UNIVERSITY OF ALBERTA

INVESTIGATION OF DIFFERENTIAL ITEM FUNCTIONING IN EXIT EXAMINATIONS ACROSS ITEM FORMAT AND SUBJECT AREA

by

DIANNE LOUISE HENDERSON ©

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Doctor of Philosophy

Department of Educational Psychology

Edmonton, Alberta

Fall, 1999

0-612-46848-8

University of Alberta

Library Release Form

Name of Author: DIANNE LOUISE HENDERSON

Title of Thesis: INVESTIGATION OF DIFFERENTIAL ITEM FUNCTIONING IN EXIT EXAMINATIONS ACROSS ITEM FORMAT AND SUBJECT AREA

Degree: DOCTOR OF PHILOSOPHY

Year this Degree Granted: 1999

R. R. # 2
Tofield, AB T0B 4J0

September 14, 1999

University of Alberta

Faculty of Graduate Studies and Research

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled INVESTIGATION OF DIFFERENTIAL ITEM FUNCTIONING IN EXIT EXAMINATIONS ACROSS ITEM FORMAT AND SUBJECT AREA submitted by DIANNE LOUISE HENDERSON in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Dr. W. Todd Rogers

Dr. Mark J. Gierl

Dr. Gay L. Bisanz

Dr. Linda M. Phillips

Dr. Marcel Bouffard

Dr. John O. Anderson

Date: Sept. 13, 1999

Abstract

The primary purposes of this study were to determine the agreement 1) between the polytomous DIF detection methods, GMH, Poly-SIB, and LDF and their respective counterparts, MH, SIB, LR; and 2) among the polytomous DIF detection methods, GMH, Poly-SIB, and LDF. The agreement was examined across samples of varying size that completed the English, Social Studies, Mathematics, and Biology Alberta Education Diploma Examinations. The variable gender was selected for the analyses, which also provided the opportunity to investigate hypothesized differences between males and females using, a common sample of examinees across subject area and item format was employed.

Results of this study indicate that both the GMH and Poly-SIB were comparable to their dichotomous counterparts, MH and SIB in small sample sizes. In contrast, while LR and LDF were comparable for the large sample sizes they were not comparable for the smaller samples. In the comparison of GMH, Poly-SIB, and LDF different results were obtained for the set of dichotomous items compared to the set that included the polytomous items due to the alteration of the total test score (matching variable) and the ability distribution of the reference and focal groups. The pattern of results, however, were similar regardless of the matching variable. For small sample sizes, LDF was the most conservative and in large sample sizes GMH was the most conservative, while Poly-SIB produced the most consistent results. In addition, Poly-SIB detected the most items, including the majority of the items identified by GMH and LDF.

Results of the analysis of the ancillary question indicate that the DIF prevalence rates for the dichotomous sections of the examinations in this study are consistent with those reported in the literature for other American high school examinations. In addition, more selection response items favored males, while all of the constructed response items identified favored females suggesting that there may be a gender by format interaction. Furthermore, previous findings suggesting that males outperform females on geometry and mathematical problem solving items and on items containing formulas, equations, symbols, or references to stereotypical male activities were not supported by this research.

Acknowledgements

At this time I would like to take this opportunity to acknowledge and thank the many individuals who have helped me to achieve this goal. I would like to first thank my family and friends for the unconditional support through my academic journey. While I would like to reassure you all that this is my last degree, I am reluctant to give that guarantee.

Secondly, I would like to thank Dr. W. Todd Rogers who rescued me and brought me to CRAME. The experience changed my life and has led me to pursue new goals and a new career that is filled with promise. Thank you. Your support, encouragement, and advice is invaluable and I look forward to working with you in the future. I would also like to thank your family, Rachel and Linnae, for sacrificing some of their time with you so that I could achieve my own goals.

I would also like to take this opportunity to thank the many CRAMErs, both past and present, which I have had the opportunity to work with in the past few years. My time in this Centre has provided me with a rich experience that I wish for every graduate student. I look forward to working with all of you as colleagues in the future.

Special thanks to Don and Mike, both exceptional teachers, that helped me to see the value in debate and confrontation, to develop confidence in my abilities in this area, and gave many thoughtful and constructive comments throughout this project. We were lucky, I think, to work together, to aspire to higher goals and to spend countless hours debating the many merits and dilemmas facing psychometricians. This experience, above all others has immeasurable influenced my decisions and will continue to influence me in the future. Thank you both. I value your friendship and the times that we spent together with your families and look forward to future encounters.

These acknowledgements would not be complete without a word (or two) of thanks to Dr. Mark Gierl who introduced me to IRT, DIF, MH, SIB, LR, SIMS, AERA, and NCME. Your influence is obvious in this project as well as in my choice of career. I am also indebted to you for advising me on the best way to experience AERA/NCME and cosmopolitans.

Table of Contents

Chapter 1: Introduction **1**

Purposes of Research 5
Definition of Terms 7

Chapter 2: Literature Review **9**

Item Response Theory *10*
Identifying DIF Using Item Response Theory *13*
Non-IRT DIF Detection Methods *17*
Mantel-Haenszel. 18
Generalized Mantel Haenzel. 21
Simultaneous Item Bias Test. 24
Poly-SIBTEST. 26
Logistic Regression. 27
Logistic Discriminant Function. 29
Gender DIF Across Examinations and Formats *32*
Studies of Gender DIF in Alberta Education Diploma Examinations *37*

Chapter 3: Method **42**

Examinations to be Studied *43*
Sample *46*
Comparisons of DIF Procedures 48
Prevalence of Gender DIF 48
Procedures *49*
Comparisons of DIF Procedures: Dichotomous versus Polytomous Versions 49
Comparisons of DIF Procedures: GMH, Poly-SIB, LDF 50
Prevalence of Gender DIF 50

Chapter 4: Results **52**

Comparability of Dichotomous and Polytomous DIF Detection Methods *52*
MH versus GMH 56
SIB versus Poly-SIB 57
LR versus LDF 58
Summary 60
Comparison of GMH, Poly-SIB, and LDF *61*
Dichotomous Items 61
Inclusion of Polytomous Items 64
Summary 71
Comparison of Gender DIF Detected Across Format and Subject *73*
Mean Differences between Males and Females 73
English 78
Social Studies 81
Mathematics 86

Biology 91
DIF Across Subject and Item Format. 94

Chapter 5: Summary and Conclusions 99

Summary of Research Questions and Methods *99*
Findings *101*
Comparison of DIF Detection Methods 101
Comparison of Polytomous DIF Detection Methods 101
Prevalence of Gender DIF across Item Format and Subject Area 104
Limitations of the Study *108*
Recommendations *109*
Future Practice 109
Future Research 111

References 113

Appendix A 123

Comparison of MH, GMH, and SIB, Poly-SIB Results for each Examination and Sample: Dichotomous Items Only 123

Appendix B 172

Comparison of LR versus LDF with Three Effect Size Measures for each Examination and Sample: Dichotomous Items Only 172

Appendix C 221

DIF Detection Results for each Examination and Sample: Polytomous and Dichotomous Items Combined 221

Appendix D 282

GMH Results: Dichotmous Items Analyzed with Combined Dichotomous and Polytomous Total Test Scores 282

Appendix E 295

Poly-SIBTEST Results for each Examination 295

List of Tables

Table 1. Average Diploma Examination by Mark. 38

Table 2. Structure Comparison of Examinations 43

Table 3. Student Combinations. 47

Table 4. Comparison of Dichotomous and Polytomous DIF Detection Methods for Moderate to Severe DIF. 53

Table 5. Percent Agreement between Dichotomous and Polytomous DIF Detection Procedures by Subject and Sample. 55

Table 6. Agreement between Polytomous DIF Detection Methods for Dichotomous Items. 62

Table 7. Agreement between Polytomous DIF Detection Methods for Combined Dichotomous and Polytomous Items. 66

Table 8. Descriptive Statistics for each Examination by Gender. 75

Table 9. DIF Items Identified by Method across Subject Area and Item Format. 77

Table 10. English DIF Items by Course Content and Cognitive Level: Dichotomous Items. 79

Table 11. Social Studies DIF Items by Course Content and Cognitive Level: D9ichotomous Items.. 83

Table 12. DIF within Item Scale Scores Detected by LDF Social Studies 1997. 85

Table 13. DIF within Item Scale Scores Detected by LDF Social Studies 1998. 86

Table 14. Mathematics DIF Items Organized by Unit and Understanding: Dichotomous Items 88

Table 15. Biology DIF Items Organized by Unit Topic: Dichotomous Items...............93

Table 16. Prevalence of DIF across Subject Area and Item Format...............................95

Table 17. MH/GMH and SIB/Poly-SIB Comparisons: English January, 1998...........124

Table 18. MH/GMH and SIB/Poly-SIB Comparisons: English June, 1998.................126

Table 19. MH/GMH and SIB/Poly-SIB Comparisons: English 1212, 1998................128

Table 20. MH/GMH and SIB/Poly-SIB Comparisons: English 2112, 1998................130

Table 21. MH/GMH and SIB/Poly-SIB Comparisons: English 2211, 1998................132

Table 22. MH/GMH and SIB/Poly-SIB Comparisons: English 2222, 1998................134

Table 23. MH/GMH and SIB/Poly-SIB Comparisons: Social January 1998...............136

Table 24. MH/GMH and SIB/Poly-SIB Comparisons: Social June 1998....................138

Table 25. MH/GMH and SIB/Poly-SIB Comparisons: Social 1212 1998...................140

Table 26. MH/GMH and SIB/Poly-SIB Comparisons: Social 2112 1998...................142

Table 27. MH/GMH and SIB/Poly-SIB Comparisons: Social 2211 1998...................144

Table 28. MH/GMH and SIB/Poly-SIB Comparisons: Social 2222 1998..................146

Table 29. MH/GMH and SIB/Poly-SIB Comparisons: Mathematics January 1998...148

Table 30. MH/GMH and SIB/Poly-SIB Comparisons: Mathematics June 1998........150

Table 31. MH/GMH and SIB/Poly-SIB Comparisons: Mathematics 1212 1998........152

Table 32. MH/GMH and SIB/Poly-SIB Comparisons: Mathematics 2112 1998........154

Table 33. MH/GMH and SIB/Poly-SIB Comparisons: Mathematics 2211 1998.......156

Table 34. MH/GMH and SIB/Poly-SIB Comparisons: Mathematics 2222 1998........158

Table 35. MH/GMH and SIB/Poly-SIB Comparisons: Biology January 1998...........160

Table 36. MH/GMH and SIB/Poly-SIB Comparisons: Biology June 1998................162

Table 37. MH/GMH and SIB/Poly-SIB Comparisons: Biology 1212 1998 164

Table 38. MH/GMH and SIB/Poly-SIB Comparisons: Biology 2112 1998 166

Table 39. MH/GMH and SIB/Poly-SIB Comparisons: Biology 2211 1998 168

Table 40. MH/GMH and SIB/Poly-SIB Comparisons: Biology 2222 1998 170

Table 41. LR/LDF Comparison English January, 1998 173

Table 42. LR/LDF Comparison: English June, 1998 175

Table 43. LR/LDF Comparison: English 1212, 1998 177

Table 44. LR/LDF Comparison: English 2112, 1998 179

Table 45. LR/LDF Comparison: English 2211, 1998 181

Table 46. LR/LDF Comparison: English 2222, 1998 183

Table 47. LR/LDF Comparison: Social January, 1998 185

Table 48. LR/LDF Comparison: Social June, 1998 187

Table 49. LR/LDF Comparison: Social 1212, 1998 189

Table 50. LR/LDF Comparison: Social 2112, 1998 191

Table 51. LR/LDF Comparison: Social 2211, 1998 193

Table 52. LR/LDF Comparison: Social 2222, 1998 195

Table 53. LR/LDF Comparison: Mathematics January, 1998 197

Table 54. LR/LDF Comparison: Mathematics June, 1998 199

Table 55. LR/LDF Comparison: Mathematics 1212, 1998 201

Table 56. LR/LDF Comparison: Mathematics 2112, 1998 203

Table 57. LR/LDF Comparison: Mathematics 2211, 1998 205

Table 58. LR/LDF Comparison: Mathematics 2222, 1998 207

Table 59. LR/LDF Comparison: Biology January, 1998 209

Table 60. LR/LDF Comparison: Biology June, 1998 211

Table 61. LR/LDF Comparison: Biology 1212, 1998 213

Table 62. LR/LDF Comparison: Biology 2112, 1998 215

Table 63. LR/LDF Comparison: Biology 2211, 1998 217

Table 64. LR/LDF Comparison: Biology 2222, 1998 219

Table 65. DIF Detection by Polytomous Method: English January, 1998 222

Table 66. DIF Detection by Polytomous Method: English June, 1998 225

Table 67. DIF Detection by Polytomous Method: English 1212, 1998 228

Table 68. DIF Detection by Polytomous Method: English 2112, 1998 231

Table 69. DIF Detection by Polytomous Method: English 2211, 1998 234

Table 70. DIF Detection by Polytomous Method: English 2222, 1998 237

Table 71. DIF Detection by Polytomous Method: Social Studies January, 1998 240

Table 72. DIF Detection by Polytomous Method: Social Studies June, 1998 243

Table 73. DIF Detection by Polytomous Method: Social Studies 1212, 1998 246

Table 74. DIF Detection by Polytomous Method: Social Studies 2112, 1998 249

Table 75. DIF Detection by Polytomous Method: Social Studies 2211, 1998 252

Table 76. DIF Detection by Polytomous Method: Social Studies 2222, 1998 255

Table 77. DIF Detection by Polytomous Method: Mathematics January, 1998 258

Table 78. DIF Detection by Polytomous Method: Mathematics June, 1998 260

Table 79. DIF Detection by Polytomous Method: Mathematics 1212, 1998 262

Table 80. DIF Detection by Polytomous Method: Mathematics 2112, 1998 264

Table 81. DIF Detection by Polytomous Method: Mathematics 2211, 1998 266

Table 82. DIF Detection by Polytomous Method: Mathematics 2222, 1998 268

Table 83. DIF Detection by Polytomous Method: Biology January, 1998..................270

Table 84. DIF Detection by Polytomous Method: Biology June, 1998.......................272

Table 85. DIF Detection by Polytomous Method: Biology 1212, 1998......................274

Table 86. DIF Detection by Polytomous Method: Biology 2112, 1998......................276

Table 87. DIF Detection by Polytomous Method: Biology 2211, 1998......................278

Table 88. DIF Detection by Polytomous Method: Biology 2222, 1998......................280

Table 89. GMH Results: Dichotomous Items with Combined Dichotomous and Polytomous Test Score, English 1998 283

Table 90. GMH Results: Dichotomous Items with Combined Dichotomous and Polytomous Test Score, Social Studies 1998 286

Table 91. GMH Results: Dichotomous Items with Combined Dichotomous and Polytomous Test Score, Mathematics 1998 289

Table 92. GMH Results: Dichotomous Items with Combined Dichotomous and Polytomous Test Score, Biology 1998..292

Table 93. Poly-SIB Results: English...296

Table 94. Poly-SIB Results: Social Studies 299

Table 95. Poly-SIB Results: Mathematics 302

Table 96. Poly-SIB Results: Biology...304

List of Figures

Figure 1. Item Characteristic Curve for a Three-Parameter Item 11

Figure 2. Item Demonstrating No DIF 14

Figure 3. Item Demonstrating Uniform DIF 15

Figure 4. Item Demonstrating Non-Uniform DIF 16

Figure 5. Contingency Table for kth level of the Matching Variable 22

Figure 6. Illustration DIF Detection using LDF 32

Chapter 1

Introduction

The goal of all test developers is to assemble a set of items that provides an estimate of an examinee's ability that is as fair and accurate as possible for all groups of the population. Thus the test development process includes a systematic item analysis to ensure that all examinees with the same underlying level of ability have the same probability of getting an item correct. Unfortunately, empirical evidence can often be found in administered tests which indicates that certain subgroups of the test taking population, matched with respect to the construct being measured, have a different probability of getting the item correct. Such items are described as having differential item functioning (DIF; Dorans & Holland, 1993; Holland & Thayer, 1988).

DIF may be attributed to either item impact, item bias, or Type I error. If the item reflects actual differences in the knowledge or ability of the examinees, then DIF may be attributed to item impact (Camilli & Shepard, 1994). On the other hand, if the item is characterized by a systematic error in how an item measures the intended construct for a distinct group of examinees (e.g., Aboriginal, female), then DIF may be attributed to item bias (Camilli & Shepard, 1994). That is, biased items contain extraneous or construct irrelevant information that makes the item unfairly difficult for an identifiable group of examinees, adversely affecting their test performance. Finally, an item may be falsely identified by chance alone.

Items identified as having DIF are currently classified as displaying small (A-level), moderate (B-level), or large (C-level) amounts of DIF in accordance with guidelines established by Educational Testing Services (ETS; Zwick & Erickan, 1989). Items that display small amounts of DIF are retained on the test. In contrast, items that display moderate or large amounts of DIF are reviewed by a panel of content experts to help ascertain the source of the performance differences (Ramsey, 1993). If the DIF is judged to be related to item impact, the item is retained without revision. If the DIF is attributed to item bias or an undetermined source then the item is either modified or deleted.

There are several statistical methods to identify DIF in both dichotomous (e.g., multiple choice) and polytomous (e.g., constructed response, essay) items, although more research has been devoted to the dichotomous case (Clauser & Mazor, 1998; Downing & Haladayna, 1997). These include methods based on item response theory (IRT), contingency tables, and logistic regression. Methods based on IRT provide both statistical methods for comparing item parameter indices and graphical methods for identifying items that are performing differently for subgroups of examinees. However, despite the visual appeal of these approaches, at the present time they are not extensively used in DIF detection studies due the difficulty associated with satisfying the associated model assumptions, the requirement of large sample sizes, and the complexity of the procedures.

Currently, the most popular methods used to detect DIF in dichotomous items are the Mantel-Haenszel (MH), Logistic Regression (LR), and Simultaneous Item Bias Test (SIB) methods. These methods provide comparable results to the IRT methods but they are more easily understood. Further, they do not rely on the explicit assumptions associated with IRT, nor do they require large sample sizes (Camilli & Shepard, 1994). Although no single method can identify all DIF items (Hambleton, Clauser, Mazer, & Jones, 1993), comparative studies indicate that results across methods tend to be similar (Clauser & Mazor, 1998; Dorans & Holland, 1993; Gierl & McEwen, 1998; Hambleton & Rogers, 1989; Roussos & Stout, 1996).

The recent inclusion of constructed response items on large-scale standardized examinations has also led to the development of methods that can detect DIF among polytomous items. IRT methods (Wainer, Sireci, & Thissen, 1991), Generalized Mantel Haenszel (GMH; Zwick, Donoghue, & Grima, 1993), Simultaneous Item Bias for polytomous items (Poly-SIB; Chang, Mazzeo, & Roussos, 1996), Logistic Discriminant Function (LDF, Miller & Spray, 1993), and LR (French & Miller, 1996; Miller & Spray, 1993) have all been generalized for use with polytomous items. However, the identification of DIF in these items is more complicated due to the increased number of score categories associated with each item and the limited number of items that are usually administered. While there are many studies investigating the prevalence of DIF among dichotomous items, there are fewer studies investigating the prevalence of DIF among polytomous items or the comparison of different methods with actual data (Clauser & Mazor, 1998; Downing & Haladyna, 1997; Potenza & Dorans, 1995).

Investigation of item-level differences between males and females on tests composed of dichotomous items has identified specific content areas that tend to favor one group over the other (Carlton & Harris, 1989; Doolittle, 1989; Doolittle & Cleary, 1987; Gierl & McEwen, 1998; Scheuneman & Gerritz, 1990; O'Neill & McPeek, 1993).

For example, males tend to perform better than females on items related to science, and on items referring to stereotypical male activities on verbal ability measures found on standardized tests like the Graduate Record Examination (GRE; O'Neill & McPeek, 1993). In addition, males tend to perform better than females on items that involve proportions, ratios, geometry, graphs, tables, or figures (Burton, 1996, Doolittle & Cleary, 1987; Harris & Carlton, 1993; O'Neill & McPeek, 1993). In contrast, females tend to perform better than males on items related to aesthetics, human rights, computation, and those that involve symbols (Burton, 1996, Doolittle & Cleary, 1987; Harris & Carlton, 1993; O'Neill & McPeek, 1993; Sadker & Sadker, 1994). Unfortunately, DIF results found in one study are frequently inconsistent with subsequent studies of the same examination (O'Neill & McPeek, 1993). These differences can be attributed to different: 1) ages and grades of cohorts; 2) constructs; and 3) types of samples (e.g., self-selected, representative, or convenience; Willingham & Cole, 1997). Furthermore, it is also possible that the prevalence of DIF may be under-identified in larger heterogeneous samples.

In addition to the hypothesis that DIF is related to specific content, there is also some evidence to indicate that the type of item scoring may be related to DIF. For example, while males generally perform better than females on dichotomous items, females perform better than males on polytomous items like essays, possibly because of better verbal fluency, reading, and writing skills (Breland, Danos, Kahn, Kubota, & Benner, 1994; Froese, 1998; Pomplun & Sundbye, 1999; Willingham & Cole, 1997). Despite this observation, systematic investigations to determine if different types of item scoring (e.g., multiple choice, performance assessment) contribute to DIF are limited (Willingham & Cole, 1997). The examination of DIF in polytomously-scored items across different examinations using a common sample has also not been completed. Nor has the investigation of DIF in examinations composed of both types of items. As more standardized examinations containing both types of items are created, a better understanding of DIF across item format and the implications that any interaction may have on test performance is required.

The existence of DIF across different examinations and formats could have serious implications for the examinees' test performance leading to unintended consequences. For example, if DIF favoring males is pervasive across different examinations and item formats in high school exit examinations which can be attributable to bias, the resulting lower performance by females may lead to lower admission rates in competitive programs or fewer merit and scholarship awards. Furthermore, the lower performance by females may discourage them from applying for admission into more competitive post secondary programs (Wightman, 1998).

Unfortunately, item characteristics identified as possibly causing DIF in studies completed to date can only be interpreted within the context of the specific examination and sample. Examining DIF items across different examinations containing both dichotomous and polytomous items within common samples should permit a more clear examination of where males outperform females, females outperform males, and where the performance is the same.

Purposes of Research

The purpose of this study, therefore, was to investigate the comparability of the polytomous DIF detection methods, GMH, Poly-SIB, and LDF, to their dichotomous counterparts and the comparability among the three methods in examinations containing dichotomous and polytomous items. The variable gender was selected to investigate the performance of the polytomous DIF detection methods for two reasons. While there are several variables to stratify the two groups for DIF detection studies such as race and school location (urban/rural), this demographic variable is uniformly collected across all the studied examinations, was readily available, and provided an adequate focal group sample across all four examinations. Second, this variable provided the opportunity to investigate hypothesized differences between males and females across subject area and item format as described by O'Neill and McPeek (1993). However, unlike the study completed by O'Neill and McPeek, a common sample of examinees across subject area and item formats was employed.

More specifically, the following questions were answered in this study:

1. Compared to their dichotomous counterparts (MH, SIB, LR), how consistently do GMH, Poly-SIB, and LDF identify gender DIF across different tests written by the same group of examinees? How comparable are the methods?
2. How consistently do GMH, Poly-SIB, and LDF identify gender DIF across different tests written by the same group of examinees in examinations consisting of dichotomous and polytomous items? How comparable are the methods?
3. What is the proportion of dichotomous and polytomous items identified with gender

DIF in each examination within a sample of student completing all of the examinations of interest? Are there any patterns of DIF across subject area and across item format?

Definition of Terms

Dichotomously Scored Items: Items that have only one correct or best answer and are scored either correct or incorrect (i.e., 0,1). Commonly used to score selection-type items.

Differential Item Functioning (DIF): An item displays DIF if examinees from different groups have different probabilities or likelihood of success on the item after conditioning or matching on the ability the item is intended to measure.

Non-uniform DIF: The difference in the probabilities of a correct response for the two groups is not the same at all levels of ability, so that an interaction between ability level and group membership exists.

Uniform DIF: The difference in the probabilities of a correct response for the two groups is the same at all levels of ability, so that there is no interaction between ability level and group membership.

Matching Variable: Variable used to equate ability in methods used to detect DIF.

External Matching Variable: Variable is not based on the test under study (e.g., school awarded mark).

Internal Matching Variable: Variable is based on the test under study (e.g., total test score).

Focal Group: The primary group of interest to compare against the reference group. The focal group is usually the minority group or the group most likely to be disadvantaged.

Item Bias: Items function differently between two specified groups but these differences are caused by systematic unfairness in how the item measures the ability of interest.

Item Impact: Items function differently between two specified groups but these differences are caused by systematic differences between the underlying ability of two identifiable groups.

Polytomously Scored Items: Items that have three or more score categories (e.g., 0, 1, 2, 3, or 4). Commonly used in supply-type questions such as essays and constructed response items in, for example, mathematics and science.

Reference Group: The standard against which the focal group is compared, usually the majority group or the group most likely to be advantaged.

The results of this study are outlined and presented in the following chapters. The literature review, presented in Chapter 2, begins with a discussion of DIF and the DIF detection methods used in this study, and concludes with a summary of previously completed studies investigating gender DIF. The sample, methods, and procedures associated with this study are outlined in Chapter 3. This is followed by the presentation of results in Chapter 4. The summary and discussion of the results, and recommendations for future research are presented in Chapter 5.

Chapter 2

Literature Review

Differences in performance levels that occur between two groups have historically been described in various ways using classical test theory (CTT). Most commonly these descriptions have investigated the differences between the average test scores of two groups or between the proportion of examinees that correctly answered an item in each group (p-values). Unfortunately, these statistics simplify the interrelation between group mean differences, item difficulty, and the ability of the item to discriminate between individuals based on the underlying abilities of interest. For example, when two groups are not equal in the construct being measured, an item that is highly discriminating will appear to be functioning differently for the two groups (i.e., large differences in p-values), when in fact it is functioning as anticipated (Camilli & Shepard, 1994).

While identifying DIF using CTT methods is problematic, item response models (IRM; Lord, 1980) have several important advantages enabling a better understanding of this phenomenon. First, the statistical properties of items are less confounded with sample characteristics and are described more precisely. Therefore, the differences between two groups can be described more precisely. Secondly, differences in performance for the two groups on a specified item can be easily and quickly observed by examining item characteristic curves (ICC). Despite these advantages, these methods are not frequently used in DIF detection studies due to the difficulty in satisfying the associated model assumptions, the requirements of large sample sizes, and the complexity associated with these procedures. IRT does, however, provide an effective method to illustrate DIF. Therefore, the next section will begin with a brief overview of IRT and how it is used to detect DIF. An overview of MH, SIB, and LR, non-IRT methods that are used to detect DIF among dichotomous items and their polytomous counterparts, is then presented. This overview is followed by a summary of the results of several studies that have investigated performance differences between males and females at the item level and a review of the examinations that will be used in this study.

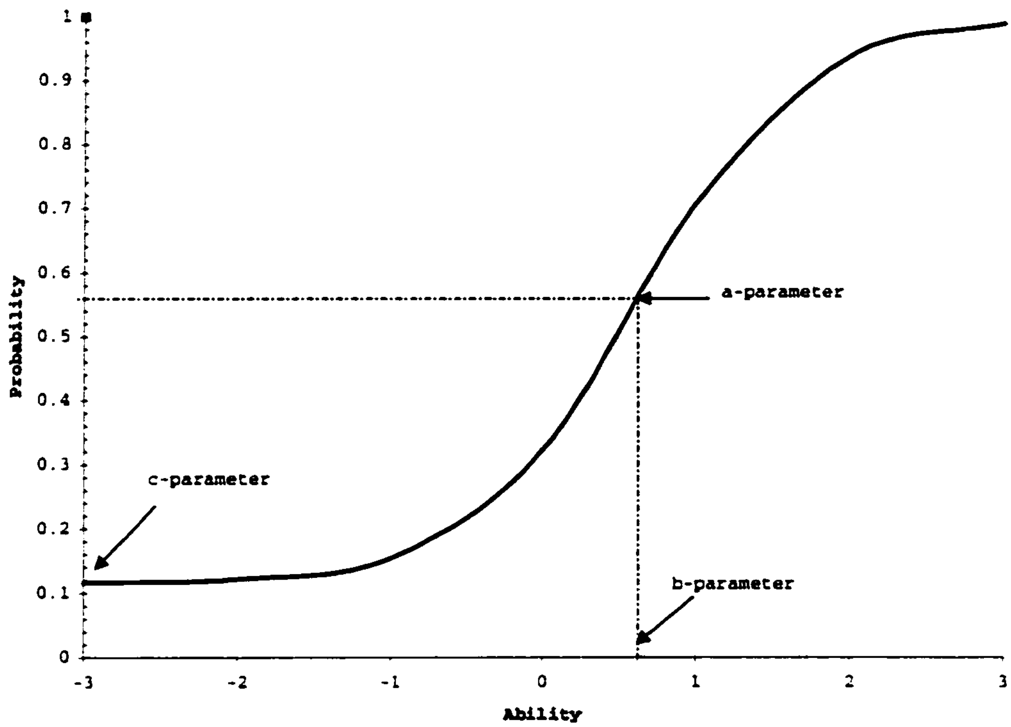
Item Response Theory

As mentioned above, DIF can be easily observed using item characteristic curves. The ICC (see Figure 1) is a

> mathematical function that relates the probability of success on an item to the ability measured by the item set or test that contains it. In simple terms, it is the nonlinear regression function of item score on the trait or ability measured by the test (Hambleton & Swaminathan, 1985, p. 9).

The probability of a correct answer $P(\theta)$ is plotted along the vertical or y-axis and ability is graphed along the horizontal or x-axis. The mathematical form of the ICC is described by an ability parameter (θ) and the number of item parameters associated with the item response model selected. These parameters include the b-parameter (b_i), a-parameter (a_i), and/or c-parameter (c_i).

Figure 1. Item Characteristic Curve for a Three-Parameter Item



Like CTT, the "true" ability of the individual cannot be directly observed with IRMs. Instead ability is estimated based on the observed item responses for an examinee. Thus θ is a measure of the latent ability underlying the test and is reflective of the construct as measured by the particular test. The scale of θ is interval and is constrained between –3 and +3.

The b-parameter (b_i) represents the difficulty of an item. It is located along the (θ) scale at the point at which the slope of the ICC is a maximum. The b-parameter is measured in the same units as (θ) and, in the two-parameter model, is "the point on the ability scale where the probability of a correct response is 0.5" (Hambleton, Swaminathan, & Rogers, 1991, p. 13). Therefore, high positive b-parameters indicate difficult items whereas large negative b-parameters indicate easy items.

The a-parameter (a_i) describes the discrimination of an item and indicates how well the item distinguishes between high and low ability examinees. It corresponds to the slope of the ICC at the point b_i. The steeper the slope the higher the a-parameter, or the better the item discriminates between high and low ability examinees. In IRM, the item discriminates best at the value of the b-parameter. At b_i, the slope of the ICC equals $.425a(1 - c)$.

The c-parameter (c_i) is the lower asymptote of the ICC and represents the probability that an examinee with an extremely low ability will correctly answer the test item. It is referred to as the pseudo-chance or guessing parameter because it is assumed that the examinee used a strategy unrelated to the ability of interest to obtain the correct answer (Lord, 1980).

As indicated earlier, the number of item parameters included in an item response model may vary. All models include the b-parameter. A one-parameter model is used to describe the data when all of the items have the same discrimination value (i.e., $a_i = 1.0$) and no pseudo-chance value (i.e., $c_i = 0.0$). If the a-parameter is included a two-parameter model is used to describe the data. The three-parameter model includes all of the parameters. Further, ICCs will vary in their shape as the b-parameter (difficulty), a-parameter (discrimination, slope), and c-parameter (pseudo-guessing, lower asymptote) are included and vary.

The mathematical form of the three-parameter logistic curve is written:

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}(i = 1,2,...n),$$

where

$P_i(\theta)$ = the probability that an examinee with ability level θ will answer item i correctly,

D = 1.7, "a scaling factor introduced to make the logistic function as close as possible to the normal ogive function" (Hambleton et al., 1991, p. 15), and

a_i, b_i, *and* c_i are as defined above.

In the case of a two-parameter model, the c-parameter is excluded:

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}(i = 1,2,....n).$$

Here it is assumed that guessing is minimal and can, therefore, be ignored. When plotted, the ICC has a lower asymptote of 0. If it can further be assumed that the items are equally discriminating, then the one-parameter model results:

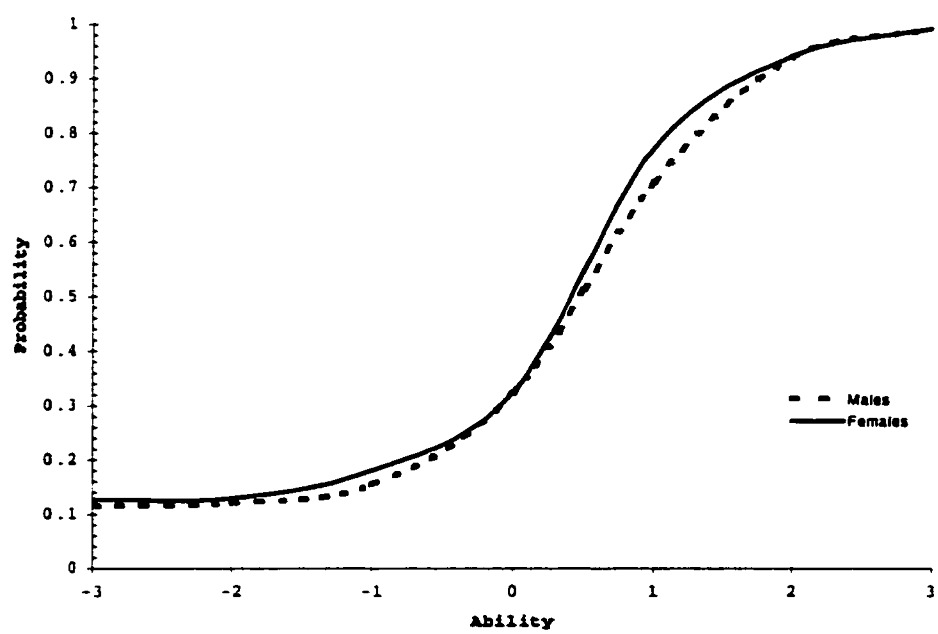
$$P_i(\theta) = \frac{e^{D\bar{a}(\theta - b_i)}}{1 + e^{D\bar{a}(\theta - b_i)}}(i = 1,2,...n).$$

In this case the ICC has a lower asymptote of 0 and the slopes of each ICC are identical and equal to 1.0.

Identifying DIF Using Item Response Theory

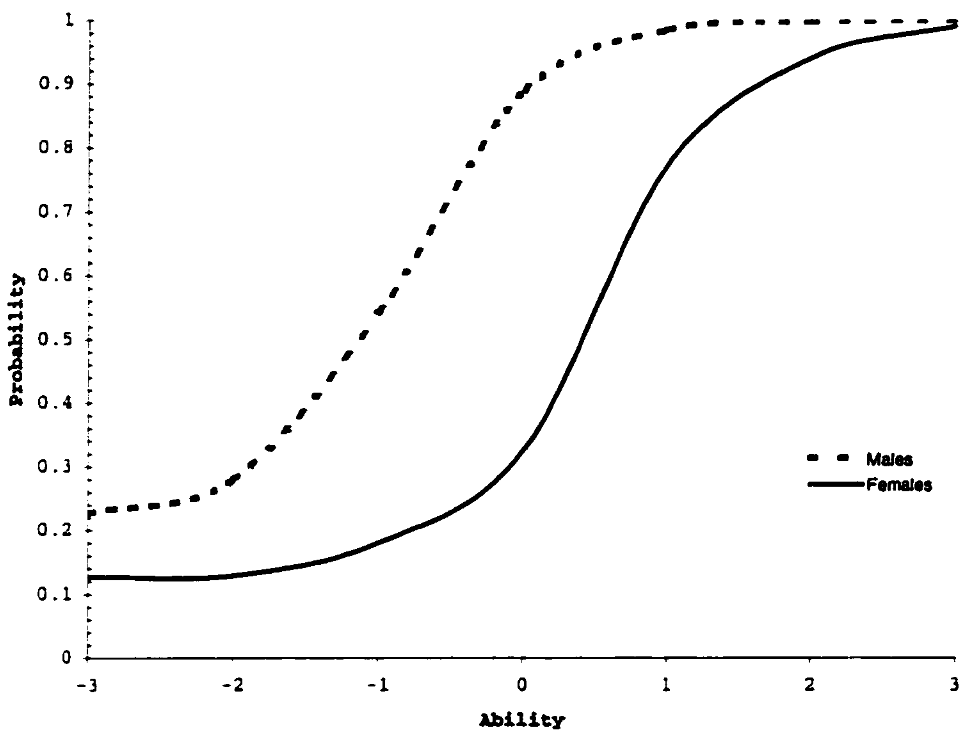
While there are several advantages of using IRT, these advantages can only be achieved when the model "fits" the data. Fitting is described as the "determination of how well a model accounts for a set of test data" (Hambleton & Swaminathan, 1985, p. 151). The ability and parameter estimates are considered to be invariant when "there is a close match between the model selected for use and the test data" (Hambleton & Swaminathan, 1985, p. 151). As a result, the same ICC can be obtained for each test item regardless of the ability distribution in the group of examinees used to estimate the item parameters (Hambleton, Swaminathan, & Rogers, 1991). By plotting the ICCs for each group (e.g., males, females), dichotomous items displaying DIF can be identified. If a test item has the same (or very close to the same) ICC for both groups then the item can be described as having no DIF. That is, members of both groups at the same ability level will have an "equal" chance of answering the item correctly as illustrated in Figure 2.

Figure 2. Item Demonstrating No DIF



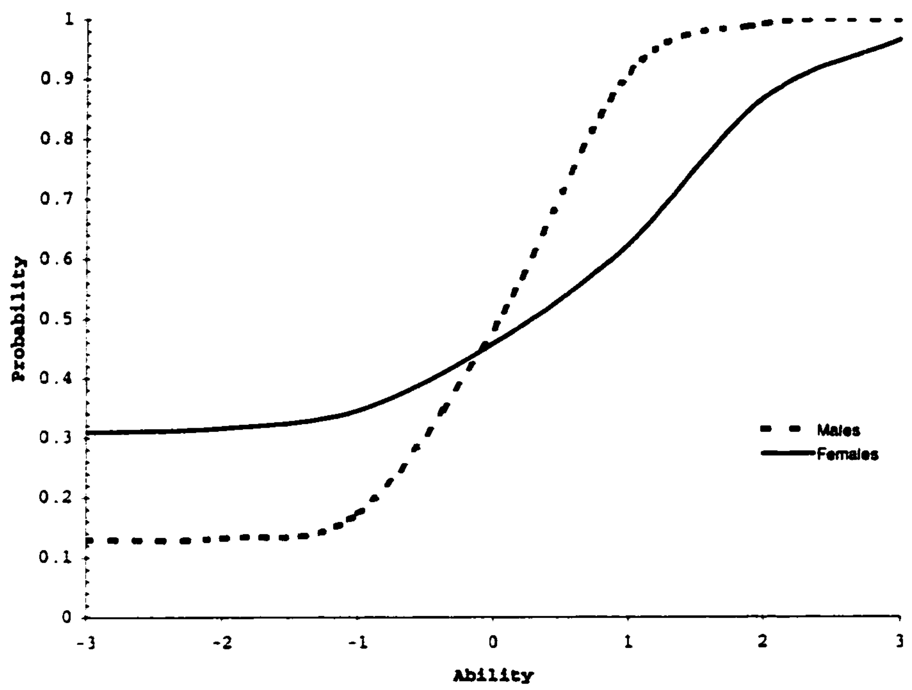
If an item is identified as displaying DIF, then the ICCs will be dissimilar. Two types of DIF can occur: uniform and nonuniform. Uniform DIF occurs when the probability of answering the item correct differs in the same way for all ability levels as seen in Figure 3. In this example, boys of all ability levels have a greater probability of answering the item correctly than girls at each ability level.

Figure 3. Item Demonstrating Uniform DIF



In contrast, non-uniform DIF occurs when there is an interaction between group membership and ability. In this case, the ICCs cross over, as illustrated in Figure 4. That is, the difference between the probabilities of each group getting the answer correct is different at different points along the ability scale. In this example, girls have a higher probability of getting the item correct than boys at lower ability levels. Conversely, boys have a higher probability of getting the item correct at higher ability levels.

Figure 4. Item Demonstrating Non-Uniform DIF



Despite the visual appeal of comparing the ICCs this method is not commonly used in DIF detection studies at the present time. There are two reasons for this. First, the magnitude of the DIF (or differences between the two ICCs) must be calculated. Second, statistical inferential procedures must be employed to determine if the calculated differences are shown to be reliable or are simply a function of chance. While procedures to calculate the area between two ICCs (Raju, 1988, 1990; Rudner, Geston, & Knight, 1980) and to test for statistical significance (e.g., Lord's [1980] chi-square; model comparison procedure [Thissen, Steinberg, & Wainer, 1993]) have been developed, these methods are not commonly used in DIF studies. This is primarily due to "the large amounts of data required to yield stable findings, it's attendant cost, and more than likely, the complexity of the model itself and the frequent failure of operational tests to satisfy the assumptions of the model" (Angoff, 1993, p. 9).

Furthermore these "area" methods are limited to use with dichotomous items. Polytomous items, like essays, are not simply scored correct or incorrect. Instead the answer is rated based on a predefined scoring rubric that may incorporate several different categories for one or more scales. Hence the item is described in terms of the probability of responding in a particular category score for the particular scale. The ICCs of polytomous items are more complex and subsequently less useful for the purposes of identifying DIF (French & Miller, 1996). Procedures have been developed that identify DIF based on statistically significant differences in the *b*-parameters for polytomous items (Muraki & Bock, 1997; Potenza & Dorans, 1995). However given the difficulty in the estimation procedures, this approach is not recommended (R. K. Hambleton, personal communication October 23, 1998).

Non-IRT DIF Detection Methods

At the current time, the most popular methods used to detect DIF are the Mantel-Haenszel (MH), Simultaneous Item Bias Test (SIB), and Logistic Regression (LR) methods. While these methods are not based on IRT, like IRT they determine DIF by matching examinees on ability. Initially these methods were developed for the detection of DIF in dichotomously scored items. More recently, extensions of each of these methods have been made for polytomous data. However, the study and application of these methods is more limited.

Several studies have been completed using MH, SIB, and LR to assess their differential effectiveness with dichotomous data. For example, Narayanan and Swaminathan (1994) found that MH and SIB were equally powerful in detecting uniform DIF items. Similarly Rogers and Swaminathan (1993) found that MH and LR identified similar uniform DIF items, although LR was more powerful in detecting nonuniform DIF items, particularly if the items were in the middle of the range of item difficulties (Mazor, Clauser & Hambleton, 1992).

While larger samples provide maximum power, all three methods can also be used with smaller samples. Each method has demonstrated sufficient power to detect DIF in sample sizes as small as 100 (e.g., MH and SIB) and 200 per group (LR) (Hambleton, et al., 1993, Li, Nandakumar, & Stout, 1995; Mazor et al., 1992; Roussos & Stout, 1996; Zwick, et al., 1993; Spray, 1989; Swaminathan & Rogers, 1990). Inflated type one error rates in DIF detection are more likely to occur with MH, SIB, and LR when there are differences in the ability distributions of the reference and focal groups, or among highly discriminating, low difficulty or high difficulty items (Rogers & Swaminathan, 1993; Roussos & Stout, 1996).

Each of these methods is discussed below. First the dichotomous form of each is presented. The generalized form of each, designed to accommodate polytomous items, is then outlined.

Mantel-Haenszel

One of the more commonly used methods used in DIF detection studies is the MH statistical procedure (Mantel & Haenszel, 1959). Holland and Thayer (1988) demonstrated that the MH procedure is related to the one-parameter IRT model when the associated underlying model assumptions are satisfied. Swaminathan and Rogers (1990) also suggest that MH can be thought of as being based on a LR model, where the ability variable is discrete and no interaction between group and ability is permitted (p. 363).

The MH DIF detection procedure uses contingency tables to compare the probability of success on each item for the two groups of interest after matching on ability. In order to compare the probabilities of a correct response, item response data for the reference (R) and focal (F) group members are arranged into a series of 2 x 2 contingency tables, one for each score level of the item. For each item, K 2 x 2 tables are constructed, where K is the number of unique scores for the test. The associated index of DIF, α_{MH}, is a constant odds ratio and is interpreted as the average factor by which the odds that an examinee from the reference group will answer the item correctly exceed the odds of an examinee from the focal group. α_{MH} is written:

$$\alpha_{MH} = \frac{\sum_{j=1}^{K} A_j D_j / N}{\sum_{j=1}^{K} C_j B_j / N_j}, (j = 1, 2, \ldots K),$$

where

α_{MH} = odds ratio,

A_j = observed number of examinees in the reference group at score level j answering the item correctly,

B_j = observed number of examinees in the reference group at score level j answering the item incorrectly,

C_j = observed number of examinees in the focal group at score level j answering the item correctly,

D_j = observed number of examinees in the focal group at score level j answering the item incorrectly, and

N = total number of items studied.

α_{MH} ranges in value from zero to infinity. To put an upper bound on the value, a simple transformation to the "delta" metric is usually performed:

$$\Delta_{MH} = -2.35 \ln [\alpha_{MH}].$$

The resulting statistic is symmetrically distributed about zero with values of zero interpreted as no DIF. Positive delta values indicate DIF favoring the focal group and negative delta values indicate DIF favoring the reference group.

The associated MH statistic, χ^2_{MH}, is calculated from the K 2 x 2 tables for each item. χ^2_{MH} is distributed approximately as a chi-square statistic with one degree of freedom and is written:

$$\chi^2_{MH} = \frac{\left\{ \left| \sum_{j=1}^{K} [A_j - E(A_j)] \right| - 1/2 \right\}^2}{\sum_{j=1}^{K} \mathrm{var}(A_j)} (j = 1,2,\ldots K),$$

where

$$E(A_j) = \frac{N_{R_j} N_{\cdot 1_j}}{N_j},$$

$$\mathrm{Var}(A_j) = \frac{N_{R_j} N_{F_j} N_{1\cdot_j} N_{0\cdot_j}}{(N_j)^2 (N_j - 1)},$$

N_{R_j} = number of examinees in the reference group with score *j*,

N_{F_j} = number of examinees in the focal group with score *j*,

N_{1_j} = total number of examinees at score level *j* answering the item correctly,

N_{0_j} = total number of examinees at score level *j* answering the item incorrectly,

and

N_j = the total number of examinees at score level j.

Guidelines for interpreting the degree of DIF in test items have been established at ETS (Zwick & Erickan, 1989). Roussos and Stout (1996) have modified these guidelines to aid in the interpretation of DIF:

- Negligible or A-level DIF: Δ_{MH} is not significantly different from 0 OR Δ_{MH} is significantly different from 0 using χ^2_{MH} AND $|\Delta_{MH}| < 1$.
- Moderate or B-level DIF: Δ_{MH} is significantly different from 0 using χ^2_{MH} AND $|\Delta_{MH}|$ at least 1 but less than 1.5.
- Large or C- level DIF: Δ_{MH} is significantly different from 0 and $|\Delta_{MH}|$ is 1.5 or greater.

Generalized Mantel Haenzel

As a result of the increased use of constructed response items in standardized tests and the subsequent need to identify DIF among this type of item, extensions of the MH procedure have been proposed (Zwick, et al., 1993). To investigate DIF in items with ordered response categories, a test of conditional association proposed by Mantel (1963) has been used to compare the item means for the two groups of interest after matching on ability. In this method, item response data for the reference (R) and focal (F) group members are arranged into a series of 2 x T x K contingency tables, one for each item at each score level. For each K, 2 x T tables are constructed, where K is the number of levels of the matching variable and T is the number of response categories for the item (see Figure 5).

Figure 5. Contingency Table for kth level of the Matching Variable

Group	Item Response Categories					Total
	y_1	y_2	y_s	...	y_T	
Reference	n_{R1k}	n_{R2k}	n_{R3k}	...	n_{RTk}	n_{R+k}
Focal	n_{F1k}	n_{F2k}	n_{F3k}	...	n_{FTk}	n_{F+k}
Total	n_{+1k}	n_{+2k}	n_{+3k}	...	n_{+Tk}	n_{++k}

The associated statistic, *Mantel* χ^2 is calculated from the K 2 x T tables for each item and is distributed approximately as a chi-square statistic with one degree of freedom and is written:

$$Mantel\chi^2 = \frac{\left(\sum_{k=1}^{K} F_k - \sum_{k=1}^{K} E(F_k)\right)^2}{\sum_{k=1}^{K} Var(F_k)} \quad (k= 1,2,3,\ldots K),$$

where

$F_k = \sum_{t=1}^{T} y_t n_{Ftk}$ is the sum of scores for the focal group at the kth level of the matching variable;

$$E(F_k) = \frac{n_{F+k}}{n_{++k}} \sum_{t=1}^{T} y_t n_{+tk} \qquad (t=1,2,3,\ldots T)$$

where

n_{F+k} = total number of focal group members at the *k*th level of the matching variable,

n_{++k} = total number of group members at the *k*th level of the matching variable

y_t = possible item scores, and

n_{+tk} = total number of group members receiving the score, *t*, for the items at the *k*th level of the matching variable; and

$$Var(F_k) = \frac{n_{R+k} n_{F+k}}{n_{++k}^2 (n_{++k} - 1)} \left\{ \left(n_{++k} \sum_{t=1}^{T} y_t^2 n_{+tk} \right) - \left(\sum_{t=1}^{T} y_t n_{+tk} \right)^2 \right\} (t=1,2,3,\ldots T),$$

where

n_{R+k} = total number of reference group members at the *k*th level of the matching variable.

In the dichotomous case, the Mantel (GMH) statistic is identical to the MH statistic without the continuity correction. An assumption associated with GMH is that the odds ratios are constant across item score categories. Consequently, GMH identifies an overall or global amount of DIF, but cannot identify DIF at each item score category. While attempts to develop an associated effect size measure similar to Δ_{MH} for use with GMH have been made, additional work in this area is still required (Zwick et al., 1993). Currently, the guidelines for the interpretation of DIF with MH can be used for GMH in the polytomous case (G. Camilli, personal communication, March 24, 1999).

Simultaneous Item Bias Test

Developed by Shealy and Stout (1993), SIB is a model-based approach that includes a test of significance and an explicit correction for guessing. With SIB the test items are split into two subtests: a "studied" subtest and a "matching" subtest. The studied subtest contains potential DIF items while the matching subtest, often but not always, contains the rest of the items (Li et al., 1995). The matching subtest contains items that measure the construct of interest and are not suspected of functioning differently. Using the total test score from the matching subtest, examinees from the reference and focal groups are matched on ability by grouping them into *K* subgroups. Then, the performance of the examinees are compared across the reference and focal groups on the studied subtest (Li et al., 1995). The weighted mean difference between the reference and the focal groups on the studied subtest score across *K* subgroups is given by:

$$\hat{\beta}_U = \sum_{k=0}^{K} p_k d_k, (k = 0,1,...K),$$

where

$\hat{\beta}_U$ = estimated amount of DIF - positive values denote DIF against the focal group; negative values denote DIF against the reference group,

$\hat{p}_k$ = proportion of focal group examinees in subgroup *k*, and

$d_k = \overline{Y}^*_{Rk}$ - $\overline{Y}^*_{Fk}$: the difference in the adjusted means of the studied subtest between the reference and the focal group for each subgroup *k*.

The means of the studied subtest for the reference and the focal groups are adjusted to correct for any differences in the ability distributions of the two groups. If the ability distributions of the reference and focal groups are equal and the item does not contain DIF, then the difference in means on the studied subtest will equal zero. However, if there are differences in the ability distributions of the reference and focal groups, then the differences in means will not equal zero even when no DIF is present. This is due to the "incompatibility in the average scores of the matching subtest for the two groups within subgroup *K*" (Li et al., 1995, p. 7). To correct this problem, SIB adjusts the observed score on the matching subtest by estimating the true score for the reference and focal groups at the *K* subgroup level. To calculate this the

> equation for the linear regression of true score on observed score from Classical Test Theory [is used] with KR20 calculated as the slope of the regression line for each group...Then the average of these two scores is calculated....The corresponding adjusted mean scores on the studied subtest, $\overline{Y}^{*}_{Rk}$ and $\overline{Y}^{*}_{Fk}$, are obtained using a first order Taylor Series approach to adjust for focal and reference group differences in the estimated true scores for subgroup [score level] *K* (Li et al., 1995, p. 8).

The statistic for testing the null hypothesis of no DIF is given by:

$$B_U = \frac{\hat{\beta}_U}{\hat{\sigma}(\hat{\beta}_U)},$$

where

$\hat{\sigma}(\hat{\beta}_U)$ is the estimated standard error of $\hat{\beta}_U$:

$$\hat{\sigma}(\hat{\beta}_U) = \left(\sum_{k=1}^{K} \hat{p}_k^2 \left(\frac{1}{N_{Rk}} \hat{\sigma}^2(Y \mid k, R) + \frac{1}{N_{Fk}} \hat{\sigma}^2(Y \mid k, F) \right) \right)^{1/2},$$

where

N_{Rk} and N_{Fk} = the number of examinees in the reference and focal groups respectively, in subgroup k, and

$\hat{\sigma}^2$ = the variances of the studied subtest scores for the focal and reference groups which are assumed to be equal.

To interpret the amount of estimated DIF, Roussos and Stout (1996, p. 220) suggest:

- Negligible or A-level DIF: $\hat{\beta}_U < 0.059$
- Moderate or B-level DIF: null hypothesis is rejected and $0.059 \leq \hat{\beta}_U < 0.088$.
- Large or C-level DIF: null hypothesis is rejected and $\hat{\beta}_U \geq 0.088$.

Poly-SIBTEST

SIB has also been generalized for use with polytomous items (Poly-SIB; Chang et al., 1996). Two modifications are required to accommodate maximum possible scores greater than one (if the maximum score is one, then the item is dichotomously scored). First the K in the SIB statistic ($\hat{\beta}_U$) used to estimated DIF is replaced with K_H, where K_H is the sum of all item scores (i.e., the maximum possible matching score). Secondly, the matching-test reliability estimate originally used in the regression correction (KR-20) is replaced with the more general coefficient alpha.

Like GMH, while Poly-SIB calculates a global amount of DIF for each item, it does not provide an indication of DIF at the individual score categories (A. G. Froelich, personal communication, July 6, 1999). In addition, Poly-SIB can be used with examinations containing both dichotomous and polytomous items. However, unlike

GMH, the associated effect size measure, $\hat{\beta}_U$, is interpreted using the same guidelines as SIB for both dichotomous and polytomous items.

Logistic Regression

Logistic regression (LR; Swaminathan & Rogers, 1990) identifies DIF by predicting the probability of a correct response to an item for the two groups of interest:

$$P(u_{ij}=1|\theta_{ij}) = \frac{e^{(\beta_{0j}+\beta_{1j}\theta_{ij}+\beta_{2j}g_i+\beta_{3j}\theta g_i)}}{1+e^{(\beta_{0j}+\beta_{1j}\theta_{ij}+\beta_{2j}g_i+\beta_{3j}\theta g_i)}},(i=1,2,...n_j);(j=1,2),$$

where

u_{ij} = the probability of the response of person i in group j to the item (0 if the response to the item is incorrect; 1 if the response is correct),

θ_{ij} = the ability of individual i in group j,

g_i = the group membership of individual i,

θg_i = the ability by group membership interaction of individual i,

β_{0j} = the weight associated with the intercept,

β_{1j} = the weight attached to the score variable θ,

β_{2j} = the weight attached to the group variable g, and

β_{3j} = is weight attached to the estimated interaction between group and ability.

A one degree-of-freedom chi-square test of significance is used to determine the presence of uniform DIF by testing the improvement in model fit when the group membership (θg_i) term is added to the model containing only ability (θ_{ij}). A second one degree-of-freedom chi-square test of significance is used to determine the presence of nonuniform

DIF by testing the improvement in model fit when the interaction ($\theta_i g_i$) term is added to the model containing only ability (θ_{ij}).

A measure of effect size, quantifying the magnitude of DIF, has recently been developed using the partitioned weighted least-squares estimate of R^2 for each term (Zumbo & Thomas, 1996). The effect size measure associated with uniform DIF is calculated by subtracting the $R^2\Delta$ for the total score from the $R^2\Delta$ for group membership (g). Similarly the effect size measure associated with nonuniform DIF is calculated by subtracting the $R^2\Delta$ for the total score from the $R^2\Delta$ for score-by-group membership (θg).

Based on Cohen's (1988) effect size measures for $R^2\Delta$ (small (0.02), medium (0.13), and large (0.26)), Zumbo and Thomas (1996, p. 8) suggested the following guidelines for the classification of $R^2\Delta$:

- Negligible or A-level DIF: chi-square test for model fit is not statistically significant or $R^2\Delta < 0.13$.
- Moderate or B-level DIF: chi-square test for model fit is statistically significant or $0.13 \leq R^2\Delta < 0.26$.
- Large or C- level DIF: chi-square test for model fit is statistically significant and/or $R^2\Delta > 0.26$.

These guidelines have been revised twice, once by Gierl and McEwen (1998) and later by Jodoin (1999). Gierl and McEwen (1998) compared MH, SIB, and LR with actual test data. They revised the the cut-offs for LR so that the results associated with LR coincided more closely with MH and SIB. Their recommended cut-offs were: negligible DIF: $R^2\Delta < 0.04$, moderate DIF: $0.04 \leq R^2\Delta < 0.08$, and severe DIF: $R^2\Delta \geq$

0.08. More recently, Jodoin (1999) proposed a further revision based on the SIBTEST effect size measures. Based on a computer simulation study, he suggested the following cut-offs: negligible DIF: $R^2\Delta < 0.035$, moderate DIF $0.035 \leq R^2\Delta < 0.070$, and severe DIF $R^2\Delta \geq 0.070$.

Logistic Discriminant Function

The LR DIF procedure has also been extended for use with polytomous items (Miller & Spray, 1993; Rogers & Swaminathan, 1993). Unfortunately, this approach is "unwieldy" and difficult to interpret as it involves a series of pairwise comparisons between score categories or combinations of score categories (Miller & French, 1996; Miller & Spray, 1993). Since LR requires a dichotomous dependent variable, the polytomous data must first be recoded into a number of dichotomous variables before completing several separate regression analyses. Three models, the continuation ratio logits, cumulative logits, and adjacent categories, have been used (French & Miller, 1996). Each involves a different coding scheme and uses the logit or the ratio of the probability of getting the category "correct" to the probability of getting the category "incorrect". In a simulation study, French and Miller (1996) used LR to detect DIF in polytomous items and suggested that the adjacent category coding method may be most useful in post hoc analyses to examine the location of the interaction between group membership and score level. In addition to the complexity associated with this method, measures of effect size to interpret the magnitude of DIF have not yet been developed.

An alternative method that identifies both uniform and nonuniform DIF among polytomous items is logistic discriminant function (LDF) which predicts group membership based on ability and item score (Miller & Spray, 1993). In this method the ratings or score points associated for each polytomous item are considered random variables and can take on any one of the values associated with the scoring scale. The discriminant function predicts the probabilities of group membership, g, given the scale score, u_{ij}, and the estimate of ability, θ_{ij}:

$$P(G|\theta_{ij},u_{ij}) = \frac{e^{(\beta_{0j}+\beta_{1j}\theta_{ij}+\beta_{2j}u_{ij}+\beta_{3j}\theta u_{ij})}}{1+e^{(\beta_{0j}+\beta_{1j}\theta_{ij}+\beta_{2j}u_{ij}+\beta_{3j}\theta u_{ij})}},(i=1,2,...n_j)(j=1,2),$$

where

G = the group membership variable,

θ_{ij} = the ability of individual i in group j,

u_{ij} = the rating of person i on scale j,

θu_{ij} = the ability x rating interaction,

β_{0j} = the weight associated with the intercept,

β_{1j} = the weight associated with the score variable, θ_{ij},

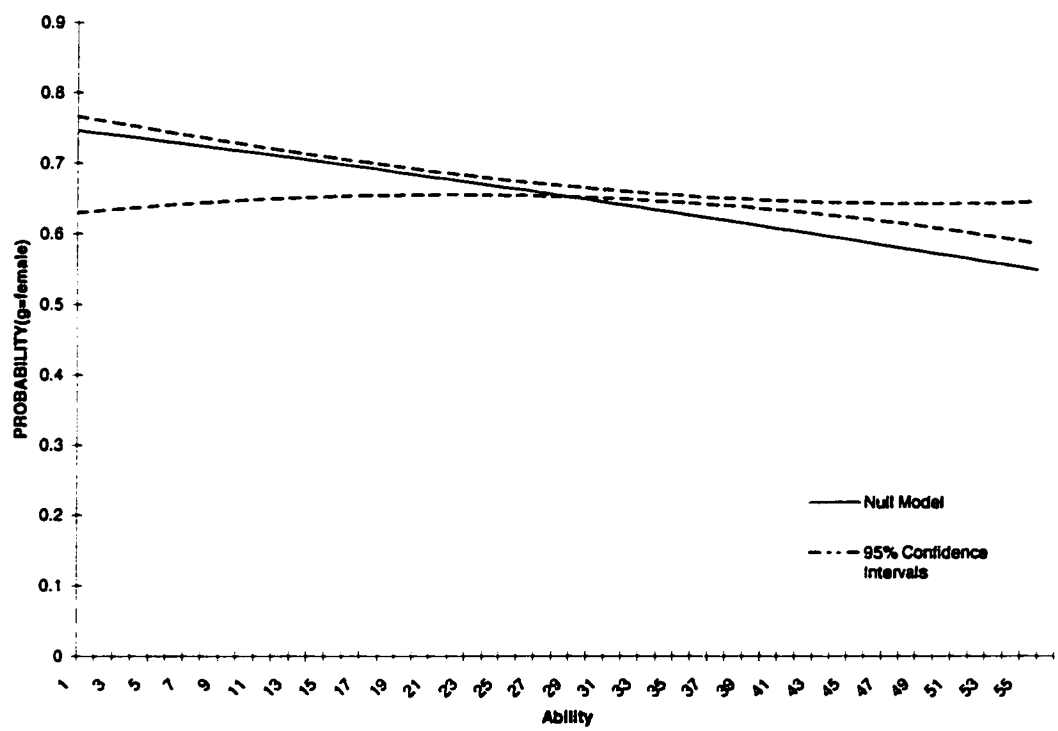
β_{2j} = the weight associated with the scale score, u_{ij}, and

β_{3j} = the weight associated with the estimated interaction between scale score and ability θu_{ij}.

A one degree-of-freedom chi-square test of significance is used to determine the presence of uniform DIF by testing the improvement in model fit when scale rating term (u_{ij}) is added to the model containing only the ability (θ_{ij}) and intercept terms. A two degree-of-freedom chi-square test of significance is used to determine the presence of nonuniform DIF by testing the improvement in model fit when the interaction between ability and scale rating (θu_{ij}) is added to the model.

Unfortunately, measures of effect size to help interpret the level of DIF have yet to be developed for this method. Instead, Miller and Spray (1993) suggest that for each item identified with DIF, simultaneous Scheffé-type confidence bands can be constructed around the estimated discriminant function for each item category score and compared to the function for the null model (i.e., the model without the coefficients β_{3j} and β_{2j}). "If the confidence band includes the line for the null model for most values of X [ability], then the actual severity of DIF for that category may not be serious" (Miller & Spray, 1993, p. 110). Where the confidence band falls above the null line, a greater proportion of focal group members than reference group members responded at the item score point. For example, in Figure 6, the limits for the 95% confidence interval associated with the discriminant function for an item score of 3 are plotted as dashed lines; the null line is solid. In this diagram the null line falls below both limits for the full model discriminant function for all scores greater than 29. This indicates that a greater proportion of focal group members with scores greater than 29 were awarded the scale score of 3 on the item. Miller and Spray (1993) also suggest that the ability distributions of the focal and reference groups should be considered in the identification of DIF within an item score category. That is, further investigation of DIF would not be required if it occurred at ability scores that none of the examinees achieved. For example, if all of the examinees received a score of 25 or more and the null line fell outside of the confidence interval for the full model discriminant function between the scores of 5 and 15, then the DIF would be of no practical consequence and the item would not be considered as displaying DIF.

Figure 6. Illustration DIF Detection using LDF



Gender DIF Across Examinations and Formats

Differences between the performance of females and the performance of males on various standardized tests have been the subject of much research. Studies completed to date have found that high school males generally outperform females in standardized tests of science, mathematics, history, and social studies even though females generally have similar or higher school awarded marks (Benbow, 1988, Doolittle & Cleary, 1987; Doolittle, 1989; Halpern, 1992; Stanley, Benbow, Brody, Dauber & Lupkowski, 1992; Wightman, 1998). Conversely, females generally outperform males in tests of verbal and written abilities, especially if constructed response items are included (Mazzeo, Schmitt, & Bleistein, 1993; Willingham & Cole, 1997). However, unlike tests of quantitative ability, the gender differences observed in tests of verbal and written abilities are usually small and of little practical significance (Wightman, 1998).

While underlying reasons for these differences continue to be debated in the literature (see Benbow, 1988; Halpern, 1992; Willingham & Cole, 1997), exploration at the item level has identified several trends. In studies of examinations designed to measure quantitative abilities, males tend to perform better than females on items related to geometry, ratio, proportions, and items containing tables, graphs, or figures. In contrast, females tend to perform better than males on items related to computation and items containing symbols (Burton, 1996; Doolittle & Cleary, 1987; Harris & Carlton, 1993; O'Neill & McPeek, 1993). DIF in mathematics and science may reflect differences in confidence levels, interests, attitudes towards the subject and the amount of course work completed (Willingham & Cole, 1997). However, these hypotheses have yet to be systematically investigated.

Relatively fewer studies have been conducted among examinations of literature, history, and the humanities. However, results from DIF studies completed on dichotomously scored measures of verbal ability indicate that males tend to perform better than females on items related to science and on items referring to stereotypical male activities. In contrast, females tend to perform better than males on items related to aesthetics and human rights and on items referring to stereotypical female activities (Mazzeo, et al., 1993; O'Neill & McPeek, 1993; Sadker & Sadker, 1994).

These trends are primarily based on a collection of studies investigating the prevalence of DIF on several different large standardized tests used for selection, placement, and scholarship awards. Unfortunately, for many items, the results are inconsistent, and in some cases, contradictory (Willingham & Cole, 1997). Furthermore, there is often no satisfactory explanation for the occurrence of DIF (Angoff, 1993; Camilli & Shepard, 1998; Gierl & McEwen, 1998). DIF may be related to an accumulation of the effects of several individual item characteristics or certain combinations of these characteristics (Bond, 1993). Or DIF may be dependent on the sample used in the study. It is likely that inconsistent and contradictory findings are related to the use of different types of samples (e.g., convenience, representative, self-selected) or samples that differ by age, grade, and courses completed. More research controlling for sample differences across examinations may contribute to the understanding of gender DIF across content areas.

In addition to the hypothesis that gender DIF is related to specific content, there is also some evidence to indicate that gender DIF is also related to item format (Bolger & Kellaghan, 1990; Burton, 1996; Dunbar, Koretz, & Hoover, 1991; Froese, 1998; Lane, Wang, & Magon, 1996; Mazzeo, et al., 1993; O'Neil & Brown, 1997; Willingham & Cole, 1997; Zwick, et al., 1993). For example, this evidence indicates that males generally perform better than females on multiple choice items while females perform better than males on essay items (Bolger & Kellaghan, 1990; Burton, 1996; Mazzeo et al., 1993; Sadker & Sadker, 1994). Even in measures of quantitative abilities, females tend to perform better than males when constructed response items are included (Burton, 1996; Lane et al., 1996). Although not statistically significant, Burton (1996) found that males performed better on the multiple choice and numerical response items of the SAT-M, while females performed better on the constructed response items. In addition, Lane et al. (1996) found that six of eight constructed response items in a grade six mathematics achievement test favored females. This difference has been attributed to the stronger writing skills and neater, more complete answers that are provided by females (Lane et al., 1996; Mazzeo et al., 1993; Willingham & Cole, 1997). However, as pointed out by Willingham and Cole (1997), these findings are still tentative primarily due to the lack of studies investigating DIF across item format and subject area.

The tentative nature of these findings is attributable to two reasons. First, few DIF studies have investigated the prevalence of DIF across item type and subject area. Most studies are completed on large-scale standardized tests of achievement, selection, placement, and scholarship consisting only of dichotomous items and administered to large heterogeneous samples. Second, results found in one study are frequently inconsistent with results from subsequent studies of the same examination (O'Neill & McPeek, 1993). This inconsistency is attributable in DIF studies to the use of different examinations in the same subject area and/or to the use of the same examination but with different samples of examinees (Garner & Englehard, 1999; O'Neill & McPeek, 1993). Therefore, exploration of the prevalence of DIF favoring males or females may be confounded by both differences between examinations and between samples. Studies designed to help disentangle these issues are required.

Although the use of a common sample was not possible, O'Neill and McPeek

(1993) attempted to determine if any patterns of gender DIF existed across several different standardized tests using previously completed DIF studies. Results from the following examinations were included: American College Testing Program Assessment (ACT), Graduate Management Admissions Test (GMAT), Graduate Record Examinations (GRE) General Test, NTE Core Battery (NTE), and the Scholastic Achievement Test (SAT). Based on the DIF results completed on each of the examinations, the authors hypothesized that differences in group performance could be explained by a number of different item characteristics. For each characteristic identified (e.g., gender stimulus) corresponding categories (e.g., the item refers to males; the item refers to females) were created. DIF items were then classified into these categories and a series of F-tests completed.

Based on their findings, O'Neill and McPeek (1993) indicated that there were certain item characteristics that were common to more than one examination. For example, males outperformed females matched for ability on reading comprehension items with science-related content. Males also outperformed females in antonyms and analogies found on the GRE and SAT items classified as "science" or "world of practical affairs" (topics in this classification included the use of tools and mechanical objects, sports, and history). When matched on quantitative score, males performed better than females on geometry and mathematics problem-solving items. In addition, males outperformed females on word problems involving novel or actual situations although females performed better on pure mathematics items such as formulas, equations, or theories.

While this is an admirable attempt to integrate DIF results, caution should be used in the interpretation and generalization of these findings due to the limitations associated with the O'Neill and McPeek study. First, the study only included those examinations in which a previous DIF study using MH had been completed. Second, the item characteristics associated with DIF in this study were based on the mean of all classified items, as few individual items had high DIF values. Third, only a small proportion of items were identified as exhibiting DIF (generally less than 10%) for each examination; however items were classified on a minimum of 30 different characteristics. As a result, a number of within-characteristic category cells were empty or small. Finally, the study was limited to selection-response items; therefore differences related to item format (e.g., multiple choice, constructed response) were not examined. Since a greater number of standardized examinations containing both types of items are being created, a better understanding of the prevalence of DIF across item format is required.

Studies of Gender DIF in Alberta Education Diploma Examinations

In Alberta, all students completing Grade 12 level courses are required to complete standard exit examinations. Each examination is subjected to a sensitivity review and content analysis before administration. However, despite the careful development of these tests, differences between the overall performance of male and female students remain. For example, the mean scores for male students are higher than the mean scores for female students on the Biology 30, Mathematics 30, and Social Studies 30 Diploma Examinations. Conversely the mean score for female students are higher than the mean scores for male students in English 30 (see Table 1).

Table 1.

Average Diploma Examination Mark by Year

Subject	Female Students		Male Students	
	1995/96	1996/97	1995/96	1996/97
Biology 30	63.0(15.6)	62.3(15.8)	65.4(16.0)	64.4(16.5)
English 30	67.3(12.7)	65.2(13.0)	65.5(12.5)	64.0(13.1)
Mathematics 30	60.9(17.8)	61.1 (17.7)	62.9(18.6)	63.4(18.1)
Social Studies 30	63.1(14.9)	62.1(14.3)	67.1(14.0)	66.5(13.9)

Note. Standard deviations are noted in brackets. From Alberta Education Annual Report 1996-97; by Alberta Education, 1996, Edmonton, AB: Author; Alberta Education Annual Report 1996-97; by Alberta Education, 1997a, Edmonton, AB: Author.

While these results are similar to the results from other examinations in these subject areas (Sadker & Sadker, 1994; Willingham & Cole, 1997), as discussed earlier, differences in overall mean scores are an unreliable method to determine if a specific group of examinees had an unfair advantage on individual test items (Camilli & Shepard, 1994).

The Alberta Education Diploma Examinations contain a mixture of selection and constructed response items and provide an opportunity to explore the prevalence of DIF across content areas and item format. While two DIF studies using the 1996 results of Social Studies 30 Diploma Examination have been completed, neither included the polytomous items. The first study was undertaken by the Alberta Education Student Evaluation Branch in response to the "consistent and worrying pattern" of higher mean scores for males as compared to females (Alberta Education, 1996, p. 26). Using only the June 1996 administration results, 7.0% of items were identified as displaying moderate DIF using MH, SIB, and the signed/unsigned area method (Alberta Education, 1996, p. 27). No items were identified with large DIF.

In a second study, Gierl and McEwen (1998) explored the prevalence of gender DIF in both the January and June 1996 Social Studies 30 examinations using MH, LR, and SIB. In each case the percentage of items identified with DIF varied by procedure. MH identified the fewest items and SIB identified the most. No nonuniform DIF was detected. Although different results were obtained with each procedure and administration, less than 20% of items were identified as displaying DIF. For the January 1996 examinations, the MH procedure identified four items with uniform DIF, whereas LR and SIB identified five and 11 items respectively. On the June 1996 examination MH identified eight items, LR nine, and SIB ten. In all cases more items favored males. While several items were identified, using judgmental review, only four items were consistently identified by content experts. Possible reasons for the presence of DIF were not postulated in this study.

Taken together, the results described above indicate that while there are many studies investigating the prevalence of DIF among dichotomous items, there are fewer studies investigating the prevalence of DIF among polytomous items or the comparison of different methods with actual data. The examination of DIF in polytomously-scored items across different examinations using a common sample has also not been completed. Nor has the investigation of DIF in examinations composed of both types of items. As more standardized examinations containing both types of items are created, a better understanding of DIF across item format and the implications that any interaction may have on test performance is required.

Unfortunately, DIF results found in one study are frequently inconsistent with subsequent studies of the same examination (O'Neill & McPeek, 1993). These differences have been attributed to different: 1) ages and grades of cohorts; 2) constructs; and 3) types of samples (e.g., self-selected, representative, or convenience; Willingham & Cole, 1997). Consequently item characteristics identified as possibly causing DIF in studies completed to date can only be interpreted within the context of the specific examination and sample.

The purpose of this study, therefore, was to investigate the polytomous DIF detection methods, GMH, Poly-SIB, and LDF. Both the comparability to their dichotomous counterparts and the comparability among the three methods in examinations containing dichotomous and polytomous items were examined. The variable gender was selected to investigate the performance of the polytomous DIF detection methods for two reasons. While there are several variables to stratify the two groups for DIF detection studies such as race, school location (urban/rural), this demographic variable is uniformly collected across all the studied examinations, was readily available, and provided an adequate focal group sample across all four examinations. Second, this variable provided the opportunity to investigate hypothesized differences between males and females. Unlike previous studies, a common sample of examinees across subject area and item formats was employed. Examining DIF items across different examinations containing both dichotomous and polytomous items within common samples should permit a more clear examination of where males outperform females, females outperform males, and where the performance is the same.

Chapter 3

Method

To answer the questions posed for this study, data from four different Alberta Education Diploma Examinations written by common samples of students were analyzed. The examinations included English, Social Studies, Mathematics and Biology. In all cases the exams included both dichotomous and polytomous items. Various samples of examinees that completed identical forms of the four examinations were analyzed to first determine if the polytomous versions of the common DIF detection methods performed similarly to their dichotomous counterparts. This analysis was followed by analyses of the comparability of the polytomous DIF detection methods. Lastly, further analyses to determine the prevalence of gender DIF across item format and subject area were completed. The purpose of this chapter is to describe, the examinations studied, the student samples, and the procedures used in this study. The examinations and student samples included in this study are described first, followed by a description of the procedures completed.

Before proceeding, it should be noted that two approaches were considered to investigate the first two questions. One procedure involved the use of simulation studies and the other procedure involved the use of empirical data from a standardized testing program. The latter approach was adopted for the following reasons. First, a number of simulation studies have investigated the dichotomous DIF detection methods with a variety of conditions (Hambleton & Rogers, 1989; Jodoin, 1999; Li, Nandakumar, & Stout, 1995; Mazor, Clauser, & Hambleton, 1992; Mazor, Kanjee, & Clauser, 1993;

Miller & Spray, 1993; Narayanan & Swaminathan, 1994; Rogers & Swaminathan, 1993; Roussos & Stout, 1996; Swaminathan & Rogers, 1990). These findings have found that DIF methods produce stable results with Type I error rates at nominal levels (e.g., < 5%). As the primary purpose of this study was to determine the comparability of the polytomous DIF detection methods to their dichotomous versions, it was felt that further simulation studies was not warranted as the dichotomous results were the point of comparison. Second, while the use of simulated data would provide the opportunity to determine the accuracy of each method in identifying "true" DIF items, the generalization to empirical data would be limited to the similarity of the generated data. Finally, simulation studies are generally recommended in situations where an analytic solution for a problem does not exist or is impractical because of its complexity (Harwell, Stone, Hsu & Kirisci, 1996). The use of empirical data provided an opportunity to investigate the functioning of the various DIF detection methods under actual rather than ideal conditions and can be viewed as an extension of simulated studies.

Examinations to be Studied

The Alberta Department of Education established the Grade 12 Diploma Examinations Program in 1984 to: 1) ensure the maintenance of provincial standards of achievement; 2) certify the level of individual student achievement in selected Grade 12 courses, and 3) report individual and group results (Alberta Education, 1997a, 1998a). The results from these examinations account for 50% of the total awarded mark in each subject for which an examination is available. Questions on the exams are based on concepts, topics, and facts from the Alberta Education Program of Studies that are to be included in the curriculum for all students in the Province of Alberta. These examinations are administered four times a year, however, the majority of students complete them in either January or June. While the examinations are carefully constructed and screened for potential sources of bias by Alberta Educational Staff, no DIF statistical analyses are completed.

For this study four examinations were chosen that represented both the humanities and the sciences: English 30, Social Studies 30, Mathematics 30, and Biology 30. Each Diploma Examination contains a mixture of dichotomous and polytomous items (see Table 2).

Table 2.

Structure Comparison of Examinations

EXAM	Number of Items		Percentage of Mark	
	Dichotomous	Polytomous	Dichotomous	Polytomous
English 30	70	6	50	50
Social Studies 30	70	4	50	50
Mathematics 30	49*	3	70	30
Biology 30	56**	2	70	30

Notes: *Includes 9 numerical responses. **Includes 8 numerical responses

The number of dichotomous items ranges from 49 items in the Mathematics 30 examination to 70 items in the English 30 examination. Dichotomous items in the English 30 and Social Studies 30 examination consist only of multiple choice items. Dichotomous items in the Mathematics 30 and Biology 30 examinations also include 9 and 8 numerical response items in which the students "grid" in their answers, respectively. While the numerical response items in Mathematics involve calculations, in Biology, these items are also used to record answers to matching, fill-in-the blank, diagram labeling, and ordering a sequence of events.

The dichotomous items are also described and classified according to the examination blueprint provided by Alberta Education. For English and Social Studies, course content and cognitive domain are used to classify these items. For Mathematics, course content and level of mathematical understanding classify the items. For Biology, items are classified by course content alone.

The number of polytomous items varied from 2 to 6 items. The fewest number of items were on the Biology 30 examination, the largest on the English 30 examination. The item types and associated scoring method included on each examination also varied. In English the items were rating scales associated with two essays. The first two items were scales associated with a minor essay; the remaining four items were associated with a major essay. In Social Studies, the six items were also rating scales, but they were associated with only one essay assignment. In both examinations the points on the rating scales ranged from 1 to 5. The mark assigned was the combined average of the scores awarded by two independent raters. In the case of imperfect agreement between the raters, scores were assigned that fell mid-way between the ratings made. For example, if rater A assigned a 2 and rater B assigned a 1, then a score of 1.5 was awarded (Alberta Education, 1997a, 1997b, 1998a, 1998b). To maintain the integrity of the five-point rating scale for data analysis, those scores that fell between the scale scores were randomly recoded to the nearest score category. For example, all scores of 1.5 were randomly recoded to either 1.0 or 2.0.

The three polytomous mathematics items complex, multi-step problems. The total possible mark for each polytomous item was four. Specific scoring rubrics were used to assign a partial or full mark based on the degree of successful completion of the problem.

In Biology, two polytomous items were included on the examination. However, only one item was included in this study. The included item required students to evaluate data, incorporate previous knowledge with new information, form new hypotheses, and make predictions regarding future trends. This item was scored holistically on a scale of 1 to 4. The deleted item was the sum of four related items totaling 12 possible marks. As this scale was too large for the DIF detection procedures used in this study and because the individual item scores were unavailable, this item was deleted.

Sample

The samples of examinees for this study were drawn from the database of students that completed all four Diploma Examinations in the same academic year. Data for two school years were considered: 1996-97 (1997) and 1997-98 (1998). The examinations are administered four times per year, with the majority of students completing the examinations in either January or June. A different parallel examination is created for each administration. While the total number of students that completed all four examinations in the same academic year was more than 5500 per year, the number of students that wrote the same examination at the same sitting (January or June) in all four subjects is much smaller. This difference is due to the number of different possible combinations that the students could choose to complete the courses and write the examinations. Depending on the combination, the sample size can be approximately 100 (e.g., completing all four examinations in January) or greater than 500 (e.g., completing all four examinations in June). Sample sizes for all possible combinations are presented in Table 3. The patterns are given as a series of four digits, either a 1 for January or a 2 for June. The digits appear in the order English 30, Social Studies 30, Mathematics 30, and Biology 30. Thus a 1212 pattern represents students who took English and Math in January and Social and Biology in June.

Table 3.

Sample Sizes for Student Combinations

Sample Pattern	1996-97	1997-98
1111	55	50
1112	205	229
1121	128	129
1122	246	248
1211	441	435
1212	575	663
1221	432	368
1222	317	273
2111	388	395
2112	484	625
2121	445	368
2122	315	234
2211	500	550
2212	346	279
2221	349	282
2222	509	426
TOTAL	5735	5554

Note. The patterns are given as a series of four digits, either a 1 for January or a 2 for June, in the order English 30, Social Studies 30, Mathematics 30, Biology 30.

Comparisons of DIF Procedures

Six samples were used to compare the polytomous DIF detection methods to the dichotomous counterparts and to compare the polytomous methods. These included the four largest samples from the 1998 academic year (patterns 1212, 2211, 2112, and 2222, see Table 3), and the total samples that wrote each examination in January and in June. Inclusion of the total samples allowed for the comparison of all results for the smaller samples with the results for the total sample, thereby obtaining a better understanding of how the polytomous DIF detection methods performed in smaller samples as compared to their dichotomous counterparts. While the dichotomous DIF detection methods employed in this study have been used with small sample sizes, fewer studies using the polytomous versions have been completed. Smaller sample sizes may decrease the power of the methods to identify DIF items.

Prevalence of Gender DIF

The sample selected to investigate the prevalence of DIF across subject and item format included those students who wrote all four examinations in June (pattern 2222). While the number of students that completed this examination is smaller than some of the other samples, this pattern includes those students enrolled in full-term classes. This group is hypothesized to be the most homogeneous sample with similar academic and extracurricular interests, and completed course work. As hypothesized earlier, the occurrence of DIF may be related to different attitudes towards the subject area and different types of completed coursework.

Procedures

Comparisons of DIF Procedures: Dichotomous versus Polytomous Versions

The polytomous DIF detection methods were compared to their dichotomous counterparts (MH and GMH, SIB and Poly-SIB, LR and LDF) using the dichotomous items included in each of the four examinations for the six samples in each year. Specific computer programs were used for MH (Shealy & Stout, 1993), SIB (Shealy & Stout, 1993), and Poly-SIB (Chang et al., 1996). SPSS was used for GMH (G. Camilli, personal communication, July 17, 1998) and LR (M. G. Jodoin, personal communication, February 1, 1999). VISDIF (Leucht & Spray, 1993) was used for LDF, including the calculation and plotting of the 95% confidence intervals. Although LR and GMH allow the user to specify a matching variable, the programs associated with MH, SIB, Poly-SIB, and VISDIF do not. These programs match examinees on the score obtained on the studied set of items. Due to these limitations, all analyses were conducted in the same manner using the same matching variable (i.e., test score associated with items analyzed).

Identified DIF items were classified into two categories: those that exhibited no DIF or a negligible amount of DIF (A-level; denoted '1' in the Appendices) and those that exhibited moderate or large levels of DIF (B- or C-level; denoted '2' in the Appendices) based on the associated effect size measure for each method. With LDF, the items were classified as exhibiting DIF if the function for the null model fell primarily outside of the 95% Scheffé-type confidence intervals for the full model of interest (i.e., uniform or nonuniform). If the function for the null model was contained within the confidence bands, then the DIF was considered negligible. The numbers of DIF items identified using the dichotomous and polytomous version were then compared to determine the consistency of DIF detection across methods.

Comparisons of DIF Procedures: GMH, Poly-SIB, LDF

The polytomous DIF detection methods, GMH, Poly-SIB, and LDF were then compared twice in each of the six examinations described above. First using only the dichotomous test items, then using the entire set of test items (dichotomous and polytomous). The matching variable for the first of these analyses was the test score associated with the set of dichotomous items. In the second analysis this was the total test score obtained over all of the items. Identified DIF items were classified into two categories as described above using the guidelines associated with the effect size measure for each method.

Prevalence of Gender DIF

Poly-SIB, supplemented by LDF, was then used to determine the prevalence of gender DIF across item format and subject area within the sample of examinees that completed all of the exams in June. This combination of methods was selected for two reasons. First, similar to previously reported studies comparing the dichotomous DIF detection methods (e.g., Gierl & McEwen, 1998), the proportion of DIF items identified varied by method with Poly-SIB detecting the most items. Perfect agreement among GMH, Poly-SIB, and LR was not expected as convergence of different methods is influenced by the unreliability of the associated statistics (Hambleton & Rogers, 1989) and variable item classification results from associated effect size measures. However, as reported in the next chapter, in this study, a common set of items was identified by all three methods, most of which were subsumed within the items identified by Poly-SIB. Fewer common items were included in the items detected by the other two methods. In addition, as the most liberal procedure, it is likely that few DIF items would remain undetected.

Second, where possible, LDF was used to analyze the scale scores for the polytomous DIF items identified by Poly-SIB to clarify the points on the score scale at which gender DIF was present. Of the three methods included in this study, only LDF provides differential performance information by scale score. Comparisons of the number of items favoring males and females across subject and item format were then completed. The following chapter presents the results of these analyses.

Chapter 4

Results

The purpose of this chapter is to present and discuss the results of the analyses described in the previous chapter. The comparisons of GMH, Poly-SIB, and LDF to their dichotomous counterparts, MH, SIB, and LR are presented and discussed in the first section. The comparisons among GMH, Poly-SIB, and LDF are presented next. This is followed by the description of the prevalence of gender DIF across item format and subject area. In this third section overall performance differences are presented and discussed first followed by the results of the DIF analyses.

Comparability of Dichotomous and Polytomous DIF Detection Methods

The comparability between the dichotomous and polytomous versions of the DIF detection methods was completed with four different sample samples, as well as the entire samples of students who wrote the examinations in both January and June from the 1998 academic year. Only the dichotomous items from each examination were included for this part of the study. As described in the previous chapter, students were matched based on their performance on this section of the examination. That is, the matching variable was the simple sum of the dichotomous items and not the total test score.

The results of the comparison between MH and GMH, SIB and Poly-SIB, and LR and LDF are summarized in Table 4 for each subject area and sample. More detailed, item specific information can be found in Appendix A for MH, GMH, SIB, and Poly-SIB, and in Appendix B for LR and LDF. As shown in Table 4, the summary is organized in terms of lower triangles. The off-diagonal elements are those items with

Table 4.

Comparison of Dichotomous and Polytomous DIF Detection Methods for Moderate to Severe DIF

		MH/GMH								SIB/POLY-SIB								LR/LDF UNIFORM								LR/LDF NON-UNIFORM							
		English (i=70)		Math (i=49)		Social (i=70)		Biology (I=56)		English (I=70)		Math (i=49)		Social (i=70)		Biology (i=56)		English (i=70)		Math (i=49)		Social (i=70)		Biology (i=56)		English (i=70)		Math (i=49)		Social (i=70)		Biology (i=56)	
		D	P	D	P	D	P	D	P	D	P	D	P	D	P	D	P	D	P	D	P	D	P	D	P	D	P	D	P	D	P	D	P
Jan	D	0		0		1		0		0		0		0		0		3		1		0		4		0		0		0		0	
	P	11	0	5	0	5	0	4	0	17	0	6	0	9	0	7	0	14	2	5	1	7	1	4	0	0	0	0	0	0	0	1	0
	N	2395		3226		2178		2577		2395		3226		2178		2577		2395		3226		2178		2577		2395		3226		2178		2577	
June	D	0		1		1		0		0		1		0		0		4		2		3		0		0		0		0		0	
	P	5	0	3	0	9	0	3	0	9	1	7	0	17	0	3	0	7	3	3	2	12	5	2	3	0	0	0	0	0	0	0	0
	N	3159		2328		3276		2977		3159		2328		3276		2977		3159		2328		3276		2977		3159		2328		3276		2977	
1212	D	0		0		0		1		0		0		0		0		5		6		6		5		2		0		2		2	
	P	10	0	8	0	11	0	4	0	12	1	12	0	14	0	7	0	3	0	3	0	2	0	1	1	0	0	0	0	0	0	0	0
	N	663		663		663		663		663		663		663		663		663		663		663		663		663		663		663		663	
2112	D	1		0		0		0		0		0		0		1		6		1		4		5		1		0		1		1	
	P	8	0	4	0	7	0	5	1	10	0	6	0	12	1	5	0	2	0	1	0	3	0	3	0	0	0	1	0	0	0	0	0
	N	625		625		625		625		625		625		625		625		625		625		625		625		625		625		625		625	
2211	D	0		0		0		1		0		0		0		0		6		5		7		8		1		2		1		1	
	P	5	0	6	0	10	0	9	1	4	0	7	0	14	0	11	2	1	0	0	0	3	0	2	0	0	0	0	0	0	0	0	0
	N	550		550		550		550		550		550		550		550		550		550		550		550		550		550		550		550	
2222	D	0		0		2		0		0		0		0		0		3		5		2		3		0		2		0		0	
	P	9	0	5	0	10	0	2	0	7	0	4	0	12	0	6	0	3	0	0	1	3	2	0	0	0	0	1	0	0	0	0	0
	N	426		426		426		426		426		426		426		426		426		426		426		426		426		426		426		426	

Notes. Where zeros are noted in the diagonals, the agreement between two methods is 100%; D = Dichotomous version; P = Polytomous version, i = the number of items in the analysis.

moderate to severe DIF that were identified by both procedures. The diagonal elements are items identified solely by one, but not both methods. Where zeros are located in the diagonal elements, the agreement is 100%. Agreement on items identified with negligible or no DIF are not included in this table. The overall agreement, in percent, for each pair of methods for each examination and sample is presented in Table 5.

To illustrate these tables, consider the first lower triangle of Table 4. This lower triangle contains the item detection agreement between MH and GMH on the January English examination. In this example, the sample size is 2395. Both MH and GMH identified 11 items with moderate or severe DIF – this is noted in the off-diagonal element of the lower triangle. There were no items identified by MH and not GMH as seen from the zero in the first diagonal element. Similarly, the zero in the second diagonal element indicates that there were no items identified by GMH but not MH. That is, when there are zeros in the diagonals, the agreement between the two methods is 100% as reported in Table 5. For a second example, consider the ninth lower triangle in the first row of Table 4. This lower triangle contains the item detection agreement between LR and LDF on the January English examination. The sample size is 2 395. Fourteen items were identified with moderate to severe DIF with both methods, however LR identified three items not identified by LDF and LDF identified two items not identified by LR. The remaining 51 items were identified as exhibiting negligible or no DIF by both methods. Consequently, in contrast to the first example, the agreement between LR and LDF is 92.8%.

Table 5.

Percent Agreement between Dichotomous and Polytomous DIF Detection Procedures by Subject and Sample

	MH/GMH				SIB/POLY-SIB				LR/LDF UNIFORM				LR/LDF NON-UNIFORM			
	English (*i*=70)	Math (*i*=49)	Social (*i*=70)	Biology (*i*=56)	English (*i*=70)	Math (*i*=49)	Social (*i*=70)	Biology (*i*=56)	English (*i*=70)	Math (*i*=49)	Social (*i*=70)	Biology (*i*=56)	English (*i*=70)	Math (*i*=49)	Social (*i*=70)	Biology (*i*=56)
Jan	100%	100%	98.6%	100%	100%	100%	100%	100%	92.8%	95.9%	98.6%	92.9%	100%	100%	100%	100%
June	100%	97.9%	98.6%	100%	98.6%	97.9%	100%	100%	90.0%	91.4%	88.6%	74.6%	100%	100%	100%	100%
1212	100%	100%	100%	98.2%	98.6%	100%	100%	100%	92.8%	91.4%	91.4%	89.2%	97.1%	100%	97.1%	96.4%
2112	98.6%	100%	100%	98.2%	100%	100%	98.6%	98.6%	91.4%	97.9%	94.3%	91.1%	98.6%	100%	98.6%	98.6%
2211	100%	100%	100%	96.4%	100%	100%	100%	96.4%	91.4%	89.8%	90.1%	85.6%	98.6%	95.9%	98.6%	98.2%
2222	100%	100%	97.1%	96.4%	100%	100%	100%	100%	95.7%	87.8%	94.3%	94.6%	100%	95.9%	100%	100%

Note. *i*= items in the analysis

MH versus GMH

The agreement between MH and GMH is good. Of the 24 analyses conducted, 16 had 100% agreement between the two methods (see MH/GMH panel, Table 4). In the remaining eight cases, MH identified one more item in six cases and two more items in the sixth. In the eighth case, both MH and GMH identified a unique item. Taking into account the total number of items both identified and not identified by the two methods, these differences are quite small. Indeed, the lowest agreement was still high at 96.4% (see MH/GMH panel, Table 5). Further, the differences that did occur were not related to sample size, as they occurred in large as well as small samples.

The discrepancies between the effect sizes in six of the cases where the agreement was less than 100% were quite small, ranging from .014 to .236. Therefore, some of the variance in the eight "nonperfect" cases may be attributable to the arbitrary nature of the established guidelines used to determine DIF magnitude because different decisions can result when effect sizes are "centred" around a specific cut-point. For example, consider item 8 of the English 2112 sample. The effect size calculated by MH was –1.00 and –0.983 by GMH. Despite the small difference between the effect sizes (0.017), only MH flagged the item because the GMH effect size was slightly below the cut-point described in the established guidelines.

In summary, GMH demonstrated comparable results to MH. Across the full set of possible contrasts, the agreement between the two methods was 99.4%. Based on this result, and the results for the individual cells, it is concluded that GMH can be confidently substituted for MH for the detection of DIF in examinations with dichotomous items.

SIB versus Poly-SIB

The agreement between SIB and Poly-SIB is also good. Total (i.e., 100%) agreement between the two methods was noted in 19 of the 24 analyses conducted. Of the remaining five cases, SIB identified one more item than Poly-SIB in two cases, Poly-SIB identified one more item than SIB in two cases and two more in one case. Hence the agreement over all items in a given set ranged from 98.6% to 100%. These differences do not appear to be related to sample size as they occurred in both large and small samples.

Like the case for MH and GMH, the differences between the effect sizes in all of the cases where the agreement was less than 100% were quite small, ranging from .001 to .006. Again, some of the variance in the eight "nonperfect" cases may be attributable to the arbitrary nature of the established guidelines used to determine DIF magnitude because different decisions can result when effect sizes are "centred" around a specific cut-point. As an example, consider item 31 and the English 1212 sample (see Appendix A). The SIB effect size was 0.056, whereas the effect size measure for Poly-SIB was 0.059. Based on these results, the item was identified as displaying moderate DIF with Poly-SIB but negligible DIF with SIB even though the difference between the two effect sizes is only 0.003.

In summary, Poly-SIB demonstrated comparable results to SIB within an acceptable level of agreement (98.6% - 100%). Discrepancies between the two methods were noted to occur due to slight differences in effect size measures at the arbitrary cut-off point distinguishing negligible and moderate-level DIF. Based on these results, it is concluded that Poly-SIB can be confidently substituted for SIB for the detection of DIF in examinations with dichotomous items.

LR versus LDF

Larger differences were found between LR and LDF procedures than the differences found between the MH and GMH and the SIB and Poly-SIB procedures. While both methods identified the same statistically significant items 100% of the time (see Appendix B), when the guidelines for determining the magnitude of DIF (e.g., effect sizes; Scheffé confidence intervals) were applied there was less agreement. As shown in Table 4, there is not one cell in which there is perfect agreement (zero in both diagonal elements) between LR and LDF uniform[1]. Further the number of items identified by one but not the other procedure ranges from 1 to 7 with a mean of 3.69. Consequently, the percentage of agreement across the full set of items is lower than that observed for both MH/GMH and SIB/Poly-SIB, ranging between 74.6% and 98.6%.

In addition, there also appears to be a relationship to sample size. For the two largest sample sizes, both LR and LDF identified items with uniform DIF that the other methods did not identify. But among the smaller samples, LR identified more items with DIF than did LDF, and in three cases LR was the only method that identified DIF items.

---

[1] Agreement between the two methods in the detection of nonuniform DIF is higher (95.9% to 100%); however, this is a function of the small numbers of items identified with nonuniform DIF. In comparison to LR, very few items were identified by LDF (3 versus 19). Of those identified, LR identified all of them. As the detection of nonuniform items by LDF is subject to the same limitations that will be discussed with regard to uniform DIF, the results are presented in the tables, but not discussed beyond this footnote.

This latter finding is likely attributable to the statistical nature of the DIF magnitude method (95% Scheffé confidence intervals) associated with the LDF procedure. As the sample size decreased, the 95% confidence interval used to determine DIF became much wider. Consequently, fewer items or no items were identified with at least moderate DIF.

Like MH, GMH, SIB, and Poly-SIB, the definitions of the cut-points are somewhat arbitrary. While this study used the cut-offs suggested by Gierl and McEwen (1998), guidelines suggested by Zumbo and Thomas (Cohen; 1996) and Jodoin (1999) did not substantially improve the agreement between the two methods (see Appendix B). In fact, in both cases the disagreement was exacerbated. Using guidelines suggested by Zumbo and Thomas (1996), LR identified very few items across all samples and subjects. In contrast, using guidelines suggested by Jodoin (1999), LR identified more items, however this did not change the agreement between methods by more than one item. For example, in the June English and January Social examination one more common item was identified; however, in others (e.g., January Math, June Math, 2112 Math, English June), no additional common items were identified.

Unlike LR, LDF has not been extensively used or studied in relation to the other DIF detection procedures. In addition, there is no comparable effect size measure to $R^2\Delta$ to determine DIF magnitude. Furthermore, the use of the 95% confidence interval method to determine DIF magnitude was not developed in relation to the effect size measures associated with other DIF detection methods nor with the area method associated with IRT methods. In addition, as confidence intervals are affected by sample size, fewer detected items would be expected in smaller samples. Hence differences between these two effect sizes may be anticipated.

In summary, the agreement between LR and LDF is poor relative to the agreement between the other methods (MH/GMH, SIB/Poly-SIB). The use of LDF may leave potential DIF items undetected, especially if smaller samples are used. It does, however, provide additional information not given by any of the other methods. Therefore, LDF is not directly comparable to LR and future developmental research is recommended. It is, however, a useful method to investigate DIF among the scale scores associated with polytomous items and is the only method that will detect nonuniform DIF items. Consequently, this method was retained for comparison to GMH and Poly-SIB when the item set examined was expanded to include polytomous items.

Summary

Based on the results of this study, it can be concluded that GMH and Poly-SIB are quite comparable to their respective dichotomous counterparts, MH and SIB, and can be substituted for use with dichotomous items. Confidence that LDF will detect all of the items detected by LR is lower, especially in smaller samples. Despite this limitation, the examination of DIF within the item score points of a polytomous item can only be accomplished with LDF as the other methods determine DIF based on mean item score differences. Consequently, while LDF may not be useful in the identification of DIF it may be useful if applied in conjunction with another method for the analysis of item scale scores, particularly if the sample size is large.

Comparison of GMH, Poly-SIB, and LDF

Dichotomous Items

A summary of the number of dichotomous items detected by the GMH, Poly-SIB, and LDF detection procedures is provided in Table 6 for each of the four examinations and six samples. As before, a lower triangle is provided for each examination and sample. However, in this case the diagonal elements represent the number of items identified by each method. The numbers in the off diagonals represent the number of items commonly identified by the other methods. For example, in January English, GMH identified 11 items, Poly-SIB identified 17 items, and LDF identified 17 uniform and 0 nonuniform items. Of the 11 items identified by GMH, Poly-SIB identified 9 items and LDF identified 9. Of the 17 items identified by Poly-SIB, 15 items were detected by LDF. As shown in Table 6, and as foreshadowed by the discussion above, the data can be divided into two sets defined by sample size. In the case of the two large samples, January and June, Poly-SIB and LDF tended to identify more DIF items than GMH. Given these differences there was less than perfect agreement between the three procedures. This is illustrated using the June sample. For English, GMH identified 5 DIF items, Poly-SIB 9 DIF items, and LDF 10 DIF items. Of the 5 GMH DIF items, Poly-SIB identified 4, as did LDF. But, of the 9 Poly-SIB items, only 4 were identified by GMH while 7 were identified by LDF. And of the 10 LDF DIF items, only 4 were identified by GMH while 7 were identified by Poly-SIB.

Table 6.

Agreement between Polytomous DIF Detection Methods for Dichotomous Items

		English ($i = 70$)			Social ($i = 70$)			Math ($i = 49$)			Biology ($i = 56$)		
		GMH	PS	LDF	GMH	PS	LDF	GMH	PS	LDF	GMH	PS	LDF
Jan	GMH	11			5			6			4		
	PS	9	17		5	9		3	6		2	7	
	LDF	9	15	17	5	7	7	3	5	6	2	4	4
June	GMH	5			9			4			3		
	PS	4	9		7	17		4	7		1	3	
	LDF	4	7	10	8	15	17	3	4	4	1	3	5
1212	GMH	10			11			8			4		
	PS	9	13		10	14		7	12		4	7	
	LDF	3	3	3	2	2	2	2	3	3	1	1	2
2112	GMH	8			7			4			6		
	PS	7	9		6	12		3	6		5	5	
	LDF	2	2	2	2	2	3	1	1	1	2	2	3
2211	GMH	5			12			6			10		
	PS	3	4		10	14		5	6		9	13	
	LDF	1	1	1	3	3	3	0	0	0	2	2	2
2222	GMH	9			10			4			2		
	PS	6	7		8	12		3	4		2	6	
	LDF	3	3	3	3	3	4	1	1	1	0	0	0

Note. i = items in the analysis.

In Social Studies, GMH identified 9 DIF items, Poly-SIB 17 DIF items, and LDF 17 DIF items. Of the 9 GMH DIF items, Poly-SIB identified 7 and LDF identified 8. But, of the 17 Poly-SIB items, only 7 were identified by GMH while 15 were identified by LDF. And of the 17 LDF DIF items, only 8 were identified by GMH while 15 were identified by Poly-SIB.

Turning to Mathematics, GMH identified 4 DIF items, Poly-SIB 7 DIF items, and LDF 4 DIF items. Of the 4 GMH DIF items, Poly-SIB identified 4, and LDF identified 3. Of the 7 Poly-SIB items, GMH and LDF identified only 4 items. And of the 4 LDF DIF items, 3 were identified by GMH and all 4 were identified by Poly-SIB.

Lastly, GMH identified 3 DIF items, Poly-SIB 3 DIF items, and LDF 5 DIF items in Biology. Of the 3 GMH DIF items, Poly-SIB identified 1, as did LDF. Of the 3 Poly-SIB items, only 1 was identified by GMH while all 3 were identified by LDF. And of the 5 LDF DIF items, 1 was identified by GMH while 3 were identified by Poly-SIB.

As might be expected, LDF identified the lowest number of DIF items in the four smaller samples while Poly-SIB identified the most. Given these differences there was also less than perfect agreement between the DIF items identified by the procedures in the smaller samples. This is illustrated, for example, in the 2222 sample. For English, GMH identified 9 DIF items, Poly-SIB 7 DIF items, and LDF 3 DIF items. Of the 9 GMH DIF items, Poly-SIB identified 6, and LDF identified 3 items. Of the 7 Poly-SIB items, 6 were identified by GMH while only 3 were identified by LDF. And of the 3 LDF DIF items, both GMH and Poly-SIB identified all 3 items.

In Social Studies, GMH identified 10 DIF items, Poly-SIB 12 DIF items, and LDF 4 DIF items. Of the 10 GMH DIF items, Poly-SIB identified 8, and LDF identified 3. Of the 12 Poly-SIB items, only 8 were identified by GMH while only 3 were identified by LDF. And of the 4 LDF DIF items, both GMH and Poly-SIB identified only 3 if the items.

In Mathematics, GMH identified 4 DIF items, Poly-SIB 4 DIF items, and LDF 1 DIF item. Of the 4 GMH DIF items, Poly-SIB identified 3, and LDF identified 1. Of the 4 Poly-SIB items, GMH identified 4 items and LDF identified 1 item. And of the 1 LDF DIF item, both GMH and Poly-SIB also identified this item.

Lastly, in Biology, GMH identified 2 DIF items and Poly-SIB identified 6 DIF items while LDF did not identify any items. Of the 2 GMH DIF items, Poly-SIB identified both items. Of the 6 Poly-SIB items, only 2 were identified by GMH.

The findings reported in Table 6 reveal an interaction between sample size and generalized DIF detection methods in the detection of DIF in dichotomous items. For small sample sizes, LDF is most conservative, in contrast to GMH, which is most conservative for large sample sizes. Poly-SIB results were not related to sample size. Before drawing further conclusions, the behavior of these three procedures with the inclusion of the polytomous items was examined. The results of these analyses are presented in the next subsection.

Inclusion of Polytomous Items

To investigate the behavior of GMH, Poly-SIB, and LDF in mixed format examinations, the analyses reported above were repeated using both the dichotomous and polytomous items for each examination. As reported in Table 2, the number of polytomous items on each examination was small, therefore, analyses with these items alone were not completed as it was felt that the results would not be "stable", making it difficult to interpret and form recommendations. Furthermore, the purposes of this research were to investigate the behavior of GMH, Poly-SIB, and LDF in mixed format examinations and to explore the prevalence of gender DIF across item format and subject area. Consequently, dichotomous items as well as the polytomous items were considered together. In this case the total test score was used to match examinees on ability.

A summary of the number of DIF items detected by the GMH, Poly-SIB, and LDF detection procedures across all items is provided in Table 7 for each of the four examinations and six samples. More detailed, item specific information is presented in Appendix C. As before, a lower triangle is provided for each examination and sample. Two numbers, separated by a comma, are presented in each cell. The first number corresponds to the number of dichotomous items identified; the second number corresponds to the number of polytomous items identified. The diagonal elements represent the number of dichotomous and polytomous items identified by each method. The numbers in the off diagonals represent the number of dichotomous and polytomous items commonly identified by the other methods. For example, in January English, GMH identified 13 dichotomous and no polytomous DIF items, Poly-SIB identified 17 dichotomous and 2 polytomous DIF items, and LDF identified 18 dichotomous and 2 polytomous DIF items. Of the 13 dichotomous DIF items identified by GMH, Poly-SIB identified 10 items and LDF identified 11. Of the 17 dichotomous DIF items identified

Table 7.

Agreement between Polytomous DIF Detection Methods for Combined Dichotomous and Polytomous Items

		English ($i = 70, 4$)			Social ($i = 70, 6$)			Math ($i = 49, 3$)			Biology ($i = 56, 1$)		
		GMH	PS	LDF	GMH	PS	LDF	GMH	PS	LDF	GMH	PS	LDF
Jan	GMH	13, 0			4, 0			5, 1			6, 0		
	PS	10, 0	17, 2		4, 0	11, 3		3, 1	7, 1		4, 0	7, 1	
	LDF	11, 0	13, 2	18, 2	4, 0	9, 3	11, 3	3, 1	5, 1	5, 2	6, 0	6, 0	10, 1
June	GMH	4, 0			12, 4			3, 0			3, 1		
	PS	3, 0	9, 1		10, 4	15, 4		3, 0	8, 2		1, 1	6, 1	
	LDF	4, 0	9, 0	10, 1	10, 4	14, 4	16, 4	1, 0	5, 1	6, 1	3, 1	6, 1	10, 1
1212	GMH	12, 0			11, 3			5, 1			4, 0		
	PS	11, 0	11, 2		11, 3	17, 4		5, 0	12, 0		4, 0	10, 1	
	LDF	5, 0	3, 2	3, 2	4, 3	4, 4	4, 4	2, 0	3, 0	3, 0	1, 0	2, 0	2, 0
2112	GMH	8, 0			11, 1			4, 0			5, 1		
	PS	4, 0	9, 3		7, 0	14, 0		2, 0	5, 2		5, 1	8, 1	
	LDF	2, 0	2, 0	2, 0	2, 0	1, 0	4, 0	0, 0	0, 1	0, 1	1, 1	2, 1	2, 1
2211	GMH	5, 0			14, 3			5, 0			6, 0		
	PS	3, 0	10, 0		12, 3	20, 4		2, 0	7, 0		6, 0	10, 1	
	LDF	1, 0	1, 0	1, 0	1, 3	1, 4	1, 4	0, 0	0, 0	1, 0	2, 0	2, 0	2, 0
2222	GMH	8, 0			14, 4			5, 0			4, 0		
	PS	4, 0	5, 0		11, 4	18, 4		3, 0	3, 1		3, 0	9, 1	
	LDF	2, 0	2, 0	2, 0	2, 4	2, 4	2, 4	2, 0	2, 0	2, 0	0, 0	0, 0	0, 0

Notes. The number of dichotomous and polytomous items with DIF are presented in pairs. The first number is the number of dichotomous

DIF items, the second number is the number of polytomous DIF items.

by Poly-SIB, 13 items were detected by LDF. The 2 polytomous DIF items identified by Poly-SIB were also detected by LDF but not by GMH.

Comparison of the dichotomous results (first number in each pair in Table 7) with the dichotomous results reported in Table 6 reveals that the inclusion of the polytomous items and the resulting change in the matching variable altered the number of dichotomous items detected. In some cases more dichotomous items were identified, in other cases fewer or the same number of dichotomous items were identified. The differences were not related to sample size, subject, or DIF detection method. They are, however, related to the inclusion of the polytomous items, which altered the total test score (matching variable) and the ability distribution of the reference and focal groups. Consequently, the use of a different matching variable in the analysis of the combined set of items produced different results for the dichotomous items than when the dichotomous items were analyzed alone.

As described earlier, in Chapter 3, limitations associated with the respective computer programs for the Poly-SIB and LDF DIF detection methods prevented the analysis of the dichotomous items using a matching variable other than the sum of the items analyzed. Only the GMH detection method afforded the opportunity to use a different matching variable. In this case it was possible to analyze only the dichotomous items using the combined dichotomous and polytomous item test score as the matching variable. In these analyses there were no differences in the results obtained when only the dichotomous items were analyzed using the combined total test score as the matching variable from the results obtained when the dichotomous and polytomous items were analyzed together (see Appendix D).

While the items detected varied with the inclusion of the polytomous items, the general pattern of the number of detected items did not change. As before, for small sample sizes, LDF is most conservative, in contrast to GMH, which is most conservative for large sample sizes. Poly-SIB results were not related to sample size.

As shown in Table 7, like the dichotomous items, the results for the combined set of items can be divided into two sets defined by sample size. In the case of the two large samples, January and June, Poly-SIB and LDF tended to identify more DIF items than GMH. Given these differences, there was less than perfect agreement between the DIF items identified by the three procedures. This is illustrated using the June sample. For English, GMH identified 4 dichotomous DIF items, Poly-SIB identified 9 dichotomous and 1 polytomous DIF items, and LDF identified 10 dichotomous and 1 polytomous DIF items. Of the 4 GMH dichotomous DIF items, Poly-SIB identified 3, and LDF identified 4. But, of the 9 Poly-SIB dichotomous DIF items, only 3 were identified by GMH while 9 were identified by LDF. And of the 10 LDF dichotomous DIF items, 4 were identified by GMH and 9 by Poly-SIB. Poly-SIB and LDF, but not GMH, which did not detect any polytomous DIF items, identified the same polytomous item.

In Social Studies, GMH identified 12 dichotomous and 4 polytomous DIF items, Poly-SIB identified 15 dichotomous and 4 polytomous DIF items, and LDF identified 16 dichotomous and 4 polytomous DIF items. Of the 12 GMH dichotomous DIF items, Poly-SIB identified 10, as did LDF. Similarly, of the 15 Poly-SIB dichotomous DIF items, 10 were identified by GMH while 14 were identified by LDF. And of the 16 LDF dichotomous DIF items, 10 were identified by GMH, while 14 were identified by Poly-SIB. The same 4 polytomous DIF items were identified by all three methods.

Turning to Mathematics, GMH identified 3 dichotomous and no polytomous DIF items, Poly-SIB identified 8 dichotomous and 2 polytomous DIF items, and LDF identified 6 dichotomous and 1 polytomous DIF item. Of the 3 GMH dichotomous DIF items, Poly-SIB identified all 3 while LDF identified 1. Of the 8 dichotomous Poly-SIB DIF items, GMH identified 3 and LDF identified 5. And of the 6 LDF dichotomous DIF items, 5 were identified by Poly-SIB and 1 was identified by GMH. Of the 2 polytomous DIF items identified by Poly-SIB, 1 was identified by LDF. GMH did not detect any polytomous DIF items.

Lastly, GMH identified 3 dichotomous and 1 polytomous DIF items, Poly-SIB identified 6 dichotomous and 1 polytomous DIF items, and LDF identified 10 dichotomous and 1 polytomous DIF items in Biology. Of the 3 GMH dichotomous DIF items, Poly-SIB identified 1 and LDF identified 3. Of the 6 Poly-SIB dichotomous DIF items, only 1 was identified by GMH while all 6 were identified by LDF. And of the 10 LDF dichotomous DIF items, 3 were identified by GMH and 6 were identified by Poly-SIB. All three DIF detection procedures identified the polytomous item.

As might be expected, based on the results presented earlier, LDF identified the lowest number of DIF items in the four smaller samples while Poly-SIB identified the most. Given these differences there was also less than perfect agreement between the DIF items identified by the procedures in the smaller samples. This is illustrated, for example, in the 2211 sample. In English, GMH identified 5 dichotomous DIF items, Poly-SIB identified 10 dichotomous DIF items, and LDF 1 dichotomous DIF item. No polytomous items were identified by any of the three DIF detection methods. Of the 5 GMH DIF items, Poly-SIB identified 3, and LDF identified 1. Of the 10 Poly-SIB items, 3 were identified by GMH while only 1 was identified by LDF. Both GMH and Poly-SIB also identified the single item identified by LDF.

In Social Studies, GMH identified 14 dichotomous and 3 polytomous DIF items, Poly-SIB identified 20 dichotomous and 4 polytomous DIF items, and LDF identified 1 dichotomous and 4 polytomous DIF items. Of the 14 GMH dichotomous DIF items, Poly-SIB identified 12, and LDF identified 1. Of the 20 Poly-SIB dichotomous DIF items, 12 were identified by GMH while only 1 was identified by LDF. Both Poly-SIB and GMH identified the single dichotomous item identified by LDF. Of the 3 GMH polytomous DIF items, Poly-SIB and LDF identified all 3 items. Of the 4 Poly-SIB polytomous DIF items, 3 were identified by GMH, while all 4 were identified by LDF.

In Mathematics, GMH identified 5 dichotomous DIF items, Poly-SIB identified 7 dichotomous DIF items, and LDF identified 1 dichotomous DIF item. No polytomous items were identified. Of the 5 GMH dichotomous DIF items, Poly-SIB identified 2, and LDF identified none. Of the 7 Poly-SIB dichotomous DIF items, GMH identified 2 items and LDF identified none. And of the 1 LDF dichotomous DIF item, neither GMH nor Poly-SIB identified this item.

Lastly, in Biology, GMH identified 6 dichotomous DIF items, Poly-SIB identified 10 dichotomous DIF items, and LDF identified 2 dichotomous DIF items. No polytomous items were identified. Of the 6 GMH dichotomous DIF items, Poly-SIB identified all 6, and LDF identified 2. Of the 10 Poly-SIB dichotomous DIF items, 6 were identified by GMH and 2 were identified by LDF. And of the 2 LDF dichotomous DIF items, GMH and Poly-SIB identified both of them.

The findings reported in Table 7 reveal an interaction between sample size and the generalized DIF detection methods in the detection of DIF in examinations consisting of both polytomous and dichotomous items. For small sample sizes, LDF is most conservative, in contrast to GMH, which is most conservative for large sample sizes. Although differences in the numbers of dichotomous and polytomous DIF items identified were noted across samples, the Poly-SIB results are the most consistent of the three methods across comparable samples.

Summary

Of the three generalized methods studied, Poly-SIB is recommended for DIF detection based on the results of this study for two reasons. First, although the proportion of DIF items identified varied by method, Poly-SIB generally detected the greatest number of items. Second, Poly-SIB results were least affected by variations in sample size, providing comparable results in both large and small samples for both dichotomous and polytomous items. Further, Poly-SIB identified the majority of DIF items also identified by GMH and LDF. Although some items may be falsely identified as exhibiting DIF (Type I error) by Poly-SIB, it is also likely that few DIF items would remain undetected with Poly-SIB (Type II error). Conversely, the possibility of not identifying items with DIF may be higher with either GMH or LDF. While a more conservative method may be desirable by test developers and administrators, this may not be most desirable by examinees and social advocates, especially if DIF items remain undetected.

Given the discrepancies between DIF detection methods, it has been suggested (e.g., Gierl, Rogers, & Klinger, 1999) that different methods should be used and only those items detected by two or more methods should be studied further. However, as illustrated in this study, many items initially detected by Poly-SIB, for example, would be eliminated from further study if this procedure was paired with either GMH or LDF. For example, in the January English sample, 7 of the 17 items initially detected by Poly-SIB would be eliminated if the item had to be detected by GMH as well as Poly-SIB. While fewer items would be eliminated if Poly-SIB was paired with LDF in the larger samples, more items would be eliminated with this pairing in the smaller samples. Furthermore, perfect agreement between the three methods is not expected as convergence of different methods is influenced by the unreliability of the associated statistics (Hambleton & Rogers, 1989). Therefore, the lack of convergence for an item may be falsely interpreted as negligible or no DIF (Type II error). Again, from the perspective of examinees and social advocates for equity and fair test administration and use, the increased probability of committing a Type II error is considered more problematic than the increased probability of committing a Type I error.

Consequently, the secondary research question regarding the prevalence of gender DIF across item format and subject area among a common sample was addressed using Poly-SIB to ensure that all potential DIF items were identified. When a polytomous item was identified as having DIF, LDF was also used in an attempt to determine the prevalence of DIF within the scale scores. Unfortunately, as discussed earlier, fewer polytomous items were identified by LDF in the smaller, common samples. As a result, LDF results are included only for the polytomous items identified by both Poly-SIB and LDF.

Comparison of Gender DIF Detected Across Format and Subject

As indicated in Chapter 1, an ancillary hypothesis regarding differences between males and females across item format and subject area was also addressed in the present study. Of importance is the fact that this hypothesis was examined using a common sample. That is, unlike previous studies where different samples were used across different examinations, the sample of students used in this study completed the same examinations in each of the four subject areas. The sample used for this analysis, 2222, contained those students that wrote all four examinations at the same time in June. This group was hypothesized to be a homogeneous sample with similar academic and extracurricular interests, completed coursework, and aspirations. The analyses were replicated, once for 1997 and once for 1998.

Mean Differences between Males and Females

Although differences between the mean scores for males and females are not adequate to identify DIF, they are commonly reported and referenced as evidence of differential performance. Further, means, together with standard deviations, provide a description of the overall performance of the groups to be studied. Hence, the mean and standard deviation for the total score, the dichotomous items, and the polytomous items for the males and females in sample 2222 are presented in Table 8 together with the effect sizes and t-test results. Significant differences between the two groups were noted on various aspects of the four examinations. To interpret the differences between the means of the males and females, effect sizes were computed using the standard deviation of the males as an estimate of the variance. These effect sizes were interpreted using Cohen's (1988) operational definitions for small (d = .2), medium (d = .5), and large (d = .8).

In the 1997 English examination, females had significantly higher mean scores ($p < .05$) only on the polytomous section of the examination; differences between the males and females on the dichotomous section of the examination was not statistically significant. In 1998, females had significantly higher mean scores ($p < .05$) on all aspects of the examination, including total test score. However, in both academic years, the effect sizes associated with the differences were all small to medium ($d \leq -.30$).
In contrast to English, males outperformed females in several aspects of the Social Studies examination in both 1997 and 1998. In both 1997 and 1998 males had a significantly ($p < .05$) higher mean score than females on the dichotomous section of the examination. The associated effect sizes for these differences were .45 for 1997 and .24 for 1998. In addition, males also had a significantly ($p < .05$) higher mean total test score than females in 1997 ($d = .35$). In contrast, females had a significantly ($p < .05$) higher mean score on the polytomous section of the examination in 1998 ($d = -.28$).

Unlike the previous subject areas, few significant differences were noted between the males and females in both the Mathematics and Biology examinations. Statistically

Table 8.

Descriptive Statistics for each Examination by Gender

			Total Score				Dichotomous Items Only				Polytomous Items Only			
		N	M	SD	t	d	M	SD	t	d	M	SD	t	d
English														
97	M	224	67.29	12.41	-1.47	-0.13	47.45	9.38	-0.41	-0.04	19.84	4.57	-3.31*	-0.28
	F	285	68.90	12.11			47.79	9.35			21.11	4.07		
98	M	183	66.52	13.42	-3.09*	-0.30	46.33	10.07	-2.83*	-0.28	20.19	4.46	-2.81*	-0.27
	F	243	70.52	13.11			49.14	10.21			21.38	4.22		
Social Studies														
97	M	224	62.17	12.59	3.88*	0.35	50.08	10.37	4.96*	0.45	12.10	3.50	-0.89	-0.07
	F	285	57.75	12.88			45.38	10.81			12.36	3.09		
98	M	183	61.32	13.04	1.45	0.14	49.34	11.27	2.51*	0.24	11.98	3.12	-2.89*	-0.28
	F	243	59.47	13.05			46.61	11.02			12.85	3.05		
Mathematics														
97	M	224	38.89	12.11	1.51	0.13	31.30	8.68	1.77	0.15	7.59	4.03	0.72	0.06
	F	285	37.30	11.53			29.96	8.33			7.34	3.82		
98	M	183	41.64	11.64	1.91	0.18	32.93	8.74	2.51*	0.23	8.70	3.55	0.25	0.02
	F	243	39.57	10.64			30.93	7.65			8.62	3.07		
Biology														
97	M	224	53.65	10.75	0.71	0.06	35.02	6.97	0.70	0.06	18.63	4.66	0.62	0.05
	F	285	53.00	9.73			34.62	5.96			18.38	4.46		
98	M	183	49.71	11.69	-1.83	-0.14	34.97	7.33	-1.04	-0.08	14.74	5.15	-2.68*	-0.21
	F	243	51.39	11.07			35.58	7.12			15.80	4.68		

Note. * $p < .05$

significant ($p < .05$) differences between the males and females were noted for only the mean score of the dichotomous section of the 1997 Mathematics examination and the mean score of the polytomous section of the 1998 Biology examination. In the 1997 Mathematics examination, males had a significantly ($p < .05$) higher mean score than females on the dichotomous section of the examination ($d = .23$). Whereas, in the 1998 Biology examination, females had a significantly ($p < .05$) higher mean score than males on the polytomous section of the examination ($d = -.21$).

While differences between mean scores indicate differential performance over the associated items, such differences do not necessarily imply the presence of differential performance at the item level. To make that determination it is necessary to conduct DIF analyses. The results of the prevalence of DIF across item format and subject area is presented in the following pages. In the following subsection, the prevalence of DIF across item format is discussed separately for each of the four subject areas. These results are then summarized across the four subject areas to address the question of interactions among subject area, gender, and item format.

The results upon which these discussions are based are summarized in Table 9. More detailed, item specific results are presented in Appendix E. As discussed earlier, the DIF item detection procedure used was Poly-SIB supplemented by LDF for identified polytomous items to determine DIF within the individual scale score points. The results for each subject area are first discussed separately. This is then followed by a discussion of the prevalence of DIF across subject area and item format.

Table 9.

DIF Items Identified by Method across Subject Area and Item Format

		1997			1998		
		N	D	P	N	D	P
English			$i = 70$	$i = 6$		$i = 70$	$i = 6$
	M	224	13	0	183	2	0
	F	285	3	1	243	3	0
Social			$i = 70$	$i = 4$		$i = 70$	$i = 4$
	M	224	6	0	183	14	0
	F	285	2	4	243	4	4
Math			$i = 49$	$i = 3$		$i = 49$	$i = 3$
	M	224	4	0	183	3	0
	F	285	4	1	243	0	1
Biology			$i = 56$	$i = 1$		$i = 56$	$i = 1$
	M	224	6	0	183	7	0
	F	285	6	0	243	2	1

Note. D = dichotomous items; P = polytomous items; M = males; F = females; i = items in the analysis.

For each examination, the dichotomous items identified with DIF are also described and classified according to the examination blueprint provided by Alberta Education. For English and Social Studies, course content and cognitive domain are used to classify these items. For Mathematics, the items are classified by course content and level of mathematical understanding. For Biology, items are classified by course content alone. Although these descriptions are useful to describe the types of items with DIF, as with previous studies, no attempt has been made to determine if the DIF is attributable to bias, impact, or Type I error. In previous studies attempts have been made to clarify the nature of DIF using panels of content experts. However, these panels are generally unsuccessful at both interpreting or predicting items that perform differently across different groups of examinees (Camilli & Shepard, 1994, Gierl & McEwen, 1988). Protocol analysis (Ericsson & Simon, 1993) holds some promise for helping researchers come to a better understanding of how the items would be interpreted by the actual examinees; however, this method is both resource and time intensive and was beyond the scope of this study.

English

The English examination consisted of 70 dichotomous and 6 polytomous items. More items were identified with DIF in 1997 than in 1998. In 1997, a total of 17 items (22.4%) were identified with DIF, whereas only 5 items (6.6%) were identified with DIF in 1998 (see Table 9). Of the 17 items identified in 1997, 13 dichotomous items favored males, 3 dichotomous items favored females and one polytomous item favored females. In 1998, 2 dichotomous items favored males, while 3 dichotomous items favored females; no polytomous items were identified in this academic year.

According to the examination blueprint, the 70 dichotomous items were "classified in two ways: according to the curricular content area being tested and according to the thinking (process) skill required to answer the question" (Alberta Education, 1997b, 1998b, p. 4). As shown in Table 10, in 1997, for course content, 30 of the 70 dichotomous items were classified "Meanings", 25 were classified "Critical Response", and 15 were classified "Human Experience and Values". For thinking skills, 43 were classified as "Inference and Application"; 22 were classified "Evaluation", and 5 were classified "Literal Understanding". In 1998, for course content, 32 were classified "Meanings", 23 were classified "Critical Response", and 15 were classified "Human Experience and Values". For thinking skills, 45 were classified as "Inference and Application"; 19 were classified "Evaluation", and 6 were classified "Literal

Table 10.

English DIF Items by Course Content and Cognitive Level: Dichotomous Items

Curricular Content	Year	Literal Understanding (i_{97} = 5; i_{98} = 6)		Inference and Application (i_{97} = 43; i_{98} = 45)		Evaluation (i_{97} = 22; i_{98} = 19)		TOTALS	
		Male	Female	Male	Female	Male	Female	**Male**	**Female**
Meanings (i_{97} = 30; i_{98} = 32)	1997	2 (2)	0 (2)	5 (22)	1 (22)	1 (6)	1 (6)	**8**	**2**
	1998	1 (6)	0 (6)	1 (19)	3 (19)	0 (7)	0 (7)	**2**	**3**
Critical Response (i_{97} = 25; i_{98} = 23)	1997	0 (3)	0 (3)	2 (14)	1 (14)	0 (8)	0 (8)	**2**	**1**
	1998	0 (0)	0 (0)	0 (17)	0 (17)	0 (6)	0 (6)	**0**	**0**
Human Experience & Values (i_{97} = 15; i_{98} = 15)	1997	0 (0)	0 (0)	2 (7)	0 (7)	1 (8)	0 (8)	**3**	**0**
	1998	0 (0)	0 (0)	0 (9)	0 (9)	0 (6)	0 (6)	**0**	**0**
TOTALS	**1997**	**2**	**0**	**9**	**2**	**2**	**1**	**13**	**3**
	1998	**1**	**0**	**1**	**3**	**0**	**0**	**2**	**3**

Notes. The numbers in parentheses are the total numbers of items within the cell classified by thinking (process) skills and curricular content. The totals are repeated in both the male and female columns. i_{97} = number of items in 1997; i_{98} = number of items in 1998.

Understanding".

For both 1997 and 1998, the polytomous items of the English Examination were related to two writing assignments designed to assess reading, writing, and thinking skills. Five point scales are used to score the student responses. Two items were associated with a short assignment and are labeled: "thought and detail" and "writing skills". The four remaining items were associated with a longer assignment requiring "the synthesis and ability to communicate regarding techniques used in the literary works studied in class" (Alberta Education, 1997b, 1998b, p. 2). These were labeled: "thought and detail", "organization", "matters of choice", and "matters of correctness". The number of dichotomous and polytomous items identified with DIF are discussed below and described according to the preceding classifications. Table 10 contains the classification of dichotomous DIF items detected in 1997 and 1998 by curricular content and thinking (process) level.

1997. In 1997, 13 dichotomous items favored males and 3 dichotomous items favored females. Of the 13 items favoring males, 8, 2, and 3 were classified, respectively, in the curricular content areas "Meanings", "Critical Response", and "Human Experience and Values". When classified according to thinking (process) level 2, 9, and 2 were classified, respectively, "Literal Understanding", "Inference and Application", and "Evaluation". Both of the items classified "Literal Understanding", were from the content area "Meanings". Of the 9 items classified "Inference and Application", 5 were from the content area "Meanings", 2 were from "Critical Response", and 2 were from "Human Experience and Values". The 2 items classified "Evaluation", were split between "Meaning" and "Human Experience and Values".

Of the 3 dichotomous items favoring females, 2 items were from the content area "Meanings" while the third item was from "Critical Response". One item from each content area was classified in the thinking (process) domain "Inference and Application". The second item from the "Meanings" content area was classified "Evaluation".

The one polytomous item identified, the short assignment designed to assess writing skills, favored females. According to the LDF results, at each scale score point, a greater proportion of females than males received the associated mark.

1998. In 1998, all five of the dichotomous DIF items were found in the content area of "Meanings", with two items favoring males and three items favoring females. 1 of the 2 items favoring males was from the thinking (process) level " Inference and Application" while the second was from "Literal Understanding" thinking (process) level. All 3 of the items favoring females were from the thinking (process) level "Inference and Application". No polytomous items with DIF were detected.

In summary, considerably more DIF items were detected with DIF in 1997 (17) than in 1998 (5). While more dichotomous items favored males than females in 1997, only 2 of the 5 items identified with DIF in 1998 favored males. In both academic years, the majority of these items were from the content area "Meanings" and the thinking (process) level "Inference and Application". Only one polytomous item was detected across the two years. "Writing Skills", a scale associated with the short assignment was detected in 1997 and favored females at all score points. This item was not detected in 1998.

Social Studies

The Social Studies examination consisted of 70 dichotomous and 4 polytomous items. Overall, fewer items were identified with DIF in 1997 than in 1998. In 1997, a total of 12 items (16.2%) were identified with DIF; 22 items (29.7%) were identified in 1998 (see Table 9). Of the 12 items identified in 1997, 6 dichotomous items favored males, 2 dichotomous items favored females, and all 4 polytomous items favored females. In 1998, 14 dichotomous items favored males, 4 dichotomous items favored females, and all 4 polytomous items favored females.

According to the examination blueprint, each dichotomous item is "classified in two ways: according to the curricular content area (topic) being tested and by the knowledge and skill objectives required to answer the question" (Alberta Education, 1997c, 1998c, p. 4). As shown in Table 11, in both 1997 and 1998, the 70 dichotomous items were equally distributed between the two content areas "Political and Economic Systems" and "Global Interaction in the 20th Century". For the knowledge and skill objective, 24 items were classified "Comprehension of Information and Ideas"; 22 were classified "Interpretation and Analysis of Information and Ideas", and 24 were classified "Synthesis and Evaluation of Information and Ideas" in 1997. In 1998, 24 items were classified "Comprehension of Information and Ideas"; 23 were classified "Interpretation and Analysis of Information and Ideas", and 23 were classified "Synthesis and Evaluation of Information and Ideas".

The four polytomous items of the Social Studies Examination were related to one writing assignment in which the student was required to "discuss the importance and complexity of an issue and rationally defend their position by using supportive and relevant evidence" (Alberta Education, 1997c, 1998c, p. 6). Each item corresponds to a five-point scoring scale. While one of the scales assessed writing skills (quality of language expression), the other scales (exploration of the issue, defense of position, and

Table 11.

Social Studies DIF Items by Course Content and Cognitive Level: Dichotomous Items

Curricular Content		Knowledge and Skill Objectives: Comprehension (i_{97} = 24; i_{98} = 24)		Interpretation and Analysis (i_{97} = 23; i_{98} = 22)		Synthesis and Evaluation (i_{97} = 24; i_{98} = 23)		**TOTALS**	
		Male	Female	Male	Female	Male	Female	**Male**	**Female**
Political & Economic Systems (i_{97}, i_{98} = 35)	1997	1 (12)	0 (12)	1 (12)	1 (12)	2 (11)	0 (11)	**4**	**1**
	1998	2 (12)	1 (12)	0 (11)	1 (11)	3 (12)	1 (12)	**5**	**3**
Global Interaction (i_{97}, i_{98} = 35	1997	1 (12)	1 (12)	1 (11)	0 (11)	0 (12)	0 (12)	**2**	**1**
	1998	3 (12)	0 (12)	4 (11)	1 (11)	2 (12)	0 (12)	**9**	**1**
TOTALS	**1997**	**2**	**1**	**2**	**1**	**2**	**0**	**6**	**2**
	1998	**5**	**1**	**4**	**2**	**5**	**1**	**14**	**4**

Notes. The numbers in parentheses are the total numbers of items within the cell classified by thinking (process) skills and curricular content. The totals are repeated in both the male and female columns. i_{97} = number of items in 1997; i_{98} = number of items in 1998.

quality of examples) assessed the ability of the student to demonstrate an understanding of course content and critical thinking skills (Alberta Education, 1997c, 1998c).

1997. Six dichotomous items favored males, while only two items favored females. Of the 6 items favoring males, 4 were found in the content area "Political and Economic Systems", and 2 were found in "Global Interaction in the 20th Century". At knowledge and skill levels, 2 items were found in each level. For both "Comprehension", and "Interpretation and Analysis", the 2 items were split between the two content areas. The two "Evaluation" items were both from "Political and Economic Systems".

The 2 dichotomous items favoring females were evenly split between the two course content areas. The "Political and Economic Systems" item was from the "Analysis and Interpretation" knowledge and skill level. The "Global Interaction in the 20th Century" item was from the "Comprehension" knowledge and skill level.

All 4 polytomous items or scales favored females. The results of the LDF analysis of each scale score are presented in Table 12. As indicated in this table, a greater proportion of males than females received the lower scores of 1 and 2 on all scales. In contrast, a greater proportion of females than males received the higher scores 4 and 5 on all scales. For the scale "Defense of Position", a greater proportion of males than females received a score of three.

Table 12.

DIF within Item Scale Scores Detected by LDF Social Studies 1997

Scoring Category	Scale Scores					
	0	1	2	3	4	5
Exploration of the Issue	-	M	M	-	F	F
Defense of Position	-	M	M	M	F	F
Quality of Examples	-	M	M	-	F	F
Quality of Language and Expression	-	M	M	-	F	F

Note. - = favors neither group; F = favors female; M = favors male.

1998. Fourteen dichotomous items favored males, while four items favored females. Of the 14 items favoring males, 5 were from the content area "Political and

Economic Systems" and 9 were from "Global Interaction in the 20th Century". At the knowledge and skill levels, 5 items were from "Comprehension", 4 items were from "Interpretation and Analysis", and 5 items were from "Synthesis and Evaluation". Of the 5 items found in the "Comprehension" knowledge and skill level, 2 items were from the content area "Political and Economic Systems" and 3 items were from "Global Interaction in the 20th Century". Of the 5 items found in the "Evaluation" knowledge and skill level, 3 items were from the content area "Political and Economic Systems" and 2 items were from "Global Interaction in the 20th Century". All four of the items classified "Interpretation and Analysis" were from "Global Interaction in the 20th Century".

Of the 4 dichotomous items favoring females, 3 items were from the content area "Political and Economic Systems", and 1 was from "Global Interaction in the 20th Century". The 3 "Political and Economic Systems" items were evenly distributed across the three knowledge and skill levels. The "Global Interaction in the 20th Century" item was classified in the "Analysis and Interpretation" knowledge and skill level.

All 4 polytomous scales favored females. The results of the LDF analysis of each scale are presented in Table 13. As indicated, a greater proportion of males than females received the lower scores, 1 and 2, on all scales. In contrast, a greater proportion of females than males received the higher scores, 4 and 5. No differences were noted for the scale score of 0. For the scale, "Defense of Position", a greater proportion of males than females received a score of three. In contrast, a greater proportion of females than males received a score of three. On the "Quality of Examples" scale.

Table 13.

DIF within Item Scale Scores Detected by LDF Social Studies 1998

Scoring Category	Scale Score					
	0	1	2	3	4	5
Exploration of the Issue	-	M	M	-	F	F
Defense of Position	-	M	M	M	F	F
Quality of Examples	-	M	M	F	F	F
Quality of Language and Expression	-	M	M	-	F	F

Note. - = favors neither group; F = favors female; M = favors male.

In summary, considerably fewer DIF items were detected with DIF in 1997 (12) than 1998 (22). In both academic years, the majority of dichotomous items favored males while all of the polytomous items favored females. In 1997, more DIF items were detected in the content area "Political and Economic Systems", whereas in 1998, more DIF items were detected in the content area "Global Interaction in the 20th Century". In both academic years the numbers of dichotomous DIF items were evenly distributed across the three knowledge and skill objectives. All four polytomous items in both years favored females. In each case a greater proportion of females than males received the higher scale scores (4 or 5). While one of the scales assessed writing skills, the other scales assessed the ability of the students to demonstrate both an understanding of course content and their critical thinking skills.

Mathematics

The Mathematics examination consisted of 49 dichotomous and 3 polytomous items. Of the 49 dichotomous items, 40 were multiple-choice items and 9 were gridded numerical response items. In 1997, a total of 9 (17.3%) items were identified with DIF, whereas only 4 items (29.7%) were identified in 1998 (see Table 9). Of the 9 items identified in 1997, 4 dichotomous items favored males, while 4 dichotomous items and 1 polytomous item favored females. Of the dichotomous items favoring males, 2 were gridded numerical response items, whereas only one of the dichotomous items favoring females was a gridded numerical response item. In 1998, 3 dichotomous items favored males, while one polytomous item favored females. Of the dichotomous items favoring males, 1 was a gridded numerical response item.

According to the examination blueprint for mathematics and as shown in Table 14, each dichotomous item is classified by one of nine unit topics or content domains and by mathematical understanding. The numbers of items in each unit topic vary from 4 to 8 (see Table 14).

While the number of items from each unit was kept constant across both 1997 and 1998, the number of items varied across the levels of mathematical understanding. For mathematical understanding, 21 items in 1997 and 15 items in 1998 were classified "Procedural"; 15 items in 1997 and 19 items in 1998 were classified "Conceptual"; and 13 items in 1997 and 15 items in 1998 were classified "Problem-Solving". As with English and Social Studies, the discussion of the items for Mathematics is organized in terms of these unit and skill dimensions.

The three polytomous items included in each examination "assess whether or not students can draw on their mathematical experiences to solve problems and to explain mathematical concepts" (Alberta Education, 1997d, 1998d, p. 5). These items may cross

Table 14.

Mathematics DIF Items Organized by Unit and Understanding: Dichotomous Items

Unit	Year	Procedural (i_{97} =21; i_{98} = 15)		Conceptual (i_{97} =15; i_{98} = 19)		Problem Solving (i_{97} =13; i_{98} = 15)		TOTALS	
		Male	Female	Male	Female	Male	Female	Male	Female
Polynomial Functions (i_{97}, i_{98} = 8)	1997	0 (1)	0 (1)	0 (4)	0 (4)	0 (2)	0 (2)	**0**	**0**
	1998	-	-	0 (4)	0 (4)	0 (4)	0 (4)	**0**	**0**
Trigonometric & Circular Functions (i_{97}, i_{98} = 8)	1997	0 (3)	2 (3)	0 (2)	0 (2)	1 (3)	0 (3)	**1**	**2**
	1998	0 (2)	0 (2)	1 (3)	0 (3)	0 (3)	0 (3)	**1**	**0**
Statistics (i_{97}, i_{98} = 4)	1997	0 (2)	0 (2)	1 (1)	0 (1)	0 (1)	0 (1)	**1**	**0**
	1998	0 (2)	0 (2)	0 (1)	0 (1)	0 (1)	0 (1)	**0**	**0**
Quadratic Relations (i_{97}, i_{98} = 7)	1997	0 (2)	1 (2)	0 (3)	1 (3)	0 (1)	0 (1)	**0**	**2**
	1998	0 (1)	0 (1)	0 (5)	0 (5)	0 (1)	0 (1)	**0**	**0**
Exponential & Logarithmic Functions (i_{97}, i_{98} = 8)	1997	2 (5)	0 (5)	0 (3)	0 (3)	-	-	**2**	**0**
	1998	0 (4)	0 (4)	0 (2)	0 (2)	0 (2)	0 (2)	**0**	**0**
Permutations & Combinations (i_{97}, i_{98} = 7)	1997	0 (4)	0 (4)	0 (2)	0 (2)	0 (1)	0 (1)	**0**	**0**
	1998	0 (4)	0 (4)	0 (3)	0 (3)	-	-	**0**	**0**
Sequences & Series (i_{97}, i_{98} = 7)	1997	0 (3)	0 (3)	-	-	0 (4)	0 (4)	**0**	**0**
	1998	0 (2)	0 (2)	0 (1)	0 (1)	0 (4)	0 (4)	**2**	**0**
TOTALS	**1997**	**2**	**3**	**1**	**1**	**1**	**0**	**4**	**4**
	1998	**1**	**0**	**2**	**0**	**0**	**0**	**3**	**0**

Notes. The numbers in parentheses are the total numbers of items within the cell classified by thinking (process) skills and curricular content. - = no items were classified in the cell; i_{97} = number of items in 1997; i_{98} = number of items in 1998.

more than one unit or may require students to make connections among mathematical concepts. Specific five-point scoring rubrics are used to evaluate the quality and completeness of the student responses to each of these items.

1997. Of the 2 multiple choice items favoring males, 1 was from the unit "Statistics" and 1 was from "Exponential and Logarithmic Functions". Of the 2 numerical response items, 1 was from "Trigonometric and Circular Functions" and 1 was from "Exponential and Logarithmic Functions". In terms of mathematical skill, 1 multiple choice and 1 numerical response item were from the "Procedural" level. The other multiple choice item was from the "Conceptual" level and the other numerical response item was from the "Problem Solving" level. Of the 2 "Procedural" level items, both were from the unit "Exponential and Logarithmic Functions". The item from the "Conceptual" level was from the unit "Statistics", and the item from the "Problem Solving" level was from the unit "Trigonometric and Circular Functions".

Of the 3 multiple choice and 1 numerical response items favoring females, 1 multiple choice and 1 numerical response item were from the unit "Trigonometric and Circular Functions" and 2 multiple choice items were from "Quadratic Relations". In terms of mathematical skill, 2 multiple choice items and 1 numerical response were from the "Procedural" level. The other multiple choice item was from the "Conceptual" level. Of the 3 items from the "Procedural" level, 1 multiple choice and 1 numerical response items were from the unit "Trigonometric and Circular Functions". The other multiple choice item was from the unit "Quadratic Relations". The item from the "Conceptual" level was from the unit "Quadratic Relations". No DIF items were detected in the units "Polynomial Functions", "Permutations and Combinations" and "Sequences and Series".

The one polytomous item identified, involving the application of trigonometric functions, favored females. However, analysis of the scale scores was not possible. The results of the LDF analysis of each score scale indicated that there was no difference in the proportion of males and females receiving the scale scores 1 through 5.

1998. All three of the dichotomous items identified favored males. Of these, 2 were multiple choice items and 1 was a numerical response item. Both multiple choice items were from the unit "Series and Sequences", and the numerical response item was from the unit "Trigonometric and Circular Functions". In terms of mathematical skill, 2 were from the "Conceptual" level and 1 was from the "Procedural" level. Of the 2 "Conceptual" level items, the multiple choice item was from the unit "Series and Sequences" and the numerical response item was from the unit "Trigonometric and Circular Functions". The multiple choice item from the "Procedural" level of mathematical skill was from the unit "Series and Sequences". No DIF items were detected in the units "Polynomial Functions", "Statistics", "Quadratic Relations", "Exponential and Logarithmic Functions", and "Permutations and Combinations".

The one polytomous item identified, involving quadratic relations and the application of these principles, favored females. Again, the results of the LDF analysis of each score scale indicated that there was no difference in the proportion of males and females receiving the scale scores 1 through 5.

In summary, more DIF items were detected with DIF in 1997 (9) than in 1998 (4). In 1997, an equal number of dichotomous items favored males and females while the only polytomous item detected favored females. In 1998, the dichotomous items favored males while the only polytomous item detected again favored females. In both academic years, no DIF items were detected in the units "Polynomial Functions", and

"Permutations and Combinations". In addition, in 1997, no DIF items were detected in the unit "Sequences and Series". Similarly, in 1998, no DIF items were detected in the units "Statistics", "Quadratic Relations", and "Exponential and Logarithmic Functions". In contrast, "Trigonometric and Circular Functions" was the only unit in which DIF items were detected in both academic years; however these items favored both males and females. In each year, only one polytomous item was detected. In both cases these favored females. In 1997 this item included content from the trigonometric functions unit and in 1998 this item included content from the quadratic relations unit. However, given the low number of items identified with DIF and the dispersion of these items across the content areas and the two groups of interest, no clear pattern of DIF could be discerned.

Biology

The Biology examination consisted of 48 multiple choice items, 8 gridded response items, and 1 polytomous item. While the gridded items in Mathematics involve calculations to obtain a specific numerical response, in Biology, these items were used to record answers to matching, fill-in-the blank, diagram labeling, and ordering sequences of events items. The prevalence of DIF was similar in both 1997 and 1998. In 1997, a total of 12 items (21.1%) were identified with DIF: 8 multiple choice items and 4 gridded items (see Table 9). In 1998, 10 items (17.5%) were identified with DIF: 7 multiple choice items and 3 gridded items (see Table 9). Of the 8 multiple choice items, 4 favored males and 4 favored females. Likewise, of the 4 gridded response items, 2 favored males and 2 favored females. No polytomous items were identified with DIF. In 1998, 7 dichotomous items favored males, while only 2 dichotomous and 1 polytomous item favored females. Of the 7 dichotomous items favoring males, 3 were gridded numerical response. No gridded response items favored females.

According to the examination blueprint, each dichotomous item is classified by general learner expectations or unit topic (Alberta Education, 1997e, 1998e, p. 2). Unlike English, Social Studies, and Mathematics, there is no classification in Biology by level of thinking. The number of items in each unit topic is listed in Table 15. As shown, the number of items varies across the topics and within topics between the years. In both years the polytomous required students to evaluate data, to incorporate previous knowledge with new information to form new hypotheses, and to make predictions regarding future trends. The student responses were scored holistically using a four-point scale.

1997. As shown in Table 15, of the 4 multiple choice items favoring males, 1 was from the unit "Nervous and Endocrine System"; 2 were from "Cell Division and Mendelian Genetics", and 1 was from "Reproductive Systems and Hormones". Of the 2 gridded response items favoring males, 1 was from the unit "Population Genetics and Interaction" and 1 was from "Nervous and Endocrine System". Of the 4 multiple choice items favoring females, 2 were from the unit "Cell Division and Mendelian Genetics", 1 was from "Nervous and Endocrine System"; and 1 was from "Molecular Genetics". Of the 2 gridded response items favoring females, 1 was from the unit "Nervous and Endocrine" and 1 was from "Cell Division and Mendelian Genetics". As reported earlier, the polytomous item was not identified with DIF in 1997.

Table 15.

Biology DIF Items Organized by Unit Topic: Dichotomous Items

Unit Topic	Year	Male	Female
Nervous & Endocrine System (i_{97} = 16; i_{98} = 17)	1997	2	2
	1998	2	1
Reproductive Systems & Hormones (i_{97} = 8; i_{98} = 4)	1997	1	0
	1998	0	0
Differentiation & Development (i_{97} = 3; i_{98} = 4)	1997	0	0
	1998	1	0
Cell Division & Mendelian Genetics (i_{97} = 14; i_{98} = 18)	1997	2	3
	1998	0	1
Molecular Genetics (i_{97} = 8; i_{98} = 9)	1997	0	1
	1998	4	0
Population Genetics & Interaction (i_{97} = 7; i_{98} = 16)	1997	1	0
	1998	0	0
TOTALS	**1997**	**6**	**6**
	1998	**7**	**2**

1998. Of the 4 multiple choice items favoring males, 2 were from the unit "Molecular Genetics", 1 was from "Nervous and Endocrine System", and 1 was from "Differentiation and Development". Of the 3 gridded response items favoring males, 2 were from the unit "Molecular Genetics" and the third was from "Nervous and Endocrine System". Of the 2 multiple choice items favoring females, 1 item was from the unit "Nervous and Endocrine System" and 1 was from "Cell Division and Mendelian

Genetics". While the polytomous item favored females, analysis of the scale scores was not possible, as LDF did not identify this item as exhibiting DIF.

In summary, a greater number of DIF items were detected with DIF in 1997 (12) than in 1998 (10). In 1997, an equal number of dichotomous items favored males and females. In 1998, a greater number of dichotomous items favored males, while the polytomous item favored females. In addition, in 1998, no DIF items were detected in the unit "Population Genetics & Interactions". In all other units DIF items were detected in both academic years. In 1998, the one polytomous item favored females and dealt with issues of human reproduction. Interestingly, none of the multiple choice items on this topic were detected. Like Mathematics, given the low number of DIF items identified, no clear pattern of the presence of DIF across the units is observed.

DIF Across Subject and Item Format.

Although several previous studies have been reported on a variety of examinations, this is the first study to examine DIF across subject area and item format using a common sample. While O'Neill and McPeek (1993) attempted to integrate these findings across several different examinations, the results are difficult to interpret because of sample differences in age, grade level, academic and extracurricular interests, completion of coursework, and confidence in ability to complete the items. One of the advantages of this study is that the hypotheses regarding gender DIF could be evaluated across the four subject areas for this sample of examinees. In addition, the impact of the number of items identified across the four subject areas could be evaluated as all four examinations were completed by the same sample of examinees.

Combining the results reported in Tables 10, 11, 14, and 15, the prevalence of DIF across subject area and item format is presented in Table 16. In general, not a lot of DIF was present in the dichotomous sections of the examinations.

Table 16.

Prevalence of DIF across Subject Area and Item Format

	1997				1998			
	Dichotomous		Polytomous		Dichotomous		Polytomous	
	M	F	M	F	M	F	M	F
English (i_D=70, i_P = 6)	14	2	0	1	2	3	0	0
Social Studies (i_D=70, i_P = 4)	6	2	0	4	14	4	0	4
Mathematics (i_D=49, i_P = 3)	4	4	0	1	3	0	0	1
Biology (i_D=56, i_P = 1)	6	6	0	0	7	2	0	1
Total (i_D=245, i_P = 14)	30	14	0	6	26	9	0	6

Notes. i_D = the number of dichotomous items, i_P = the number of polytomous items.

The prevalence of DIF in the dichotomous sections of the examinations analyzed is similar to other measures of high school achievement in which 15% to 25% of items were identified with DIF (Hambleton et al., 1993). In 1997, the overall prevalence of DIF across the four examinations was 17.9%; in 1998, the prevalence was 14.3%. Although DIF studies are not routinely conducted on these examinations, they are carefully constructed and screened for potential sources of bias. This likely contributed to the level of DIF observed in this study in each subject.

Of the dichotomous items identified, more items favoring males than females were noted across the four examinations in both 1997 and 1998. In 1997, 16 more dichotomous items favored males than females. Similarly, 15 more dichotomous itemsfavored males than females in 1998. However, the prevalence of these items is related to subject area and examination. In 1997, the incidence of DIF items favoring males was greatest in English and Social Studies. However in both Mathematics and Biology, the incidence of DIF items favoring males and females was equal. In 1998, the incidence of DIF items favoring males was greatest in Social Studies, Mathematics, and Biology. In English, the number of DIF items favoring females and males was almost equal (2 versus 3).

Taken together, the results from the ancillary analysis investigating the prevalence of gender DIF across subject area and item format support some of the previous research and hypotheses regarding differential performance between males and females (e.g., Cole & Willingdon, 1998; O'Neill & McPeek, 1993), whereas other hypotheses are not supported. First, previous findings suggesting that males outperform females on geometry and mathematical problem solving items (O'Neill & McPeek, 1993) were not found in this study. While the majority of these types of items favored neither group, those that were identified with DIF were balanced between the two groups. For example, although three items identified with DIF in 1997 were classified "Trigonometric and Circular Functions", one item favored males and 2 items favored females. Furthermore, across both academic years only one of the 28 dichotomous items classified as "Problem Solving" was identified with DIF. While this item favored males, the other 27 items favored neither males nor females. In addition, the polytomous items requiring problem solving skills and the application of mathematical concepts in practical and applied settings either favored females or did not favor either group.

Second, in this study, mathematics items containing graphs, figures, or tables did not necessarily favor males as has been previously hypothesized (Burton, 1996; Doolittle & Cleary, 1987; Harris & Carlton, 1993; O'Neill & McPeek, 1993). Similarly, mathematics items containing formulas, equations, or symbols did not necessarily favor females (Burton, 1996; Doolittle & Cleary, 1987; Harris & Carlton, 1993; O'Neill & McPeek, 1993). Although items containing these features were identified with DIF, they did not consistently favor one group or the other. Furthermore, the majority of items on the examinations with these features favored neither group.

Third, consistent with previous findings, there were no clear patterns of DIF related to the gridded response items (Burton, 1996). Gridded response items were found on both the Mathematics and Biology examinations. In Mathematics, these items involved calculations to obtain a specific numerical response; whereas, in Biology, these items were used to record answers to matching, fill-in-the blank, diagram labeling, and the ordering of events. While these gridded response items were identified with DIF in both Mathematics and Biology, they did not consistently favor males or females.

Fourth, references to stereotypical male or female activities (O'Neill & McPeek, 1993) were either not identified as DIF items or did not consistently favor one group or the other. For example, a set of items on the 1997 English examination was based on a narrative passage about a father, his sons, and their experience with fly-fishing. Of the 8 items relating to this passage, none of the items were identified with DIF. However, it should also be noted that, in the majority of cases, the items found on the examinations studied did not refer to stereotypical activities of either group. Further, reading passages, problems, and questions that contained references to people had names that were either gender neutral or included both a male and a female name.

In contrast to the dichotomous sections of the examinations, 42.9% (n = 6) of the polytomous items administered in both 1997 and 1998 were identified with DIF, all favoring females. Of the six polytomous items identified, four were from Social Studies and one was from Mathematics in both academic years. The sixth item was from English in 1997 and Biology in 1998. These findings support previous research suggesting that there is an item by format interaction where females perform better on constructed response items even in measures of quantitative ability (Bolger & Kellaghan, 1990; Burton, 1996; Lane et al., 1996; Sadker & Sadker, 1994). While the underlying reasons for these differences are not known, it may be related to stronger writing skills, or more complete answers provided by females (Lane et al., 1996; Willingham & Cole, 1997)

Chapter 5

Summary and Conclusions

This chapter is organized in four sections. In the first section, the research questions and a brief description of the methods used in this study are presented. A summary and discussion of the key findings are presented in the second section. The limitations of the study are then presented in the third section. The last section contains the implications for practice and recommendations for future research.

Summary of Research Questions and Methods

The primary purposes of this study were to determine the agreement 1) between the polytomous DIF detection methods, GMH, Poly-SIB, and LDF and their respective counterparts, MH, SIB, LR; and 2) among the polytomous DIF detection methods, GMH, Poly-SIB, and LDF. The agreement was examined across samples of varying size. The variable gender was selected for the analyses conducted for two reasons. While there are several variables to stratify the two groups for DIF detection studies such as race and school location (urban/rural), gender was uniformly collected across all the studied examinations and provided an adequate focal group sample size across all four examinations. Second, this variable provided the opportunity to investigate hypothesized differences between males and females as described by O'Neill and McPeek (1993). However, unlike the study completed by O'Neill and McPeek, a common sample of examinees across subject area and item format was employed. Hence an ancillary purpose was to investigate the prevalence of gender DIF across four different examinations containing both dichotomous and polytomous items.

More specifically, the following questions were answered in this study:

1. Compared to their dichotomous counterparts (MH, SIB, LR), how consistently do GMH, Poly-SIB, and LDF identify gender DIF across different tests written by the same groups of examinees? How comparable are the methods?
2. How consistently do GMH, Poly-SIB, and LDF identify gender DIF across different tests written by the same group of examinees in examinations consisting of dichotomous and polytomous items? How comparable are the methods?
3. What is the proportion of dichotomous and polytomous items identified with gender DIF in each examination within a sample of student completing all of the examinations of interest? Are there any patterns of DIF across subject area and across item format?

To answer the first question, item responses to the dichotomous section of the English, Social Studies, Mathematics, and Biology Alberta Education Diploma Examinations from six different samples were used to investigate the comparability of the methods within each pair. The four largest samples from the 1998 academic year, as well as the entire sample that wrote the January and June examinations were used for this part of the study. To answer the second question, the same six samples were used to compare GMH, Poly-SIB and LDF, again using gender as a stratification variable. These comparisons were completed twice. First, only the dichotomous item responses for each of the examinations were considered for the comparisons between the three methods. Second, both the dichotomous and polytomous item responses were considered to assess the equivalence of the three generalized procedures.

The DIF detection method Poly-SIB was then used to investigate the prevalence of gender DIF across item format and subject area within a common sample in each of the past two academic years. LDF was used to identify DIF within the scale scores for the polytomous items identified with DIF by Poly-SIB. The sample selected to investigate this question contained those students that wrote all four examinations in June. This pattern included those students enrolled in full-term classes and was hypothesized to have similar academic and extracurricular interests, and completed coursework.

Findings

Comparison of DIF Detection Methods

Both the GMH and Poly-SIB were comparable to their dichotomous counterparts, MH and SIB in sample sizes as low as n = 426. In contrast, while LR and LDF were comparable for the large sample sizes (n = 2395 and n = 3159) they were not comparable for the smaller samples (n = 426 to n = 663). While likely a conservative approach, confidence that LDF will detect all items with moderate to severe DIF is lower, especially in smaller samples.

Comparison of Polytomous DIF Detection Methods

GMH, Poly-SIB, and LDF yielded different results for the set of dichotomous items compared to the set that included the polytomous items. In some cases more dichotomous items were identified, in other cases fewer or the same number of dichotomous items were identified. The differences were not related to sample size, subject, or DIF detection method. They were, however, related to the inclusion of the polytomous items, which altered the total test score (matching variable) and the ability distribution of the reference and focal groups. Consequently, the use of a different matching variable in the analysis of the combined set of items produced different results for the dichotomous items than when the dichotomous items were analyzed alone.

Unfortunately, limitations associated with the respective computer programs for the Poly-SIB and LDF DIF detection methods prevented the analysis of the dichotomous items using a matching variable other than the sum of the items analyzed. Only the GMH detection method afforded the opportunity to use a different matching variable. In this case it was possible to analyze only the dichotomous items using the combined dichotomous and polytomous item test score as the matching variable. In these analyses there were no differences in the results obtained when only the dichotomous items were analyzed using the combined total test score as the matching variable from the results obtained when the dichotomous and polytomous items were analyzed together.

While the items detected varied with the inclusion of the polytomous items, the general pattern of the number of detected items was the same for both analyses. For small sample sizes, LDF was the most conservative. In large sample sizes GMH was the most conservative. Although differences in the numbers of dichotomous and polytomous DIF items identified were noted across the six samples, Poly-SIB produced the most consistent results. In addition, Poly-SIB detected the most items, including the majority of the items identified by GMH and LDF in both the dichotomous set of items and the set of items containing both dichotomous and polytomous items. As the most liberal procedure, it is likely that few DIF items would remain undetected with Poly-SIB, although some items may be falsely identified as exhibiting DIF (Type I error). Conversely, the possibility of not identifying items with DIF (Type II error) may be higher with either GMH or LDF, as would the use of convergence between two methods to identify items.

Identifying items based on convergence is problematic for two reasons. First, items uniquely identified by either method would be eliminated from further study. Second, perfect agreement between the three methods is not expected as convergence of different methods is influenced by the unreliability of the associated statistics (Hambleton & Rogers, 1989). Therefore, the lack of convergence may be falsely interpreted as negligible or no DIF.

The DIF detection method selected is most likely a direct consequence of the purpose of the DIF study and the perspective of the researchers. A more conservative method may be most desirable by test developers and administrators to minimize the cost associated with investigating the identified items. By reducing the number of false positive results, the number of DIF items identified that require further review and/or revision is also reduced, as is the cost associated with this review process. But the minimization of Type I error may not be most desirable by examinees and social advocates who are most concerned about minimizing Type II errors. That is, ensuring that there are no undetected DIF items that might adversely affect the performance of the focal group of interest. Therefore, researchers must be cognizant of the implications of the detection method selected when completing DIF studies.

Prevalence of Gender DIF across Item Format and Subject Area

Other studies investigating the prevalence of DIF in Canadian high school exit examinations composed of dichotomous and polytomous items have not been completed prior to this study. Furthermore, DIF studies are not routinely completed on Alberta Education Diploma examinations. The majority of published DIF research has been completed on American examinations composed of only dichotomous items. Therefore comparisons of the results of this study to previous studies are limited.

Although DIF studies have not been routinely conducted on previous versions of the examinations in this study, the prevalence of DIF on the dichotomous section of the 1997 Social Studies examination is similar to the prevalence rates reported for the 1996 Social Studies examination (Alberta Education, 1996; Gierl & McEwen, 1998). In addition, based on previous studies, the DIF prevalence rates for the dichotomous sections of the examinations in this study are consistent with those reported in the literature for other American high school examinations (15% to 25%; Hambleton, et al. 1993). Only the 1998 English and Mathematics examinations had prevalence rates that were lower (7.1%, 6.2%, respectively). These generally "low" levels of DIF across the four examinations in both academic years are likely related to the test construction process and sensitivity reviews completed on these examinations prior to administration. However, the small number of DIF items identified make it difficult to discern any consistent pattern of gender DIF at the subtest level within each subject due to the sparseness of data across the levels of content and cognitive domains. This is illustrated in the following summary of findings for the examinations in this study.

Overall, more DIF items were detected with DIF in 1997 than 1998, with the exception of Social Studies where more items were detected in 1998 than 1997. In English, the majority of these items were from the curricular content area "Meanings" and the thinking (process) level "Inference and Application" and required students to infer meaning from context. Only one polytomous item, "Writing Skills", was detected in 1997. This item was not detected in 1998. In Social Studies, more DIF items were detected in the curricular content area "Political and Economic Systems", in 1997 whereas in 1998, more DIF items were detected in the curricular content area "Global Interaction in the 20th Century". In both academic years the dichotomous items were distributed across the three knowledge and skill objectives. All four of the polytomous items in both academic years favored females. In Mathematics and Biology, the DIF items were distributed across the content areas and the two groups of interest with no clear observable pattern of DIF. In each year, one polytomous item favoring females was detected in Mathematics. In 1997, this item involved the understanding and application of trigonometric functions, whereas, the 1998 item involved the understanding and application of quadratic relations. The 1998 Biology polytomous item, involving the understanding of issues related to human reproduction, favored females. However, none of the dichotomous items with similar content were identified with DIF.

Taken together, the results from the ancillary analysis investigating the prevalence of gender DIF across subject area and item format support some of the previous research and hypotheses regarding differential performance between males and females (e.g., Cole & Willingdon, 1998; O'Neill & McPeek, 1993), whereas other hypotheses are not supported. First, previous findings suggesting that males outperform females on geometry and mathematical problem solving items (O'Neill & McPeek, 1993) were not found in this study. While the majority of these types of items favored neither group, those that were identified with DIF were balanced between the two groups. For example, although three items identified with DIF in 1997 were classified "Trigonometric and Circular Functions", one item favored males and 2 items favored females. Furthermore, across both academic years only one of the 28 dichotomous items classified as "Problem Solving" was identified with DIF. While this item favored males, the other 27 items favored neither males nor females. In addition, the polytomous items requiring problem solving skills and the application of mathematical concepts in practical and applied settings either favored females or did not favor either group.

Second, in this study, mathematics items containing graphs, figures, or tables did not necessarily favor males as has been previously hypothesized (Burton, 1996; Doolittle & Clear, 1987; Harris & Carlton, 1993; O'Neill & McPeek, 1993). Similarly, mathematics items containing formulas, equations, or symbols did not necessarily favor females (Burton, 1996; Doolittle & Clear, 1987; Harris & Carlton, 1993; O'Neill & McPeek, 1993). Although items containing these features were identified with DIF, they did not consistently favor one group or the other. Furthermore, the majority of items on the examinations with these features favored neither group.

Third, consistent with previous findings, there were no clear patterns of DIF related to the gridded response items (Burton, 1996). Gridded response items were found on both the Mathematics and Biology examinations. In Mathematics, these items involved calculations to obtain a specific numerical response; whereas, in Biology, these items were used to record answers to matching, fill-in-the blank, diagram labeling, and the ordering of events. While these gridded response items were identified with DIF in both Mathematics and Biology, they did not consistently favor males or females.

Fourth, references to stereotypical male or female activities (O'Neill & McPeek, 1993) were either not identified as DIF items or did not consistently favor one group or the other. For example, a set of items on the 1997 English examination was based on a narrative passage about a father, his sons, and their experience with fly-fishing. Of the eight items relating to this passage, none of the items were identified with DIF. However, it should also be noted that, in the majority of cases, the items found on the examinations studied did not refer to stereotypical activities of either group. Further, reading passages, problems, and questions that contained references to people had names that were either gender neutral or included both a male and a female name.

Fifth, while the majority of dichotomous items favored males, all of the polytomous items favored females. These findings support previous research suggesting that there is an item by format interaction where females perform better on constructed response items even in measures of quantitative ability (Bolger & Kellaghan, 1990; Lane et al., 1996; Sadker & Sadker, 1994). While the underlying reasons for these differences are not known, these differences may be related to stronger writing skills, or neater, more complete answers provided by females (Lane et al., 1996; Mazzeo et al., 1993; Willingham & Cole, 1997).

Limitations of the Study

The primary purpose of this study was to compare, in a psychometric sense, polytomous DIF detection methods in examinations consisting of dichotomous and polytomous items. The prevalence of gender DIF across item format and subject area within a common sample of examinees was also completed; however, further analysis to determine underlying reasons for the performance differences was not performed. In most studies this analysis is completed using panels of content experts to help ascertain the source of the performance differences (Ramsey, 1993). Unfortunately these panels are generally unsuccessful at both interpreting or predicting items that perform differently across different groups of examinees (Camilli & Shepard, 1994; Gierl & McEwen, 1998). Protocol analysis (Ericsson & Simon, 1993) holds some promise for helping researchers come to a better understanding of how the items would be interpreted by the actual examinees. However, this method is both resource and time intensive.

While the Alberta Education Diploma Examinations provide a source of data to investigate the prevalence of DIF across different examinations within the same group of examinees, comparisons of individual test items across time were not possible because new examinations are developed for each administration. Furthermore, the findings related to gender DIF are based on the analysis of a specific group of examinees that completed all four examinations at the same time. While the results of this study can be compared to previous research, caution should be used in the generalization of these findings beyond similar samples of examinees. As discussed earlier, it is likely that inconsistent and contradictory findings are related to the use of different types of samples (e.g., convenience, representative, self-selected) or samples that differ by age, grade, and courses completed (Willingham & Cole, 1997). Contradictory findings may also be related to the accumulation of the effects of several individual item characteristics or certain combinations of these characteristics (Bond, 1993) or may be simply an artifact of an imperfect system of measurement (Willingham & Cole, 1997).

In addition, while the use of empirical data provided an opportunity to determine the comparability of the polytomous DIF detection procedures to their dichotomous counterparts and among each other, it should be noted that, in all cases, Type I and Type II error rates could not be definitively determined. The use of a simulation study, comparing these methods could have been used to determine these rates and provide more insight into the comparability of these methods. However, the use of empirical data provided an opportunity to investigate the functioning of the various DIF detection methods under actual rather than ideal conditions.

Recommendations

Future Practice

Based on the results of this study, DIF studies involving only the dichotomous portion of mixed format examinations ignore potential sources of DIF, especially against males. Instead, DIF studies involving the entire set of items should be completed on all examinations. This will help ensure that the set of items administered provides an estimate of ability that is as fair and accurate as possible, regardless of the examinees' gender.

Of the three DIF detection methods investigated in this study, Poly-SIB is recommended for DIF detection in mixed format examinations. Poly-SIB analyzes both dichotomous and polytomous items concurrently and can be easily implemented in the test development and analysis processes. Poly- SIB produced the most consistent results regardless of sample size, but also detected the most items, including the majority of the items identified by GMH and LDF. As the most liberal procedure, it is likely that few DIF items would remain undetected with Poly-SIB, although some items may be falsely identified as exhibiting DIF (Type I error). Conversely, the possibility of not identifying items with DIF (Type II error) may be higher with either GMH or LDF, as would the use of convergence between two methods to identify items.

A more conservative method may be most desirable for test developers and administrators to minimize the cost associated with investigating the identified items. By reducing the number of false positive results, the number of DIF items identified that requiring review is also reduced, as is the cost associated with this review process. But the minimization of Type I error may not be most desirable for examinees and social advocates who are most concerned about minimizing Type II errors and ensuring that there are no undetected DIF items that might adversely affect the performance of the focal group of interest.

For polytomous items identified with DIF, LDF is recommended to examine DIF within the item score points as the other methods determine DIF based on mean item score differences. LDF is most useful when applied in conjunction with another method for the analysis of item scale scores, particularly if the sample size is large.

Future Research

Several implications for future research are evident from the results of this study. First, the generalized versions of MH and SIB (GMH and Poly-SIB) are quite comparable to each other and detect similar items with moderate to severe DIF, even using smaller sample sizes (n = 426 to n = 663). The lack of agreement between LR and LDF is more of a concern. Confidence that LDF will detect all items with moderate to severe DIF is lower, especially in smaller samples. This is attributed to lack of agreement in the determination of DIF severity between the two methods. While effect size measures and cut-points associated with LR are comparable to both GMH and Poly-SIB, they are not comparable to the 95% Scheffé confidence intervals associated with LDF, especially among small sample sizes. Future research investigating and developing an effect size measure and cut-points for LDF that are more comparable to other DIF detection methods and less dependent on sample size is recommended.

Second the results from this study support the need for comprehensive DIF analyses involving both item types in examinations of mixed item format with similar and different samples. In this study, regardless of subject, the majority of dichotomous items favored males while all of the polytomous items favored females. Future research investigating the underlying reason for the performance difference between males and females on polytomous items is also required.

Third, relatively fewer dichotomous and polytomous items were detected in Mathematics and Biology as compared to English and Social Studies. In addition, while the focus of many studies investigating performance differences at both the test and item level have concentrated on mathematics and science more research on the prevalence of DIF on examinations of English and Social Studies is warranted.

References

Alberta Education (1996). Alberta Education Annual Report 1995-96. Edmonton, AB: Author.

Alberta Education (1997a). Alberta Education Annual Report 1996-97. Edmonton, AB: Author.

Alberta Education (1997b). English 30 Diploma Examination Results Examiners' Report for June 1997. Edmonton, AB: Author.

Alberta Education (1997c). Social Studies 30 Diploma Examination Results Examiners' Report for June 1997. Edmonton, AB: Author.

Alberta Education (1997d). Mathematics 30 Diploma Examination Results Examiners' Report for June 1997. Edmonton, AB: Author.

Alberta Education (1997e). Biology 30 Diploma Examination Results Examiners' Report for June 1997. Edmonton, AB: Author.

Alberta Education (1998a). Alberta Education Annual Report 1997-98. Edmonton, AB: Author.

Alberta Education (1998b). English 30 Diploma Examination Results Examiners' Report for June 1998. Edmonton, AB: Author.

Alberta Education (1998c). Social Studies 30 Diploma Examination Results Examiners' Report for June 1998. Edmonton, AB: Author.

Alberta Education (1998d). Mathematics 30 Diploma Examination Results Examiners' Report for June 1998. Edmonton, AB: Author.

Alberta Education (1998e). Biology 30 Diploma Examination Results Examiners' Report for June 1998. Edmonton, AB: Author.

Angoff, W. (1993). Perspectives on differential item functioning methodology. In P. W. Holland & H. Wainer (Eds.), Differential Item Functioning (pp. 3-24). Hillsdale, NJ: Lawrence Erlbaum.

Benbow, C. P. (1988). Sex differences in mathematical reasoning ability in intellectually talented preadolescents: Their nature, effects, and possible causes. Behavioral and Brain Sciences, 11, 169-232.

Bennett, R. E., Rock, D. A., & Wang, M. (1991). Equivalence of free-response and multiple choice items. Journal of Educational Measurement, 28, 77-92.

Bolger, N. & Kellaghan, T. (1990). Method of measurement and gender differences in scholastic achievement. Journal of Educational Measurement, 27, 165-174.

Bond, L. (1993). Comments on the O'Neill & McPeek paper. In P. W. Holland & H. Wainer (Eds.), Differential Item Functioning (pp. 277-280). Hillsdale, NJ: Lawrence Erlbaum.

Breland, H. & Danos, D., Kahn, H., Kubota, M., & Bonner, M., (1994). Performance versus objective testing and gender: An exploratory study of an Advanced Placement History Examination. Journal of Educational Measurement, 31, 275-393.

Burton, N. M. (1996). How have changes in the SAT affected women's math scores? Educational Measurement: Issues and Practice, 15(4), 5-9.

Camilli, G. & Shepard, L. A. (1994). Methods for identifying biased test items. Newbury Park, CA: Sage.

Carlton, S. T., & Harris, A. M. (1989, March). Female/male performance differences on the SAT: Causes and correlates. Paper presented at the annual meeting of the American Educational Research Association, San Francisco, CA.

Chang, H.-H., Mazzeo, J., & Roussos, L. (1996). Detecting DIF for polytomously scored items: An adaptation of the SIB procedure. Journal of Educational Measurement, 33, 333-353.

Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items: An NCME instructional module. Educational Measurement: Issues and Practice, 17(1), 31-44.

Clauser, B. E., Mazor, K. M., & Hambleton, R. (1991). Influence of the criterion variable on the identification of differentially functioning test items using the Mantel-Haenszel statistic. Applied Psychological Measurement, 15, 353-359.

Cohen, A. S., Kim, S.-H., & Baker, F. B. (1993). Detection of differential item functioning in the graded response model. Applied Psychological Measurement, 17, 335-350.

Cohen, J. (1988). Statistical power for the behavioral sciences (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Doolittle, A. E. (1989). Gender differences in performance on mathematics achievement items. Applied Measurement in Education, 2 161-177.

Doolittle, A. E. & Cleary, T. A. (1987). Gender differences in performance on mathematics achievement items. Journal of Educational Measurement, 24, 157-166.

Dorans, N., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), Differential Item Functioning (pp. 35-66). Hillsdale, NJ: Lawrence Erlbaum.

Downing, S. M., & Haladyna, T. M. (1997). Test item development: Validity evidence from quality assurance procedures. Applied Measurement in Education, 10, 61-82.

Dunbar, S. B., Koretz, D. M., & Hoover, H. D. (1991). Quality control in the development and use of performance assessments. Applied Measurement in Education, 4, 289-303.

Ericsson, K. A., & Simon, H. A. (1993). Protocol analyses: Verbal reports as data (rev. ed.). Cambridge, MA: MIT Press.

French, A. W. & Miller, T. R. (1996). Logistic regression and its use in detecting differential item functioning in polytomous items. Journal of Educational Measurement, 33, 315-332.

Garner, M., & Engelhard, G. J. (1999). Gender differences in performance on multiple-choice and constructed response items. Applied Measurement in Education, 12, 29-51.

Gierl, M. J. & McEwen, N. (1998, May). Differential item functioning on the Alberta Education Social Studies 30 examination. Paper presented at the annual meeting of the Canadian Society for the Study of Education, Ottawa, ON.

Gierl, M. J., Rogers, W. T. & Klinger, D. A. (in press). Using statistical and substantive reviews to identify and interpret translation DIF. Alberta Journal of Educational Research [Special Issue: "Measurement and Evaluation: Current and Future Research Directions for the New Millenium].

Harris, A. M., & Carlton, S. T. (1993). Patterns of gender differences on mathematics items on the Scholastic Aptitude Test. Applied Measurement in Education, 6, 137-151.

Halpern, D. F. (1992). Sex differences in cognitive abilities (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

Hambleton, R. K., Clauser B. E., Mazor, K. M., & Jones, R. W. (1993). Advances in the detection of differentially functioning test items. European Journal of Psychological Assessment, 9, 1-18.

Hambleton, R. K., & Rogers, H. J. (1989). Detection of potentially biased test items: Comparison of IRT area and Mantel-Haenszel methods. Applied Measurement in Education, 2, 313-334.

Hambleton, R. K., & Swaminathan, H. (1985). Item response theory: Principles and applications. Boston, MA: Kluwer-Nijhoff.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). Fundamentals of item response theory. Newbury Park, CA: Sage.

Harwell, M., Stone, C. A., Hsu, T.-C. & Kirisci, L. (1996). Monte Carlo studies in item response theory. Applied Psychological Measurement, 20, 101-125.

Holland, P. W., & Thayer, D. T. (1988). Differential item functioning and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.) Test Validity (pp. 129-145). Hillsdale, NJ: Lawrence Erlbaum.

Jodoin, M. G. (1999). Reducing type I error rates using an effect size measure with the logistic regression procedure for DIF detection. Unpublished master's thesis, University of Alberta, Edmonton, Alberta, Canada.

Lane, S., Wang, N., & Magon, M. (1996). Gender-related differential item functioning on a middle-school mathematics performance assessment. Educational Researcher, 15, 21-27, 31.

Leucht, R. & Spray, J. (1993). Visual DIF (VISDIF; Version 2.0) [Computer software]. Iowa City, IA: American College Testing.

Li, H-H., Nandakumar, R., & Stout, W. (1995, April). Application of SIB in dealing with issues of DIF in the context of multidimensional data. Paper presented at the annual meeting of the National Council of Measurement in Education, San Francisco, CA.

Lord, F. M. (1980). Applications of item response theory to practical testing problems. Hillsdale, NJ: Lawrence Erlbaum.

Lord, F. M., & Novick, M. R. (1968). Statistical theories of mental test scores. Reading, MA: Addison-Wesley.

Mazor, K. M., Clauser, B. E., & Hambleton, R. (1992). The effect of sample size on the functioning of the M-H statistic. Educational and Psychological Measurement, 52, 443-451.

Mazor, K., Kanjee, A., & Clauser, B. E. (1993, April). Using logistic regression with multiple ability estimates to detect differential item functioning. Journal of Educational Measurement, 32, 134-144.

Mazzeo, J., Schmitt, A. P., & Bleistein, C. A. (1993). Sex-related performance differences on constructed-response and multiple-choice sections of Advanced Placement Examinations (College Board Report No. 97-2). New York, NY: College Entrance Examination Board.

Miller, T. R. & Spray, J. A. (1993). Logistic discriminant function analysis for DIF identification of polytomously scored items. Journal of Educational Measurement, 30, 107-122.

Muraki, E. & Bock, R. D. (1997). Parscale: IRT item analysis and test scoring for rating-scale data. Chicago, IL: Scientific Software International.

Narayanan, P. & Swaminathan, H. (1994). Performance of Mantel-Haenszel and simultaneous item bias procedure for detecting differential item functioning. Applied Psychological Measurement, 18, 315-328.

O'Neil, H. F., Jr. & Brown, R. S. (1997). Differential effects of question formats in math assessment on metacognition and affect. Applied Measurement in Education, 11, 331-352.

O'Neill, K. A. & McPeek, W. M. (1993). Item and test characteristics that are associated with differential item functioning. In P. W. Holland & H. Wainer (Eds.), Differential Item Functioning (pp. 255-276). Hillsdale, NJ: Lawrence Erlbaum.

Pomplun, M., & Sundbye, N. (1991). Gender differences in constructed response reading items. Applied Measurement in Education, 12, 95-109.

Potenza, M. T. & Dorans, N. J. (1995). DIF assessment for polytomously scored items: A framework for classification and evaluation. Applied Psychological Measurement, 19, 23-37.

Raju, N. S. (1988). The area between two item characteristic curves. Psychometrika, 53, 495-502.

Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. Applied Psychological Measurement, 14, 197-207.

Ramsey, P. A. (1993). Sensitivity review: The ETS experience as a case study. In P. W. Holland & H. Wainer (Eds.), Differential Item Functioning (pp. 367 – 388). Hillsdale, NJ: Lawrence Erlbaum.

Rogers, H. J. & Swaminathan, H. (1993). A comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning. Applied Psychological Measurement, 17, 105-116.

Roussos, L., & Stout, W. (1996). Simulation studies of the effects of small sample size and studied item parameters on SIB and Mantel-Haenszel type I error performance. Journal of Educational Measurement, 33, 215-230.

Rudner, L. M., Geston, P. R., & Knight, D. L. (1980) Biased item detection techniques. Journal of Educational Measurement, 17, 213-233.

Sadker, M. & Sadker, D. (1994). Failing at fairness. How our schools cheat girls. Toronto, ON: Simon & Schuster.

Scheuneman, J. D., & Gerritz, K. (1990). Using differential item functioning procedures to explore sources of item difficulty and group performance characteristics. Journal of Educational Measurement, 24, 97-118.

Scheuneman, J. D., & Grima, A. (1997). Characteristics of quantitative word items associated with differential performance for female and black examinees. Applied Measurement in Education, 10, 299-319.

Shealy, R., & Stout, W. F. (1993). An item response theory model for test bias. In P. W. Holland & H. Wainer (Eds.), Differential Item Functioning (pp. 197-239). Hillsdale, NJ: Lawrence Erlbaum.

Stanley, J. C., Benbow, C. P., Brody, L. E., Dauber, S., & Lupkowski, A. E. (1992). Gender differences on eighty-six nationally standardized aptitude and achievement tests. In N. Colangelo, S. G. Assouline, & D. L. Ambroson (Eds.), Talent Development: Proceedings from the 1991 Henry B. and Jocelyn Wallace National Research Symposium on Talent Development (pp. 41-48). Iowa City, IA: Trillium.

Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression. Journal of Educational Measurement, 27, 361-370.

Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In H. Wainer & H. I. Braun (Eds.), Test Validity (pp. 147-169). Hillsdale, NJ: Lawrence Erlbaum.

Wainer, H., Sireci, S. G., & Thissen, D. (1991). Differential item functioning: Definitions and detection. Journal of Educational Measurement, 28, 197-219.

Wightman, L. F. (1998). An examination of sex differences in LSAT scores from the perspective of social consequences. Applied Measurement in Education, 11, 255-277.

Willingham, W. W. & Cole, N. S. (1997). Fairness issues in test design and use. In W.W. Willingham & N. S. Cole (Eds.), Gender and Fair Assessment (pp. 227 - 346). Hillsdale, NJ: Lawrence Erlbaum.

Zumbo, B. D. & Thomas, D. R. (1996, October). A measure of effect size using logistic regression procedures. Paper presented at the National Board of Medical Examiners, Philadelphia, PA.

Zwick, R., Donoghue, J. R., & Grima, A. (1993). Assessment of differential item functioning for performance tasks. Journal of Educational Measurement, 30, 233-351

Zwick, R., & Erickan, K. (1989). Analyses of differential item functioning in the NAEP history assessment. Journal of Educational Measurement, 26, 53-66.

Appendix A

Comparison of MH, GMH and SIB, Poly-SIB Results for each Examination and Sample:

Dichotomous Items Only

Table 17.

MH/GMH and SIB/Poly-SIB Comparisons: English January, 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	-1.740	2	-1.758	2	-0.099	2	-0.099	2
2	-0.270	1	-0.273	1	-0.027	1	-0.028	1
3	-0.170	1	-0.168	1	-0.011	1	-0.012	1
4	-0.450	1	-0.459	1	-0.024	1	-0.024	1
5	-0.200	1	-0.203	1	-0.006	1	-0.006	1
6	-0.320	1	-0.320	1	-0.016	1	-0.016	1
7	-0.530	1	-0.544	1	-0.016	1	-0.017	1
8	-0.280	1	-0.284	1	-0.023	1	-0.024	1
9	-0.330	1	-0.338	1	-0.010	1	-0.01	1
10	-0.150	1	-0.154	1	-0.018	1	-0.018	1
11	-0.440	1	-0.449	1	-0.025	1	-0.025	1
12	-1.510	2	-1.418	2	-0.027	1	-0.027	1
13	-0.630	1	-0.621	1	-0.038	1	-0.038	1
14	0.400	1	0.392	1	0.031	1	0.031	1
15	0.890	1	0.869	1	0.079	2	0.078	2
16	0.470	1	0.471	1	0.042	1	0.04	1
17	0.980	1	0.976	1	0.085	2	0.084	2
18	1.000	2	0.996	2	0.102	2	0.102	2
19	-0.220	1	-0.224	1	-0.009	1	-0.01	1
20	-0.690	1	-0.698	1	-0.046	1	-0.047	1
21	-0.050	1	-0.053	1	-0.007	1	-0.007	1
22	-0.860	1	-0.854	1	-0.083	2	-0.083	2
23	0.420	1	0.412	1	0.035	1	0.035	1
24	-0.160	1	-0.157	1	-0.006	1	-0.007	1
25	0.170	1	0.167	1	0.010	1	0.009	1
26	-0.970	1	-0.977	1	-0.071	2	-0.071	2
27	0.910	1	0.895	1	0.062	2	0.062	2
28	0.490	1	0.487	1	0.049	1	0.049	1
29	0.500	1	0.496	1	0.017	1	0.017	1
30	0.320	1	0.320	1	0.026	1	0.025	1
31	-0.630	1	-0.638	1	-0.040	1	-0.041	1
32	0.340	1	0.338	1	0.029	1	0.029	1
33	-0.030	1	-0.030	1	-0.006	1	-0.006	1
34	0.120	1	0.116	1	0.007	1	0.007	1
35	0.950	1	0.935	1	0.078	2	0.078	2
36	-0.010	1	-0.014	1	-0.005	1	-0.005	1
37	0.940	1	0.909	1	0.065	2	0.065	2
38	0.580	1	0.581	1	0.051	1	0.05	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
39	1.040	2	1.033	2	0.080	2	0.08	2
40	1.500	2	1.461	2	0.098	2	0.098	2
41	-0.260	1	-0.263	1	-0.018	1	-0.018	1
42	1.510	2	1.478	2	0.082	2	0.082	2
43	-0.330	1	-0.330	1	-0.023	1	-0.023	1
44	-1.110	2	-1.135	2	-0.063	2	-0.063	2
45	0.070	1	0.069	1	0.006	1	0.005	1
46	-0.200	1	-0.196	1	-0.018	1	-0.018	1
47	-0.550	1	-0.545	1	-0.025	1	-0.025	1
48	-1.350	2	-1.348	2	-0.120	2	-0.12	2
49	0.100	1	0.100	1	0.013	1	0.013	1
50	-1.100	2	-1.116	2	-0.060	2	-0.06	2
51	-0.890	1	-0.894	1	-0.064	2	-0.065	2
52	-1.160	2	-1.164	2	-0.078	2	-0.078	2
53	0.470	1	0.458	1	0.020	1	0.02	1
54	0.020	1	0.022	1	0.001	1	0.001	1
55	0.050	1	0.045	1	0.004	1	0.003	1
56	-1.420	2	-1.444	2	-0.055	1	-0.055	1
57	0.420	1	0.422	1	0.033	1	0.033	1
58	-0.010	1	-0.011	1	-0.003	1	-0.003	1
59	-0.270	1	-0.274	1	-0.020	1	-0.02	1
60	0.560	1	0.549	1	0.038	1	0.037	1
61	0.350	1	0.343	1	0.026	1	0.026	1
62	0.630	1	0.632	1	0.048	1	0.047	1
63	-0.550	1	-0.550	1	-0.034	1	-0.035	1
64	0.030	1	0.028	1	0.000	1	0	1
65	-0.380	1	-0.378	1	-0.033	1	-0.034	1
66	0.690	1	0.677	1	0.058	1	0.057	1
67	-0.380	1	-0.376	1	-0.043	1	-0.043	1
68	-0.090	1	-0.092	1	-0.008	1	-0.009	1
69	0.290	1	0.291	1	0.021	1	0.02	1
70	-0.920	1	-0.936	1	-0.049	1	-0.049	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 18.

MH/GMH and SIB/Poly-SIB Comparisons: English June, 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	-0.750	1	-0.759	1	-0.017	1	-0.017	1
2	-0.280	1	-0.279	1	-0.024	1	-0.024	1
3	0.890	1	0.882	1	0.079	2	0.079	2
4	0.560	1	0.563	1	0.044	1	0.044	1
5	0.290	1	0.287	1	0.030	1	0.03	1
6	0.260	1	0.260	1	0.020	1	0.02	1
7	-0.520	1	-0.516	1	-0.035	1	-0.036	1
8	-0.140	1	-0.135	1	-0.017	1	-0.017	1
9	-0.660	1	-0.676	1	-0.054	1	-0.054	1
10	-0.040	1	-0.045	1	-0.008	1	-0.008	1
11	-0.440	1	-0.438	1	-0.040	1	-0.039	1
12	-0.320	1	-0.326	1	-0.027	1	-0.028	1
13	0.830	1	0.827	1	0.077	2	0.076	2
14	-0.140	1	-0.144	1	-0.003	1	-0.003	1
15	0.920	1	0.894	1	0.024	1	0.024	1
16	0.400	1	0.398	1	0.016	1	0.016	1
17	0.470	1	0.467	1	0.027	1	0.027	1
18	-0.020	1	-0.025	1	-0.001	1	-0.001	1
19	-0.170	1	-0.173	1	-0.013	1	-0.013	1
20	-1.040	2	-1.041	2	-0.091	2	-0.091	2
21	-1.480	2	-1.486	2	-0.126	2	-0.126	2
22	-0.260	1	-0.265	1	-0.024	1	-0.024	1
23	-0.240	1	-0.244	1	-0.025	1	-0.025	1
24	-0.140	1	-0.137	1	-0.012	1	-0.012	1
25	0.140	1	0.143	1	0.013	1	0.013	1
26	-0.020	1	-0.019	1	-0.006	1	-0.006	1
27	0.140	1	0.135	1	0.005	1	0.005	1
28	1.210	2	1.176	2	0.106	2	0.107	2
29	0.350	1	0.343	1	0.026	1	0.026	1
30	0.600	1	0.596	1	0.054	1	0.054	1
31	-0.700	1	-0.711	1	-0.058	1	-0.059	2
32	1.260	2	1.226	2	0.036	1	0.036	1
33	-0.090	1	-0.091	1	0.000	1	0	1
34	0.210	1	0.206	1	0.017	1	0.017	1
35	-0.010	1	-0.012	1	0.001	1	0.001	1
36	0.060	1	0.064	1	0.007	1	0.007	1
37	-0.380	1	-0.382	1	-0.031	1	-0.031	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
38	0.620	1	0.605	1	0.028	1	0.028	1
39	0.290	1	0.289	1	0.019	1	0.02	1
40	0.240	1	0.245	1	0.021	1	0.021	1
41	-0.140	1	-0.138	1	-0.006	1	-0.007	1
42	-0.230	1	-0.231	1	-0.008	1	-0.009	1
43	-0.090	1	-0.093	1	-0.008	1	-0.008	1
44	-0.220	1	-0.222	1	-0.016	1	-0.016	1
45	-0.090	1	-0.089	1	-0.003	1	-0.003	1
46	0.350	1	0.342	1	0.034	1	0.034	1
47	0.590	1	0.592	1	0.055	1	0.055	1
48	0.360	1	0.362	1	0.036	1	0.036	1
49	-0.950	1	-0.957	1	-0.059	2	-0.059	2
50	1.140	2	1.119	2	0.100	2	0.1	2
51	-0.120	1	-0.115	1	-0.003	1	-0.004	1
52	0.250	1	0.248	1	0.019	1	0.019	1
53	0.340	1	0.334	1	0.037	1	0.037	1
54	-0.890	1	-0.903	1	-0.071	2	-0.072	2
55	-0.150	1	-0.151	1	-0.015	1	-0.015	1
56	-0.270	1	-0.268	1	-0.017	1	-0.017	1
57	-0.480	1	-0.489	1	-0.014	1	-0.014	1
58	-0.660	1	-0.661	1	-0.063	2	-0.063	2
59	0.120	1	0.125	1	0.004	1	0.004	1
60	-0.010	1	-0.008	1	-0.004	1	-0.004	1
61	-0.230	1	-0.228	1	-0.021	1	-0.021	1
62	-0.150	1	-0.151	1	-0.007	1	-0.006	1
63	0.140	1	0.141	1	0.013	1	0.014	1
64	-0.040	1	-0.037	1	-0.003	1	-0.002	1
65	-0.100	1	-0.102	1	-0.008	1	-0.008	1
66	-0.070	1	-0.069	1	-0.001	1	-0.002	1
67	0.120	1	0.117	1	0.021	1	0.021	1
68	-0.290	1	-0.292	1	-0.025	1	-0.025	1
69	0.230	1	0.234	1	0.023	1	0.025	1
70	-0.150	1	-0.147	1	-0.008	1	-0.008	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 19.

MH/GMH and SIB/Poly-SIB Comparisons: English 1212, 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	-2.030	2	-2.017	2	-0.096	2	-0.096	2
2	-1.120	2	-1.124	2	-0.089	2	-0.09	2
3	0.300	1	0.305	1	0.011	1	0.008	1
4	-0.400	1	-0.406	1	-0.015	1	-0.014	1
5	0.750	1	0.748	1	0.008	1	0.01	1
6	-0.430	1	-0.415	1	-0.011	1	-0.014	1
7	-1.110	1	-1.152	1	-0.030	1	-0.03	1
8	-0.590	1	-0.593	1	-0.035	1	-0.038	1
9	-0.600	1	-0.632	1	-0.007	1	-0.005	1
10	-0.460	1	-0.459	1	-0.019	1	-0.017	1
11	-0.680	1	-0.667	1	-0.034	1	-0.034	1
12	-0.960	1	-0.942	1	-0.017	1	-0.017	1
13	-0.510	1	-0.523	1	-0.057	1	-0.056	1
14	-0.200	1	-0.197	1	-0.015	1	-0.017	1
15	0.620	1	0.595	1	0.071	1	0.07	1
16	0.280	1	0.280	1	0.039	1	0.038	1
17	0.970	1	0.965	1	0.090	2	0.09	2
18	1.000	2	1.016	2	0.105	2	0.103	2
19	-0.130	1	-0.130	1	-0.005	1	-0.005	1
20	-0.820	1	-0.820	1	-0.023	1	-0.022	1
21	-0.290	1	-0.290	1	-0.036	1	-0.036	1
22	-0.830	1	-0.810	1	-0.078	2	-0.08	2
23	0.340	1	0.351	1	0.022	1	0.023	1
24	-0.070	1	-0.073	1	0.009	1	0.011	1
25	-0.250	1	-0.249	1	-0.016	1	-0.017	1
26	-0.680	1	-0.681	1	-0.039	1	-0.041	1
27	0.940	1	0.909	1	0.047	1	0.044	1
28	1.160	2	1.137	2	0.120	2	0.121	2
29	-0.080	1	-0.076	1	-0.019	1	-0.021	1
30	0.060	1	0.057	1	0.001	1	-0.001	1
31	-1.170	2	-1.155	2	-0.056	1	-0.056	1
32	0.710	1	0.706	1	0.059	1	0.059	1
33	0.200	1	0.210	1	0.008	1	0.006	1
34	-0.010	1	-0.014	1	0.001	1	-0.001	1
35	1.150	2	1.132	2	0.102	2	0.099	2
36	0.090	1	0.089	1	-0.013	1	-0.013	1
37	0.850	1	0.828	1	0.062	1	0.061	1
38	0.230	1	0.223	1	0.017	1	0.017	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
39	1.350	2	1.363	2	0.100	2	0.101	2
40	1.660	2	1.619	2	0.124	2	0.123	2
41	0.030	1	0.029	1	-0.016	1	-0.017	1
42	1.810	2	1.715	2	0.097	2	0.096	2
43	-0.300	1	-0.302	1	-0.021	1	-0.021	1
44	-0.560	1	-0.576	1	-0.032	1	-0.031	1
45	-0.020	1	-0.019	1	-0.003	1	-0.005	1
46	0.570	1	0.566	1	0.048	1	0.049	1
47	-0.250	1	-0.248	1	-0.010	1	-0.009	1
48	-1.230	2	-1.197	2	-0.102	2	-0.104	2
49	0.600	1	0.598	1	0.034	1	0.035	1
50	-0.840	1	-0.822	1	-0.056	1	-0.059	2
51	-0.820	1	-0.840	1	-0.054	1	-0.058	1
52	-0.820	1	-0.803	1	-0.048	1	-0.05	1
53	1.010	1	0.938	1	0.062	2	0.062	2
54	-0.820	1	-0.794	1	-0.030	1	-0.03	1
55	0.270	1	0.276	1	0.009	1	0.01	1
56	-0.800	1	-0.767	1	-0.023	1	-0.023	1
57	0.520	1	0.532	1	0.052	1	0.052	1
58	-0.430	1	-0.440	1	-0.025	1	-0.025	1
59	-0.190	1	-0.185	1	-0.024	1	-0.027	1
60	0.140	1	0.137	1	-0.014	1	-0.016	1
61	0.210	1	0.207	1	0.008	1	0.005	1
62	-0.240	1	-0.236	1	-0.011	1	-0.013	1
63	-0.080	1	-0.082	1	-0.038	1	-0.042	1
64	0.050	1	0.047	1	-0.003	1	-0.003	1
65	-0.190	1	-0.183	1	-0.025	1	-0.029	1
66	0.810	1	0.787	1	0.054	1	0.052	1
67	-0.390	1	-0.400	1	-0.056	1	-0.056	1
68	-0.490	1	-0.483	1	-0.041	1	-0.044	1
69	0.540	1	0.532	1	0.047	1	0.047	1
70	-0.950	1	-0.950	1	-0.036	1	-0.038	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 20.

MH/GMH and SIB/Poly-SIB Comparisons: English 2112, 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	0.730	1	0.728	1	0.022	1	0.026	1
2	-0.230	1	-0.225	1	-0.004	1	0	1
3	1.120	2	1.093	2	0.116	2	0.121	2
4	0.530	1	0.545	1	0.057	1	0.063	1
5	0.320	1	0.317	1	0.053	1	0.055	1
6	0.950	1	0.918	1	0.037	1	0.038	1
7	-0.830	1	-0.826	1	-0.065	2	-0.061	2
8	-1.000	2	-0.983	1	-0.072	1	-0.071	1
9	-0.310	1	-0.306	1	-0.014	1	-0.014	1
10	0.410	1	0.400	1	0.032	1	0.032	1
11	0.010	1	0.005	1	0.003	1	0.006	1
12	-0.350	1	-0.355	1	-0.020	1	-0.014	1
13	1.240	2	1.230	2	0.108	2	0.107	2
14	1.520	1	1.501	1	0.034	1	0.033	1
15	0.910	1	0.888	1	0.036	1	0.037	1
16	0.210	1	0.207	1	0.002	1	0.003	1
17	1.600	2	1.562	2	0.070	2	0.074	2
18	-0.460	1	-0.461	1	-0.040	1	-0.039	1
19	-0.590	1	-0.598	1	-0.030	1	-0.03	1
20	-0.460	1	-0.459	1	-0.028	1	-0.027	1
21	-1.190	2	-1.188	2	-0.103	2	-0.102	2
22	-0.180	1	-0.185	1	-0.013	1	-0.011	1
23	-0.530	1	-0.531	1	-0.056	1	-0.049	1
24	0.740	1	0.721	1	0.032	1	0.029	1
25	0.270	1	0.268	1	0.027	1	0.033	1
26	0.160	1	0.157	1	0.016	1	0.023	1
27	-0.140	1	-0.146	1	-0.021	1	-0.022	1
28	0.440	1	0.454	1	0.058	1	0.061	1
29	0.070	1	0.068	1	0.008	1	0.015	1
30	0.450	1	0.441	1	0.038	1	0.045	1
31	-0.520	1	-0.517	1	-0.018	1	-0.015	1
32	0.930	1	0.888	1	0.020	1	0.021	1
33	-1.310	1	-1.309	1	-0.036	1	-0.035	1
34	0.500	1	0.495	1	0.037	1	0.041	1
35	-0.690	1	-0.687	1	-0.028	1	-0.028	1
36	-0.250	1	-0.256	1	0.001	1	0.002	1
37	0.070	1	0.074	1	0.000	1	-0.001	1
38	-0.130	1	-0.129	1	-0.009	1	-0.004	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
39	0.360	1	0.345	1	0.019	1	0.027	1
40	0.590	1	0.599	1	0.028	1	0.035	1
41	0.000	1	0.003	1	-0.017	1	-0.009	1
42	-0.740	1	-0.770	1	-0.028	1	-0.029	1
43	-0.300	1	-0.311	1	-0.020	1	-0.018	1
44	-0.190	1	-0.182	1	-0.010	1	-0.006	1
45	-0.110	1	-0.112	1	0.001	1	0.002	1
46	0.100	1	0.102	1	0.007	1	0.009	1
47	0.930	1	0.918	1	0.090	2	0.091	2
48	-0.180	1	-0.179	1	0.007	1	0.004	1
49	-1.170	2	-1.177	2	-0.079	2	-0.075	2
50	1.640	2	1.565	2	0.129	2	0.132	2
51	-0.150	1	-0.151	1	-0.023	1	-0.015	1
52	0.360	1	0.353	1	0.024	1	0.029	1
53	-0.150	1	-0.153	1	0.000	1	0.001	1
54	-0.550	1	-0.548	1	-0.051	1	-0.046	1
55	-0.410	1	-0.400	1	-0.043	1	-0.043	1
56	0.340	1	0.339	1	0.010	1	0.013	1
57	-1.980	2	-2.038	2	-0.057	1	-0.055	1
58	-0.500	1	-0.511	1	-0.074	1	-0.074	1
59	-0.230	1	-0.236	1	-0.018	1	-0.015	1
60	-0.610	1	-0.609	1	-0.060	1	-0.057	1
61	-0.470	1	-0.471	1	-0.036	1	-0.033	1
62	-0.210	1	-0.214	1	-0.024	1	-0.022	1
63	0.710	1	0.708	1	0.084	2	0.089	2
64	-0.130	1	-0.131	1	-0.005	1	-0.001	1
65	-0.260	1	-0.261	1	-0.018	1	-0.017	1
66	-0.240	1	-0.238	1	-0.015	1	-0.011	1
67	1.040	2	1.033	2	0.100	2	0.105	2
68	-0.190	1	-0.184	1	-0.026	1	-0.022	1
69	-0.360	1	-0.365	1	-0.044	1	-0.044	1
70	0.300	1	0.302	1	0.007	1	0.014	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 21.

MH/GMH and SIB/Poly-SIB Comparisons: English 2211, 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	-0.550	1	-0.581	1	-0.011	1	-0.012	1
2	-0.930	1	-0.953	1	-0.117	2	-0.117	2
3	0.470	1	0.456	1	0.049	1	0.052	1
4	0.810	1	0.800	1	0.053	1	0.054	1
5	-0.040	1	-0.041	1	-0.004	1	-0.004	1
6	-0.020	1	-0.016	1	0.000	1	0.001	1
7	-0.340	1	-0.348	1	-0.045	1	-0.044	1
8	0.450	1	0.449	1	0.046	1	0.046	1
9	-0.210	1	-0.207	1	-0.035	1	-0.034	1
10	-0.160	1	-0.158	1	-0.020	1	-0.019	1
11	0.760	1	0.754	1	0.042	1	0.043	1
12	-0.380	1	-0.371	1	-0.022	1	-0.021	1
13	0.510	1	0.514	1	0.080	1	0.069	1
14	-1.640	1	-1.613	1	-0.027	1	-0.026	1
15	-0.320	1	-0.327	1	-0.013	1	-0.005	1
16	0.410	1	0.400	1	0.013	1	0.017	1
17	-0.170	1	-0.169	1	-0.002	1	-0.004	1
18	-0.010	1	-0.012	1	0.008	1	0.012	1
19	0.110	1	0.110	1	-0.013	1	-0.011	1
20	-1.000	2	-0.999	2	-0.077	1	-0.08	1
21	-1.380	2	-1.407	2	-0.124	2	-0.129	2
22	-0.790	1	-0.809	1	-0.050	1	-0.046	1
23	-0.190	1	-0.193	1	-0.001	1	0.000	1
24	-0.900	1	-0.905	1	-0.057	1	-0.056	1
25	-0.750	1	-0.750	1	-0.069	1	-0.072	1
26	-0.650	1	-0.680	1	-0.079	1	-0.079	1
27	0.550	1	0.546	1	0.050	1	0.049	1
28	0.880	1	0.855	1	0.074	1	0.076	1
29	1.450	2	1.374	2	0.098	2	0.101	2
30	0.490	1	0.482	1	0.037	1	0.037	1
31	-0.900	1	-0.926	1	-0.075	1	-0.076	1
32	1.490	1	1.420	1	0.027	1	0.026	1
33	-1.200	1	-1.161	1	-0.027	1	-0.027	1
34	-0.490	1	-0.497	1	-0.029	1	-0.036	1
35	-0.120	1	-0.116	1	-0.005	1	-0.002	1
36	0.250	1	0.242	1	0.010	1	0.019	1
37	-0.490	1	-0.495	1	-0.055	1	-0.055	1
38	0.840	1	0.807	1	0.017	1	0.018	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
39	-0.370	1	-0.394	1	-0.045	1	-0.045	1
40	1.050	1	1.050	1	0.069	1	0.069	1
41	-0.610	1	-0.610	1	-0.022	1	-0.024	1
42	1.100	1	1.075	1	0.024	1	0.025	1
43	-0.270	1	-0.267	1	-0.002	1	-0.002	1
44	-0.670	1	-0.668	1	-0.069	1	-0.068	1
45	-0.020	1	-0.023	1	-0.013	1	-0.012	1
46	-0.070	1	-0.066	1	-0.008	1	-0.008	1
47	0.170	1	0.171	1	0.043	1	0.043	1
48	-0.060	1	-0.056	1	0.023	1	0.024	1
49	-1.840	2	-1.843	2	-0.105	2	-0.105	2
50	1.080	2	1.057	2	0.066	1	0.068	1
51	0.150	1	0.149	1	-0.020	1	-0.02	1
52	0.160	1	0.162	1	0.005	1	0.008	1
53	0.320	1	0.325	1	0.036	1	0.033	1
54	-0.210	1	-0.211	1	0.001	1	0.001	1
55	0.290	1	0.291	1	0.031	1	0.033	1
56	-0.250	1	-0.251	1	-0.027	1	-0.026	1
57	0.310	1	0.310	1	0.009	1	0.008	1
58	-0.590	1	-0.586	1	-0.067	1	-0.070	1
59	0.310	1	0.306	1	0.023	1	0.025	1
60	0.470	1	0.469	1	0.063	1	0.06	1
61	-0.500	1	-0.493	1	-0.036	1	-0.036	1
62	0.070	1	0.075	1	0.016	1	0.015	1
63	0.470	1	0.469	1	0.038	1	0.035	1
64	-0.080	1	-0.079	1	-0.002	1	-0.003	1
65	0.540	1	0.532	1	0.045	1	0.049	1
66	-0.070	1	-0.069	1	0.014	1	0.013	1
67	0.460	1	0.444	1	0.047	1	0.044	1
68	0.070	1	0.069	1	0.008	1	0.018	1
69	0.690	1	0.679	1	0.082	1	0.081	1
70	0.310	1	0.306	1	0.030	1	0.030	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 22.

MH/GMH and SIB/Poly-SIB Comparisons: English 2222, 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	0.180	1	0.186	1	-0.002	1	-0.004	1
2	-0.040	1	-0.043	1	-0.002	1	0.001	1
3	0.670	1	0.657	1	0.067	1	0.067	1
4	0.700	1	0.681	1	0.059	1	0.066	1
5	-0.100	1	-0.098	1	0.016	1	0.017	1
6	0.280	1	0.261	1	0.023	1	0.027	1
7	-0.980	1	-1.006	1	-0.066	1	-0.064	1
8	0.140	1	0.135	1	0.023	1	0.023	1
9	-0.110	1	-0.118	1	-0.005	1	-0.005	1
10	-0.140	1	-0.142	1	0.004	1	0.004	1
11	-0.610	1	-0.601	1	-0.064	1	-0.064	1
12	-0.220	1	-0.214	1	-0.026	1	-0.018	1
13	0.530	1	0.541	1	0.047	1	0.044	1
14	0.750	1	0.779	1	-0.004	1	-0.002	1
15	2.020	1	1.917	1	0.011	1	0.014	1
16	-0.110	1	-0.111	1	-0.007	1	-0.004	1
17	-0.590	1	-0.589	1	-0.004	1	-0.002	1
18	1.340	2	1.278	2	0.086	1	0.09	1
19	0.030	1	0.025	1	0.025	1	0.024	1
20	-0.650	1	-0.657	1	-0.046	1	-0.046	1
21	-1.740	2	-1.732	2	-0.142	2	-0.143	2
22	0.130	1	0.126	1	0.026	1	0.035	1
23	0.090	1	0.088	1	0.026	1	0.028	1
24	-0.600	1	-0.592	1	-0.037	1	-0.037	1
25	0.660	1	0.652	1	0.063	1	0.061	1
26	0.520	1	0.513	1	0.030	1	0.029	1
27	-0.100	1	-0.104	1	-0.027	1	-0.029	1
28	1.820	2	1.860	2	0.157	2	0.158	2
29	-0.440	1	-0.411	1	-0.013	1	-0.008	1
30	0.170	1	0.166	1	0.042	1	0.046	1
31	0.080	1	0.082	1	-0.010	1	-0.008	1
32	2.010	2	1.911	2	0.035	1	0.035	1
33	0.220	1	0.214	1	0.023	1	0.021	1
34	-0.500	1	-0.498	1	0.000	1	0.001	1
35	0.050	1	0.047	1	0.028	1	0.028	1
36	0.360	1	0.343	1	0.025	1	0.03	1
37	-0.660	1	-0.648	1	-0.048	1	-0.047	1
38	1.020	1	1.054	1	0.075	2	0.078	2

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
39	0.540	1	0.523	1	0.013	1	0.016	1
40	-0.010	1	-0.010	1	0.014	1	0.018	1
41	-0.090	1	-0.093	1	-0.003	1	-0.002	1
42	-0.410	1	-0.412	1	-0.031	1	-0.03	1
43	-0.580	1	-0.611	1	-0.022	1	-0.019	1
44	-0.240	1	-0.244	1	-0.022	1	-0.021	1
45	-0.910	1	-0.943	1	-0.059	1	-0.059	1
46	1.030	2	1.031	2	0.124	2	0.127	2
47	0.500	1	0.496	1	0.044	1	0.043	1
48	-0.380	1	-0.380	1	-0.003	1	-0.004	1
49	-0.990	1	-1.021	1	-0.052	1	-0.05	1
50	1.440	2	1.371	2	0.140	2	0.143	2
51	-0.330	1	-0.341	1	-0.021	1	-0.018	1
52	-0.810	1	-0.826	1	-0.051	1	-0.047	1
53	-0.180	1	-0.177	1	-0.013	1	-0.015	1
54	-1.680	2	-1.700	2	-0.170	2	-0.168	2
55	1.140	2	1.115	2	0.108	2	0.11	2
56	-0.440	1	-0.434	1	-0.039	1	-0.036	1
57	-0.970	1	-0.927	1	-0.029	1	-0.025	1
58	-0.550	1	-0.556	1	-0.026	1	-0.021	1
59	0.050	1	0.046	1	0.030	1	0.031	1
60	-0.720	1	-0.717	1	-0.064	1	-0.057	1
61	-0.060	1	-0.062	1	0.040	1	0.042	1
62	-0.580	1	-0.571	1	-0.025	1	-0.021	1
63	0.160	1	0.159	1	0.026	1	0.031	1
64	0.590	1	0.597	1	0.012	1	0.014	1
65	-0.040	1	-0.039	1	-0.004	1	-0.004	1
66	0.380	1	0.378	1	0.039	1	0.04	1
67	-0.050	1	-0.047	1	0.003	1	0.003	1
68	-0.180	1	-0.180	1	-0.029	1	-0.026	1
69	0.550	1	0.531	1	0.044	1	0.046	1
70	-1.500	2	-1.452	2	-0.094	1	-0.093	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 23.

MH/GMH and SIB/Poly-SIB Comparisons: Social January 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	-0.100	1	-0.102	1	-0.006	1	-0.01	1
2	-0.140	1	-0.139	1	-0.010	1	-0.01	1
3	0.440	1	0.435	1	0.015	1	0.017	1
4	-0.320	1	-0.315	1	-0.021	1	-0.02	1
5	0.860	1	0.834	1	0.050	1	0.051	1
6	0.020	1	0.019	1	0.009	1	0.01	1
7	-0.500	1	-0.495	1	-0.031	1	-0.03	1
8	0.390	1	0.395	1	0.028	1	0.03	1
9	0.130	1	0.130	1	0.004	1	0.006	1
10	-0.430	1	-0.437	1	-0.023	1	-0.02	1
11	0.180	1	0.182	1	0.012	1	0.015	1
12	-0.040	1	-0.037	1	0.004	1	0.004	1
13	-0.460	1	-0.460	1	-0.038	1	-0.04	1
14	-0.960	1	-0.952	1	-0.088	2	-0.09	2
15	-0.470	1	-0.472	1	-0.038	1	-0.04	1
16	-0.520	1	-0.516	1	-0.053	1	-0.05	1
17	-0.280	1	-0.274	1	-0.019	1	-0.02	1
18	0.370	1	0.370	1	0.030	1	0.031	1
19	-0.430	1	-0.432	1	-0.031	1	-0.03	1
20	-0.200	1	-0.200	1	-0.022	1	-0.02	1
21	-0.310	1	-0.312	1	-0.010	1	-0.01	1
22	-0.350	1	-0.346	1	-0.032	1	-0.03	1
23	-0.130	1	-0.129	1	-0.003	1	-0	1
24	-0.390	1	-0.395	1	-0.034	1	-0.03	1
25	-0.010	1	-0.013	1	-0.002	1	-0	1
26	-1.120	2	-1.121	2	-0.089	2	-0.09	2
27	-0.130	1	-0.134	1	-0.011	1	-0.01	1
28	-0.890	1	-0.891	1	-0.074	2	-0.07	2
29	-0.170	1	-0.179	1	-0.015	1	-0.01	1
30	0.750	1	0.746	1	0.047	1	0.048	1
31	-0.140	1	-0.135	1	-0.006	1	-0.01	1
32	-1.800	2	-1.816	2	-0.123	2	-0.12	2
33	-0.370	1	-0.374	1	-0.025	1	-0.02	1
34	-0.480	1	-0.479	1	-0.019	1	-0.02	1
35	-0.240	1	-0.238	1	-0.013	1	-0.01	1
36	0.450	1	0.443	1	0.038	1	0.039	1
37	0.270	1	0.269	1	0.020	1	0.021	1
38	0.200	1	0.203	1	0.012	1	0.013	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
39	-0.550	1	-0.551	1	-0.051	1	-0.05	1
40	0.300	1	0.294	1	0.017	1	0.017	1
41	0.680	1	0.676	1	0.057	1	0.058	1
42	-0.350	1	-0.349	1	-0.031	1	-0.03	1
43	0.220	1	0.224	1	0.026	1	0.026	1
44	-0.170	1	-0.173	1	-0.018	1	-0.02	1
45	0.490	1	0.470	1	0.012	1	0.015	1
46	0.640	1	0.640	1	0.046	1	0.046	1
47	-0.240	1	-0.237	1	-0.007	1	-0.01	1
48	-0.160	1	-0.160	1	-0.008	1	-0.01	1
49	0.420	1	0.415	1	0.022	1	0.023	1
50	-0.510	1	-0.515	1	-0.026	1	-0.02	1
51	-0.160	1	-0.155	1	-0.008	1	-0.01	1
52	0.710	1	0.703	1	0.050	1	0.051	1
53	1.290	2	1.269	2	0.106	2	0.107	2
54	-0.520	1	-0.522	1	-0.031	1	-0.03	1
55	0.440	1	0.431	1	0.044	1	0.043	1
56	0.260	1	0.257	1	0.023	1	0.024	1
57	0.740	1	0.742	1	0.057	1	0.057	1
58	0.140	1	0.141	1	0.020	1	0.021	1
59	-0.330	1	-0.328	1	-0.029	1	-0.03	1
60	1.520	2	1.442	2	0.086	2	0.087	2
61	0.580	1	0.577	1	0.056	1	0.056	1
62	-0.580	1	-0.585	1	-0.043	1	-0.04	1
63	-0.360	1	-0.356	1	-0.017	1	-0.02	1
64	0.260	1	0.257	1	0.018	1	0.018	1
65	-0.320	1	-0.324	1	-0.008	1	-0.01	1
66	0.570	1	0.567	1	0.031	1	0.032	1
67	0.200	1	0.196	1	0.025	1	0.025	1
68	1.070	2	1.045	2	0.079	2	0.078	2
69	0.760	1	0.757	1	0.065	2	0.066	2
70	1.010	2	0.996	1	0.078	2	0.078	2

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 24.

MH/GMH and SIB/Poly-SIB Comparisons: Social June 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	-0.020	1	-0.022	1	-0.004	1	-0.003	1
2	-0.540	1	-0.539	1	-0.036	1	-0.035	1
3	-0.070	1	-0.074	1	-0.012	1	-0.01	1
4	0.370	1	0.370	1	0.022	1	0.023	1
5	1.130	2	1.106	2	0.080	2	0.08	2
6	-0.770	1	-0.771	1	-0.066	2	-0.067	2
7	-0.080	1	-0.079	1	-0.015	1	-0.014	1
8	-0.340	1	-0.342	1	-0.026	1	-0.026	1
9	0.050	1	0.048	1	-0.002	1	-0.002	1
10	-0.270	1	-0.265	1	-0.023	1	-0.023	1
11	0.890	1	0.897	1	0.069	2	0.069	2
12	0.000	1	0.002	1	-0.004	1	-0.003	1
13	-0.330	1	-0.331	1	-0.033	1	-0.034	1
14	0.050	1	0.053	1	0.002	1	0.002	1
15	-1.030	2	-1.042	2	-0.067	2	-0.066	2
16	-0.390	1	-0.388	1	-0.026	1	-0.025	1
17	-0.140	1	-0.137	1	-0.011	1	-0.011	1
18	-0.120	1	-0.117	1	-0.008	1	-0.007	1
19	0.110	1	0.108	1	0.010	1	0.01	1
20	0.240	1	0.236	1	0.021	1	0.021	1
21	0.990	1	0.994	1	0.079	2	0.079	2
22	0.230	1	0.225	1	0.017	1	0.016	1
23	-0.090	1	-0.087	1	-0.007	1	-0.006	1
24	-0.830	1	-0.813	1	-0.082	2	-0.082	2
25	0.110	1	0.112	1	-0.001	1	0.001	1
26	-0.880	1	-0.879	1	-0.077	2	-0.079	2
27	-0.870	1	-0.883	1	-0.050	1	-0.05	1
28	-0.220	1	-0.215	1	-0.010	1	-0.011	1
29	-0.200	1	-0.204	1	-0.019	1	-0.019	1
30	-0.260	1	-0.261	1	-0.014	1	-0.014	1
31	-0.900	1	-0.889	1	-0.080	2	-0.08	2
32	-0.030	1	-0.030	1	0.002	1	0.003	1
33	-0.020	1	-0.019	1	0.009	1	0.008	1
34	-0.840	1	-0.842	1	-0.077	2	-0.078	2
35	-0.560	1	-0.563	1	-0.048	1	-0.047	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
36	0.300	1	0.298	1	0.016	1	0.017	1
37	-0.220	1	-0.221	1	-0.024	1	-0.023	1
38	-0.790	1	-0.804	1	-0.047	1	-0.046	1
39	0.000	1	0.001	1	-0.007	1	-0.006	1
40	-0.710	1	-0.710	1	-0.072	2	-0.073	2
41	-1.070	2	-1.072	2	-0.092	2	-0.091	2
42	0.340	1	0.336	1	0.028	1	0.028	1
43	0.580	1	0.578	1	0.052	1	0.052	1
44	-0.380	1	-0.379	1	-0.040	1	-0.04	1
45	1.060	2	1.037	2	0.063	2	0.064	2
46	-0.370	1	-0.376	1	-0.040	1	-0.04	1
47	-0.760	1	-0.767	1	-0.038	1	-0.036	1
48	1.040	2	1.046	2	0.103	2	0.102	2
49	0.670	1	0.657	1	0.038	1	0.038	1
50	2.230	2	2.140	2	0.149	2	0.15	2
51	-0.180	1	-0.179	1	-0.004	1	-0.003	1
52	1.230	2	1.193	2	0.069	2	0.069	2
53	-0.170	1	-0.167	1	0.005	1	0.005	1
54	1.010	2	0.989	1	0.073	2	0.073	2
55	0.480	1	0.481	1	0.033	1	0.033	1
56	-0.560	1	-0.563	1	-0.045	1	-0.045	1
57	0.630	1	0.626	1	0.054	1	0.054	1
58	0.820	1	0.800	1	0.040	1	0.041	1
59	0.920	1	0.919	1	0.083	2	0.085	2
60	1.190	2	1.147	2	0.056	1	0.058	1
61	-0.320	1	-0.314	1	-0.030	1	-0.03	1
62	1.110	2	1.069	2	0.042	1	0.045	1
63	-0.180	1	-0.179	1	-0.016	1	-0.015	1
64	0.470	1	0.472	1	0.017	1	0.018	1
65	0.330	1	0.330	1	0.020	1	0.02	1
66	0.080	1	0.077	1	0.006	1	0.007	1
67	-0.590	1	-0.595	1	-0.035	1	-0.034	1
68	-0.020	1	-0.024	1	-0.006	1	-0.004	1
69	-0.480	1	-0.476	1	-0.041	1	-0.04	1
70	-0.260	1	-0.267	1	-0.021	1	-0.02	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 25.

MH/GMH and SIB/Poly-SIB Comparisons: Social 1212 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	-0.210	1	-0.209	1	-0.013	1	-0.009	1
2	-0.100	1	-0.102	1	0.008	1	0.008	1
3	-1.020	1	-1.009	1	-0.031	1	-0.031	1
4	-0.060	1	-0.061	1	0.001	1	-0.001	1
5	1.040	2	1.048	2	0.101	2	0.099	2
6	-0.920	1	-0.915	1	-0.101	2	-0.102	2
7	0.270	1	0.276	1	0.012	1	0.013	1
8	-0.290	1	-0.286	1	-0.023	1	-0.024	1
9	0.570	1	0.565	1	0.040	1	0.043	1
10	0.150	1	0.151	1	0.017	1	0.02	1
11	0.630	1	0.634	1	0.085	2	0.081	2
12	0.190	1	0.200	1	0.003	1	0.003	1
13	-0.210	1	-0.209	1	0.000	1	-0.005	1
14	0.210	1	0.216	1	-0.007	1	-0.002	1
15	-0.850	1	-0.814	1	-0.047	1	-0.049	1
16	-0.620	1	-0.615	1	-0.048	1	-0.046	1
17	0.170	1	0.173	1	0.036	1	0.041	1
18	-0.910	1	-0.888	1	-0.043	1	-0.043	1
19	-0.490	1	-0.477	1	-0.050	1	-0.049	1
20	-0.610	1	-0.603	1	-0.046	1	-0.047	1
21	1.250	2	1.251	2	0.114	2	0.109	2
22	0.630	1	0.622	1	0.051	1	0.046	1
23	-0.280	1	-0.281	1	-0.019	1	-0.018	1
24	-1.430	2	-1.419	2	-0.122	2	-0.131	2
25	-0.100	1	-0.095	1	0.009	1	0.004	1
26	-0.790	1	-0.772	1	-0.058	1	-0.056	1
27	-0.880	1	-0.893	1	-0.064	1	-0.06	1
28	-0.230	1	-0.227	1	-0.015	1	-0.021	1
29	-0.520	1	-0.530	1	-0.038	1	-0.039	1
30	-0.390	1	-0.396	1	-0.016	1	-0.018	1
31	-0.650	1	-0.646	1	-0.067	1	-0.065	1
32	-0.230	1	-0.233	1	-0.016	1	-0.021	1
33	0.340	1	0.341	1	0.036	1	0.036	1
34	-0.610	1	-0.627	1	-0.062	1	-0.061	1
35	-0.180	1	-0.186	1	-0.023	1	-0.021	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
36	0.920	1	0.922	1	0.078	2	0.079	2
37	-0.520	1	-0.533	1	-0.049	1	-0.055	1
38	-1.270	2	-1.352	2	-0.076	2	-0.083	2
39	-0.310	1	-0.306	1	-0.014	1	-0.012	1
40	-0.510	1	-0.516	1	-0.048	1	-0.054	1
41	-1.210	2	-1.201	2	-0.079	2	-0.081	2
42	0.190	1	0.189	1	0.037	1	0.035	1
43	0.420	1	0.411	1	0.033	1	0.034	1
44	-0.300	1	-0.295	1	-0.030	1	-0.028	1
45	0.780	1	0.770	1	0.048	1	0.047	1
46	-0.600	1	-0.596	1	-0.027	1	-0.035	1
47	-0.200	1	-0.195	1	-0.020	1	-0.022	1
48	1.060	2	1.054	2	0.102	2	0.102	2
49	0.940	1	0.913	1	0.057	1	0.059	1
50	2.080	2	2.006	2	0.123	2	0.132	2
51	0.230	1	0.232	1	0.015	1	0.017	1
52	0.540	1	0.538	1	0.037	1	0.04	1
53	0.090	1	0.086	1	0.007	1	0.001	1
54	1.310	2	1.294	2	0.133	2	0.131	2
55	0.420	1	0.420	1	0.033	1	0.032	1
56	-0.770	1	-0.756	1	-0.052	1	-0.051	1
57	1.160	2	1.151	2	0.131	2	0.13	2
58	0.230	1	0.227	1	0.012	1	0.013	1
59	0.660	1	0.655	1	0.073	1	0.074	1
60	1.300	2	1.282	2	0.094	2	0.095	2
61	0.380	1	0.375	1	0.032	1	0.029	1
62	1.820	2	1.673	2	0.090	2	0.09	2
63	0.040	1	0.039	1	0.032	1	0.038	1
64	0.290	1	0.292	1	0.017	1	0.02	1
65	0.300	1	0.301	1	0.023	1	0.022	1
66	0.330	1	0.332	1	0.047	1	0.047	1
67	-0.430	1	-0.454	1	-0.018	1	-0.024	1
68	0.920	1	0.871	1	0.010	1	0.011	1
69	-0.540	1	-0.541	1	-0.048	1	-0.044	1
70	-0.620	1	-0.622	1	-0.006	1	-0.004	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 26.

MH/GMH and SIB/Poly-SIB Comparisons: Social 2112 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	0.110	1	0.112	1	0.018	1	0.018	1
2	0.170	1	0.166	1	-0.001	1	0.014	1
3	0.220	1	0.215	1	0.035	1	0.035	1
4	-0.100	1	-0.104	1	0.004	1	0.001	1
5	1.160	1	1.092	1	0.056	1	0.056	1
6	0.400	1	0.406	1	0.017	1	0.03	1
7	-0.410	1	-0.419	1	-0.060	1	-0.047	1
8	0.600	1	0.620	1	0.045	1	0.057	1
9	0.180	1	0.181	1	0.021	1	0.019	1
10	0.200	1	0.200	1	0.030	1	0.038	1
11	-0.260	1	-0.257	1	-0.024	1	-0.021	1
12	-0.260	1	-0.271	1	-0.048	1	-0.039	1
13	-0.390	1	-0.396	1	-0.057	1	-0.049	1
14	-1.100	2	-1.082	2	-0.127	2	-0.113	2
15	-0.420	1	-0.417	1	-0.057	1	-0.046	1
16	-0.940	1	-0.907	1	-0.091	2	-0.081	2
17	-1.030	1	-1.018	1	-0.076	2	-0.068	2
18	0.230	1	0.223	1	0.014	1	0.029	1
19	-0.250	1	-0.248	1	-0.020	1	-0.023	1
20	-0.220	1	-0.222	1	-0.048	1	-0.042	1
21	-0.050	1	-0.055	1	-0.006	1	-0.001	1
22	-0.190	1	-0.195	1	-0.021	1	-0.024	1
23	-0.090	1	-0.094	1	-0.010	1	-0.012	1
24	-0.330	1	-0.330	1	-0.050	1	-0.04	1
25	-0.280	1	-0.298	1	0.005	1	-0.004	1
26	-0.990	1	-0.980	1	-0.102	2	-0.094	2
27	-0.560	1	-0.569	1	-0.041	1	-0.046	1
28	-0.450	1	-0.443	1	-0.049	1	-0.041	1
29	-0.030	1	-0.028	1	-0.029	1	-0.028	1
30	0.800	1	0.821	1	0.041	1	0.043	1
31	-0.950	1	-0.951	1	-0.105	2	-0.097	2
32	-1.390	2	-1.378	2	-0.093	2	-0.09	2
33	-0.070	1	-0.071	1	-0.031	1	-0.032	1
34	-0.630	1	-0.626	1	-0.053	1	-0.042	1
35	-0.490	1	-0.489	1	-0.031	1	-0.02	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
36	0.910	1	0.898	1	0.057	1	0.061	1
37	-0.010	1	-0.014	1	-0.031	1	-0.016	1
38	-0.360	1	-0.369	1	-0.044	1	-0.031	1
39	-0.950	1	-0.943	1	-0.112	2	-0.102	2
40	0.840	1	0.799	1	0.053	1	0.049	1
41	0.780	1	0.780	1	0.024	1	0.038	1
42	0.100	1	0.101	1	-0.024	1	-0.01	1
43	0.970	1	0.949	1	0.052	1	0.061	1
44	-0.560	1	-0.572	1	-0.060	1	-0.049	1
45	-0.970	1	-0.919	1	-0.024	1	-0.025	1
46	1.440	2	1.428	2	0.089	2	0.096	2
47	-0.470	1	-0.457	1	-0.002	1	-0.008	1
48	-0.970	1	-0.951	1	-0.070	2	-0.078	2
49	0.320	1	0.328	1	0.013	1	0.019	1
50	-0.260	1	-0.261	1	-0.025	1	-0.023	1
51	-0.440	1	-0.440	1	-0.037	1	-0.042	1
52	1.010	1	1.034	1	0.047	1	0.053	1
53	0.530	1	0.526	1	0.037	1	0.044	1
54	-0.190	1	-0.191	1	-0.032	1	-0.016	1
55	0.430	1	0.431	1	0.010	1	0.021	1
56	1.360	2	1.288	2	0.091	2	0.092	2
57	0.560	1	0.553	1	0.001	1	0.019	1
58	0.600	1	0.575	1	0.038	1	0.05	1
59	-0.330	1	-0.323	1	-0.034	1	-0.02	1
60	2.010	2	1.922	2	0.099	2	0.093	2
61	1.180	2	1.146	2	0.072	1	0.083	2
62	-0.520	1	-0.522	1	-0.027	1	-0.034	1
63	-0.820	1	-0.868	1	-0.030	1	-0.035	1
64	-0.180	1	-0.172	1	-0.024	1	-0.017	1
65	-1.490	1	-1.548	1	-0.028	1	-0.028	1
66	-0.780	1	-0.782	1	-0.026	1	-0.03	1
67	0.170	1	0.173	1	0.017	1	0.015	1
68	1.630	2	1.547	2	0.103	2	0.11	2
69	0.450	1	0.457	1	0.050	1	0.058	1
70	0.940	1	0.944	1	0.065	1	0.064	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 27.

MH/GMH and SIB/Poly-SIB Comparisons: Social 2211 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	-0.130	1	-0.125	1	-0.005	1	-0.005	1
2	-0.280	1	-0.289	1	-0.028	1	-0.028	1
3	-0.520	1	-0.538	1	-0.035	1	-0.035	1
4	0.180	1	0.178	1	0.005	1	0.005	1
5	1.480	2	1.487	2	0.110	2	0.11	2
6	-0.390	1	-0.387	1	-0.026	1	-0.026	1
7	-2.030	2	-1.973	2	-0.133	2	-0.133	2
8	-1.000	1	-1.030	1	-0.054	1	-0.054	1
9	-0.800	1	-0.796	1	-0.068	1	-0.068	1
10	-0.960	1	-0.944	1	-0.075	1	-0.075	1
11	0.940	1	0.978	1	0.106	2	0.106	2
12	-0.370	1	-0.380	1	-0.004	1	-0.004	1
13	0.260	1	0.265	1	0.015	1	0.015	1
14	0.120	1	0.121	1	-0.030	1	-0.03	1
15	-2.210	2	-2.140	2	-0.119	2	-0.119	2
16	-0.710	1	-0.707	1	-0.076	2	-0.076	2
17	-0.910	1	-0.912	1	-0.045	1	-0.045	1
18	-0.780	1	-0.801	1	-0.040	1	-0.04	1
19	0.330	1	0.311	1	0.021	1	0.021	1
20	0.410	1	0.405	1	0.042	1	0.042	1
21	1.570	2	1.555	2	0.143	2	0.143	2
22	0.100	1	0.095	1	-0.004	1	-0.004	1
23	-0.340	1	-0.329	1	-0.038	1	-0.038	1
24	-0.590	1	-0.584	1	-0.063	1	-0.063	1
25	-0.050	1	-0.044	1	-0.021	1	-0.021	1
26	-0.890	1	-0.906	1	-0.108	2	-0.108	2
27	-1.040	1	-1.026	1	-0.062	1	-0.062	1
28	-0.040	1	-0.045	1	-0.023	1	-0.023	1
29	0.630	1	0.653	1	0.062	1	0.062	1
30	-0.490	1	-0.472	1	-0.020	1	-0.02	1
31	-0.990	1	-0.967	1	-0.091	2	-0.091	2
32	0.520	1	0.514	1	0.005	1	0.005	1
33	-0.220	1	-0.222	1	-0.023	1	-0.023	1
34	-1.410	2	-1.371	2	-0.151	2	-0.151	2
35	-0.670	1	-0.650	1	-0.050	1	-0.05	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
36	0.000	1	-0.001	1	-0.011	1	-0.011	1
37	-0.370	1	-0.373	1	-0.054	1	-0.054	1
38	-0.530	1	-0.533	1	-0.014	1	-0.014	1
39	0.240	1	0.241	1	0.006	1	0.006	1
40	-0.430	1	-0.435	1	-0.046	1	-0.046	1
41	-1.400	2	-1.426	2	-0.113	2	-0.113	2
42	0.190	1	0.178	1	0.028	1	0.028	1
43	1.270	2	1.235	2	0.065	1	0.065	1
44	0.580	1	0.606	1	0.043	1	0.043	1
45	1.510	2	1.425	2	0.075	2	0.075	2
46	0.040	1	0.036	1	-0.019	1	-0.019	1
47	-0.370	1	-0.397	1	-0.033	1	-0.033	1
48	0.730	1	0.735	1	0.088	1	0.088	1
49	0.750	1	0.707	1	0.035	1	0.035	1
50	3.660	2	3.185	2	0.213	2	0.213	2
51	-0.160	1	-0.163	1	-0.025	1	-0.025	1
52	1.360	1	1.297	1	0.065	2	0.065	2
53	-0.970	1	-0.976	1	-0.058	1	-0.058	1
54	0.590	1	0.609	1	0.043	1	0.043	1
55	0.080	1	0.078	1	-0.001	1	-0.001	1
56	0.570	1	0.541	1	-0.007	1	-0.007	1
57	0.530	1	0.525	1	0.030	1	0.03	1
58	0.550	1	0.509	1	0.001	1	0.001	1
59	0.820	1	0.804	1	0.049	1	0.049	1
60	0.530	1	0.533	1	0.039	1	0.039	1
61	-0.760	1	-0.758	1	-0.069	1	-0.069	1
62	2.200	2	1.998	2	0.072	2	0.072	2
63	0.080	1	0.084	1	-0.008	1	-0.008	1
64	1.430	1	1.308	1	0.049	1	0.049	1
65	0.790	1	0.769	1	0.005	1	0.005	1
66	0.250	1	0.250	1	0.006	1	0.006	1
67	-0.110	1	-0.110	1	-0.043	1	-0.043	1
68	-1.310	1	-1.309	1	-0.021	1	-0.021	1
69	-0.900	1	-0.903	1	-0.059	1	-0.059	1
70	0.160	1	0.159	1	0.012	1	0.012	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 28.

MH/GMH and SIB/Poly-SIB Comparisons: Social 2222 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	0.980	1	1.113	1	0.037	1	0.037	1
2	0.370	1	0.358	1	0.023	1	0.023	1
3	-0.890	1	-0.872	1	-0.049	1	-0.047	1
4	0.950	1	1.042	1	0.039	1	0.038	1
5	1.400	2	1.356	2	0.091	2	0.093	2
6	-0.710	1	-0.361	1	-0.082	1	-0.09	1
7	1.170	2	1.635	2	0.110	2	0.115	2
8	-0.830	1	-0.135	1	-0.070	1	-0.072	1
9	-0.080	1	0.678	1	0.006	1	0.006	1
10	-0.110	1	0.016	1	-0.031	1	-0.031	1
11	1.210	2	1.711	2	0.094	1	0.094	1
12	-0.130	1	0.397	1	-0.007	1	-0.011	1
13	-1.370	2	-0.794	1	-0.135	2	-0.139	2
14	1.130	1	1.091	1	0.087	1	0.088	1
15	-1.200	1	-1.217	1	-0.085	1	-0.085	1
16	-0.080	1	0.535	1	0.029	1	0.027	1
17	0.280	1	-0.060	1	0.000	1	-0.001	1
18	1.460	1	1.087	1	0.061	1	0.06	1
19	0.190	1	0.861	1	0.022	1	0.021	1
20	0.280	1	0.795	1	0.057	1	0.06	1
21	-0.010	1	0.126	1	-0.019	1	-0.017	1
22	-0.320	1	-0.074	1	-0.006	1	-0.007	1
23	0.540	1	0.292	1	0.010	1	0.018	1
24	-1.380	2	-1.276	2	-0.121	2	-0.13	2
25	1.030	1	1.501	1	0.024	1	0.036	1
26	-2.070	2	-1.656	2	-0.204	2	-0.209	2
27	-0.140	1	-0.440	1	-0.037	1	-0.038	1
28	0.100	1	0.432	1	0.018	1	0.018	1
29	-0.960	1	-1.096	1	-0.054	1	-0.049	1
30	-0.560	1	0.033	1	-0.038	1	-0.039	1
31	-0.730	1	-0.192	1	-0.098	2	-0.097	2
32	0.630	1	0.762	1	0.033	1	0.033	1
33	-0.150	1	0.205	1	-0.026	1	-0.028	1
34	-1.020	1	-0.347	1	-0.098	1	-0.095	1
35	-0.660	1	-0.784	1	-0.093	1	-0.086	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
36	-0.380	1	0.554	1	-0.053	1	-0.053	1
37	-0.610	1	0.179	1	-0.083	1	-0.084	1
38	-0.750	1	-0.712	1	-0.041	1	-0.04	1
39	0.440	1	0.419	1	0.020	1	0.024	1
40	-1.190	2	-1.162	2	-0.081	1	-0.088	1
41	-0.450	1	-0.567	1	-0.061	1	-0.058	1
42	0.410	1	1.172	1	0.010	1	0.01	1
43	0.510	1	0.895	1	0.064	1	0.064	1
44	-0.980	1	-0.557	1	-0.065	1	-0.066	1
45	2.230	2	2.290	2	0.136	2	0.143	2
46	-0.640	1	-0.977	1	-0.033	1	-0.035	1
47	-1.340	1	-1.477	1	-0.052	1	-0.051	1
48	1.380	2	1.584	2	0.112	2	0.112	2
49	0.820	1	1.315	1	0.062	1	0.057	1
50	2.870	2	1.969	2	0.182	2	0.193	2
51	-0.470	1	0.286	1	-0.028	1	-0.03	1
52	0.870	1	1.583	1	0.069	1	0.074	1
53	-0.230	1	0.072	1	-0.041	1	-0.041	1
54	0.770	1	0.724	1	0.038	1	0.035	1
55	0.860	1	0.514	1	0.051	1	0.054	1
56	-1.200	1	-1.226	1	-0.095	2	-0.093	2
57	0.420	1	0.607	1	0.036	1	0.035	1
58	1.340	1	0.721	1	0.046	1	0.044	1
59	1.550	2	2.061	2	0.090	1	0.088	1
60	1.040	1	1.528	1	0.092	2	0.095	2
61	-0.210	1	-0.255	1	-0.039	1	-0.042	1
62	0.540	1	1.392	1	0.002	1	0.004	1
63	-0.480	1	-0.479	1	-0.052	1	-0.055	1
64	0.670	1	0.979	1	0.013	1	0.013	1
65	0.060	1	0.261	1	0.019	1	0.013	1
66	-0.720	1	-0.613	1	-0.076	1	-0.076	1
67	-1.780	2	-0.687	1	-0.111	2	-0.11	2
68	1.880	1	1.540	1	0.058	1	0.057	1
69	-0.750	1	-0.832	1	-0.063	1	-0.061	1
70	0.150	1	0.280	1	-0.012	1	-0.012	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 29.

MH/GMH and SIB/Poly-SIB Comparisons: Mathematics January 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	-0.250	1	-0.253	1	-0.017	1	-0.017	1
2	-1.350	2	-1.372	2	-0.046	1	-0.046	1
3	-0.230	1	-0.232	1	-0.014	1	-0.014	1
4	-0.560	1	-0.561	1	-0.050	1	-0.05	1
5	0.290	1	0.286	1	0.027	1	0.027	1
6	-0.090	1	-0.095	1	-0.008	1	-0.008	1
7	0.050	1	0.050	1	0.011	1	0.011	1
8	0.130	1	0.134	1	0.015	1	0.015	1
9	0.050	1	0.053	1	0.006	1	0.005	1
10	-0.640	1	-0.635	1	-0.056	1	-0.055	1
11	0.490	1	0.486	1	0.030	1	0.03	1
12	-0.090	1	-0.094	1	-0.007	1	-0.007	1
13	-0.680	1	-0.678	1	-0.060	2	-0.06	2
14	0.740	1	0.723	1	0.025	1	0.025	1
15	0.200	1	0.197	1	0.022	1	0.022	1
16	-2.080	2	-2.164	2	-0.042	1	-0.042	1
17	-0.770	1	-0.778	1	-0.042	1	-0.042	1
18	-0.290	1	-0.286	1	-0.025	1	-0.025	1
19	0.430	1	0.427	1	0.029	1	0.029	1
20	1.500	2	1.445	2	0.103	2	0.103	2
21	-0.460	1	-0.463	1	-0.031	1	-0.031	1
22	-0.010	1	-0.013	1	0.002	1	0.002	1
23	1.470	2	1.405	2	0.086	2	0.086	2
24	-0.080	1	-0.084	1	-0.003	1	-0.003	1
25	0.250	1	0.248	1	0.006	1	0.006	1
26	0.010	1	0.007	1	0.004	1	0.004	1
27	-0.330	1	-0.333	1	-0.026	1	-0.025	1
28	0.390	1	0.384	1	0.019	1	0.019	1
29	-0.420	1	-0.419	1	-0.036	1	-0.036	1
30	0.390	1	0.384	1	0.024	1	0.025	1
31	0.110	1	0.112	1	0.008	1	0.008	1
32	0.860	1	0.855	1	0.083	2	0.083	2
33	-0.340	1	-0.336	1	-0.026	1	-0.026	1
34	-0.280	1	-0.286	1	-0.014	1	-0.014	1
35	0.590	1	0.577	1	0.017	1	0.017	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
36	-0.130	1	-0.127	1	-0.005	1	-0.005	1
37	-0.370	1	-0.372	1	-0.032	1	-0.032	1
38	0.350	1	0.350	1	0.021	1	0.021	1
39	0.040	1	0.042	1	-0.001	1	0	1
40	0.840	1	0.834	1	0.075	2	0.075	2
41	-1.220	2	-1.207	2	-0.087	2	-0.087	2
42	-0.710	1	-0.713	1	-0.052	1	-0.052	1
43	0.140	1	0.144	1	0.008	1	0.008	1
44	0.300	1	0.301	1	0.021	1	0.021	1
45	0.280	1	0.284	1	0.017	1	0.018	1
46	0.330	1	0.332	1	0.016	1	0.016	1
47	-0.560	1	-0.561	1	-0.050	1	-0.05	1
48	0.330	1	0.329	1	0.017	1	0.017	1
49	0.210	1	0.213	1	0.008	1	0.009	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 30.

MH/GMH and SIB/Poly-SIB Comparisons: Mathematics June 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	-0.170	1	-0.166	1	-0.014	1	-0.014	1
2	-1.080	2	-1.089	2	-0.092	2	-0.092	2
3	-0.340	1	-0.338	1	-0.038	1	-0.039	1
4	0.970	1	0.947	1	0.072	2	0.072	2
5	-0.290	1	-0.292	1	-0.027	1	-0.027	1
6	0.450	1	0.447	1	0.039	1	0.039	1
7	-0.630	1	-0.643	1	-0.054	1	-0.054	1
8	-0.040	1	-0.039	1	-0.004	1	-0.005	1
9	-0.300	1	-0.300	1	-0.030	1	-0.029	1
10	-0.330	1	-0.327	1	-0.036	1	-0.036	1
11	0.020	1	0.018	1	0.006	1	0.006	1
12	0.180	1	0.184	1	0.004	1	0.004	1
13	0.430	1	0.425	1	0.027	1	0.027	1
14	-0.670	1	-0.683	1	-0.037	1	-0.038	1
15	0.390	1	0.386	1	0.040	1	0.04	1
16	-0.590	1	-0.595	1	-0.040	1	-0.041	1
17	0.530	1	0.524	1	0.041	1	0.041	1
18	-0.130	1	-0.128	1	-0.012	1	-0.013	1
19	-0.300	1	-0.304	1	-0.016	1	-0.016	1
20	0.030	1	0.031	1	0.004	1	0.004	1
21	0.390	1	0.384	1	0.017	1	0.017	1
22	0.040	1	0.038	1	0.001	1	0.001	1
23	0.450	1	0.445	1	0.038	1	0.038	1
24	0.280	1	0.279	1	0.021	1	0.021	1
25	0.690	1	0.681	1	0.038	1	0.038	1
26	-0.380	1	-0.381	1	-0.038	1	-0.038	1
27	-0.200	1	-0.199	1	-0.015	1	-0.016	1
28	-0.100	1	-0.104	1	-0.007	1	-0.007	1
29	-0.350	1	-0.346	1	-0.032	1	-0.032	1
30	-0.070	1	-0.071	1	-0.010	1	-0.01	1
31	-0.550	1	-0.549	1	-0.049	1	-0.05	1
32	2.040	2	1.978	2	0.183	2	0.183	2
33	1.110	2	1.082	2	0.067	2	0.067	2
34	-0.670	1	-0.674	1	-0.037	1	-0.038	1
35	-0.420	1	-0.422	1	-0.034	1	-0.034	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
36	1.010	2	0.980	1	0.072	2	0.072	2
37	0.130	1	0.128	1	0.012	1	0.011	1
38	-0.580	1	-0.591	1	-0.035	1	-0.035	1
39	0.820	1	0.816	1	0.059	2	0.058	1
40	-0.120	1	-0.116	1	-0.008	1	-0.009	1
41	0.230	1	0.230	1	0.010	1	0.009	1
42	-0.130	1	-0.127	1	-0.014	1	-0.014	1
43	-0.460	1	-0.459	1	-0.043	1	-0.043	1
44	-0.270	1	-0.273	1	-0.019	1	-0.02	1
45	0.480	1	0.483	1	0.031	1	0.031	1
46	-0.960	1	-0.952	1	-0.070	2	-0.071	2
47	0.050	1	0.051	1	0.003	1	0.003	1
48	-0.800	1	-0.801	1	-0.073	2	-0.073	2
49	0.300	1	0.302	1	0.023	1	0.023	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 31.

MH/GMH and SIB/Poly-SIB Comparisons: Mathematics 1212 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	0.090	1	0.092	1	0.016	1	0.014	1
2	-2.540	2	-2.443	2	-0.093	2	-0.093	2
3	-0.290	1	-0.293	1	-0.012	1	-0.011	1
4	-0.820	1	-0.825	1	-0.080	2	-0.08	2
5	0.390	1	0.392	1	0.053	1	0.053	1
6	-0.350	1	-0.342	1	-0.035	1	-0.032	1
7	-0.540	1	-0.547	1	-0.048	1	-0.048	1
8	0.040	1	0.036	1	0.018	1	0.023	1
9	0.890	1	0.903	1	0.082	2	0.085	2
10	-0.360	1	-0.368	1	-0.025	1	-0.025	1
11	-0.240	1	-0.237	1	-0.012	1	-0.007	1
12	-0.640	1	-0.642	1	-0.045	1	-0.04	1
13	-0.380	1	-0.373	1	-0.020	1	-0.02	1
14	1.290	1	1.212	1	0.036	1	0.036	1
15	0.370	1	0.362	1	0.024	1	0.019	1
16	-2.420	2	-2.296	2	-0.042	1	-0.041	1
17	-1.440	2	-1.448	2	-0.075	2	-0.07	2
18	-0.630	1	-0.609	1	-0.052	1	-0.049	1
19	0.790	1	0.759	1	0.062	2	0.064	2
20	1.770	2	1.709	2	0.123	2	0.124	2
21	-0.510	1	-0.516	1	-0.039	1	-0.04	1
22	0.050	1	0.045	1	0.003	1	0.004	1
23	1.530	2	1.465	2	0.065	2	0.069	2
24	-0.130	1	-0.132	1	-0.003	1	-0.004	1
25	0.310	1	0.301	1	-0.017	1	-0.013	1
26	-0.560	1	-0.558	1	-0.020	1	-0.02	1
27	-0.820	1	-0.801	1	-0.045	1	-0.041	1
28	0.850	1	0.838	1	0.043	1	0.044	1
29	-0.440	1	-0.440	1	-0.033	1	-0.034	1
30	1.190	2	1.181	2	0.081	2	0.081	2
31	0.930	1	0.953	1	0.077	2	0.076	2
32	0.890	1	0.875	1	0.084	2	0.088	2
33	-0.550	1	-0.539	1	-0.047	1	-0.042	1
34	-0.270	1	-0.277	1	-0.029	1	-0.027	1
35	0.330	1	0.329	1	0.041	1	0.044	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
36	-0.250	1	-0.245	1	-0.010	1	-0.013	1
37	-0.230	1	-0.229	1	-0.019	1	-0.021	1
38	0.360	1	0.356	1	0.019	1	0.021	1
39	-0.130	1	-0.129	1	-0.027	1	-0.026	1
40	1.190	2	1.221	2	0.135	2	0.134	2
41	-1.770	2	-1.679	2	-0.109	2	-0.108	2
42	-0.830	1	-0.823	1	-0.049	1	-0.044	1
43	0.040	1	0.036	1	-0.011	1	-0.006	1
44	0.470	1	0.465	1	0.028	1	0.036	1
45	0.120	1	0.123	1	-0.001	1	0.004	1
46	0.830	1	0.828	1	0.034	1	0.043	1
47	-0.500	1	-0.496	1	-0.047	1	-0.047	1
48	0.140	1	0.138	1	0.012	1	0.014	1
49	0.410	1	0.409	1	0.001	1	-0.002	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 32.

MH/GMH and SIB/Poly-SIB Comparisons: Mathematics 2112 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	-0.860	1	-0.863	1	-0.062	1	-0.061	1
2	-0.820	1	-0.810	1	-0.011	1	-0.012	1
3	-0.620	1	-0.614	1	-0.024	1	-0.022	1
4	-0.820	1	-0.802	1	-0.066	1	-0.066	1
5	0.470	1	0.467	1	0.035	1	0.033	1
6	-0.050	1	-0.051	1	-0.005	1	-0.007	1
7	0.270	1	0.263	1	0.044	1	0.045	1
8	0.160	1	0.157	1	0.017	1	0.015	1
9	-0.190	1	-0.196	1	-0.025	1	-0.023	1
10	-1.010	2	-1.007	2	-0.093	2	-0.092	2
11	0.760	1	0.745	1	0.046	1	0.047	1
12	0.000	1	0.005	1	-0.007	1	-0.006	1
13	-0.250	1	-0.252	1	-0.021	1	-0.022	1
14	0.800	1	0.791	1	0.051	1	0.055	1
15	0.220	1	0.221	1	0.022	1	0.018	1
16	-3.060	2	-3.213	2	-0.054	1	-0.054	1
17	-0.310	1	-0.315	1	-0.015	1	-0.013	1
18	0.650	1	0.638	1	0.053	1	0.052	1
19	-0.360	1	-0.354	1	-0.023	1	-0.022	1
20	1.000	1	0.987	1	0.071	2	0.071	2
21	0.580	1	0.588	1	0.048	1	0.046	1
22	0.020	1	0.016	1	0.026	1	0.026	1
23	1.340	2	1.266	2	0.085	2	0.085	2
24	0.650	1	0.650	1	0.044	1	0.045	1
25	-0.080	1	-0.081	1	0.008	1	0.008	1
26	0.200	1	0.196	1	0.019	1	0.016	1
27	-0.460	1	-0.464	1	-0.017	1	-0.022	1
28	-0.630	1	-0.641	1	-0.028	1	-0.03	1
29	-0.440	1	-0.445	1	-0.054	1	-0.054	1
30	0.060	1	0.055	1	0.024	1	0.025	1
31	-0.050	1	-0.047	1	0.000	1	0.001	1
32	0.650	1	0.638	1	0.074	1	0.073	1
33	-0.220	1	-0.212	1	0.013	1	0.015	1
34	0.330	1	0.343	1	0.016	1	0.019	1
35	1.260	1	1.210	1	0.044	1	0.045	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
36	-0.190	1	-0.200	1	-0.024	1	-0.02	1
37	-1.010	2	-1.019	2	-0.072	2	-0.075	2
38	0.500	1	0.499	1	0.022	1	0.023	1
39	0.890	1	0.895	1	0.072	2	0.07	2
40	0.690	1	0.677	1	0.065	1	0.064	1
41	-0.980	1	-0.953	1	-0.089	2	-0.086	2
42	-0.360	1	-0.361	1	-0.025	1	-0.022	1
43	0.030	1	0.031	1	0.000	1	0.002	1
44	0.340	1	0.345	1	0.021	1	0.024	1
45	0.460	1	0.449	1	0.030	1	0.029	1
46	-0.050	1	-0.056	1	-0.006	1	-0.006	1
47	-0.480	1	-0.491	1	-0.041	1	-0.042	1
48	-0.190	1	-0.194	1	0.004	1	0.004	1
49	0.240	1	0.232	1	0.022	1	0.028	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 33.

MH/GMH and SIB/Poly-SIB Comparisons: Mathematics 2211 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	0.100	1	0.105	1	0.001	1	0	1
2	-0.460	1	-0.454	1	-0.011	1	-0.012	1
3	0.010	1	0.009	1	-0.002	1	-0.002	1
4	-0.520	1	-0.522	1	-0.052	1	-0.053	1
5	0.600	1	0.610	1	0.077	1	0.078	1
6	0.490	1	0.487	1	0.021	1	0.02	1
7	0.160	1	0.160	1	-0.001	1	-0.001	1
8	-0.200	1	-0.195	1	0.002	1	0	1
9	-0.280	1	-0.275	1	-0.006	1	-0.008	1
10	-0.640	1	-0.639	1	-0.058	1	-0.059	1
11	0.020	1	0.024	1	0.012	1	0.011	1
12	-0.380	1	-0.376	1	-0.039	1	-0.04	1
13	-1.720	2	-1.696	2	-0.135	2	-0.133	2
14	0.510	1	0.500	1	0.012	1	0.012	1
15	0.050	1	0.048	1	0.006	1	0.004	1
16	0.340	1	0.310	1	-0.004	1	0.002	1
17	-0.780	1	-0.782	1	-0.037	1	-0.036	1
18	-0.490	1	-0.468	1	-0.031	1	-0.032	1
19	0.540	1	0.532	1	0.031	1	0.03	1
20	1.950	2	1.816	2	0.135	2	0.134	2
21	-1.440	2	-1.450	2	-0.096	2	-0.095	2
22	-0.170	1	-0.176	1	-0.016	1	-0.018	1
23	1.260	2	1.220	2	0.066	2	0.064	2
24	-0.820	1	-0.832	1	-0.068	1	-0.069	1
25	1.840	1	1.555	1	0.030	1	0.031	1
26	-0.380	1	-0.372	1	-0.042	1	-0.043	1
27	-0.360	1	-0.357	1	-0.027	1	-0.031	1
28	-0.160	1	-0.159	1	0.009	1	0.01	1
29	-0.270	1	-0.271	1	0.001	1	-0.002	1
30	0.610	1	0.597	1	0.039	1	0.038	1
31	0.230	1	0.238	1	0.026	1	0.024	1
32	1.100	2	1.056	2	0.091	2	0.089	2
33	-1.310	2	-1.368	2	-0.094	2	-0.094	2
34	-0.840	1	-0.877	1	-0.040	1	-0.041	1
35	0.710	1	0.699	1	0.006	1	0.005	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
36	0.310	1	0.315	1	0.040	1	0.038	1
37	0.520	1	0.515	1	0.043	1	0.042	1
38	0.530	1	0.524	1	0.041	1	0.04	1
39	0.220	1	0.213	1	0.016	1	0.016	1
40	0.550	1	0.540	1	0.057	1	0.054	1
41	0.000	1	0.000	1	0.010	1	0.009	1
42	-1.030	1	-1.008	1	-0.075	2	-0.073	2
43	0.440	1	0.433	1	0.043	1	0.041	1
44	0.320	1	0.307	1	0.036	1	0.034	1
45	0.340	1	0.325	1	0.022	1	0.02	1
46	0.730	1	0.730	1	0.047	1	0.048	1
47	-0.750	1	-0.762	1	-0.067	1	-0.068	1
48	0.500	1	0.490	1	0.018	1	0.019	1
49	0.460	1	0.442	1	0.020	1	0.018	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 34.

MH/GMH and SIB/Poly-SIB Comparisons: Mathematics 2222 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	0.500	1	0.499	1	0.051	1	0.049	1
2	-1.130	2	-1.161	2	-0.092	1	-0.095	1
3	-0.900	1	-0.896	1	-0.102	2	-0.099	2
4	0.700	1	0.697	1	0.056	1	0.054	1
5	-0.600	1	-0.632	1	-0.069	1	-0.067	1
6	0.840	1	0.840	1	0.098	1	0.1	1
7	-0.220	1	-0.224	1	0.001	1	0.003	1
8	0.360	1	0.361	1	0.062	1	0.061	1
9	-0.430	1	-0.425	1	-0.011	1	-0.008	1
10	-0.380	1	-0.384	1	-0.013	1	-0.009	1
11	-0.390	1	-0.382	1	-0.029	1	-0.023	1
12	-0.190	1	-0.210	1	0.004	1	0.001	1
13	0.580	1	0.586	1	0.057	1	0.058	1
14	-0.700	1	-0.710	1	-0.017	1	-0.019	1
15	1.500	2	1.504	2	0.153	2	0.158	2
16	-0.210	1	-0.208	1	0.015	1	0.014	1
17	0.590	1	0.576	1	0.038	1	0.035	1
18	0.140	1	0.131	1	-0.011	1	-0.013	1
19	0.400	1	0.388	1	0.021	1	0.021	1
20	-0.330	1	-0.323	1	-0.033	1	-0.034	1
21	0.990	1	0.951	1	0.032	1	0.034	1
22	0.590	1	0.572	1	0.044	1	0.051	1
23	0.380	1	0.382	1	0.025	1	0.027	1
24	-0.190	1	-0.196	1	-0.007	1	-0.005	1
25	0.160	1	0.157	1	0.030	1	0.026	1
26	-0.130	1	-0.129	1	-0.003	1	0	1
27	0.150	1	0.153	1	0.013	1	0.013	1
28	-0.760	1	-0.738	1	-0.072	1	-0.066	1
29	-0.620	1	-0.604	1	-0.072	1	-0.064	1
30	0.380	1	0.372	1	0.026	1	0.024	1
31	-0.550	1	-0.532	1	-0.024	1	-0.036	1
32	1.540	2	1.472	2	0.166	2	0.172	2
33	-0.280	1	-0.288	1	-0.018	1	-0.018	1
34	-1.190	1	-1.141	1	-0.023	1	-0.03	1
35	-0.850	1	-0.847	1	-0.069	1	-0.073	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
36	1.970	2	1.833	2	0.138	2	0.125	2
37	0.340	1	0.334	1	0.042	1	0.044	1
38	-0.780	1	-0.789	1	-0.018	1	-0.016	1
39	0.020	1	0.017	1	0.005	1	0.007	1
40	-0.790	1	-0.810	1	-0.028	1	-0.031	1
41	0.810	1	0.770	1	0.034	1	0.027	1
42	-0.370	1	-0.351	1	-0.032	1	-0.026	1
43	-0.170	1	-0.174	1	-0.023	1	-0.022	1
44	0.570	1	0.584	1	0.036	1	0.034	1
45	0.300	1	0.299	1	0.017	1	0.026	1
46	-0.400	1	-0.396	1	-0.016	1	-0.024	1
47	0.520	1	0.498	1	0.026	1	0.026	1
48	-1.200	2	-1.195	2	-0.099	1	-0.097	1
49	-0.140	1	-0.136	1	-0.001	1	0.001	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 35.

MH/GMH and SIB/Poly-SIB Comparisons: Biology January 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	0.520	1	0.525	1	0.035	1	0.036	1
2	0.330	1	0.404	1	0.026	1	0.026	1
3	-0.240	1	-0.177	1	-0.013	1	-0.013	1
4	-0.580	1	-0.537	1	-0.031	1	-0.031	1
5	0.000	1	0.020	1	-0.001	1	-0.001	1
6	0.240	1	0.293	1	0.018	1	0.018	1
7	-0.220	1	-0.198	1	-0.020	1	-0.02	1
8	0.070	1	0.105	1	0.002	1	0.003	1
9	0.210	1	0.179	1	0.008	1	0.008	1
10	0.650	1	0.669	1	0.034	1	0.034	1
11	-0.170	1	-0.193	1	-0.008	1	-0.008	1
12	0.610	1	0.647	1	0.034	1	0.034	1
13	-0.150	1	-0.157	1	-0.017	1	-0.017	1
14	-1.150	2	-1.164	2	-0.054	1	-0.054	1
15	-0.360	1	-0.298	1	-0.024	1	-0.024	1
16	-0.720	1	-0.725	1	-0.046	1	-0.046	1
17	-0.380	1	-0.375	1	-0.034	1	-0.034	1
18	0.600	1	0.600	1	0.020	1	0.02	1
19	1.030	2	1.091	2	0.055	1	0.055	1
20	-0.290	1	-0.236	1	-0.004	1	-0.004	1
21	-0.360	1	-0.321	1	-0.020	1	-0.02	1
22	0.270	1	0.390	1	0.011	1	0.011	1
23	0.810	1	0.806	1	0.074	2	0.074	2
24	-0.120	1	-0.112	1	-0.011	1	-0.011	1
25	-0.730	1	-0.698	1	-0.038	1	-0.038	1
26	-0.280	1	-0.175	1	-0.010	1	-0.01	1
27	1.100	2	1.056	2	0.078	2	0.078	2
28	-0.080	1	-0.119	1	-0.007	1	-0.007	1
29	-1.120	2	-1.164	2	-0.076	2	-0.076	2
30	-0.340	1	-0.283	1	-0.019	1	-0.019	1
31	0.610	1	0.652	1	0.041	1	0.041	1
32	-0.040	1	0.150	1	0.004	1	0.004	1
33	0.910	1	0.921	1	0.055	1	0.055	1
34	-0.190	1	-0.174	1	-0.010	1	-0.01	1
35	0.720	1	0.756	1	0.058	1	0.058	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
36	-0.690	1	-0.659	1	-0.067	2	-0.067	2
37	0.880	1	0.885	1	0.039	1	0.039	1
38	-0.790	1	-0.749	1	-0.072	2	-0.072	2
39	-0.090	1	-0.090	1	-0.013	1	-0.013	1
40	-0.510	1	-0.367	1	-0.023	1	-0.023	1
41	-0.040	1	-0.017	1	-0.005	1	-0.005	1
42	0.460	1	0.540	1	0.026	1	0.026	1
43	0.200	1	0.176	1	0.017	1	0.017	1
44	-0.770	1	-0.798	1	-0.070	2	-0.07	2
45	-0.130	1	-0.066	1	-0.008	1	-0.008	1
46	0.220	1	0.232	1	0.023	1	0.023	1
47	0.340	1	0.255	1	0.008	1	0.008	1
48	-0.590	1	-0.561	1	-0.058	1	-0.058	1
49	0.980	1	0.932	1	0.080	2	0.08	2
50	0.210	1	0.223	1	0.022	1	0.022	1
51	0.410	1	0.376	1	0.032	1	0.032	1
52	-0.230	1	-0.232	1	-0.021	1	-0.02	1
53	-0.510	1	-0.451	1	-0.036	1	-0.036	1
54	-0.390	1	-0.304	1	-0.032	1	-0.032	1
55	0.070	1	0.056	1	0.002	1	0.002	1
56	0.730	1	0.691	1	0.047	1	0.048	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 36.

MH/GMH and SIB/Poly-SIB Comparisons: Biology June 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	-0.050	1	-0.050	1	0.001	1	0	1
2	0.850	1	0.821	1	0.027	1	0.027	1
3	-0.180	1	-0.182	1	-0.013	1	-0.013	1
4	-0.430	1	-0.427	1	-0.027	1	-0.027	1
5	0.620	1	0.616	1	0.054	1	0.054	1
6	-0.120	1	-0.118	1	-0.008	1	-0.008	1
7	0.830	1	0.820	1	0.064	2	0.064	2
8	0.260	1	0.255	1	0.021	1	0.021	1
9	0.870	1	0.849	1	0.035	1	0.035	1
10	-0.640	1	-0.652	1	-0.045	1	-0.045	1
11	-0.800	1	-0.806	1	-0.047	1	-0.047	1
12	-0.250	1	-0.254	1	-0.008	1	-0.008	1
13	-0.280	1	-0.283	1	-0.020	1	-0.02	1
14	-0.450	1	-0.451	1	-0.035	1	-0.035	1
15	-0.010	1	-0.010	1	-0.004	1	-0.004	1
16	-0.690	1	-0.694	1	-0.050	1	-0.05	1
17	0.600	1	0.596	1	0.044	1	0.044	1
18	-0.260	1	-0.259	1	-0.023	1	-0.023	1
19	0.460	1	0.465	1	0.039	1	0.038	1
20	-0.660	1	-0.662	1	-0.045	1	-0.045	1
21	0.390	1	0.387	1	0.024	1	0.024	1
22	-1.050	2	-1.065	2	-0.072	2	-0.072	2
23	0.650	1	0.648	1	0.038	1	0.037	1
24	0.420	1	0.409	1	0.025	1	0.025	1
25	-0.310	1	-0.308	1	-0.024	1	-0.024	1
26	0.110	1	0.112	1	0.015	1	0.015	1
27	0.460	1	0.455	1	0.021	1	0.021	1
28	0.010	1	0.009	1	-0.002	1	-0.002	1
29	-0.750	1	-0.768	1	-0.035	1	-0.035	1
30	-1.060	2	-1.066	2	-0.052	1	-0.053	1
31	-0.170	1	-0.175	1	-0.004	1	-0.004	1
32	-1.100	2	-1.114	2	-0.037	1	-0.037	1
33	0.080	1	0.075	1	0.006	1	0.006	1
34	0.480	1	0.481	1	0.033	1	0.033	1
35	0.350	1	0.340	1	0.022	1	0.022	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
36	0.260	1	0.262	1	0.012	1	0.012	1
37	-0.750	1	-0.748	1	-0.042	1	-0.042	1
38	-0.650	1	-0.660	1	-0.026	1	-0.027	1
39	-0.290	1	-0.292	1	-0.016	1	-0.016	1
40	0.210	1	0.208	1	0.010	1	0.01	1
41	0.670	1	0.665	1	0.054	1	0.054	1
42	-0.920	1	-0.929	1	-0.074	2	-0.075	2
43	-0.060	1	-0.061	1	-0.006	1	-0.006	1
44	0.550	1	0.553	1	0.055	1	0.055	1
45	0.300	1	0.295	1	0.024	1	0.024	1
46	0.480	1	0.482	1	0.009	1	0.009	1
47	0.690	1	0.676	1	0.015	1	0.015	1
48	-0.100	1	-0.097	1	-0.008	1	-0.008	1
49	0.290	1	0.290	1	0.028	1	0.027	1
50	0.660	1	0.009	1	0.032	1	0.032	1
51	0.780	1	0.000	1	0.056	1	0.056	1
52	0.200	1	0.321	1	0.018	1	0.018	1
53	0.070	1	0.750	1	0.010	1	0.009	1
54	-0.690	1	0.001	1	-0.047	1	-0.048	1
55	-0.360	1	0.082	1	-0.025	1	-0.025	1
56	-0.310	1	0.219	1	-0.020	1	-0.02	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 37.

MH/GMH and SIB/Poly-SIB Comparisons: Biology 1212 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	0.170	1	0.169	1	0.023	1	0.022	1
2	1.400	1	1.323	1	0.029	1	0.029	1
3	-0.310	1	-0.319	1	-0.032	1	-0.03	1
4	0.640	1	0.627	1	0.039	1	0.039	1
5	0.980	1	0.961	1	0.084	2	0.081	2
6	-0.230	1	-0.218	1	-0.014	1	-0.02	1
7	0.500	1	0.486	1	0.018	1	0.019	1
8	0.210	1	0.210	1	0.011	1	0.011	1
9	1.560	2	1.506	2	0.065	2	0.063	2
10	0.070	1	0.067	1	0.016	1	0.014	1
11	-0.200	1	-0.195	1	0.007	1	0.002	1
12	-0.780	1	-0.778	1	-0.020	1	-0.02	1
13	-0.190	1	-0.193	1	-0.001	1	-0	1
14	-1.250	2	-1.257	2	-0.107	2	-0.11	2
15	-0.140	1	-0.138	1	-0.015	1	-0.02	1
16	-1.230	2	-1.246	2	-0.093	2	-0.09	2
17	0.520	1	0.516	1	0.022	1	0.022	1
18	-0.480	1	-0.484	1	-0.060	1	-0.06	1
19	0.710	1	0.697	1	0.056	1	0.057	1
20	-0.240	1	-0.239	1	-0.025	1	-0.03	1
21	0.610	1	0.596	1	0.020	1	0.016	1
22	-1.470	2	-1.486	2	-0.087	2	-0.09	2
23	0.720	1	0.692	1	0.052	1	0.049	1
24	0.680	1	0.683	1	0.036	1	0.029	1
25	0.390	1	0.397	1	0.036	1	0.03	1
26	-0.240	1	-0.235	1	-0.001	1	-0	1
27	-0.150	1	-0.141	1	-0.003	1	-0	1
28	0.030	1	0.030	1	-0.010	1	-0.01	1
29	-1.090	1	-1.122	1	-0.049	1	-0.05	1
30	-0.940	1	-0.938	1	-0.042	1	-0.04	1
31	-0.840	1	-0.905	1	-0.032	1	-0.03	1
32	-0.210	1	-0.218	1	-0.015	1	-0.02	1
33	-0.350	1	-0.355	1	-0.036	1	-0.04	1
34	0.500	1	0.499	1	0.048	1	0.048	1
35	0.250	1	0.255	1	0.005	1	0.006	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
36	-0.160	1	-0.161	1	-0.004	1	-0	1
37	-0.550	1	-0.537	1	-0.024	1	-0.02	1
38	-1.130	1	-1.135	1	-0.047	1	-0.05	1
39	-0.560	1	-0.561	1	-0.015	1	-0.02	1
40	0.430	1	0.417	1	0.023	1	0.023	1
41	0.370	1	0.363	1	0.035	1	0.033	1
42	-0.930	1	-0.910	1	-0.077	2	-0.08	2
43	-0.360	1	-0.359	1	-0.019	1	-0.02	1
44	0.260	1	0.256	1	0.029	1	0.027	1
45	1.010	2	0.972	1	0.076	2	0.077	2
46	-1.340	1	-1.251	1	-0.017	1	-0.02	1
47	0.340	1	0.318	1	0.003	1	0.003	1
48	0.310	1	0.302	1	-0.001	1	-0	1
49	0.330	1	0.335	1	0.029	1	0.03	1
50	0.640	1	0.613	1	0.020	1	0.018	1
51	0.770	1	0.784	1	0.055	1	0.057	1
52	0.780	1	0.793	1	0.058	1	0.06	1
53	-0.320	1	-0.326	1	-0.034	1	-0.04	1
54	-0.390	1	-0.395	1	-0.028	1	-0.03	1
55	-0.460	1	-0.463	1	-0.042	1	-0.04	1
56	-0.730	1	-0.739	1	-0.019	1	-0.02	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 38.

MH/GMH and SIB/Poly-SIB Comparisons: Biology 2112 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	0.050	1	-0.173	1	0.011	1	0.017	1
2	1.250	1	1.201	1	0.038	1	0.043	1
3	-0.220	1	-0.069	1	-0.002	1	0.003	1
4	-2.250	2	-2.315	2	-0.112	2	-0.115	2
5	0.120	1	0.523	1	0.038	1	0.043	1
6	-0.400	1	-0.622	1	-0.043	1	-0.05	1
7	1.260	2	1.323	2	0.096	2	0.095	2
8	0.350	1	0.397	1	0.028	1	0.019	1
9	0.760	1	0.732	1	0.034	1	0.042	1
10	-0.940	1	-0.789	1	-0.062	1	-0.061	1
11	-1.370	2	-0.878	1	-0.036	1	-0.028	1
12	0.700	1	0.770	1	0.024	1	0.029	1
13	-0.380	1	-0.407	1	-0.026	1	-0.021	1
14	-0.570	1	-0.627	1	-0.034	1	-0.037	1
15	-0.980	1	-1.054	1	-0.063	2	-0.057	1
16	-0.430	1	-0.414	1	-0.027	1	-0.02	1
17	0.810	1	0.648	1	0.049	1	0.055	1
18	-0.350	1	-0.265	1	-0.027	1	-0.031	1
19	0.390	1	0.345	1	0.026	1	0.027	1
20	-0.650	1	-0.527	1	-0.023	1	-0.025	1
21	0.910	1	0.746	1	0.037	1	0.043	1
22	0.000	1	-0.048	1	-0.022	1	-0.024	1
23	1.030	1	1.198	1	0.049	1	0.057	1
24	0.030	1	-0.048	1	0.014	1	0.019	1
25	-0.060	1	-0.167	1	-0.004	1	0.002	1
26	0.200	1	0.007	1	0.030	1	0.022	1
27	-0.120	1	-0.372	1	-0.008	1	-0.006	1
28	-0.130	1	-0.130	1	-0.015	1	-0.009	1
29	-1.310	1	-1.524	1	-0.050	1	-0.047	1
30	-1.220	1	-0.735	1	-0.057	1	-0.052	1
31	-0.990	1	-1.742	1	-0.013	1	-0.014	1
32	-1.080	1	-1.120	1	-0.034	1	-0.031	1
33	0.160	1	0.132	1	0.026	1	0.023	1
34	0.380	1	0.417	1	0.033	1	0.038	1
35	0.940	1	1.067	1	0.056	1	0.056	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
36	0.210	1	0.325	1	0.010	1	0.013	1
37	-0.760	1	-0.769	1	-0.044	1	-0.041	1
38	-0.700	1	-0.770	1	-0.031	1	-0.027	1
39	-0.380	1	-0.592	1	-0.029	1	-0.023	1
40	0.900	1	0.767	1	0.046	1	0.044	1
41	1.460	2	1.196	2	0.080	2	0.081	2
42	-1.730	2	-1.681	2	-0.133	2	-0.132	2
43	-0.020	1	-0.029	1	0.013	1	0.009	1
44	0.290	1	0.345	1	0.031	1	0.034	1
45	0.190	1	0.604	1	0.037	1	0.033	1
46	1.500	1	1.445	1	0.035	1	0.044	1
47	-0.310	1	-0.226	1	-0.003	1	-0.002	1
48	0.520	1	0.424	1	0.049	1	0.051	1
49	0.460	1	0.208	1	0.051	1	0.043	1
50	1.300	1	0.950	1	0.049	1	0.057	1
51	0.660	1	0.593	1	0.043	1	0.042	1
52	0.150	1	-0.045	1	0.016	1	0.011	1
53	-0.080	1	-0.170	1	-0.011	1	-0.011	1
54	-1.550	2	-1.553	2	-0.105	2	-0.102	2
55	-0.440	1	-0.426	1	-0.031	1	-0.018	1
56	0.600	1	0.083	1	0.004	1	0.004	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 39.

MH/GMH and SIB/Poly-SIB Comparisons: Biology 2211 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	0.550	1	0.622	1	0.040	1	0.045	1
2	0.880	1	0.932	1	0.078	2	0.089	2
3	-1.150	1	-1.028	1	-0.017	1	-0.01	1
4	0.190	1	0.155	1	0.004	1	0.004	1
5	0.670	1	0.677	1	0.063	1	0.066	2
6	0.420	1	0.515	1	0.040	1	0.042	1
7	-0.280	1	-0.263	1	-0.003	1	-0.001	1
8	0.420	1	0.459	1	0.044	1	0.046	1
9	-0.480	1	-0.529	1	-0.030	1	-0.031	1
10	1.180	2	1.299	2	0.066	1	0.069	2
11	-0.580	1	-0.723	1	-0.019	1	-0.019	1
12	2.080	2	2.011	2	0.093	2	0.094	2
13	1.090	2	0.915	1	0.103	2	0.105	2
14	-0.930	1	-1.005	1	-0.034	1	-0.033	1
15	-0.170	1	-0.171	1	0.014	1	0.012	1
16	-1.410	2	-1.640	2	-0.088	2	-0.087	2
17	0.020	1	-0.192	1	0.009	1	0.019	1
18	-0.060	1	-0.083	1	-0.010	1	-0.011	1
19	0.710	1	0.593	1	0.033	1	0.034	1
20	0.800	1	0.164	1	0.017	1	0.019	1
21	-0.830	1	-0.889	1	-0.048	1	-0.047	1
22	0.630	1	0.418	1	0.026	1	0.027	1
23	0.550	1	0.491	1	0.048	1	0.047	1
24	0.010	1	-0.054	1	0.015	1	0.017	1
25	-1.070	1	-0.605	1	-0.030	1	-0.026	1
26	0.420	1	0.446	1	0.028	1	0.029	1
27	1.760	2	1.454	2	0.082	2	0.086	2
28	0.270	1	0.508	1	0.021	1	0.021	1
29	-2.330	2	-2.313	2	-0.121	2	-0.117	2
30	-0.980	1	-0.576	1	-0.028	1	-0.026	1
31	0.990	1	0.724	1	0.056	1	0.06	1
32	-0.830	1	-0.424	1	-0.017	1	-0.013	1
33	1.490	2	1.321	2	0.080	2	0.081	2
34	-0.480	1	-0.607	1	-0.011	1	-0.008	1
35	0.760	1	0.287	1	0.048	1	0.056	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
36	-1.230	2	-1.120	2	-0.105	2	-0.105	2
37	1.570	2	1.618	2	0.074	2	0.077	2
38	-1.350	2	-1.346	2	-0.127	2	-0.122	2
39	-0.440	1	-0.490	1	-0.041	1	-0.032	1
40	-0.020	1	0.072	1	-0.012	1	-0.009	1
41	0.180	1	0.181	1	0.023	1	0.027	1
42	0.570	1	0.767	1	0.042	1	0.04	1
43	-0.500	1	-0.586	1	-0.043	1	-0.039	1
44	-0.300	1	-0.643	1	-0.009	1	-0.009	1
45	1.100	1	1.328	1	0.059	2	0.059	2
46	-0.050	1	-0.233	1	0.011	1	0.014	1
47	0.320	1	0.170	1	0.006	1	0.007	1
48	-0.520	1	-0.363	1	-0.049	1	-0.047	1
49	0.480	1	0.352	1	0.042	1	0.044	1
50	0.300	1	0.137	1	0.036	1	0.04	1
51	-0.260	1	-0.199	1	0.004	1	0.007	1
52	0.030	1	0.103	1	0.019	1	0.024	1
53	-0.990	1	-1.226	2	-0.068	1	-0.067	1
54	-0.680	1	-0.947	1	-0.034	1	-0.033	1
55	-0.190	1	-0.096	1	0.004	1	0.005	1
56	0.310	1	0.342	1	0.027	1	0.029	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Table 40.

MH/GMH and SIB/Poly-SIB Comparisons: Biology 2222 1998

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
1	-0.220	1	0.156	1	0.025	1	0.023	1
2	1.660	1	1.402	1	0.070	2	0.07	2
3	-0.560	1	-0.313	1	-0.045	1	-0.054	1
4	-0.520	1	-0.391	1	-0.052	1	-0.052	1
5	0.690	1	0.914	1	0.067	1	0.071	1
6	-0.380	1	-0.113	1	-0.042	1	-0.045	1
7	0.950	1	0.874	1	0.079	1	0.081	1
8	0.300	1	0.274	1	0.047	1	0.045	1
9	0.340	1	0.161	1	-0.007	1	-0.018	1
10	-1.410	2	-1.220	2	-0.109	2	-0.108	2
11	-0.950	1	-1.101	1	-0.061	1	-0.06	1
12	-0.330	1	-0.349	1	-0.017	1	-0.015	1
13	0.430	1	0.415	1	0.026	1	0.023	1
14	-0.270	1	-0.250	1	-0.047	1	-0.049	1
15	0.090	1	0.427	1	0.002	1	0.006	1
16	-0.610	1	-0.419	1	-0.045	1	-0.046	1
17	0.910	1	0.754	1	0.049	1	0.052	1
18	-0.590	1	-0.382	1	-0.038	1	-0.04	1
19	-0.320	1	-0.363	1	-0.013	1	-0.017	1
20	-1.340	1	-1.118	1	-0.070	1	-0.064	1
21	-0.060	1	0.072	1	0.008	1	0.008	1
22	-0.360	1	-0.406	1	-0.030	1	-0.028	1
23	1.730	2	1.443	2	0.085	2	0.089	2
24	0.080	1	0.051	1	-0.016	1	-0.013	1
25	-0.810	1	-1.058	1	-0.077	1	-0.073	1
26	-0.210	1	-0.118	1	0.001	1	-0.004	1
27	0.320	1	0.559	1	0.015	1	0.019	1
28	-0.700	1	-0.722	1	-0.047	1	-0.042	1
29	-0.210	1	-0.164	1	-0.021	1	-0.018	1
30	-0.710	1	-0.613	1	-0.082	2	-0.078	2
31	1.500	1	0.961	1	0.022	1	0.021	1
32	-1.080	2	-1.577	2	-0.061	2	-0.064	2
33	0.520	1	0.175	1	0.028	1	0.025	1
34	-0.060	1	-0.128	1	-0.002	1	0	1
35	0.270	1	0.251	1	0.035	1	0.038	1

Item	MH		GMH		SIB		PSIB	
	Δ	DIF	Δ	DIF	$\hat{\beta}_U$	DIF	$\hat{\beta}_U$	DIF
36	0.400	1	0.575	1	0.030	1	0.030	1
37	-0.830	1	-0.742	1	-0.037	1	-0.038	1
38	-0.770	1	-1.198	1	-0.031	1	-0.030	1
39	0.370	1	-0.049	1	-0.003	1	-0.003	1
40	0.230	1	0.256	1	0.024	1	0.028	1
41	0.420	1	0.477	1	0.052	1	0.054	1
42	-0.440	1	-0.220	1	-0.012	1	-0.006	1
43	-0.230	1	-0.432	1	-0.034	1	-0.032	1
44	0.870	1	0.759	1	0.087	1	0.084	1
45	1.150	1	0.884	1	0.103	2	0.111	2
46	1.800	1	1.316	1	0.006	1	0.008	1
47	1.100	1	1.212	1	0.029	1	0.028	1
48	-0.330	1	-0.160	1	-0.008	1	-0.006	1
49	-0.190	1	-0.095	1	-0.024	1	-0.021	1
50	1.260	1	0.980	1	0.065	1	0.069	1
51	0.110	1	0.333	1	0.003	1	0.014	1
52	-0.490	1	-0.598	1	-0.054	1	-0.053	1
53	0.730	1	0.668	1	0.063	1	0.064	1
54	-0.560	1	-0.312	1	-0.029	1	-0.022	1
55	0.010	1	0.198	1	0.012	1	0.010	1
56	-0.070	1	-0.110	1	-0.025	1	-0.024	1

Note. 1 = negligible or no DIF; 2 = moderate to severe DIF.

Appendix B

Comparison of LR versus LDF with Three Effect Size Measures for each Examination and Sample: Dichotomous Items Only

Table 41.

LR/LDF Comparison English January, 1998

Item	LR χ^2		R^2	Cohen	Gierl	Jodoin	LDF χ^2		95% CI
1	42.067	*	0.182	2	2	2	41.475	*	2
2	1.412		0.007	1	1	1	1.351		1
3	0.438		0.001	1	1	1	0.313		1
4	2.568		0.013	1	1	1	2.470		1
5	0.558		0.003	1	1	1	0.533		1
6	2.196		0.009	1	1	1	2.019		1
7	1.815		0.006	1	1	1	1.614		1
8	1.121		0.003	1	1	1	0.945		1
9	1.373		0.008	1	1	1	1.323		1
10	0.539		0.002	1	1	1	0.518		1
11	2.670		0.012	1	1	1	2.527		1
12	11.061	*	0.072	1	2	2	11.048	*	2
13	5.699	*	0.047	1	2	2	5.712	*	2
14	2.947		0.011	1	1	1	3.186		1
15	14.942	*	0.039	1	1	2	15.562	*	2
16	4.774	*	0.010	1	1	1	5.177	*	1
17	17.411	*	0.028	1	1	1	17.775	*	2
18	23.620	*	0.088	1	2	2	23.764	*	2
19	1.202		0.005	1	1	1	1.099		1
20	8.703	*	0.036	1	1	2	8.464	*	1
21	0.083		0.000	1	1	1	0.079		1
22	18.562	*	0.066	1	2	2	18.628	*	2
23	2.805		0.011	1	1	1	2.923		1
24	0.422		0.002	1	1	1	0.389		1
25	0.443		0.001	1	1	1	0.580		1
26	17.616	*	0.059	1	2	2	17.087	*	2
27	15.306	*	0.048	1	2	2	15.784	*	2
28	6.341	*	0.032	1	1	1	6.412	*	1
29	2.359		0.008	1	1	1	2.535		1
30	2.147		0.007	1	1	1	2.306		1
31	6.458	*	0.016	1	1	1	5.779	*	1
32	1.980		0.006	1	1	1	2.240		1
33	0.078		0.000	1	1	1	0.042		1
34	0.270		0.001	1	1	1	0.287		1
35	16.689	*	0.030	1	1	1	17.852	*	2
36	0.000		0.000	1	1	1	0.000		1
37	12.686	*	0.051	1	2	2	12.953	*	2
38	7.595	*	0.035	1	1	2	7.714	*	1

Item	LR						LDF		
	χ^2		R^2	Cohen	Gierl	Jodoin	χ^2		95% CI
39	21.196	*	0.046	1	2	2	22.206	*	2
40	36.870	*	0.098	1	2	2	37.624	*	2
41	1.397		0.005	1	1	1	1.214		1
42	30.689	*	0.077	1	2	2	31.360	*	2
43	1.957		0.009	1	1	1	1.846		1
44	18.088	*	0.077	1	2	2	17.725	*	1
45	0.116		0.001	1	1	1	0.128		1
46	1.079		0.003	1	1	1	0.940		1
47	4.455	*	0.025	1	1	1	4.267	*	1
48	39.499	*	0.116	1	2	2	38.670	*	2
49	0.221		0.001	1	1	1	0.256		1
50	16.888	*	0.087	1	2	2	16.605	*	1
51	12.682	*	0.031	1	1	1	11.733	*	1
52	21.438	*	0.047	1	2	2	19.844	*	2
53	1.979		0.010	1	1	1	2.084		1
54	0.012		0.000	1	1	1	0.000		1
55	0.003		0.000	1	1	1	0.013		1
56	17.342	*	0.062	1	2	2	16.528	*	1
57	3.810		0.014	1	1	1	4.063	*	1
58	0.000		0.000	1	1	1	0.009		1
59	0.829		0.003	1	1	1	0.619		1
60	5.713	*	0.017	1	1	1	6.101	*	1
61	1.591		0.008	1	1	1	1.675		1
62	6.867	*	0.022	1	1	1	7.228	*	1
63	4.790	*	0.014	1	1	1	4.291	*	1
64	0.055		0.000	1	1	1	0.005		1
65	3.631		0.008	1	1	1	3.219		1
66	9.348	*	0.025	1	1	1	9.851	*	1
67	4.722	*	0.028	1	1	1	4.668	*	1
68	0.100		0.001	1	1	1	0.088		1
69	1.020		0.004	1	1	1	1.109		1
70	12.012	*	0.045	1	2	2	11.517	*	1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 42.

LR/LDF Comparison: English June, 1998

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
1	4.408	* 0.022	1	1	1	3.597	1
2	1.978	0.007	1	1	1	1.744	1
3	21.302	* 0.045	1	2	2	22.191	* 2
4	7.597	* 0.018	1	1	1	8.129	* 2
5	1.895	0.004	1	1	1	2.111	1
6	1.949	0.006	1	1	1	2.440	1
7	5.435	* 0.024	1	1	1	5.002	* 1
8	0.839	0.003	1	1	1	0.785	1
9	12.037	* 0.051	1	2	2	11.593	* 1
10	0.217	0.001	1	1	1	0.258	1
11	5.779	* 0.020	1	1	1	5.565	* 1
12	3.179	0.006	1	1	1	2.266	1
13	19.962	* 0.063	1	2	2	19.982	* 2
14	0.073	0.001	1	1	1	0.004	1
15	6.340	* 0.026	1	1	1	7.491	* 1
16	2.665	0.010	1	1	1	3.314	1
17	2.934	0.006	1	1	1	4.742	* 1
18	0.099	0.000	1	1	1	0.054	1
19	0.833	0.002	1	1	1	0.598	1
20	32.983	* 0.066	1	2	2	32.111	* 2
21	63.894	* 0.100	1	2	2	62.111	* 2
22	2.308	0.008	1	1	1	2.145	1
23	2.190	0.005	1	1	1	1.922	1
24	0.767	0.006	1	1	1	0.721	1
25	0.364	0.001	1	1	1	0.453	1
26	0.036	0.000	1	1	1	0.026	1
27	0.488	0.005	1	1	1	0.492	1
28	39.523	* 0.114	1	2	2	40.335	* 2
29	2.264	0.004	1	1	1	3.102	1
30	8.069	* 0.013	1	1	1	8.590	* 1
31	12.227	* 0.045	1	2	2	11.565	* 1
32	14.000	* 0.042	1	2	2	16.365	* 1
33	0.012	0.000	1	1	1	0.020	1
34	0.755	0.002	1	1	1	1.036	1
35	0.020	0.000	1	1	1	0.001	1
36	0.024	0.000	1	1	1	0.049	1
37	4.635	* 0.011	1	1	1	4.005	* 1
38	5.861	* 0.018	1	1	1	7.207	* 1

Item	LR						LDF	
	χ^2		R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
39	0.822		0.002	1	1	1	1.248	1
40	1.357		0.003	1	1	1	1.741	1
41	0.199		0.000	1	1	1	0.024	1
42	1.059		0.002	1	1	1	0.325	1
43	0.506		0.002	1	1	1	0.418	1
44	1.839		0.004	1	1	1	1.191	1
45	0.090		0.000	1	1	1	0.046	1
46	3.553		0.011	1	1	1	3.593	1
47	7.514	*	0.012	1	1	1	7.225	* 2
48	3.427		0.009	1	1	1	3.218	1
49	20.142	*	0.043	1	2	2	17.490	* 1
50	34.444	*	0.062	1	2	2	35.241	* 2
51	0.354		0.001	1	1	1	0.088	1
52	1.314		0.003	1	1	1	2.015	1
53	3.324		0.008	1	1	1	3.087	1
54	16.600	*	0.056	1	2	2	15.785	* 2
55	0.517		0.001	1	1	1	0.419	1
56	1.568		0.004	1	1	1	0.929	1
57	2.628		0.007	1	1	1	1.196	1
58	15.277	*	0.032	1	1	1	15.808	* 2
59	0.302		0.001	1	1	1	0.274	1
60	0.116		0.000	1	1	1	0.139	1
61	1.832		0.004	1	1	1	1.472	1
62	0.729		0.002	1	1	1	0.544	1
63	0.667		0.003	1	1	1	0.680	1
64	0.111		0.000	1	1	1	0.095	1
65	0.525		0.001	1	1	1	0.568	1
66	0.433		0.001	1	1	1	0.243	1
67	0.214		0.000	1	1	1	0.277	1
68	2.465		0.007	1	1	1	2.166	1
69	1.465		0.007	1	1	1	1.459	1
70	1.131		0.002	1	1	1	0.849	1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 43.

LR/LDF Comparison: English 1212, 1998

Item	LR χ^2		LR R^2	Cohen	Gierl	Jodoin	LDF χ^2		LDF 95% CI
1	15.402	*	0.138	2	2	2	15.180	*	1
2	5.982	*	0.048	1	2	2	5.884	*	1
3	0.112		0.000	1	1	1	0.185		1
4	0.416		0.004	1	1	1	0.392		1
5	0.303		0.003	1	1	1	0.307		1
6	0.836		0.007	1	1	1	0.782		1
7	0.954		0.006	1	1	1	0.864		1
8	1.593		0.007	1	1	1	1.441		1
9	0.563		0.004	1	1	1	0.553		1
10	1.065		0.008	1	1	1	1.043		1
11	1.785		0.019	1	1	1	1.720		1
12	2.250		0.020	1	1	1	2.208		1
13	1.967		0.019	1	1	1	1.972		1
14	0.281		0.002	1	1	1	0.206		1
15	2.369		0.018	1	1	1	2.562		1
16	0.268		0.001	1	1	1	0.357		1
17	4.052	*	0.011	1	1	1	4.237	*	1
18	9.157	*	0.068	1	2	2	9.214	*	2
19	0.388		0.003	1	1	1	0.325		1
20	2.011		0.015	1	1	1	1.917		1
21	0.483		0.005	1	1	1	0.476		1
22	4.025	*	0.031	1	1	1	4.015	*	1
23	0.190		0.002	1	1	1	0.210		1
24	0.008		0.000	1	1	1	0.004		1
25	0.348		0.002	1	1	1	0.288		1
26	1.999		0.014	1	1	1	1.859		1
27	4.502	*	0.037	1	1	2	4.695	*	1
28	8.377	*	0.078	1	2	2	8.429	*	1
29	0.001		0.000	1	1	1	0.000		1
30	0.075		0.000	1	1	1	0.112		1
31	5.018	*	0.033	1	1	1	4.553	*	1
32	2.319		0.016	1	1	1	2.494		1
33	0.064		0.000	1	1	1	0.093		1
34	0.068		0.001	1	1	1	0.063		1
35	7.882	*	0.033	1	1	1	8.599	*	1
36	0.037		0.000	1	1	1	0.038		1
37	3.414		0.024	1	1	1	3.610		1
38	0.213		0.002	1	1	1	0.225		1

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
39	8.811	* 0.044	1	2	2	9.391	* 1
40	14.710	* 0.087	1	2	2	15.119	* 2
41	0.023	0.000	1	1	1	0.005	1
42	11.939	* 0.076	1	2	2	12.163	* 2
43	0.341	0.003	1	1	1	0.320	1
44	2.113	0.019	1	1	1	2.031	1
45	0.024	0.000	1	1	1	0.020	1
46	0.925	0.007	1	1	1	1.018	1
47	0.108	0.001	1	1	1	0.093	1
48	9.021	* 0.065	1	2	2	8.580	* 1
49	0.944	0.006	1	1	1	1.052	1
50	3.412	0.027	1	1	1	3.277	1
51	2.339	0.013	1	1	1	1.972	1
52	2.422	0.015	1	1	1	2.028	1
53	3.561	0.031	1	1	1	3.620	1
54	2.271	0.019	1	1	1	2.055	1
55	0.375	0.003	1	1	1	0.406	1
56	0.589	0.006	1	1	1	0.557	1
57	1.804	0.015	1	1	1	1.907	1
58	1.135	0.009	1	1	1	1.064	1
59	0.212	0.001	1	1	1	0.139	1
60	0.170	0.001	1	1	1	0.106	1
61	0.087	0.001	1	1	1	0.103	1
62	0.278	0.003	1	1	1	0.245	1
63	0.205	0.001	1	1	1	0.124	1
64	0.001	0.000	1	1	1	0.007	1
65	0.273	0.001	1	1	1	0.183	1
66	2.655	0.016	1	1	1	2.895	1
67	2.128	0.024	1	1	1	2.109	1
68	0.977	0.009	1	1	1	0.939	1
69	1.637	0.015	1	1	1	1.708	1
70	2.389	0.017	1	1	1	2.234	1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 44.

LR/LDF Comparison: English 2112, 1998

Item	LR χ^2		LR R^2	Cohen	Gierl	Jodoin	LDF χ^2		LDF 95% CI
1	0.345		0.003	1	1	1	0.458		1
2	0.001		0.000	1	1	1	0.000		1
3	8.565	*	0.066	1	2	2	8.593	*	1
4	1.885		0.013	1	1	1	1.966		1
5	0.694		0.004	1	1	1	0.666		1
6	2.853		0.026	1	1	1	3.053		1
7	3.896	*	0.037	1	1	2	3.741		1
8	4.203	*	0.041	1	2	2	4.180	*	1
9	0.228		0.002	1	1	1	0.207		1
10	0.685		0.006	1	1	1	0.635		1
11	0.008		0.000	1	1	1	0.008		1
12	0.683		0.005	1	1	1	0.599		1
13	9.897	*	0.072	1	2	2	9.806	*	2
14	2.414		0.019	1	1	1	2.652		1
15	2.038		0.020	1	1	1	2.445		1
16	0.056		0.001	1	1	1	0.095		1
17	7.578	*	0.043	1	2	2	8.609	*	1
18	1.079		0.009	1	1	1	1.054		1
19	0.999		0.006	1	1	1	0.968		1
20	1.135		0.007	1	1	1	1.146		1
21	8.822	*	0.050	1	2	2	8.907	*	1
22	0.392		0.003	1	1	1	0.393		1
23	1.371		0.011	1	1	1	1.337		1
24	1.172		0.011	1	1	1	1.176		1
25	0.112		0.002	1	1	1	0.121		1
26	0.060		0.000	1	1	1	0.064		1
27	0.007		0.000	1	1	1	0.006		1
28	1.370		0.010	1	1	1	1.412		1
29	0.104		0.000	1	1	1	0.173		1
30	1.186		0.006	1	1	1	1.127		1
31	1.010		0.010	1	1	1	0.949		1
32	1.075		0.009	1	1	1	1.571		1
33	2.952		0.031	1	1	1	2.358		1
34	1.480		0.011	1	1	1	1.552		1
35	1.182		0.013	1	1	1	1.091		1
36	0.333		0.002	1	1	1	0.330		1
37	0.006		0.000	1	1	1	0.001		1
38	0.351		0.004	1	1	1	0.231		1

Item	LR						LDF	
	χ^2		R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
39	0.086		0.001	1	1	1	0.132	1
40	1.258		0.009	1	1	1	1.342	1
41	0.204		0.001	1	1	1	0.127	1
42	1.538		0.012	1	1	1	1.078	1
43	0.457		0.004	1	1	1	0.446	1
44	0.001		0.000	1	1	1	0.006	1
45	0.009		0.000	1	1	1	0.015	1
46	0.013		0.000	1	1	1	0.010	1
47	3.787		0.022	1	1	1	3.523	1
48	0.095		0.001	1	1	1	0.123	1
49	7.731	*	0.061	1	2	2	7.334	* 1
50	9.972	*	0.052	1	2	2	9.744	* 2
51	0.503		0.004	1	1	1	0.362	1
52	0.524		0.003	1	1	1	0.679	1
53	0.179		0.001	1	1	1	0.218	1
54	1.294		0.011	1	1	1	1.210	1
55	1.809		0.019	1	1	1	1.798	1
56	0.131		0.001	1	1	1	0.215	1
57	6.465	*	0.045	1	2	2	5.121	* 1
58	2.365		0.015	1	1	1	2.547	1
59	0.407		0.003	1	1	1	0.449	1
60	2.362		0.017	1	1	1	2.435	1
61	1.262		0.009	1	1	1	1.207	1
62	0.307		0.002	1	1	1	0.280	1
63	3.070		0.031	1	1	1	3.069	1
64	0.039		0.000	1	1	1	0.029	1
65	0.273		0.002	1	1	1	0.301	1
66	0.294		0.002	1	1	1	0.293	1
67	3.832		0.017	1	1	1	3.648	1
68	0.004		0.000	1	1	1	0.001	1
69	1.061		0.012	1	1	1	1.068	1
70	0.301		0.002	1	1	1	0.317	1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 45.

LR/LDF Comparison: English 2211, 1998

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
1	1.035	0.012	1	1	1	41.475 *	1
2	5.081 *	0.050	1	2	2	1.351	1
3	0.567	0.004	1	1	1	0.313	1
4	2.132	0.015	1	1	1	2.470	1
5	0.073	0.000	1	1	1	0.533	1
6	0.033	0.000	1	1	1	2.019	1
7	0.731	0.007	1	1	1	1.614	1
8	1.235	0.015	1	1	1	0.945	1
9	0.776	0.008	1	1	1	1.323	1
10	0.259	0.003	1	1	1	0.518	1
11	1.492	0.015	1	1	1	2.527	1
12	1.247	0.008	1	1	1	11.048 *	1
13	2.742	0.023	1	1	1	5.712 *	1
14	3.661	0.041	1	2	2	3.186	1
15	0.452	0.004	1	1	1	15.562 *	1
16	1.179	0.011	1	1	1	5.177 *	1
17	0.061	0.001	1	1	1	17.775 *	1
18	0.006	0.000	1	1	1	23.764 *	1
19	0.008	0.000	1	1	1	1.099	1
20	5.079 *	0.035	1	1	2	8.464 *	1
21	9.051 *	0.053	1	2	2	0.079	2
22	2.414	0.020	1	1	1	18.628 *	1
23	0.279	0.002	1	1	1	2.923	1
24	2.993	0.039	1	1	2	0.389	1
25	1.908	0.016	1	1	1	0.580	1
26	3.256	0.026	1	1	1	17.087 *	1
27	2.142	0.025	1	1	1	15.784 *	1
28	3.138	0.027	1	1	1	6.412 *	1
29	6.301 *	0.043	1	2	2	2.535	1
30	0.354	0.003	1	1	1	2.306	1
31	3.934 *	0.045	1	2	2	5.779 *	1
32	2.849	0.024	1	1	1	2.240	1
33	1.633	0.019	1	1	1	0.042	1
34	0.963	0.007	1	1	1	0.287	1
35	0.122	0.001	1	1	1	17.852 *	1
36	0.157	0.001	1	1	1	0.000	1
37	2.198	0.018	1	1	1	12.953 *	1
38	1.353	0.011	1	1	1	7.714 *	1

Item	LR						LDF	
	χ^2		R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
39	0.849		0.006	1	1	1	22.206 *	1
40	4.298	*	0.028	1	1	1	37.624 *	1
41	0.697		0.005	1	1	1	1.214	1
42	1.409		0.012	1	1	1	31.360 *	1
43	0.603		0.004	1	1	1	1.846	1
44	2.104		0.013	1	1	1	17.725 *	1
45	0.055		0.000	1	1	1	0.128	1
46	0.003		0.000	1	1	1	0.940	1
47	0.007		0.000	1	1	1	4.267 *	1
48	0.013		0.000	1	1	1	38.670 *	1
49	12.936	*	0.103	1	2	2	0.256	1
50	4.021	*	0.026	1	1	1	16.605 *	1
51	0.003		0.000	1	1	1	11.733 *	1
52	0.001		0.000	1	1	1	19.844 *	1
53	0.739		0.006	1	1	1	2.084	1
54	0.001		0.000	1	1	1	0.000	1
55	0.246		0.003	1	1	1	0.013	1
56	0.478		0.004	1	1	1	16.528 *	1
57	0.505		0.004	1	1	1	4.063 *	1
58	3.132		0.022	1	1	1	0.009	1
59	0.421		0.003	1	1	1	0.619	1
60	1.224		0.010	1	1	1	6.101 *	1
61	0.671		0.007	1	1	1	1.675	1
62	0.011		0.000	1	1	1	7.228 *	1
63	0.754		0.007	1	1	1	4.291 *	1
64	0.199		0.001	1	1	1	0.005	1
65	1.163		0.009	1	1	1	3.219	1
66	0.018		0.000	1	1	1	9.851 *	1
67	1.297		0.009	1	1	1	4.668 *	1
68	0.009		0.000	1	1	1	0.088	1
69	3.781		0.041	1	2	2	1.109	1
70	0.737		0.006	1	1	1	11.517 *	1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 46.

LR/LDF Comparison: English 2222, 1998

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
1	0.006	0.000	1	1	1	0.055	1
2	0.029	0.001	1	1	1	0.021	1
3	2.933	0.024	1	1	1	3.032	1
4	0.945	0.008	1	1	1	1.005	1
5	0.007	0.000	1	1	1	0.002	1
6	0.061	0.000	1	1	1	0.118	1
7	1.594	0.016	1	1	1	1.461	1
8	0.060	0.001	1	1	1	0.071	1
9	0.038	0.000	1	1	1	0.027	1
10	0.104	0.001	1	1	1	0.086	1
11	0.650	0.006	1	1	1	0.612	1
12	0.780	0.004	1	1	1	0.534	1
13	1.078	0.009	1	1	1	1.061	1
14	1.263	0.009	1	1	1	1.640	1
15	3.778	0.042	1	2	2	4.411 *	1
16	0.264	0.002	1	1	1	0.382	1
17	0.807	0.006	1	1	1	0.434	1
18	4.042 *	0.031	1	1	1	4.195 *	1
19	0.088	0.001	1	1	1	0.133	1
20	3.033	0.023	1	1	1	2.932	1
21	8.112 *	0.047	1	2	2	8.003 *	1
22	0.049	0.000	1	1	1	0.059	1
23	0.038	0.000	1	1	1	0.027	1
24	0.757	0.010	1	1	1	0.756	1
25	0.916	0.009	1	1	1	0.975	1
26	1.971	0.021	1	1	1	2.001	1
27	0.387	0.004	1	1	1	0.388	1
28	14.806 *	0.106	1	2	2	15.043 *	2
29	0.288	0.002	1	1	1	0.174	1
30	1.351	0.005	1	1	1	1.299	1
31	0.105	0.001	1	1	1	0.077	1
32	2.552	0.022	1	1	1	3.254	1
33	0.374	0.004	1	1	1	0.601	1
34	0.737	0.005	1	1	1	0.634	1
35	0.385	0.003	1	1	1	0.530	1
36	0.582	0.006	1	1	1	0.630	1
37	1.028	0.008	1	1	1	0.913	1

Item	LR						LDF		
	χ^2		R^2	Cohen	Gierl	Jodoin	χ^2		95% CI
38	3.248		0.022	1	1	1	3.714		1
39	0.067		0.000	1	1	1	0.098		1
40	0.708		0.005	1	1	1	0.600		1
41	0.026		0.000	1	1	1	0.001		1
42	1.005		0.007	1	1	1	0.430		1
43	1.177		0.011	1	1	1	1.140		1
44	1.134		0.008	1	1	1	0.959		1
45	2.095		0.019	1	1	1	1.938		1
46	3.705		0.033	1	1	1	3.698		1
47	1.070		0.005	1	1	1	0.922		1
48	0.078		0.001	1	1	1	0.104		1
49	4.452	*	0.033	1	1	1	3.833		1
50	10.836	*	0.073	1	2	2	11.760	*	2
51	0.318		0.003	1	1	1	0.100		1
52	2.339		0.016	1	1	1	1.595		1
53	0.453		0.003	1	1	1	0.436		1
54	10.731	*	0.093	1	2	2	9.995	*	2
55	3.036		0.021	1	1	1	3.340		1
56	1.574		0.010	1	1	1	1.051		1
57	1.922		0.014	1	1	1	1.058		1
58	2.347		0.017	1	1	1	2.152		1
59	0.092		0.001	1	1	1	0.140		1
60	2.112		0.016	1	1	1	2.077		1
61	0.035		0.000	1	1	1	0.087		1
62	1.743		0.013	1	1	1	1.456		1
63	0.148		0.002	1	1	1	0.125		1
64	0.501		0.004	1	1	1	0.592		1
65	0.035		0.000	1	1	1	0.038		1
66	0.266		0.001	1	1	1	0.343		1
67	0.212		0.002	1	1	1	0.152		1
68	0.156		0.001	1	1	1	0.095		1
69	1.335		0.015	1	1	1	1.199		1
70	7.758	*	0.064	1	2	2	7.055	*	1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 47.

LR/LDF Comparison: Social January, 1998

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
1	0.021	0.000	1	1	1	0.019	1
2	0.776	0.002	1	1	1	0.751	1
3	1.469	0.004	1	1	1	1.673	1
4	1.609	0.007	1	1	1	1.613	1
5	9.496	* 0.024	1	1	1	9.483	* 1
6	0.003	0.000	1	1	1	0.005	1
7	4.385	* 0.010	1	1	1	4.462	* 1
8	2.260	0.006	1	1	1	2.222	1
9	0.265	0.001	1	1	1	0.237	1
10	2.135	0.009	1	1	1	2.105	1
11	0.357	0.001	1	1	1	0.368	1
12	0.035	0.000	1	1	1	0.022	1
13	4.746	* 0.021	1	1	1	4.877	* 1
14	20.288	* 0.043	1	2	2	20.942	* 2
15	4.205	* 0.015	1	1	1	4.349	* 1
16	6.234	* 0.013	1	1	1	6.523	* 1
17	1.109	0.002	1	1	1	1.149	1
18	2.922	0.008	1	1	1	2.849	1
19	3.192	0.012	1	1	1	3.242	1
20	0.924	0.003	1	1	1	0.999	1
21	0.797	0.003	1	1	1	0.729	1
22	2.885	0.011	1	1	1	2.898	1
23	0.391	0.001	1	1	1	0.395	1
24	3.197	0.011	1	1	1	3.243	1
25	0.061	0.000	1	1	1	0.059	1
26	23.767	* 0.138	2	2	2	23.811	* 2
27	0.274	0.001	1	1	1	0.290	1
28	15.531	* 0.052	1	2	2	15.659	* 2
29	0.358	0.001	1	1	1	0.349	1
30	7.850	* 0.014	1	1	1	7.490	* 1
31	0.289	0.001	1	1	1	0.318	1
32	49.835	* 0.139	2	2	2	49.862	* 2
33	1.770	0.005	1	1	1	1.784	1
34	2.415	0.006	1	1	1	2.263	1
35	0.297	0.001	1	1	1	0.282	1
36	3.192	0.011	1	1	1	3.194	1
37	1.121	0.002	1	1	1	1.074	1

Item	LR						LDF	
	χ^2		R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
38	0.879		0.003	1	1	1	0.881	1
39	6.300	*	0.020	1	1	1	6.434	* 1
40	0.915		0.004	1	1	1	0.947	1
41	8.211	*	0.037	1	1	2	8.213	* 1
42	1.867		0.004	1	1	1	1.961	1
43	1.041		0.004	1	1	1	1.028	1
44	0.281		0.001	1	1	1	0.272	1
45	1.756		0.006	1	1	1	1.966	1
46	6.421	*	0.015	1	1	1	6.267	* 1
47	0.508		0.001	1	1	1	0.492	1
48	0.409		0.001	1	1	1	0.416	1
49	2.482		0.008	1	1	1	2.531	1
50	2.427		0.012	1	1	1	2.355	1
51	0.258		0.001	1	1	1	0.279	1
52	7.056	*	0.019	1	1	1	7.032	* 1
53	29.387	*	0.067	1	2	2	29.100	* 2
54	3.738		0.007	1	1	1	3.844	* 1
55	4.047	*	0.011	1	1	1	3.919	* 1
56	1.312		0.003	1	1	1	1.221	1
57	8.692	*	0.017	1	1	1	8.289	* 1
58	0.497		0.001	1	1	1	0.426	1
59	3.036		0.009	1	1	1	3.132	1
60	29.112	*	0.066	1	2	2	28.897	* 2
61	6.788	*	0.018	1	1	1	6.656	* 1
62	4.714	*	0.015	1	1	1	4.785	* 1
63	1.085		0.002	1	1	1	0.931	1
64	1.421		0.003	1	1	1	1.288	1
65	0.812		0.003	1	1	1	0.477	1
66	6.745	*	0.017	1	1	1	6.682	* 1
67	1.113		0.003	1	1	1	1.007	1
68	16.838	*	0.044	1	2	2	16.656	* 2
69	12.247	*	0.038	1	1	2	12.189	* 1
70	16.551	*	0.037	1	1	2	16.278	* 2

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 48.

LR/LDF Comparison: Social June, 1998

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
1	0.006	0.000	1	1	1	0.091	1
2	6.496*	0.023	1	1	1	6.088*	1
3	0.075	0.000	1	1	1	0.006	1
4	2.579	0.010	1	1	1	2.809	1
5	27.271*	0.041	1	2	2	27.981*	1
6	18.789*	0.057	1	2	2	19.212*	1
7	0.232	0.000	1	1	1	0.178	1
8	2.012	0.004	1	1	1	1.730	1
9	0.095	0.000	1	1	1	0.167	1
10	2.051	0.004	1	1	1	1.891	1
11	20.148*	0.025	1	1	1	20.098*	2
12	0.034	0.000	1	1	1	0.002	1
13	4.589*	0.019	1	1	1	4.682*	1
14	0.018	0.000	1	1	1	0.036	1
15	22.214*	0.048	1	2	2	20.826*	2
16	4.026*	0.017	1	1	1	3.865*	1
17	0.541	0.002	1	1	1	0.514	1
18	0.124	0.000	1	1	1	0.041	1
19	0.437	0.001	1	1	1	0.325	1
20	1.617	0.003	1	1	1	1.745	1
21	25.907*	0.038	1	1	2	24.728*	2
22	1.080	0.003	1	1	1	0.877	1
23	0.177	0.000	1	1	1	0.092	1
24	21.601*	0.074	1	2	2	22.039*	2
25	0.140	0.000	1	1	1	0.271	1
26	19.684*	0.046	1	2	2	19.814*	2
27	14.506*	0.047	1	2	2	13.749*	1
28	1.492	0.012	1	1	1	1.508	1
29	0.903	0.002	1	1	1	0.818	1
30	1.238	0.003	1	1	1	0.969	1
31	24.651*	0.051	1	2	2	25.996*	2
32	0.003	0.000	1	1	1	0.003	1
33	0.018	0.000	1	1	1	0.029	1
34	23.501*	0.067	1	2	2	23.975*	2
35	8.562*	0.014	1	1	1	8.364*	1
36	1.322	0.002	1	1	1	1.589	1
37	1.950	0.006	1	1	1	1.996	1

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
38	11.429*	0.029	1	1	1	10.291*	1
39	0.000	0.000	1	1	1	0.017	1
40	16.184*	0.042	1	2	2	16.503*	2
41	31.384*	0.082	1	2	2	30.393*	2
42	2.792	0.007	1	1	1	2.969	1
43	8.284*	0.025	1	1	1	8.249*	1
44	5.697	0.012	1	1	1	5.954*	1
45	22.788*	0.028	1	1	1	24.107*	2
46	5.141*	0.011	1	1	1	5.242*	1
47	6.807*	0.016	1	1	1	5.587*	1
48	31.856*	0.070	1	2	2	31.241*	2
49	8.644*	0.027	1	1	1	8.957*	1
50	101.834*	0.146	2	2	2	103.90*	2
51	0.353	0.001	1	1	1	0.158	1
52	25.117*	0.051	1	2	2	25.945*	2
53	0.221	0.000	1	1	1	0.272	1
54	24.304*	0.053	1	2	2	24.756*	2
55	4.254*	0.010	1	1	1	4.545*	1
56	8.544*	0.016	1	1	1	8.153*	1
57	10.881*	0.025	1	1	1	11.147*	1
58	9.886*	0.021	1	1	1	10.722*	1
59	24.801*	0.046	1	2	2	24.903*	2
60	24.129*	0.036	1	1	2	25.798*	2
61	3.165	0.007	1	1	1	3.172	1
62	17.284*	0.029	1	1	1	18.892*	2
63	1.209	0.004	1	1	1	1.104	1
64	3.773	0.007	1	1	1	4.717*	1
65	2.181	0.005	1	1	1	2.290	1
66	0.059	0.000	1	1	1	0.087	1
67	5.901*	0.011	1	1	1	4.649*	1
68	0.002	0.000	1	1	1	0.063	1
69	5.846*	0.012	1	1	1	5.495*	1
70	1.572	0.006	1	1	1	1.405	1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 49.

LR/LDF Comparison: Social 1212, 1998

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
1	0.000	0.000	1	1	1	0.012	1
2	0.020	0.000	1	1	1	0.029	1
3	1.715	0.009	1	1	1	2.048	1
4	0.158	0.002	1	1	1	0.108	1
5	6.132*	0.034	1	1	1	5.729*	1
6	6.499*	0.042	1	2	2	6.508*	1
7	0.266	0.002	1	1	1	0.190	1
8	0.468	0.003	1	1	1	0.553	1
9	2.118	0.015	1	1	1	1.924	1
10	0.041	0.000	1	1	1	0.012	1
11	3.699	0.019	1	1	1	3.703	1
12	0.044	0.000	1	1	1	0.013	1
13	0.194	0.001	1	1	1	0.207	1
14	0.039	0.001	1	1	1	0.029	1
15	2.021	0.013	1	1	1	2.239	1
16	2.963	0.022	1	1	1	3.021	1
17	0.698	0.006	1	1	1	0.605	1
18	1.273	0.007	1	1	1	1.527	1
19	0.239	0.001	1	1	1	0.249	1
20	1.466	0.011	1	1	1	1.558	1
21	8.483*	0.034	1	1	1	8.696*	1
22	2.073	0.014	1	1	1	2.150	1
23	0.160	0.001	1	1	1	0.255	1
24	11.429*	0.086	1	2	2	11.422*	2
25	0.179	0.001	1	1	1	0.280	1
26	2.866	0.017	1	1	1	3.107	1
27	2.762	0.025	1	1	1	2.845	1
28	0.354	0.005	1	1	1	0.359	1
29	1.354	0.010	1	1	1	1.488	1
30	0.098	0.001	1	1	1	0.184	1
31	3.859*	0.023	1	1	1	3.962*	1
32	0.609	0.003	1	1	1	0.706	1
33	1.445	0.014	1	1	1	1.487	1
34	1.753	0.013	1	1	1	1.843	1
35	0.365	0.002	1	1	1	0.428	1
36	3.116	0.015	1	1	1	2.789	1
37	2.970	0.021	1	1	1	2.940	1

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
38	6.376*	0.035	1	1	2	6.680*	1
39	0.234	0.002	1	1	1	0.301	1
40	1.132	0.007	1	1	1	1.221	1
41	6.926*	0.042	1	2	2	7.112*	1
42	0.186	0.001	1	1	1	0.149	1
43	0.462	0.004	1	1	1	0.407	1
44	0.839	0.005	1	1	1	0.858	1
45	1.615	0.007	1	1	1	1.369	1
46	1.331	0.009	1	1	1	1.408	1
47	1.287	0.008	1	1	1	1.616	1
48	7.556*	0.043	1	2	2	7.541	1
49	2.606	0.022	1	1	1	2.495	1
50	18.955*	0.083	1	2	2	17.770*	2
51	0.179	0.001	1	1	1	0.089	1
52	1.423	0.008	1	1	1	1.144	1
53	0.030	0.000	1	1	1	0.042	1
54	11.704*	0.078	1	2	2	11.384*	1
55	0.488	0.003	1	1	1	0.422	1
56	3.623	0.021	1	1	1	3.905*	1
57	8.264*	0.044	1	2	2	8.148*	1
58	0.071	0.000	1	1	1	0.029	1
59	2.653	0.017	1	1	1	2.647	1
60	7.255*	0.036	1	1	2	6.662*	1
61	0.455	0.003	1	1	1	0.378	1
62	10.381*	0.044	1	2	2	9.105*	1
63	0.006	0.000	1	1	1	0.002	1
64	0.447	0.002	1	1	1	0.251	1
65	0.168	0.001	1	1	1	0.102	1
66	1.839	0.010	1	1	1	1.721	1
67	0.846	0.005	1	1	1	1.150	1
68	1.473	0.013	1	1	1	1.298	1
69	0.930	0.006	1	1	1	1.096	1
70	0.744	0.008	1	1	1	0.780	1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 50.

LR/LDF Comparison: Social 2112, 1998

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
1	0.835	0.006	1	1	1	1.155	1
2	0.019	0.000	1	1	1	0.060	1
3	1.412	0.008	1	1	1	2.121	1
4	0.000	0.000	1	1	1	0.004	1
5	6.147*	0.039	1	1	2	6.660*	1
6	1.138	0.006	1	1	1	1.133	1
7	1.232	0.006	1	1	1	1.040	1
8	2.329	0.014	1	1	1	2.471	1
9	0.249	0.002	1	1	1	0.311	1
10	0.631	0.004	1	1	1	0.792	1
11	0.117	0.001	1	1	1	0.107	1
12	0.463	0.002	1	1	1	0.528	1
13	1.601	0.013	1	1	1	1.666	1
14	10.262*	0.051	1	2	2	10.781*	2
15	1.533	0.011	1	1	1	1.524	1
16	4.847*	0.027	1	1	1	5.198	1
17	3.826	0.026	1	1	1	3.391	1
18	0.523	0.004	1	1	1	0.521	1
19	0.566	0.004	1	1	1	0.543	1
20	1.062	0.007	1	1	1	1.127	1
21	0.018	0.000	1	1	1	0.085	1
22	0.167	0.001	1	1	1	0.147	1
23	0.250	0.002	1	1	1	0.221	1
24	1.349	0.010	1	1	1	1.331	1
25	0.795	0.007	1	1	1	0.667	1
26	6.886*	0.066	1	2	2	6.854*	1
27	1.110	0.008	1	1	1	1.050	1
28	0.882	0.006	1	1	1	0.910	1
29	0.974	0.006	1	1	1	0.761	1
30	3.268	0.015	1	1	1	3.228	1
31	5.564*	0.033	1	1	1	5.755*	1
32	8.615*	0.056	1	2	2	8.304*	1
33	0.213	0.002	1	1	1	0.178	1
34	1.684	0.012	1	1	1	1.305	1
35	0.601	0.004	1	1	1	0.398	1
36	4.355*	0.029	1	1	1	4.434*	1
37	0.045	0.001	1	1	1	0.053	1

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
38	0.926	0.006	1	1	1	0.872	1
39	3.702	0.029	1	1	1	3.751	1
40	2.703	0.026	1	1	1	2.834	1
41	2.267	0.017	1	1	1	2.379	1
42	0.023	0.000	1	1	1	0.026	1
43	6.329*	0.048	1	2	2	6.333*	2
44	1.533	0.010	1	1	1	1.390	1
45	0.582	0.005	1	1	1	0.263	1
46	10.196*	0.058	1	2	2	10.137*	2
47	0.118	0.001	1	1	1	0.065	1
48	4.560*	0.034	1	1	1	4.344*	1
49	1.356	0.011	1	1	1	1.574	1
50	0.302	0.003	1	1	1	0.204	1
51	1.880	0.016	1	1	1	1.867	1
52	2.437	0.014	1	1	1	2.767	1
53	1.713	0.010	1	1	1	1.722	1
54	0.605	0.003	1	1	1	0.479	1
55	0.138	0.001	1	1	1	0.119	1
56	6.574*	0.036	1	1	2	6.459*	1
57	0.648	0.004	1	1	1	0.515	1
58	1.521	0.009	1	1	1	1.471	1
59	0.484	0.003	1	1	1	0.550	1
60	8.308*	0.042	1	2	2	8.860*	1
61	4.791*	0.026	1	1	1	4.556*	1
62	1.094	0.008	1	1	1	1.027	1
63	2.078	0.013	1	1	1	1.316	1
64	0.030	0.000	1	1	1	0.061	1
65	3.421	0.024	1	1	1	1.163	1
66	0.792	0.006	1	1	1	0.629	1
67	0.691	0.003	1	1	1	0.521	1
68	10.299*	0.049	1	2	2	10.328*	1
69	1.681	0.010	1	1	1	1.662	1
70	4.750*	0.027	1	1	1	4.706*	1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 51.

LR/LDF Comparison: Social 2211, 1998

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
1	0.065	0.000	1	1	1	0.122	1
2	0.157	0.002	1	1	1	0.108	1
3	0.223	0.001	1	1	1	0.102	1
4	0.117	0.000	1	1	1	0.138	1
5	7.830*	0.038	1	1	2	8.238*	1
6	0.826	0.007	1	1	1	0.817	1
7	17.307*	0.097	1	2	2	16.400*	2
8	2.689	0.014	1	1	1	2.242	1
9	5.476*	0.030	1	1	1	4.991*	1
10	3.814	0.027	1	1	1	3.552	1
11	5.418*	0.024	1	1	1	5.690*	1
12	0.025	0.000	1	1	1	0.004	1
13	0.038	0.000	1	1	1	0.038	1
14	0.002	0.000	1	1	1	0.008	1
15	10.138*	0.064	1	2	2	9.208*	1
16	2.002	0.020	1	1	1	1.887	1
17	2.755	0.024	1	1	1	2.660	1
18	2.247	0.017	1	1	1	1.970	1
19	0.942	0.007	1	1	1	0.904	1
20	0.902	0.005	1	1	1	1.066	1
21	9.167*	0.052	1	2	2	9.080*	2
22	0.048	0.001	1	1	1	0.030	1
23	0.354	0.003	1	1	1	0.257	1
24	3.033	0.024	1	1	1	3.041	1
25	0.381	0.003	1	1	1	0.232	1
26	3.422	0.021	1	1	1	3.316	1
27	5.053*	0.045	1	2	2	4.740*	1
28	0.012	0.000	1	1	1	0.013	1
29	1.883	0.010	1	1	1	2.018	1
30	0.126	0.002	1	1	1	0.062	1
31	3.584	0.023	1	1	1	3.709	1
32	0.226	0.002	1	1	1	0.305	1
33	0.498	0.005	1	1	1	0.527	1
34	10.548*	0.089	1	2	2	10.511*	1
35	2.382	0.018	1	1	1	2.288	1
36	0.100	0.000	1	1	1	0.039	1
37	2.522	0.020	1	1	1	2.535	1

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
38	0.754	0.005	1	1	1	0.554	1
39	0.330	0.002	1	1	1	0.391	1
40	2.117	0.013	1	1	1	2.170	1
41	8.518*	0.055	1	2	2	8.107	1
42	0.350	0.003	1	1	1	0.388	1
43	6.226*	0.047	1	2	2	6.319*	1
44	2.175	0.014	1	1	1	2.205	1
45	5.300*	0.023	1	1	1	5.830*	1
46	0.001	0.000	1	1	1	0.001	1
47	0.876	0.006	1	1	1	0.672	1
48	2.596	0.017	1	1	1	2.591	1
49	3.105	0.027	1	1	1	3.306	1
50	38.078*	0.170	2	2	2	38.984*	2
51	0.053	0.000	1	1	1	0.001	1
52	6.550*	0.042	1	2	2	6.889*	1
53	3.364	0.021	1	1	1	3.010	1
54	2.716	0.018	1	1	1	2.885	1
55	0.089	0.001	1	1	1	0.126	1
56	0.250	0.002	1	1	1	0.293	1
57	2.395	0.019	1	1	1	2.482	1
58	0.246	0.002	1	1	1	0.350	1
59	1.446	0.010	1	1	1	1.548	1
60	0.930	0.006	1	1	1	1.271	1
61	1.235	0.007	1	1	1	1.203	1
62	7.615*	0.043	1	2	2	8.349	1
63	0.002	0.000	1	1	1	0.008	1
64	4.427*	0.031	1	1	1	4.888*	1
65	1.034	0.006	1	1	1	1.208	1
66	0.001	0.000	1	1	1	0.017	1
67	0.159	0.001	1	1	1	0.084	1
68	0.749	0.006	1	1	1	0.441	1
69	2.695	0.015	1	1	1	2.257	1
70	0.152	0.001	1	1	1	0.183	1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 52.

LR/LDF Comparison: Social 2222, 1998

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
1	1.806	0.010	1	1	1	1.690	1
2	0.822	0.003	1	1	1	0.803	1
3	1.312	0.007	1	1	1	1.308	1
4	2.652	0.012	1	1	1	2.611	1
5	7.387*	0.027	1	1	1	7.226*	1
6	1.132	0.006	1	1	1	1.135	1
7	5.755*	0.023	1	1	1	5.693*	1
8	0.256	0.001	1	1	1	0.267	1
9	0.307	0.001	1	1	1	0.294	1
10	0.027	0.000	1	1	1	0.031	1
11	8.151*	0.031	1	1	1	8.077*	1
12	0.280	0.001	1	1	1	0.267	1
13	4.748*	0.024	1	1	1	4.758*	1
14	4.526*	0.019	1	1	1	4.475*	1
15	2.886	0.013	1	1	1	2.863	1
16	0.011	0.001	1	1	1	0.011	1
17	0.004	0.000	1	1	1	0.005	1
18	4.265*	0.017	1	1	1	4.128*	1
19	1.373	0.006	1	1	1	1.369	1
20	1.867	0.009	1	1	1	1.856	1
21	0.880	0.003	1	1	1	0.829	1
22	0.125	0.001	1	1	1	0.127	1
23	0.582	0.003	1	1	1	0.557	1
24	5.532*	0.028	1	1	1	5.534*	1
25	3.712	0.014	1	1	1	3.603	1
26	11.473*	0.053	1	2	2	11.555*	2
27	0.030	0.000	1	1	1	0.034	1
28	0.326	0.002	1	1	1	0.326	1
29	1.340	0.006	1	1	1	1.354	1
30	0.216	0.001	1	1	1	0.223	1
31	1.086	0.004	1	1	1	1.101	1
32	1.495	0.006	1	1	1	1.456	1
33	0.036	0.000	1	1	1	0.036	1
34	4.100*	0.021	1	1	1	4.121*	2
35	1.290	0.005	1	1	1	1.303	1
36	0.002	0.000	1	1	1	0.002	1
37	0.097	0.000	1	1	1	0.099	1

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
38	0.614	0.003	1	1	1	0.617	1
39	1.810	0.008	1	1	1	1.768	1
40	1.250	0.005	1	1	1	1.274	1
41	0.448	0.002	1	1	1	0.455	1
42	0.977	0.004	1	1	1	0.955	1
43	2.256	0.011	1	1	1	2.238	1
44	0.764	0.003	1	1	1	0.783	1
45	14.014*	0.049	1	2	2	13.564*	1
46	0.482	0.002	1	1	1	0.495	1
47	1.078	0.004	1	1	1	1.052	1
48	10.238*	0.046	1	2	2	10.226*	2
49	3.517	0.017	1	1	1	3.477	1
50	24.199*	0.102	1	2	2	23.996*	2
51	0.002	0.000	1	1	1	0.001	1
52	3.795	0.016	1	1	1	3.733	1
53	0.216	0.001	1	1	1	0.205	1
54	2.951	0.013	1	1	1	2.919	1
55	1.308	0.006	1	1	1	1.280	1
56	2.390	0.009	1	1	1	2.406	1
57	1.177	0.005	1	1	1	1.166	1
58	4.303*	0.018	1	1	1	4.218*	1
59	8.693*	0.036	1	1	2	8.612*	1
60	6.963*	0.025	1	1	1	6.780*	1
61	0.005	0.000	1	1	1	0.004	1
62	1.018	0.005	1	1	1	0.981	1
63	0.653	0.003	1	1	1	0.662	1
64	2.100	0.009	1	1	1	1.995	1
65	0.635	0.003	1	1	1	0.614	1
66	0.635	0.003	1	1	1	0.650	1
67	4.941*	0.021	1	1	1	4.740*	2
68	6.009*	0.028	1	1	1	5.801*	1
69	1.382	0.007	1	1	1	1.398	1
70	0.092	0.000	1	1	1	0.089	1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 53.

LR/LDF Comparison: Mathematics January, 1998

Item	LR						LDF		
	χ^2		R^2	Cohen	Gierl	Jodoin	χ^2		95% CI
1	1.796		.0060	1	1	1	1.799		1
2	19.892	*	.0700	1	2	2	19.433	*	2
3	0.601		.0020	1	1	1	.570		1
4	9.652	*	.0390	1	1	2	9.655	*	1
5	2.588		.0050	1	1	1	2.514		1
6	0.388		.0010	1	1	1	.378		1
7	0.093		.0000	1	1	1	.089		1
8	0.744		.0020	1	1	1	.725		1
9	0.118		.0000	1	1	1	.105		1
10	11.281	*	.0200	1	1	1	11.387	*	1
11	5.149	*	.0170	1	1	1	5.159	*	1
12	0.376		.0010	1	1	1	.397		1
13	12.985	*	.0230	1	1	1	13.190	*	1
14	6.513	*	.0150	1	1	1	6.452	*	1
15	1.552		.0030	1	1	1	1.515		1
16	26.317	*	.0960	1	2	2	24.629	*	2
17	10.897	*	.0280	1	1	1	10.620	*	1
18	2.383		.0050	1	1	1	2.368		1
19	4.552	*	.0090	1	1	1	4.543	*	1
20	49.704	*	.1120	1	2	2	49.544	*	2
21	4.256	*	.0080	1	1	1	4.256	*	1
22	0.016		.0000	1	1	1	.017		1
23	39.498	*	.0940	1	2	2	39.357	*	2
24	0.111		.0000	1	1	1	.111		1
25	0.494		.0020	1	1	1	.530		1
26	0.000		.0000	1	1	1	.000		1
27	2.595		.0040	1	1	1	2.560		1
28	1.945		.0050	1	1	1	1.970		1
29	3.685		.0070	1	1	1	3.743		1
30	3.475		.0060	1	1	1	3.432		1
31	0.328		.0000	1	1	1	.277		1
32	22.469	*	.0740	1	2	2	22.422	*	2
33	2.768		.0060	1	1	1	2.738		1
34	1.370		.0040	1	1	1	1.321		1
35	2.825		.0100	1	1	1	2.873		1
36	0.391		.0010	1	1	1	.408		1
37	3.389		.0070	1	1	1	3.456		1

Item	LR						LDF		
	χ^2		R^2	Cohen	Gierl	Jodoin	χ^2		95% CI
38	3.010		.0070	1	1	1	3.008		1
39	0.014		.0000	1	1	1	.017		1
40	19.677	*	.0360	1	1	2	19.542	*	1
41	32.795	*	.0430	1	2	2	32.441	*	2
42	9.869	*	.0170	1	1	1	9.657	*	2
43	0.504		.0010	1	1	1	.474		1
44	2.350		.0030	1	1	1	2.260		1
45	1.395		.0030	1	1	1	1.407		1
46	1.725		.0050	1	1	1	1.750		1
47	7.981	*	.0140	1	1	1	7.969	*	1
48	1.921		.0070	1	1	1	1.943		1
49	0.909		.0010	1	1	1	.915		1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 54.

LR/LDF Comparison: Mathematics June, 1998

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
1	0.792	0.004	1	1	1	0.930	1
2	21.894*	0.070	1	2	2	22.569*	2
3	2.258	0.004	1	1	1	2.110	1
4	18.304*	0.084	1	2	2	17.687*	2
5	1.822	0.006	1	1	1	1.840	1
6	4.637*	0.013	1	1	1	4.735*	1
7	7.007*	0.028	1	1	1	7.669*	1
8	0.011	0.000	1	1	1	0.049	1
9	2.149	0.004	1	1	1	2.162	1
10	2.386	0.007	1	1	1	2.284	1
11	0.155	0.001	1	1	1	0.206	1
12	0.177	0.000	1	1	1	0.022	1
13	3.269	0.019	1	1	1	3.026	1
14	4.818*	0.017	1	1	1	6.028*	1
15	2.760	0.015	1	1	1	2.852	1
16	4.380*	0.010	1	1	1	5.439*	1
17	5.023*	0.010	1	1	1	4.704*	1
18	0.118	0.001	1	1	1	0.311	1
19	0.793	0.002	1	1	1	1.235	1
20	0.205	0.001	1	1	1	0.154	1
21	1.298	0.006	1	1	1	0.842	1
22	0.209	0.000	1	1	1	0.138	1
23	4.642*	0.013	1	1	1	4.952*	1
24	1.270	0.008	1	1	1	1.155	1
25	6.377*	0.027	1	1	1	5.556*	1
26	2.379	0.011	1	1	1	2.545	1
27	0.531	0.002	1	1	1	0.835	1
28	0.001	0.000	1	1	1	0.000	1
29	2.009	0.004	1	1	1	1.992	1
30	0.069	0.000	1	1	1	0.129	1
31	5.277*	0.022	1	1	1	5.503*	1
32	83.108*	0.191	2	2	2	81.827*	2
33	18.537*	0.053	1	2	2	16.484*	1
34	5.984*	0.019	1	1	1	7.306*	1
35	3.426	0.011	1	1	1	3.853	1
36	16.670*	0.042	1	2	2	15.060*	1
37	0.625	0.002	1	1	1	0.573	1

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
38	4.473*	0.024	1	1	1	5.101	1
39	12.194*	0.026	1	1	1	11.176*	1
40	0.338	0.000	1	1	1	0.628	1
41	1.268	0.007	1	1	1	0.849	1
42	0.247	0.000	1	1	1	0.200	1
43	3.586	0.007	1	1	1	3.209	1
44	0.833	0.002	1	1	1	1.097	1
45	4.811*	0.011	1	1	1	5.966*	1
46	14.280*	0.026	1	1	1	14.878*	2
47	0.021	0.000	1	1	1	0.003	1
48	13.837*	0.032	1	1	1	13.756*	2
49	1.733	0.004	1	1	1	1.736	1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 55.

LR/LDF Comparison: Mathematics 1212, 1998

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
1	0.104	0.001	1	1	1	0.116	1
2	13.352*	0.111	1	2	2	12.485*	1
3	0.432	0.003	1	1	1	0.327	1
4	3.556	0.031	1	1	2	3.562	1
5	0.801	0.006	1	1	1	0.753	1
6	1.068	0.013	1	1	1	0.991	1
7	2.028	0.025	1	1	1	2.044	1
8	0.267	0.002	1	1	1	0.275	1
9	5.555*	0.044	1	2	1	5.405*	2
10	1.462	0.008	1	1	1	1.457	1
11	0.073	0.000	1	1	1	0.038	1
12	2.938	0.016	1	1	1	3.006	1
13	0.799	0.005	1	1	1	0.838	1
14	3.319	0.023	1	1	1	3.882*	1
15	0.515	0.004	1	1	1	0.485	1
16	5.097*	0.052	1	2	1	3.587	1
17	7.909*	0.041	1	2	1	7.030*	1
18	2.005	0.012	1	1	1	1.852	1
19	4.346*	0.031	1	1	1	4.657*	1
20	14.133*	0.093	1	2	1	14.574*	2
21	2.042	0.014	1	1	1	1.969	1
22	0.005	0.001	1	1	1	0.001	1
23	6.420*	0.041	1	2	1	6.801*	1
24	0.016	0.000	1	1	1	0.007	1
25	0.021	0.000	1	1	1	0.001	1
26	1.217	0.010	1	1	1	1.270	1
27	3.279	0.020	1	1	1	2.929	1
28	2.950	0.023	1	1	1	3.178	1
29	0.829	0.005	1	1	1	0.809	1
30	6.119	0.032	1	1	1	6.327*	1
31	3.787	0.020	1	1	1	3.421	1
32	5.246*	0.052	1	2	2	5.301*	1
33	1.770	0.012	1	1	2	1.619	1
34	0.290	0.002	1	1	1	0.193	1
35	0.509	0.003	1	1	1	0.722	1
36	0.472	0.003	1	1	2	0.554	1
37	0.436	0.003	1	1	1	0.434	1

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
38	0.673	0.005	1	1	1	0.761	1
39	0.277	0.001	1	1	1	0.183	1
40	10.129*	0.049	1	2	1	10.082*	2
41	12.341*	0.057	1	2	1	12.462*	1
42	2.671	0.014	1	1	1	2.229	1
43	0.011	0.000	1	1	1	0.008	1
44	0.926	0.005	1	1	1	0.824	1
45	0.000	0.000	1	1	1	0.012	1
46	1.193	0.008	1	1	1	1.462	1
47	1.706	0.013	1	1	1	1.610	1
48	0.320	0.002	1	1	1	0.466	1
49	0.325	0.001	1	1	1	0.436	1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 56.

LR/LDF Comparison: Mathematics 2112, 1998

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
1	3.742	0.037	1	1	2	3.783	1
2	1.262	0.013	1	1	1	1.197	1
3	0.553	0.004	1	1	1	0.566	1
4	2.400	0.024	1	1	1	2.417	1
5	1.085	0.006	1	1	1	1.034	1
6	0.117	0.001	1	1	1	0.124	1
7	0.425	0.005	1	1	1	0.417	1
8	0.149	0.001	1	1	1	0.131	1
9	0.146	0.001	1	1	1	0.159	1
10	5.415*	0.039	1	1	2	5.588*	1
11	1.346	0.012	1	1	1	1.330	1
12	0.014	0.000	1	1	1	0.021	1
13	0.515	0.003	1	1	1	0.561	1
14	1.760	0.013	1	1	1	1.734	1
15	0.509	0.004	1	1	1	0.470	1
16	15.515*	0.141	2	2	2	14.452*	1
17	0.999	0.007	1	1	1	0.998	1
18	1.354	0.007	1	1	1	1.267	1
19	0.125	0.001	1	1	1	0.140	1
20	4.591*	0.034	1	1	1	4.499*	1
21	1.930	0.013	1	1	1	1.856	1
22	0.076	0.001	1	1	1	0.073	1
23	9.073*	0.081	1	2	2	9.004*	2
24	1.787	0.013	1	1	1	1.727	1
25	0.026	0.000	1	1	1	0.033	1
26	0.366	0.004	1	1	1	0.355	1
27	1.130	0.007	1	1	1	1.181	1
28	1.402	0.011	1	1	1	1.405	1
29	1.606	0.009	1	1	1	1.732	1
30	0.428	0.003	1	1	1	0.364	1
31	0.014	0.000	1	1	1	0.038	1
32	3.557	0.030	1	1	1	3.491	1
33	0.008	0.000	1	1	1	0.012	1
34	0.164	0.001	1	1	1	0.158	1
35	3.510	0.026	1	1	1	3.509	1
36	0.253	0.002	1	1	1	0.251	1
37	4.070*	0.020	1	1	1	4.260*	1

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
38	0.924	0.005	1	1	1	0.871	1
39	2.906	0.020	1	1	1	2.792	1
40	2.392	0.016	1	1	1	2.249	1
41	5.614*	0.033	1	1	1	6.005*	1
42	1.046	0.007	1	1	1	1.083	1
43	0.015	0.000	1	1	1	0.036	1
44	0.716	0.004	1	1	1	0.587	1
45	1.292	0.011	1	1	1	1.252	1
46	0.181	0.002	1	1	1	0.186	1
47	0.950	0.005	1	1	1	1.043	1
48	0.463	0.006	1	1	1	0.465	1
49	0.115	0.001	1	1	1	0.091	1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 57.

LR/LDF Comparison: Mathematics 2211, 1998

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
1	0.001	0.000	1	1	1	0.000	1
2	0.310	0.002	1	1	1	0.318	1
3	0.013	0.000	1	1	1	0.007	1
4	1.637	0.021	1	1	1	1.612	1
5	1.831	0.013	1	1	1	1.869	1
6	0.601	0.006	1	1	1	0.625	1
7	0.122	0.002	1	1	1	0.124	1
8	0.145	0.002	1	1	1	0.130	1
9	0.000	0.000	1	1	1	0.001	1
10	2.012	0.015	1	1	1	1.823	1
11	0.019	0.000	1	1	1	0.022	1
12	0.488	0.004	1	1	1	0.456	1
13	9.893*	0.054	1	2	2	9.468*	1
14	0.212	0.001	1	1	1	0.175	1
15	0.011	0.000	1	1	1	0.008	1
16	1.234	0.035	1	1	2	1.365	1
17	2.547	0.029	1	1	1	2.498	1
18	1.417	0.011	1	1	1	1.284	1
19	0.260	0.002	1	1	1	0.298	1
20	13.239*	0.125	2	2	2	13.350*	1
21	6.096*	0.045	1	2	2	5.762*	1
22	0.024	0.000	1	1	1	0.021	1
23	5.929*	0.054	1	2	2	5.987*	1
24	3.541	0.031	1	1	1	3.390	1
25	3.549	0.035	1	1	2	3.388	1
26	0.952	0.011	1	1	1	0.946	1
27	0.894	0.009	1	1	1	0.820	1
28	0.001	0.000	1	1	1	0.001	1
29	0.162	0.001	1	1	1	0.124	1
30	1.351	0.010	1	1	1	1.550	1
31	0.200	0.001	1	1	1	0.291	1
32	3.798	0.035	1	1	2	3.853	1
33	7.441*	0.061	1	2	2	7.192*	1
34	1.069	0.011	1	1	1	1.046	1
35	0.367	0.003	1	1	1	0.304	1
36	0.026	0.000	1	1	1	0.022	1
37	0.301	0.003	1	1	1	0.368	1

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
38	0.393	0.004	1	1	1	0.425	1
39	0.978	0.008	1	1	1	1.047	1
40	2.343	0.018	1	1	1	2.465	1
41	0.008	0.000	1	1	1	0.060	1
42	4.580*	0.033	1	1	1	4.146*	1
43	0.923	0.005	1	1	1	1.217	1
44	0.655	0.005	1	1	1	0.788	1
45	0.020	0.000	1	1	1	0.035	1
46	3.914*	0.034	1	1	1	3.930*	1
47	2.600	0.018	1	1	1	2.365	1
48	0.263	0.003	1	1	1	0.271	1
49	0.736	0.005	1	1	1	0.830	1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 58.

LR/LDF Comparison: Mathematics 2222, 1998

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
1	1.060	0.009	1	1	1	0.816	1
2	4.535*	0.040	1	1	1	4.965*	1
3	2.769	0.016	1	1	1	2.466	1
4	1.935	0.021	1	1	1	1.725	1
5	0.835	0.006	1	1	1	0.841	1
6	4.191*	0.034	1	1	1	4.387*	1
7	0.200	0.002	1	1	1	0.418	1
8	0.456	0.004	1	1	1	0.292	1
9	0.144	0.001	1	1	1	0.133	1
10	0.202	0.001	1	1	1	0.170	1
11	0.192	0.001	1	1	1	0.128	1
12	0.064	0.001	1	1	1	0.396	1
13	0.555	0.005	1	2	2	0.381	1
14	0.922	0.010	1	1	1	1.665	1
15	10.116*	0.086	1	1	1	10.502*	2
16	0.000	0.000	1	1	2	0.143	1
17	0.492	0.003	1	1	1	0.317	1
18	0.022	0.000	1	1	1	0.237	1
19	0.384	0.004	1	1	1	0.112	1
20	0.062	0.000	1	2	2	0.102	1
21	1.622	0.016	1	2	2	1.119	1
22	1.102	0.012	1	1	1	0.882	1
23	1.876	0.013	1	2	2	2.097	1
24	0.085	0.001	1	1	1	0.152	1
25	0.486	0.006	1	1	2	0.246	1
26	0.058	0.001	1	1	1	0.032	1
27	0.092	0.001	1	1	1	0.011	1
28	1.079	0.009	1	1	1	0.970	1
29	0.625	0.004	1	1	1	0.644	1
30	0.249	0.003	1	1	1	0.182	1
31	1.047	0.014	1	1	1	1.114	1
32	10.226*	0.068	1	1	2	9.372*	1
33	0.054	0.000	1	2	2	0.323	1
34	3.663	0.035	1	1	1	5.225*	1
35	4.156*	0.031	1	1	1	4.834*	1
36	9.702*	0.076	1	1	1	8.417*	1
37	0.733	0.007	1	1	1	0.631	1

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
38	0.720	0.008	1	1	1	1.152	1
39	0.129	0.001	1	1	1	0.043	1
40	1.407	0.008	1	1	1	2.123	1
41	0.876	0.006	1	1	1	0.527	1
42	0.139	0.001	1	1	1	0.101	1
43	0.721	0.004	1	1	1	0.545	1
44	0.824	0.004	1	1	1	0.534	1
45	0.149	0.001	1	1	1	0.551	1
46	0.461	0.003	1	1	1	0.645	1
47	0.045	0.000	1	1	1	0.000	1
48	4.977*	0.038	1	1	1	4.620*	1
49	0.014	0.000	1	1	1	0.008	1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 59.

LR/LDF Comparison: Biology January, 1998

Item	LR						LDF		
	χ^2		R^2	Cohen	Gierl	Jodoin	χ^2		95% CI
1	4.900	*	.0140	1	1	1	4.960	*	1
2	3.141		.0100	1	1	1	3.195		1
3	0.283		.0010	1	1	1	0.237		1
4	4.865	*	.0300	1	1	1	4.792	*	1
5	0.000		.0000	1	1	1	0.000		1
6	1.345		.0030	1	1	1	1.379		1
7	1.156		.0040	1	1	1	1.168		1
8	0.237		.0010	1	1	1	0.272		1
9	0.571		.0010	1	1	1	0.618		1
10	6.902	*	.0130	1	1	1	6.930	*	1
11	0.358		.0020	1	1	1	0.316		1
12	5.722	*	.0170	1	1	1	5.825	*	1
13	0.693		.0070	1	1	1	0.691		1
14	16.140	*	.0390	1	1	2	15.385	*	1
15	1.700		.0080	1	1	1	1.663		1
16	8.190	*	.0330	1	1	1	8.025	*	1
17	3.471		.0090	1	1	1	3.454		1
18	3.427		.0190	1	1	1	3.551		1
19	16.750	*	.0360	1	1	2	16.579	*	1
20	0.385		.0030	1	1	1	0.320		1
21	1.497		.0050	1	1	1	1.419		1
22	1.196		.0050	1	1	1	1.292		1
23	16.095	*	.0770	1	2	2	16.111	*	2
24	0.402		.0010	1	1	1	0.393		1
25	6.746	*	.0190	1	1	1	6.419	*	1
26	0.684		.0020	1	1	1	0.619		1
27	21.352	*	.0690	1	2	2	21.443	*	2
28	0.119		.0010	1	1	1	0.096		1
29	22.600	*	.0530	1	2	2	22.177	*	2
30	0.954		.0030	1	1	1	0.889		1
31	8.397	*	.0270	1	1	1	8.480	*	1
32	0.065		.0010	1	1	1	0.102		1
33	13.508	*	.0410	1	2	2	13.605	*	1
34	0.437		.0010	1	1	1	0.406		1
35	12.616	*	.0320	1	1	1	12.662	*	1
36	11.593	*	.0330	1	1	1	11.649	*	1
37	9.631	*	.0410	1	2	2	9.782	*	1

Item	LR						LDF		
	χ^2		R^2	Cohen	Gierl	Jodoin	χ^2		95% CI
38	12.960	*	.0400	1	2	2	13.078	*	1
39	0.150		.0000	1	1	1	0.144		1
40	2.240		.0080	1	1	1	2.117		1
41	0.021		.0000	1	1	1	0.023		1
42	3.884	*	.0120	1	1	1	3.972	*	1
43	0.853		.0040	1	1	1	0.851		1
44	14.569	*	.0900	1	2	2	14.531	*	1
45	0.134		.0010	1	1	1	0.110		1
46	1.385		.0040	1	1	1	1.350		1
47	0.711		.0050	1	1	1	0.742		1
48	8.063	*	.0340	1	1	1	8.070	*	1
49	18.575	*	.0640	1	2	2	18.633	*	2
50	1.444		.0030	1	1	1	1.382		1
51	3.387		.0090	1	1	1	1.382		1
52	0.859		.0050	1	1	1	3.397		1
53	3.729		.0200	1	1	1	0.842		1
54	2.171		.0050	1	1	1	3.675		1
55	0.052		.0000	1	1	1	2.194		1
56	7.849	*	.0250	1	1	1	0.056		1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 60.

LR/LDF Comparison: Biology June, 1998

Item	LR						LDF		
	χ^2		R^2	Cohen	Gierl	Jodoin	χ^2		95% CI
1	0.003		.0000	1	1	1	0.017		1
2	5.816	*	.0160	1	1	1	6.119	*	1
3	0.635		.0010	1	1	1	0.521		1
4	3.759		.0080	1	1	1	3.455		1
5	10.419	*	.0230	1	1	1	10.958	*	1
6	0.292		.0000	1	1	1	0.235		1
7	17.379	*	.0410	1	2	2	18.059	*	2
8	1.433		.0040	1	1	1	1.553		1
9	8.842	*	.0200	1	1	1	9.264	*	1
10	8.788	*	.0160	1	1	1	8.475	*	1
11	12.043	*	.0200	1	1	1	11.391	*	2
12	0.627		.0010	1	1	1	0.442		1
13	1.540		.0030	1	1	1	1.376		1
14	4.571	*	.0090	1	1	1	4.331	*	1
15	0.073		.0000	1	1	1	0.034		1
16	10.194	*	.0270	1	1	1	9.877	*	1
17	8.159	*	.0160	1	1	1	8.627	*	1
18	1.907		.0040	1	1	1	1.782		1
19	4.929	*	.0080	1	1	1	5.088	*	1
20	9.616	*	.0180	1	1	1	9.174	*	1
21	2.425		.0040	1	1	1	2.722		1
22	23.558	*	.0510	1	2	2	22.998	*	2
23	7.336		.0120	1	1	1	7.920	*	1
24	2.075		.0030	1	1	1	2.389		1
25	1.619		.0020	1	1	1	1.427		1
26	0.707		.0020	1	1	1	0.799		1
27	3.067		.0060	1	1	1	3.421		1
28	0.015		.0000	1	1	1	0.046		1
29	8.568	*	.0210	1	1	1	7.962	*	1
30	14.953	*	.0220	1	1	1	13.566	*	1
31	0.227		.0010	1	1	1	0.115		1
32	13.177	*	.0340	1	1	1	12.349	*	1
33	0.266		.0010	1	1	1	0.315		1
34	3.901	*	.0060	1	1	1	4.378	*	1
35	2.774		.0090	1	1	1	2.867		1
36	1.154		.0020	1	1	1	1.334		1
37	7.113	*	.0140	1	1	1	6.626	*	1

Item	LR						LDF		
	χ^2		R^2	Cohen	Gierl	Jodoin	χ^2		95% CI
38	4.975	*	.0130	1	1	1	4.520	*	1
39	1.836		.0040	1	1	1	1.577		1
40	0.866		.0020	1	1	1	1.024		1
41	10.238	*	.0170	1	1	1	10.793	*	2
42	21.015	*	.0330	1	1	1	20.680	*	2
43	0.175		.0000	1	1	1	0.145		1
44	8.590	*	.0160	1	1	1	8.915	*	1
45	2.802		.0060	1	1	1	2.997		1
46	1.380		.0040	1	1	1	1.473		1
47	3.345		.0100	1	1	1	3.444		1
48	0.126		.0000	1	1	1	0.089		1
49	2.770		.0040	1	1	1	2.503		1
50	5.417		.0100	1	1	1	5.456	*	1
51	12.752	*	.0280	1	1	1	12.296	*	1
52	1.360		.0020	1	1	1	1.198		1
53	0.302		.0010	1	1	1	0.319		1
54	9.630	*	.0140	1	1	1	8.872	*	1
55	2.096		.0030	1	1	1	2.013		1
56	2.275		.0060	1	1	1	2.120		1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 61.

LR/LDF Comparison: Biology 1212, 1998

Item	LR						LDF		
	χ^2		R^2	Cohen	Gierl	Jodoin	χ^2		95% CI
1	0.411		.0030	1	1	1	0.278		1
2	3.008		.0290	1	1	1	1.969		1
3	0.483		.0030	1	1	1	0.697		1
4	1.664		.0130	1	1	1	1.211		1
5	6.998	*	.0460	1	2	2	6.613	*	1
6	0.002		.0000	1	1	1	0.002		1
7	1.332		.0110	1	1	1	1.201		1
8	0.453		.0040	1	1	1	0.419		1
9	5.182	*	.0400	1	2	2	4.338		1
10	0.461		.0030	1	1	1	0.350	*	1
11	0.038		.0000	1	1	1	0.138		1
12	1.085		.0090	1	1	1	1.778		1
13	0.027		.0010	1	1	1	0.074		1
14	6.508	*	.0420	1	2	2	6.918	*	2
15	0.001		.0000	1	1	1	0.017		1
16	7.432	*	.0530	1	2	2	7.967	*	1
17	1.840		.0130	1	1	1	1.547		1
18	1.400		.0120	1	1	1	1.512		1
19	3.717		.0280	1	1	1	3.872	*	1
20	0.094		.0010	1	1	1	0.206		1
21	1.445		.0110	1	1	1	0.924		1
22	6.109	*	.0450	1	2	2	6.806	*	1
23	2.495		.0150	1	1	1	1.810		1
24	2.520		.0150	1	1	1	1.703		1
25	1.151		.0080	1	1	1	0.867		1
26	0.055		.0000	1	1	1	0.061		1
27	0.004		.0000	1	1	1	0.027		1
28	0.005		.0000	1	1	1	0.063		1
29	2.821		.0220	1	1	1	3.534		1
30	2.106		.0130	1	1	1	3.035		1
31	1.771		.0130	1	1	1	2.540		1
32	0.113		.0010	1	1	1	0.370		1
33	0.942		.0070	1	1	1	0.991		1
34	1.554		.0090	1	1	1	1.175		1
35	0.325		.0050	1	1	1	0.272		1
36	0.027		.0000	1	1	1	0.093		1
37	0.148		.0010	1	1	1	0.289		1

Item	LR						LDF		
	χ^2		R^2	Cohen	Gierl	Jodoin	χ^2		95% CI
38	1.975		.0150	1	1	1	2.507		1
39	0.402		.0030	1	1	1	0.677		1
40	1.294		.0110	1	1	1	0.976		1
41	0.895		.0060	1	1	1	0.804		1
42	3.956	*	.0260	1	1	1	4.190	*	1
43	0.408		0.004	1	1	1	0.394		1
44	0.695		0.005	1	1	1	0.658		1
45	5.855	*	0.054	1	2	2	5.475	*	1
46	1.349		0.019	1	1	1	1.684		1
47	0.220		0.004	1	1	1	0.150		1
48	0.249		0.001	1	1	1	0.168		1
49	1.119		0.005	1	1	1	1.404		1
50	1.234		0.011	1	1	1	0.939		1
51	3.670		0.038	1	1	2	3.350		1
52	4.106	*	0.024	1	1	1	4.302	*	2
53	0.541		0.005	1	1	1	0.699		1
54	0.465		0.003	1	1	1	0.672		1
55	0.534		0.002	1	1	1	0.467		1
56	1.254		0.015	1	1	1	1.492		1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 62.

LR/LDF Comparison: Biology 2112, 1998

Item	LR						LDF		
	χ^2		R^2	Cohen	Gierl	Jodoin	χ^2		95% CI
1	0.136		.0010	1	1	1	0.133		1
2	1.614		.0180	1	1	1	1.871		1
3	0.402		.0030	1	1	1	0.357		1
4	14.136	*	.0900	1	2	2	13.561	*	1
5	0.151		.0010	1	1	1	0.162		1
6	0.762		.0080	1	1	1	0.708		1
7	9.469	*	.0750	1	2	2	9.567	*	2
8	0.328		.0040	1	1	1	0.325		1
9	0.819		.0060	1	1	1	0.975		1
10	1.404		.0130	1	1	1	1.383		1
11	3.002		.0200	1	1	1	2.672		1
12	1.351		.0130	1	1	1	1.608		1
13	0.826		.0050	1	1	1	0.915		1
14	1.331		.0110	1	1	1	1.286		1
15	3.314		.0270	1	1	1	3.071		1
16	0.827		.0080	1	1	1	0.756		1
17	1.987		.0140	1	1	1	2.032		1
18	0.575		.0050	1	1	1	0.613		1
19	0.097		.0010	1	1	1	0.054		1
20	1.747		.0140	1	1	1	'.665		1
21	2.475		.0160	1	1	1	2.621		1
22	0.335		.0030	1	1	1	0.299		1
23	6.382	*	.0430	1	2	2	6.341	*	2
24	0.190		.0010	1	1	1	0.181		1
25	0.073		.0000	1	1	1	0.070		1
26	0.046		.0000	1	1	1	0.059		1
27	0.002		.0000	1	1	1	0.030		1
28	0.001		.0000	1	1	1	0.002		1
29	4.105	*	.0350	1	1	2	3.333		1
30	2.964		.0190	1	1	1	2.596		1
31	0.924		.0110	1	1	1	0.314		1
32	3.645		.0400	1	2	2	3.050		1
33	0.181		.0020	1	1	1	0.169		1
34	1.399		.0080	1	1	1	1.177		1
35	3.103		.0510	1	2	2	3.192		1
36	0.368		.0030	1	1	1	0.423		1
37	1.462		.0150	1	1	1	1.211		1

Item	LR						LDF		
	χ^2		R^2	Cohen	Gierl	Jodoin	χ^2		95% CI
38	1.147		.0150	1	1	1	0.808		1
39	0.838		.0080	1	1	1	0.650		1
40	2.776		.0230	1	1	1	2.878		1
41	7.456	*	.0510	1	2	2	7.294	*	1
42	10.246	*	.0690	1	2	2	10.349	*	1
43	0.014		.0000	1	1	1	0.025		1
44	0.280		.0010	1	1	1	0.236		1
45	0.956		.0080	1	1	1	0.986		1
46	1.627		.0190	1	1	1	1.870		1
47	0.014		.0000	1	1	1	0.053		1
48	0.669		.0040	1	1	1	0.648		1
49	1.426		.0070	1	1	1	1.171		1
50	2.011		0.017	1	1	1	2.251		1
51	1.989		0.026	1	1	1	2.053		1
52	0.118		0.001	1	1	1	0.073		1
53	0.057		0.001	1	1	1	0.093		1
54	11.853	*	0.075	1	2	2	11.920	*	2
55	0.995		0.006	1	1	1	1.175		1
56	0.194		0.002	1	1	1	0.253		

<u>Note.</u> * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 63.

LR/LDF Comparison: Biology 2211, 1998

Item	LR						LDF		
	χ^2		R^2	Cohen	Gierl	Jodoin	χ^2		95% CI
1	1.186		.0100	1	1	1	1.229		1
2	2.426		.0230	1	1	1	2.477		1
3	1.576		.0160	1	1	1	1.503		1
4	0.054		.0010	1	1	1	0.057		1
5	1.749		.0180	1	1	1	1.810		1
6	0.751		.0070	1	1	1	0.774		1
7	0.489		.0040	1	1	1	0.498		1
8	0.862		.0110	1	1	1	0.883		1
9	1.104		.0120	1	1	1	1.064		1
10	4.560	*	.0290	1	1	1	4.698	*	1
11	1.745		.0260	1	1	1	1.704		1
12	8.416	*	.0790	1	2	2	8.441	*	1
13	5.850	*	.0810	1	2	2	5.850	*	1
14	1.433		.0120	1	1	1	1.302		1
15	0.134		.0010	1	1	1	0.126		1
16	7.284	*	.0830	1	2	2	7.197	*	1
17	0.026		.0000	1	1	1	0.024		1
18	0.173		.0030	1	1	1	0.166		1
19	1.018		.0110	1	1	1	1.076		1
20	0.178		.0020	1	1	1	0.189		1
21	2.158		.0260	1	1	1	2.085		1
22	0.701		.0090	1	1	1	0.733		1
23	1.423		.0190	1	1	1	1.428		1
24	0.045		.0000	1	1	1	0.038		1
25	2.098		.0220	1	1	1	1.994		1
26	0.446		.0040	1	1	1	0.484		1
27	6.789	*	.0600	1	2	2	6.828	*	1
28	0.205		.0040	1	1	1	0.207		1
29	19.191	*	.1560	2	2	2	18.709	*	2
30	1.259		.0130	1	1	1	1.189		1
31	2.012		.0210	1	1	1	2.048		1
32	0.003		.0000	1	1	1	0.000		1
33	5.351	*	.0400	1	2	2	5.423	*	1
34	0.816		.0080	1	1	1	0.767		1
35	0.829		.0050	1	1	1	0.868		1
36	6.873	*	.0560	1	2	2	6.877	*	1
37	5.840	*	.0550	1	2	2	5.884	*	1

Item	LR						LDF		
	χ^2		R^2	Cohen	Gierl	Jodoin	χ^2		95% CI
38	9.724	*	.0570	1	2	2	9.786	*	2
39	1.144		.0120	1	1	1	1.123		1
40	0.006		.0000	1	1	1	0.010		1
41	0.034		.0010	1	1	1	0.035		1
42	1.605		.0180	1	1	1	1.650		1
43	2.291		.0250	1	1	1	2.286		1
44	1.131		.0160	1	1	1	1.126		1
45	2.713		.0240	1	1	1	2.775		1
46	0.055		.0000	1	1	1	0.058		1
47	0.084		.0010	1	1	1	0.090		1
48	1.106		.0110	1	1	1	1.109		1
49	1.013		.0110	1	1	1	1.037		1
50	0.009		.0000	1	1	1	0.005		1
51	0.395		.0030	1	1	1	0.379		1
52	0.004		.0000	1	1	1	0.006		1
53	4.405	*	.0480	1	2	2	4.329	*	1
54	2.983		.0220	1	1	1	2.952		1
55	0.047		.0000	1	1	1	0.048		1
56	0.494		.0050	1	1	1	0.517		1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 64.

LR/LDF Comparison: Biology 2222, 1998

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
1	0.658	0.007	1	1	1	0.651	1
2	2.552	0.024	1	1	1	2.560	1
3	0.360	0.004	1	1	1	0.359	1
4	0.776	0.007	1	1	1	0.776	1
5	4.008	* 0.03	1	1	1	3.970	* 1
6	0.136	0.002	1	1	1	0.135	1
7	3.037	0.028	1	1	1	3.024	1
8	0.671	0.009	1	1	1	0.667	1
9	0.228	0.002	1	1	1	0.242	1
10	4.835	* 0.045	1	2	2	4.873	* 1
11	4.479	* 0.036	1	1	2	4.374	* 1
12	0.278	0.003	1	1	1	0.226	1
13	0.306	0.003	1	1	1	0.299	1
14	0.203	0.002	1	1	1	0.207	1
15	0.258	0.002	1	1	1	0.267	1
16	0.850	0.008	1	1	1	0.849	1
17	2.133	0.021	1	1	1	2.126	1
18	0.598	0.006	1	1	1	0.612	1
19	0.391	0.003	1	1	1	0.437	1
20	4.337	* 0.039	1	1	2	4.293	* 1
21	0.097	0	1	1	1	0.090	1
22	1.222	0.012	1	1	1	1.212	1
23	5.421	* 0.04	1	2	2	5.360	* 1
24	0.094	0.001	1	1	1	0.091	1
25	2.693	0.018	1	1	1	2.700	1
26	0.082	0	1	1	1	0.085	1
27	0.548	0.004	1	1	1	0.570	1
28	1.244	0.01	1	1	1	1.254	1
29	0.116	0.002	1	1	1	0.104	1
30	0.960	0.007	1	1	1	0.912	1
31	1.895	0.015	1	1	1	1.943	1
32	6.236	* 0.064	1	2	2	5.896	* 1
33	0.744	0.007	1	1	1	0.721	1
34	0.121	0.001	1	1	1	0.127	1
35	0.177	0.002	1	1	1	0.177	1
36	0.922	0.008	1	1	1	0.937	1
37	1.764	0.018	1	1	1	1.708	1

Item	LR					LDF	
	χ^2	R^2	Cohen	Gierl	Jodoin	χ^2	95% CI
38	1.722	0.017	1	1	1	1.644	1
39	0.024	0.001	1	1	1	0.032	1
40	0.151	0.001	1	1	1	0.156	1
41	0.997	0.009	1	1	1	0.970	1
42	0.309	0.002	1	1	1	0.330	1
43	1.042	0.012	1	1	1	1.061	1
44	2.250	0.022	1	1	1	2.229	1
45	3.658	0.035	1	1	2	3.613	1
46	3.547	0.032	1	1	1	3.538	1
47	2.405	0.022	1	1	1	2.415	1
48	0.008	0	1	1	1	0.010	1
49	0.026	0	1	1	1	0.036	1
50	2.461	0.023	1	1	1	2.467	1
51	0.401	0.004	1	1	1	0.398	1
52	1.302	0.01	1	1	1	1.442	1
53	2.427	0.019	1	1	1	2.423	1
54	0.367	0.003	1	1	1	0.407	1
55	0.042	0.001	1	1	1	0.036	1
56	0.060	0.001	1	1	1	0.060	1

Note. * $p < 0.05$; 1 = negligible or no DIF; 2= moderate to severe DIF.

Appendix C

DIF Detection Results for each Examination and Sample: Polytomous and Dichotomous Items Combined

Table 65.

DIF Detection by Polytomous Method: English January, 1998

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
1	-1.619	*	2	-0.095	*	2	37.2	*	2
2	-0.21		1	-0.015		1	0.767		1
3	-0.054		1	-0.005		1	0.002		1
4	-0.332		1	-0.028		1	1.658		1
5	-0.108		1	-0.005		1	0.262		1
6	-0.213		1	-0.018		1	1.096		1
7	-0.333		1	-0.009		1	0.691		1
8	-0.076		1	-0.013		1	0.165		1
9	-0.144		1	-0.009		1	0.787		1
10	-0.016		1	-0.007		1	0.121		1
11	-0.511		1	-0.024		1	1.52		1
12	-1.481	*	2	-0.026	*	1	12.97	*	2
13	-0.65	*	1	-0.043	*	1	6.413	*	1
14	0.561	*	1	0.034		1	4.883	*	1
15	0.968	*	1	0.082	*	2	19.65	*	2
16	0.62	*	1	0.054	*	1	7.681	*	2
17	1.164	*	2	0.094	*	2	23.42	*	2
18	0.979	*	1	0.108	*	2	26.93	*	2
19	-0.236		1	-0.015		1	0.515		1
20	-0.56	*	1	-0.051	*	1	6.594	*	1
21	-0.014		1	-0.001		1	0.008		1
22	-0.848	*	1	-0.071	*	2	15.41	*	1
23	0.462	*	1	0.033		1	4.307	*	2
24	-0.041		1	-0.006		1	0.077		1
25	0.317		1	0.019		1	1.659		1
26	-0.762	*	1	-0.068	*	2	13.75	*	1
27	0.991	*	1	0.069	*	2	19.05	*	1
28	0.505	*	1	0.055	*	1	7.816	*	2
29	0.516		1	0.021		1	4.008	*	1
30	0.425	*	1	0.037		1	3.912	*	1
31	-0.402		1	-0.035	*	1	3.301		1
32	0.498	*	1	0.035		1	4.033	*	1
33	0.012		1	0.003		1	0.065		1

	GMH		Poly-SIB		LDF	
Item	Δ	DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
34	0.164	1	0.017	1	0.649	1
35	1.082	* 2	0.089	* 2	23.83	* 2
36	-0.006	1	-0.008	1	0.045	1
37	1.078	* 2	0.07	* 2	15.51	* 2
38	0.645	* 1	0.059	* 2	9.051	* 2
39	1.12	* 2	0.09	* 2	27.74	* 2
40	1.604	* 2	0.107	* 2	42.82	* 2
41	-0.111	1	-0.017	1	0.362	1
42	1.497	* 2	0.086	* 2	35.98	* 2
43	-0.199	1	-0.018	1	1.081	1
44	-1.096	* 2	-0.061	* 2	15.13	* 2
45	0.084	1	0.015	1	0.445	1
46	-0.064	1	-0.013	1	0.283	1
47	-0.453	1	-0.025	1	2.952	1
48	-1.245	* 2	-0.113	* 2	32.96	* 2
49	0.106	1	0.016	1	0.779	1
50	-1.099	* 2	-0.057	* 1	14.29	* 1
51	-0.698	* 1	-0.046	* 1	8.026	* 1
52	-1.016	* 2	-0.066	* 2	14.51	* 2
53	0.493	1	0.022	1	3.093	1
54	0.109	1	0.002	1	0.294	2
55	0.164	1	0.004	1	0.229	1
56	-1.412	* 2	-0.05	* 1	13.06	* 1
57	0.537	* 1	0.038	* 1	5.719	* 1
58	0.108	1	0.008	1	0.275	1
59	-0.089	1	-0.009	1	0.046	1
60	0.644	* 1	0.053	* 1	8.378	* 1
61	0.413	1	0.028	1	2.498	1
62	0.703	* 1	0.047	* 1	9.531	* 1
63	-0.438	1	-0.027	1	2.35	1
64	0.158	1	0.011	1	0.425	1
65	-0.33	1	-0.023	1	1.401	1
66	0.734	* 1	0.065	* 2	13.06	* 1
67	-0.362	1	-0.039	1	3.571	1
68	-0.022	1	-0.001	1	0	1
69	0.262	1	0.022	1	1.948	1

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
70	-0.821	*	1	-0.044	*	1	8.885	*	1
71	-0.198		1	-0.03		1	1.427		1
72	-0.242		1	-0.041		1	2.539		1
73	-0.578	*	1	-0.147	*	2	19.48	*	2
74	-0.251		1	-0.035		1	2.72		1
75	-0.269		1	-0.031		1	2.837		1
76	-0.739	*	1	-0.147	*	2	26.05	*	2

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Table 66.

DIF Detection by Polytomous Method: English June, 1998

	GMH		Poly-SIB		LDF	
Item	Δ	DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
1	-0.622	1	-0.014	1	2.826	1
2	-0.25	1	-0.023	1	1.159	1
3	0.992	1	0.083	2	26.370	2
4	0.613	1	0.055	1	10.380	1
5	0.341	1	0.029	1	3.026	1
6	0.385	1	0.024	1	3.436	1
7	-0.45	1	-0.031	1	4.062	1
8	-0.169	1	-0.006	1	0.501	1
9	-0.661	1	-0.049	1	10.170	1
10	0.03	1	-0.004	1	0.055	1
11	-0.39	1	-0.033	1	4.218	1
12	-0.178	1	-0.014	1	0.850	1
13	0.807	1	0.083	2	21.390	2
14	-0.082	1	0.001	1	0.027	1
15	0.886	1	0.027	1	8.978	1
16	0.489	1	0.02	1	4.147	1
17	0.623	1	0.029	1	7.054	1
18	0.07	1	0	1	0.010	1
19	-0.055	1	-0.005	1	0.109	1
20	-0.957	1	-0.09	2	27.370	2
21	-1.394	2	-0.122	2	53.260	2
22	-0.209	1	-0.017	1	1.518	1
23	-0.236	1	-0.024	1	1.175	1
24	-0.124	1	-0.011	1	0.615	1
25	0.223	1	0.009	1	0.762	1
26	-0.064	1	-0.006	1	0.007	1
27	0.145	1	0.007	1	0.521	1
28	1.211	2	0.106	2	42.600	2
29	0.441	1	0.033	1	4.999	1
30	0.671	1	0.057	1	11.250	2
31	-0.669	1	-0.049	1	9.529	1
32	1.278	2	0.041	1	18.920	1
33	-0.052	1	0.002	1	0.186	1

	GMH		Poly-SIB		LDF	
Item	Δ	DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
34	0.3	1	0.02	1	2.094	1
35	0.073	1	0.004	1	0.086	1
36	0.081	1	0.012	1	0.257	1
37	-0.372	1	-0.024	1	2.504	1
38	0.815	1	0.032	1	8.826	1
39	0.315	1	0.025	1	2.496	1
40	0.332	1	0.025	1	2.940	1
41	0.051	1	-0.001	1	0.067	1
42	-0.114	1	-0.007	1	0.024	1
43	-0.001	1	-0.009	1	0.110	1
44	-0.192	1	-0.008	1	0.270	1
45	0.011	1	-0.001	1	0.002	1
46	0.355	1	0.035	1	5.054	1
47	0.669	1	0.061	2	9.683	2
48	0.404	1	0.037	1	4.503	1
49	-0.878	1	-0.05	1	13.840	1
50	1.177	2	0.104	2	41.050	2
51	0.042	1	-0.001	1	0.097	1
52	0.366	1	0.029	1	3.493	1
53	0.379	1	0.043	1	4.732	1
54	-0.786	1	-0.062	2	13.310	2
55	-0.119	1	-0.006	1	0.101	1
56	-0.222	1	-0.007	1	0.254	1
57	-0.346	1	-0.011	1	0.399	1
58	-0.602	1	-0.06	2	12.400	2
59	0.183	1	0.011	1	0.718	1
60	-0.038	1	-0.005	1	0.004	1
61	-0.166	1	-0.015	1	0.550	1
62	-0.094	1	0	1	0.110	1
63	0.185	1	0.013	1	0.847	1
64	0.02	1	0.001	1	0.018	1
65	-0.047	1	-0.002	1	0.158	1
66	-0.024	1	0.003	1	0.001	1
67	0.153	1	0.024	1	1.034	1
68	-0.227	1	-0.018	1	1.231	1
69	0.238	1	0.021	1	1.697	1

	GMH		Poly-SIB		LDF	
Item	Δ	DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
70	-0.07	1	-0.006	1	0.214	1
71	0.08	1	0.023	1	0.026	1
72	-0.272	1	-0.041	1	5.982	1
73	-0.29	1	-0.026	1	2.929	1
74	-0.397	1	-0.038	1	4.338	1
75	-0.522	1	-0.033	1	4.430	1
76	-0.76	1	-0.14	2	37.11	2

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Table 67.

DIF Detection by Polytomous Method: English 1212, 1998

	GMH			Poly-SIB		LDF	
Item	Δ		DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
1	-1.924	*	2	-0.1	2	13.88	1
2	-1.233	*	2	-0.077	2	5.058	1
3	0.193		1	0.02	1	0.535	1
4	-0.166		1	-0.029	1	0.223	1
5	0.69		1	0.014	1	0.46	1
6	0.091		1	-0.018	1	0.48	1
7	-0.991		1	-0.027	1	0.462	1
8	-0.595		1	-0.044	1	0.787	1
9	-0.213		1	-0.003	1	0.35	1
10	-0.129		1	-0.058	1	0.677	1
11	-0.746		1	-0.046	1	1.28	1
12	-1.315		1	-0.019	1	2.807	1
13	-0.953		1	-0.053	1	2.148	1
14	-0.34		1	0.007	1	0.027	1
15	0.942	*	1	0.07	1	3.509	1
16	0.38		1	0.028	1	0.852	1
17	1.286	*	2	0.098	2	6.002	1
18	1.377	*	2	0.127	2	10.23	2
19	-0.287		1	-0.021	1	0.112	1
20	-0.509		1	-0.042	1	1.392	1
21	-0.035		1	-0.015	1	0.337	1
22	-0.751		1	-0.072	1	3.172	1
23	0.41		1	0.018	1	0.417	1
24	-0.008		1	0.003	1	0.016	1
25	-0.386		1	-0.017	1	0.073	1
26	-0.38		1	-0.047	1	1.249	1
27	0.961		1	0.063	1	5.647	1
28	1.078	*	2	0.115	2	9.11	1
29	-0.389		1	-0.009	1	0.039	1
30	0.044		1	0.018	1	0.39	1
31	-0.858		1	-0.058	1	3.301	1
32	0.753		1	0.059	1	3.424	1
33	0.36		1	0.006	1	0.31	1

	GMH			Poly-SIB		LDF	
Item	Δ		DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
34	0.072		1	0.008	1	0.007	1
35	1.652	*	2	0.13	2	10.79	1
36	0.247		1	-0.014	1	0.066	1
37	1.023	*	1	0.067	1	4.538	1
38	0.38		1	0.012	1	0.396	1
39	1.295	*	2	0.139	2	11.45	2
40	2.232	*	2	0.133	2	17.18	1
41	0.228		1	0.002	1	0.052	1
42	1.951	*	2	0.101	2	13.51	2
43	-0.307		1	-0.021	1	0.17	1
44	-0.867		1	-0.022	1	1.528	1
45	0.149		1	0.004	1	0	1
46	0.479		1	0.038	1	1.564	1
47	-0.706		1	-0.003	1	0.014	1
48	-1.062	*	2	-0.098	2	6.851	1
49	0.576		1	0.047	1	1.726	1
50	-1.02		1	-0.037	1	2.568	1
51	-0.503		1	-0.054	1	1.109	1
52	-0.628		1	-0.023	1	1.143	1
53	1.354	*	2	0.058	1	4.294	1
54	-0.668		1	-0.052	1	1.344	1
55	0.414		1	0.011	1	0.652	1
56	-0.631		1	-0.012	1	0.301	1
57	0.968	*	1	0.044	1	2.53	1
58	-0.55		1	-0.015	1	0.684	1
59	0.117		1	-0.006	1	0.004	1
60	0.216		1	0.025	1	0.001	1
61	0.322		1	0.02	1	0.247	1
62	-0.281		1	-0.008	1	0.089	1
63	0.085		1	-0.016	1	0.002	1
64	0.247		1	0.003	1	0.215	1
65	-0.368		1	-0.012	1	0.01	1
66	1.019	*	2	0.077	2	4.032	1
67	-0.415		1	-0.052	1	1.778	1
68	-0.515		1	-0.021	1	0.617	1
69	0.586		1	0.037	1	2.218	1

	GMH			Poly-SIB		LDF	
Item	Δ		DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
70	-0.401		1	-0.034	1	1.581	1
71	0.076		1	-0.058	1	0.864	1
72	0.161		1	0.004	1	0.001	1
73	0.256	*	1	-0.138	2	6.449	2
74	0.172		1	0.004	1	0.246	1
75	0.313	*	1	-0.094	1	3.094	1
76	0.377	*	1	-0.192	2	12.14	2

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Table 68.

DIF Detection by Polytomous Method: English 2112, 1998

Item	GMH Δ	DIF	Poly-SIB $\hat{\beta}_U$	DIF	LDF χ^2	DIF
1	1.787	1	-0.015	1	0.53	1
2	-0.323	1	-0.02	1	0.025	1
3	1.474	* 2	0.081	* 2	9.978	* 1
4	0.486	1	0.058	* 1	2.558	1
5	0.599	1	0.034	1	0.904	1
6	0.749	1	0.026	1	3.501	1
7	-0.891	1	-0.026	1	3.246	1
8	-0.796	1	-0.005	1	3.714	1
9	-0.025	1	-0.049	* 1	0.079	1
10	0.37	1	-0.004	1	0.818	1
11	-0.183	1	-0.033	1	0.005	1
12	-0.406	1	-0.012	1	0.248	1
13	1.207	* 2	0.077	* 2	9.963	* 2
14	0.758	1	0	1	3.01	1
15	1.252	1	0.027	* 1	2.996	1
16	0.753	1	0.024	* 1	0.167	1
17	1.553	* 2	0.035	* 1	9.792	* 1
18	-0.675	1	0.001	1	0.699	1
19	-0.078	1	0	1	0.596	1
20	-0.686	1	-0.084	* 2	0.713	1
21	-1.076	* 2	-0.115	* 2	7.547	* 1
22	0.303	1	-0.014	1	0.281	1
23	0.006	1	-0.021	1	0.957	1
24	1.215	* 2	-0.008	1	1.202	1
25	0.166	1	0.016	1	0.178	1
26	0.492	1	-0.004	1	0.076	1
27	-0.185	1	0.006	1	0.011	1
28	0.444	1	0.107	* 2	1.657	1
29	0.2	1	0.032	* 1	0.421	1
30	0.463	1	0.06	* 2	1.515	1
31	-0.275	1	-0.046	* 1	0.754	1
32	0.794	1	0.04	* 1	2.001	1
33	-0.959	1	0.004	1	1.978	1

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
34	0.874		1	0.023		1	2.079		1
35	-0.829		1	0.003		1	0.92		1
36	-0.13		1	0.01		1	0.212		1
37	0.031		1	-0.024		1	0.015		1
38	-0.065		1	0.034	*	1	0.153		1
39	0.568		1	0.03	*	1	0.313		1
40	0.647		1	0.026		1	1.631		1
41	0.271		1	0.003		1	0.041		1
42	-0.607		1	-0.005		1	0.794		1
43	-0.133		1	-0.003		1	0.256		1
44	-0.203		1	-0.005		1	0.118		1
45	-0.156		1	0		1	0.053		1
46	-0.017		1	0.04	*	1	0.067		1
47	0.85		1	0.066	*	2	4.149	*	1
48	-0.248		1	0.037	*	1	0.036		1
49	-1.424	*	2	-0.049	*	1	6.382	*	1
50	1.31	*	2	0.109	*	2	10.52	*	2
51	-0.166		1	0.001		1	0.119		1
52	0.754		1	0.031	*	1	1.156		1
53	-0.118		1	0.039	*	1	0.085		1
54	-0.693		1	-0.064	*	2	0.923		1
55	-0.547		1	-0.007		1	1.457		1
56	0.317		1	-0.006		1	0.511		1
57	-2.215	*	2	-0.008		1	4.496	*	1
58	-0.685		1	-0.056	*	1	1.991		1
59	-0.426		1	0.014		1	0.248		1
60	-0.604		1	-0.001		1	2.075		1
61	-0.215		1	-0.011		1	0.758		1
62	-0.356		1	-0.004		1	0.126		1
63	0.827		1	0.016		1	3.239		1
64	0.003		1	0.001		1	0.116		1
65	-0.117		1	-0.002		1	0.184		1
66	-0.24		1	0.003		1	0.068		1
67	0.999	*	1	0.027		1	4.683	*	1
68	0.15		1	-0.018		1	0.007		1
69	-0.442		1	0.023		1	0.964		1

	GMH		Poly-SIB			LDF		
Item	Δ	DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
70	0.228	1	-0.004		1	0.65		1
71	-0.095	1	0.034		1	0.54		1
72	-0.387	1	-0.029		1	5.469	*	1
73	0.246	1	-0.049		1	0.092		1
74	-0.015	1	-0.062	*	2	0.323		1
75	-0.122	1	-0.065	*	2	0.924		1
76	-0.153	1	-0.128	*	2	14.822	*	1

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Table 69.

DIF Detection by Polytomous Method: English 2211, 1998

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
1	-0.505		1	0.014		1	0.554		1
2	-0.985		1	0.002		1	4.525	*	1
3	1.026	*	2	0.126	*	2	1.174		1
4	0.745		1	0.039		1	3.273		1
5	0.289		1	0.041		1	0.241		1
6	0.128		1	0.05		1	0.179		1
7	-0.034		1	-0.054		1	0.468		1
8	0.335		1	-0.102	*	2	1.321		1
9	-0.76		1	-0.03		1	0.534		1
10	-0.008		1	0.012		1	0.109		1
11	0.759		1	-0.006		1	1.832		1
12	-0.054		1	-0.022		1	0.526		1
13	0.591		1	0.144	*	2	3.096		1
14	-1.839		2	0.033		1	2.868		1
15	-0.366		1	0.044	*	1	0.197		1
16	1.156		1	0.008		1	1.854		1
17	0.927		1	0.081	*	2	0.539		1
18	0.462		1	-0.041		1	0.033		1
19	0.163		1	-0.018		1	0.098		1
20	-0.827		1	-0.033		1	3.801		1
21	-1.186	*	2	-0.098	*	2	6.719	*	2
22	-0.387		1	-0.028		1	1.683		1
23	0.084		1	-0.05		1	0.092		1
24	-1.23		1	0.019		1	2.543		1
25	-0.663		1	0.02		1	1.523		1
26	-0.989		1	-0.004		1	2.791		1
27	0.72		1	-0.007		1	2.276		1
28	0.853		1	0.054		1	3.706		1
29	1.213	*	2	0.011		1	8.735	*	1
30	0.799		1	0.048		1	0.768		1
31	-0.997		1	-0.006		1	3.209		1
32	1.467		1	0.024		1	3.711		1
33	-0.821		1	-0.035		1	0.979		1

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
34	-0.286		1	0.045		1	0.365		1
35	-0.49		1	-0.038		1	0.046		1
36	0.139		1	-0.008		1	0.467		1
37	-0.595		1	-0.014		1	1.46		1
38	1.609		1	-0.004		1	1.849		1
39	-0.62		1	0.025		1	0.265		1
40	1.289	*	2	0.041		1	6.026	*	1
41	-0.362		1	-0.001		1	0.186		1
42	1.171		1	-0.037		1	2.601		1
43	-0.026		1	-0.013		1	0.238		1
44	-0.845		1	0.012		1	0.753		1
45	-0.114		1	0.001		1	0.018		1
46	-0.088		1	-0.001		1	0.061		1
47	0.85		1	0.097	*	2	0.024		1
48	0.026		1	-0.005		1	0.054		1
49	-1.755	*	2	-0.097	*	2	10.146	*	1
50	0.651		1	0.122	*	2	5.321	*	1
51	-0.249		1	-0.02		1	0.36		1
52	0.157		1	0.026		1	0.189		1
53	0.558		1	-0.011		1	1.121		1
54	-0.082		1	-0.017		1	0.042		1
55	0.507		1	-0.046		1	0.463		1
56	-0.84		1	0		1	0.085		1
57	1.434		1	-0.039		1	1.493		1
58	-0.535		1	-0.07		1	2.264		1
59	0.613		1	-0.002		1	0.717		1
60	0.839		1	-0.039		1	1.828		1
61	-0.132		1	-0.05		1	0.227		1
62	0.044		1	-0.02		1	0.194		1
63	0.427		1	0.085	*	2	0.806		1
64	-0.149		1	0.022		1	0.014		1
65	0.23		1	-0.02		1	1.726		1
66	0.503		1	-0.011		1	0.046		1
67	0.562		1	0.102	*	2	2.038		1
68	0.03		1	-0.002		1	0.12		1
69	0.936	*	1	-0.065		1	3.792		1

	GMH			Poly-SIB		LDF		
Item	Δ		DIF	$\hat{\beta}_U$	DIF	χ^2		DIF
70	0.494		1	0.005	1	1.116		1
71	-0.246		1	-0.024	1	0.282		1
72	-0.699		1	-0.098	1	1.833		1
73	-0.279		1	0.093	1	1.306		1
74	-0.206		1	0.016	1	1.749		1
75	-0.792	*	1	0.033	1	3.091		1
76	-0.919	*	1	-0.009	1	8.292	*	1

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Table 70.

DIF Detection by Polytomous Method: English 2222, 1998

	GMH		Poly-SIB		LDF	
Item	Δ	DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
1	-0.148	1	-0.023	1	0.07	1
2	-0.016	1	-0.015	1	0.004	1
3	1.185	2	0.091	1	3.105	1
4	0.86	1	0.03	1	1.189	1
5	-0.034	1	-0.014	1	0	1
6	0.025	1	-0.007	1	0.124	1
7	-1.139	1	-0.099	2	1.352	1
8	-0.036	1	-0.004	1	0.086	1
9	-0.165	1	-0.019	1	0.031	1
10	0.214	1	0.004	1	0.095	1
11	-0.392	1	-0.055	1	0.487	1
12	-0.232	1	-0.058	1	0.356	1
13	0.413	1	0.042	1	1.038	1
14	0.772	1	-0.004	1	1.646	1
15	2.846	2	0.016	1	4.527	1
16	0.57	1	-0.025	1	0.399	1
17	0.065	1	-0.028	1	0.345	1
18	0.895	1	0.083	1	4.435	1
19	0.291	1	-0.004	1	0.184	1
20	-0.479	1	-0.075	1	2.754	1
21	-1.343	2	-0.174	2	7.33	1
22	0.298	1	0.047	1	0.075	1
23	0.088	1	-0.049	1	0.017	1
24	0.24	1	-0.036	1	0.82	1
25	0.807	1	0.057	1	0.929	1
26	0.904	1	0.003	1	1.853	1
27	-0.318	1	-0.056	1	0.428	1
28	2.367	2	0.134	2	14.732	2
29	-0.588	1	-0.074	1	0.102	1
30	0.744	1	0.009	1	1.451	1
31	-0.269	1	0.008	1	0.041	1
32	0.839	1	0.017	1	3.508	1
33	0.038	1	0.008	1	0.63	1

	GMH		Poly-SIB		LDF	
Item	Δ	DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
34	-0.487	1	-0.012	1	0.525	1
35	0.658	1	0.012	1	0.616	1
36	0.124	1	0.036	1	0.652	1
37	-0.901	1	-0.085	1	0.735	1
38	1.732	2	0.036	1	3.87	1
39	0.176	1	-0.004	1	0.185	1
40	-0.261	1	-0.057	1	0.503	1
41	-0.068	1	-0.032	1	0.009	1
42	0.076	1	-0.041	1	0.283	1
43	-0.664	1	-0.095	1	1.029	1
44	-0.75	1	-0.047	1	0.766	1
45	-0.838	1	-0.079	1	1.79	1
46	0.966	1	0.053	1	3.874	1
47	0.954	1	0.081	1	0.935	1
48	0.084	1	0.004	1	0.084	1
49	-0.571	1	-0.051	1	3.404	1
50	1.838	2	0.152	2	11.758	2
51	-0.615	1	-0.04	1	0.1	1
52	-0.504	1	-0.03	1	1.595	1
53	-0.962	1	-0.018	1	0.436	1
54	-1.893	2	-0.21	2	9.995	1
55	0.979	1	0.039	1	3.34	1
56	0.24	1	-0.055	1	1.051	1
57	-1.06	1	-0.045	1	1.058	1
58	-1.074	1	-0.053	1	2.152	1
59	0.302	1	-0.013	1	0.14	1
60	-0.892	1	-0.06	1	2.077	1
61	0.444	1	0.012	1	0.087	1
62	-0.677	1	-0.077	1	1.456	1
63	0.535	1	0.009	1	0.125	1
64	0.39	1	0.031	1	0.592	1
65	-0.532	1	-0.015	1	0.038	1
66	-0.021	1	0.021	1	0.343	1
67	0.309	1	-0.031	1	0.152	1
68	-0.268	1	0.007	1	0.095	1
69	0.449	1	0.075	1	1.199	1

	GMH		Poly-SIB		LDF	
Item	Δ	DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
70	-1.322	2	-0.097	1	7.055	1
71	0.562	1	0.028	1	1.104	1
72	0.195	1	-0.022	1	0.006	1
73	-0.532	1	-0.002	1	0.536	1
74	-0.257	1	-0.064	1	0.52	1
75	0.325	1	0.045	1	0	1
76	-0.947	1	-0.073	1	2.57	1

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Table 71.

DIF Detection by Polytomous Method: Social Studies January, 1998

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
1	0.109		1	0.003		1	0.054		1
2	-0.235		1	-0.016		1	0.302		1
3	0.554		1	0.017		1	2.512		1
4	-0.222		1	-0.019		1	1.151		1
5	0.947	*	1	0.055	*	1	11.27	*	2
6	0.052		1	0.002		1	0.056		1
7	-0.436		1	-0.033		1	3.233		1
8	0.532	*	1	0.04	*	1	3.187		1
9	0.142		1	0.01		1	0.644		1
10	-0.363		1	-0.018		1	1.565		1
11	0.193		1	0.015		1	0.687		1
12	0.109		1	0.011		1	0.206		1
13	-0.49	*	1	-0.041		1	4.409	*	1
14	-0.933	*	1	-0.076	*	2	18.302	*	2
15	-0.398		1	-0.034		1	3.498		1
16	-0.467	*	1	-0.051	*	1	5.072	*	1
17	-0.101		1	-0.011		1	0.539		1
18	0.428		1	0.035		1	3.895	*	1
19	-0.355		1	-0.023		1	2.498		1
20	-0.216		1	-0.018		1	0.626		1
21	-0.141		1	-0.009		1	0.359		1
22	-0.28		1	-0.022		1	2.174		1
23	-0.007		1	-0.007		1	0.178		1
24	-0.309		1	-0.029		1	2.491		1
25	0.044		1	0.004		1	0.004		1
26	-1.063	*	2	-0.097	*	2	22.695	*	2
27	-0.039		1	-0.009		1	0.072		1
28	-0.85	*	1	-0.071	*	2	14.111	*	2
29	-0.218		1	-0.004		1	0.135		1
30	0.729	*	1	0.058	*	1	9.725	*	2
31	-0.069		1	-0.002		1	0.145		1
32	-1.676	*	2	-0.122	*	2	46.559	*	2
33	-0.282		1	-0.017		1	1.217		1

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
34	-0.321		1	-0.02		1	1.406		1
35	-0.139		1	-0.009		1	0.039		1
36	0.501	*	1	0.04	*	1	4.327	*	1
37	0.328		1	0.023		1	1.784		1
38	0.208		1	0.016		1	1.285		1
39	-0.476	*	1	-0.045	*	1	5.391	*	1
40	0.314		1	0.022		1	1.346		1
41	0.665	*	1	0.06	*	2	9.395	*	1
42	-0.274		1	-0.017		1	1.13		1
43	0.293		1	0.03		1	1.608		1
44	-0.203		1	-0.009		1	0.092		1
45	0.505		1	0.013		1	2.678		1
46	0.685	*	1	0.055	*	1	7.823	*	1
47	-0.122		1	-0.01		1	0.184		1
48	-0.073		1	-0.009		1	0.163		1
49	0.471		1	0.028		1	3.351		1
50	-0.349		1	-0.021		1	1.947		1
51	0.014		1	-0.002		1	0.123		1
52	0.75	*	1	0.049	*	1	8.375	*	1
53	1.315	*	2	0.109	*	2	32.382	*	2
54	-0.441		1	-0.024		1	2.386		1
55	0.559	*	1	0.042	*	1	5.06	*	1
56	0.319		1	0.029		1	1.918		1
57	0.841	*	1	0.063	*	2	10.53	*	1
58	0.216		1	0.019		1	0.975		1
59	-0.346		1	-0.033		1	2.487		1
60	1.474	*	2	0.092	*	2	32.427	*	2
61	0.678	*	1	0.057	*	1	7.991	*	1
62	-0.535	*	1	-0.031		1	3.865	*	1
63	-0.329		1	-0.013		1	0.445		1
64	0.291		1	0.03		1	2.027		1
65	-0.498		1	-0.008		1	0.207		1
66	0.696	*	1	0.041	*	1	7.933	*	1
67	0.276		1	0.032		1	1.524		1
68	0.994	*	1	0.083	*	2	19.112	*	2
69	0.819	*	1	0.071	*	2	14.115	*	2

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
70	0.994	*	1	0.078	*	2	18.717	*	2
71	-0.51	*	2	-0.133	*	2	12.924	*	2
72	-0.506	*	2	-0.085	*	2	10.929	*	2
73	-0.178		1	-0.045		1	2.133		1
74	-0.612	*	2	-0.085	*	2	12.902	*	2

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Table 72.

DIF Detection by Polytomous Method: Social Studies June, 1998

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
1	0.342		1	0.004		1	1.472		1
2	-0.458	*	1	-0.024		1	3.084		1
3	0.162		1	0.009		1	1.014		1
4	0.538	*	1	0.025		1	5.390	*	1
5	1.35	*	2	0.1	*	2	43.859	*	2
6	-0.559	*	1	-0.056	*	1	13.955	*	1
7	0.164		1	0.011		1	0.417		1
8	-0.098		1	-0.011		1	0.043		1
9	0.278		1	0.021		1	1.896		1
10	-0.086		1	-0.006		1	0.068		1
11	1.146	*	2	0.095	*	2	35.471	*	2
12	0.179		1	0.011		1	0.849		1
13	-0.213		1	-0.025		1	2.865		1
14	0.239		1	0.013		1	1.599		1
15	-0.742	*	1	-0.056	*	1	12.874	*	1
16	-0.272		1	-0.012		1	2.043		1
17	-0.001		1	-0.002		1	0.004		1
18	0.069		1	0.007		1	0.711		1
19	0.311		1	0.028		1	2.468		1
20	0.435	*	1	0.041	*	1	5.295	*	1
21	1.167	*	2	0.1	*	2	40.449	*	2
22	0.475	*	1	0.037	*	1	3.779		1
23	0.121		1	0.012		1	0.516		1
24	-0.697	*	1	-0.069	*	2	16.805	*	2
25	0.314		1	0.018		1	3.593		1
26	-0.647	*	1	-0.05	*	1	13.100	*	1
27	-0.78	*	1	-0.042	*	1	8.907	*	1
28	-0.157		1	-0.005		1	0.881		1
29	0.054		1	-0.006		1	0.005		1
30	-0.03		1	0.005		1	0.001		1
31	-0.663	*	1	-0.064	*	2	18.639	*	2
32	0.193		1	0.016		1	1.139		1
33	0.099		1	0.019		1	0.101		1

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
34	-0.705	*	1	-0.058	*	1	17.578	*	2
35	-0.355		1	-0.026		1	2.852		1
36	0.462	*	1	0.038	*	1	6.879	*	1
37	-0.093		1	-0.007		1	0.506		1
38	-0.524	*	1	-0.031	*	1	5.167	*	1
39	0.235		1	0.014		1	1.778		1
40	-0.653	*	1	-0.056	*	1	10.897	*	2
41	-0.953	*	1	-0.075	*	2	22.251	*	2
42	0.559	*	1	0.044	*	1	7.322	*	1
43	0.689	*	1	0.063	*	2	13.479	*	1
44	-0.192		1	-0.026		1	1.913		1
45	1.371	*	2	0.091	*	2	43.414	*	2
46	-0.232		1	-0.026		1	1.720		1
47	-0.527		1	-0.014		1	1.662		1
48	1.2	*	2	0.119	*	2	41.428	*	2
49	0.778	*	1	0.053	*	1	13.833	*	1
50	2.327	*	2	0.164	*	2	131.583	*	2
51	0.176		1	0.005		1	0.415		1
52	1.414	*	2	0.08	*	2	37.358	*	2
53	0.261		1	0.019		1	2.120		1
54	1.19	*	2	0.096	*	2	35.100	*	2
55	0.682	*	1	0.043	*	1	9.218	*	1
56	-0.416	*	1	-0.029		1	3.137		1
57	0.758	*	1	0.072	*	2	18.240	*	2
58	1.048	*	2	0.051	*	1	18.799	*	1
59	1.19	*	2	0.108	*	2	37.016	*	2
60	1.474	*	2	0.079	*	2	40.687	*	2
61	-0.114		1	-0.004		1	0.679		1
62	1.248	*	2	0.058	*	1	31.373	*	1
63	-0.024		1	-0.001		1	0.055		1
64	0.755	*	1	0.035	*	1	11.727	*	1
65	0.54	*	1	0.033	*	1	6.169	*	1
66	0.287		1	0.022		1	2.117		1
67	-0.25		1	-0.019		1	0.776		1
68	0.311		1	0.004		1	1.078		1
69	-0.274		1	-0.017		1	1.877		1

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
70	-0.073		1	-0.011		1	0.268		1
71	-1.336	*	2	-0.311	*	2	123.726	*	2
72	-1.184	*	2	-0.258	*	2	93.336	*	2
73	-1.203	*	2	-0.299	*	2	106.925	*	2
74	-1.275	*	2	-0.235	*	2	94.585	*	2

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Table 73.

DIF Detection by Polytomous Method: Social Studies 1212, 1998

	GMH		Poly-SIB		LDF	
Item	Δ	DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
1	0.115	1	-0.003	1	0.066	1
2	0.005	1	0.021	1	0.031	1
3	-0.337	1	-0.044	1	0.879	1
4	0.041	1	0.031	1	0.388	1
5	1.084 *	2	0.105 *	2	8.363 *	1
6	-1.017 *	2	-0.091 *	2	5.679 *	1
7	0.41	1	0.021	1	0.737	1
8	-0.17	1	-0.004	1	0.103	1
9	0.633	1	0.054	1	2.999	1
10	0.458	1	0.035	1	0.312	1
11	0.769	1	0.079 *	2	5.892 *	1
12	0.069	1	0.009	1	0.262	1
13	-0.175	1	-0.024	1	0.078	1
14	0.272	1	0.019	1	0.357	1
15	-0.819	1	-0.044	1	1.143	1
16	-0.855	1	-0.069 *	2	2.363	1
17	0.492	1	0.065	1	1.103	1
18	-0.215	1	-0.042	1	0.531	1
19	-0.013	1	-0.032	1	0.008	1
20	-1.003 *	1	-0.046	1	0.798	1
21	1.75 *	2	0.129 *	2	12.31 *	2
22	0.936 *	1	0.059	1	3.333	1
23	-0.685	1	-0.019	1	0.012	1
24	-1.192 *	2	-0.121 *	2	10.06 *	2
25	0.236	1	-0.009	1	0	1
26	-0.556	1	-0.024	1	2.017	1
27	-0.506	1	-0.038	1	2.005	1
28	-0.592	1	-0.021	1	0.24	1
29	-0.096	1	-0.026	1	0.738	1
30	0.407	1	0.012	1	0	1
31	-0.722	1	-0.06	1	2.857	1
32	-0.184	1	-0.011	1	0.269	1
33	0.552	1	0.048	1	2.023	1

	GMH		Poly-SIB		LDF	
Item	Δ	DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
34	-0.494	1	-0.053	1	1.14	1
35	-0.111	1	0.004	1	0.051	1
36	1.511 *	2	0.109 *	2	5.023 *	1
37	-0.437	1	-0.059	1	1.893	1
38	-1.079	1	-0.069 *	2	5.145 *	1
39	-0.165	1	-0.022	1	0.024	1
40	-0.467	1	-0.033	1	0.474	1
41	-0.942	1	-0.08 *	2	4.978 *	1
42	-0.053	1	0.032	1	0.587	1
43	0.667	1	0.04	1	0.915	1
44	-0.3	1	-0.004	1	0.219	1
45	0.869	1	0.071 *	2	3.269	1
46	-0.175	1	-0.038	1	0.701	1
47	-0.817	1	-0.009	1	0.709	1
48	1.426 *	2	0.114 *	2	9.889 *	2
49	0.983	1	0.077 *	2	3.318	1
50	2.24 *	3	0.163 *	2	22.65 *	2
51	0.552	1	0.031	1	0.491	1
52	1.218 *	1	0.056	1	2.292	1
53	0.05	1	0.01	1	0.364	1
54	1.703 *	2	0.13 *	2	13.62 *	1
55	0.059	1	0.021	1	0.872	1
56	-0.595	1	-0.05	1	2.053	1
57	1.482 *	2	0.125 *	2	10.67 *	1
58	0.634	1	0.034	1	0.357	1
59	0.892 *	1	0.076	1	4.147 *	1
60	1.842 *	2	0.089 *	2	8.962 *	1
61	0.458	1	0.059	1	1.157	1
62	2.176 *	2	0.115 *	2	12.66 *	1
63	0.304	1	0.015	1	0.131	1
64	0.926	1	0.03	1	1.066	1
65	0.214	1	0.031	1	0.56	1
66	0.943 *	1	0.057	1	2.877	1
67	-0.404	1	0.01	1	0.253	1
68	1.355	1	0.025	1	1.765	1
69	-0.259	1	-0.026	1	0.461	1

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
70	0.009		1	0.005		1	0.42		1
71	-1.138	*	2	-0.262	*	2	15.94	*	2
72	-1.16	*	2	-0.233	*	2	15.69	*	2
73	-0.938	*	1	-0.24	*	2	12.88	*	2
74	-1.195	*	2	-0.194	*	2	15.39	*	2

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Table 74.

DIF Detection by Polytomous Method: Social Studies 2112, 1998

	GMH		Poly-SIB		LDF	
Item	Δ	DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
1	0.582	1	0.034	1	1.525	1
2	-0.298	1	0.011	1	0.187	1
3	0.719	1	0.041	1	2.645	1
4	0.011	1	0.004	1	0.033	1
5	1.318 *	2	0.072 *	2	7.404 *	1
6	0.25	1	0.04	1	1.593	1
7	-0.255	1	-0.005	1	0.711	1
8	0.375	1	0.049	1	2.929	1
9	0.422	1	0.03	1	0.508	1
10	0.162	1	0.018	1	0.953	1
11	-0.386	1	-0.029	1	0.041	1
12	-0.483	1	-0.024	1	0.289	1
13	-0.526	1	-0.035	1	1.522	1
14	-1.257 *	2	-0.095 *	2	9.657 *	1
15	-0.159	1	-0.035	1	1.28	1
16	-0.586	1	-0.085 *	2	4.387 *	1
17	-1.108	1	-0.056	1	2.731	1
18	0.565	1	0.027	1	0.776	1
19	-0.439	1	-0.022	1	0.365	1
20	-0.657	1	-0.044	1	0.838	1
21	-0.071	1	-0.003	1	0.188	1
22	-0.216	1	-0.007	1	0.057	1
23	-0.005	1	-0.012	1	0.13	1
24	-0.107	1	-0.037	1	1.058	1
25	-0.407	1	-0.018	1	0.532	1
26	-0.884	1	-0.097 *	2	6.356 *	1
27	-0.42	1	-0.03	1	0.784	1
28	-0.196	1	-0.046	1	0.625	1
29	-0.273	1	-0.016	1	0.485	1
30	1.04 *	2	0.059	1	3.945 *	1
31	-0.925 *	1	-0.088 *	2	5.153 *	1
32	-1.303 *	2	-0.08 *	2	7.577 *	1
33	0.285	1	-0.003	1	0.091	1

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
34	-0.136		1	-0.014		1	0.986		1
35	-0.541		1	-0.013		1	0.188		1
36	0.752		1	0.08	*	2	5.028	*	1
37	-0.023		1	-0.011		1	0.168		1
38	-0.049		1	-0.028		1	0.664		1
39	-0.899	*	1	-0.102	*	2	3.527		1
40	1.164	*	2	0.059		1	3.077		1
41	0.934		1	0.065		1	2.666		1
42	0.311		1	0.022		1	0.12		1
43	0.825		1	0.084	*	2	7.035	*	1
44	-0.78		1	-0.06		1	1.077		1
45	-0.539		1	-0.026		1	0.143		1
46	1.179	*	2	0.103	*	2	11.281	*	2
47	0.117		1	-0.017		1	0.01		1
48	-0.822		1	-0.068		1	3.725		1
49	0.218		1	0.027		1	1.94		2
50	-0.344		1	-0.001		1	0.118		1
51	-0.542		1	-0.04		1	1.575		2
52	1.156		1	0.06		1	3.309		1
53	0.579		1	0.049		1	2.088		1
54	0.023		1	0.007		1	0.205		1
55	0.475		1	0.031		1	0.243		1
56	1.095	*	2	0.104	*	2	7.316	*	1
57	0.557		1	0.026		1	0.888		1
58	0.617		1	0.054		1	1.966		1
59	-0.183		1	-0.016		1	0.349		1
60	1.491	*	2	0.088	*	2	9.874	*	1
61	1.304	*	2	0.076		1	5.33	*	2
62	-0.202		1	-0.033		1	0.751		1
63	-0.798		1	-0.035		1	0.951		1
64	-0.006		1	0.001		1	0.012		1
65	-2.629	*	2	-0.021		1	0.8		1
66	-0.66		1	-0.024		1	0.409		1
67	0.426		1	0.038		1	0.785		1
68	1.297	*	2	0.12	*	2	11.467	*	1
69	0.634		1	0.051		1	2.088		1

	GMH		Poly-SIB		LDF	
Item	Δ	DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
70	0.973 *	1	0.082 *	2	5.388 *	1
71	-0.24	2	-0.074	1	5.818 *	1
72	-0.276 *	2	-0.026	1	2.275	1
73	-0.152	1	-0.06	1	1.52	1
74	-0.152	1	-0.064	1	4.942 *	1

Note. * = p < .05 for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Table 75.

DIF Detection by Polytomous Method: Social Studies 2211, 1998

Item	GMH Δ		DIF	Poly-SIB $\hat{\beta}_U$		DIF	LDF χ^2		DIF
1	1.443		1	0.018		1	0.497		1
2	-0.307		1	-0.007		1	0		1
3	0.153		1	0.009		1	0.049		1
4	0.642		1	0.029		1	0.32		1
5	2.128	*	2	0.151	*	2	11.896	*	1
6	-0.114		1	-0.03		1	0.303		1
7	-1.784	*	2	-0.103	*	2	11.639	*	1
8	-0.817		1	-0.029		1	0.79		1
9	-0.997		1	-0.025		1	2.734		1
10	-0.772		1	-0.042		1	1.899		1
11	1.302	*	2	0.136	*	2	9.378	*	1
12	-0.149		1	0.014		1	0.089		1
13	0.469		1	0.023		1	0.245		1
14	-0.048		1	0.01		1	0.283		1
15	-1.84	*	2	-0.105	*	2	6.408	*	1
16	-0.979		1	-0.051		1	1.11		1
17	-0.388		1	-0.029		1	1.741		1
18	-0.265		1	-0.038		1	0.931		1
19	0.43		1	0.05		1	2.149		1
20	1.171	*	2	0.041		1	2.371		1
21	1.443	*	2	0.149	*	2	12.626	*	1
22	0.48		1	0.025		1	0.42		1
23	0.223		1	-0.006		1	0.004		1
24	-0.715		1	-0.049		1	1.898		1
25	-0.228		1	0.026		1	0.021		1
26	-0.755		1	-0.051		1	1.832		1
27	-1.623	*	2	-0.065		1	3.452		1
28	0.071		1	-0.018		1	0.003		1
29	1.003		1	0.074	*	2	3.506		1
30	-0.585		1	0.019		1	0.05		1
31	-0.494		1	-0.034		1	2.428		1
32	0.258		1	0.034		1	0.938		1
33	-0.113		1	-0.034		1	0.284		1

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
34	-0.952		1	-0.107	*	2	8.532	*	1
35	-0.359		1	-0.025		1	1.123		1
36	0.143		1	0.004		1	0.202		1
37	-0.281		1	-0.053		1	1.631		1
38	0.082		1	-0.01		1	0.056		1
39	0.457		1	0.056		1	1.166		1
40	-0.172		1	-0.048		1	1.457		1
41	-1.428	*	2	-0.104	*	2	6.075	*	1
42	0.376		1	0.04		1	1.082		1
43	1.368	*	2	0.109	*	2	8.023	*	1
44	1.216	*	2	0.098	*	2	3.76		1
45	1.319	*	2	0.117	*	2	10.152	*	1
46	0.526		1	0.026		1	0.311		1
47	-1.298		1	0.004		1	0.18		1
48	1.011	*	1	0.119	*	2	4.158	*	1
49	1.232	*	1	0.054		1	4.561	*	1
50	3.392	*	2	0.213	*	2	45.353	*	1
51	0.988		1	0.011		1	0.312		1
52	1.501	*	2	0.071	*	2	9.398	*	2
53	-0.953		1	-0.009		1	1.382		1
54	0.875		1	0.067		1	4.456	*	1
55	0.762		1	0.017		1	0.552		1
56	0.116		1	0.044		1	0.941		1
57	0.462		1	0.09	*	2	3.896	*	1
58	0.634		1	0.024		1	1.121		1
59	0.889		1	0.106	*	2	3.324		1
60	0.858		1	0.064	*	2	3.046		1
61	-0.007		1	-0.011		1	0.395		1
62	1.951	*	2	0.088	*	2	12.075	*	1
63	-0.102		1	0.009		1	0.223		1
64	1.52	*	1	0.071	*	2	7.344	*	1
65	0.622		1	0.069	*	2	2.491		1
66	0.379		1	0.038		1	0.473		1
67	0.717		1	-0.005		1	0.034		1
68	-1.167		1	-0.008		1	0.042		1
69	-0.808		1	-0.032		1	0.796		1

Item	GMH			Poly-SIB			LDF		
	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
70	0.513		1	0.043		1	0.535		1
71	-1.329	*	2	-0.378	*	2	24.651	*	2
72	-1.093	*	2	-0.284	*	2	16.409	*	2
73	-0.925	*	1	-0.225	*	2	15.052	*	2
74	-1.406	*	2	-0.269	*	2	20.7	*	2

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 76.

DIF Detection by Polytomous Method: Social Studies 2222, 1998

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
1	1.113		1	0.025		1	1.69		1
2	0.358		1	0.062		1	0.803		1
3	-0.872		1	-0.011		1	1.308		1
4	1.042		1	0.066		1	2.611		1
5	1.356	*	2	0.103	*	2	7.226	*	1
6	-0.361		1	-0.059		1	1.135		1
7	1.635	*	2	0.117	*	2	5.693	*	1
8	-0.135		1	0.009		1	0.267		1
9	0.678		1	0.027		1	0.294		1
10	0.016		1	-0.012		1	0.031		1
11	1.711	*	2	0.124	*	2	8.077	*	1
12	0.397		1	0.036		1	0.267		1
13	-0.794		1	-0.129	*	2	4.758	*	1
14	1.091		1	0.152	*	2	4.475	*	1
15	-1.217		2	-0.055		1	2.863		1
16	0.535		1	0.037		1	0.011		1
17	-0.06		1	0.035		1	0.005		1
18	1.087		1	0.088	*	2	4.128	*	1
19	0.861		1	0.06		1	1.369		1
20	0.795		1	0.08		1	1.856		1
21	0.126		1	0.028		1	0.829		1
22	-0.074		1	-0.007		1	0.127		1
23	0.292		1	0.038		1	0.557		1
24	-1.276	*	2	-0.136	*	2	5.534	*	1
25	1.501	*	2	0.008		1	3.603		1
26	-1.656	*	2	-0.132	*	2	11.555	*	1
27	-0.44		1	-0.013		1	0.034		1
28	0.432		1	0.053		1	0.326		1
29	-1.096		1	-0.041		1	1.354		1
30	0.033		1	-0.028		1	0.223		1
31	-0.192		1	-0.044		1	1.101		1
32	0.762		1	0.076		1	1.456		1
33	0.205		1	0.01		1	0.036		1

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
34	-0.347		1	-0.057		1	4.121	*	1
35	-0.784		1	-0.028		1	1.303		1
36	0.554		1	0.004		1	0.002		1
37	0.179		1	0.009		1	0.099		1
38	-0.712		1	-0.04		1	0.617		1
39	0.419		1	0.043		1	1.768		1
40	-1.162	*	1	-0.056		1	1.274		1
41	-0.567		1	-0.014		1	0.455		1
42	1.172		1	0.039		1	0.955		1
43	0.895		1	0.081		1	2.238		1
44	-0.557		1	-0.05		1	0.783		1
45	2.29	*	2	0.166	*	2	13.564	*	2
46	-0.977		1	-0.041		1	0.495		1
47	-1.477		1	0.001		1	1.052		1
48	1.584	*	2	0.115	*	2	10.226	*	2
49	1.315		1	0.075		1	3.477		1
50	1.969	*	2	0.194	*	2	23.996	*	1
51	0.286		1	0.003		1	0.001		1
52	1.583	*	2	0.088	*	2	3.733		1
53	0.072		1	0.016		1	0.205		1
54	0.724		1	0.068		1	2.919		1
55	0.514		1	0.055		1	1.28		1
56	-1.226	*	1	-0.076		1	2.406		1
57	0.607		1	-0.002		1	1.166		1
58	0.721		1	0.093	*	2	4.218	*	1
59	2.061	*	2	0.123	*	2	8.612	*	1
60	1.528	*	2	0.128	*	2	6.78	*	1
61	-0.255		1	0.015		1	0.004		1
62	1.392		2	0.041		1	0.981		1
63	-0.479		1	-0.047		1	0.662		1
64	0.979		1	0.076	*	2	1.995		1
65	0.261		1	0.058		1	0.614		1
66	-0.613		1	-0.028		1	0.65		1
67	-0.687		1	-0.078	*	2	4.74	*	1
68	1.54		1	0.066	*	2	5.801	*	1
69	-0.832		1	-0.027		1	1.398		1

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
70	0.28		1	0.016		1	0.089		1
71	-1.673	*	2	-0.371	*	2	27.53	*	2
72	-1.212	*	2	-0.293	*	2	12.242	*	2
73	-1.567	*	2	-0.479	*	2	32.715	*	2
74	-1.232	*	2	-0.248	*	2	12.26	*	2

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Table 77.

DIF Detection by Polytomous Method: Mathematics January, 1998

	GMH		Poly-SIB		LDF	
Item	Δ	DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
1	-0.205	1	-0.015	1	1.338	1
2	-1.305 *	2	-0.044 *	1	18.225 *	1
3	-0.084	1	-0.007	1	0.366	1
4	-0.576 *	1	-0.052 *	1	8.821 *	1
5	0.304	1	0.03	1	3.044	1
6	-0.1	1	-0.012	1	0.207	1
7	0.067	1	0.009	1	0.154	1
8	0.181	1	0.019	1	1.031	1
9	0.122	1	0.008	1	0.28	1
10	-0.606 *	1	-0.054 *	1	9.598 *	1
11	0.551 *	1	0.03 *	1	5.719 *	1
12	-0.057	1	-0.009	1	0.183	1
13	-0.619 *	1	-0.06 *	2	11.826 *	2
14	0.829 *	1	0.029 *	1	6.99 *	1
15	0.245	1	0.026	1	1.913	1
16	-1.99 *	2	-0.038 *	1	24.192 *	2
17	-0.727 *	1	-0.041 *	1	9.687 *	1
18	-0.322	1	-0.022	1	1.927	1
19	0.497 *	1	0.032 *	1	5.204 *	1
20	1.455 *	2	0.104 *	2	51.246 *	2
21	-0.405	1	-0.031 *	1	3.642	1
22	-0.033	1	0	1	0	1
23	1.368 *	2	0.085 *	2	40.692 *	2
24	-0.064	1	-0.006	1	0.044	1
25	0.386	1	0.006	1	0.691	1
26	-0.006	1	0.004	1	0.009	1
27	-0.256	1	-0.023	1	1.922	1
28	0.343	1	0.018	1	2.366	1
29	-0.34	1	-0.031	1	3.038	1
30	0.463 *	1	0.029	1	4.036 *	1
31	0.179	1	0.01	1	0.557	1
32	0.875 *	1	0.079 *	2	23.337 *	2
33	-0.329	1	-0.024	1	2.2	1

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
34	-0.28		1	-0.013		1	1.024		1
35	0.562		1	0.02	*	1	3.233		1
36	-0.104		1	-0.008		1	0.25		1
37	-0.326		1	-0.03		1	2.725		1
38	0.356		1	0.022		1	3.533		1
39	0.077		1	0		1	0.072		1
40	0.845	*	1	0.07	*	2	20.648	*	1
41	-1.168	*	2	-0.086	*	2	29.637	*	2
42	-0.651	*	1	-0.047	*	1	8.381	*	1
43	0.163		1	0.01		1	0.756		1
44	0.333		1	0.023		1	2.733		1
45	0.296		1	0.017		1	1.748		1
46	0.32		1	0.018		1	2.014		1
47	-0.525	*	1	-0.04	*	1	7.008	*	1
48	0.422		1	0.014		1	2.156		1
49	0.259		1	0.011		1	1.174		1
50	0.937	*	1	-0.073		1	3.404		1
51	-1.413	*	2	0.111	*	2	7.288	*	-2
52	0.574		1	-0.157	*	2	13.247	*	2

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Table 78.

DIF Detection by Polytomous Method: Mathematics June, 1998

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
1	-0.007		1	-0.003		1	0.930		1
2	-0.939	*	1	-0.079	*	2	22.569	*	2
3	-0.14		1	-0.014		1	2.110		1
4	0.931	*	1	0.077	*	2	17.687	*	2
5	-0.161		1	-0.018		1	1.840		1
6	0.485	*	1	0.046	*	1	4.735	*	1
7	-0.476		1	-0.03		1	7.669	*	1
8	0.1		1	0.012		1	0.049		1
9	-0.193		1	-0.016		1	2.162		1
10	-0.266		1	-0.025		1	2.284		1
11	0.123		1	0.016		1	0.206		1
12	0.28		1	0.011		1	0.022		1
13	0.427		1	0.046	*	1	3.026		1
14	-0.442		1	-0.024		1	6.028	*	1
15	0.364		1	0.04		1	2.852		1
16	-0.484		1	-0.025		1	5.439	*	1
17	0.684	*	1	0.061	*	2	4.704	*	1
18	0.036		1	0.002		1	0.311		1
19	-0.049		1	-0.005		1	1.235		1
20	0.147		1	0.015		1	0.154		1
21	0.616		1	0.02		1	0.842		1
22	0.202		1	0.015		1	0.138		1
23	0.533	*	1	0.047	*	1	4.952	*	1
24	0.335		1	0.023		1	1.155		1
25	0.817	*	1	0.045	*	1	5.556	*	1
26	-0.269		1	-0.029		1	2.545		1
27	-0.072		1	-0.006		1	0.835		1
28	0.042		1	0.003		1	0.000		1
29	-0.145		1	-0.015		1	1.992		1
30	-0.004		1	0		1	0.129		1
31	-0.477	*	1	-0.036		1	5.503	*	1
32	2.157	*	2	0.2	*	2	81.827	*	1
33	1.27	*	2	0.076	*	2	16.484	*	1

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
34	-0.532		1	-0.028		1	7.306	*	1
35	-0.34		1	-0.027		1	3.853	*	1
36	1.138	*	2	0.081	*	2	15.060	*	2
37	0.236		1	0.024		1	0.573		1
38	-0.467		1	-0.031		1	5.101	*	1
39	0.921	*	1	0.072	*	2	11.176	*	2
40	0.112		1	0.003		1	0.628		1
41	0.486		1	0.016		1	0.849		1
42	-0.028		1	-0.001		1	0.200		1
43	-0.279		1	-0.022		1	3.209		1
44	-0.05		1	-0.005		1	1.097		1
45	0.701	*	1	0.043	*	1	5.966	*	2
46	-0.633	*	1	-0.058	*	1	14.878	*	2
47	0.247		1	0.011		1	0.003		1
48	-0.642	*	1	-0.064	*	2	13.756	*	2
49	0.478	*	1	0.035		1	1.736		1
50	-0.523	*	1	-0.322	*	2	37.109	*	2
51	-0.223	*	1	-0.074		1	5.776	*	1
52	-0.272	*	1	-0.116	*	2	9.938	*	1

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Table 79.

DIF Detection by Polytomous Method: Mathematics 1212, 1998

	GMH		Poly-SIB		LDF	
Item	Δ	DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
1	0.239	1	0.001	1	0.192	1
2	-2.295	2	-0.083 *	2	11.737 *	1
3	-0.056	1	-0.017	1	0.329	1
4	-0.683	1	-0.073	1	3.18	1
5	0.457	1	0.023	1	0.802	1
6	-0.528	1	-0.034	1	0.912	1
7	-0.571	1	-0.042	1	1.949	1
8	0.168	1	0.016	1	0.307	1
9	0.851	1	0.083 *	2	5.547 *	2
10	-0.312	1	-0.034	1	1.257	1
11	0.068	1	-0.008	1	0.028	1
12	-0.575	1	-0.042	1	2.666	1
13	-0.249	1	-0.035	1	0.83	1
14	1.019	1	0.043	1	3.781	1
15	0.208	1	0.013	1	0.575	1
16	-1.619 *	2	-0.028	1	3.658	1
17	-1.544 *	2	-0.07 *	2	6.998 *	1
18	-0.767	1	-0.054	1	1.79	1
19	0.81	1	0.055	1	4.638 *	1
20	1.719 *	2	0.131 *	2	14.839 *	2
21	-0.816	1	-0.038	1	1.876	1
22	0.047	1	-0.011	1	0.001	1
23	1.113	1	0.08 *	2	6.609 *	1
24	-0.112	1	0.003	1	0.016	1
25	-0.347	1	-0.004	1	0.002	1
26	-0.617	1	-0.038	1	1.318	1
27	-0.828	1	-0.07 *	2	2.732	1
28	0.88	1	0.058	1	3.205	1
29	-0.552	1	-0.051	1	0.709	1
30	1.028 *	1	0.095 *	2	6.112 *	1
31	0.734	1	0.081 *	2	3.534	1
32	0.84	1	0.08 *	2	5.231 *	1
33	-0.515	1	-0.046	1	1.62	1

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
34	-0.289		1	-0.019		1	0.18		1
35	0.838		1	0.025		1	0.665		1
36	-0.238		1	-0.004		1	0.485		1
37	-0.275		1	-0.046		1	0.383		1
38	0.225		1	0.023		1	0.783		1
39	-0.336		1	-0.026		1	0.199		1
40	1.373	*	2	0.114	*	2	10.078	*	2
41	-1.756	*	2	-0.132	*	2	12.726	*	1
42	-0.8		1	-0.073	*	2	2.337		1
43	-0.01		1	-0.008		1	0.004		1
44	0.363		1	0.03		1	0.776		1
45	0.066		1	-0.012		1	0.012		1
46	0.499		1	0.022		1	1.495		1
47	-0.458		1	-0.054		1	1.525		1
48	0.24		1	0.005		1	0.442		1
49	0.252		1	0.008		1	0.433		1
50	0.05		1	-0.019		1	0.25		1
51	-3.553	*	2	0.152		1	2.987		1
52	0.953		1	-0.091		1	2.072		1

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Table 80.

DIF Detection by Polytomous Method: Mathematics 2112, 1998

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
1	-0.863		1	-0.066		1	3.377		1
2	-0.81		1	-0.019		1	1.128		1
3	-0.614		1	-0.013		1	0.481		1
4	-0.802		1	-0.058		1	2.283		1
5	0.467		1	0.061		1	1.162		1
6	-0.051		1	-0.017		1	0.11		1
7	0.263		1	0.046		1	0.418		1
8	0.157		1	0.028		1	0.2		1
9	-0.196		1	-0.005		1	0.11		1
10	-1.007	*	2	-0.087	*	2	5.183	*	1
11	0.745		1	0.037		1	1.431		1
12	0.005		1	-0.005		1	0.01		1
13	-0.252		1	-0.017		1	0.469		1
14	0.791		1	0.035		1	1.793		1
15	0.221		1	0.014		1	0.542		1
16	-3.213	*	2	-0.042	*	1	14.511	*	1
17	-0.315		1	-0.026		1	0.882		1
18	0.638		1	0.041		1	1.313		1
19	-0.354		1	-0.017		1	0.12		1
20	0.987	*	1	0.064		1	4.62	*	1
21	0.588		1	0.044		1	1.99		1
22	0.016		1	0.006		1	0.101		1
23	1.266	*	2	0.076	*	2	9.196	*	1
24	0.65		1	0.038		1	1.829		1
25	-0.081		1	0.013		1	0.041		1
26	0.196		1	0.053		1	0.391		1
27	-0.464		1	-0.041		1	1.018		1
28	-0.641		1	-0.032		1	1.278		1
29	-0.445		1	-0.033		1	1.643		1
30	0.055		1	0.027		1	0.438		1
31	-0.047		1	0.006		1	0.017		1
32	0.638		1	0.07		1	3.621		1
33	-0.212		1	-0.006		1	0.005		1

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
34	0.343		1	0.006		1	0.209		1
35	1.21		1	0.044	*	1	3.648		1
36	-0.2		1	-0.011		1	0.189		1
37	-1.019	*	2	-0.07		1	3.909	*	1
38	0.499		1	0.028		1	0.987		1
39	0.895		1	0.065	*	2	2.886		1
40	0.677		1	0.085	*	2	2.333		1
41	-0.953	*	1	-0.08	*	2	5.67	*	1
42	-0.361		1	-0.015		1	0.958		1
43	0.031		1	0.023		1	0.022		1
44	0.345		1	0.03		1	0.678		1
45	0.449		1	0.035		1	1.343		1
46	-0.056		1	0.006		1	0.141		1
47	-0.491		1	-0.036		1	0.927		1
48	-0.194		1	-0.031		1	0.451		1
49	0.232		1	0.015		1	0.096		1
50	0.071		1	-0.045		1	0.039		1
51	0.289		1	0.25	*	2	2.859		1
52	-0.282		1	-0.217	*	2	5.517	*	2

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Table 81.

DIF Detection by Polytomous Method: Mathematics 2211, 1998

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
1	-0.253		1	0.013		1	0.001		1
2	-1.372	*	2	-0.017		1	0.254		1
3	-0.232		1	-0.011		1	0.001		1
4	-0.561	*	1	-0.047		1	1.53		1
5	0.286		1	0.057		1	2.008		1
6	-0.095		1	0.039		1	0.697		1
7	0.05		1	0.029		1	0.143		1
8	0.134		1	-0.021		1	0.098		1
9	0.053		1	-0.024		1	0.008		1
10	-0.635	*	1	-0.036		1	1.61		1
11	0.486	*	1	0.016		1	0.035		1
12	-0.094		1	-0.026		1	0.382		1
13	-0.678	*	1	-0.156	*	2	8.789	*	1
14	0.723	*	1	0.028		1	0.23		1
15	0.197		1	0.005		1	0.002		1
16	-2.164	*	2	0.008		1	1.268		1
17	-0.778	*	1	-0.028		1	2.369		1
18	-0.286		1	-0.039		1	1.119		1
19	0.427		1	0.035		1	0.37		1
20	1.445	*	2	0.124	*	2	13.618	*	1
21	-0.463	*	1	-0.122	*	2	5.336	*	1
22	-0.013		1	-0.044		1	0.011		1
23	1.405	*	2	0.072	*	2	6.214	*	1
24	-0.084		1	-0.058		1	3.194		1
25	0.248		1	0.015		1	3.533		1
26	0.007		1	-0.038		1	0.887		1
27	-0.333		1	-0.037		1	0.683		1
28	0.384		1	-0.008		1	0.011		1
29	-0.419	*	1	-0.028		1	0.078		1
30	0.384		1	0.056		1	1.734		1
31	0.112		1	0.017		1	0.386		1
32	0.855	*	1	0.087	*	2	4.001	*	1
33	-0.336		1	-0.099	*	2	6.866	*	1

	GMH			Poly-SIB			LDF		
Item	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
34	-0.286		1	-0.018		1	0.923		1
35	0.577		1	0.011		1	0.375		1
36	-0.127		1	0.012		1	0.038		1
37	-0.372		1	0.035		1	0.454		1
38	0.35		1	0.046		1	0.495		1
39	0.042		1	0.013		1	1.193		1
40	0.834	*	1	0.057		1	2.637		1
41	-1.207	*	2	-0.003		1	0.116		1
42	-0.713	*	1	-0.077	*	2	3.771		1
43	0.144		1	0.036		1	1.429		1
44	0.301		1	0.015		1	0.922		1
45	0.284		1	0.018		1	0.064		1
46	0.332		1	0.051		1	4.045	*	1
47	-0.561	*	1	-0.055		1	2.134		1
48	0.329		1	0.015		1	0.307		1
49	0.213		1	0.014		1	0.967		1
50	-0.092		1	-0.134		1	1.945		1
51	0.187	*	1	0.152		1	2.846		1
52	-0.197	*	1	-0.171		1	1.708		1

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Table 82.

<u>DIF Detection by Polytomous Method: Mathematics 2222, 1998</u>

	GMH		Poly-SIB		LDF	
Item	Δ	DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
1	0.704	1	0.076	1	1.238	1
2	-1.07	1	-0.053	1	4.006 *	1
3	-0.652	1	-0.061	1	1.668	1
4	0.764	1	0.057	1	2.173	1
5	-0.493	1	-0.067	1	0.481	1
6	1.113 *	2	0.108	1	5.457 *	1
7	-0.006	1	-0.004	1	0.183	1
8	0.814	1	0.057	1	0.655	1
9	-0.065	1	-0.021	1	0.005	1
10	-0.444	1	-0.026	1	0.067	1
11	-0.367	1	-0.073	1	0.001	1
12	0.411	1	0.014	1	0.17	1
13	0.337	1	0.066	1	0.664	1
14	-0.526	1	-0.024	1	1.08	1
15	1.404 *	2	0.165 *	2	11.95 *	2
16	-0.099	1	0.016	1	0.001	1
17	0.475	1	0.05	1	0.901	1
18	0.382	1	-0.021	1	0.031	1
19	0.978	1	0.033	1	0.352	1
20	0.103	1	0.001	1	0.006	1
21	1.225	1	0.049	1	1.439	1
22	0.886	1	0.07	1	1.251	1
23	0.614	1	0.042	1	2.746	1
24	-0.207	1	-0.019	1	0.045	1
25	0.501	1	0.067	1	0.446	1
26	-0.295	1	0.001	1	0.128	1
27	0.33	1	0.015	1	0.12	1
28	-0.583	1	-0.044	1	0.544	1
29	-0.418	1	-0.08	1	0.313	1
30	0.067	1	0.038	1	0.307	1
31	-0.472	1	-0.054	1	0.881	1
32	2.15 *	2	0.165 *	2	11.14 *	2
33	-0.548	1	0.01	1	0.082	1

	GMH		Poly-SIB		LDF	
Item	Δ	DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
34	-0.602	1	-0.027	1	4.371 *	1
35	-1.227 *	2	-0.074	1	3.847 *	1
36	1.936 *	2	0.152 *	2	10.09 *	1
37	0.336	1	0.022	1	1.007	1
38	-0.703	1	-0.031	1	0.833	1
39	0.764	1	0.029	1	0.244	1
40	-0.24	1	-0.028	1	1.166	1
41	0.18	1	0.045	1	0.804	1
42	-0.365	1	-0.022	1	0.003	1
43	0.203	1	-0.052	1	0.194	1
44	0.688	1	0.039	1	1.202	1
45	0.258	1	0.02	1	1.123	1
46	-0.063	1	-0.001	1	0.223	1
47	0.777	1	0.025	1	0.015	1
48	-0.938	1	-0.079	1	3.639	1
49	-0.072	1	0.008	1	0.046	1
50	-0.653 *	1	-0.415 *	2	9.56 *	1
51	-0.241	1	-0.018	1	0.024	1
52	-0.203	1	-0.05	1	1.324	1

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Table 83.

DIF Detection by Polytomous Method: Biology January, 1998

Item	GMH			Poly-SIB			LDF		
	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
1	0.603	*	1	0.038	*	1	6.669	*	1
2	0.459		1	0.028		1	4.542	*	1
3	-0.005		1	-0.007		1	0.002		1
4	-0.637	*	1	-0.031	*	1	4.028	*	1
5	0.085		1	0.002		1	0.135		1
6	0.333		1	0.02		1	2.319		1
7	-0.152		1	-0.019		1	0.577		1
8	0.202		1	0.006		1	0.64		1
9	0.373		1	0.015		1	1.269		1
10	0.679	*	1	0.044	*	1	9.529	*	1
11	-0.042		1	-0.009		1	0.072		1
12	0.835	*	1	0.035	*	1	7.326	*	1
13	-0.171		1	-0.015		1	0.541		1
14	-0.938	*	1	-0.053	*	1	11.84	*	2
15	-0.222		1	-0.018		1	1.062		1
16	-0.689	*	1	-0.039	*	1	6.44	*	1
17	-0.278		1	-0.032		1	2.175		1
18	0.677	*	1	0.024	*	1	4.406	*	1
19	1.214	*	2	0.059	*	2	20.56	*	2
20	-0.076		1	-0.005		1	0.126		1
21	-0.225		1	-0.016		1	0.738		1
22	0.462		1	0.013		1	2.004		1
23	0.855	*	1	0.079	*	2	17.65	*	2
24	-0.012		1	-0.007		1	0.042		1
25	-0.605	*	1	-0.035	*	1	4.479	*	1
26	-0.117		1	-0.006		1	0.172		1
27	1.042	*	2	0.082	*	2	24.19	*	2
28	0.039		1	-0.005		1	0.019		1
29	-1.057	*	2	-0.071	*	2	18.02	*	2
30	-0.054		1	-0.016		1	0.304		1
31	0.661	*	1	0.046	*	1	10.33	*	1
32	0.482		1	0.008		1	0.504		1
33	1.092	*	2	0.058	*	1	16.23	*	2

Item	GMH			Poly-SIB			LDF		
	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
34	-0.039		1	-0.009		1	0.057		1
35	0.883	*	1	0.058	*	1	15.44	*	2
36	-0.589	*	1	-0.058	*	1	9.209	*	1
37	1.062	*	2	0.04	*	1	11.33	*	2
38	-0.722	*	1	-0.07	*	2	10.84	*	1
39	0.037		1	-0.005		1	0		1
40	-0.254		1	-0.022		1	1.252		1
41	0.1		1	0		1	0.102		1
42	0.673	*	1	0.026		1	5.563	*	1
43	0.201		1	0.018		1	1.317		1
44	-0.68	*	1	-0.068	*	2	13.19	*	2
45	0.058		1	-0.003		1	0.004		1
46	0.295		1	0.022		1	2.442		1
47	0.353		1	0.012		1	1.051		1
48	-0.505	*	1	-0.052	*	1	6.664	*	1
49	1.034	*	2	0.082	*	2	22.65	*	2
50	0.354		1	0.022		1	3.257		1
51	0.599	*	1	0.033		1	5.899	*	1
52	-0.153		1	-0.013		1	0.352		1
53	-0.383		1	-0.03		1	2.602		1
54	-0.224		1	-0.024		1	0.9		1
55	0.104		1	0.006		1	0.259		1
56	0.642	*	1	0.048	*	1	11.12	*	2
57	0.104		1	-0.137	*	2	0.264		1

<u>Note.</u> * = $p < .05$ for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Table 84.

DIF Detection by Polytomous Method: Biology June, 1998

Item	GMH			Poly-SIB			LDF		
	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
1	0.034		1	0.013		1	0.007		1
2	0.637		1	0.029	*	1	5.960	*	1
3	-0.086		1	-0.001		1	0.584		1
4	-0.461	*	1	-0.021		1	3.592		1
5	0.591	*	1	0.058	*	1	10.500	*	2
6	-0.082		1	-0.003		1	0.271		1
7	0.794	*	1	0.076	*	2	17.450	*	2
8	0.246		1	0.026		1	1.441		1
9	0.702	*	1	0.042	*	1	9.018	*	1
10	-0.593	*	1	-0.036	*	1	8.629	*	1
11	-0.818	*	1	-0.034	*	1	11.500	*	1
12	-0.158		1	-0.006		1	0.508		1
13	-0.225		1	-0.009		1	1.484		1
14	-0.42	*	1	-0.027		1	4.472	*	1
15	-0.086		1	0.009		1	0.053		1
16	-0.68	*	1	-0.04	*	1	10.030	*	1
17	0.645	*	1	0.054	*	1	8.260	*	1
18	-0.314		1	-0.015		1	1.879		1
19	0.379		1	0.052	*	1	4.833	*	1
20	-0.6	*	1	-0.037	*	1	9.347	*	1
21	0.452		1	0.031	*	1	2.564		1
22	-1.006	*	2	-0.063	*	2	23.140	*	2
23	0.642	*	1	0.046	*	1	7.506	*	2
24	0.335		1	0.039	*	1	2.191		1
25	-0.23		1	-0.011		1	1.528		1
26	0.142		1	0.021		1	0.711		1
27	0.475		1	0.026	*	1	3.233		1
28	0.084		1	0.014		1	0.029		1
29	-0.79	*	1	-0.027	*	1	8.114	*	1
30	-1.002	*	2	-0.038	*	1	13.850	*	2
31	-0.28		1	0.002		1	0.135		1
32	-1.119	*	2	-0.03	*	1	12.530	*	2
33	0.053		1	0.015		1	0.271		1

Item	GMH			Poly-SIB			LDF		
	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
34	0.357		1	0.05	*	1	3.992	*	1
35	0.372		1	0.029	*	1	2.807		1
36	0.222		1	0.025		1	1.235		1
37	-0.612	*	1	-0.032	*	1	6.782	*	1
38	-0.762	*	1	-0.021		1	4.647	*	1
39	-0.281		1	-0.008		1	1.659		1
40	0.275		1	0.021		1	0.929		1
41	0.626	*	1	0.063	*	2	10.300	*	2
42	-0.907	*	1	-0.062	*	2	20.820	*	2
43	-0.063		1	0.002		1	0.176		1
44	0.633	*	1	0.061	*	2	8.609	*	2
45	0.336		1	0.038	*	1	2.846		1
46	0.42		1	0.013		1	1.452		1
47	0.472		1	0.017		1	3.438		1
48	-0.064		1	0.002		1	0.113		1
49	0.359		1	0.041	*	1	2.586		1
50	0.537	*	1	0.043	*	1	5.596	*	1
51	0.657	*	1	0.065	*	2	12.880	*	2
52	0.263		1	0.032		1	1.278		1
53	0.187		1	0.018		1	0.345		1
54	-0.614	*	1	-0.033	*	1	9.424	*	1
55	-0.243		1	-0.012		1	2.126		1
56	-0.38		1	-0.011		1	2.143		1
57	3.179	*	2	-0.455	*	2	80.230	*	2

<u>Note.</u> * = $p < .05$ for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Table 85.

DIF Detection by Polytomous Method: Biology 1212, 1998

Item	GMH			Poly-SIB			LDF		
	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
1	0.62		1	0.015		1	0.174		1
2	1.292		1	0.031		1	1.723		1
3	-0.519		1	-0.033		1	0.859		1
4	0.356		1	0.036		1	0.954		1
5	0.967	*	1	0.083	*	2	6.068	*	1
6	-0.229		1	-0.016		1	0.013		1
7	0.297		1	0.032		1	1.021		1
8	0.118		1	0.007		1	0.337		1
9	1.352	*	1	0.065	*	2	4.063	*	1
10	0.18		1	0.013		1	0.228		1
11	-0.19		1	-0.006		1	0.227		1
12	-0.278		1	-0.024		1	2.007		1
13	-0.135		1	-0.007		1	0.139		1
14	-1.189	*	2	-0.092	*	2	7.404	*	2
15	-0.117		1	-0.004		1	0.045		1
16	-1.517	*	2	-0.09	*	2	8.327	*	1
17	0.521		1	0.032		1	1.325		1
18	-0.638		1	-0.034		1	1.749		1
19	0.782		1	0.079	*	2	3.571		1
20	-0.278		1	-0.033		1	0.306		1
21	0.238		1	0.028		1	0.712		1
22	-1.282	*	2	-0.093	*	2	7.114	*	1
23	0.577		1	0.049		1	1.475		1
24	0.553		1	0.04		1	1.333		1
25	0.229		1	0.034		1	0.658		1
26	-0.357		1	-0.012		1	0.116		1
27	0.017		1	-0.01		1	0.06		1
28	-0.096		1	0.006		1	0.131		1
29	-1.166		1	-0.034		1	3.797		1
30	-1.008		1	-0.032		1	3.481		1
31	-1.116		1	-0.025		1	2.78		1
32	-0.653		1	-0.002		1	0.467		1
33	-0.684		1	-0.04		1	1.162		1

Item	GMH			Poly-SIB			LDF		
	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
34	0.388		1	0.037		1	0.903		1
35	0.061		1	0.015		1	0.234		1
36	0.08		1	0.011		1	0.157		1
37	-0.518		1	-0.011		1	0.384		1
38	-0.914		1	-0.042		1	2.715		1
39	-0.492		1	-0.01		1	0.811		1
40	0.142		1	0.035		1	0.794		1
41	0.444		1	0.024		1	0.622		1
42	-1.006	*	1	-0.077	*	2	4.527	*	1
43	-0.215		1	-0.004		1	0.484		1
44	0.507		1	0.024		1	0.527		1
45	1.185	*	2	0.074	*	2	5.142	*	1
46	-1.311		1	-0.023		1	1.798		1
47	0.318		1	0.004		1	0.13		1
48	-0.058		1	0.021		1	0.09		1
49	0.566		1	0.028		1	1.171		1
50	0.374		1	0.017		1	0.818		1
51	0.933		1	0.072	*	2	3.222		1
52	0.79		1	0.086	*	2	4.151	*	2
53	-0.208		1	-0.011		1	0.845		1
54	-0.613		1	-0.024		1	0.931		1
55	-0.8		1	-0.037		1	0.688		1
56	-0.385		1	-0.026		1	1.642		1
57	1.976		1	-0.22		2	2.692		1

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Table 86.

DIF Detection by Polytomous Method: Biology 2112, 1998

Item	GMH		Poly-SIB		LDF	
	Δ	DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
1	0.21	1	0.006	1	0.087	1
2	1.264	1	0.029	1	1.969	1
3	-0.286	1	0.026	1	0.277	1
4	-2.182	* 2	-0.107	* 2	12.686	* 1
5	-0.005	1	0.034	1	0.207	1
6	-0.544	1	-0.02	1	0.655	1
7	1.321	* 2	0.108	* 2	9.716	* 2
8	0.748	1	0.03	1	0.372	1
9	0.613	1	0.041	1	1.056	1
10	-0.544	1	-0.043	1	1.237	1
11	-0.533	1	-0.018	1	2.405	1
12	0.711	1	0.035	1	1.741	1
13	-0.196	1	-0.017	1	0.752	1
14	-0.593	1	-0.028	1	1.16	1
15	-1.052	1	-0.036	1	2.781	1
16	-0.138	1	-0.021	1	0.656	1
17	0.638	1	0.048	1	2.19	1
18	-0.131	1	-0.03	1	0.503	1
19	0.368	1	0.031	1	0.091	1
20	-0.608	1	-0.034	1	1.484	1
21	1.04	1	0.052	1	2.858	1
22	-0.183	1	-0.013	1	0.236	1
23	0.946	1	0.066	* 2	6.648	* 2
24	-0.055	1	0.023	1	0.263	1
25	0.207	1	0.012	1	0.114	1
26	0.052	1	0.018	1	0.034	1
27	-0.092	1	-0.01	1	0.061	1
28	0.071	1	-0.002	1	0.003	1
29	-1.415	1	-0.036	1	3.1	1
30	-1.245	1	-0.046	1	2.276	1
31	-1.618	1	-0.015	1	0.252	1
32	-1.119	1	-0.026	1	2.897	1
33	0.254	1	0.032	1	0.212	1

Item	GMH		Poly-SIB		LDF	
	Δ	DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
34	0.748	1	0.036	1	1.348	1
35	0.837	1	0.065	* 2	3.276	1
36	0.042	1	0.018	1	0.513	1
37	-1.034	1	-0.031	1	1.09	1
38	-0.522	1	-0.022	1	0.716	1
39	-0.573	1	-0.01	1	0.56	1
40	0.788	1	0.068	* 2	3.078	1
41	1.282	* 2	0.09	* 2	7.813	* 1
42	-1.902	* 2	-0.125	* 2	9.722	* 1
43	0.211	1	0.002	1	0.01	1
44	0.393	1	0.035	1	0.31	1
45	0.551	1	0.04	1	1.057	1
46	1.218	1	0.034	1	1.96	1
47	-0.236	1	0.001	1	0.069	1
48	0.605	1	0.046	1	0.755	1
49	0.674	1	0.049	1	1.429	1
50	0.74	1	0.051	* 1	2.505	1
51	0.695	1	0.042	1	2.25	1
52	0.263	1	0.027	1	0.124	1
53	-0.08	1	-0.001	1	0.125	1
54	-1.663	* 2	-0.083	* 2	12.119	* 1
55	-0.621	1	-0.032	1	1.056	1
56	0.716	1	0.016	1	0.312	1
57	-1.743	* 2	-0.28	* 2	32.56	* 2

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 87.

DIF Detection by Polytomous Method: Biology 2211, 1998

Item	GMH		Poly-SIB		LDF	
	Δ	DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
1	0.763	1	0.026	1	1.943	1
2	0.973	1	0.061	1	3.404	1
3	-0.858	1	-0.023	1	0.939	1
4	0.243	1	-0.007	1	0.155	1
5	1.145	1	0.063	1	2.443	1
6	0.558	1	0.044	1	1.213	1
7	-0.082	1	-0.006	1	0.186	1
8	0.675	1	0.051	1	1.205	1
9	-0.544	1	-0.025	1	0.654	1
10	1.102	1	0.064	1	6.737	* 1
11	-0.445	1	-0.017	1	1.186	1
12	2.184	* 2	0.097	* 2	9.617	* 1
13	0.919	* 1	0.108	* 2	6.067	* 1
14	-0.771	1	-0.033	1	0.695	1
15	-0.176	1	0.015	1	0.028	1
16	-1.281	* 1	-0.074	* 2	6.169	* 1
17	0.005	1	0.012	1	0.007	1
18	-0.413	1	0.002	1	0.095	1
19	1.053	1	0.052	1	1.834	1
20	1.506	1	0.014	1	0.381	1
21	-0.745	1	-0.035	1	1.394	1
22	0.453	1	0.021	1	1.155	1
23	0.635	1	0.037	1	1.773	1
24	0.319	1	0.022	1	0.003	1
25	-0.202	1	-0.027	1	1.398	1
26	0.755	1	0.031	1	0.983	1
27	1.701	* 2	0.096	* 2	8.04	* 1
28	0.352	1	0.013	1	0.234	1
29	-2.049	* 2	-0.124	* 2	16.019	* 2
30	-0.254	1	-0.021	1	0.646	1
31	0.858	1	0.062	1	2.859	1
32	-0.127	1	0.002	1	0.103	1
33	1.415	* 2	0.081	* 2	6.68	* 1

Item	GMH		Poly-SIB		LDF	
	Δ	DIF	$\hat{\beta}_U$	DIF	χ^2	DIF
34	-0.472	1	-0.005	1	0.307	1
35	0.811	1	0.058	1	1.54	1
36	-0.951	* 1	-0.099	* 2	5.555	* 1
37	1.76	* 2	0.085	* 2	6.838	* 1
38	-1.288	* 2	-0.125	* 2	8.207	* 2
39	-0.378	1	-0.022	1	0.575	1
40	0.251	1	-0.002	1	0.129	1
41	0.379	1	0.031	1	0.28	1
42	0.314	1	0.027	1	2.507	1
43	-0.625	1	-0.049	1	1.735	1
44	-0.189	1	-0.034	1	0.889	1
45	1.204	1	0.052	* 2	3.74	1
46	0.252	1	0.012	1	0.011	1
47	0.085	1	0.01	1	0.177	1
48	-0.68	1	-0.049	1	0.724	1
49	0.587	1	0.039	1	1.829	1
50	0.117	1	0.017	1	0.377	1
51	0.126	1	-0.006	1	0.028	1
52	0.341	1	0.021	1	0.174	1
53	-0.967	1	-0.065	1	3.155	1
54	-0.588	1	-0.041	1	1.722	1
55	-0.111	1	0	1	0.158	1
56	0.018	1	0.021	1	1.214	1
57	-0.423	1	-0.098	2	2.768	1

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Table 88.

DIF Detection by Polytomous Method: Biology 2222, 1998

Item	GMH			Poly-SIB			LDF		
	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
1	0.399		1	0.024		1	1.058		1
2	1.906	*	1	0.064	*	2	3.291		1
3	-0.011		1	-0.056		1	0.179		1
4	-0.399		1	-0.041		1	0.46		1
5	1.064		1	0.093	*	2	4.479	*	1
6	-0.467		1	-0.047		1	0.054		1
7	1.127		1	0.081		1	3.465		1
8	0.204		1	0.047		1	0.756		1
9	0.239		1	0.004		1	0.58		1
10	-1.058		1	-0.105	*	2	3.812		1
11	-1.669	*	2	-0.075		1	3.214		1
12	0.638		1	-0.031		1	0.041		1
13	0.254		1	0.035		1	0.481		1
14	-0.195		1	-0.024		1	0.083		1
15	0.112		1	0.004		1	0.551		1
16	-0.091		1	-0.077		1	0.495		1
17	1.07		1	0.033		1	2.597		1
18	-0.639		1	-0.032		1	0.335		1
19	-0.455		1	0.002		1	0.147		1
20	-0.885		1	-0.038		1	3.394		1
21	0.076		1	0.005		1	0		1
22	-0.683		1	-0.032		1	0.811		1
23	1.723	*	2	0.088	*	2	7.238	*	1
24	0.103		1	0.012		1	0.001		1
25	-0.151		1	-0.059		1	1.83		1
26	-0.549		1	0.004		1	0.023		1
27	0.811		1	0.017		1	1.112		1
28	0.037		1	-0.035		1	0.535		1
29	-0.23		1	-0.018		1	0.014		1
30	-0.673		1	-0.058		1	0.313		1
31	1.042		1	0.023		1	2.896		1
32	-1.047		1	-0.063	*	2	4.665	*	1
33	0.143		1	0.051		1	1.066		1

Item	GMH			Poly-SIB			LDF		
	Δ		DIF	$\hat{\beta}_U$		DIF	χ^2		DIF
34	-0.067		1	-0.003		1	0.01		1
35	0.559		1	0.022		1	0.296		1
36	0.978		1	0.029		1	1.5		1
37	-0.501		1	-0.029		1	1.12		1
38	-0.952		1	-0.033		1	1.182		1
39	-0.276		1	0.004		1	0.196		1
40	0.657		1	0.03		1	0.391		1
41	0.196		1	0.045		1	1.261		1
42	0.074		1	0		1	0.099		1
43	-0.212		1	0.005		1	0.742		1
44	1.181	*	2	0.128	*	2	2.707		1
45	1.243	*	2	0.099	*	2	4.437	*	1
46	1.57		1	0.023		1	4.326	*	1
47	0.672		1	0.026		1	3.211		1
48	0.127		1	0.019		1	0.014		1
49	-0.105		1	0.017		1	0.025		1
50	0.752		1	0.088	*	2	3.493		1
51	0.319		1	0.022		1	0.821		1
52	-0.671		1	-0.038		1	0.6		1
53	1.142		1	0.084	*	2	3.655		1
54	0.262		1	0.006		1	0.053		1
55	0.166		1	0.042		1	0.357		1
56	-0.12		1	-0.002		1	0		1
57	-0.87	*	1	-0.268	*	2	9.716	*	1

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF;

2= moderate to severe DIF.

Appendix D

GMH Results: Dichotomous Items Analyzed with Combined Dichotomous and Polytomous Total Test Score

Table 89.

GMH Results: Dichotomous Items with Combined Dichotomous and Polytomous Test Score, English 1998

	January			June			1212			2112			2211			2222		
	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF
1	-1.619	*	2	-0.622	*	1	-1.924	*	2	1.787		1	-0.505		1	-0.148		1
2	-0.21		1	-0.25	*	1	-1.233	*	2	-0.323		1	-0.985		1	-0.016		1
3	-0.054		1	0.992	*	1	0.193		1	1.474	*	2	1.026	*	2	1.185	*	2
4	-0.332		1	0.613	*	1	-0.166		1	0.486		1	0.745		1	0.86		1
5	-0.108		1	0.341	*	1	0.69		1	0.599		1	0.289		1	-0.034		1
6	-0.213		1	0.385	*	1	0.091		1	0.749		1	0.128		1	0.025		1
7	-0.333		1	-0.45	*	1	-0.991		1	-0.891		1	-0.034		1	-1.139		1
8	-0.076		1	-0.169	*	1	-0.595		1	-0.796		1	0.335		1	-0.036		1
9	-0.144		1	-0.661	*	1	-0.213		1	-0.025		1	-0.76		1	-0.165		1
10	-0.016		1	0.03	*	1	-0.129		1	0.37		1	-0.008		1	0.214		1
11	-0.511		1	-0.39	*	1	-0.746		1	-0.183		1	0.759		1	-0.392		1
12	-1.481	*	2	-0.178	*	1	-1.315		1	-0.406		1	-0.054		1	-0.232		1
13	-0.65	*	1	0.807	*	1	-0.953		1	1.207	*	2	0.591		1	0.413		1
14	0.561	*	1	-0.082	*	1	-0.34		1	0.758		1	-1.839		2	0.772		1
15	0.968	*	1	0.886	*	1	0.942	*	1	1.252		1	-0.366		1	2.846	*	2
16	0.62	*	1	0.489	*	1	0.38		1	0.753		1	1.156		1	0.57		1
17	1.164	*	2	0.623	*	1	1.286	*	2	1.553	*	2	0.927		1	0.065		1
18	0.979	*	1	0.07	*	1	1.377	*	2	-0.675		1	0.462		1	0.895		1
19	-0.236		1	-0.055	*	1	-0.287		1	-0.078		1	0.163		1	0.291		1
20	-0.56	*	1	-0.957	*	1	-0.509		1	-0.686		1	-0.827		1	-0.479		1
21	-0.014		1	-1.394	*	2	-0.035		1	-1.076	*	2	-1.186	*	2	-1.343	*	2

	January			June			1212			2112			2211			2222		
	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF
22	-0.848	*	1	-0.209	*	1	-0.751		1	0.303		1	-0.387		1	0.298		1
23	0.462	*	1	-0.236	*	1	0.41		1	0.006		1	0.084		1	0.088		1
24	-0.041		1	-0.124	*	1	-0.008		1	1.215	*	2	-1.23		1	0.24		1
25	0.317		1	0.223	*	1	-0.386		1	0.166		1	-0.663		1	0.807		1
26	-0.762	*	1	-0.064	*	1	-0.38		1	0.492		1	-0.989		1	0.904		1
27	0.991	*	1	0.145	*	1	0.961		1	-0.185		1	0.72		1	-0.318		1
28	0.505	*	1	1.211	*	2	1.078	*	2	0.444		1	0.853		1	2.367	*	2
29	0.516		1	0.441	*	1	-0.389		1	0.2		1	1.213	*	2	-0.588		1
30	0.425	*	1	0.671	*	1	0.044		1	0.463		1	0.799		1	0.744		1
31	-0.402		1	-0.669	*	1	-0.858		1	-0.275		1	-0.997		1	-0.269		1
32	0.498	*	1	1.278	*	2	0.753		1	0.794		1	1.467		1	0.839		1
33	0.012		1	-0.052	*	1	0.36		1	-0.959		1	-0.821		1	0.038		1
34	0.164		1	0.3	*	1	0.072		1	0.874		1	-0.286		1	-0.487		1
35	1.082	*	2	0.073	*	1	1.652	*	2	-0.829		1	-0.49		1	0.658		1
36	-0.006		1	0.081	*	1	0.247		1	-0.13		1	0.139		1	0.124		1
37	1.078	*	2	-0.372	*	1	1.023	*	1	0.031		1	-0.595		1	-0.901		1
38	0.645	*	1	0.815	*	1	0.38		1	-0.065		1	1.609		1	1.732	*	2
39	1.12	*	2	0.315	*	1	1.295	*	2	0.568		1	-0.62		1	0.176		1
40	1.604	*	2	0.332	*	1	2.232	*	2	0.647		1	1.289	*	2	-0.261		1
41	-0.111		1	0.051	*	1	0.228		1	0.271		1	-0.362		1	-0.068		1
42	1.497	*	2	-0.114	*	1	1.951	*	2	-0.607		1	1.171		1	0.076		1
43	-0.199		1	-0.001	*	1	-0.307		1	-0.133		1	-0.026		1	-0.664		1
44	-1.096	*	2	-0.192	*	1	-0.867		1	-0.203		1	-0.845		1	-0.75		1
45	0.084		1	0.011	*	1	0.149		1	-0.156		1	-0.114		1	-0.838		1
46	-0.064		1	0.355	*	1	0.479		1	-0.017		1	-0.088		1	0.966		1

	January			June			1212			2112			2211			2222		
	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF
47	-0.453		1	0.669	*	1	-0.706		1	0.85		1	0.85		1	0.954		1
48	-1.245	*	2	0.404	*	1	-1.062	*	2	-0.248		1	0.026		1	0.084		1
49	0.106		1	-0.878	*	1	0.576		1	-1.424	*	2	-1.755	*	2	-0.571		1
50	-1.099	*	2	1.177	*	2	-1.02		1	1.31	*	2	0.651		1	1.838	*	2
51	-0.698	*	1	0.042	*	1	-0.503		1	-0.166		1	-0.249		1	-0.615		1
52	-1.016	*	2	0.366	*	1	-0.628		1	0.754		1	0.157		1	-0.504		1
53	0.493		1	0.379	*	1	1.354	*	2	-0.118		1	0.558		1	-0.962		1
54	0.109		1	-0.786	*	1	-0.668		1	-0.693		1	-0.082		1	-1.893	*	2
55	0.164		1	-0.119	*	1	0.414		1	-0.547		1	0.507		1	0.979		1
56	-1.412	*	2	-0.222	*	1	-0.631		1	0.317		1	-0.84		1	0.24		1
57	0.537	*	1	-0.346	*	1	0.968	*	1	-2.215	*	2	1.434		1	-1.06		1
58	0.108		1	-0.602	*	1	-0.55		1	-0.685		1	-0.535		1	-1.074		1
59	-0.089		1	0.183	*	1	0.117		1	-0.426		1	0.613		1	0.302		1
60	0.644	*	1	-0.038	*	1	0.216		1	-0.604		1	0.839		1	-0.892		1
61	0.413		1	-0.166	*	1	0.322		1	-0.215		1	-0.132		1	0.444		1
62	0.703	*	1	-0.094	*	1	-0.281		1	-0.356		1	0.044		1	-0.677		1
63	-0.438		1	0.185	*	1	0.085		1	0.827		1	0.427		1	0.535		1
64	0.158		1	0.02	*	1	0.247		1	0.003		1	-0.149		1	0.39		1
65	-0.33		1	-0.047	*	1	-0.368		1	-0.117		1	0.23		1	-0.532		1
66	0.734	*	1	-0.024	*	1	1.019	*	2	-0.24		1	0.503		1	-0.021		1
67	-0.362		1	0.153	*	1	-0.415		1	0.999	*	1	0.562		1	0.309		1
68	-0.022		1	-0.227	*	1	-0.515		1	0.15		1	0.03		1	-0.268		1
69	0.262		1	0.238	*	1	0.586		1	-0.442		1	0.936	*	1	0.449		1
70	-0.821	*	1	-0.07	*	1	-0.401		1	0.228		1	0.494		1	-1.322	*	2

Table 90

GMH Results: Dichotomous Items with Combined Dichotomous and Polytomous Test Score, Social Studies 1998

	January			June			1212			2112			2211			2222		
	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF
1	0.109		1	0.342		1	0.115		1	0.582		1	1.443		1	1.113		1
2	-0.235		1	-0.458	*	1	0.005		1	-0.298		1	-0.307		1	0.358		1
3	0.554		1	0.162		1	-0.337		1	0.719		1	0.153		1	-0.872		1
4	-0.222		1	0.538	*	1	0.041		1	0.011		1	0.642		1	1.042		1
5	0.947	*	1	1.35	*	2	1.084	*	2	1.318	*	2	2.128	*	2	1.356	*	2
6	0.052		1	-0.559	*	1	-1.017	*	2	0.25		1	-0.114		1	-0.361		1
7	-0.436		1	0.164		1	0.41		1	-0.255		1	-1.784	*	2	1.635	*	2
8	0.532	*	1	-0.098		1	-0.17		1	0.375		1	-0.817		1	-0.135		1
9	0.142		1	0.278		1	0.633		1	0.422		1	-0.997		1	0.678		1
10	-0.363		1	-0.086		1	0.458		1	0.162		1	-0.772		1	0.016		1
11	0.193		1	1.146	*	2	0.769		1	-0.386		1	1.302	*	2	1.711	*	2
12	0.109		1	0.179		1	0.069		1	-0.483		1	-0.149		1	0.397		1
13	-0.49	*	1	-0.213		1	-0.175		1	-0.526		1	0.469		1	-0.794		1
14	-0.933	*	1	0.239		1	0.272		1	-1.257	*	2	-0.048		1	1.091		1
15	-0.398		1	-0.742	*	1	-0.819		1	-0.159		1	-1.84	*	2	-1.217		2
16	-0.467	*	1	-0.272		1	-0.855		1	-0.586		1	-0.979		1	0.535		1
17	-0.101		1	-0.001		1	0.492		1	-1.108		1	-0.388		1	-0.06		1
18	0.428		1	0.069		1	-0.215		1	0.565		1	-0.265		1	1.087		1
19	-0.355		1	0.311		1	-0.013		1	-0.439		1	0.43		1	0.861		1
20	-0.216		1	0.435	*	1	-1.003	*	1	-0.657		1	1.171	*	2	0.795		1
21	-0.141		1	1.167	*	2	1.75	*	2	-0.071		1	1.443	*	2	0.126		1
22	-0.28		1	0.475	*	1	0.936	*	1	-0.216		1	0.48		1	-0.074		1

	January			June			1212			2112			2211			2222		
	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF
23	-0.007		1	0.121		1	-0.685		1	-0.005		1	0.223		1	0.292		1
24	-0.309		1	-0.697	*	1	-1.192	*	2	-0.107		1	-0.715		1	-1.276	*	2
25	0.044		1	0.314		1	0.236		1	-0.407		1	-0.228		1	1.501	*	2
26	-1.063	*	2	-0.647	*	1	-0.556		1	-0.884		1	-0.755		1	-1.656	*	2
27	-0.039		1	-0.78	*	1	-0.506		1	-0.42		1	-1.623	*	2	-0.44		1
28	-0.85	*	1	-0.157		1	-0.592		1	-0.196		1	0.071		1	0.432		1
29	-0.218		1	0.054		1	-0.096		1	-0.273		1	1.003		1	-1.096		1
30	0.729	*	1	-0.03		1	0.407		1	1.04	*	2	-0.585		1	0.033		1
31	-0.069		1	-0.663	*	1	-0.722		1	-0.925	*	1	-0.494		1	-0.192		1
32	-1.676	*	2	0.193		1	-0.184		1	-1.303	*	2	0.258		1	0.762		1
33	-0.282		1	0.099		1	0.552		1	0.285		1	-0.113		1	0.205		1
34	-0.321		1	-0.705	*	1	-0.494		1	-0.136		1	-0.952		1	-0.347		1
35	-0.139		1	-0.355		1	-0.111		1	-0.541		1	-0.359		1	-0.784		1
36	0.501	*	1	0.462	*	1	1.511	*	2	0.752		1	0.143		1	0.554		1
37	0.328		1	-0.093		1	-0.437		1	-0.023		1	-0.281		1	0.179		1
38	0.208		1	-0.524	*	1	-1.079		1	-0.049		1	0.082		1	-0.712		1
39	-0.476	*	1	0.235		1	-0.165		1	-0.899	*	1	0.457		1	0.419		1
40	0.314		1	-0.653	*	1	-0.467		1	1.164	*	2	-0.172		1	-1.162	*	1
41	0.665	*	1	-0.953	*	1	-0.942		1	0.934		1	-1.428	*	2	-0.567		1
42	-0.274		1	0.559	*	1	-0.053		1	0.311		1	0.376		1	1.172		1
43	0.293		1	0.689	*	1	0.667		1	0.825		1	1.368	*	2	0.895		1
44	-0.203		1	-0.192		1	-0.3		1	-0.78		1	1.216	*	2	-0.557		1
45	0.505		1	1.371	*	2	0.869		1	-0.539		1	1.319	*	2	2.29	*	2
46	0.685	*	1	-0.232		1	-0.175		1	1.179	*	2	0.526		1	-0.977		1
47	-0.122		1	-0.527		1	-0.817		1	0.117		1	-1.298		1	-1.477		1

	January			June			1212			2112			2211			2222		
	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF
48	-0.073		1	1.2	*	2	1.426	*	2	-0.822		1	1.011	*	1	1.584	*	2
49	0.471		1	0.778	*	1	0.983		1	0.218		1	1.232	*	1	1.315		1
50	-0.349		1	2.327	*	2	2.24	*	2	-0.344		1	3.392	*	2	1.969	*	2
51	0.014		1	0.176		1	0.552		1	-0.542		1	0.988		1	0.286		1
52	0.75	*	1	1.414	*	2	1.218	*	1	1.156		1	1.501	*	2	1.583	*	2
53	1.315	*	2	0.261		1	0.05		1	0.579		1	-0.953		1	0.072		1
54	-0.441		1	1.19	*	2	1.703	*	2	0.023		1	0.875		1	0.724		1
55	0.559	*	1	0.682	*	1	0.059		1	0.475		1	0.762		1	0.514		1
56	0.319		1	-0.416	*	1	-0.595		1	1.095	*	2	0.116		1	-1.226	*	1
57	0.841	*	1	0.758	*	1	1.482	*	2	0.557		1	0.462		1	0.607		1
58	0.216		1	1.048	*	2	0.634		1	0.617		1	0.634		1	0.721		1
59	-0.346		1	1.19	*	2	0.892	*	1	-0.183		1	0.889		1	2.061	*	2
60	1.474	*	2	1.474	*	2	1.842	*	2	1.491	*	2	0.858		1	1.528	*	2
61	0.678	*	1	-0.114		1	0.458		1	1.304	*	2	-0.007		1	-0.255		1
62	-0.535	*	1	1.248	*	2	2.176	*	2	-0.202		1	1.951	*	2	1.392		2
63	-0.329		1	-0.024		1	0.304		1	-0.798		1	-0.102		1	-0.479		1
64	0.291		1	0.755	*	1	0.926		1	-0.006		1	1.52	*	1	0.979		1
65	-0.498		1	0.54	*	1	0.214		1	-2.629	*	2	0.622		1	0.261		1
66	0.696	*	1	0.287		1	0.943	*	1	-0.66		1	0.379		1	-0.613		1
67	0.276		1	-0.25		1	-0.404		1	0.426		1	0.717		1	-0.687		1
68	0.994	*	1	0.311		1	1.355		1	1.297	*	2	-1.167		1	1.54		1
69	0.819	*	1	-0.274		1	-0.259		1	0.634		1	-0.808		1	-0.832		1
70	0.994	*	1	-0.073		1	0.009		1	0.973	*	1	0.513		1	0.28		1

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 91.

GMH Results: Dichotomous Items with Combined Dichotomous and Polytomous Test Score, Mathematics 1998

	January		June		1212		2112		2211		2222	
	Δ	DIF	Δ	DIF	Δ	DIF	Δ	DIF	Δ	DIF	Δ	DIF
1	-0.205	1	-0.007	1	0.239	1	-0.863	1	-0.253	1	0.704	1
2	-1.305 *	2	-0.939 *	1	-2.295	2	-0.81	1	-1.372 *	2	-1.07	1
3	-0.084	1	-0.14	1	-0.056	1	-0.614	1	-0.232	1	-0.652	1
4	-0.576 *	1	0.931 *	1	-0.683	1	-0.802	1	-0.561 *	1	0.764	1
5	0.304	1	-0.161	1	0.457	1	0.467	1	0.286	1	-0.493	1
6	-0.1	1	0.485 *	1	-0.528	1	-0.051	1	-0.095	1	1.113 *	2
7	0.067	1	-0.476	1	-0.571	1	0.263	1	0.05	1	-0.006	1
8	0.181	1	0.1	1	0.168	1	0.157	1	0.134	1	0.814	1
9	0.122	1	-0.193	1	0.851	1	-0.196	1	0.053	1	-0.065	1
10	-0.606 *	1	-0.266	1	-0.312	1	-1.007 *	2	-0.635 *	1	-0.444	1
11	0.551 *	1	0.123	1	0.068	1	0.745	1	0.486 *	1	-0.367	1
12	-0.057	1	0.28	1	-0.575	1	0.005	1	-0.094	1	0.411	1
13	-0.619 *	1	0.427	1	-0.249	1	-0.252	1	-0.678 *	1	0.337	1
14	0.829 *	1	-0.442	1	1.019	1	0.791	1	0.723 *	1	-0.526	1
15	0.245	1	0.364	1	0.208	1	0.221	1	0.197	1	1.404 *	2
16	-1.99 *	2	-0.484	1	-1.619	4	-3.213 *	2	-2.164 *	2	-0.099	1
17	-0.727 *	1	0.684 *	1	-1.544	2	-0.315	1	-0.778 *	1	0.475	1
18	-0.322	1	0.036	1	-0.767	1	0.638	1	-0.286	1	0.382	1
19	0.497 *	1	-0.049	1	0.81	1	-0.354	1	0.427	1	0.978	1
20	1.455 *	2	0.147	1	1.719	2	0.987 *	1	1.445 *	2	0.103	1
21	-0.405	1	0.616	1	-0.816	1	0.588	1	-0.463 *	1	1.225	1

	January			June			1212			2112			2211			2222		
	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF
22	-0.033		1	0.202		1	0.047		1	0.016		1	-0.013		1	0.886		1
23	1.368	*	2	0.533	*	1	1.113		1	1.266	*	2	1.405	*	2	0.614		1
24	-0.064		1	0.335		1	-0.112		1	0.65		1	-0.084		1	-0.207		1
25	0.386		1	0.817	*	1	-0.347		1	-0.081		1	0.248		1	0.501		1
26	-0.006		1	-0.269		1	-0.617		1	0.196		1	0.007		1	-0.295		1
27	-0.256		1	-0.072		1	-0.828		1	-0.464		1	-0.333		1	0.33		1
28	0.343		1	0.042		1	0.88		1	-0.641		1	0.384		1	-0.583		1
29	-0.34		1	-0.145		1	-0.552		1	-0.445		1	-0.419	*	1	-0.418		1
30	0.463	*	1	-0.004		1	1.028		1	0.055		1	0.384		1	0.067		1
31	0.179		1	-0.477	*	1	0.734		1	-0.047		1	0.112		1	-0.472		1
32	0.875	*	1	2.157	*	2	0.84		1	0.638		1	0.855	*	1	2.15	*	2
33	-0.329		1	1.27	*	2	-0.515		1	-0.212		1	-0.336		1	-0.548		1
34	-0.28		1	-0.532		1	-0.289		1	0.343		1	-0.286		1	-0.602		1
35	0.562		1	-0.34		1	0.838		1	1.21		1	0.577		1	-1.227	*	2
36	-0.104		1	1.138	*	2	-0.238		1	-0.2		1	-0.127		1	1.936	*	2
37	-0.326		1	0.236		1	-0.275		1	-1.019	*	2	-0.372		1	0.336		1
38	0.356		1	-0.467		1	0.225		1	0.499		1	0.35		1	-0.703		1
39	0.077		1	0.921	*	1	-0.336		1	0.895		1	0.042		1	0.764		1
40	0.845	*	1	0.112		1	1.373		2	0.677		1	0.834	*	1	-0.24		1
41	-1.168	*	2	0.486		1	-1.756		2	-0.953	*	1	-1.207	*	2	0.18		1
42	-0.651	*	1	-0.028		1	-0.8		1	-0.361		1	-0.713	*	1	-0.365		1
43	0.163		1	-0.279		1	-0.01		1	0.031		1	0.144		1	0.203		1
44	0.333		1	-0.05		1	0.363		1	0.345		1	0.301		1	0.688		1
45	0.296		1	0.701	*	1	0.066		1	0.449		1	0.284		1	0.258		1
46	0.32		1	-0.633	*	1	0.499		1	-0.056		1	0.332		1	-0.063		1

	January		June		1212		2112		2211		2222	
	Δ	DIF	Δ	DIF	Δ	DIF	Δ	DIF	Δ	DIF	Δ	DIF
47	-0.525 *	1	0.247	1	-0.458	1	-0.491	1	-0.561 *	1	0.777	1
48	0.422	1	-0.642 *	1	0.24	1	-0.194	1	0.329	1	-0.938	1
49	0.259	1	0.478 *	1	0.252	1	0.232	1	0.213	1	-0.072	1

Note. * = $p < .05$ for associated significance test; 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 92.

GMH Results: Dichotomous Items with Combined Dichotomous and Polytomous Test Score, Biology 1998

	January		June		1212		2112		2211		2222	
	Δ	DIF	Δ	DIF	Δ	DIF	Δ	DIF	Δ	DIF	Δ	DIF
1	0.603 *	1	0.034	1	0.62	1	0.21	1	0.763	1	0.399	1
2	0.459	1	0.637	1	1.292	1	1.264	1	0.973	1	1.906 *	1
3	-0.01	1	-0.09	1	-0.52	1	-0.29	1	-0.86	1	-0.01	1
4	-0.64 *	1	-0.46 *	1	0.356	1	-2.18 *	2	0.243	1	-0.4	1
5	0.085	1	0.591 *	1	0.967 *	1	-0.01	1	1.145	1	1.064	1
6	0.333	1	-0.08	1	-0.23	1	-0.54	1	0.558	1	-0.47	1
7	-0.15	1	0.794 *	1	0.297	1	1.321 *	2	-0.08	1	1.127	1
8	0.202	1	0.246	1	0.118	1	0.748	1	0.675	1	0.204	1
9	0.373	1	0.702 *	1	1.352 *	1	0.613	1	-0.54	1	0.239	1
10	0.679 *	1	-0.59 *	1	0.18	1	-0.54	1	1.102	1	-1.06	1
11	-0.04	1	-0.82 *	1	-0.19	1	-0.53	1	-0.45	1	-1.67 *	2
12	0.835 *	1	-0.16	1	-0.28	1	0.711	1	2.184 *	2	0.638	1
13	-0.17	1	-0.23	1	-0.14	1	-0.2	1	0.919 *	1	0.254	1
14	-0.94 *	1	-0.42 *	1	-1.19 *	2	-0.59	1	-0.77	1	-0.2	1
15	-0.22	1	-0.09	1	-0.12	1	-1.05	1	-0.18	1	0.112	1
16	-0.69 *	1	-0.68 *	1	-1.52 *	2	-0.14	1	-1.28 *	1	-0.09	1
17	-0.28	1	0.645 *	1	0.521	1	0.638	1	0.005	1	1.07	1
18	0.677 *	1	-0.31	1	-0.64	1	-0.13	1	-0.41	1	-0.64	1
19	1.214 *	2	0.379	1	0.782	1	0.368	1	1.053	1	-0.46	1
20	-0.08	1	-0.6 *	1	-0.28	1	-0.61	1	1.506	1	-0.89	1
21	-0.23	1	0.452	1	0.238	1	1.04	1	-0.75	1	0.076	1

	January			June			1212			2112			2211			2222		
	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF	Δ		DIF
22	0.462		1	-1.01	*	2	-1.28	*	2	-0.18		1	0.453		1	-0.68		1
23	0.855	*	1	0.642	*	1	0.577		1	0.946		1	0.635		1	1.723	*	2
24	-0.01		1	0.335		1	0.553		1	-0.06		1	0.319		1	0.103		1
25	-0.61	*	1	-0.23		1	0.229		1	0.207		1	-0.2		1	-0.15		1
26	-0.12		1	0.142		1	-0.36		1	0.052		1	0.755		1	-0.55		1
27	1.042	*	2	0.475		1	0.017		1	-0.09		1	1.701	*	2	0.811		1
28	0.039		1	0.084		1	-0.1		1	0.071		1	0.352		1	0.037		1
29	-1.06	*	2	-0.79	*	1	-1.17		1	-1.42		1	-2.05	*	2	-0.23		1
30	-0.05		1	-1	*	2	-1.01		1	-1.25		1	-0.25		1	-0.67		1
31	0.661	*	1	-0.28		1	-1.12		1	-1.62		1	0.858		1	1.042		1
32	0.482		1	-1.12	*	2	-0.65		1	-1.12		1	-0.13		1	-1.05		1
33	1.092	*	2	0.053		1	-0.68		1	0.254		1	1.415	*	2	0.143		1
34	-0.04		1	0.357		1	0.388		1	0.748		1	-0.47		1	-0.07		1
35	0.883	*	1	0.372		1	0.061		1	0.837		1	0.811		1	0.559		1
36	-0.59	*	1	0.222		1	0.08		1	0.042		1	-0.95	*	1	0.978		1
37	1.062	*	2	-0.61	*	1	-0.52		1	-1.03		1	1.76	*	2	-0.5		1
38	-0.72	*	1	-0.76	*	1	-0.91		1	-0.52		1	-1.29	*	2	-0.95		1
39	0.037		1	-0.28		1	-0.49		1	-0.57		1	-0.38		1	-0.28		1
40	-0.25		1	0.275		1	0.142		1	0.788		1	0.251		1	0.657		1
41	0.1		1	0.626	*	1	0.444		1	1.282	*	2	0.379		1	0.196		1
42	0.673	*	1	-0.91	*	1	-1.01	*	1	-1.9	*	2	0.314		1	0.074		1
43	0.201		1	-0.06		1	-0.22		1	0.211		1	-0.63		1	-0.21		1
44	-0.68	*	1	0.633	*	1	0.507		1	0.393		1	-0.19		1	1.181	*	2
45	0.058		1	0.336		1	1.185	*	2	0.551		1	1.204		1	1.243	*	2
46	0.295		1	0.42		1	-1.31		1	1.218		1	0.252		1	1.57		1

	January			June			1212		2112			2211		2222	
	Δ		DIF	Δ		DIF	Δ	DIF	Δ		DIF	Δ	DIF	Δ	DIF
47	0.353		1	0.472		1	0.318	1	-0.24		1	0.085	1	0.672	1
48	-0.51	*	1	-0.06		1	-0.06	1	0.605		1	-0.68	1	0.127	1
49	1.034	*	2	0.359		1	0.566	1	0.674		1	0.587	1	-0.11	1
50	0.354		1	0.537	*	1	0.374	1	0.74		1	0.117	1	0.752	1
51	0.599	*	1	0.657	*	1	0.933	1	0.695		1	0.126	1	0.319	1
52	-0.15		1	0.263		1	0.79	1	0.263		1	0.341	1	-0.67	1
53	-0.38		1	0.187		1	-0.21	1	-0.08		1	-0.97	1	1.142	1
54	-0.22		1	-0.61	*	1	-0.61	1	-1.66	*	3	-0.59	1	0.262	1
55	0.104		1	-0.24		1	-0.8	1	-0.62		1	-0.11	1	0.166	1
56	0.642	*	1	-0.38		1	-0.39	1	0.716		1	0.018	1	-0.12	1

<u>Note.</u> * = $p < .05$ for associated significance test; 1 = negligible or no DIF; 2= moderate to severe DIF.

Appendix E

Poly-SIBTEST Results for Each Examination

Table 93.

Poly-SIB Results: English

	1997			1998		
Item	$\hat{\beta}_U$		DIF	$\hat{\beta}_U$		DIF
1	0.086	*	2	-0.023		1
2	0.038		1	-0.015		1
3	0.132	*	2	0.091		1
4	0.003		1	0.03		1
5	-0.011		1	-0.014		1
6	0.028		1	-0.007		1
7	0.095	*	2	-0.099	*	2
8	0.021		1	-0.004		1
9	-0.054		1	-0.019		1
10	0.047		1	0.004		1
11	0.043		1	-0.055		1
12	0.016		1	-0.058		1
13	0.146	*	2	0.042		1
14	-0.011		1	-0.004		1
15	0.069	*	2	0.016		1
16	-0.059		1	-0.025		1
17	0.028		1	-0.028		1
18	-0.065		1	0.083		1
19	-0.053		1	-0.004		1
20	-0.004		1	-0.075		1
21	0.044		1	-0.174	*	2
22	0.094	*	2	0.047		1
23	0		1	-0.049		1
24	-0.022		1	-0.036		1
25	-0.028		1	0.057		1
26	0.103	*	2	0.003		1
27	0.001		1	-0.056		1
28	0.012		1	0.134	*	2
29	0.027		1	-0.074		1
30	-0.12	*	2	0.009		1
31	-0.023		1	0.008		1
32	-0.062		1	0.017		1
33	-0.045		1	0.008		1

	1997			1998		
Item	$\hat{\beta}_U$		DIF	$\hat{\beta}_U$		DIF
34	0.029		1	-0.012		1
35	-0.04		1	0.012		1
36	0.008		1	0.036		1
37	0.034		1	-0.085		1
38	0.056		1	0.036		1
39	-0.016		1	-0.004		1
40	-0.038		1	-0.057		1
41	-0.059		1	-0.032		1
42	-0.041		1	-0.041		1
43	0.054		1	-0.095		1
44	-0.042		1	-0.047		1
45	0.003		1	-0.079		1
46	0.071		1	0.053		1
47	0.037		1	0.081		1
48	-0.003		1	0.004		1
49	-0.114	*	2	-0.051		1
50	0.04		1	0.152	*	2
51	0.067		1	-0.04		1
52	0.049		1	-0.03		1
53	-0.145	*	2	-0.018		1
54	0.001		1	-0.21	*	2
55	0.092	*	2	0.039		1
56	0.025		1	-0.055		1
57	0.081	*	2	-0.045		1
58	0	*	1	-0.053		1
59	0.05		1	-0.013		1
60	0.091	*	2	-0.06		1
61	0.105	*	2	0.012		1
62	-0.013		1	-0.077		1
63	0.044		1	0.009		1
64	0.112	*	2	0.031		1
65	0.023		1	-0.015		1
66	0.074		1	0.021		1
67	0.082	*	2	-0.031		1
68	-0.012		1	0.007		1
69	-0.008		1	0.075		1

	1997			1998	
Item	$\hat{\beta}_U$		DIF	$\hat{\beta}_U$	DIF
70	0.035		1	-0.097	1
71	-0.037		1	0.028	1
72	-0.107		1	-0.022	1
73	-0.134		1	-0.002	1
74	-0.105		1	-0.064	1
75	-0.105		1	0.045	1
76	-0.284	*	2	-0.073	1

Note. 1 = negligible or no DIF; 2= moderate to severe DIF..

Table 94.

Poly-SIB Results: Social Studies

	1997			1998		
Item	$\hat{\beta}_U$		DIF	$\hat{\beta}_U$		DIF
1	0.049		1	0.025		1
2	-0.051		1	0.062		1
3	0.031		1	-0.011		1
4	-0.042		1	0.066		1
5	-0.035		1	0.103	*	2
6	0.001		1	-0.059		1
7	0.148	*	2	0.117	*	2
8	0.041		1	0.009		1
9	0.039		1	0.027		1
10	-0.008		1	-0.012		1
11	-0.086	*	2	0.124	*	2
12	0.065		1	0.036		1
13	0.033		1	-0.129	*	2
14	0.002		1	0.152	*	2
15	-0.005		1	-0.055		1
16	0.063		1	0.037		1
17	0.009		1	0.035		1
18	0.056		1	0.088	*	2
19	0.031		1	0.06		1
20	0.044		1	0.08		1
21	0.064		1	0.028		1
22	0.034		1	-0.007		1
23	0.109	*	2	0.038		1
24	0.056		1	-0.136	*	2
25	0.016		1	0.008		1
26	0.009		1	-0.132	*	2
27	0.065		1	-0.013		1
28	0.064		1	0.053		1
29	0.035		1	-0.041		1
30	-0.046		1	-0.028		1
31	0.104	*	2	-0.044		1
32	-0.013		1	0.076		1
33	0.003		1	0.01		1

	1997			1998		
Item	$\hat{\beta}_U$		DIF	$\hat{\beta}_U$		DIF
34	0.132	*	2	-0.057		1
35	0.02		1	-0.028		1
36	0.053		1	0.004		1
37	-0.088	*	2	0.009		1
38	-0.086		1	-0.04		1
39	-0.033		1	0.043		1
40	0.038		1	-0.056		1
41	0.014		1	-0.014		1
42	0.084	*	2	0.039		1
43	0.015		1	0.081		1
44	0.07	*	2	-0.05		1
45	0.005		1	0.166	*	2
46	-0.048		1	-0.041		1
47	0.049		1	0.001		1
48	-0.074		1	0.115	*	2
49	0.03		1	0.075		1
50	0.044		1	0.194	*	2
51	0.053		1	0.003		1
52	0.065		1	0.088	*	2
53	0.031		1	0.016		1
54	0.022		1	0.068		1
55	-0.001		1	0.055		1
56	0.036		1	-0.076		1
57	-0.002		1	-0.002		1
58	-0.004		1	0.093	*	2
59	0.049		1	0.123	*	2
60	0.048		1	0.128	*	2
61	0.043		1	0.015		1
62	0.028		1	0.041		1
63	0.057		1	-0.047		1
64	-0.025		1	0.076	*	2
65	-0.022		1	0.058		1
66	0.042		1	-0.028		1
67	-0.067		1	-0.078	*	2
68	0.058		1	0.066	*	2
69	-0.001		1	-0.027		1

	1997			1998		
Item	$\hat{\beta}_U$		DIF	$\hat{\beta}_U$		DIF
70	0.003		1	0.016		1
71	-0.308	*	2	-0.371	*	2
72	-0.343	*	2	-0.293	*	2
73	-0.344	*	2	-0.479	*	2
74	-0.263	*	2	-0.248	*	2

Note. 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 95.

Poly-SIB Results: Mathematics

	1997			1998		
Item	$\hat{\beta}_U$		DIF	$\hat{\beta}_U$		DIF
1	0.002		1	0.076		1
2	0.08	*	2	-0.053		1
3	0.013		1	-0.061		1
4	0.042		1	0.057		1
5	-0.02		1	-0.067		1
6	-0.01		1	0.108		1
7	0.068		1	-0.004		1
8	0.047		1	0.057		1
9	-0.04		1	-0.021		1
10	-0.06		1	-0.026		1
11	0.018		1	-0.073		1
12	-0.03		1	0.014		1
13	-0.07		1	0.066		1
14	-0.09		1	-0.024		1
15	0.003		1	0.165	*	2
16	0.018		1	0.016		1
17	-0.09	*	2	0.05		1
18	-0		1	-0.021		1
19	0.09	*	2	0.033		1
20	-0.1	*	2	0.001		1
21	-0.02		1	0.049		1
22	0.055		1	0.07		1
23	0.101	*	2	0.042		1
24	-0.01		1	-0.019		1
25	-0.02		1	0.067		1
26	-0.01		1	0.001		1
27	0.001		1	0.015		1
28	0.069	*	2	-0.044		1
29	0.058		1	-0.08		1
30	-0.04		1	0.038		1
31	-0.12	*	2	-0.054		1
32	-0.09	*	2	0.165	*	2
33	-0.02		1	0.01		1

	1997			1998		
Item	$\hat{\beta}_U$		DIF	$\hat{\beta}_U$		DIF
34	0.07		1	-0.027		1
35	-0.07		1	-0.074		1
36	-0.02		1	0.152	*	2
37	0.012		1	0.022		1
38	0.039		1	-0.031		1
39	-0.05		1	0.029		1
40	-0.08		1	-0.028		1
41	-0.04		1	0.045		1
42	-0.02		1	-0.022		1
43	-0.02		1	-0.052		1
44	0.072		1	0.039		1
45	0.006		1	0.02		1
46	0.091		1	-0.001		1
47	0.094		1	0.025		1
48	0.055		1	-0.079		1
49	0.125		1	0.008		1
50	-0.08		1	-0.415	*	2
51	0.112		1	-0.018		1
52	-0.32	*	2	-0.05		1

Note. 1 = negligible or no DIF; 2= moderate to severe DIF.

Table 96.

Poly-SIB Results: Biology

Item	1997 $\hat{\beta}_U$		1997 DIF	1998 $\hat{\beta}_U$		1998 DIF
1	0.003		1	0.024		1
2	-0.02		1	0.064	*	2
3	0.117	*	2	-0.06		1
4	-0.019		1	-0.04		1
5	-0.016		1	0.093	*	2
6	-0.021		1	-0.05		1
7	-0.031		1	0.081		1
8	-0.061	*	2	0.047		1
9	0.101	*	2	0.004		1
10	-0.014		1	-0.11	*	2
11	-0.008		1	-0.08		1
12	-0.009		1	-0.03		1
13	0.068		1	0.035		1
14	0.008		1	-0.02		1
15	-0.097	*	2	0.004		i
16	0.035		1	-0.08		1
17	0.057		1	0.033		1
18	-0.031		1	-0.03		1
19	0.011		1	0.002		1
20	0.091	*	2	-0.04		1
21	-0.04		1	0.005		1
22	0.086	*	2	-0.03		1
23	-0.052		1	0.088	*	2
24	0.01		1	0.012		1
25	-0.048		1	-0.06		1
26	0.028		1	0.004		1
27	0.002		1	0.017		1
28	-0.101	*	2	-0.04		1
29	0.111	*	2	-0.02		1
30	-0.018		1	-0.06		1
31	-0.096	*	2	0.023		1
32	-0.033		1	-0.06	*	2
33	0.017		1	0.051		1

	1997			1998		
Item	$\hat{\beta}_U$		DIF	$\hat{\beta}_U$		DIF
34	-0.093	*	2	-0		1
35	0.033		1	0.022		1
36	-0.057		1	0.029		1
37	0.047		1	-0.03		1
38	-0.041		1	-0.03		1
39	-0.009		1	0.004		1
40	-0.022		1	0.03		1
41	0.064		1	0.045		1
42	-0.038		1	0		1
43	0.076		1	0.005		1
44	-0.071		1	0.128	*	2
45	-0.015		1	0.099	*	2
46	0.053		1	0.023		1
47	-0.119	*	2	0.026		1
48	-0.017		1	0.019		1
49	0.014		1	0.017		1
50	0.103	*	2	0.088	*	2
51	-0.082		1	0.022		1
52	-0.01		1	-0.04		1
53	-0.046		1	0.084	*	2
54	-0.078		1	0.006		1
55	-0.003		1	0.042		1
56	0.018		1	-0		1
57	0.072		1	-0.27	*	2

Note. 1 = negligible or no DIF; 2= moderate to severe DIF.