17 PROBLEM command	454
5.3.18 RELIABILITY command	459
5.3.19 RESPONSE command	460
5.3.20 SAVE command.....	461
5.3.21 SCORE command.....	474
5.3.22 SELECT command	480
5.3.23 SIMULATE command.....	482
5.3.24 STOP command.....	488
5.3.25 SUBTEST command.....	489
5.3.26 TECHNICAL command.....	491
5.3.27 TETRACHORIC command	499

5.3.28 TITLE command	502
5.3.29 Variable format statement	502
6 IRT GRAPHICS.....	505
6.1 INTRODUCTION.....	505
6.2 MAIN MENU.....	505
6.2.1 The ICC option.....	506
6.2.2 The Information option	507
6.2.3 The ICC and Info option.....	508
6.2.4 The Total Info option.....	509
6.2.5 Matrix Plot option	510
6.2.6 The Histogram option	512
6.2.7 The Bivariate Plot option.....	513
6.2.8 The Exit option	514
6.3 MANIPULATING AND MODIFYING GRAPHS	514
6.3.1 File menu.....	514
6.3.2 Edit menu	515
6.3.3 Options menu	515
6.3.4 Graphs menu	516
6.3.5 Axis Labels dialog box	517
6.3.6 Bar Graph Parameters dialog box	518
6.3.7 Legend Parameters dialog box.....	520
6.3.8 Line Parameters dialog box	521
6.3.9 Plot Parameters dialog box.....	522
6.3.10 Text Parameters dialog box	522
6.4 ITEM CHARACTERISTIC CURVES.....	523
6.5 ITEM INFORMATION CURVES.....	524

6.6 TEST INFORMATION CURVES.....	526
7 OVERVIEW AND MODELS	528
7.1 OVERVIEW OF IRT PROGRAMS	528
7.1.1 BILOG-MG	528
7.1.2 PARSCALE.....	528
7.1.3 MULTILOG.....	529
7.1.4 TESTFACT.....	529
7.2 MODELS IN BILOG-MG	530
7.2.1 Introduction	530
7.2.2 Multiple-group analyses	531
7.2.3 Technical details	538
7.2.4 Statistical tests.....	543
7.3 MODELS IN PARSCALE.....	544
7.3.1 Introduction	544
7.3.2 Samejima's graded response model.....	546
7.3.3 Masters' partial credit model	550
7.3.4 Scoring function of generalized partial credit model	557
7.3.5 Multiple-group polytomous item response models.....	560
7.3.6 Constraints for group parameters.....	560
7.3.7 Test of goodness-of-fit	561
7.3.8 Initial parameter estimates.....	562
7.4 MODELS IN MULTILOG	567
7.4.1 Introduction	567
7.4.2 The graded model	567
7.4.3 The one- and two-parameter logistic models.....	567

7.4.4 The multiple response model	568
7.4.5 The multiple-choice model.....	569
7.4.6 The three-parameter logistic model	569
7.4.7 The nominal model.....	570
7.4.8 Contrasts.....	570
7.4.9 Equality constraints and fixed parameters.....	575
7.5 OPTIONS AND STATISTICS IN TESTFACT	575
7.5.1 Introduction	575
7.5.2 Classical item analysis and test scoring	575
7.5.3 Classical descriptive statistics	576
7.5.4 Item statistics	577
7.5.5 Fractile tables	580
7.5.6 Plots.....	582
7.5.7 Correction for guessing	582
7.5.8 Internal consistency	582
7.5.9 Tetrachoric correlations and factor analysis	583
7.5.10 IRT based item factor analysis	584
7.5.11 Full information factor analysis.....	585
7.5.12 Bifactor analysis.....	586
7.5.13 Not-reached items in factor analysis	586
7.5.14 Constraints on item parameter estimates	586
7.5.15 Statistical test of the number of factors	587
7.5.16 Factor scores.....	588
7.5.17 Number of quadrature points.....	589
7.5.18 Monte Carlo integration.....	591
7.5.19 Applications.....	591

8 ESTIMATION	592
8.1 INTRODUCTION.....	592
8.1.1 Trait estimation with Item Response Theory.....	593
8.1.2 Information.....	597
8.2 ESTIMATION IN BILOG-MG	599
8.2.1 Item calibration.....	599
8.2.2 Test scoring	605
8.2.3 Test and item information	608
8.2.4 Effects of guessing	610
8.2.5 Aggregate-level IRT models.....	610
8.3 ESTIMATION IN PARSCALE.....	611
8.3.1 Prior densities for item parameters.....	612
8.3.2 Rescaling the parameters	612
8.3.3 The information function.....	613
8.3.4 Warm's weighted ML estimation of ability parameters	615
8.4 ESTIMATION IN MULTILOG	616
8.4.1 Item parameter estimation	616
9 USES OF ITEM RESPONSE THEORY	618
9.1 INTRODUCTION.....	618
9.2 SELECTION TESTING	618
9.3 QUALIFICATION TESTING.....	619
9.4 PROGRAM EVALUATION AND ASSESSMENT TESTING	619
9.5 CLINICAL TESTING.....	619
9.6 MEASUREMENT METHODS AND RESEARCH	620

9.7 APPROACHES TO ANALYSIS OF ITEM RESPONSE DATA	620
9.7.1 Test scoring	621
9.7.2 Test generalizability	622
9.7.3 Item analysis	623
9.7.4 Estimating the population distribution	625
9.7.5 Differential item functioning	626
9.7.6 Forms equating	626
9.7.7 Vertical equating	627
9.7.8 Construct definition	629
9.7.9 Analysis and scoring of rated responses	629
9.7.10 Matrix sampling	630
9.7.11 Estimating domain scores	631
9.7.12 Adaptive testing	632
10 BILOG-MG EXAMPLES	634
10.1 CONVENTIONAL SINGLE-GROUP IRT ANALYSIS	634
10.2 DIFFERENTIAL ITEM FUNCTIONING	638
10.3 DIFFERENTIAL ITEM FUNCTIONING	650
10.4 EQUIVALENT GROUPS EQUATING	652
10.5 VERTICAL EQUATING	658
10.6 MULTIPLE MATRIX SAMPLING DATA	666
10.7 ANALYSIS OF VARIANT ITEMS	670
10.8 GROUP-WISE ADAPTIVE TESTING	674
10.9 TWO-STAGE SPELLING TEST	679
10.10 ESTIMATING AND SCORING TESTS OF INCREASING LENGTH	685
10.11 COMMANDS FOR PARALLEL-FORM CORRELATIONS	685

10.12 EAP SCORING OF THE NAEP FORMS AND STATE MAIN AND VARIANT TESTS.....	686
10.13 DOMAIN SCORES.....	688
11 PARSCALE EXAMPLES	692
11.1 ITEM CALIBRATION AND EXAMINEE BAYES SCORING WITH THE RATING-SCALE GRADED MODEL.....	692
11.2 EXAMINEE MAXIMUM LIKELIHOOD SCORING FROM EXISTING PARAMETERS	708
11.3 CALIBRATION AND SCORING WITH THE GENERALIZED PARTIAL CREDIT RATING-SCALE MODEL: COLLAPSING OF CATEGORIES.....	709
11.4 TWO-GROUP DIFFERENTIAL ITEM FUNCTIONING (DIF) ANALYSIS WITH THE PARTIAL CREDIT MODEL.....	710
11.5 A TEST WITH 26 MULTIPLE-CHOICE ITEMS AND ONE 4-CATEGORY ITEM: THREE-PARAMETER LOGISTIC AND GENERALIZED PARTIAL CREDIT MODEL.....	720
11.6 ANALYSIS OF THREE TESTS CONTAINING ITEMS WITH TWO AND THREE CATEGORIES: CALCULATION OF COMBINED SCORES	722
11.7 RATER-EFFECT MODEL: MULTI-RECORD INPUT FORMAT WITH VARYING NUMBERS OF RATERS PER EXAMINEE.....	723
11.8 RATER-EFFECT MODEL: ONE-RECORD INPUT FORMAT WITH SAME NUMBER OF RATERS PER EXAMINEE	727
11.9 RATERS-EFFECT MODEL: ONE-RECORD INPUT FORMAT WITH VARYING NUMBERS OF RECORDS PER EXAMINEE	728
12 MULTILOG EXAMPLES	730
12.1 ONE-PARAMETER LOGISTIC MODEL FOR A FIVE-ITEM BINARY-SCORED TEST (LSAT6).....	730
12.2 TWO-PARAMETER MODEL FOR THE FIVE-ITEM TEST.....	732
12.3 THREE-PARAMETER (AND GUESSING) MODEL FOR THE FIVE-ITEM TEST.....	733
12.4 THREE-CATEGORY GRADED LOGISTIC MODEL FOR A TWO-ITEM QUESTIONNAIRE	735
12.5 THREE-CATEGORY PARTIAL CREDIT MODEL FOR THE TWO-ITEM QUESTIONNAIRE	738
12.6 FOUR-CATEGORY GRADED MODEL FOR A TWO-ITEM INTERVIEW SCALE	740

12.7 A GRADED MODEL ANALYSIS OF ITEM-WORDING EFFECT ON RESPONSES TO AN OPINION SURVEY	741
12.8 GRADED-MODEL SCORES FOR INDIVIDUAL RESPONDENTS	748
12.9 FIVE-CATEGORY RATINGS OF AUDIOGENIC SEIZURES IN MICE IN FOUR EXPERIMENTAL CONDITIONS.....	749
12.10 A NOMINAL MODEL FOR RESPONSES TO MULTIPLE-CHOICE ALTERNATIVES.....	751
12.11 A CONSTRAINED NONLINEAR MODEL FOR MULTIPLE-CHOICE ALTERNATIVES.....	757
12.12 A NOMINAL MODEL FOR TESTLETS	759
12.13 A CONSTRAINED NOMINAL MODEL FOR QUESTIONNAIRE ITEMS	761
12.14 A CONSTRAINED GENERALIZED PARTIAL CREDIT MODEL.....	762
12.15 A MIXED NOMINAL AND GRADED MODEL FOR SELF-REPORT INVENTORY ITEMS	765
12.16 A MIXED THREE-PARAMETER LOGISTIC AND PARTIAL CREDIT MODEL FOR A 26-ITEM TEST	767
12.17 EQUIVALENT GROUPS EQUATING OF TWO FORMS OF A FOUR-ITEM PERSONALITY INVENTORY	768
12.18 DIFFERENTIAL ITEM FUNCTIONING (DIF) ANALYSIS OF EIGHT ITEMS FROM THE 100-ITEM SPELLING TEST.....	770
12.19 INDIVIDUAL SCORES FOR A SKELETAL MATURITY SCALE BASED ON GRADED RATINGS OF OSSIFICATION SITES IN THE KNEE.....	772
13 TESTFACT EXAMPLES.....	775
13.1 CLASSICAL ITEM ANALYSIS AND SCORING ON A GEOGRAPHY TEST WITH AN EXTERNAL CRITERION	775
13.2 TWO-FACTOR NON-ADAPTIVE FULL INFORMATION FACTOR ANALYSIS ON A FIVE-ITEM TEST (LSAT7)	778
13.3 ONE-FACTOR NON-ADAPTIVE FULL INFORMATION ITEM FACTOR ANALYSIS OF THE FIVE-ITEM TEST.....	780

13.4 A THREE-FACTOR ADAPTIVE ITEM FACTOR ANALYSIS WITH BAYES (EAP) ESTIMATION OF FACTOR SCORES: 32 ITEMS FROM AN ACTIVITY SURVEY	780
13.4.1 Discussion of output.....	782
13.5 ADAPTIVE ITEM FACTOR ANALYSIS AND BAYES MODAL (MAP) FACTOR SCORE ESTIMATION FOR THE ACTIVITY SURVEY.....	802
13.6 SIX-FACTOR ANALYSIS OF THE ACTIVITY SURVEY BY MONTE CARLO FULL INFORMATION ANALYSIS	803
13.7 ITEM BIFACTOR ANALYSIS OF A 12TH-GRADE SCIENCE ASSESSMENT TEST.....	804
13.7.1 Discussion of bifactor analysis output	805
13.8 CONVENTIONAL THREE-FACTOR ANALYSIS OF THE 12TH-GRADE SCIENCE ASSESSMENT TEST	814
13.9 COMPUTING EXAMINEE GENERAL FACTOR SCORES FROM PARAMETERS OF A PREVIOUS BIFACTOR ANALYSIS.....	815
13.10 ONE-FACTOR ANALYSIS OF THE 12TH-GRADE SCIENCE ASSESSMENT TEST	817
13.11 ITEM FACTOR ANALYSIS OF A USER-SUPPLIED CORRELATION MATRIX	818
13.12 SIMULATING EXAMINEE RESPONSES TO A THREE-FACTOR TEST WITH USER-SUPPLIED PARAMETERS	819
13.13 SIMULATING EXAMINEE RESPONSES IN THE PRESENCE OF GUESSING AND NON-ZERO FACTOR MEANS.....	820
13.14 THREE-FACTOR ANALYSIS WITH PROMAX ROTATION: 32 ITEMS FROM THE SCIENCE ASSESSMENT TEST	823
13.15 PRINCIPAL FACTOR SOLUTION OF A FACTOR ANALYSIS ON SIMULATED DATA: NO GUESSING	825
13.16 NON-ADAPTIVE FACTOR ANALYSIS OF SIMULATED DATA: PRINCIPAL FACTOR SOLUTION, NO GUESSING.....	826
13.17 ADAPTIVE ITEM FACTOR ANALYSIS OF 25 SPELLING ITEMS FROM THE 100-ITEM SPELLING TEST.....	827
13.18 CLASSICAL ITEM FACTOR ANALYSIS OF SPELLING DATA FROM A TETRACHORIC CORRELATION MATRIX	828

14 APPENDIX A: A BRIEF HISTORY OF ITEM RESPONSE THEORY	830
14.1 ANTECEDENTS.....	830
14.2 CONNECTIONS	833
14.3 IRT TEST SCORING	836
14.4 IRT ITEM ANALYSIS	840
14.5 CURRENT TRENDS.....	844
15 REFERENCES	848