

ITEM PARAMETER ESTIMATION USING MARGINAL MAXIMUM LIKELIHOOD
ESTIMATION UNDER TARGETED TESTING DESIGN

by

ZIWEI ZHANG

(Under the Direction of Shiyu Wang)

ABSTRACT

Marginal maximum likelihood estimation (MMLE) is a popular item calibration approach used in the targeted testing design, where missing responses commonly exist due to the testing design structure. Previous research studies investigated the performance of single group MMLE, that is MMLE without using students' background information, and the multiple group MMLE, that is MMLE using students' background information, in some large-scale targeted testing designs. However, the practical educational settings often imply small-scale testing scenarios, that is the sample size is small or median. This research investigated the performance of these two versions of MMLE under different small-scale targeted testing designs. A series of simulation studies were conducted across a variety of sample sizes, multiple form designs, number of test forms, number of anchor items, and anchor item selection strategies. Our results shed light on the selection of item parameter estimation in different small-scale targeted testing designs.

INDEX WORDS: Item responses theory, marginal maximum likelihood estimation, targeted testing design, small-scale testing

ITME PARAMETER ESTIMATION USING MARGINAL MAXIMUM LIKELIHOOD
ESTIMATION UNDER TARGETED TESTING DESIGN

by

Ziwei Zhang

B.S., East China Normal University, China, 2017

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment
of the Requirements for the Degree

MASTER OF ARTS

ATHENS, GEORGIA

2020

© 2020

Ziwei Zhang

All Rights Reserved

ITME PARAMETER ESTIMATION USING MARGINAL MAXIMUM LIKELIHOOD
ESTIMATION UNDER TARGETED TESTING DESIGN

by

ZIWEI ZHANG

Major Professor:	Shiyu Wang
Committee:	Seock-ho Kim
	Amanda. E. Ferster

Electronic Version Approved:

Ron Walcott
Interim Dean of the Graduate School
The University of Georgia
May 2020

DEDICATION

To my family.

ACKNOWLEDGEMENTS

First of all, I want to appreciate my dear advisor, Dr. Shiyu Wang. It was under her guidance, I stepped into this field, have made progress so far and will go further into the future. She is my mentor who sets an example for me of what a scholar should be like. The precious advice she provided will benefit me all my life. I also want to appreciate my committee members, Dr. Seock-ho Kim and Dr. Amanda Ferster. Not only did they help me with my thesis, but I also learned a lot of professional skills in their classes. They have witnessed my growth and have been always there whenever I need them.

It is a lonely journey to do research. But people around me make me feel joy and warm all the time. Thank my family. It's impossible for me to make any achievement without their unconditional love and support. Thank Zifan, Yawei, and Yanyan. Their lovely company has lightened my world. 2020 has been a tough year because of Coronavirus. So many scientists are making contributions to beat the outbreak. Hopefully one day I can use my knowledge to make the world better as well.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
LIST OF TABLES	xi
LIST OF FIGURES	xii
 CHAPTER	
1 INTRODUCTION	1
2 LITERATURE REVIEW	4
Targeted Testing Design	4
Missing Data Issues in Targeted Testing Design.....	5
Item Response Theory Model.....	7
3 METHODS	3
Monte Carlo Simulation Designs.....	3
4 RESULTS	18
NEAT Design.....	18
Multiform Design.....	64
Panel Missing Design	89
Panel Chained Design	95
BIB Design.....	103
Comparison between 2-Form and 5-Form Designs	111
Comparison among Different Multiple Form Designs	112

5 DISCUSSION	118
REFERENCES	121
APPENDICES	
A1 Anchor Item Selection for 5-Form NEAT Design.....	124
A2 Item Parameters	125
B1 Number of Extreme Estimation Values for a Parameters under 5-Form NEAT Design	127
B2 Five Number Summaries for Bias for a Parameters under 5-Form NEAT Design..	128
B3 Five Number Summaries for RMSE for a Parameters under 5-Form NEAT Design	129
B4 Number of Extreme Estimation Values for b Parameters under 5-Form NEAT Design	130
B5 Five Number Summaries for Bias for b Parameters under 5-Form NEAT Design..	131
B6 Five Number Summaries for RMSE for b Parameters under 5-Form NEAT Design	132
C1 Number of Extreme Estimation Values for a Parameters under 2-Form NEAT Design	133
C2 Five Number Summaries for Bias for a Parameters under 2-Form NEAT Design..	134
C3 Bias for a Parameters under 2-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes.....	135
C4 Bias for a Parameters under 2-Form NEAT Design with a Sample Size of 200 across Different Number of Uniformly Selected Anchor Items	136

C5 Bias for a Parameters under 2-Form NEAT Design with 20 Anchor Items and a Sample Size of 200 across Different Anchor Item Selection Strategies.....	136
C6 Five Number Summaries for RMSE for a Parameters under 2-Form NEAT Design	137
C7 RMSE for a Parameters under 2-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes.....	138
C8 RMSE (from 0 to 2) for a Parameters under 2-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes	139
C9 RMSE for a Parameters under 2-Form NEAT Design with a Sample Size of 200 across Different Number of Uniformly Selected Anchor Items	140
C10 RMSE for a Parameters under 2-Form NEAT Design with 20 Anchor Items and a Sample Size of 200 across Different Anchor Item Selection Strategies.....	140
C11 Number of Extreme Estimation Values for b Parameters under 2-Form NEAT Design	141
C12 Five Number Summaries for Bias for b Parameters under 2-Form NEAT Design	142
C13 Five Number Summaries for RMSE for b Parameters under 2-Form NEAT Design	144
D1 Number of Extreme Estimation Values for a Parameters under 5-Form Multiform Design	146
D2 Five Number Summaries for Bias for a Parameters under 5-Form Multiform Design	147
D3 Five Number Summaries for RMSE for a Parameters under 5-Form Multiform Design	148

D4	Number of Extreme Estimation Values for b Parameters under 5-Form Multiform	
	Design	149
D5	Five Number Summaries for Bias for b Parameters under 5-Form Multiform	
	Design	150
D6	Five Number Summaries for RMSE for b Parameters under 5-Form Multiform	
	Design	151
E1	Number of Extreme Estimation Values for a Parameters under 5-Form Panel Missing	
	Design	152
E2	Five Number Summaries for Bias for a Parameters under 5-Form Panel Missing	
	Design	152
E3	Five Number Summaries for RMSE for a Parameters under 5-Form Panel Missing	
	Design	152
E4	Number of Extreme Estimation Values for b Parameters under 5-Form Panel Missing	
	Design	153
E5	Five Number Summaries for Bias for b Parameters under 5-Form Panel Missing	
	Design	153
E6	Five Number Summaries for RMSE for b Parameters under 5-Form Panel Missing	
	Design	153
F1	Number of Extreme Estimation Values for a Parameters under 5-Form Panel Chained	
	Design	154
F2	Five Number Summaries for Bias for a Parameters under 5-Form Panel Chained	
	Design	154

F3 Five Number Summaries for RMSE for a Parameters under 5-Form Panel Chained	
Design	154
F4 Number of Extreme Estimation Values for b Parameters under 5-Form Panel Chained	
Design	155
F5 Five Number Summaries for Bias for b Parameters under 5-Form Panel Chained	
Design	155
F6 Five Number Summaries for RMSE for b Parameters under 5-Form Panel Chained	
Design	155
G1 Number of Extreme Estimation Values for a Parameters under 5-Form BIB	
Design	156
G2 Five Number Summaries for Bias for a Parameters under 5-Form BIB Design.....	156
G3 Five Number Summaries for RMSE for a Parameters under 5-Form BIB Design..	156
G4 Number of Extreme Estimation Values for b Parameters under 5-Form BIB	
Design	157
G5 Five Number Summaries for Bias for b Parameters under 5-Form BIB Design.....	157
G6 Five Number Summaries for RMSE for b Parameters under 5-Form BIB Design..	157

LIST OF TABLES

	Page
Table 3.1: 2-Form NEAT Design	9
Table 3.2: 5-Form NEAT Design	10
Table 3.3: 5-Form Multiform Design	10
Table 3.4: 5-Form Panel Missing Design	11
Table 3.5: 5-Form Panel Chained Design.....	12
Table 3.6: 5-Form BIB Design	12
Table 3.7: Missing Percentages for Each Multiple Form Design	15

LIST OF FIGURES

	Page
Figure 3.1: Anchor Item Selection for 2-Form NEAT Design	14
Figure 3.2: Item Formation Process.....	16
Figure 4.1: Number of Extreme Estimation Values for a Parameters under 5-Form NEAT Design	19
Figure 4.2: Bias for a Parameters under 5-Form NEAT Design	22
Figure 4.3: Bias for a Parameters under 5-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes.....	23
Figure 4.4: Bias for a Parameters under 5-Form NEAT Design with a Sample Size of 100 across Different Number of Uniformly Selected Anchor Items	24
Figure 4.5: Bias for a Parameters under 5-Form NEAT Design with 20 Anchor Items and a Sample Size of 100 across Different Anchor Item Selection Strategies.....	25
Figure 4.6: RMSE for a Parameters under 5-Form NEAT Design	26
Figure 4.7: RMSE for a Parameters under 5-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes.....	27
Figure 4.8: RMSE (from 0 to 2) for a Parameters under 5-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes	28
Figure 4.9: RMSE for a Parameters under 5-Form NEAT Design with a Sample Size of 100 across Different Number of Uniformly Selected Anchor Items	29

Figure 4.10: RMSE for a Parameters under 5-Form NEAT Design with 20 Anchor Items and a Sample Size of 100 across Different Anchor Item Selection Strategies.....	30
Figure 4.11: Number of Extreme Estimation Values for b Parameters under 5-Form NEAT Design	31
Figure 4.12: Bias for b Parameters under 5-Form NEAT Design	34
Figure 4.13: Bias for b Parameters under 5-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes.....	35
Figure 4.14: Bias for b Parameters under 5-Form NEAT Design with a Sample Size of 100 across Different Number of Uniformly Selected Anchor Items	36
Figure 4.15: Bias for b Parameters under 5-Form NEAT Design with 20 Anchor Items and a Sample Size of 100 across Different Anchor Item Selection Strategies.....	37
Figure 4.16: RMSE for b Parameters under 5-Form NEAT Design	38
Figure 4.17: RMSE for b Parameters under 5-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes.....	39
Figure 4.18: RMSE (from 0 to 2) for b Parameters under 5-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes	40
Figure 4.19: RMSE for b Parameters under 5-Form NEAT Design with a Sample Size of 100 across Different Number of Uniformly Selected Anchor Items	41
Figure 4.20: RMSE (from 0 to 2) for b Parameters under 5-Form NEAT Design with a Sample Size of 100 across Different Number of Uniformly Selected Anchor Items.....	42
Figure 4.21: RMSE for b Parameters under 5-Form NEAT with 20 Anchor Items and a Sample Size of 100 across Different Anchor Item Selection Strategies.....	43

Figure 4.22: RMSE (from 0 to 2) for b Parameters under 5-Form NEAT Design with 20 Anchor Items and a Sample Size of 100 across Different Anchor Item Selection Strategies	43
Figure 5.1: Number of Extreme Estimation Values for a Parameters under 2-Form NEAT Design	44
Figure 5.2: Bias for a Parameters under 2-Form NEAT Design	47
Figure 5.3: RMSE for a Parameters under 2-Form NEAT Design	49
Figure 5.4: Number of Extreme Estimation Values for b Parameters under 2-Form NEAT Design	51
Figure 5.5: Bias for b Parameters under 2-Form NEAT Design	54
Figure 5.6: Bias for b Parameters under 2-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes.....	55
Figure 5.7: Bias for b Parameters under 2-Form NEAT Design with a Sample Size of 200 across Different Number of Uniformly Selected Anchor Items	56
Figure 5.8: Bias for b Parameters under 2-Form NEAT Design with 50 Anchor Items and a Sample Size of 200 across Different Anchor Item Selection Strategies.....	57
Figure 5.9: RMSE for b Parameters under 2-Form NEAT Design	58
Figure 5.10: RMSE for b Parameters under 2-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes.....	60
Figure 5.11: RMSE (from 0 to 2) for b Parameters under 2-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes	61
Figure 5.12: RMSE for b Parameters under 2-Form NEAT Design with a Sample Size of 200 across Different Number of Uniformly Selected Anchor Items	62

Figure 5.13: RMSE (from 0 to 2) for b Parameters under 2-Form NEAT Design with a Sample Size of 200 across Different Number of Uniformly Selected Anchor Items	63
Figure 5.14: RMSE for b Parameters under 2-Form NEAT Design with 20 Anchor Items and a Sample Size of 200 across Different Anchor Item Selection Strategies.....	64
Figure 5.15: RMSE (from 0 to 2) for b Parameters under 2-Form NEAT Design with 20 Anchor Items and a Sample Size of 200 across Different Anchor Item Selection Strategies	64
Figure 6.1: Number of Extreme Estimation Values for a Parameters under 5-Form Multiform Design	65
Figure 6.2: Bias for a Parameters under 5-Form Multiform Design	67
Figure 6.3: Bias for a Parameters under 5-Form Multiform Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes.....	69
Figure 6.4: Bias for a Parameters under 5-Form Multiform Design with a Sample Size of 100 across Different Number of Uniformly Selected Anchor Items	70
Figure 6.5: Bias for a Parameters under 5-Form Multiform Design with 20 Anchor Items and a Sample Size of 100 across Different Anchor Item Selection Strategies.....	71
Figure 6.6: RMSE for a Parameters under 5-Form Multiform Design	72
Figure 6.7: RMSE for a Parameters under 5-Form Multiform Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes.....	73
Figure 6.8: RMSE (from 0 to 2) for a Parameters under 5-Form Multiform Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes	74
Figure 6.9: RMSE for a Parameters under 5-Form Multiform Design with a Sample Size of 100 across Different Number of Uniformly Selected Anchor Items	75

Figure 6.10: RMSE for a Parameters under 5-Form Multiform Design with 20 Anchor Items and a Sample Size of 100 across Different Anchor Item Selection Strategies	76
Figure 6.11: Number of Extreme Estimation Values for b Parameters under 5-Form Multiform Design	78
Figure 6.12: Bias for b Parameters under 5-Form Multiform Design	79
Figure 6.13: Bias for b Parameters under 5-Form Multiform Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes.....	80
Figure 6.14: Bias for b Parameters under 5-Form Multiform Design with a Sample Size of 100 across Different Number of Uniformly Selected Anchor Items	81
Figure 6.15: Bias for b Parameters under 5-Form Multiform Design with 20 Anchor Items and a Sample Size of 100 across Different Anchor Item Selection Strategies.....	82
Figure 6.16: RMSE for b Parameters under 5-Form Multiform Design	83
Figure 6.17: RMSE for b Parameters under 5-Form Multiform Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes	84
Figure 6.18: RMSE (from 0 to 2) for b Parameters under 5-Form Multiform Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes	85
Figure 6.19: RMSE for b Parameters under 5-Form Multiform Design with a Sample Size of 100 across Different Number of Uniformly Selected Anchor Items	86
Figure 6.20: RMSE (from 0 to 1) for b Parameters under 5-Form Multiform Design with a Sample Size of 100 across Different Number of Uniformly Selected Anchor Items	87
Figure 6.21: RMSE for b Parameters under 5-Form Multiform Design with 20 Anchor Items and a Sample Size of 20 across Different Anchor Item Selection Strategies	88

Figure 6.22: RMSE (from 0 to 2) for b Parameters under 5-Form Multiform Design with 20 Anchor Items and a Sample Size of 20 across Different Anchor Item Selection Strategies.....	88
Figure 7.1: Bias for a Parameters under 5-Form Panel Missing Design across Different Sample Sizes	90
Figure 7.2: RMSE for a Parameters under 5-Form Panel Missing Design across Different Sample Sizes	91
Figure 7.3: Bias for b Parameters under 5-Form Panel Missing Design across Different Sample Sizes	92
Figure 7.4: RMSE for b Parameters under 5-Form Panel Missing Design across Different Sample Sizes	94
Figure 7.5: RMSE (from 0 to 2) for b Parameters under 5-Form Panel Missing Design across Different Sample Sizes	95
Figure 8.1: Bias for a Parameters under 5-Form Panel Chained Design across Different Sample Sizes	97
Figure 8.2: RMSE for a Parameters under 5-Form Panel Chained Design across Different Sample Sizes	98
Figure 8.3: RMSE (from 0 to 2) for a Parameters under 5-Form Panel Chained Design across Different Sample Sizes	99
Figure 8.4: Bias for b Parameters under 5-Form Panel Chained Design across Different Sample Sizes	101
Figure 8.5: RMSE for b Parameters under 5-Form Panel Chained Design across Different Sample Sizes	102

Figure 8.6: RMSE (from 0 to 2) for b Parameters under 5-Form Panel Chained Design across Different Sample Sizes	103
Figure 9.1: Bias for a Parameters under 5-Form BIB Design across Different Sample Sizes ...	105
Figure 9.2: RMSE for a Parameters under 5-Form BIB Design across Different Sample Sizes	106
Figure 9.3: RMSE (from 0 to 2) for a Parameters under 5-Form BIB Design across Different Sample Sizes	107
Figure 9.4: Bias for b Parameters under 5-Form BIB Design across Different Sample Size.....	109
Figure 9.5: RMSE for b Parameters under 5-Form BIB Design across Different Sample Sizes	110
Figure 9.6: RMSE (from 0 to 2) for b Parameters under 5-Form BIB Design across Different Sample Sizes	111

CHAPTER 1

INTRODUCTION

Incomplete designs, that is to administer different subsets of items to different group of examinees, are frequently used to expedite item parameter calibration efficiency within the item response theory (IRT) framework (Eggen & Verhelst, 2011). Targeted testing design is a type of an incomplete design in which the assignments of test questions to students are determined a priori on the basis of students' background information Y . The Y variable is typically related to students' abilities. For example, easier test forms are administered to lower ability groups, while more difficult forms are distributed to higher ability groups. Because each item is only administered to part of the students according to Y , the item responses from students who do not access to the item are missing in a targeted testing design. Researchers discussed whether Y should be ignored when missing data presents in marginal maximum likelihood estimation (MMLE) based on Rubin's (1976) work. They demonstrated that the background information should be considered when using MMLE to calibrate item parameters under targeted testing design (Mislevy & Sheenan, 1989; DeMars, 2002; Eggen & Verhelst, 2011).

However, limitations exist in previous studies. First of all, the sample size for each group was large, for example, 1000 examinees for each group of examinees (DeMars, 2002; Eggen & Verhelst, 2011). And the test form designs were mostly in a simple structure. For example, only two test forms were investigated (DeMars, 2002; Eggen & Verhelst, 2011). When applying the testing design to a small-scale pretesting scenario, such as a classroom testing, the sample size can be typically small. In addition, the targeted testing design has to be matched to practical

curriculum design, thus required more complicated multiple form designs. In multiple form designs, we considered the number of test forms and specific anchor designs (Hulin, Lissak, and Drasgow, 1982; Kim, & Cohen, 1998; Kolen & Brennan, 2004; Sinharay & Holland, 2006; Sinharay, Haberman, Holland, & Lewis, 2012). A motivation example from a recent study is such a case (Wang, 2018). To develop an item bank for a computerized adaptive testing (CAT) for a graduate statistic course, five test forms with different difficulty levels were developed using 148 multiple choice items. These forms were distributed to five graduate statistics courses based on the instructors' prior knowledge about the students' background. Each pair of the test forms had a set of anchor items and the sample size for each student group was smaller than 30. Besides the aforementioned two limitations, previous studies only investigated the Rasch model (DeMars, 2002; Eggen & Verhelst, 2011). However, other general IRT models, such as two-parameter logistic (2-PL) model, is widely used in reality and should be studied (Birnbaum, 1968).

In summary, the overall objective of this thesis is to investigate the performance of MMLE for IRT item parameter estimation under the 2-PL model in small-scale targeted testing scenarios. Our research question is that under small-scale targeted testing scenarios, how MMLE performs when the background information is taken or not taken into account for item calibration under the 2-PL model? This includes the following sub-questions:

(1) How different MMLE estimation procedures perform when different levels of sample size that are in small or median range are used?

(2) How different MMLE estimation procedures perform when different multiple form designs are used?

To answer these questions, Monte Carlo simulation studies were conducted. Based on simulation results, we provided recommendations on item calibration under the small-scale targeted testing scenarios, including the selection of sample sizes, multiple form designs, and MMLE approaches.

CHAPTER 2

LITERATURE REVIEW

Marginal maximum likelihood estimation (MMLE) is a popular estimation method for item calibration in item response theory (IRT) framework. When missing data exist under the targeted testing design, there are some issues to use MMLE to estimate item parameters. In this section, we provide an overview on targeted testing design, missing data issues in targeted testing design and IRT models.

TARGETED TESTING DESIGN

A targeted testing design is a type of testing design that uses background information, such as a prior knowledge about difficulties of items and abilities of examinees, to distribute items to students. With such a design, test forms are assigned aligned with examinees' abilities. For example, test forms with easy items are distributed to groups with low abilities while test forms with difficult items are distributed to groups with high abilities. Based on how to use some background information, Eggen and Verhelst (2011) introduced two types of targeted testing designs. The first one is the targeted testing with sample from one population (TTOP). In this design, the background information is only used in the assignment of items or tests to examinees but not in the sampling of examinees. The second is the targeted testing with student samples from multiple (sub) populations (TTMP). In this design, the background variable is used in the sampling of examinees as well as in the assignment of tests to examinees. Comparing these two targeted testing designs, in TTOP, a random sample was sampled from the total population; while

in TTMP, a random sample was sampled from every subpopulation. To simplify the design, our study stressed on TTOP situation.

To develop test forms for targeted testing design, several multiple form designs are considered. The first is non-equivalent group with anchor test (NEAT) design (von Davier & Wilson, 2007; von Davier, 2010). This design includes multiple test forms and an anchor item set X is selected for each pair of form. The second is multiform design (Pokropek, 2011; Little & Rhemtulla, 2013; Rhemtulla & Hancock, 2016). This design includes three or more test forms and each form is composed with an anchor item set X and another two item sets that are overlapped with other two test forms. The third is panel missing design (Graham and his colleagues, 2001; Pokropek, 2011). In this design, each test form leaves one item set to be untouched except one test form is composed with all item sets. The fourth one is panel chained design (Pokropek, 2011). Within this design, two of the test forms include the easiest and the hardest item set respectively, and the rest of the test forms is composed with two adjacent item sets. The last one is balanced incomplete blocks (BIB) design (Campbell, Sengupta, Santos, & Lorig, 1995; Pokropek, 2011). This design has the same number of test forms and item sets. The number of item sets in each test form is the same. The number of test forms that each item set is assigned to is the same. These five designs will be investigated in the later simulation section on the impact of item parameter estimation.

MISSING DATA ISSUES IN TARGETED TESTING DESIGN

In a targeted testing design, because each item is only administered to part of the students according to background information, the item responses from students who do not access to the item are missing. The exist of missing data affects the process of item calibration. Rubin (1976) described missingness as a probabilistic phenomenon and is led by the “process that causes

missing data”. He proposed an ignorability principle about making inferences when missing data exist. In his work, the data is called missing at random (MAR) when probability of missing data is related to an observed variable but unrelated to the unobserved variable. Rubin (1976) mathematically proved that when making direct-likelihood or Bayesian inferences about the parameter of the data, it is appropriate to ignore the process that causes missing data if the missing data is MAR, and the parameter of the missing data process is distinct from the parameter of the data.

Based on Rubin’s work, Mislevy and Sheenan (1989) further mathematically proved that in targeted testing design, when the background information Y is used in both examinee and item sampling, MAR does not hold because the value of Y for each examinee, which is unobserved, determines which items he or she would be assigned. MMLE is a direct-likelihood estimation method to estimate item parameters when missing data present. They concluded that background information shouldn’t be ignored when calibrate items through MMLE in this situation.

Eggen and Verhelst (2011) demonstrated this conclusion is true with the case of TTOP. That is, because the assignment of items is depending on the values of Y , ability distribution will be different if background information Y is not used. Thus, MAR is not fulfilled in TTOP as well. They stated that in TTOP, background information should be considered when calibrate items through MMLE. To prove their conjecture, Eggen & Verhelst (2011) carried out a simulation study. They had two groups of 1000 students which were drawn from $N(-1,1)$ and $N(1,1)$. Two test forms were administered to the two groups and each of the form was consisted with 9 items and 3 of the items were anchor items. They compared the Rasch model item parameter estimation results from MMLE when ignoring the background information or not. The results indicate that when background information is ignored, the difficulty parameters are

overestimated for the items that are only distributed to the less able group, while underestimated for the items that are only distributed to the more able group. However, it is free from systematic bias when the background information is considered.

DeMars (2002) conducted a simulation study to investigate the performance of MMLE under the Rasch model in a targeted test design. Two groups of 1000 students were drawn from $N(-0.5,1)$ and $N(0.5,1)$, or $N(-1,1)$ and $N(1,1)$. Two test forms were administered to the two groups according to the difficulty levels and each of the form was consisted with 60 items and 20 of the items were anchor items. He found that if group information is ignored in MMLE, the difficulty parameters are overestimated for the items on the easy form but are underestimated for the items on the difficult form. But when group information is used in MMLE, degree of bias is far less than that obtained from MMLE without the group information.

ITEM RESPONSE THOERY MODEL

IRT is a class of probabilistic model to predict the performance of an examinee on a test item when examinee's ability and item parameters are given. There are several unidimensional models can be used. Previous studies only investigated the performance of MMLE under the Rasch model (1-PL; Rasch, 1960; DeMars, 2002; Eggen & Verhelst, 2011), which is presented in Formula (1). In this study, we focused on two-parameter logistic model (2-PL; Birnbaum, 1968), which is presented in Formula (2). In the two formulas, P_j indicates the probability of item j answered correctly by an examinee at a given level of ability θ , b_j indicates the difficulty parameter of item j , and a_j indicates the discrimination parameter of item j .

$$P_j(\theta) = \frac{1}{1+\exp [-(\theta-b_j)]} \quad (1)$$

$$P_j(\theta) = \frac{1}{1+\exp [-a_j(\theta-b_j)]} \quad (2)$$

CHAPTER 3

METHODS

In this research, Monte Carlo simulation studies were conducted to investigate the performances of single group MMLE and multiple group MMLE for IRT item calibration under small-scale targeted testing scenarios. We stressed on TTOP situation and applied the 2-PL model in the simulation.

MONTE CARLO SIMULATION DESIGNS

In the simulation designs, performances of single group MMLE and multiple group MMLE were compared across different group sample sizes and different multiple test form designs. 100 simulation replications were conducted.

Multiple form designs

The five multiple form designs reviewed in Chapter 2 were used to develop the targeted testing design. That is the NEAT design, multiform design, panel missing design, panel chained design and BIB design. For multiple form designs, we considered different number of test forms. Specifically, each of these designs includes 5 test forms, that is a 5-form design, except for the NEAT design. We developed a 2-form NEAT design which includes 2 test forms as well as a 5-form NEAT design which includes 5 test forms. We also considered different number of anchor items and anchor item selection strategies when the multiple form design had a set of anchor items among all test forms. The details of these designs are presented in the following sub-sections.

NEAT design. A typical NEAT design is composed with 2 test forms. We developed a 2-form NEAT design as shown in table 3.1. The total test items were divided into three item sets: An anchor item set X, item sets A and B. Difficulties of items of anchor item set X covered all difficulty levels, and item sets A and B were consisted of easy items and hard items, respectively. Item sets X and A composed test form 1, and item sets X and B composed test form 2. Thus, test form 1 was an easy form and test form 2 was a hard form. These two test forms were assigned to 2 group of examinees with low and high abilities, respectively.

Table 3.1
2-Form NEAT Design

	X	A	B
Test form 1	√	√	
Test form 2	√		√

Note: Test forms 1-2 were assigned to 2 groups with abilities from low to high

On the basis of 2-form NEAT design, we further extended a 5-form NEAT design, which is presented in Table 3.2. The total test items were divided into six item sets: Anchor item set X, item sets A, B, C, D, E. Difficulties of items of anchor item set X covered all difficulty levels, and item sets A-E were consisted of items from easy to hard. Test forms 1-5 were composed with not only the anchor item set X, but also another item set A, B, C, D, or E. The difficulties of them were from easy to hard. They were assigned to 5 groups with low to high abilities, respectively.

Table 3.2

5-Form NEAT Design

	X	A	B	C	D	E
Test form 1	√	√				
Test form 2	√		√			
Test form 3	√			√		
Test form 4	√				√	
Test form 5	√					√

Note: Test forms 1-5 were assigned to 5 groups with abilities from low to high

Multiform design. We designed a 5-form multiform design, as Table 3.3. The total test items were divided into six item sets: Anchor item set X, item sets A, B, C, D, and E. Difficulties of items of anchor item set X covered all difficulty levels, and item sets A-E were consisted of items from easy to hard. Besides anchor item set X, each test form concluded another two item sets: test form 1 included item sets X, A, and B; test form 2 included item sets X, A, and C; test form 3 included item sets X, B, and D; test form 4 included item sets X, C, and E; test form 5 included item sets X, D, and E. The difficulties of test forms 1-5 were increased. Test forms 1-5 were administered to 5 groups with abilities from low to high, respectively.

Table 3.3

5-Form Multiform Design

	X	A	B	C	D	E
Test form 1	√	√	√			
Test form 2	√	√		√		
Test form 3	√		√		√	
Test form 4	√			√		√
Test form 5	√				√	√

Note: Test forms 1-5 were assigned to 5 groups with abilities from low to high

Panel missing design. We developed a 5-form panel missing design. As shown in Table 3.4, the total test items were divided into four item sets: A, B, C, and D. Items in these item sets were items from easy to hard. Test form 1 included all four sets while test form 2 included item sets A, B, and C; test form 3 included item sets A, B, and D; test form 4 included item sets A, C, and D; and test form 5 included item sets B, C, and D. Test form 2, test form 3, test form 1, test form 4, and test form 5 were test forms in order of difficulties from low to high. They were correspondingly assigned to 5 groups with abilities from low to high.

Table 3.4

5-Form Panel Missing Design

	A	B	C	D
Test form 1	√	√	√	√
Test form 2	√	√	√	
Test form 3	√	√		√
Test form 4	√		√	√
Test form 5		√	√	√

Note: Test form 2, 3, 1, 4, 5 were assigned to 5 groups with abilities from low to high

Panel chained design. We developed a 5-form panel chained design. As shown in table 3.5, the total test items were divided into four item sets: A, B, C, and D. Items in these item sets were items from easy to hard. Test form 1 only included one item set, item set A; test form 2 was composed with item sets A and B; test form 3 was composed with item sets B and C; test form 4 was composed with item sets C and D; and test form 5 only included item set D. The difficulties of test forms 1-5 were increased. We assigned test forms 1-5 to 5 groups with different abilities from low to high.

Table 3.5

5-Form Panel Chained Design

	A	B	C	D
Test form 1	√			
Test form 2	√	√		
Test form 3		√	√	
Test form 4			√	√
Test form 5				√

Note: Test forms 1-5 were assigned to 5 groups with abilities from low to high

BIB design. We developed a 5-form BIB design which was shown in Table 3.6. the total test items were divided into four item sets: A, B, C, D, and E. Items in these item sets were items from easy to hard. Test form 1 was composed with item sets A and C; test form 2 was composed with item sets A and D; test form 3 was composed with item sets B and D; test form 4 was composed with item sets B and E; and test form 5 was composed with item sets C and E. The difficulties of five test forms were increased. Test forms 1-5 were assigned to 5 groups with different abilities from low to high.

Table 3.6

5-Form BIB Design

	A	B	C	D	E
Test form 1	√		√		
Test form 2	√			√	
Test form 3		√		√	
Test form 4		√			√
Test form 5			√		√

Note: Test forms 1-5 were assigned to 5 groups with abilities from low to high

Anchor item. Among above multiple form designs, only the NEAT design and multiform design had an anchor item set X. When an anchor item set X existed, we considered different number of anchor items in it: 20, 30, and 50. To generate anchor items in X which could cover all levels of difficulties, we divided items into item strata based on item difficulty and discrimination levels. Then we used two selection strategies to select anchor items from these item strata. One strategy was the uniform selection strategy, that was to select the same number anchor items from each item strata; another was the non-uniform selection strategy, that was to select most anchor items, 80% for example, from the non-extreme strata.

Take selecting 20 anchor items for the 2-form NEAT design as an example (Figure 3.1). There were 8 item strata: very low b low a , very low b high a , low b low a , low b high a , high b low a , high b high a , very high b low a , and very high b high a . According to difficulty levels, the first two and the last two strata were extreme item strata. When uniform selection strategy was applied, we selected 2 items from each item strata to make up the anchor item set X. When non-uniform selection strategy was used, we selected 4 items from extreme item strata and 16 items from non-extreme strata to make up the anchor item set X. The left items composed item sets A and B correspondingly. An example for selecting 20 anchor items for the 5-form NEAT design can be found in Appendix A1.

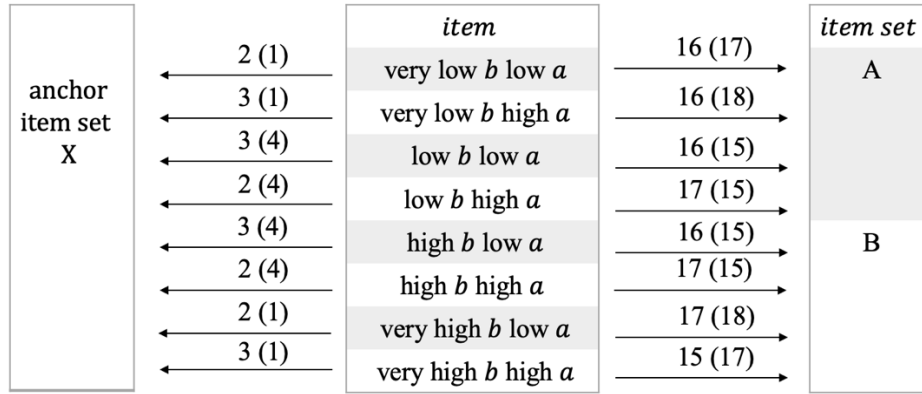


Figure 3.1. Anchor Item Selection for the 2-Form NEAT Design

Missing percentage of each test form design. Missing percentage was the percentage of missing in the data matrix of examinees' responses. For 2-form NEAT design, when there were 20, 30, 50 items in the anchor item set X, the missing percentages were 43.33%, 40% and 33.33%. For 5-form NEAT design, when there were 20, 30, 50 items in the anchor item set X, the missing percentages were 69.33%, 64%, 53.33%. For multiform design, when there were 20, 30, 50 items in the anchor item set X, the missing percentages were 52%, 48%, 40%. Panel missing design had a missing percentage of 20%. Panel chained design had a missing percentage of 60.27%. BIB design had a missing percentage of 60%.

Table 3.7

Missing Percentages for Each Multiple Form Design

Multiple form design	Number of test form	Number of anchor items	Missing percentage
NEAT design	5-form	20	69.33%
		30	64%
		50	53.33%
	2-form	20	44.33%
		30	40%
		50	33.33%
Multiform design	5-form	20	52%
		30	48%
		50	40%
Panel missing design	5-form	0	20%
Panel chained design	5-form	0	60.27%
BIB design	5-form	0	60%

Item parameter generation

A set of 150 items were generated with b parameters simulated from $N(0, 1)$ and a parameters from $\log N(0, .5)$ (Appendix A2). These item parameters were kept the same for each simulation replication. To make items with different difficulty levels had similar discrimination levels, a and b parameters were ordered from low to high. low a parameters were randomly assigned to b parameters that were located at odd positions, and high a parameters were randomly assigned to b parameters that were located at even positions, as shown in Figure 3.2. Thus, while the b parameters of item 1-150 were increased gradually, the a parameters of every two adjacent items were polarized.

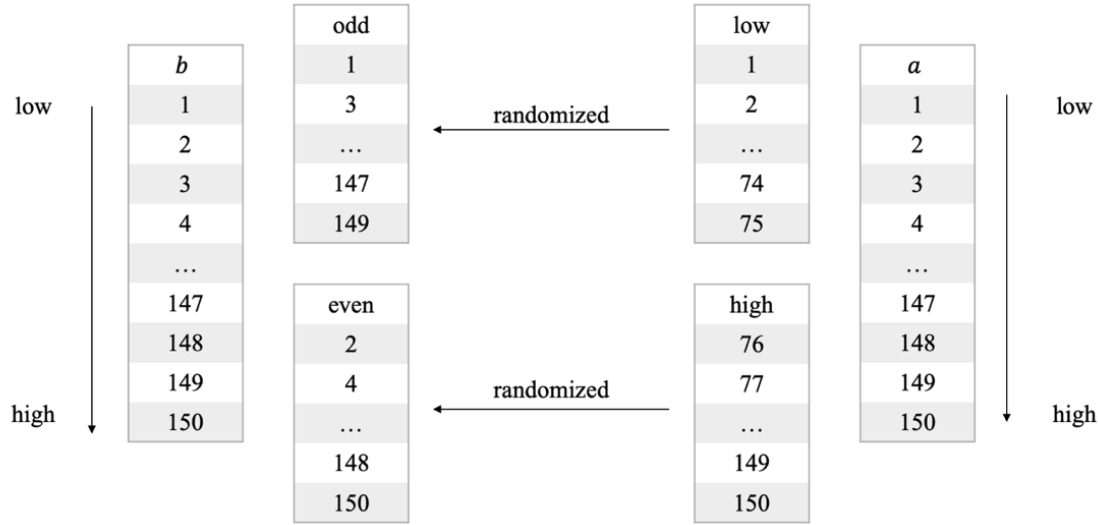


Figure 3.2. Item Formation Process

Sample size and Examinees' ability parameter generation

For each of 100 replications, a new set of examinee abilities was generated. When there were 2 test forms, 2 groups of examinees were needed and their abilities were simulated from $N(-0.5, 1)$ and $N(0.5, 1)$, respectively. The sample size for each group was 20, 50, 100, or 200. When there were 5 test forms, 5 groups of examinees were needed and their abilities were simulated from $N(-1, 1)$, $N(-0.5, 1)$, $N(0, 1)$, $N(0.5, 1)$, and $N(1, 1)$, respectively. The sample size for each group was 20, 50, or 100.

Response generation

To generate item responses, we drew random numbers from a uniform distribution between 0 and 1. The item was scored as 1 (which meant correct) if the random number was less than the calculated results by using Formula (1). Otherwise, the item was scored as 0 (which meant wrong).

Item parameter estimation

We used flexMIRT to estimate item parameters under the 2-PL model by MMLE. When ignoring the background information Y , we used single group MMLE. When considering the background information Y , we used multiple group MMLE.

Evaluation criteria

Bias and root mean squared error (RMSE) for each type of item parameters of each item were calculated by using Formula (3) and (4) below. In these two formulas, π represents the original parameter obtained by the simulation; $\hat{\pi}_r$ represents the estimated parameter by MMLE under 2-PL IRT model for the r th replication; $\hat{\pi}$ represents the mean of parameter estimates across replications; and R represents the number of the replications, which was 100 in our research.

$$bias(\hat{\pi}) = \frac{1}{R} \sum_{r=1}^R (\hat{\pi}_r - \pi) \quad (3)$$

$$RMSE(\hat{\pi}) = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\pi}_r - \pi)^2} \quad (4)$$

In addition, to calculate bias and RMSE for a and b parameters based on a more stable environment, extreme estimation values were firstly identified and excluded from calculation. The number of extreme estimation values reflects stability of the estimation: the less the number is, the more stable is the estimation. We defined extreme estimation values as estimates of parameters that were larger than the maximum original value $+ 3 \times$ standard deviation of the distribution of true values or smaller than the minimum original value $- 3 \times$ standard deviation of the distribution of true values.

CHAPTER 4

RESULTS

NEAT DESIGN

As a reminder, four factors were considered with the NEAT design. That is, the number of test forms, sample size, number of anchor items in anchor item set X and anchor item selection strategy for items in anchor item set X . To compare performances of single group MMLE and multiple group MMLE, the number of extreme estimation values, bias, and RMSE were compared across different sample sizes, different number of anchor items and different anchor item selection strategies. The results were presented in following two parts: 5-form NEAT design and 2-form NEAT design.

5-form NEAT design

Estimation of α parameters. For α parameter estimation, number of extreme estimation values, bias, and RMSE were presented.

Number of extreme estimation values. Figure 4.1 and Appendix B1 document the number of extreme estimations values when estimating α parameters by single group MMLE and multiple group MMLE.

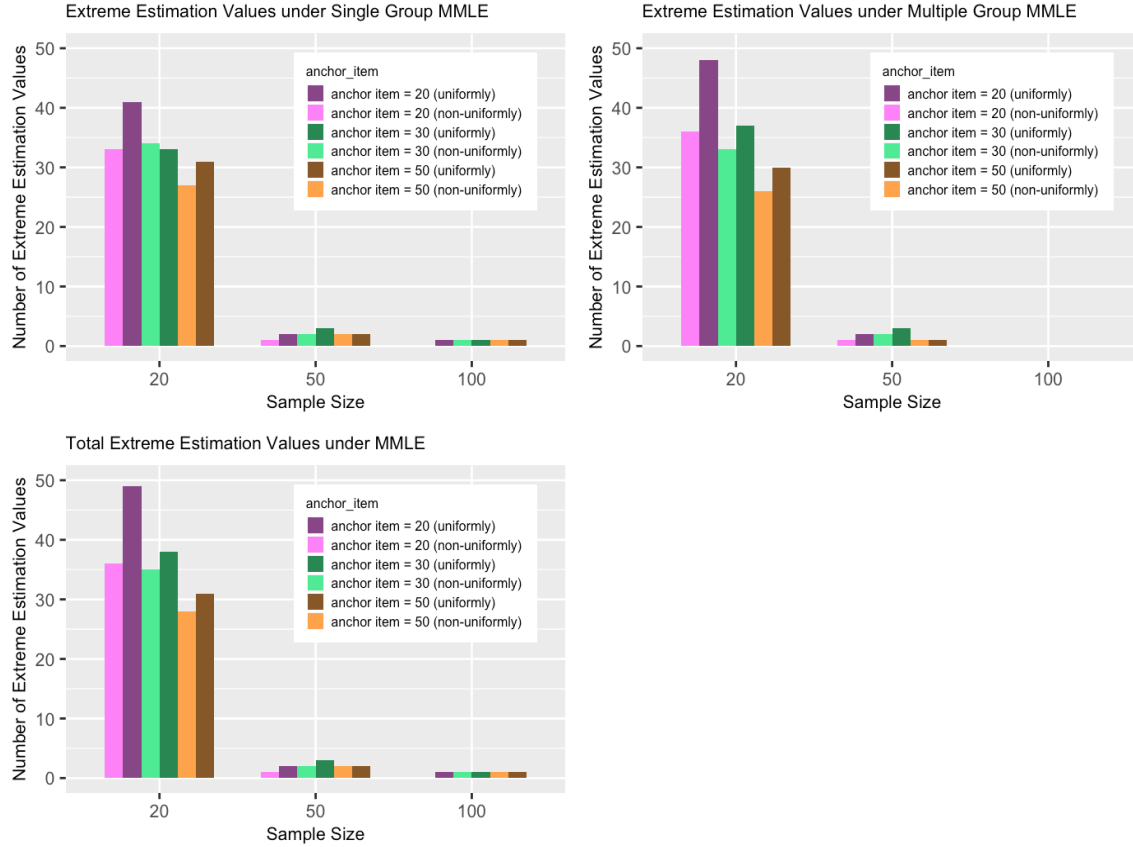


Figure 4.1. Number of Extreme Estimation Values for a Parameters under 5-Form NEAT Design

Comparison across different sample sizes. When sample size was 100, single group MMLE identified one extreme estimation value (except when using 20 non-uniformly selected anchor items, none was identified) while multiple group MMLE identified none. When sample size was 50, single group MMLE identified the same number of extreme estimation values as multiple group MMLE when there were 20 or 30 anchor items; but it identified more than multiple group MMLE when there were 50 anchor items. When sample size was 20, single group MMLE identified less extreme estimation values than multiple group MMLE when using 20 anchor items and 30 uniformly selected anchor items; but it identified more than multiple group MMLE when using 50 anchor items and 30 non-uniformly selected anchor items. Overall, for

both MMLE approaches, as sample size increased, the number of extreme estimation values decreased.

Comparison across different number of anchor items. When there were 50 anchor items, single group MMLE identified more extreme estimation values than multiple group MMLE. When there were 30 anchor items, single group MMLE identified more extreme estimation values as multiple group when sample size was 100; it identified the same number of extreme estimation values as multiple group when sample size was 50; it identified less than multiple group MMLE when sample size was 30 and anchor items were uniformly selected and identified more when sample size was 30 and anchor items were non-uniformly selected. When there were 20 anchor items, single group MMLE identified less extreme estimation values than multiple group MMLE when sample size was 20; it identified the same number of extreme estimation values as multiple group MMLE when sample size was 50; and it identified more extreme estimation values when sample size was 100 and using uniformly selected anchor items, and identified the same number as multiple group MMLE when sample size was 100 and using non-uniformly selected anchor items. Specifically, for both MMLE approaches, as the number of anchor items increased, the number of extreme estimation values decreased.

Comparison across different anchor item selection strategies. When anchor items were uniformly selected, single group MMLE identified less extreme estimation values than multiple group MMLE when sample size was 20 (except when there were 50 anchor items, it identified more); it identified the same number of extreme estimation values as multiple group MMLE when sample size was 50 (except when there were 50 anchor items, it identified more); and it identified more extreme estimation values than multiple group MMLE when sample size was 100. When anchor items were non-uniformly selected, single group MMLE identified more

extreme estimation values than multiple group MMLE when sample size was 20 (except when there were 20 anchor items, it identified less); it identified the same number of extreme estimation values as multiple group MMLE when sample size was 50 (except when there were 50 anchor items, it identified more); and it identified more extreme estimation values than multiple group MMLE when sample size was 100 (except when there were 50 anchor items, it identified the same number with multiple group MMLE).

In general, multiple group MMLE was more stable than single group MMLE when 50 anchor items were used, or when sample size was 50 or 100.

Bias. Figure 4.2 presents boxplots of bias of estimated α parameters from single group MMLE and multiple group MMLE. The five number summaries can be found in Appendix B2. According to Figure 4.2, it can be seen that α parameters were overestimated by both MMLE approaches.

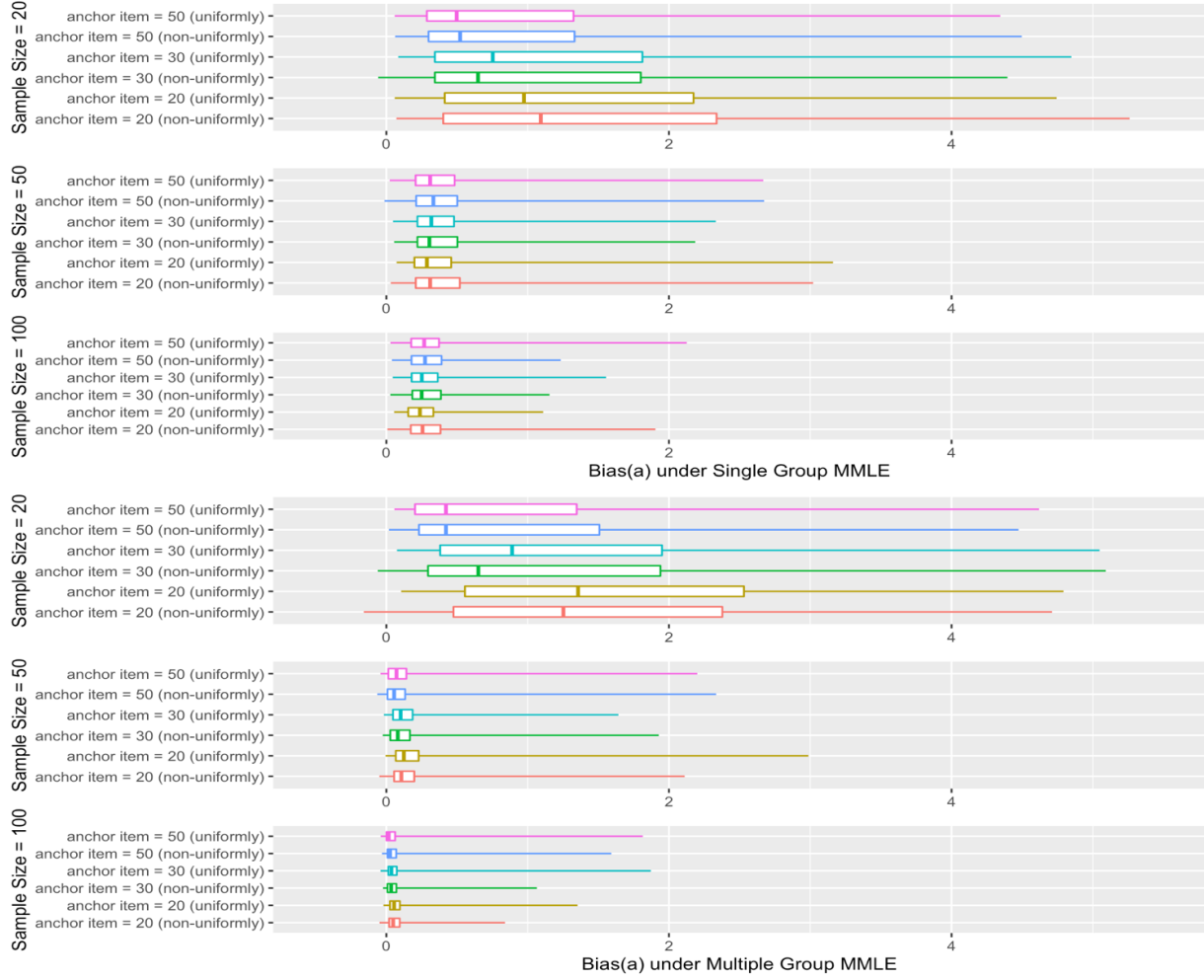


Figure 4.2. Bias for a Parameters under 5-Form NEAT Design

Comparison across different sample sizes. Figure 4.2 shows that bias observed from multiple group MMLE was closer to zero than bias observed from single group MMLE when sample size was 50 or 100. But when sample size was 20, bias obtained from these two MMLE approaches were close. In addition, for both MMLE approaches, bias was closer to zero as sample size increased.

Figure 4.3 presents bias of estimated a parameters for each item when there were 20 uniformly selected anchor items. When sample size was 100 or 50, the absolute value of the bias from single group MMLE was larger than that from multiple group MMLE, and bias dots from

these two MMLE approaches lied apart. But when sample size was 20, the absolute value of the bias from single group MMLE was smaller than that from multiple group MMLE, and bias dots from these two MMLE approaches were mixed together. In addition, as sample size increased, the range of bias from both MMLE approaches decreased.

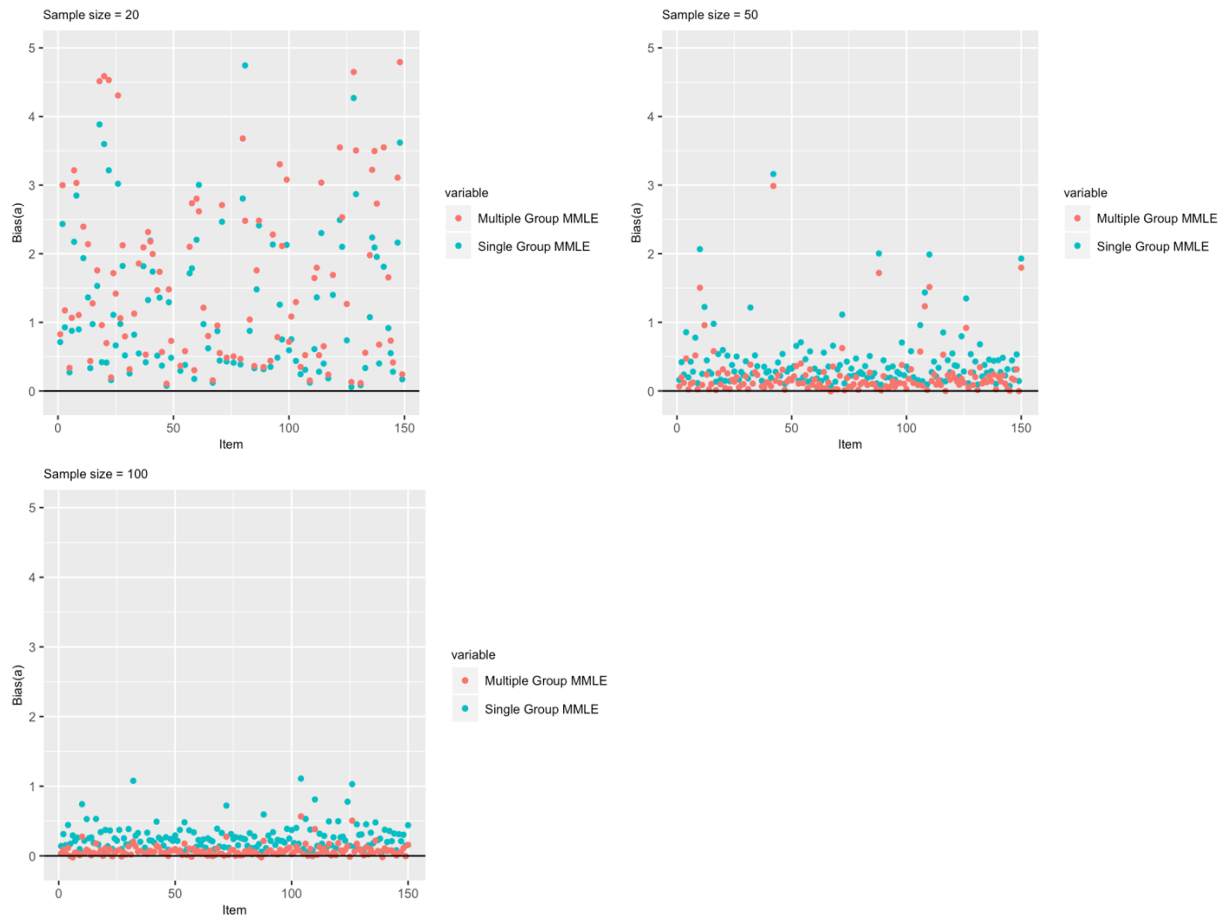


Figure 4.3. Bias for a Parameters under 5-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes

Comparison across different number of anchor items. Figure 4.2 shows that for all number of anchor items conditions, bias obtained from multiple group MMLE was closer to zero than bias obtained from single group MMLE except when sample size was 20. Specifically, when sample size was 20, as number of anchor items increased, bias was closer to zero for these

two MMLE approaches. When sample size was 50 or 100, bias didn't change as sample size changed.

Figure 4.4 presents bias of estimated a parameters for each item when using uniformly selected anchor items and a sample size of 100. Bias observed for multiple group MMLE were closer to zero than bias for single group MMLE regardless of the number of anchor items. Difference of bias was tiny when using different number of anchor items conditions for both MMLE approaches.

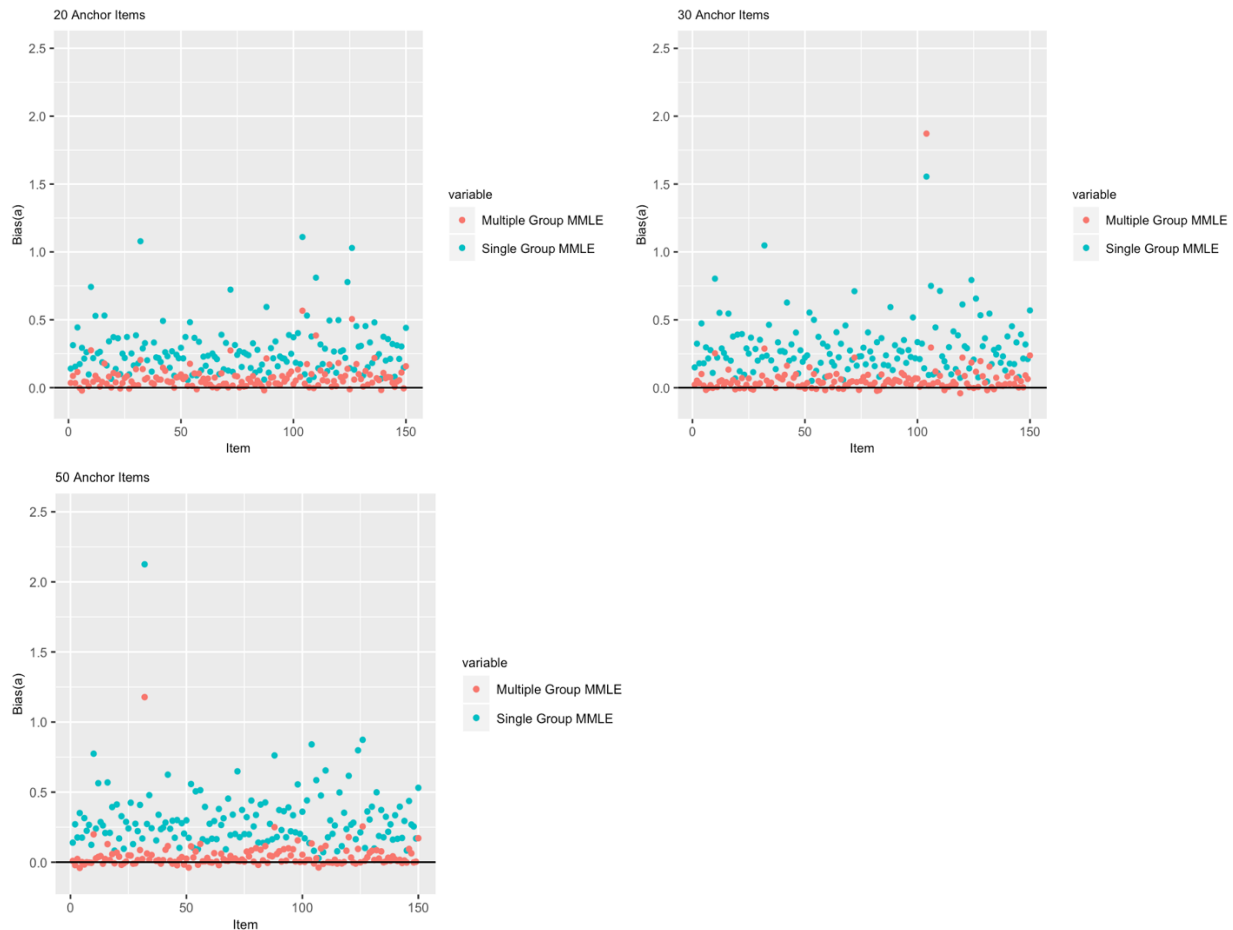


Figure 4.4. Bias for a Parameters under 5-Form NEAT Design with a Sample Size of 100 across Different Number of Uniformly Selected Anchor Items

Comparison across different anchor item selection strategies. Figure 4.2 shows that for both anchor item selection strategies, bias obtained from multiple group MMLE was closer to zero than bias obtained from single group MMLE except when sample size was 20. When sample size was 20, the absolute value of bias from these two MMLE approaches were close.

Figure 4.5 presents estimated bias of estimated a parameters for each item when using 20 anchor items and a sample size of 100. The bias of multiple group MMLE was closer to zero than bias of single group MMLE regardless of the anchor item selection strategies. Difference of bias was tiny when using different anchor item selection strategies for both MMLE approaches.

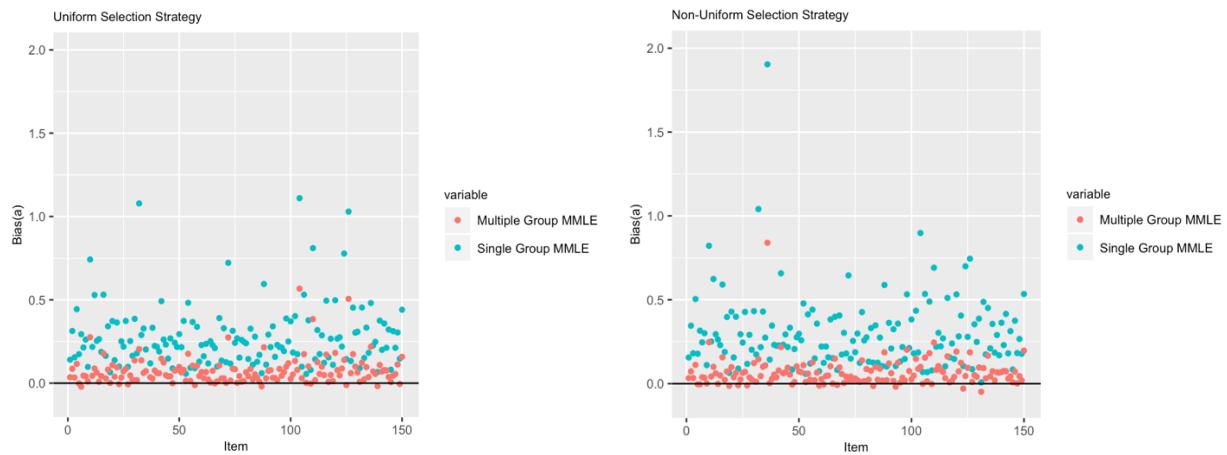


Figure 4.5. Bias for a Parameters under 5-Form NEAT Design with 20 Anchor Items and a Sample Size of 100 across Different Anchor Item Selection Strategies

RMSE. Figure 4.6 presents boxplots of RMSE of estimated a parameters from single group MMLE and multiple group MMLE. The five number summaries can be found in Appendix B3.

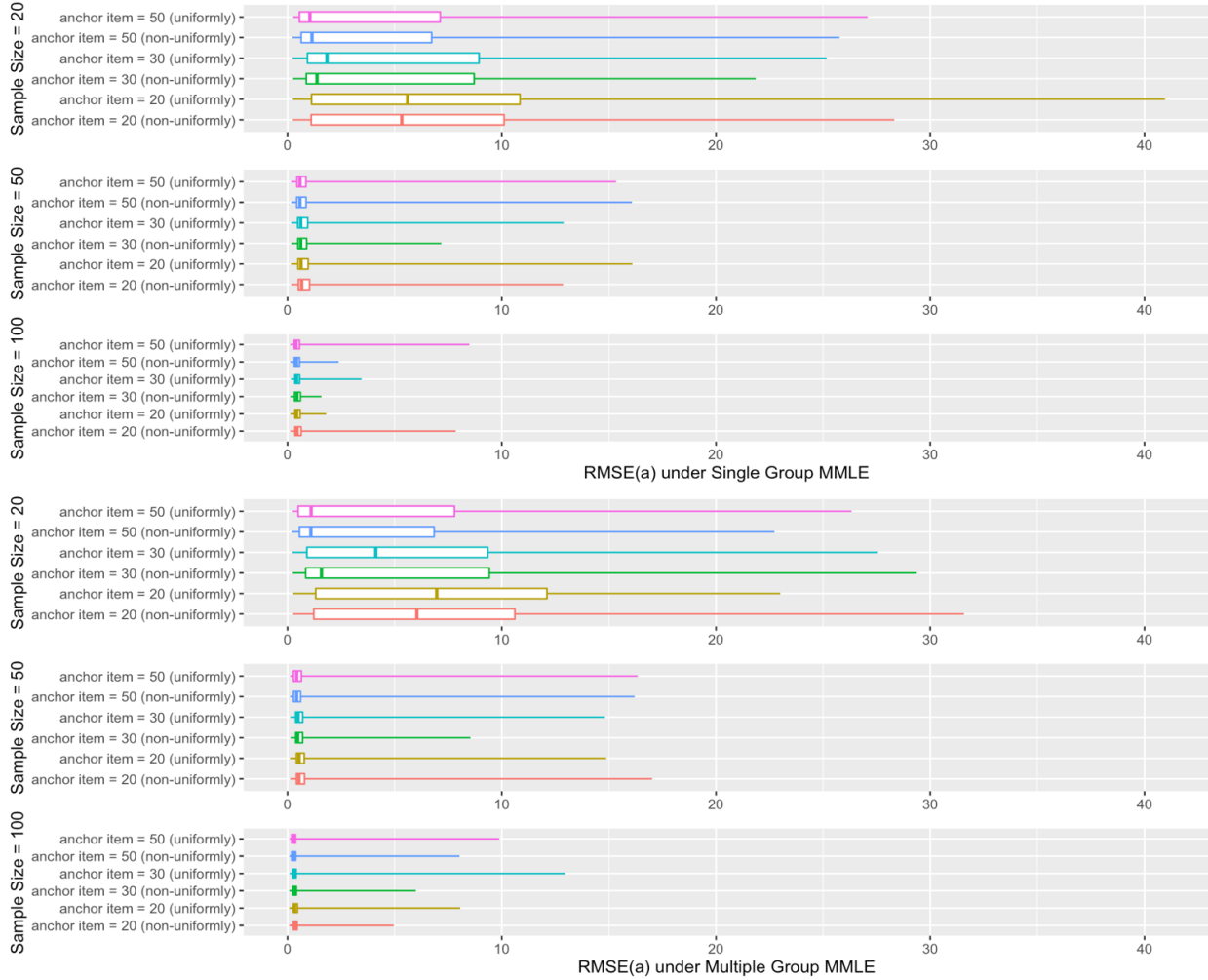


Figure 4.6. RMSE for a Parameters under 5-Form NEAT Design

Comparison across different sample sizes. From Figure 4.6, it can be seen that larger RMSE were observed for single group MMLE when sample size was 50 or 100. But RMSE observed for two MMLE approaches were close when sample size was 20. In addition, for both MMLE approaches, RMSE decreased with the sample size increased.

Figure 4.7 shows RMSE of estimated a parameters for each item when there were 20 uniformly selected anchor items. Figure 4.8 shows the RMSE that was smaller than 2 when sample size was 50 or 100 to provide a close look at the bottom of the plots in Figure 4.8. It can be seen that when sample size was 100 or 50, RMSE obtained from single group MMLE was

larger than RMSE obtained from multiple group MMLE. When sample size was 20, RMSE obtained from single group MMLE was close to RMSE obtained from multiple group MMLE. In addition, as sample size increased, RMSE dots for two MMLE approaches lied apart further and range of RMSE decreased for both MMLE approaches.

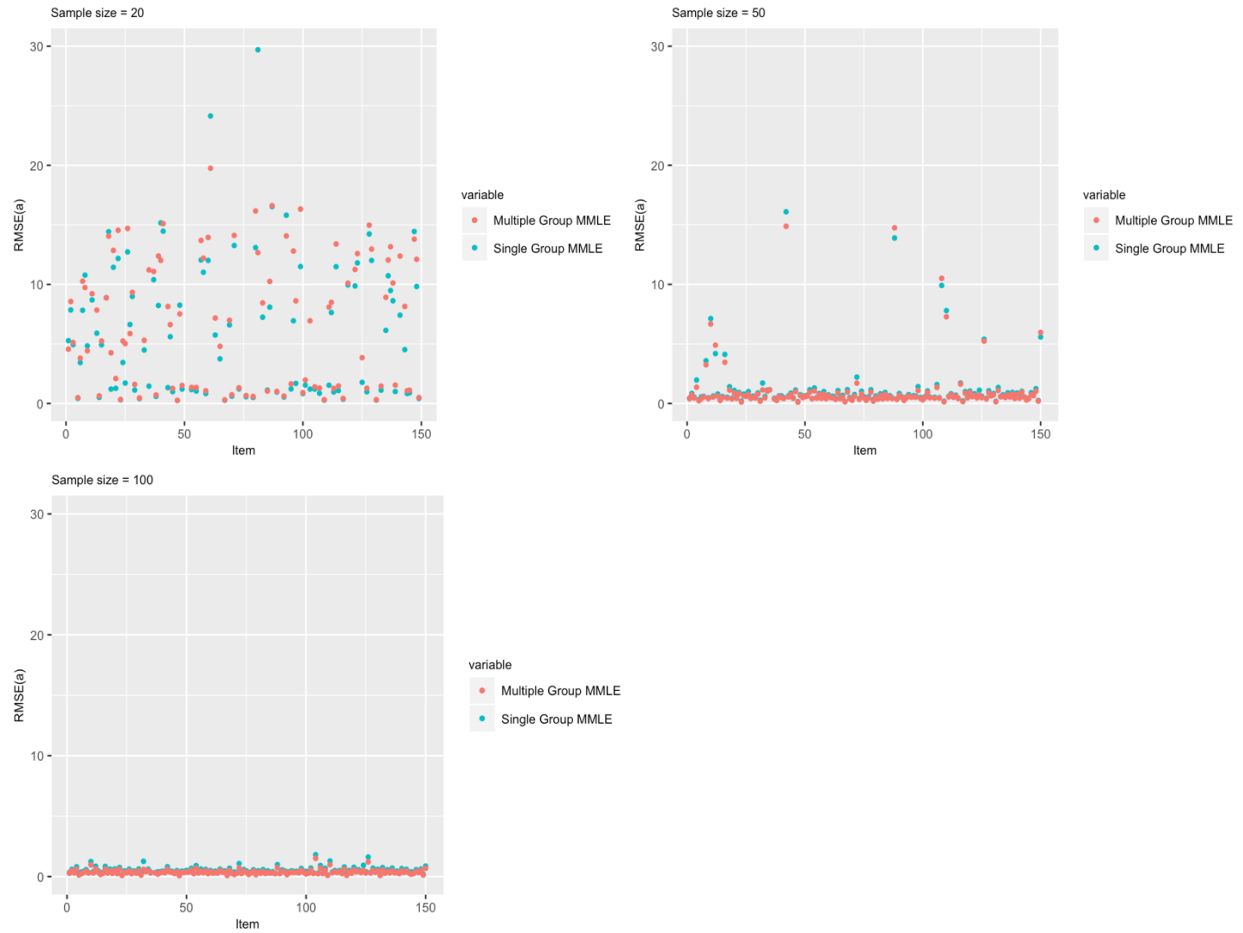


Figure 4.7. RMSE for a Parameters under 5-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes

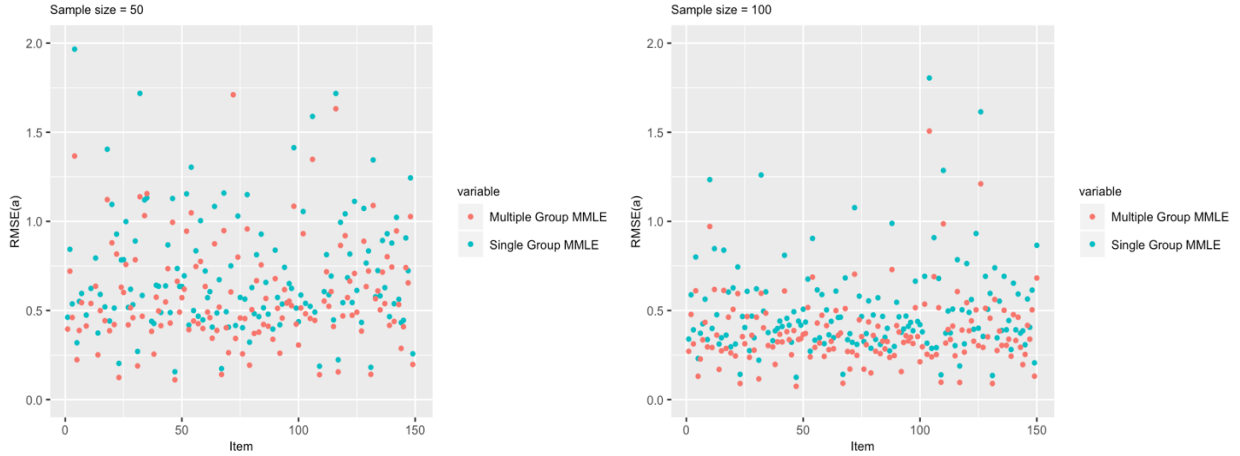


Figure 4.8. RMSE (from 0 to 2) for α Parameters under 5-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes

Comparison across different number of anchor items. Figure 4.6 shows that for all number of anchor items conditions, RMSE obtained from multiple group MMLE was smaller than RMSE obtained from single group MMLE except when sample size was 20. Specifically, when sample size was 20, RMSE was smaller as number of anchor items increased.

Figure 4.9 presents RMSE of estimated α parameters for each item when using uniformly selected anchor items and a sample size of 100. RMSE observed for multiple group MMLE was smaller than RMSE for single group MMLE regardless of number of anchor items. Difference of RMSE was tiny when using different number of anchor items conditions for both MMLE approaches.

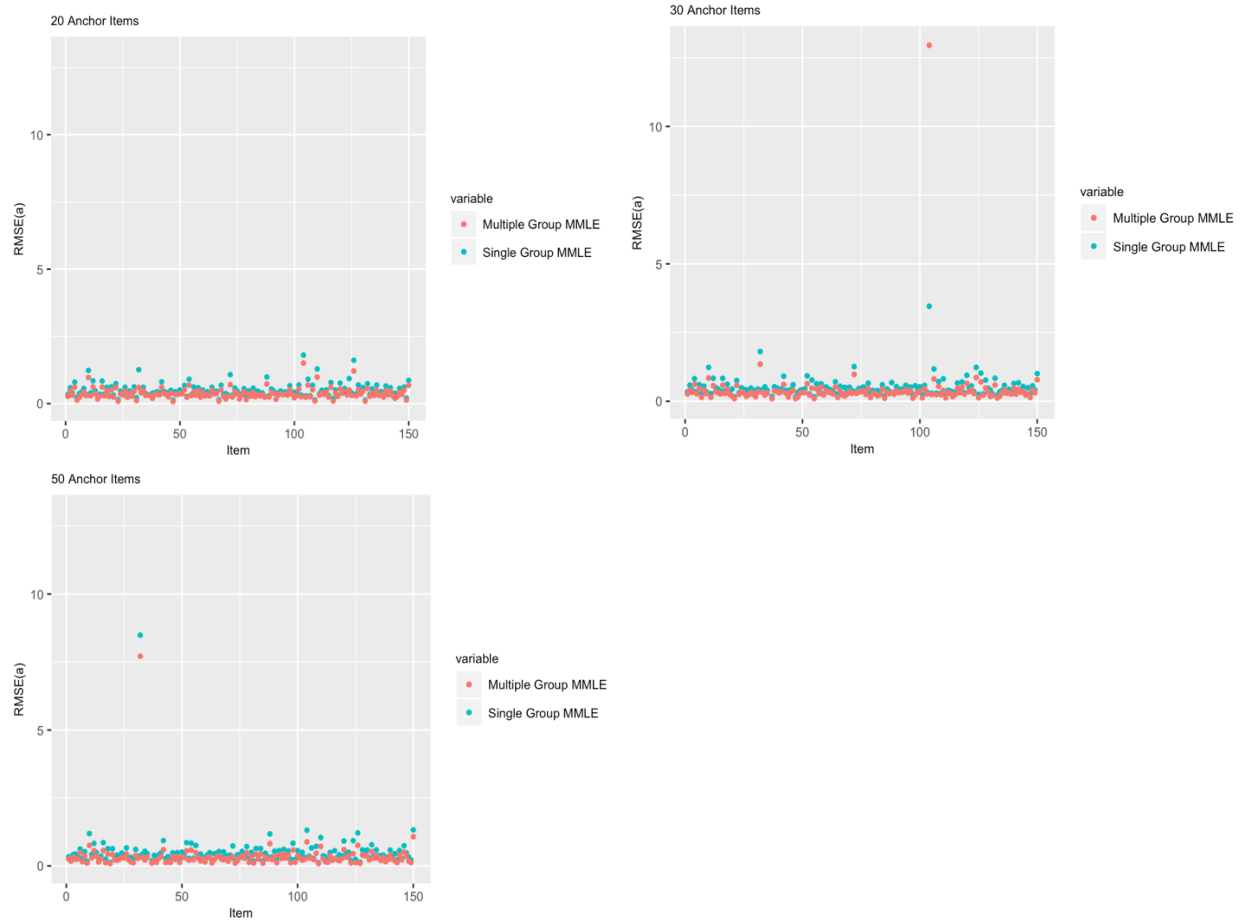


Figure 4.9. RMSE for a Parameters under 5-Form NEAT Design with a Sample Size of 100 across Different Number of Uniformly Selected Anchor Items

Comparison across different anchor item selection strategies. Figure 4.6 shows that for both anchor item selection strategies, RMSE obtained from multiple group MMLE was smaller than RMSE obtained from single group MMLE except when sample size was 20. When sample size was 20, RMSE obtained from these two MMLE approaches were similar.

Figure 4.10 presents estimated RMSE of estimated a parameters for each item when 20 anchor items and a sample size of 100. RMSE for multiple group MMLE was smaller than RMSE dots for single group MMLE regardless of anchor item selection strategies. Difference of RMSE was tiny when using different anchor item selection strategies for both MMLE approaches.



Figure 4.10. RMSE for a Parameters under 5-Form NEAT Design with 20 Anchor Items and a Sample Size of 100 across Different Anchor Item Selection Strategies

Estimation of b parameters. For b parameter estimation, number of extreme estimation values, bias, and RMSE were presented.

Number of extreme estimation values. Figure 4.11 and Appendix B4 show the number of extreme estimations values when estimating b parameters by single group MMLE and multiple group MMLE.

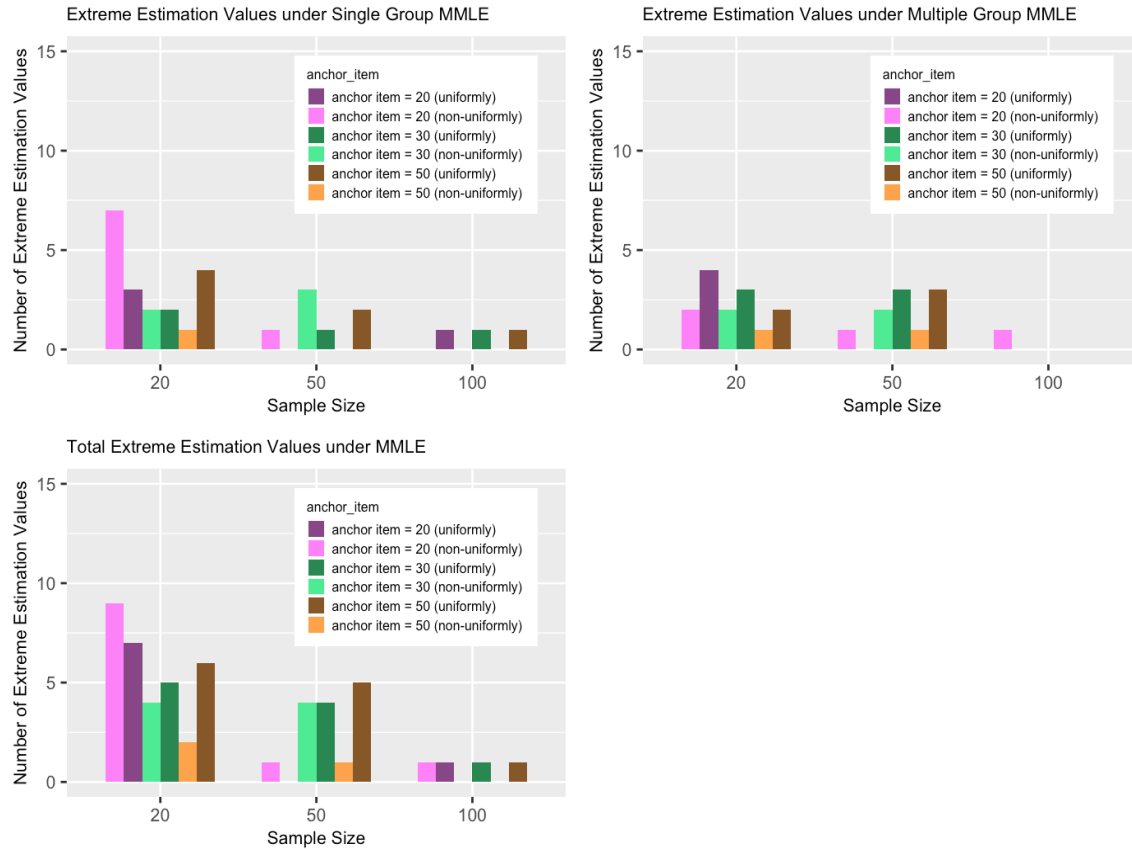


Figure 4.11. Number of Extreme Estimation Values for b Parameters under 5-Form NEAT Design

Comparison across different sample sizes. When sample size was 20, there was no consistent pattern between numbers of extreme estimation values identified by single group MMLE and multiple group MMLE. When sample size was 50, single group MMLE identified the same number of extreme estimation values as multiple group MMLE when there were 20 anchor items. When there were 30 and 50 anchor items, single group MMLE identified less extreme estimation values than multiple group MMLE (except when 30 non-uniformly selected anchor items were used, single group MMLE identified more). When sample size was 100, if anchor items were selected uniformly, single group MMLE identified more extreme estimation values than multiple group MMLE; but if anchor items were selected non-uniformly, both MMLE approaches identified none (except when there were 20 anchor items, multiple group

MMLE identified more). In addition, for both MMLE approaches, as sample size increased, the number of extreme estimation values decreased.

Comparison across different number of anchor items. When there were 50 anchor items and a sample size of 50 or 20, single group MMLE identified more extreme estimation values than multiple group MMLE when anchor items were uniformly selected; but two MMLE approaches identified the same number of extreme estimation values when anchor items were non-uniformly selected. When using a sample size of 30, single group MMLE identified less extreme estimation values than multiple group MMLE. When there were 30 anchor items and using a sample size of 50 or 20, single group MMLE identified more extreme estimation values than multiple group MMLE when anchor items were uniformly selected; but two MMLE approaches identified the same number of extreme estimation values when anchor items were non-uniformly selected. When using a sample size of 30, single group MMLE identified less extreme estimation values than multiple group MMLE when anchor items were selected uniformly; while it identified more than multiple group MMLE when anchor items were selected non-uniformly. When there were 20 anchor items and using a sample size of 20, single group MMLE identified less extreme estimation values than multiple group MMLE when anchor items were uniformly selected; while it identified more than multiple group MMLE when anchor items were non-uniformly selected. When sample size was 50, two MMLE approaches identified the same number of extreme estimation values. When sample size was 100, single group MMLE identified more extreme estimation values than multiple group MMLE when anchor items were uniformly selected; while it identified less than multiple group MMLE when anchor items were non-uniformly selected.

Comparison across different anchor item selection strategies. When anchor items were uniformly selected, single group MMLE identified less extreme estimation values than multiple group MMLE when sample size was 20 (except when 50 anchor items were used, single group MMLE identified more); it identified less extreme estimation values than multiple group MMLE when sample size was 50 (except when 20 anchor items were used, both MMLE approaches identified none); and it identified more extreme estimation values than multiple group MMLE when sample size was 100. When anchor items were non-uniformly selected, two MMLE approaches identified the same number of extreme estimation values when sample size was 20 (except when 20 anchor items were used, single group MMLE identified more). When sample size was 50, two MMLE approaches identified the same number of extreme estimation values when using 20 anchor items; single group MMLE identified more than multiple group MMLE when using 30 anchor items; and single group MMLE identified less than multiple group MMLE when using 50 anchor items. When sample size was 100, single group MMLE identified no extreme estimation values while multiple group MMLE identified one when using 20 anchor items; and both MMLE approaches identified none when using 30 or 50 anchor items.

In general, when sample size was 20, single group MMLE was more stable than multiple group MMLE when 20 and 30 anchor items were uniformly selected; and multiple group MMLE was more stable when 50 anchor items were uniformly selected. When sample size was 50, single group was more stable. When sample size was 100, multiple group was more stable.

Bias. Figure 4.12 presents boxplots of bias of estimated b parameters from single group MMLE and multiple group MMLE. The five number summaries can be found in Appendix B5.

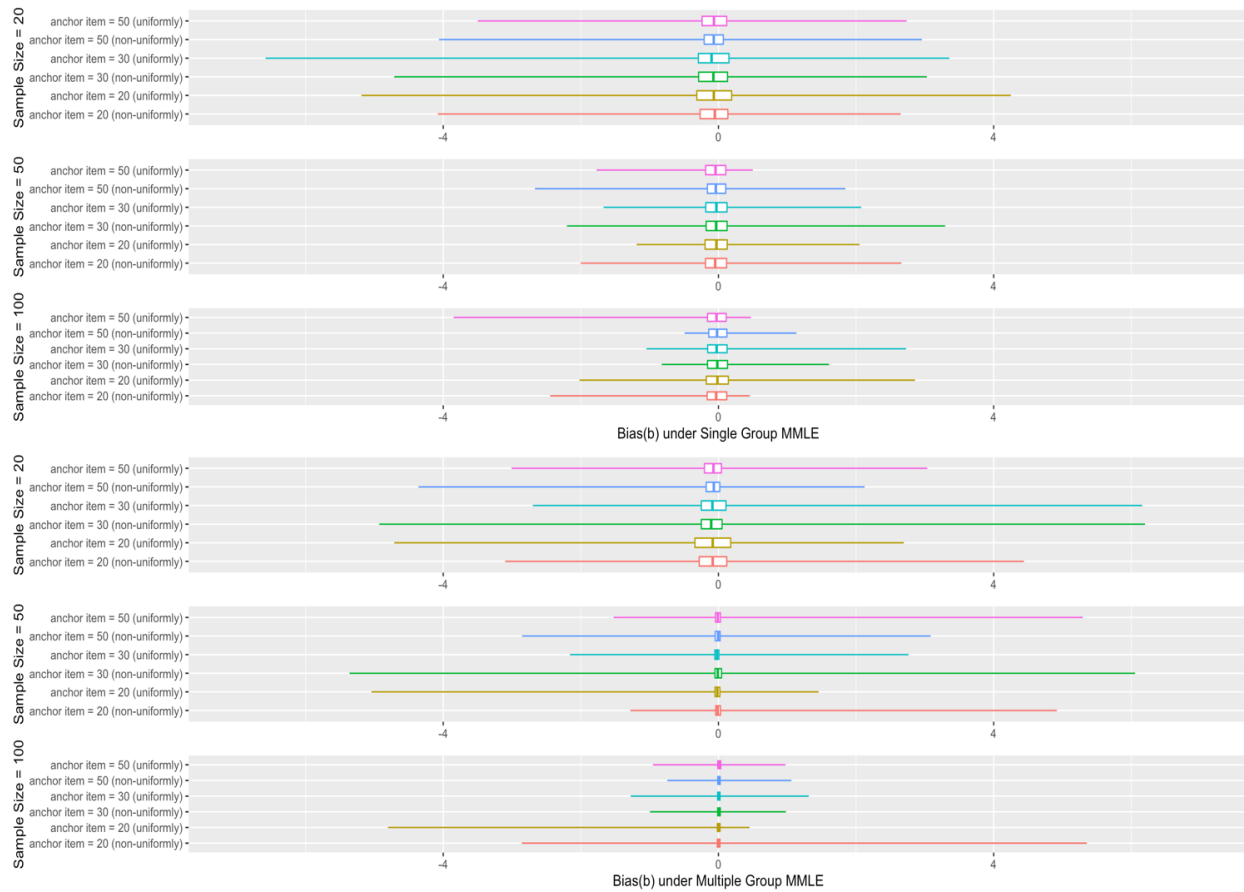


Figure 4.12. Bias for b Parameters under 5-Form NEAT Design

Comparison across different sample sizes. From Figure 4.12, it can be seen that when sample size was 50 or 100, bias obtained from multiple group MMLE was closer to zero than bias obtained from single group MMLE. When sample size was 20, difference between bias obtained from two MMLE approaches was not that significant. With sample size increased, bias was closer to zero for multiple group MMLE, but bias didn't change that much for single group MMLE.

Figure 4.13 presents bias of estimated b parameters for each item when there were 20 uniformly selected anchor items. Items 1-150 were items from easy to hard. When sample size was 20, b parameters were overestimated for easy items and were underestimated for hard items by these two MMLE approaches. When sample size was 50 or 100, while b parameters were

overestimated for easy items and were underestimated for hard items by single group MMLE, it was almost unbiased by multiple group MMLE. Overall, for both MMLE approaches, as sample size increased, bias dots were less scattered, and the range of bias decreased.

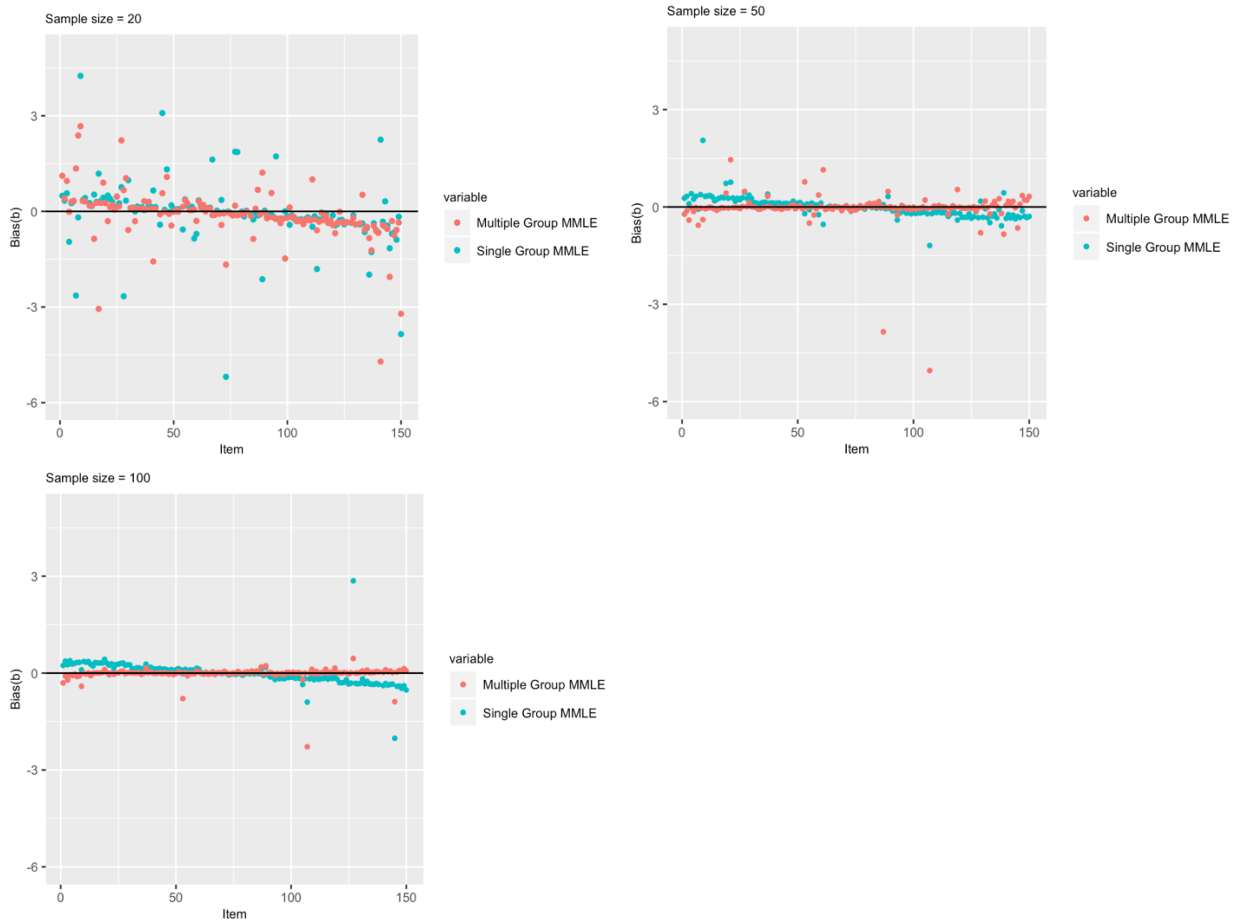


Figure 4.13. Bias for b Parameters under 5-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes

Comparison across different number of anchor items. Figure 4.12 shows that for all number of anchor items conditions, bias obtained from multiple group MMLE was closer to zero than bias obtained from single group MMLE except when sample size was 20. When sample size was 20, bias obtained from these two MMLE approaches were close.

Figure 4.14 presents bias of estimated b parameters for each item when using uniformly selected anchor items and a sample size of 100. Single group MMLE overestimated b parameters

for easy items and underestimated b parameters for hard items while multiple group MMLE was almost unbiased. Difference of bias was tiny when using different number of anchor items conditions for both MMLE approaches.

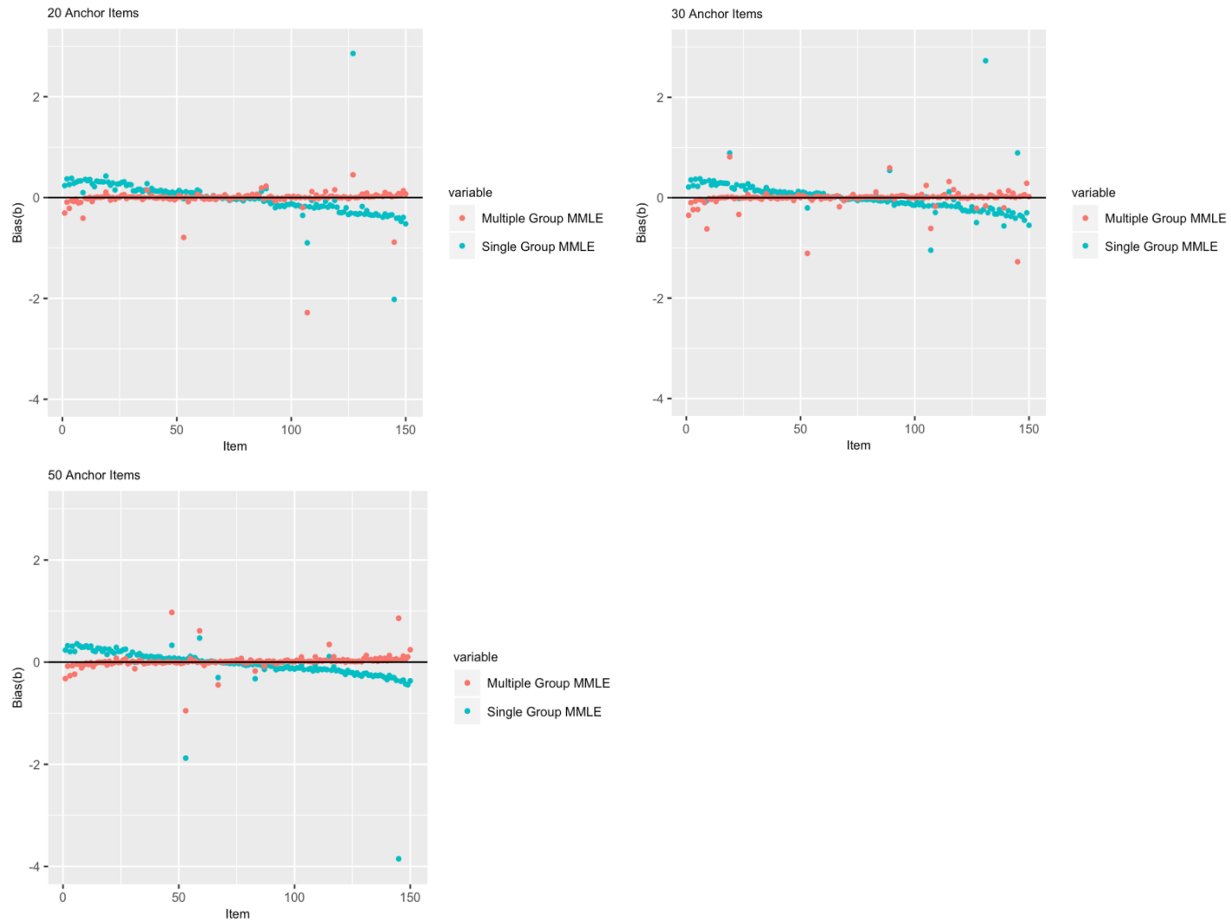


Figure 4.14. Bias for b Parameters under 5-Form NEAT Design with a Sample Size of 100 across Different Number of Uniformly Selected Anchor Items

Comparison across different anchor item selection strategies. Figure 4.12 shows that for both anchor item selection strategies, bias obtained from multiple group MMLE was closer to zero than bias obtained from single group MMLE except when sample size was 20. When sample size was 20, bias obtained from these two MMLE approaches were close.

Figure 4.15 presents estimated RMSE of estimated b parameters for each item when 20 anchor items and a sample size of 100 were used. Single group MMLE overestimated b

parameters for easy items and underestimated b parameters for hard items while multiple group MMLE was almost unbiased. Difference of bias was tiny when using different anchor item selection strategies for both MMLE approaches.

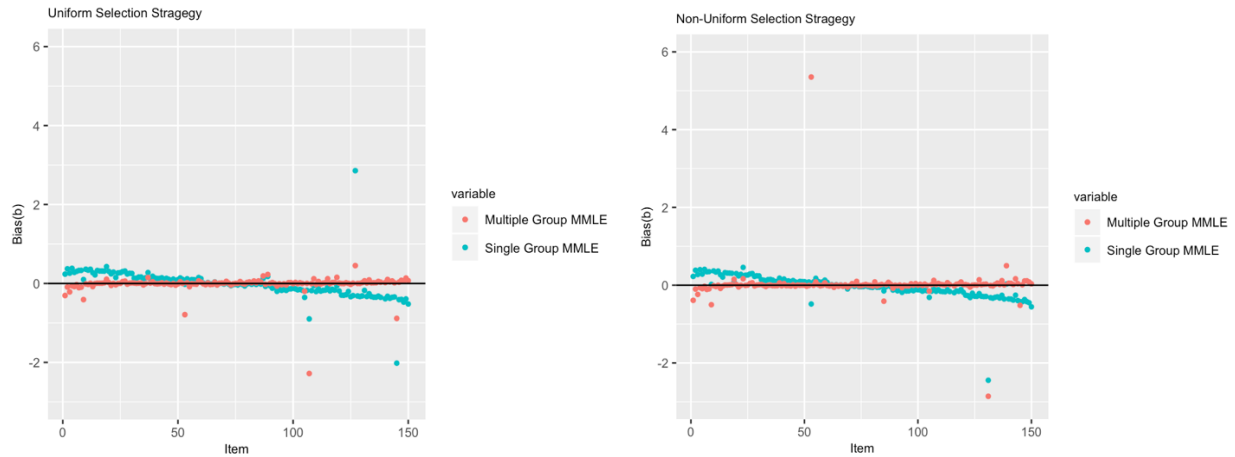


Figure 4.15. Bias for b Parameters under 5-Form NEAT Design with 20 Anchor Items and a Sample Size of 100 across Different Anchor Item Selection Strategies

RMSE. Figure 4.16 presents boxplots of RMSE of estimated b parameters from single group MMLE and multiple group MMLE. The five number summaries can be found in Appendix B6.

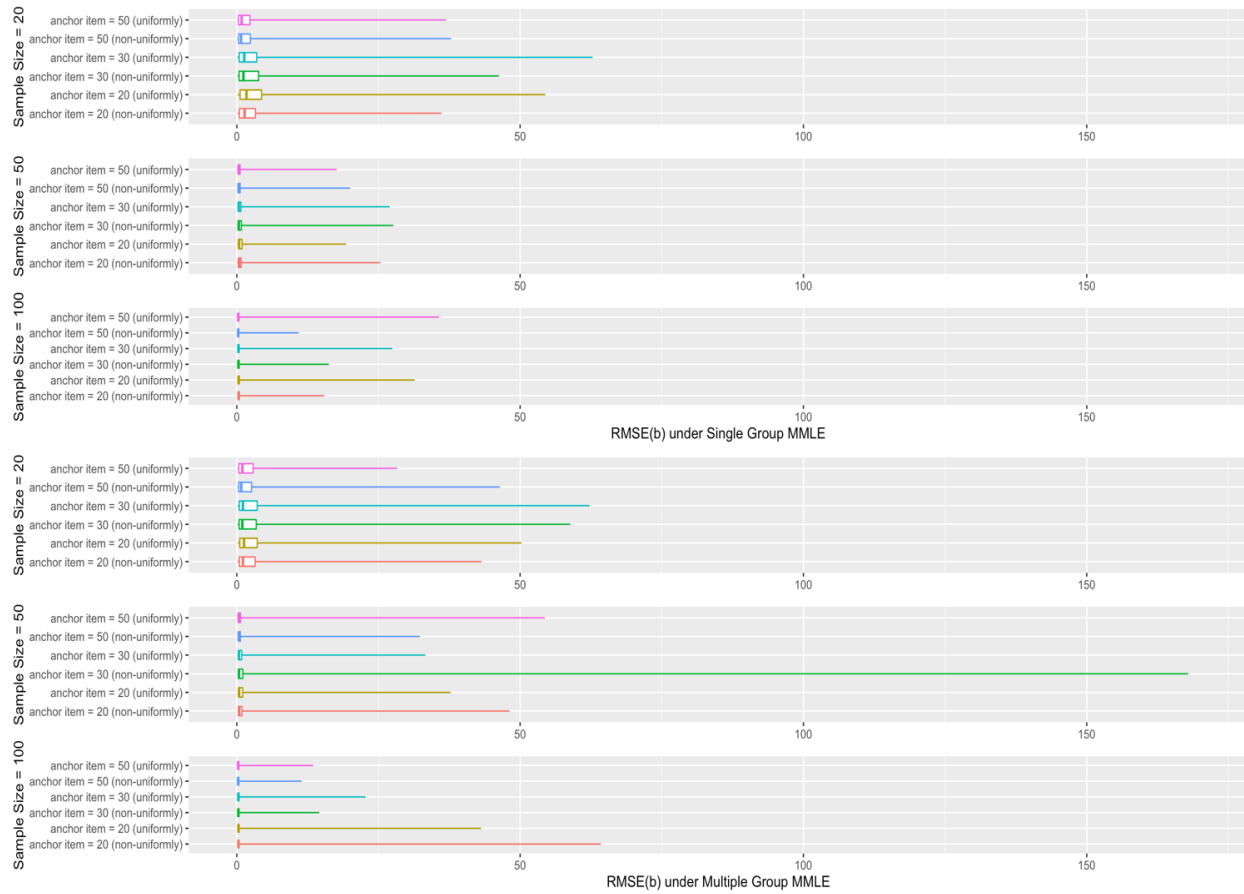


Figure 4.16. RMSE for b Parameters under 5-Form NEAT Design

Comparison across different sample sizes. From Figure 4.16, it can be seen that when sample size was 50 or 100, RMSE obtained from multiple group MMLE was smaller than bias obtained from single group MMLE. When sample size was 20, difference between RMSE obtained from two MMLE approaches was not that significant. In addition, for both MMLE approaches, RMSE was smaller when sample size increased.

Figure 4.17 presents RMSE of estimated b parameters for each item when there were 20 uniformly selected anchor items. Figure 4.18 shows RMSE that smaller than 2 to take a close look at the bottom of plots in Figure 4.17. Items 1-150 were items from easy to hard. When sample size was 50 or 100, larger RMSE was observed from single group MMLE than from multiple group MMLE for easy and hard items; but for middle difficult items, smaller RMSE

was observed from single group MMLE than that from multiple group MMLE. When sample size was 20, RMSE dots obtained from two MMLE approaches were mixed together and there was no clear magnitude pattern existed. As sample size increased to 50 or 100, RMSE dots obtained from two MMLE approaches lied apart and the dots were less scattered.

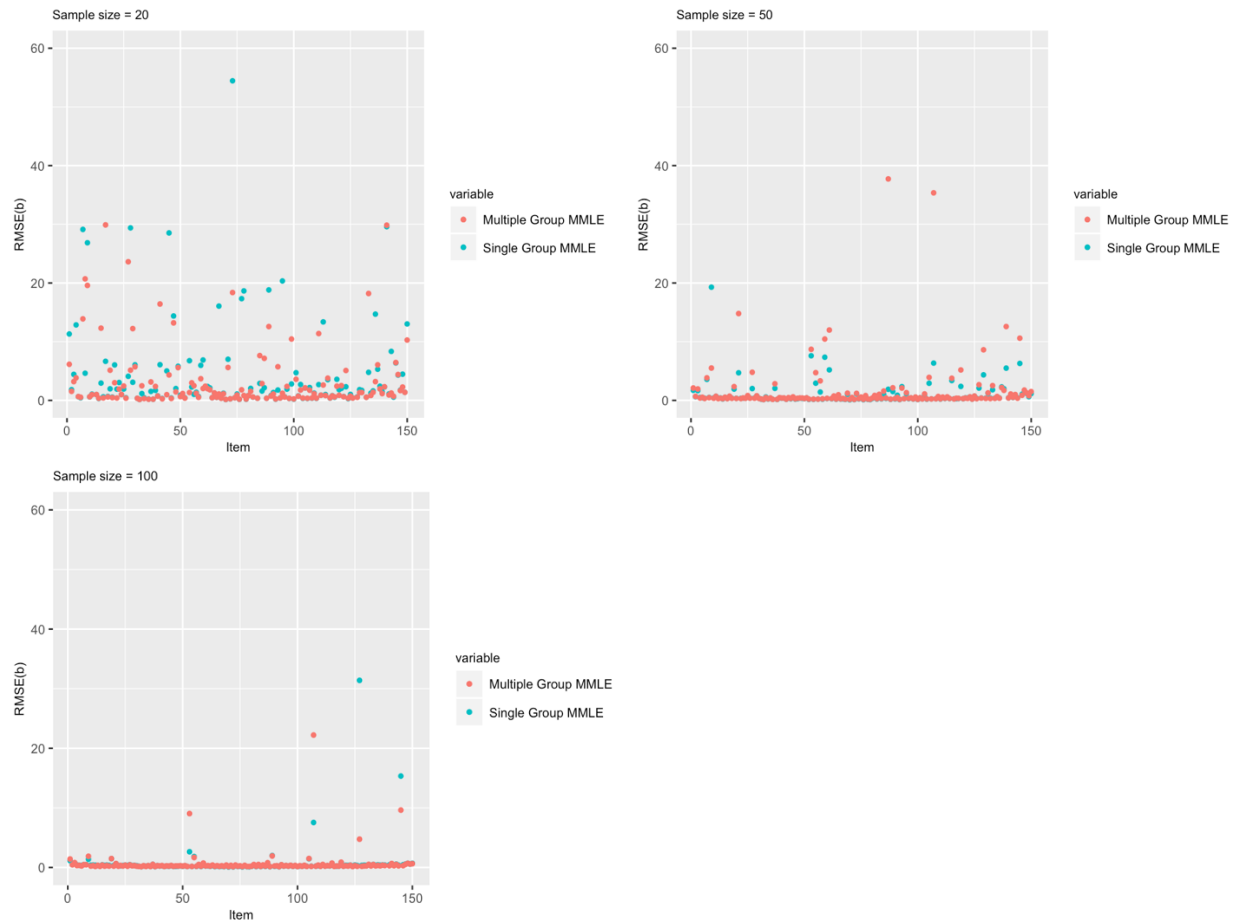


Figure 4.17. RMSE for b Parameters under 5-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes

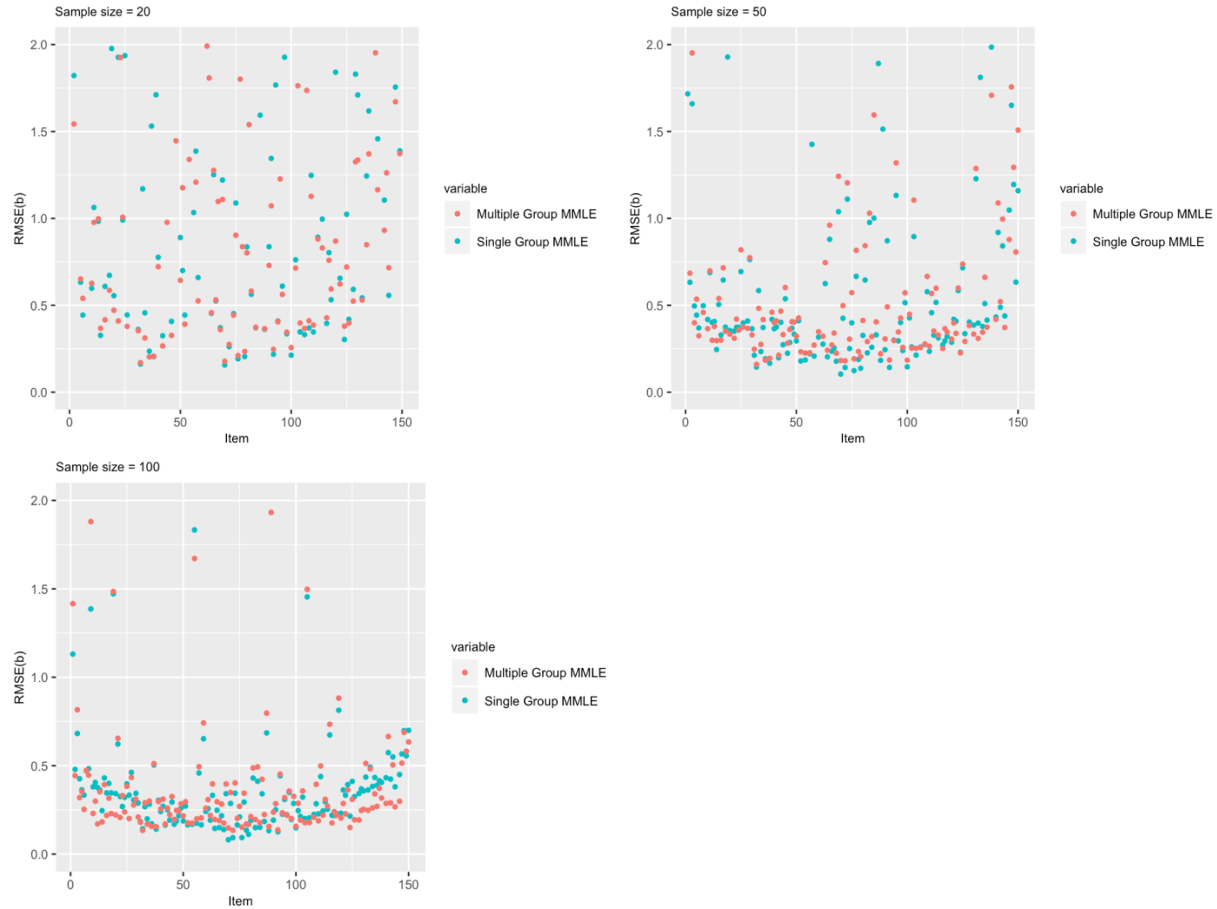


Figure 4.18. RMSE (from 0 to 2) for b Parameters under 5-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes

Comparison across different number of anchor items. Figure 4.16 shows that for all number of anchor items conditions, RMSE obtained from multiple group MMLE was smaller than RMSE obtained from single group MMLE except when sample size was 20. When sample size was 20, difference between RMSE obtained from two MMLE approaches was not that significant.

Figure 4.19 presents RMSE of estimated b parameters for each item when using uniformly selected anchor items and a sample size of 100. Figure 4.20 presents RMSE smaller than 2 to take a close look at the bottom of plots in Figure 4.19. For easy and hard items, larger RMSE was observed for single group MMLE than multiple group MMLE; but for middle

difficult items, smaller RMSE was observed for single group MMLE than multiple group MMLE. No significant difference of RMSE existed when using different number of anchor items for both two MMLE approaches.

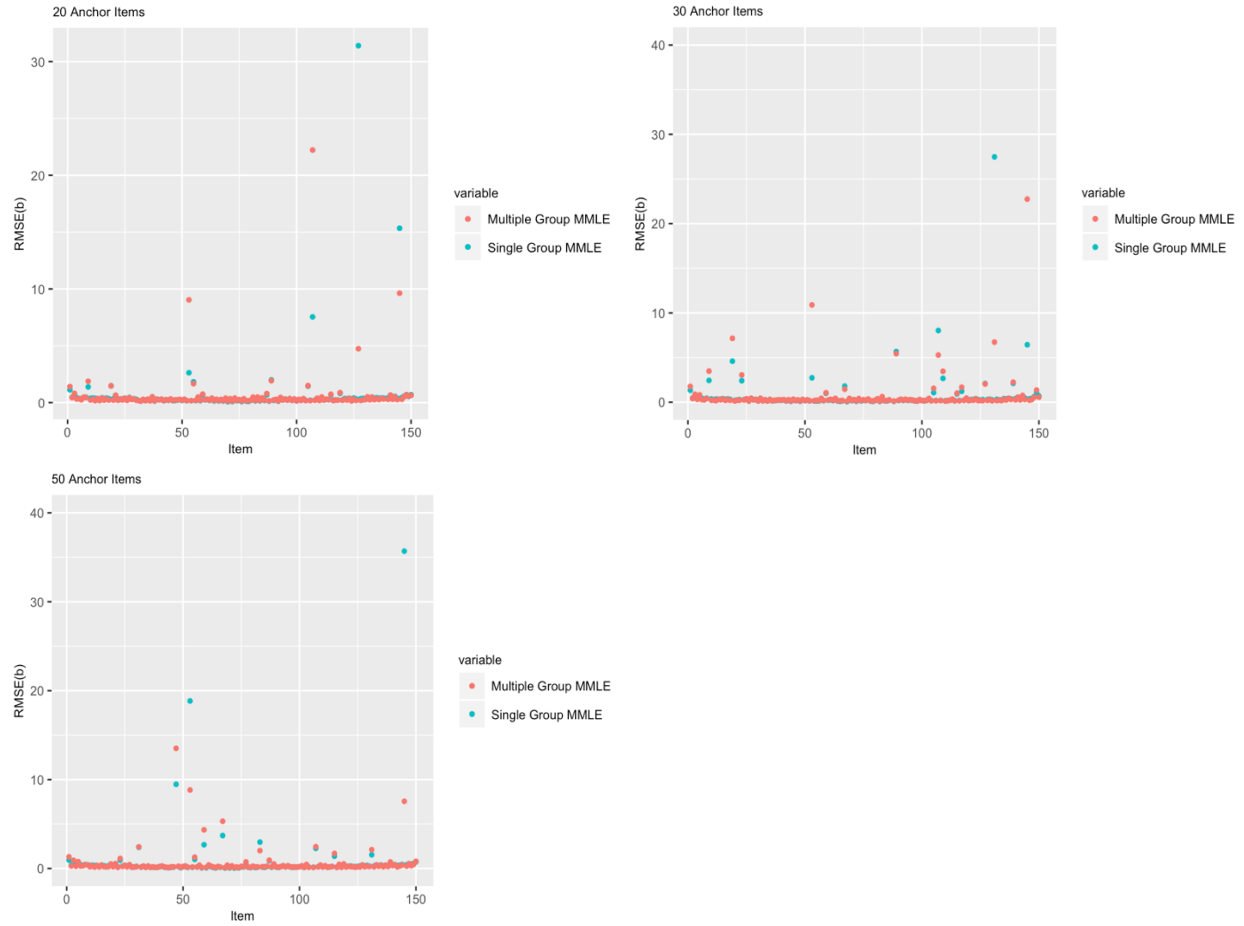


Figure 4.19. RMSE for b Parameters under 5-Form NEAT Design with a Sample Size of 100 across Different Number of Uniformly Selected Anchor Items

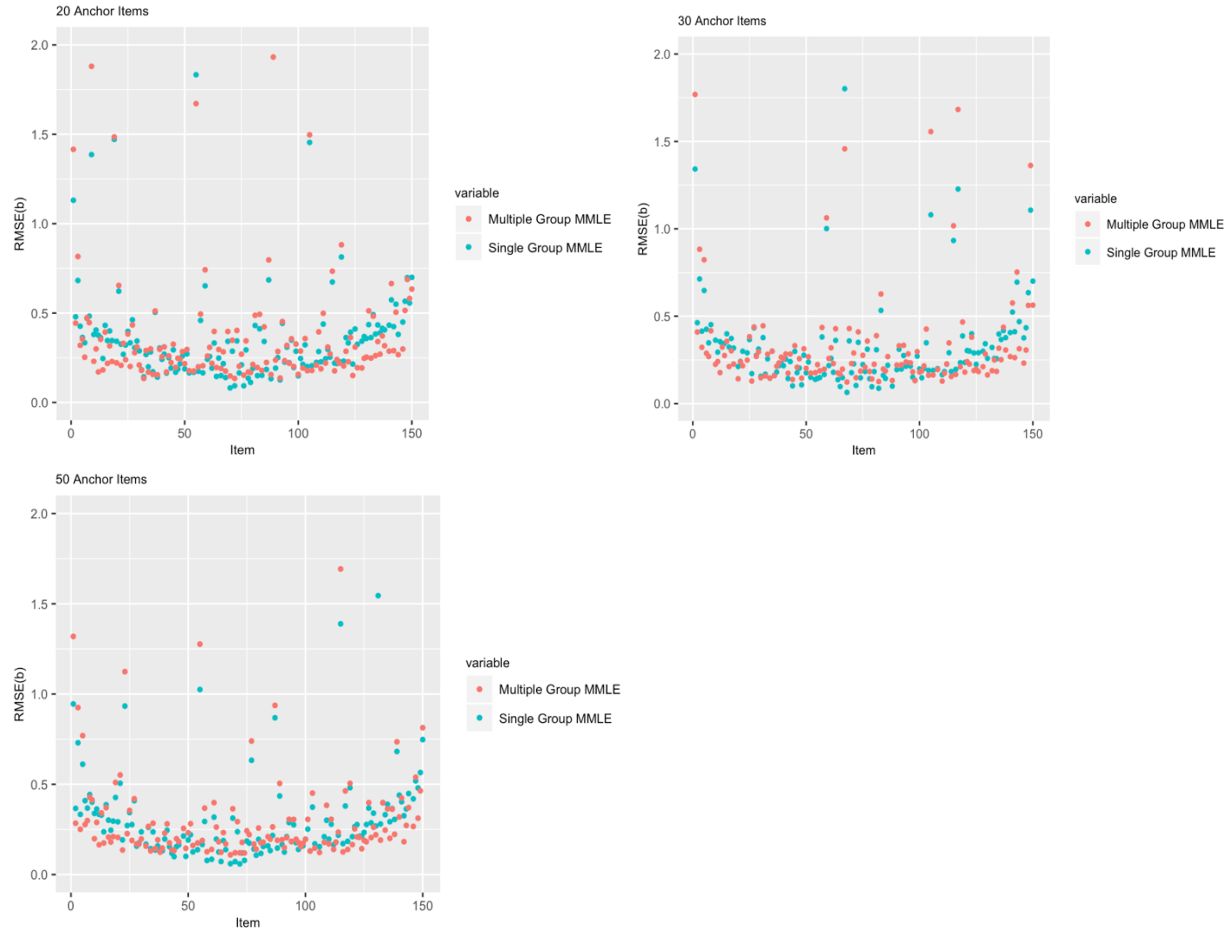


Figure 4.20. RMSE (from 0 to 2) for b Parameters under 5-Form NEAT Design with a Sample Size of 100 across Different Number of Uniformly Selected Anchor Items

Comparison across different anchor item selection strategies. Figure 4.12 shows that for two anchor item selection strategies, RMSE obtained from multiple group MMLE was smaller than RMSE obtained from single group MMLE except when sample size was 20. When sample size was 20, difference between RMSE obtained from two MMLE approaches was not that significant.

Figure 4.21 presents RMSE of estimated b parameters for each item when 20 anchor items and a sample size of 100. Figure 4.22 presents RMSE that was smaller than 2 to take a close look at the bottom of the plots in Figure 4.21. For easy and hard items, larger RMSE was observed from single group MMLE than from multiple group MMLE; but for middle difficult

items, smaller RMSE was observed from single group MMLE than from multiple group MMLE. No significant difference of RMSE existed when using different anchor item selection strategies for both MMLE approaches.

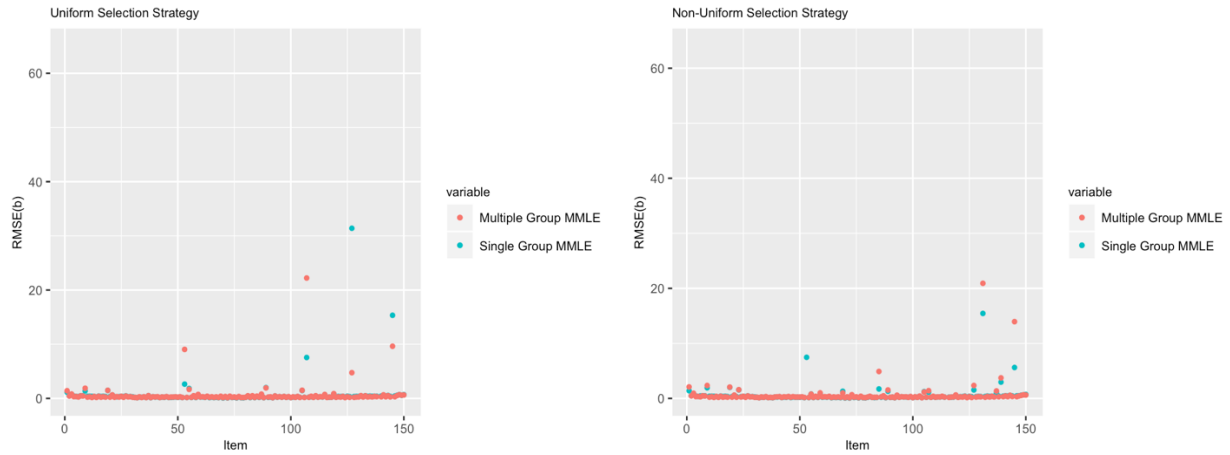


Figure 4.21. RMSE for b Parameters under 5-Form NEAT with 20 Anchor Items and a Sample Size of 100 across Different Anchor Item Selection Strategies



Figure 4.22. RMSE (from 0 to 2) for b Parameters under 5-Form NEAT Design with 20 Anchor Items and a Sample Size of 100 across Different Anchor Item Selection Strategies

2-form NEAT design

Estimation of a parameters. For a parameter estimation, number of extreme estimation values, bias, and RMSE were presented.

Number of extreme estimation values. Figure 5.1 and Appendix C1 document the number of extreme estimations values when estimating a parameters by single group MMLE and multiple group MMLE.

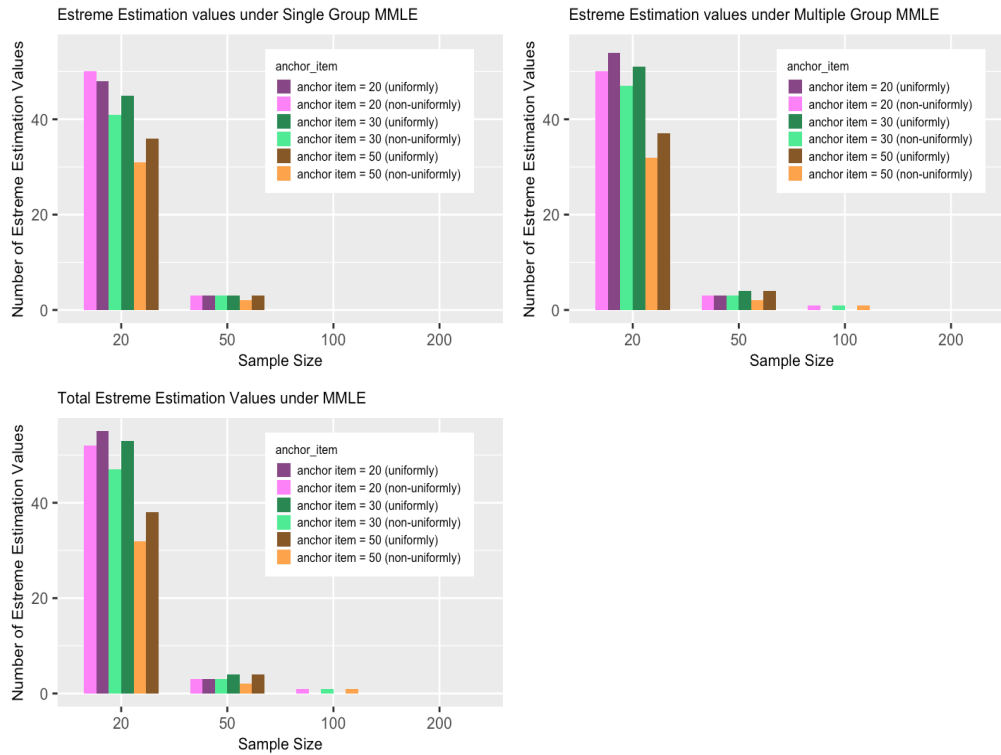


Figure 5.1. Number of Extreme Estimation Values for a Parameters under 2-Form NEAT Design

Comparison across different sample sizes. When sample size was 20, single group MMLE identified less extreme estimation values than multiple group MMLE (except when using 20 non-uniformly anchor items, two MMLE approaches identified the same number). When sample size was 50, single group MMLE identified the same number of extreme estimation values as multiple group MMLE (except when using 30 uniformly selected anchor items and 50

uniformly selected anchor items, single group MMLE identified less). When sample size was 100, two MMLE approaches identified no extreme estimation value when anchor items were uniformly selected. When anchor items were non-uniformly selected, multiple group MMLE identified one extreme estimation value while single group MMLE identified none. When sample size was 200, both MMLE approaches identified no extreme estimation value. In addition, as sample size increased, the number of extreme estimation values decreased for both MMLE approaches.

Comparison across different number of anchor items. When there were 20 anchor items, single group MMLE identified less extreme estimation values than multiple group MMLE when sample size was 20 and anchor items were uniformly selected, or when sample size was 100 and anchor items were uniformly selected. In the rest cases, two MMLE approaches identified the same number of extreme estimation values. When there were 30 anchor items, single group MMLE identified less extreme estimation values than multiple group MMLE when sample size was 20, or when sample size was 50 and anchor items were uniformly selected, or when sample size was 100 and anchor items were non-uniformly selected. In the rest cases, two MMLE approaches identified the same number of extreme estimation values. When there were 50 anchor items, single group MMLE identified less extreme estimation values than multiple group MMLE when sample size was 20, or when sample size was 50 and anchor items were uniformly selected, or when sample size was 100 and anchor items were non-uniformly selected. In the rest cases, two MMLE approaches identified the same number of extreme estimation values.

Comparison across different anchor item selection strategies. When anchor items were uniformly selected, single group MMLE identified less extreme estimation values than multiple group MMLE when sample size was 20, or when sample size was 50 and 30 anchor items were

used, or when sample size was 50 and 50 anchor items were used. In the rest cases, two MMLE approaches identified the same number of extreme estimation values. When anchor items were non-uniformly selected, single group MMLE identified less extreme estimation values than multiple group MMLE when sample size was 20 and 30 anchor items were used, or when sample size was 20 and 50 anchor items were used, or when sample size was 100. In the rest cases, two MMLE approaches identified the same number of extreme estimation values.

In general, single group MMLE was more stable than multiple group MMLE when sample size was 20. Two MMLE approaches were the same stable when sample size was 30 (except when there were 30 or 50 uniformly selected anchor items, single group MMLE was more stable). When sample size was 100, two MMLE approaches were the same stable when there were uniformly selected anchor items; single group MMLE was more stable when there were non-uniformly selected anchor items. When sample size was 200, two MMLE approaches were the same stable.

Bias. Figure 5.2 presents boxplots of bias of estimated α parameters from single group MMLE and multiple group MMLE. The five number summaries can be found in Appendix C2. According to Figure 5.2, it can be seen that α parameters were overestimated by both MMLE approaches.



Figure 5.2. Bias for a Parameters under 2-Form NEAT Design

In short, the magnitude relationship between bias for estimated a parameters obtained from two MMLE approaches of 2-form NEAT design was similar to that of 5-form NEAT design. That is, when sample size was 20 or 50, bias observed for two MMLE approaches were close. When sample size was 100 or 200, bias observed for multiple group MMLE were closer to zero than bias observed for single group MMLE. Besides, same as 5-form NEAT design, there

was no significant difference of bias when using different number of anchor items or different anchor item selection strategies for both MMLE approaches. Detailed comparisons to compare bias across different sample sizes, different number of anchor items, and different anchor item selection strategies were presented in Appendix C3-C5. They respectively showed results under these three situations: when 20 uniformly selected anchor items were used, when a sample size of 200 and uniformly selected anchor items were used, and when a sample size of 200 and 20 uniformly selected anchor items were used.

RMSE. Figure 5.3 presents boxplots of RMSE of estimated a parameters from single group MMLE and multiple group MMLE. The five number summaries were documented in Appendix C6.

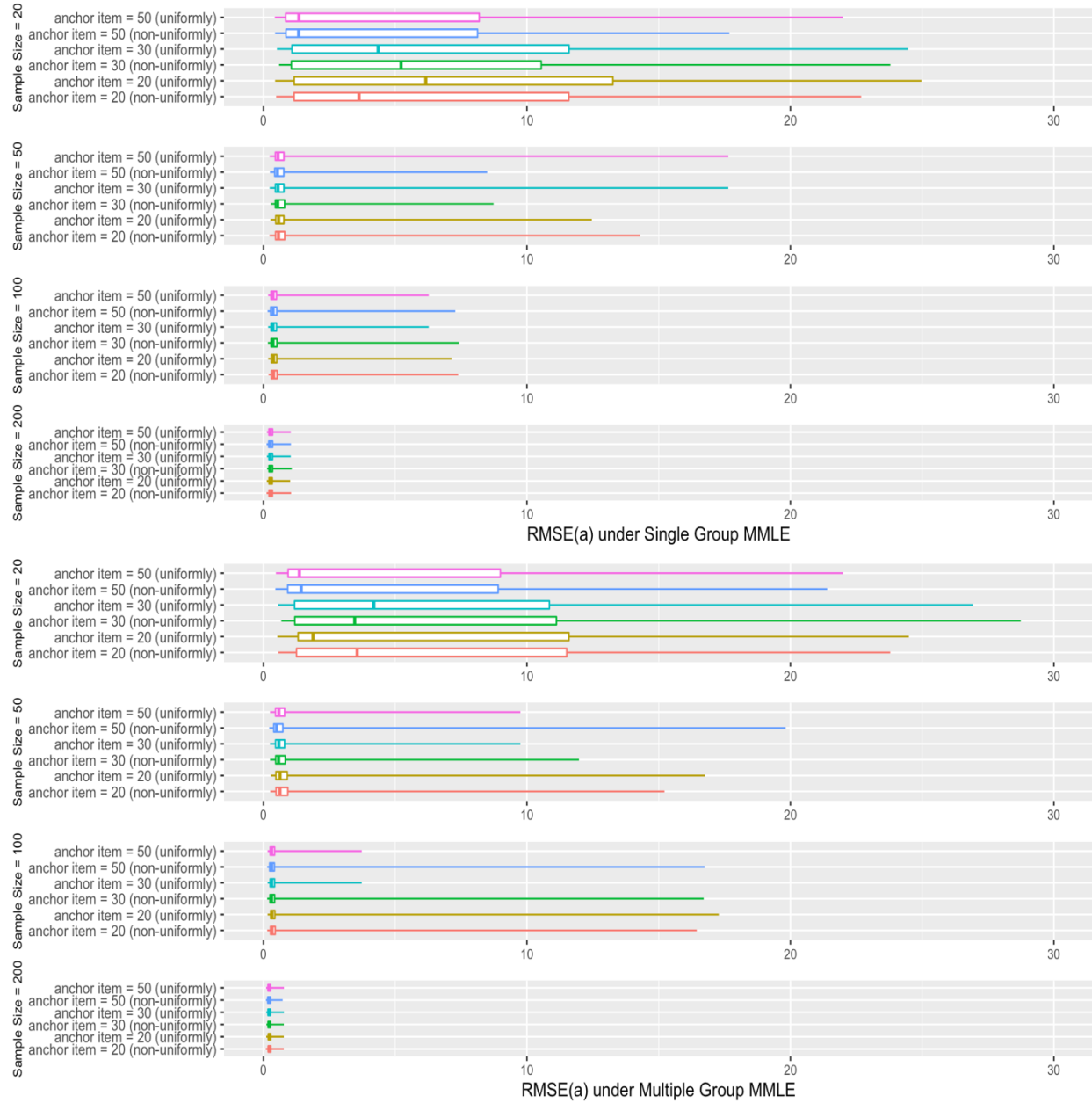


Figure 5.3. RMSE for a Parameters under 2-Form NEAT Design

In short, the magnitude relationship between RMSE for estimated a parameters obtained from two MMLE approaches of 2-form NEAT design was similar to that of 5-form NEAT design. That is, when sample size was 20 or 50, RMSE obtained for single group MMLE and multiple group MMLE were close. When sample size was 100 or 200, RMSE obtained from single group MMLE was larger than RMSE obtained from multiple group MMLE. RMSE dots

lied apart further as sample size increased. There was no significant difference of RMSE when using different number of anchor items or different anchor item selection strategies when sample size was 50 or 100 for both MMLE approaches. Detailed comparisons to compare RMSE across different sample sizes, different number of anchor items, and different anchor item selection strategies were presented in Appendix C7-C10. They respectively showed results under these three situations: when 20 uniformly selected anchor items were used, when a sample size of 200 and uniformly selected anchor items were used, and when a sample size of 200 and 20 uniformly selected anchor items were used.

Estimation of b parameters. For b parameter estimation, number of extreme estimation values, bias, and RMSE were presented.

Number of extreme estimation values. Figure 5.4 and Appendix C11 document the number of extreme estimations values when estimating b parameters by single group MMLE and multiple group MMLE.

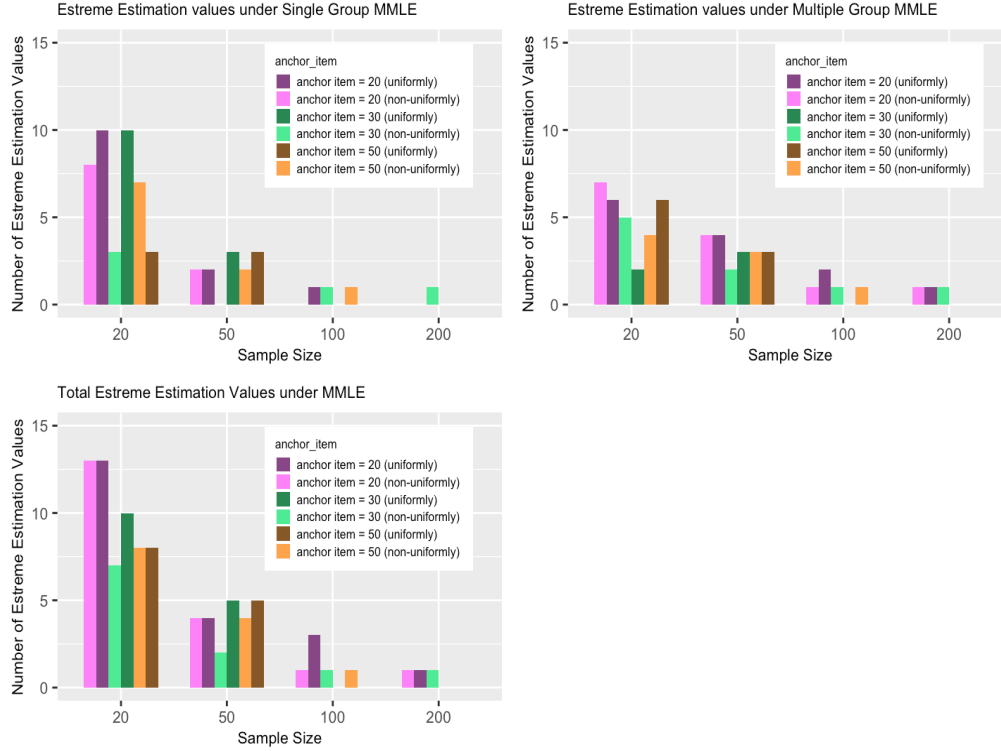


Figure 5.4. Number of Extreme Estimation Values for b Parameters under 2-Form NEAT Design

Comparison across different sample sizes. When sample size was 20, single group MMLE identified more extreme estimation values than multiple group MMLE when there were 20 anchor items, 30 uniformly selected anchor items or 50 non-uniformly selected anchor items. In the rest cases, single group MMLE identified less extreme estimation values than multiple group MMLE. When sample size was 50, single group MMLE identified less extreme estimation values than multiple group MMLE (except when there were 30 and 50 uniformly selected anchor items, two MMLE approaches identified the same number). When sample size was 100 or 200, single group MMLE identified less extreme estimation values than multiple group MMLE when there were 20 anchor items. But two MMLE approaches identified the same number of extreme estimation values when there were 30 or 50 anchor items. In addition, for both MMLE approaches, as sample size increased, the number of extreme estimation values decreased.

Comparison across different number of anchor items. When there were 20 anchor items, single group MMLE identified more extreme estimation values than multiple group MMLE when sample size was 20; while it identified less extreme estimation values than multiple group MMLE when sample size was 50, 100, or 200. When there were 30 anchor items, single group MMLE identified more extreme estimation values than multiple group MMLE when sample size was 20 and anchor items were uniformly selected; it identified less extreme estimation values than multiple group MMLE when sample size was 20 and anchor items were non-uniformly selected, or when sample size was 20 and anchor items were non-uniformly selected. In the rest cases, two MMLE approaches identified the same number of extreme estimation values. When there were 50 anchor items, single group MMLE identified more extreme estimation values than multiple group MMLE when sample size was 20 and anchor items were non-uniformly selected; it identified less extreme estimation values than multiple group MMLE when sample size was 20 and anchor items were uniformly selected, or when sample size was 50 and anchor items were non-uniformly selected. In the rest cases, two MMLE approaches identified the same number of extreme estimation values.

Comparison across different anchor item selection strategies. When anchor items were uniformly selected, single group MMLE identified more extreme estimation values than multiple group MMLE when sample size was 20 (except when there were 50 anchor items, single group MMLE identified less); while two approaches identified the same number of extreme estimation values when sample size was 50, 100 or 200 (except when there were 20 anchor items, single group MMLE identified less). When anchor items were non-uniformly selected, single group MMLE identified more extreme estimation values than multiple group MMLE when sample size was 20 (except when there were 30 anchor items, single group MMLE identified less); it

identified less extreme estimation values than multiple group MMLE when sample size was 50; and two MMLE approaches identified the same number of extreme estimation values when sample size was 100 or 200 (except when there were 20 anchor items, single group MMLE identified less).

In general, multiple group MMLE was more stable than single group MMLE when sample size was 20 (except when 30 non-uniformly selected anchor items were used or when 50 uniformly selected anchor items were used, multiple single MMLE was more stable). Single group MMLE was more stable than multiple group MMLE when sample size was 50(except when sample size was 50 and 30 or 50 uniformly selected anchor items were used, two MMLE had the same stabilities). Two MMLE approaches had the same stabilities when sample size was 100 or 200 (except when 20 anchor items were used, single group MMLE was more stable).

Bias. Figure 5.5 presents boxplots of bias of estimated b parameters from single group MMLE and multiple group MMLE. The five number summaries can be found in Appendix C12.

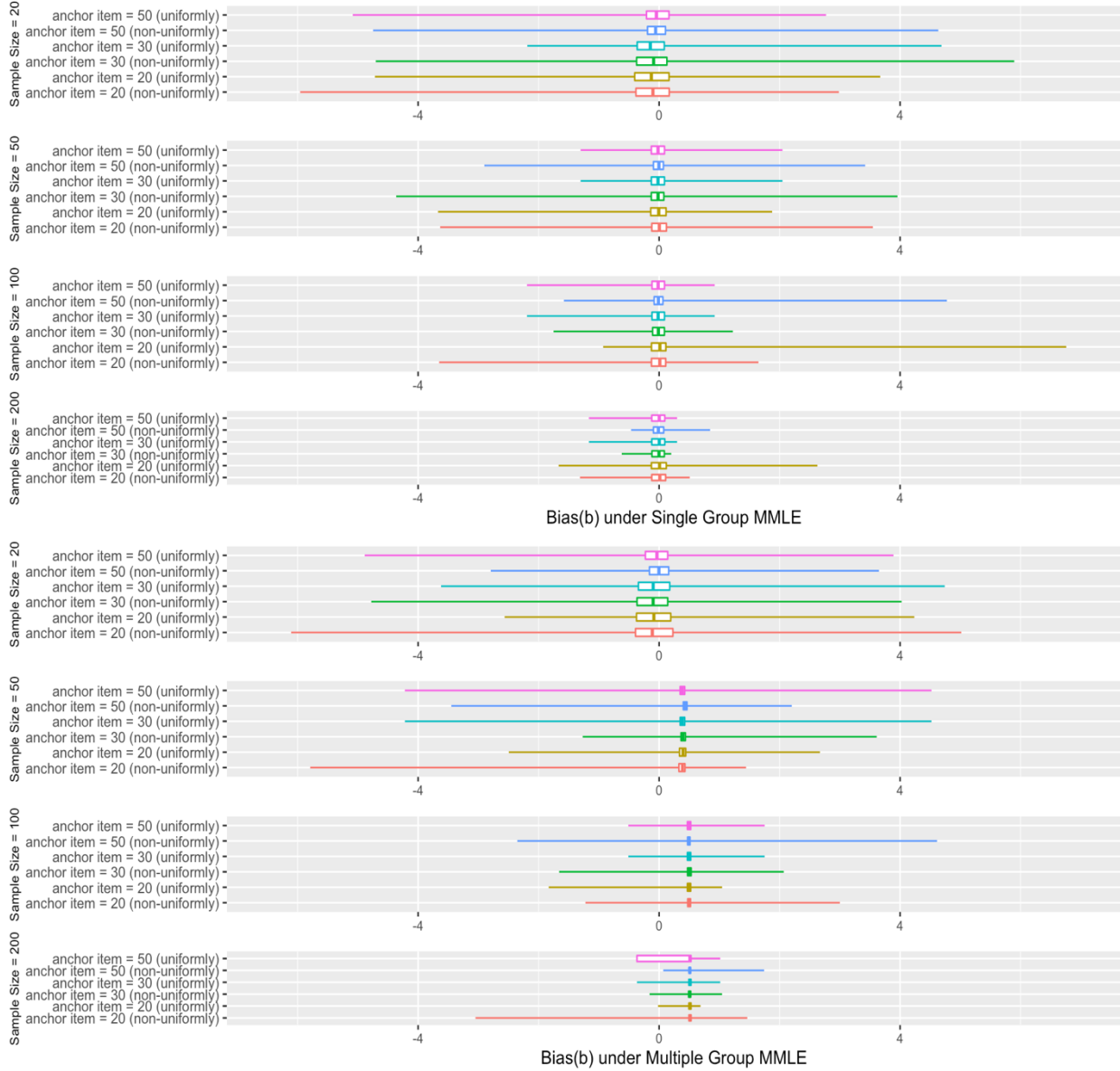


Figure 5.5. Bias for b Parameters under 2-Form NEAT Design

Comparison across different sample sizes. According to Figure 5.5, it was found that bias obtained from single group MMLE was closer to zero than bias obtained from multiple group MMLE when sample size was 50, 100, or 200. When sample size was 20, there didn't exist significant difference between bias obtained from two MMLE approaches. The median was over zero when sample size was 50, 100, or 200 by multiple group MMLE, while the median in the rest of cases was around zero.

Figure 5.6 shows bias of estimated b parameters for each item when there were 20 uniformly selected anchor items. 1-150 item were items from easy to hard. When sample size was 20, both MMLE approaches overestimated b parameters for easy items and underestimated them for hard items. Bias dots obtained from two MMLE approaches were overlapped especially for easy items. When sample size was 50, 100, or 200, single group MMLE overestimated b parameters for easy items and underestimated them for hard items; while multiple group MMLE overestimated b parameters all the time. Bias from single group MMLE was closer to zero than bias from multiple group MMLE. In addition, as sample size increased, bias dots for both MMLE approaches were less scattered.

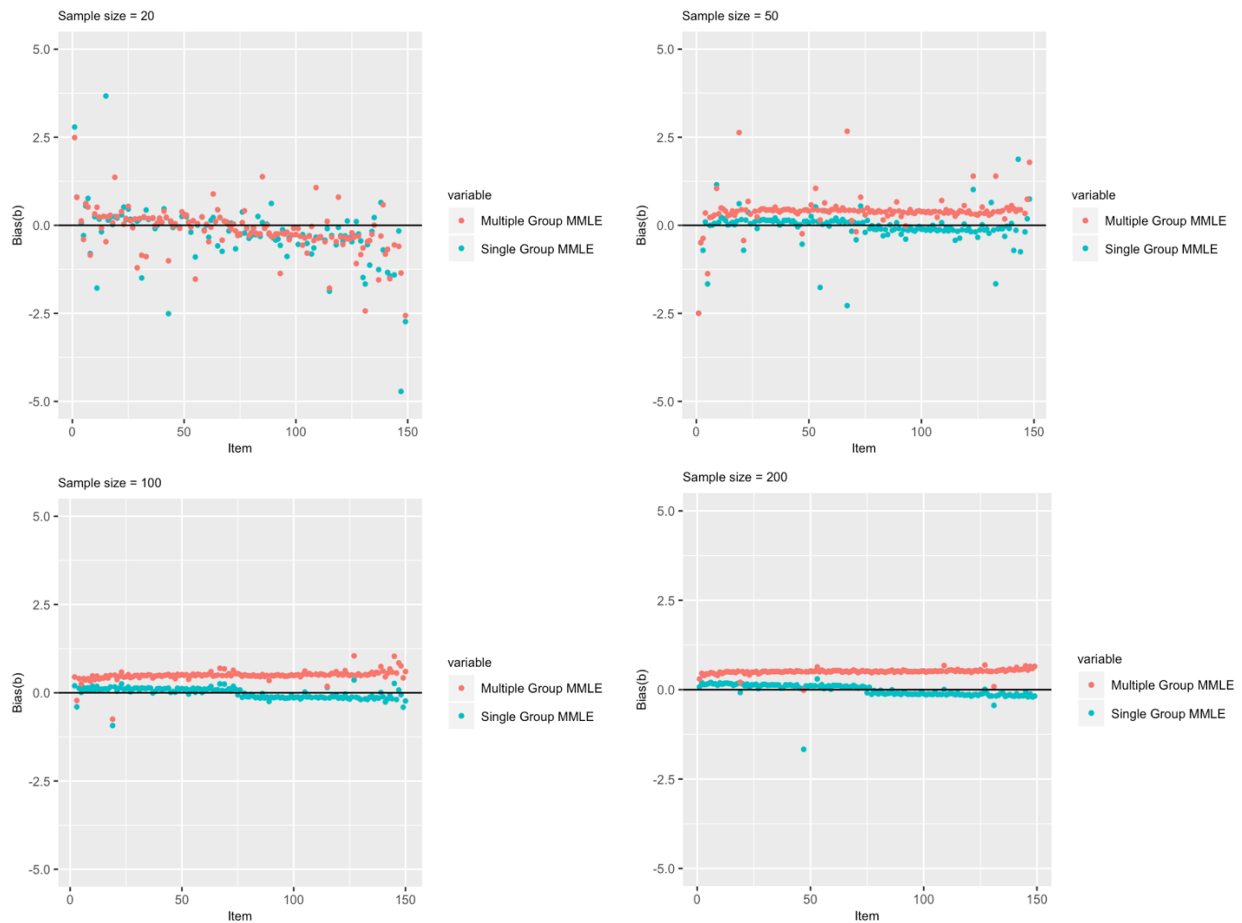


Figure 5.6. Bias for b Parameters under 2-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes

Comparison across different number of anchor items. Figure 5.5 presents that no significant difference of bias existed when using different number of anchor items for these two MMLE approaches.

Figure 5.7 presents bias of estimated b parameters for each item when using uniformly selected anchor items and a sample size of 200. b parameters were overestimated for easy items and were underestimated for hard items by single group MMLE and were overestimated for all items by multiple group MMLE. Bias from single group MMLE was closer to zero than bias from multiple group MMLE. When the number of anchor items changed, bias obtained from two MMLE approaches changed little.

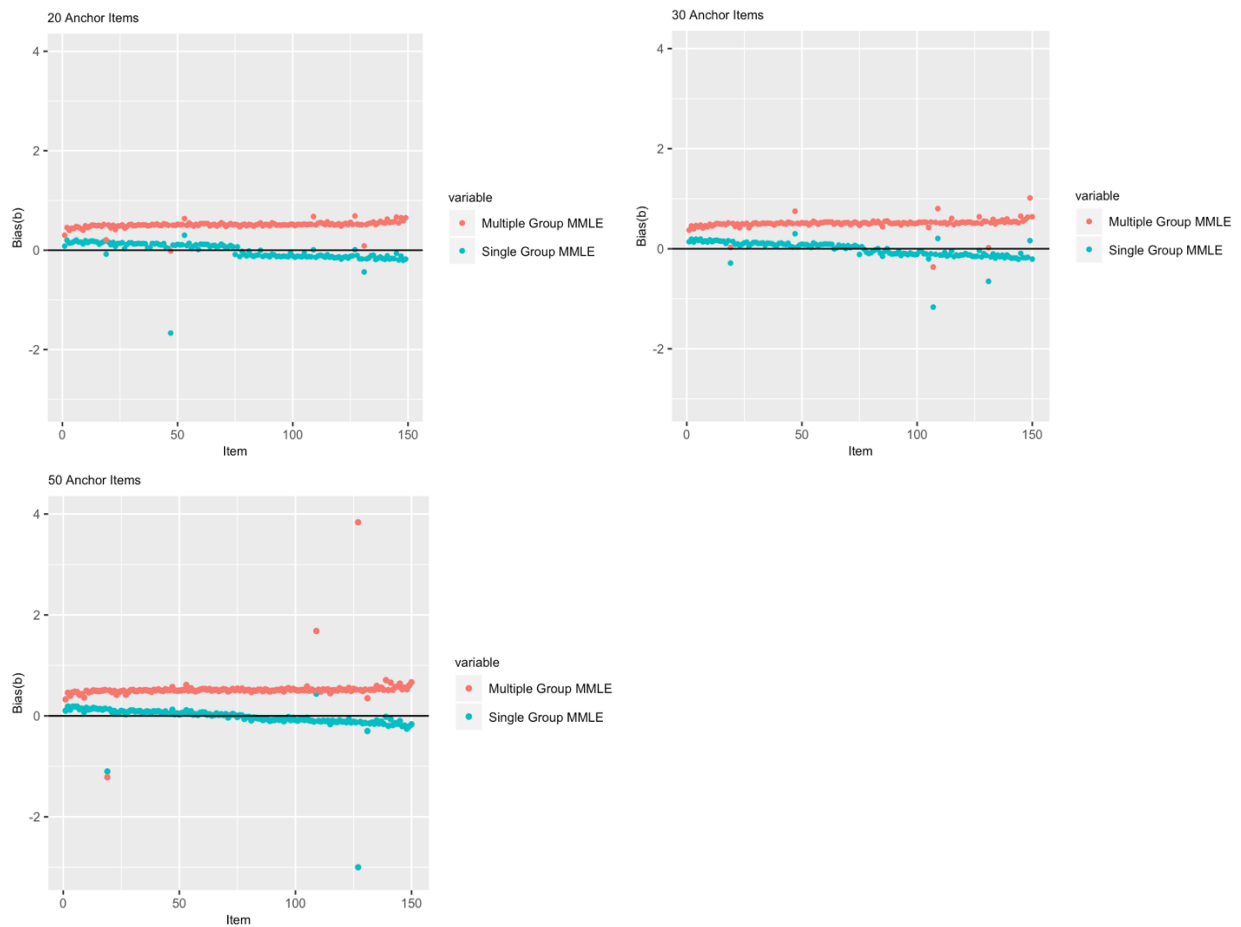


Figure 5.7. Bias for b Parameters under 2-Form NEAT Design with a Sample Size of 200 across Different Number of Uniformly Selected Anchor Items

Comparison across different anchor item selection strategies. Figure 5.5 shows that difference of bias obtained when using two anchor item selection strategies for two MMLE approaches was not that significant. But when there were 50 anchor items, bias obtained when using non-uniformly selected anchor items was closer to zero than using uniformly selected anchor items for two MMLE approaches.

Figure 5.8 presents estimated RMSE of estimated b parameters for each item when 50 anchor items and a sample size of 200 were used. Single group MMLE overestimated b parameters for easy items and underestimated them for hard items, while multiple group MMLE overestimated for all items. Bias observed for single group MMLE was closer to zero than bias observed for multiple group MMLE. Difference between two anchor item strategies was tiny.

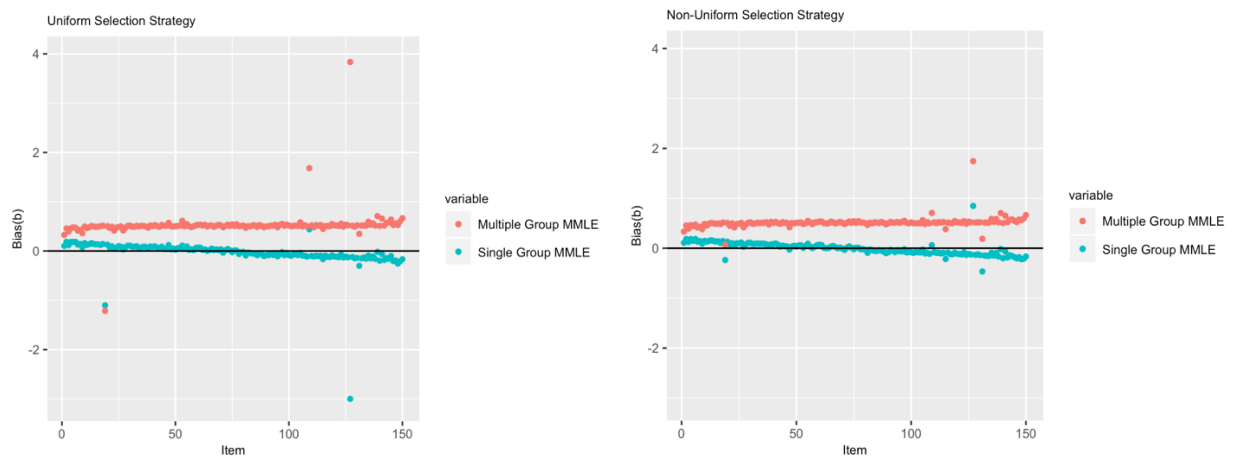


Figure 5.8. Bias for b Parameters under 2-Form NEAT Design with 50 Anchor Items and a Sample Size of 200 across Different Anchor Item Selection Strategies

RMSE. Figure 5.9 presents boxplots of RMSE of estimated b parameters from single group MMLE and multiple group MMLE. The five number summaries can be found in Appendix C13.



Figure 5.9. RMSE for b Parameters under 2-Form NEAT Design

Comparison across different sample sizes. Figure 5.9 shows that RMSE obtained from single group MMLE was not much different from RMSE obtained from multiple group MMLE regardless of sample sizes. As sample size increase, RMSE obtained from both MMLE approaches increased.

Figure 5.10 presents RMSE of estimated b parameters for each item when there were 20 uniformly selected anchor items. Figure 5.11 presents RMSE smaller than 2 to take a close look at the bottom of plots in Figure 5.10. When sample size was 20, RMSE observed for single group MMLE was larger than RMSE observed for multiple group MMLE. RMSE dots were mixed together. When sample size was 50, 100, or 200, RMSE observed for single group MMLE was smaller than RMSE observed for multiple group MMLE. RMSE dots observed for the two MMLE approaches lied apart. 1-150 item were items from easy to hard. For single group MMLE, RMSE for middle difficult items was closer to zero than RMSE for easy or hard items. For multiple group MMLE, RMSE for most items were closer to zero, except for very easy or very hard items, whose RMSE was very large. In addition, as sample size increased, RMSE dots were less scattered.

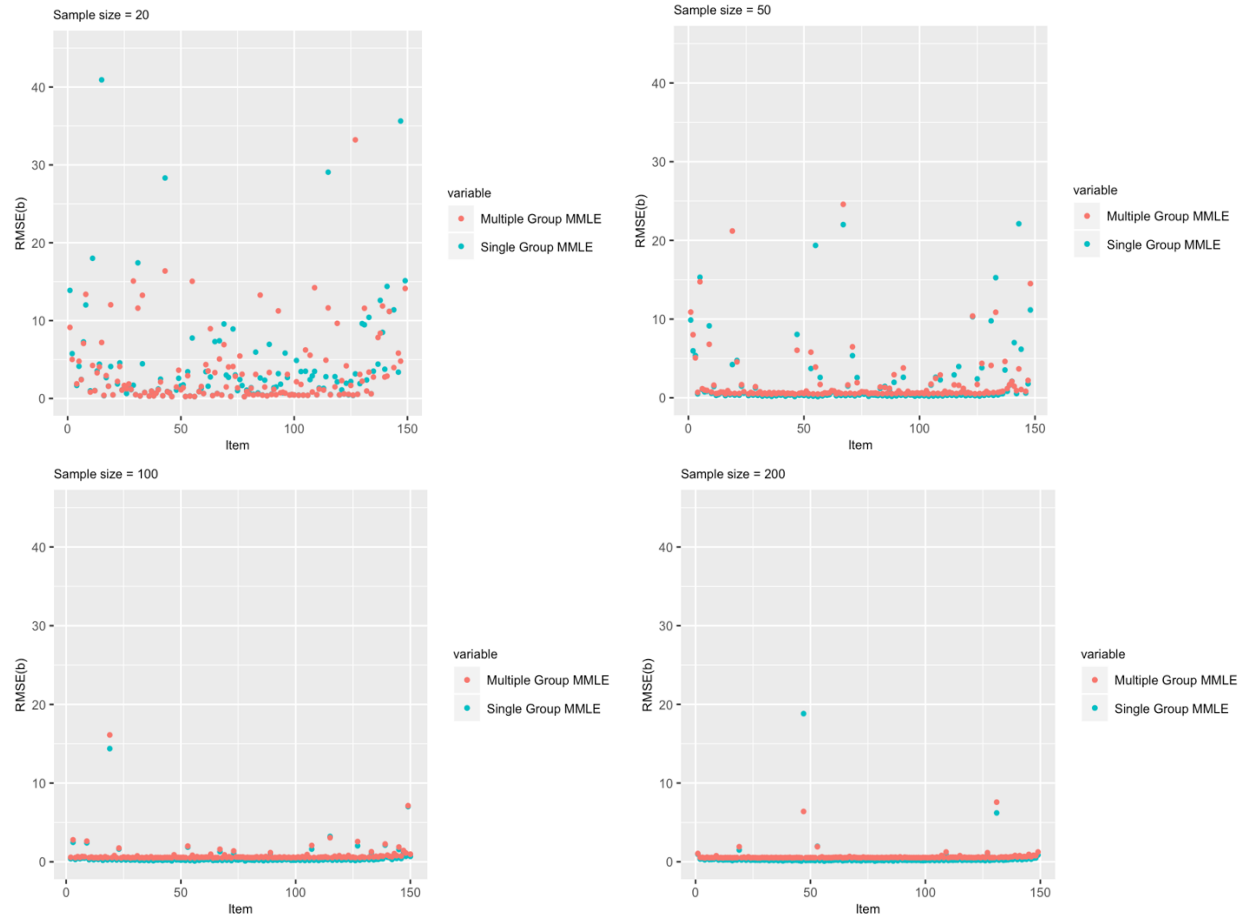


Figure 5.10. RMSE for b Parameters under 2-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes

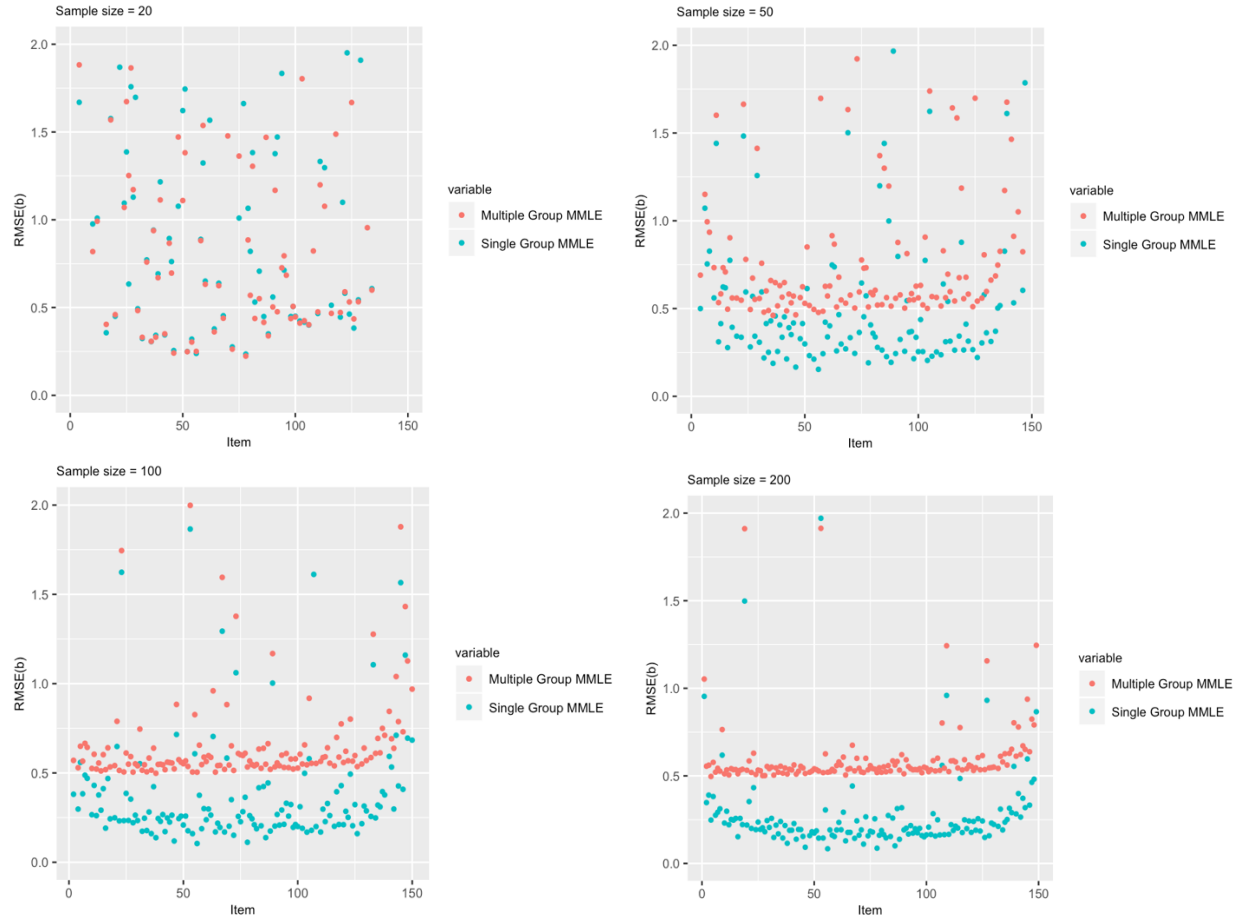


Figure 5.11. RMSE (from 0 to 2) for b Parameters under 2-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes

Comparison across different number of anchor items. Figure 5.9 shows that there existed no significant difference of bias when using different number of anchor items for two MMLE approaches.

Figure 5.12 and Figure 5.13 presents RMSE of estimated b parameters for each item when using uniformly selected anchor items and a sample size of 200. RMSE from single group MMLE was smaller than RMSE from multiple group MMLE. There was no significant difference when using different number of anchor items conditions for both MMLE approaches.

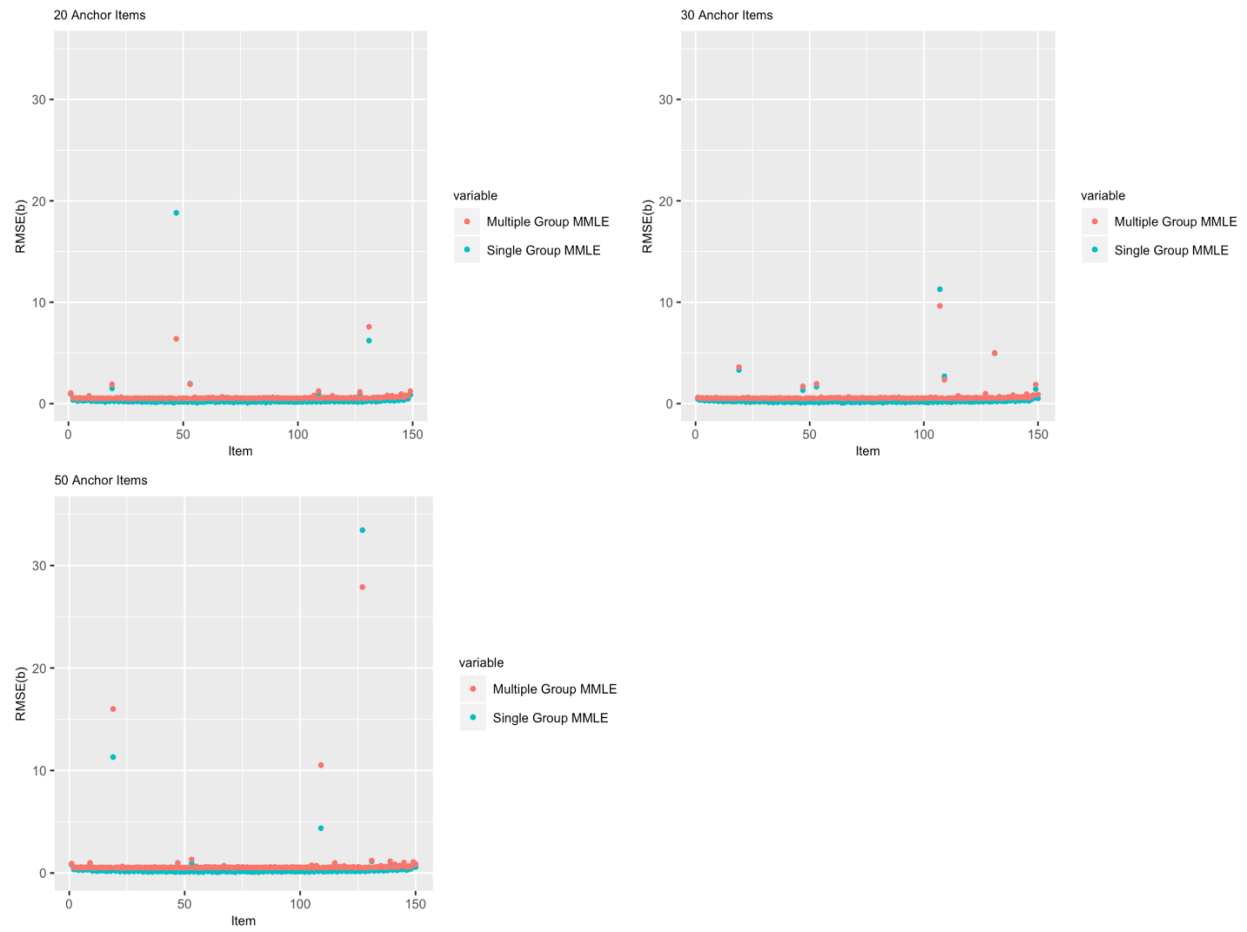


Figure 5.12. RMSE for b Parameters under 2-Form NEAT Design with a Sample Size of 200 across Different Number of Uniformly Selected Anchor Items

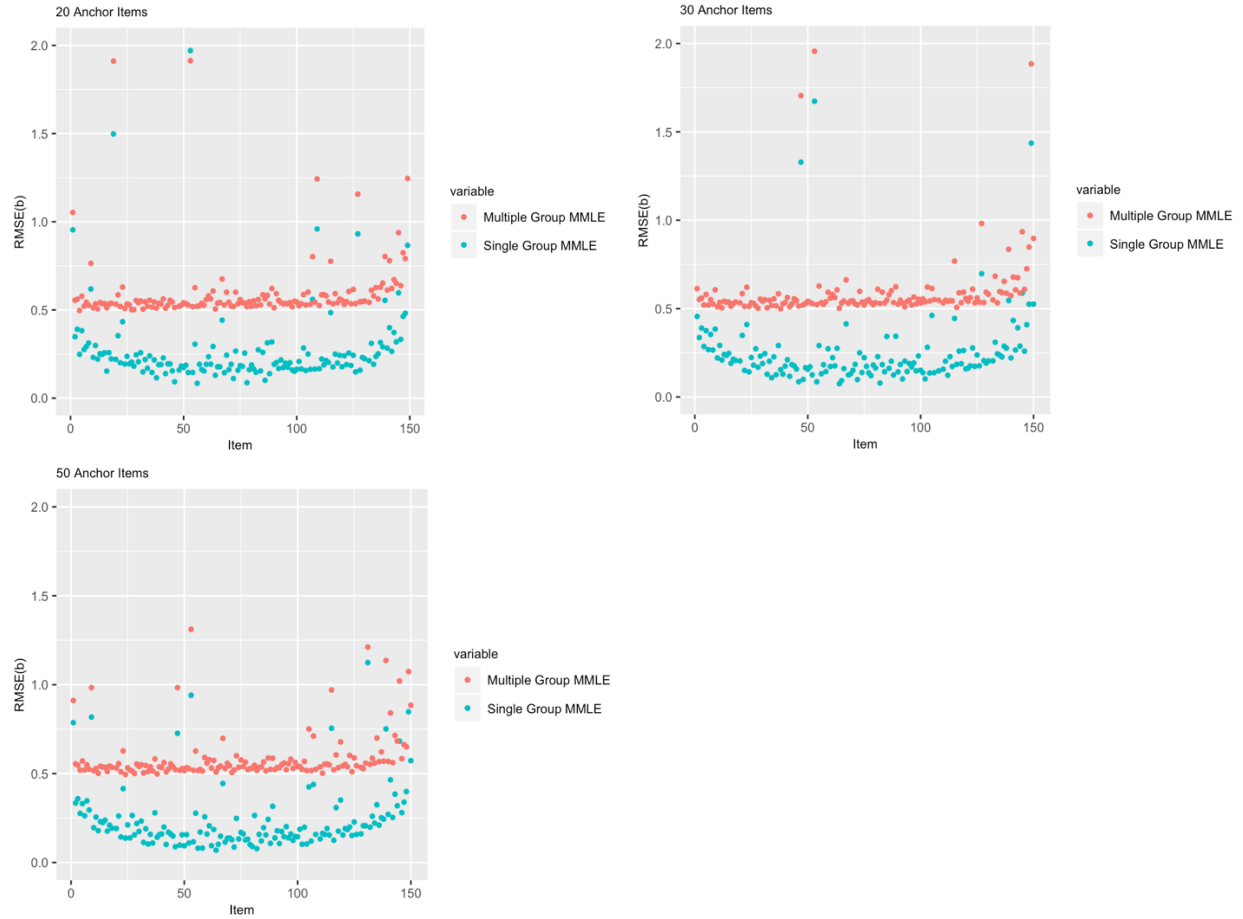


Figure 5.13. RMSE (from 0 to 2) for b Parameters under 2-Form NEAT Design with a Sample Size of 200 across Different Number of Uniformly Selected Anchor Items

Comparison across different anchor item selection strategies. Figure 5.9 shows that RMSE was close when using two anchor item selection strategies for both MMLE approaches.

Figure 5.14 presents estimated RMSE of estimated b parameters for each item when 20 anchor items and a sample size of 200. Figure 5.15 presents the RMSE that was smaller than 2 to take a close look at the bottom of the plots in Figure 5.14. RMSE obtained from single group MMLE was closer to that from multiple group MMLE. It was found that the difference of RMSE when using different anchor item selection strategies was not that significant for both two MMLE approaches.

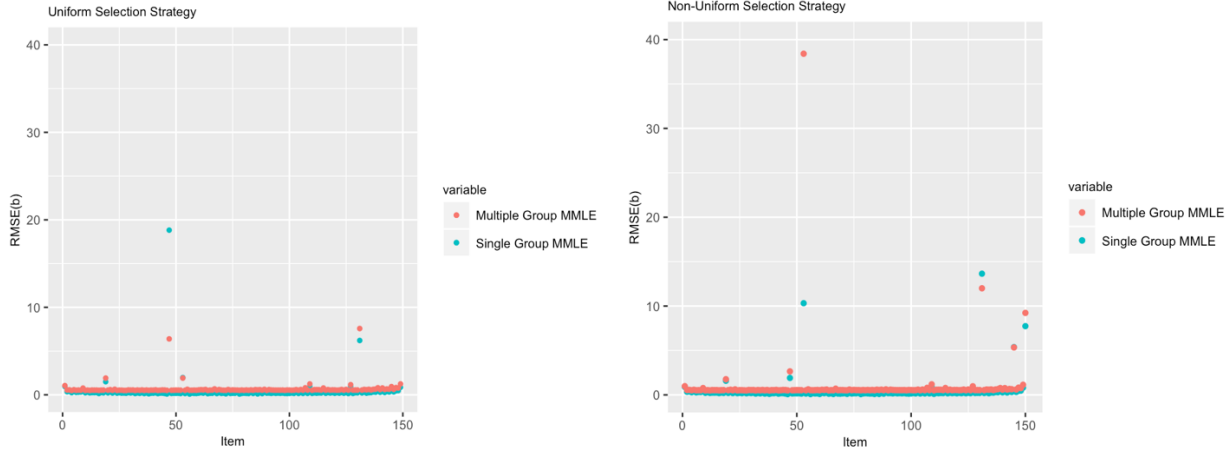


Figure 5.14. RMSE for b Parameters under 2-Form NEAT Design with 20 Anchor Items and a Sample Size of 200 across Different Anchor Item Selection Strategies

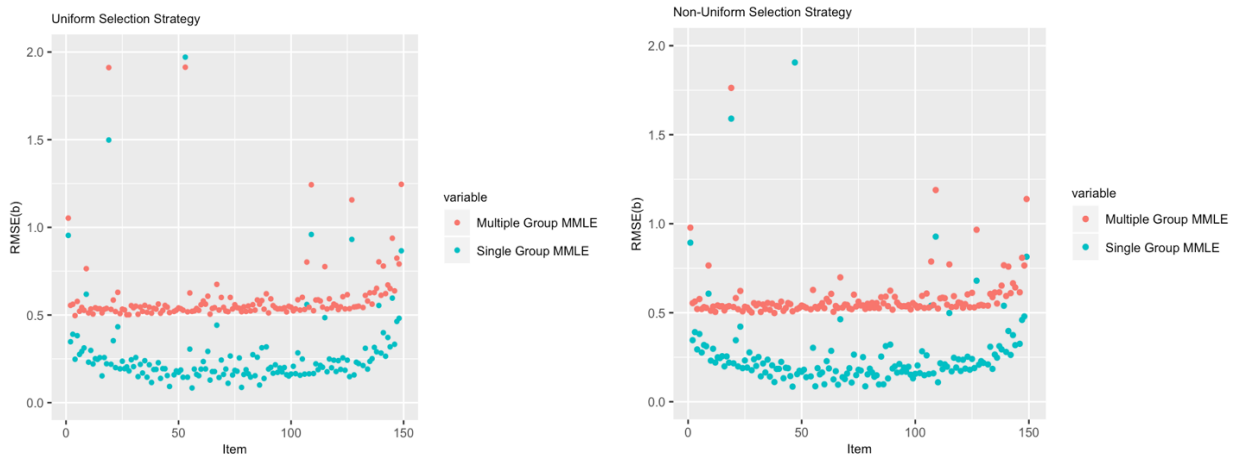


Figure 5.15. RMSE (from 0 to 2) for b Parameters under 2-Form NEAT Design with 20 Anchor Items and a Sample Size of 200 across Different Anchor Item Selection Strategies

MULTIFORM DESIGN

Under multiform design, only 5-form test design was used. As a reminder, three factors were investigated: sample size, number of anchor items, and anchor item strategy. To compare performances of single group MMLE and multiple group MMLE, the number of extreme estimation values, bias and RMSE obtained from two MMLE approaches were compared across different sample sizes, different number of anchor items and different anchor item strategies.

5-form multiform design

Estimation of a parameters. For a parameter estimation, number of extreme estimation values, bias, and RMSE were presented.

Number of extreme estimation values. Figure 6.1 and Appendix D1 document the number of extreme estimations values when estimating a parameters by single group MMLE and multiple group MMLE.

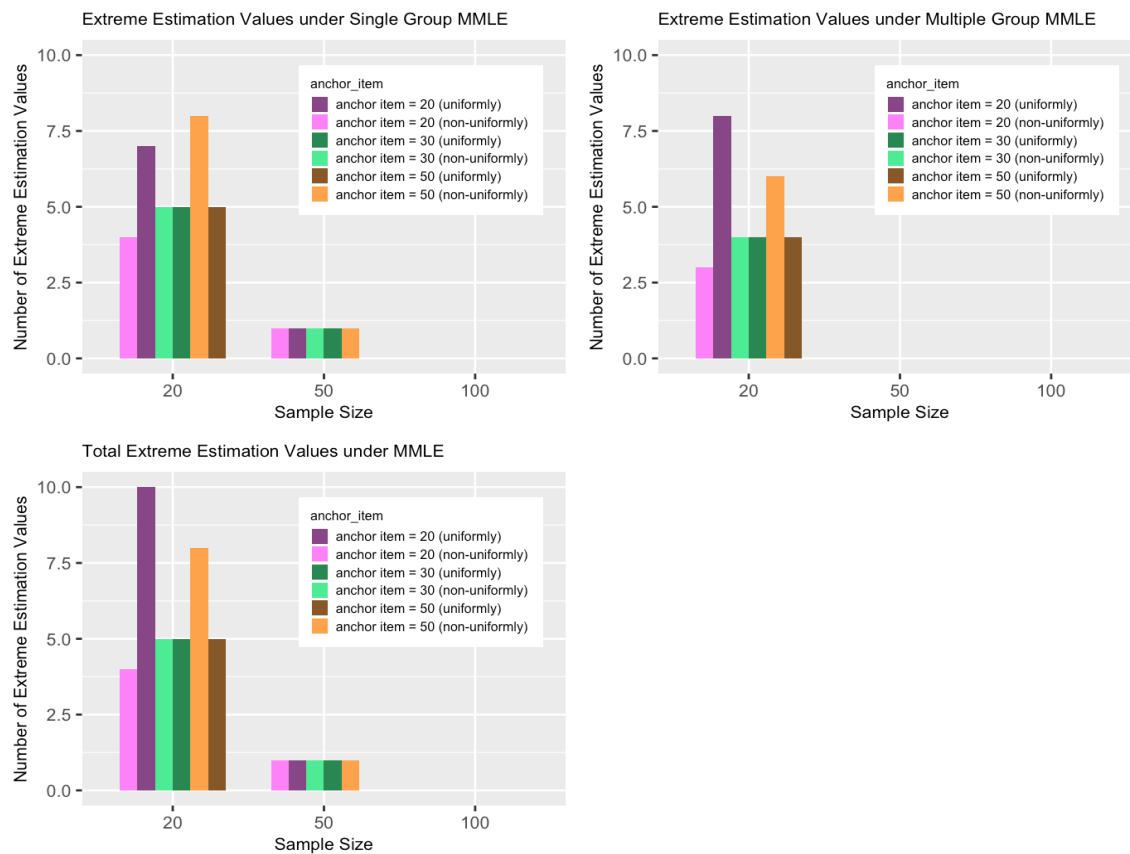


Figure 6.1. Number of Extreme Estimation Values for a Parameters under 5-Form Multiform Design

Comparison across different sample sizes. When sample size was 20, single group MMLE identified more extreme estimation values than multiple group MMLE (except when 20 uniformly selected anchor items were used, single group MMLE identified more). When sample size was 50, single group MMLE identified more extreme estimation values than multiple group

MMLE (except when 50 uniformly selected anchor items were used, two MMLE approaches identified the same number). When sample size was 100, both MMLE approaches identified no extreme estimation value. In addition, as sample size increased, the number of extreme estimations values decreased for both MMLE approaches.

Comparison across different number of anchor items. When there were 20 anchor items, single group MMLE identified less extreme estimation values than multiple group MMLE when sample size was 20 and anchor items were uniformly selected; it identified more extreme estimation values than multiple group MMLE when sample size was 20 and anchor items were non-uniformly selected, or when sample size was 50; both MMLE approaches identified no extreme estimation value when sample size was 100. When there were 30 anchor items, single group MMLE identified more extreme estimation values than multiple group MMLE when sample size was 20 or 50; both MMLE approaches identified no extreme estimation value when sample size was 100. When there were 50 anchor items, single group MMLE identified more extreme estimation values than multiple group MMLE when sample size was 20; single group MMLE identified more extreme estimation values than multiple group MMLE when sample size was 50 and anchor items were non-uniformly selected; two MMLE approaches identified no extreme estimation value when sample size was 50 and anchor items were uniformly selected, or when sample size was 100.

Comparison across different anchor item selection strategies. When anchor items were uniformly selected, single group MMLE identified more extreme estimation values than multiple group MMLE when sample size was 20 (except when 20 anchor items were used, single group MMLE identified less); it identified more extreme estimation values than multiple group MMLE when sample size was 50 (except when 50 anchor items were used, both MMLE approaches

identified none); both MMLE approaches identified none when sample size was 100. When anchor items were non-uniformly selected, single group MMLE identified more extreme estimation values than multiple group MMLE when sample size was 20 or 50; both MMLE approaches identified no extreme estimation values when sample size was 100.

In general, when sample size was 20 or 50, multiple group MMLE was more stable than single group MMLE. When sample size as 100, two MMLE approaches had the same stabilities.

Bias. Figure 6.2 presents boxplots of bias of estimated a parameters from single group MMLE and multiple group MMLE. The five number summaries can be found in Appendix D2. According to Figure 6.2, it can be seen that a parameters were overestimated by both MMLE approaches.

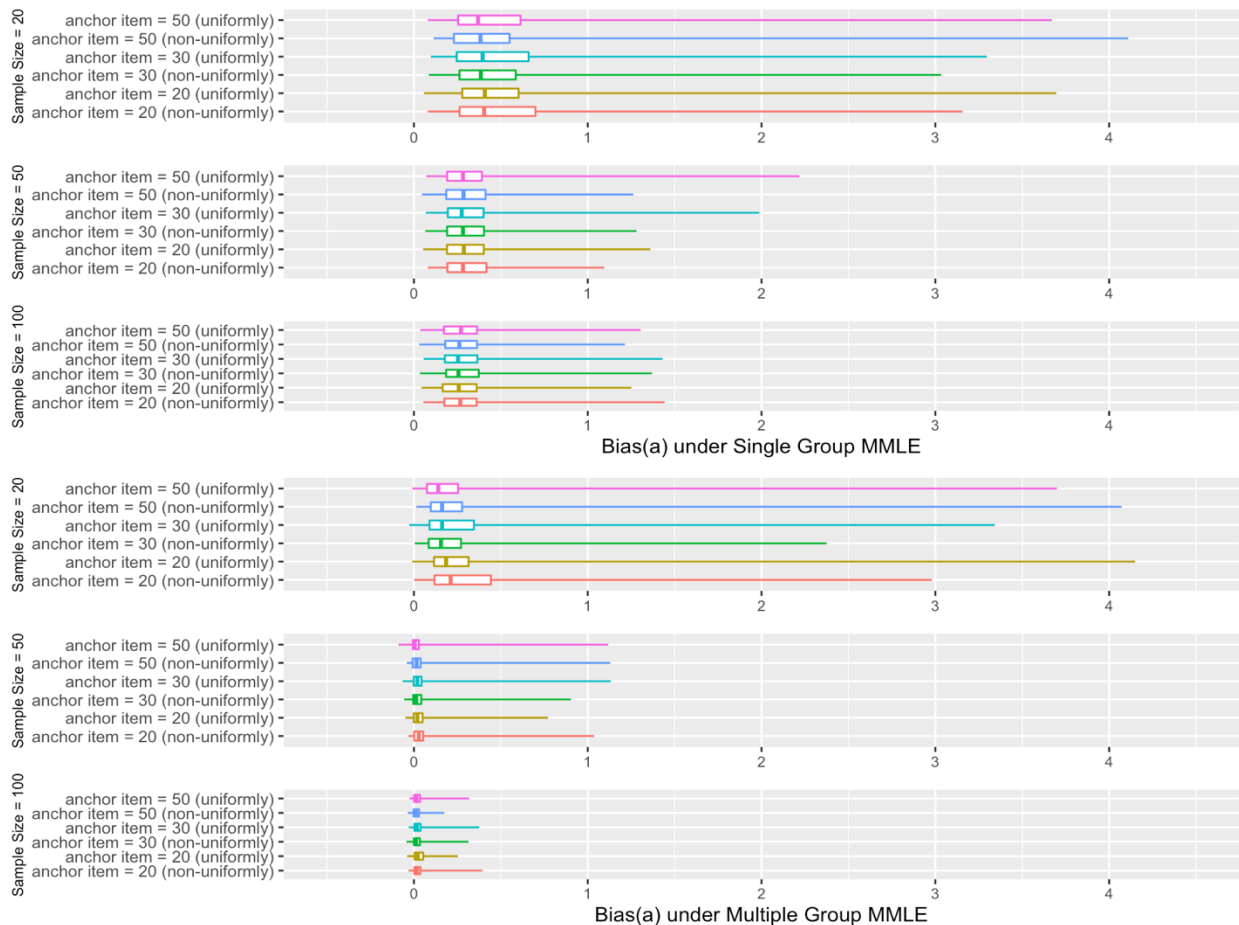


Figure 6.2. Bias for a Parameters under 5-Form Multiform Design

Comparison across different sample sizes. Figure 6.2 shows that bias obtained from multiple group MMLE was closer to zero than bias obtained from single group MMLE irrespective of sample size. When sample size was 50 or 100, range of bias obtained from multiple group MMLE was smaller than that from single group MMLE. In addition, for both MMLE approaches, as sample size increased, bias was closer to zero and the range of bias decreased.

Figure 6.3 presents bias of estimated a parameters for each item when there were 20 uniformly selected anchor items. Regardless of sample size, bias observed for single group MMLE was closer to zero than bias observed for multiple group MMLE. Bias observed for single group MMLE was closer to zero than bias observed for multiple group MMLE. As sample size increased, bias was closer to zero and bias range decreased for both MMLE approaches.

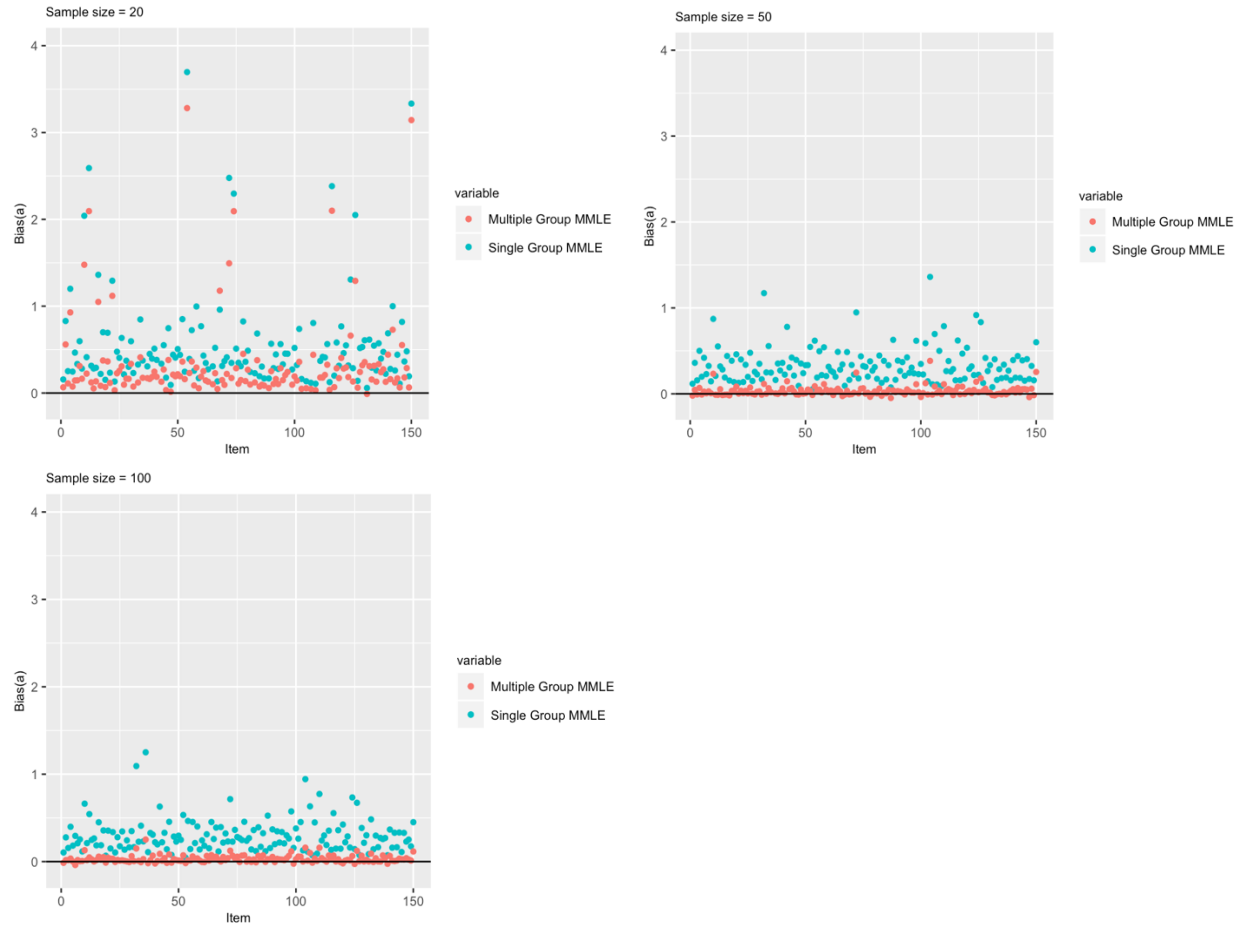


Figure 6.3. Bias for a Parameters under 5-Form Multiform Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes

Comparison across different number of anchor items. According to Figure 6.2, it can be seen that when using different number of anchor items, bias obtained from single group MMLE was closer to zero than bias obtained from multiple group MMLE.

Figure 6.4 shows bias of estimated a parameters for each item when using uniformly selected anchor items and a sample size of 100. Bias obtained from multiple group MMLE were closer to zero than bias obtained from single group MMLE. There was no significant difference of bias when using different number of anchor items for these two MMLE approaches.

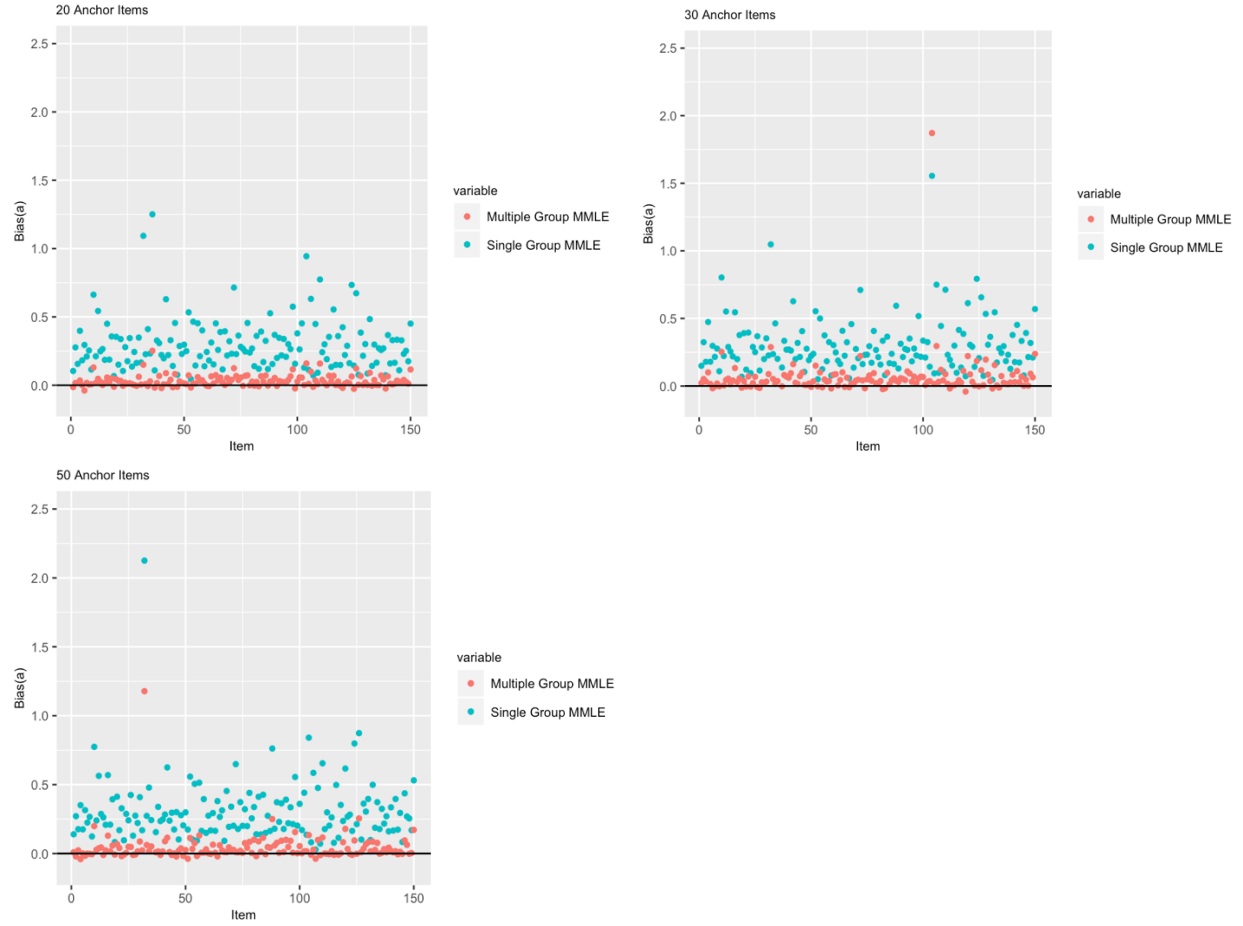


Figure 6.4. Bias for a Parameters under 5-Form Multiform Design with a Sample Size of 100 across Different Number of Uniformly Selected Anchor Items

Comparison across different anchor item selection strategies. Figure 6.2 presents that regardless of anchor item selection strategies, bias obtained from multiple group MMLE was closer to zero than bias obtained from single group MMLE. Specifically, when sample size was 20, there existed differences when using different anchor item selection strategies for both MMLE approaches. But the patterns of the differences were not consistent.

Figure 6.5 presents estimated bias of estimated a parameters for each item when using 20 anchor items and a sample size of 20. For both anchor item selection strategies, bias observed for multiple group MMLE was closer to zero than bias observed for single group MMLE. The

difference of bias was tiny when using different anchor item selection strategies for these two MMLE approaches.

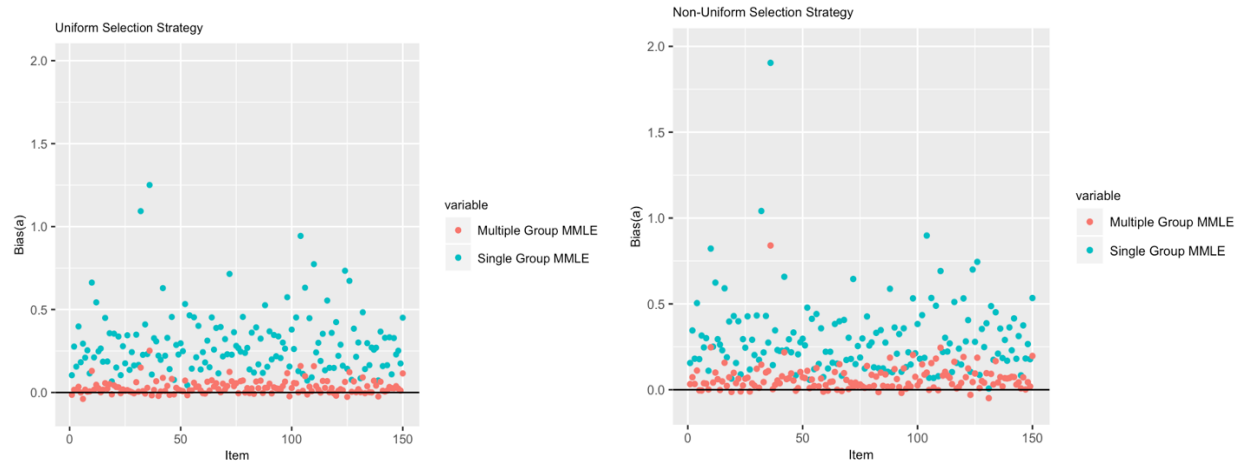


Figure 6.5. Bias for a Parameters under 5-Form Multiform Design with 20 Anchor Items and a Sample Size of 100 across Different Anchor Item Selection Strategies

RMSE. Figure 6.6 presents boxplots of RMSE of estimated a parameters from single group MMLE and multiple group MMLE. The five number summaries can be found in Appendix D3.

Comparison across different sample sizes. From Figure 6.6, it was found that when sample size was 50 or 100, RMSE observed for single group MMLE was larger than RMSE observed for multiple group MMLE. But when sample size was 20, RMSE observed for two MMLE approaches weren't much different. In addition, as sample size increased, RMSE decreased for both MMLE approaches.

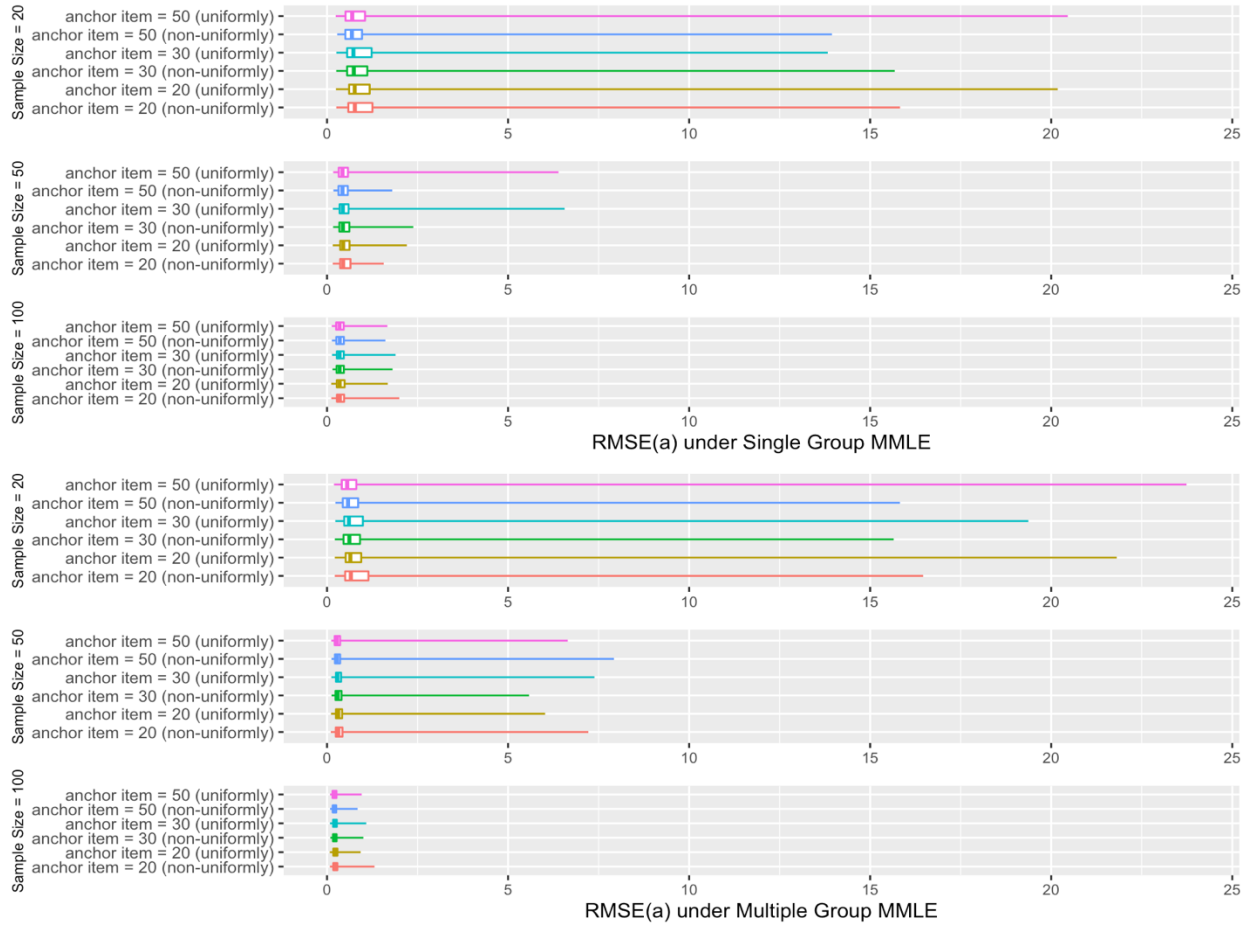


Figure 6.6. RMSE for a Parameters under 5-Form Multiform Design

Figure 6.7 shows RMSE of estimated a parameters for each item when there were 20 uniformly selected anchor items. Figure 6.8 shows RMSE of estimated a parameters after dropping very large values. When sample size was 20, RMSE obtained from single group MMLE was larger than RMSE obtained from multiple group MMLE. RMSE dots obtained from two MMLE approaches mixed together. When sample size was 50, RMSE obtained from single group MMLE was larger than RMSE obtained from multiple group MMLE. RMSE dots obtained from two MMLE approaches lied apart further as sample size increased. In addition, RMSE was closer to zero and the range of it was decreased as sample size increased for both MMLE approaches.

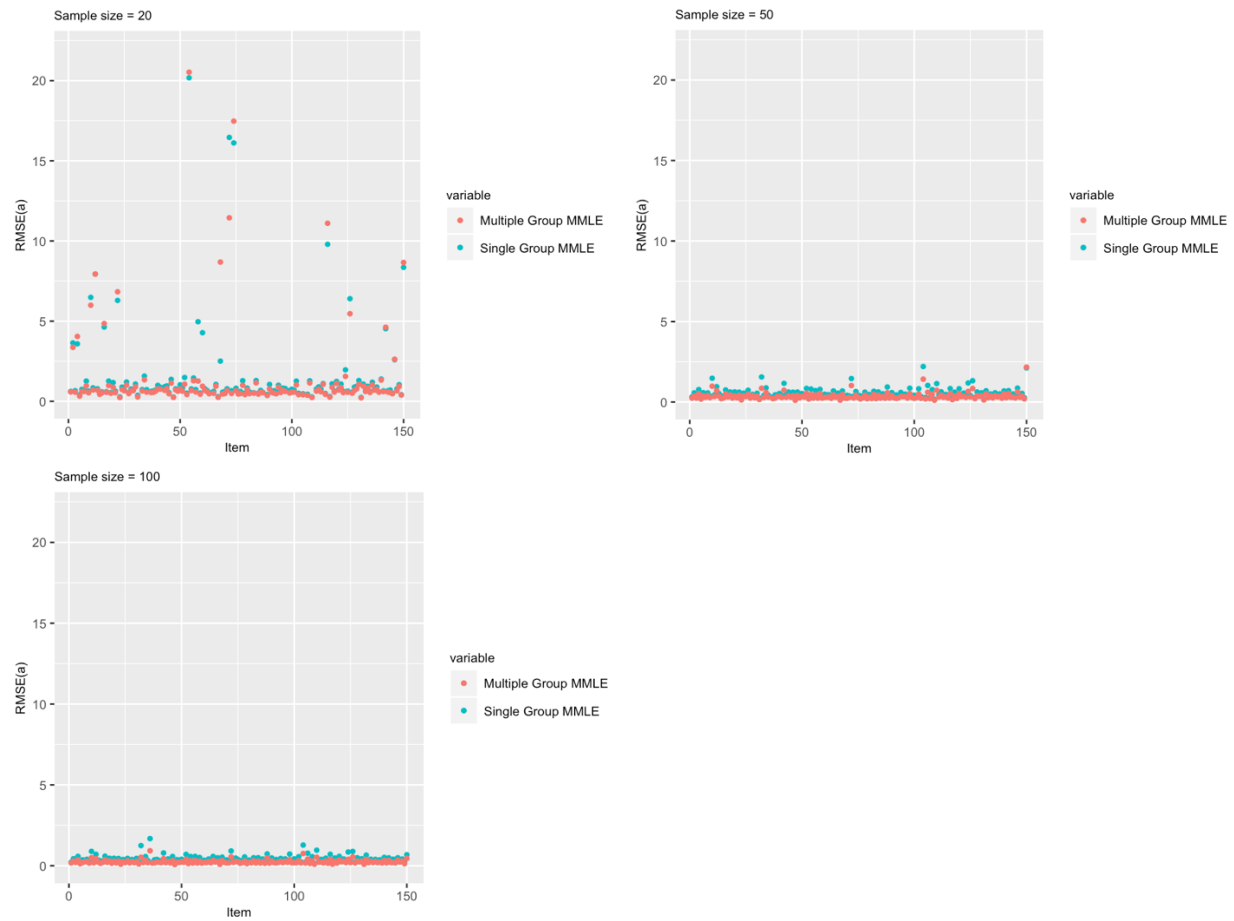


Figure 6.7. RMSE for a Parameters under 5-Form Multiform Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes

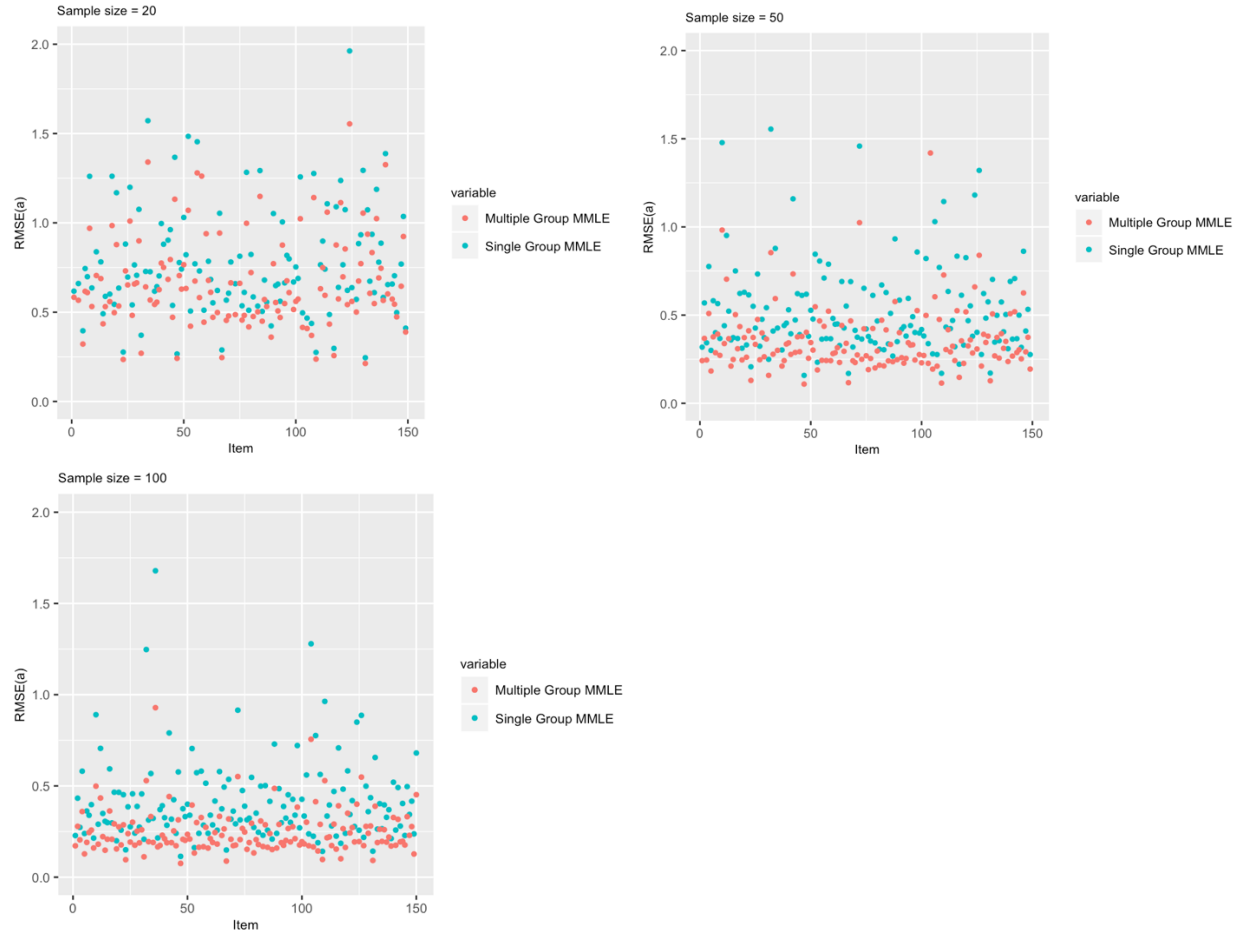


Figure 6.8. RMSE (from 0 to 2) for a Parameters under 5-Form Multiform Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes

Comparison across different number of anchor items. Figure 6.6 shows that regardless of number of anchor items, RMSE observed for single group MMLE was larger than multiple group MMLE except when sample size was 20.

Figure 6.9 presents RMSE of estimated a parameters for each item when using uniformly selected anchor items and a sample size of 100. RMSE obtained from multiple group MMLE was smaller than RMSE obtained from single group MMLE. There was no significant difference of RMSE when using different number of anchor items for these two MMLE approaches.

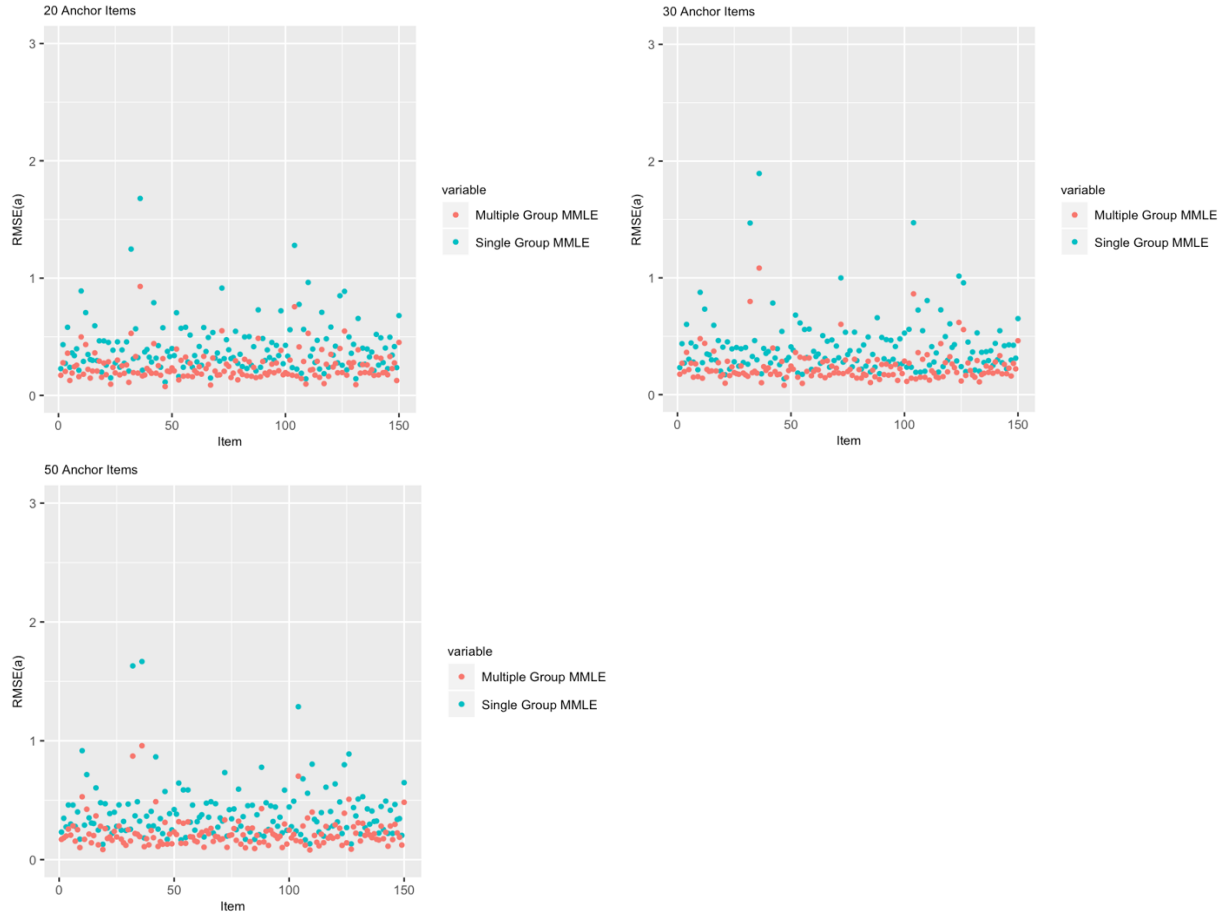


Figure 6.9. RMSE for a Parameters under 5-Form Multiform Design with a Sample Size of 100 across Different Number of Uniformly Selected Anchor Items

Comparison across different anchor item selection strategies. Figure 6.6 shows that regardless of anchor item selection strategies, RMSE observed for single group MMLE was larger than multiple group MMLE except when sample size was 20.

Figure 6.10 presents estimated RMSE of estimated a parameters for each item when 20 anchor items and a sample size of 100. RMSE obtained from multiple group MMLE was smaller than RMSE obtained from single group MMLE. The difference of RMSE was tiny when using different anchor item selection strategies for the two MMLE approaches.



Figure 6.10. RMSE for a Parameters under 5-Form Multifactor Design with 20 Anchor Items and a Sample Size of 100 across Different Anchor Item Selection Strategies

Estimation of b parameters. For b parameter estimation, number of extreme estimation values, bias, and RMSE were presented.

Number of extreme estimation values. Figure 6.11 and Appendix D4 document the number of extreme estimations values when estimating b parameters by single group MMLE and multiple group MMLE.

Comparison across different sample sizes. When sample size was 20, single group MMLE identified more extreme estimation values than multiple group MMLE (except when 30 uniformly selection anchor items were used, it identified less; or when 50 non-uniformly anchor items were used, two MMLE approaches identified the same number). When sample size was 50, single group MMLE identified more extreme estimation values than multiple group MMLE (except when 20 uniformly selection anchor items or 50 non-uniformly anchor items were used, both MMLE approaches identified none). When sample size was 100, both MMLE approaches identified no extreme estimation value. In addition, when sample size decreased, both MMLE approaches identified less extreme estimation values.

Comparison across different number of anchor items. When there were 20 anchor items, single group MMLE identified more extreme estimation values than multiple group MMLE when sample size was 20, or when sample size was 50 and anchor items were non-uniformly selected. Two MMLE approaches identified no extreme estimation value when sample size was 100, or when sample size was 50 and anchor items were uniformly selected. When there were 30 anchor items, single group MMLE identified less extreme estimation values than multiple group MMLE when sample size was 20 and anchor items were uniformly selected. But it identified more than multiple group MMLE when sample size was 20 and anchor items were non-uniformly selected, or when sample size was 50. When sample size was 100, both MMLE approaches identified no extreme estimation values. When there were 50 anchor items, single group MMLE identified more extreme estimation values than multiple group MMLE when sample size was 20 and anchor items were uniformly selected, or when sample size was 50 and anchor items were uniformly selected; In the rest of cases, two MMLE approaches identified the same number of extreme estimation values.

Comparison across different anchor item selection strategies. When anchor items were uniformly selected, single group MMLE identified more extreme estimation values than multiple group MMLE when sample size was 20 (except when 30 anchor items were used, single group MMLE identified less). Single group MMLE identified more extreme estimation values than multiple group MMLE when sample size was 50 (except when 20 anchor items were used, two MMLE approaches identified none). Both MMLE approaches identified no extreme estimation values when sample size was 100. When anchor items were non-uniformly selected, single group MMLE identified more extreme estimation values than multiple group MMLE when sample size was 20 or 50 (except when 50 anchor items were used, two MMLE approaches identified the

same number). Both MMLE approaches identified no extreme estimation values when sample size was 100.

In general, multiple group MMLE was more stable than single group MMLE when sample size was 20 or 50. Two MMLE approaches have the same stabilities when sample size was 100.

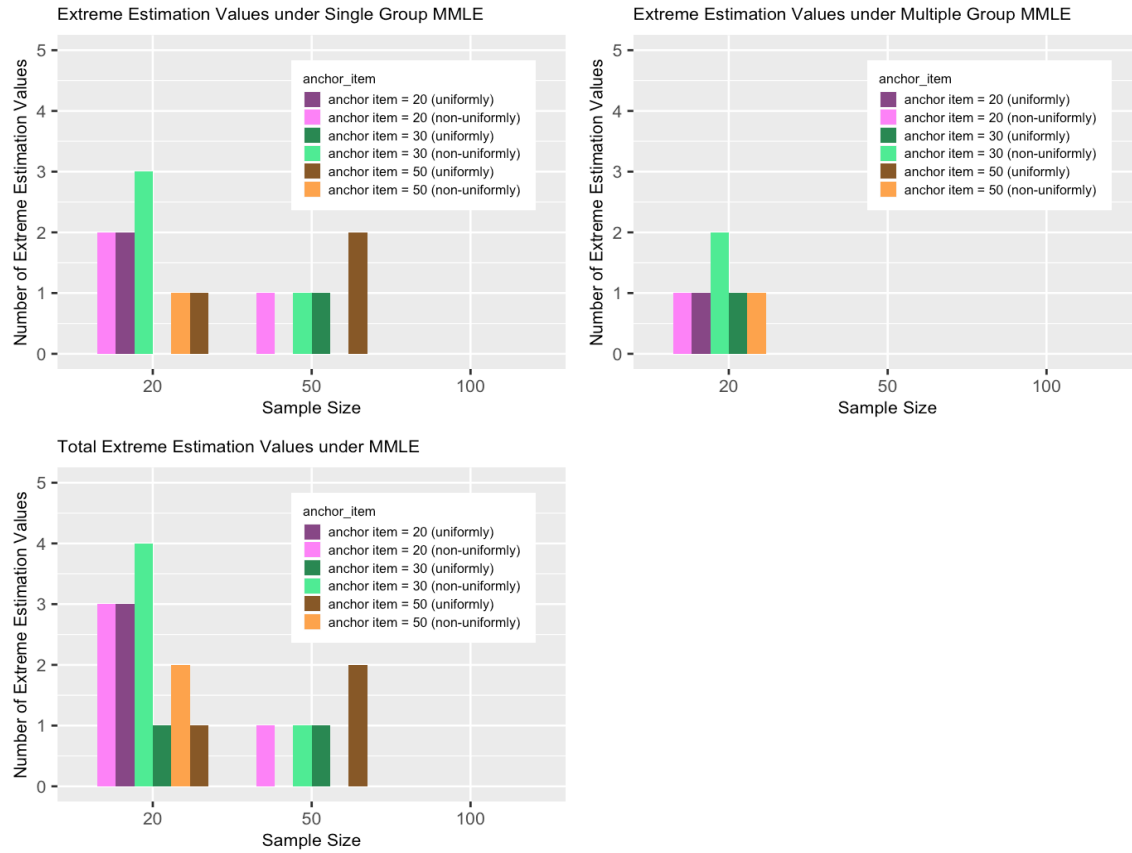


Figure 6.11. Number of Extreme Estimation Values for b Parameters under 5-Form Multiform Design

Bias. Figure 6.12 presents boxplots of bias of estimated b parameters from single group MMLE and multiple group MMLE. The five number summaries can be found in Appendix D5.

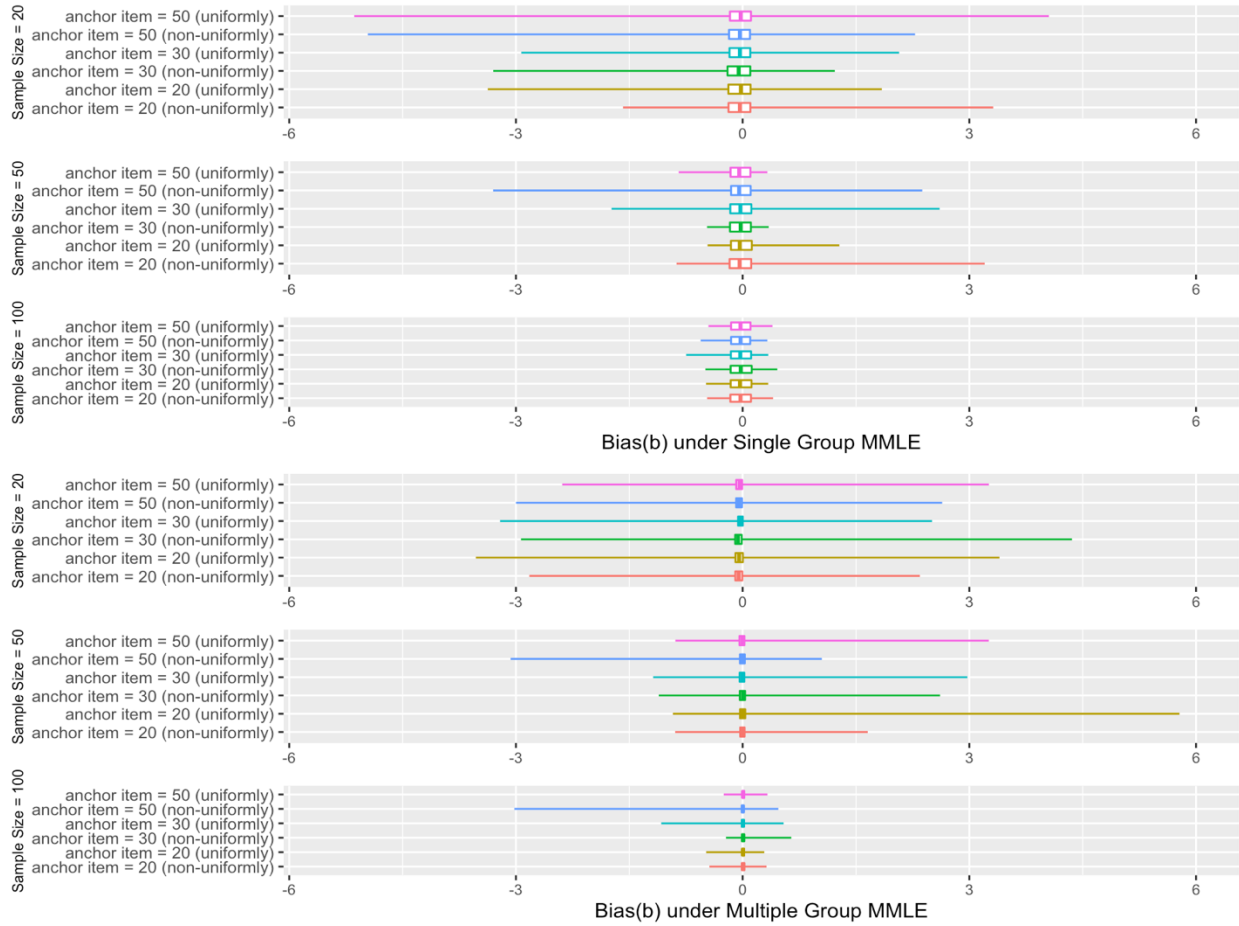


Figure 6.12. Bias for b Parameters under 5-Form Multiform Design

Comparison across different sample sizes. From Figure 6.12, it can be seen that regardless of sample size, bias obtained from multiple group MMLE was closer to zero than bias obtained from single group MMLE. For multiple group MMLE, as sample size increased, bias was closer to zero.

Figure 6.13 presents bias of estimated b parameters for each item when there were 20 uniformly selected anchor items. Items 1-150 were items from easy to hard. Irrespective of sample size, single group MMLE overestimated b parameters for easy items and underestimated b parameters for hard items while multiple group MMLE was almost unbiased. In addition, as

sample size increased, for both MMLE approaches, range of bias decreased, and bias dots were less scattered.

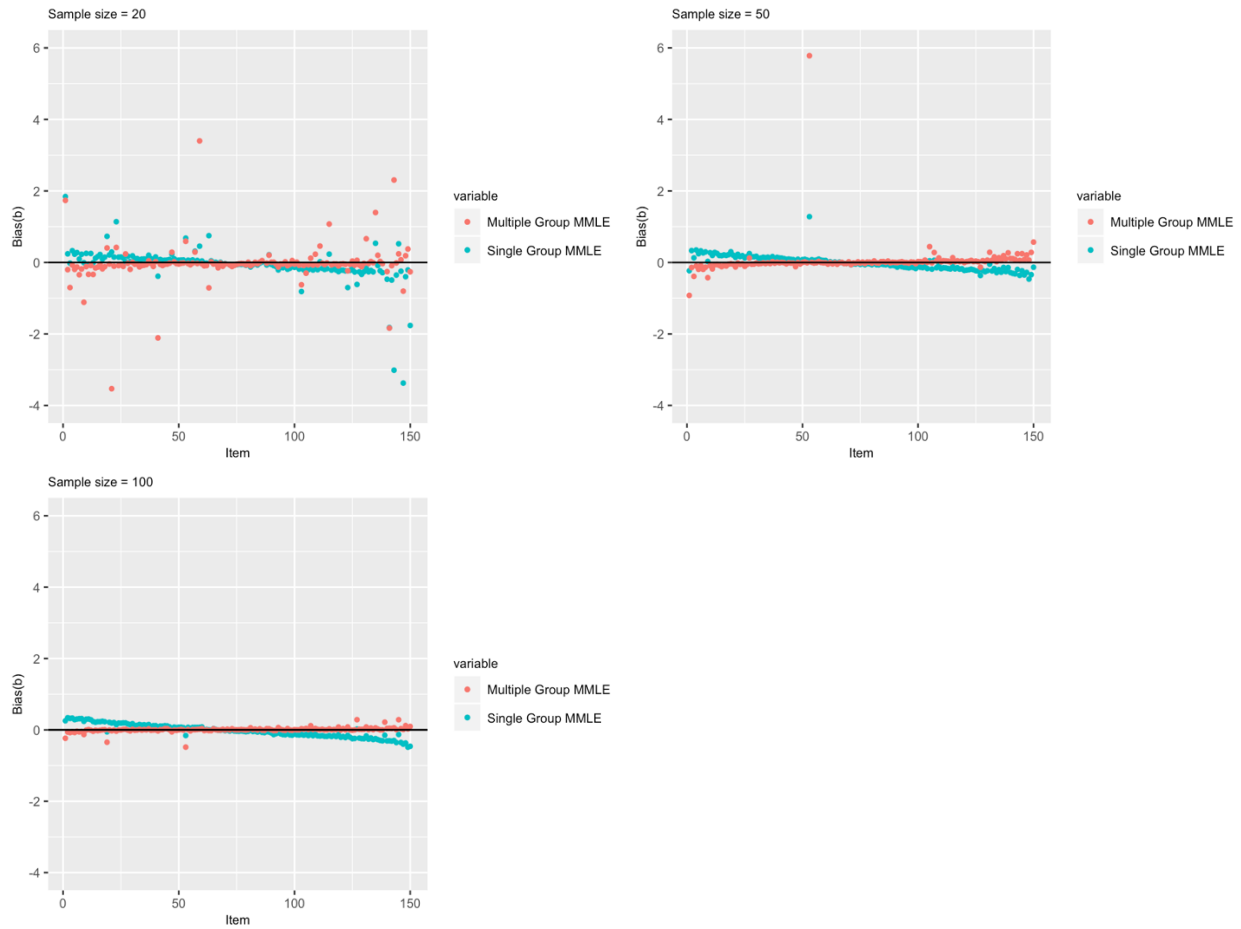


Figure 6.13. Bias for b Parameters under 5-Form Multiform Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes

Comparison across different number of anchor items. Figure 6.12 shows that for all number of anchor items conditions, bias obtained from multiple group MMLE was closer to zero than bias obtained from single group MMLE.

Figure 6.14 presents bias of estimated b parameters for each item when using uniformly selected anchor items and a sample size of 100. Bias obtained from single group MMLE was overestimated for easy items and were underestimated for hard items. Multiple group MMLE

was unbiased for all items. When different number of anchor items were used, the difference of bias was tiny for the two MMLE approaches.

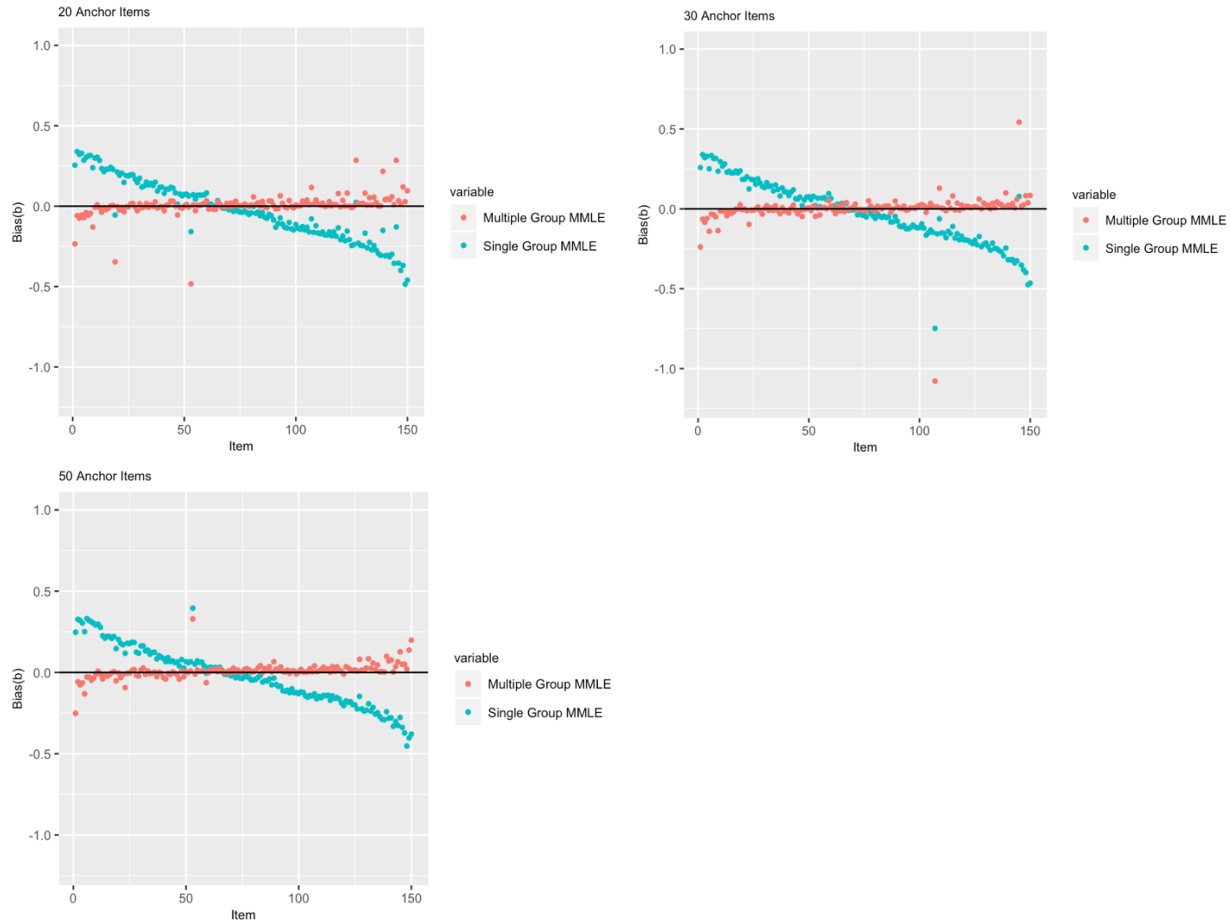


Figure 6.14. Bias for b Parameters under 5-Form Multiform Design with a Sample Size of 100 across Different Number of Uniformly Selected Anchor Items

Comparison across different anchor item selection strategies. Figure 6.12 shows that for both anchor item selection strategies, bias obtained from multiple group MMLE was closer to zero than bias obtained from single group MMLE.

Figure 6.15 presents estimated RMSE of estimated b parameters for each item when 20 anchor items and a sample size of 100 were used. Bias obtained from single group MMLE was overestimated for easy items and were underestimated for hard items. Multiple group MMLE

was unbiased for all items. There was no significant difference of bias when using different anchor item selection strategies for these two MMLE approaches.

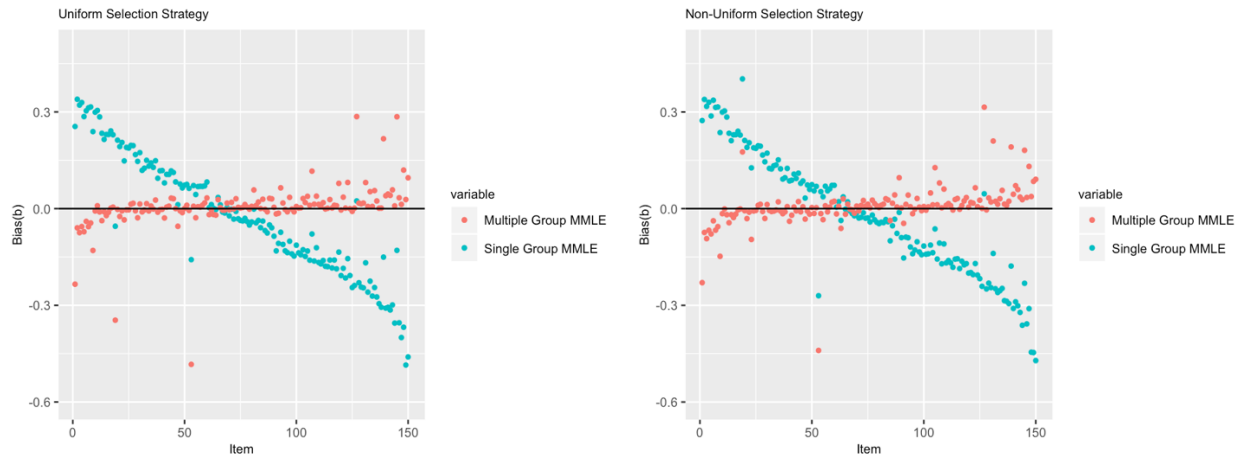


Figure 6.15. Bias for b Parameters under 5-Form Multifactor Design with 20 Anchor Items and a Sample Size of 100 across Different Anchor Item Selection Strategies

RMSE. Figure 6.16 presents boxplots of RMSE of estimated b parameters from single group MMLE and multiple group MMLE. The five number summaries can be found in Appendix D6.

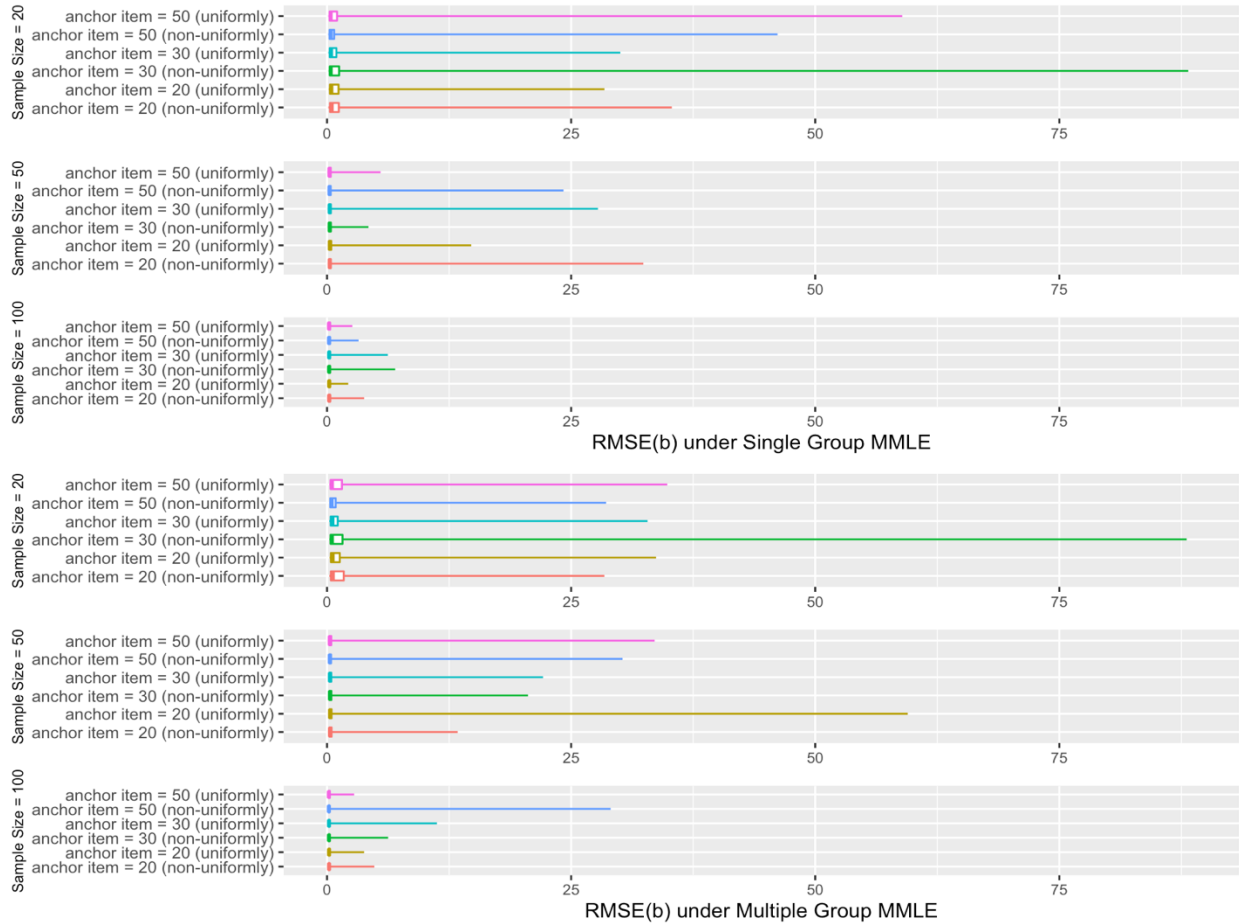


Figure 6.16. RMSE for b Parameters under 5-Form Multiform Design

Comparison across different sample sizes. From Figure 6.16, it was found that the difference of RMSE obtained from two MMLE approaches was not significant when using different sample size. For both MMLE approaches, as sample size increased, RMSE decreased.

Figure 6.17 presents RMSE obtained from two MMLE approaches. Figure 6.18 presents the RMSE that smaller than 2 to take a close look at the bottom of plots in Figure 6.17. When sample size was 20, smaller RMSE was observed from single group MMLE than from multiple group MMLE. When sample size was 50 or 100, RMSE obtained from single group MMLE was smaller than RMSE obtained from multiple group MMLE for middle difficult items, but it was larger than RMSE obtained from multiple group MMLE for easy or hard items. For both MMLE

approaches, RMSE of middle difficult items was closer to zero than RMSE of easy or hard items. Besides, as sample size increased, RMSE dots were less scattered and RMSE range decreased for both MMLE approaches.

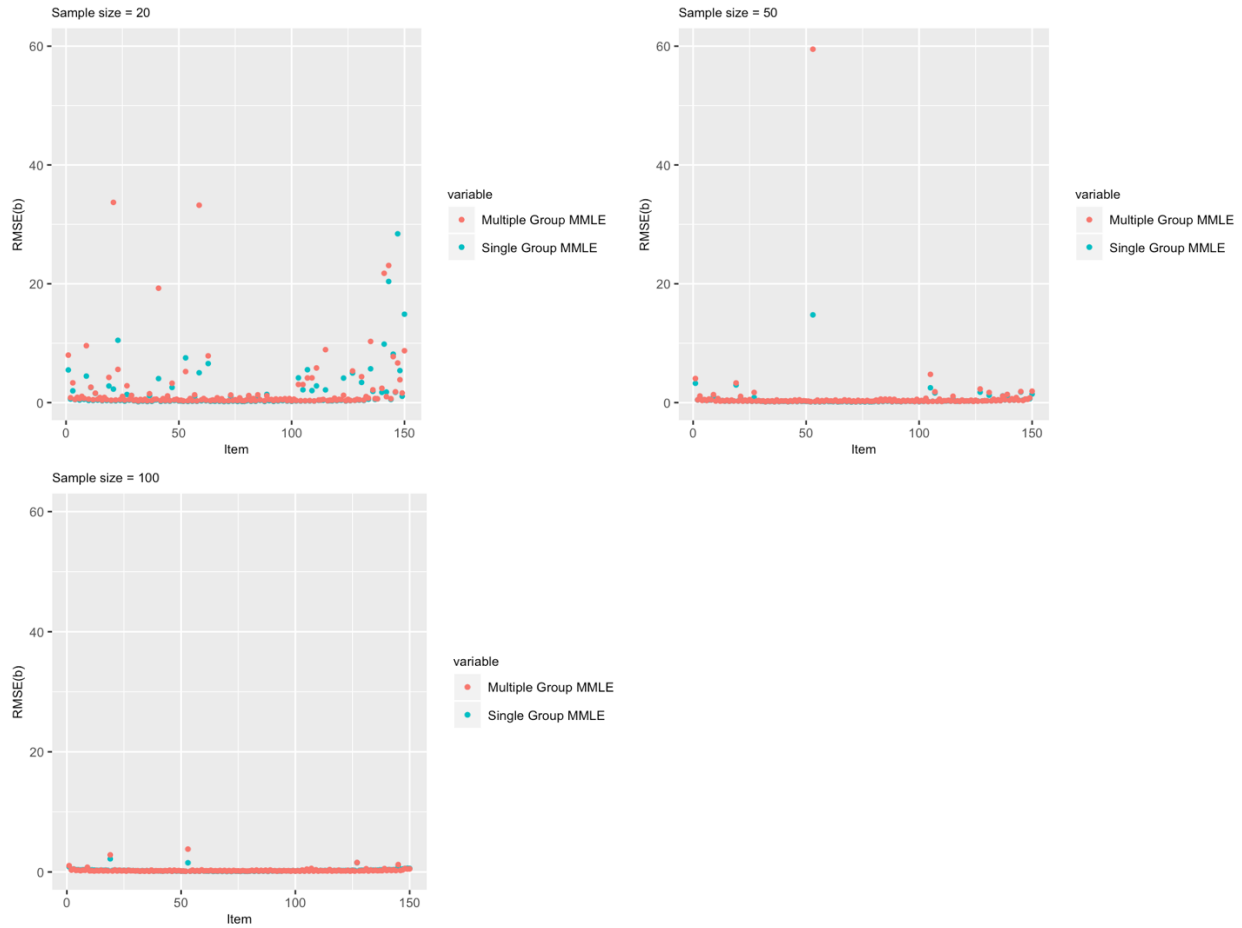


Figure 6.17. RMSE for b Parameters under 5-Form Multiform Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes

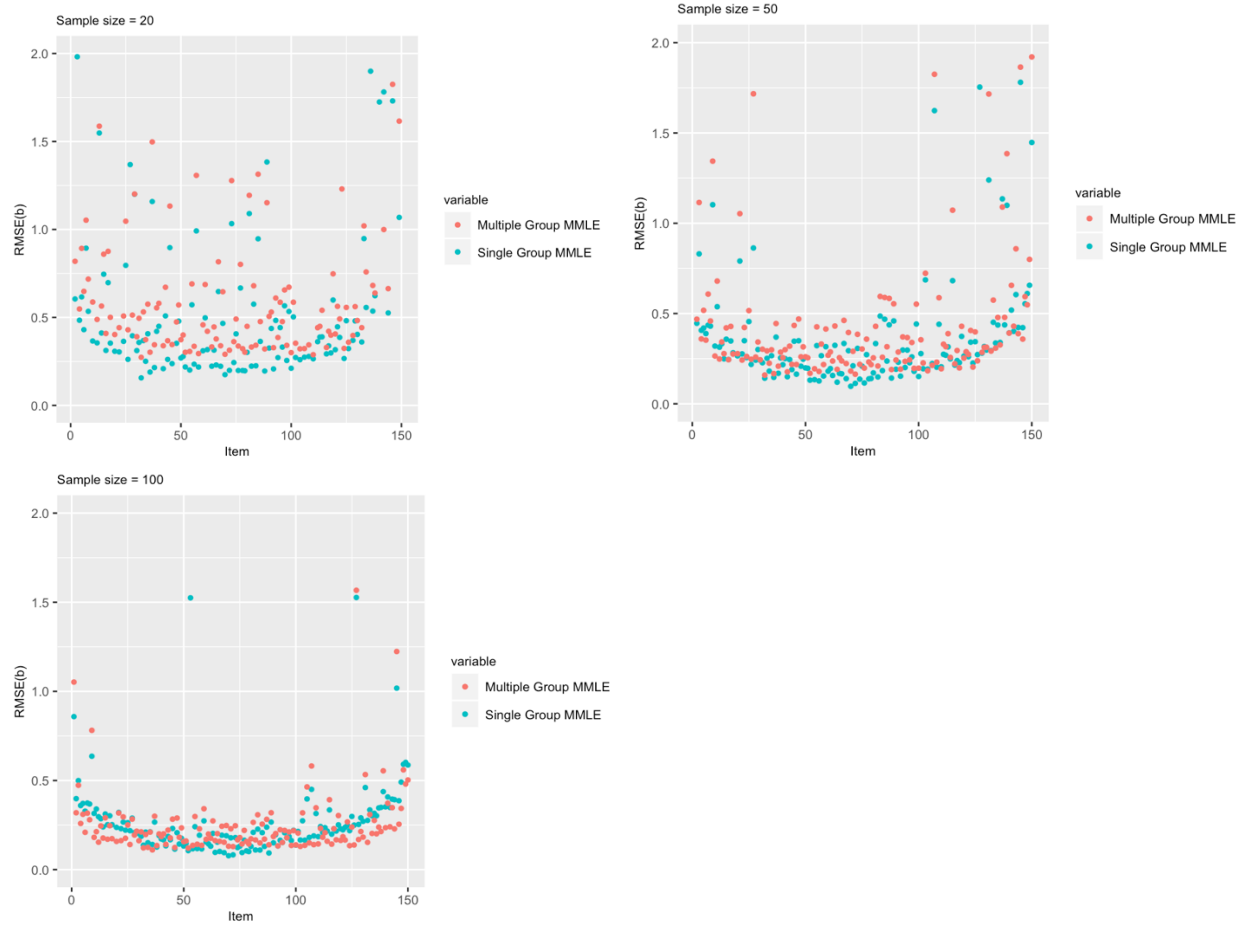


Figure 6.18. RMSE (from 0 to 2) for b Parameters under 5-Form Multiform Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes

Comparison across different number of anchor items. From Figure 6.16, it was found that the difference of RMSE obtained from two MMLE approaches was not significant when using different number of anchor items.

Figure 6.19 presents RMSE of estimated b parameters for each item when using uniformly selected anchor items and a sample size of 100. Figure 6.20 presents RMSE that was smaller than 2 to take a close look at the bottom of plots in Figure 6.19. RMSE observed for single group MMLE was smaller than RMSE observed for multiple group MMLE for middle difficult items. RMSE observed for single group MMLE was larger than RMSE observed for

multiple group MMLE for easy and hard items. When using different number of anchor items, difference of RMSE were tiny for both MMLE approaches.

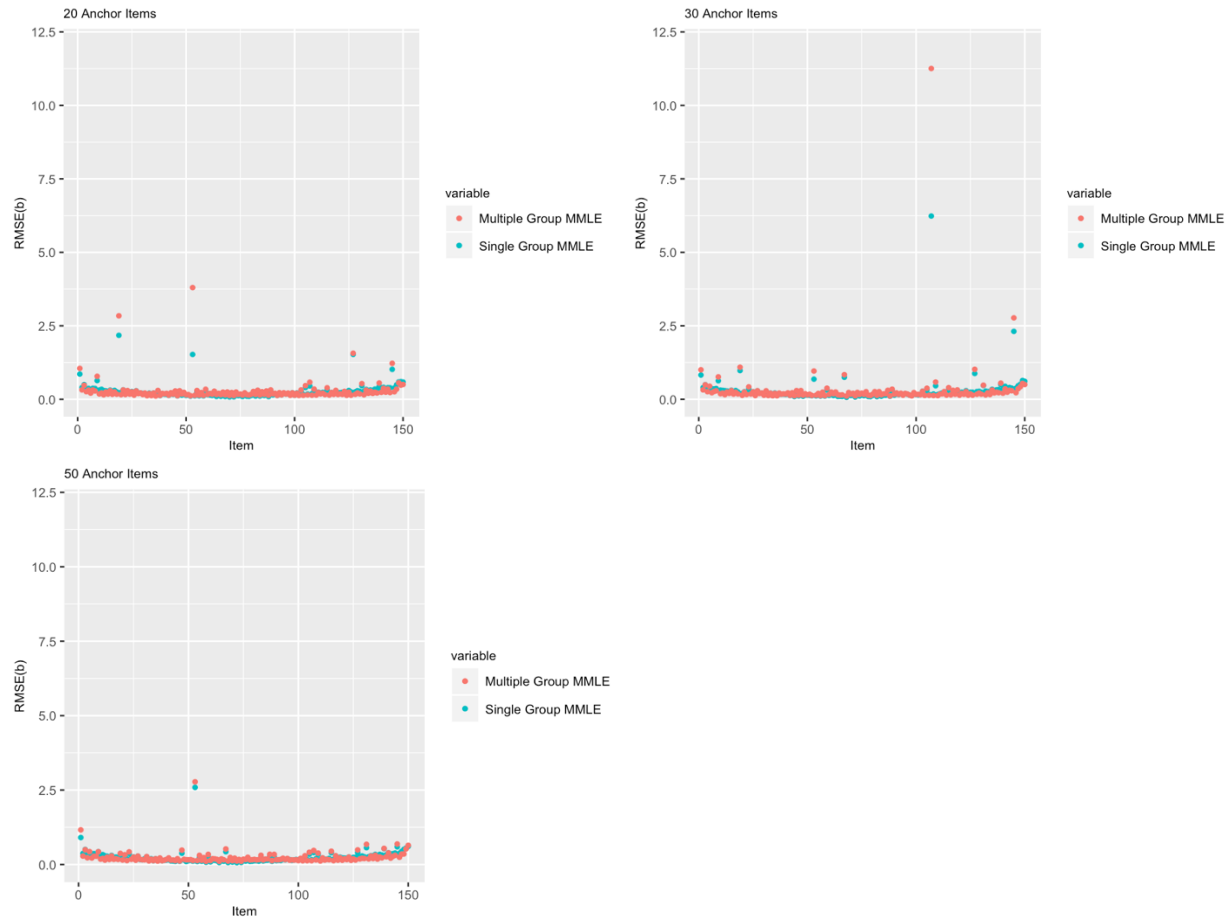


Figure 6.19. RMSE for b Parameters under 5-Form Multiform Design with a Sample Size of 100 across Different Number of Uniformly Selected Anchor Items

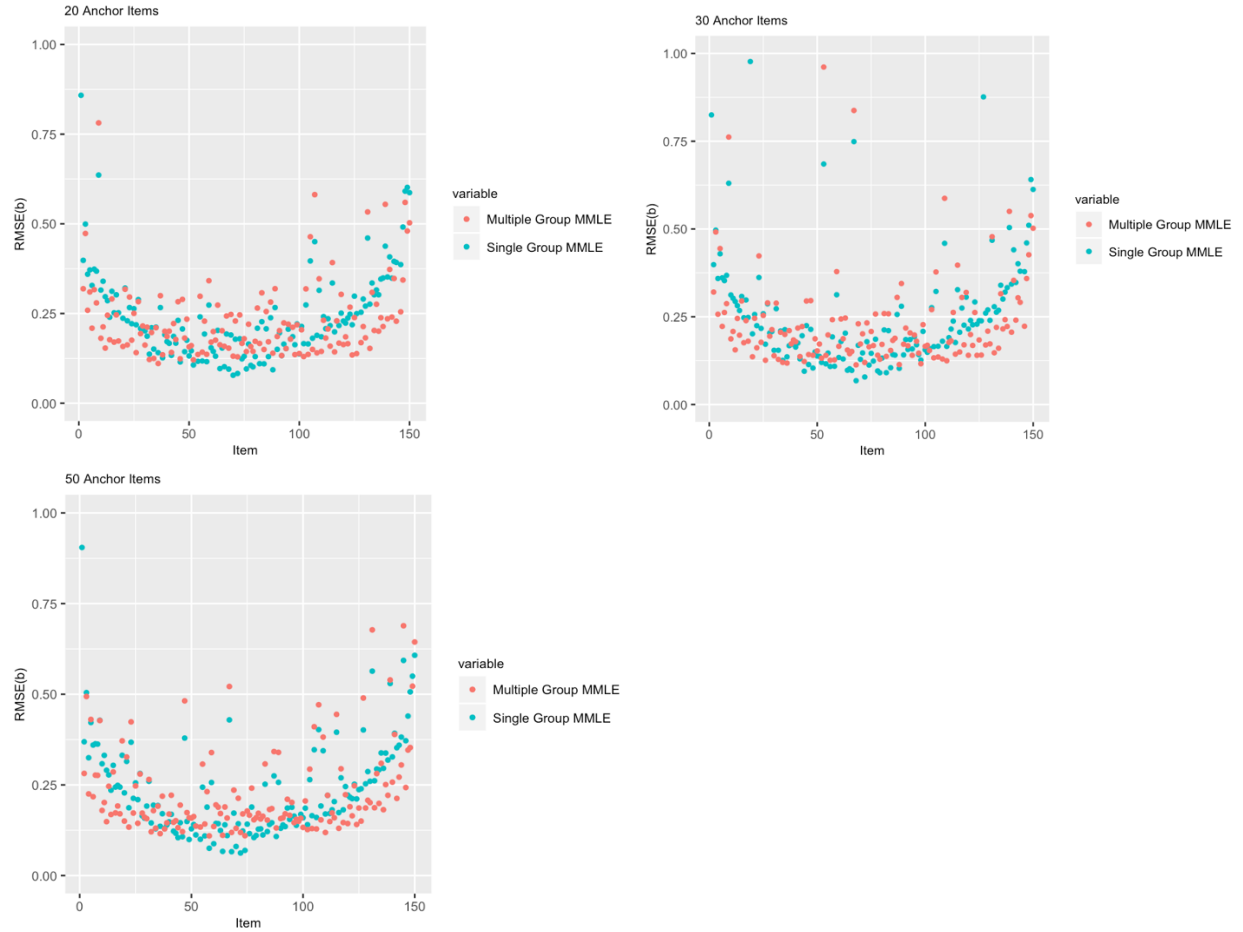


Figure 6.20. RMSE (from 0 to 1) for b Parameters under 5-Form Multiform Design with a Sample Size of 100 across Different Number of Uniformly Selected Anchor Items

Comparison across different anchor item selection strategies. Figure 6.16 presents that regardless of strategies of item selection, RMSE obtained from single group MMLE was similar to RMSE obtained from multiple group MMLE. Specifically, for multiple group MMLE, when sample size was 20, differences of RMSE existed when using different anchor selection strategies, but the differences didn't have a consistent pattern.

Figure 6.21 presents RMSE of estimated b parameters for each item when 20 anchor items and a sample size of 20. Figure 6.22 presents RMSE that was smaller than 2 to take a close look the bottom of plots in Figure 6.21. RMSE obtained from single group MMLE was closer to zero than RMSE obtained from multiple group MMLE. RMSE dots for two MMLE approaches

were less scattered when using uniformly selected anchor items than when using non-uniformly selected anchor items. The difference of RMSE was very tiny when different anchor item selection strategies were used for the two MMLE approaches.

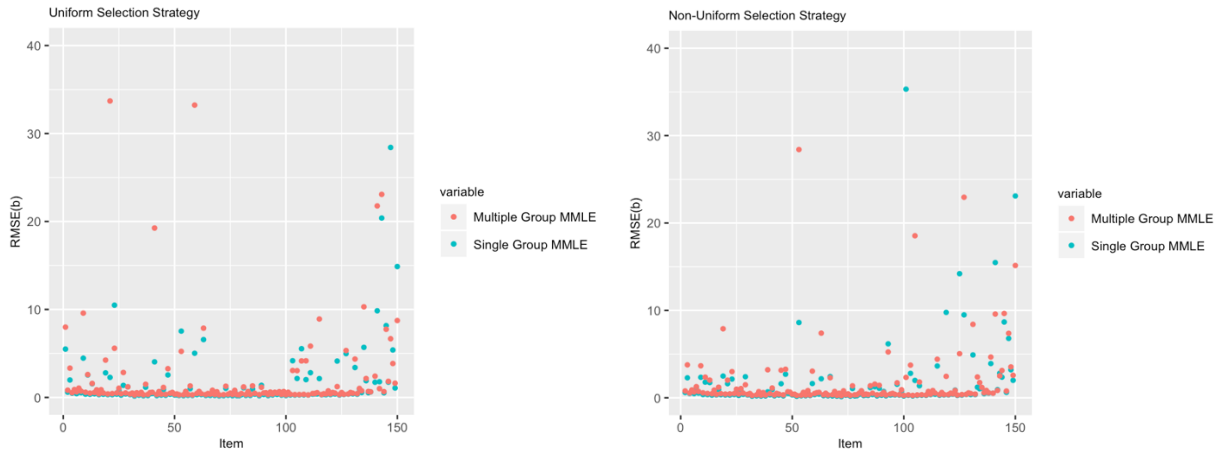


Figure 6.21. RMSE for b Parameters under 5-Form Multiform Design with 20 Anchor Items and a Sample Size of 20 across Different Anchor Item Selection Strategies



Figure 6.22. RMSE (from 0 to 2) for b Parameters under 5-Form Multiform Design with 20 Anchor Items and a Sample Size of 20 across Different Anchor Item Selection Strategies

PANEL MISSING DESIGN

Because only 5-form test design was used for the panel missing design. Number of extreme estimation values, bias and RMSE obtained from single group MMLE and multiple group MMLE for a parameter estimation and b parameter estimation were compared across different sample sizes.

5-form panel missing design

Estimation of a parameters. For a parameter estimation, number of extreme estimation values, bias, and RMSE were presented.

Number of extreme estimation values. When sample size was 20, single group MMLE identified more extreme estimation values than multiple group MMLE. When sample size was 50 or 100, these two MMLE approaches identified no extreme estimation value. Detailed number can be found in Appendix E1. In general, multiple group MMLE was more stable than single group MMLE when sample size was 20 and two MMLE approaches have the same stability when sample size was 50 or 100.

Bias. Figure 7.1 presents bias for a parameters obtained from multiple group MMLE and single group MMLE for each item across different sample sizes. Bias obtained from multiple group MMLE was closer to zero than bias obtained from single group MMLE regardless of sample size. As sample size increased, bias was closer to zero and the range of bias decreased for these two MMLE approaches. Bias dots observed for two MMLE approaches lied apart further from each other as sample size increased. Five number summaries of bias of estimated a parameters are summarized in Appendix E2.

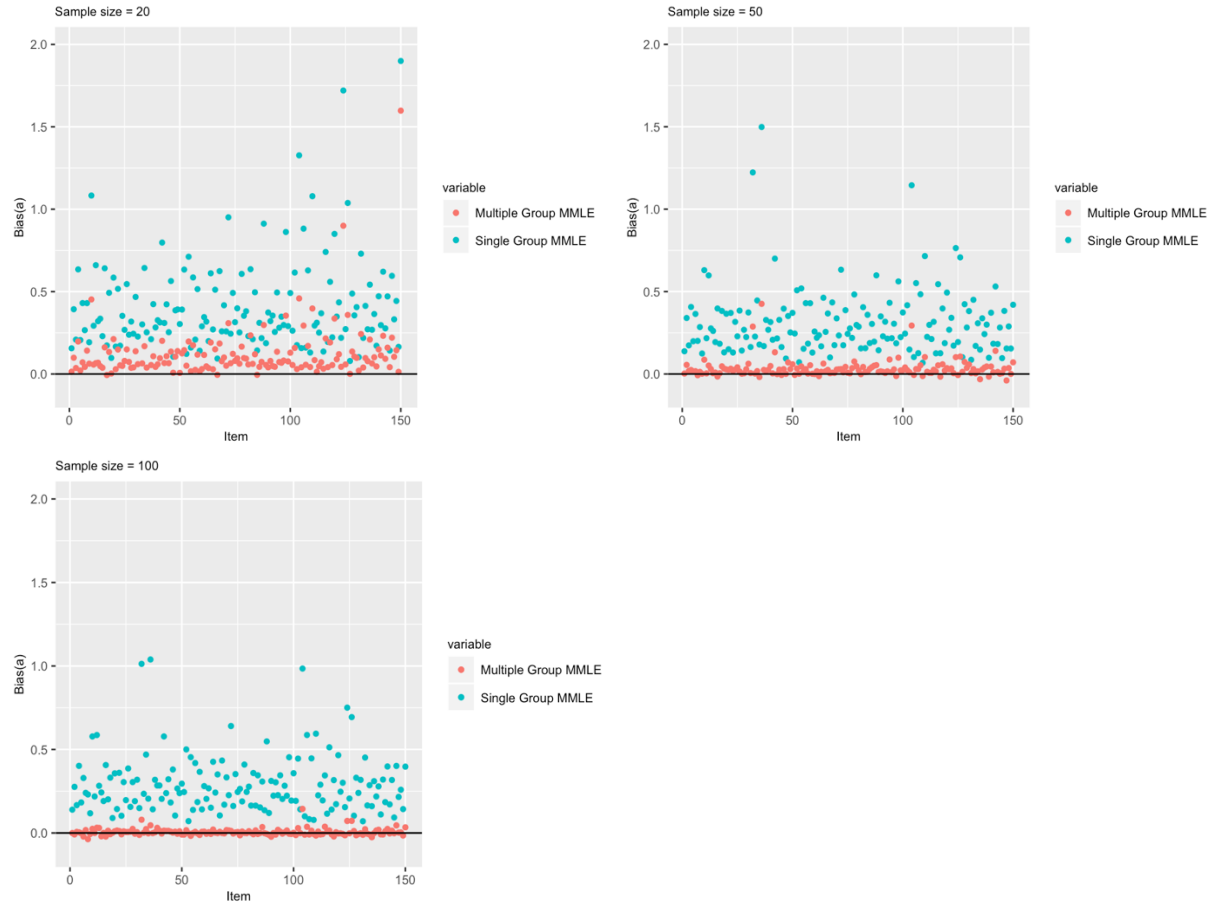


Figure 7.1. Bias for a Parameters under 5-Form Panel Missing Design across Different Sample Sizes

RMSE. Figure 7.2 shows RMSE for a parameters obtained from multiple group MMLE and single group MMLE for each item across different sample sizes. It was found that larger RMSE were observed when using single group MMLE than when using multiple group MMLE irrespective of sample size. In addition, for both MMLE approaches, as sample size increased, RMSE was closer to zero and the range of RMSE decreased. RMSE dots observed for two MMLE approaches were further apart as sample size increased. Five number summaries of RMSE of estimated a parameters are summarized in Appendix E3.

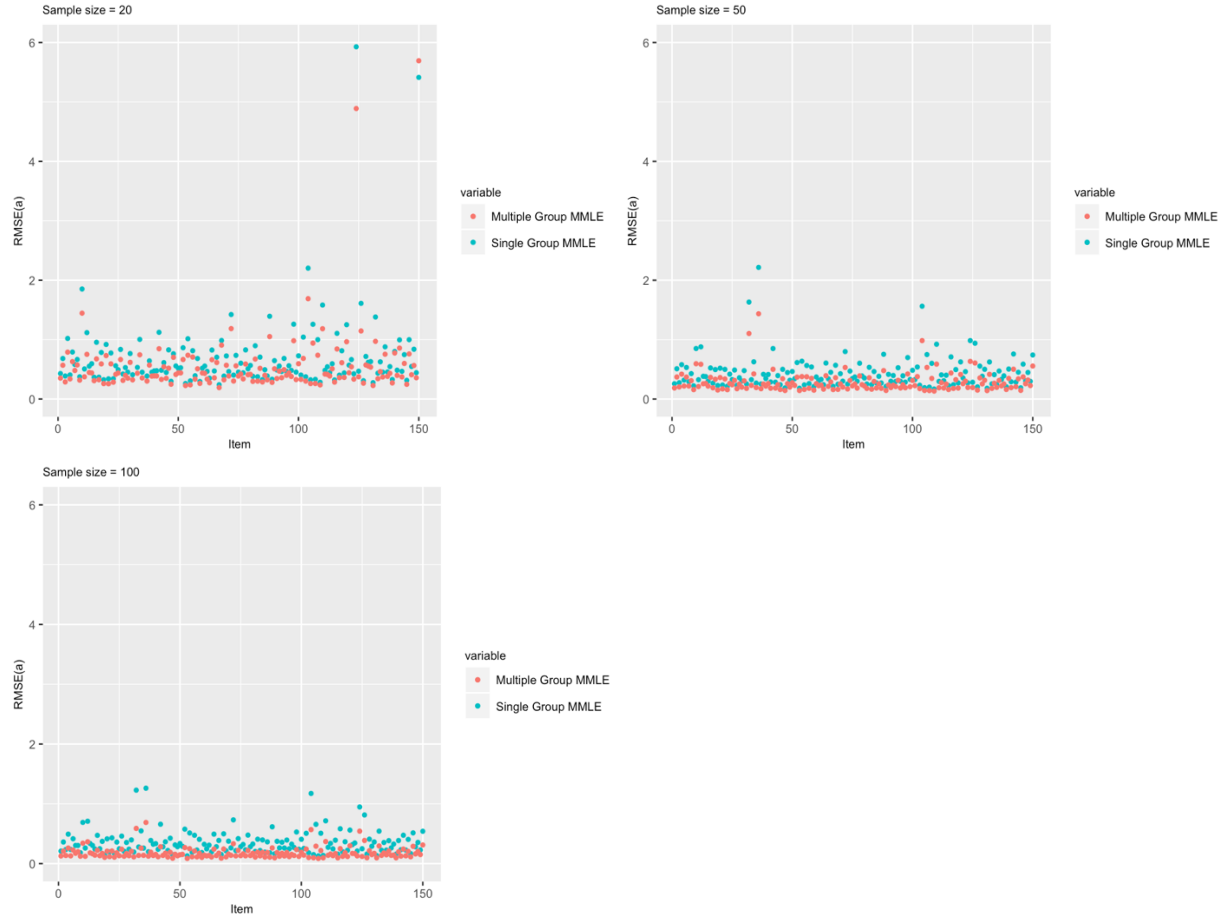


Figure 7.2. RMSE for a Parameters under 5-Form Panel Missing Design across Different Sample Sizes

Estimation of b parameters. For b parameter estimation, number of extreme estimation values, bias, and RMSE were presented.

Number of extreme estimation values. When sample size was 20, multiple group MMLE identified one extreme estimation value while single group MMLE screened out none. When sample size was 50 or 100, no extreme estimation value was identified when using both MMLE approaches. Table of the number can be found in Appendix E4. In general, multiple group MMLE was more stable than single group MMLE when sample size was 20 and two MMLE approaches have the same stability when sample size was 50 or 100.

Bias. Figure 7.3 presents bias for b parameters obtained from multiple group MMLE and single group MMLE for each item across different sample sizes. Since items 1-150 were items from easy to hard, it was found that easy items were overestimated while hard items were underestimated for single group MMLE. However, it was almost unbiased for multiple group MMLE. Besides, it was very close to unbiased estimation for multiple group MMLE but not for single group MMLE. As sample size increased, bias dots observed for two MMLE approaches were less scattered. Five number summaries of bias of estimated b parameters are summarized in Appendix E5.

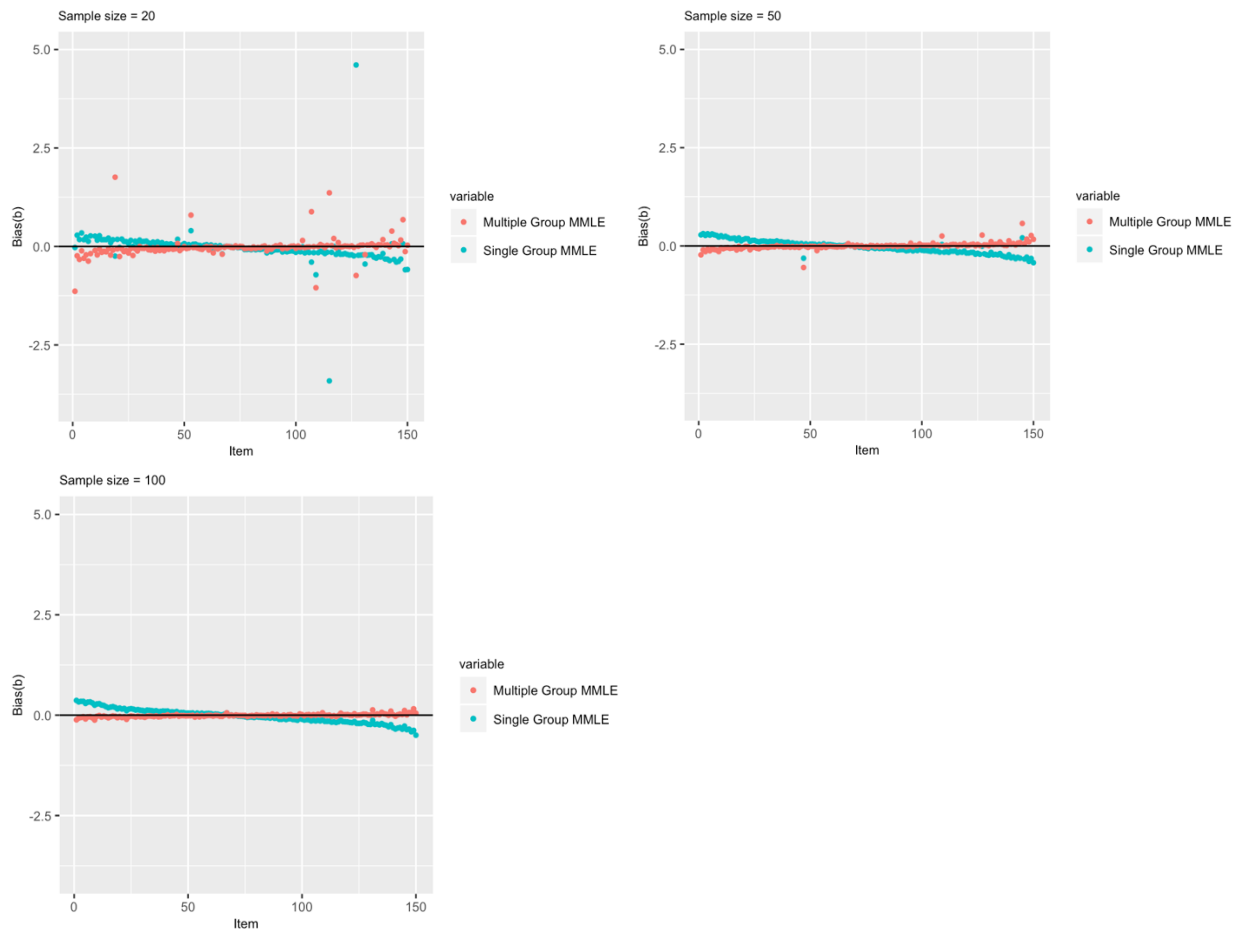


Figure 7.3. Bias for b Parameters under 5-Form Panel Missing Design across Different Sample Sizes

RMSE. Figure 7.4 presents RMSE obtained from single group MMLE and multiple group MMLE for each item across different sample sizes. Figure 7.5 presents the RMSE that was smaller than 2 to take a close look at the bottom of the plots in Figure 7.4. Items 1-150 were items from easy to hard. When sample size was 20, larger RMSE was observed for multiple group MMLE than for single group MMLE. When sample size was 50, RMSE obtained from multiple group MMLE was larger than that from single group MMLE for middle difficult items; but for easy and hard items, RMSE obtained from both MMLE approaches were similar. When sample size was 100, for middle difficult items, larger RMSE was observed for single group MMLE; while for easy and hard items, larger RMSE was observed for multiple group MMLE. In addition, as sample size increased, range of RMSE for both MMLE approaches decreased.

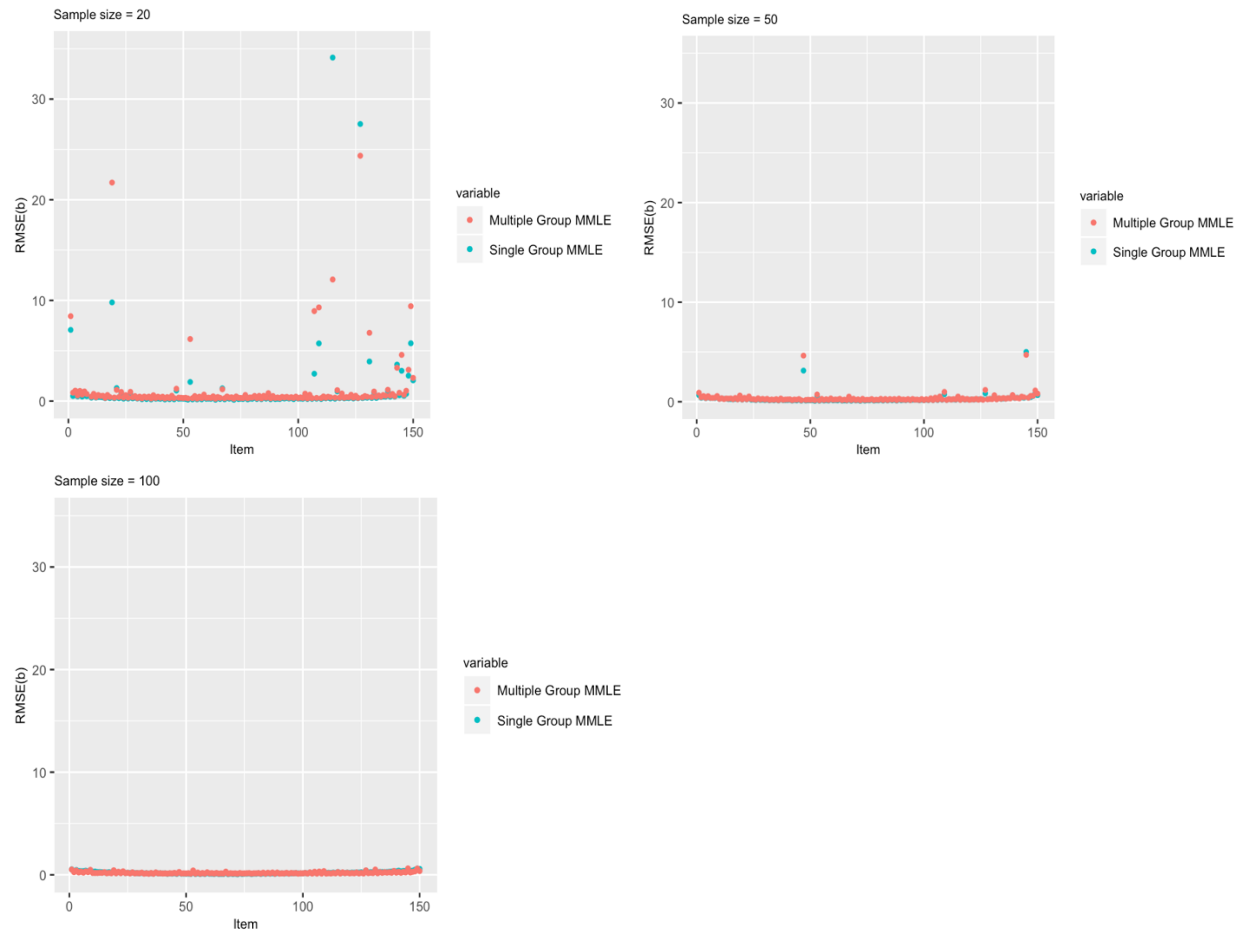


Figure 7.4. RMSE for b Parameters under 5-Form Panel Missing Design across Different Sample Sizes

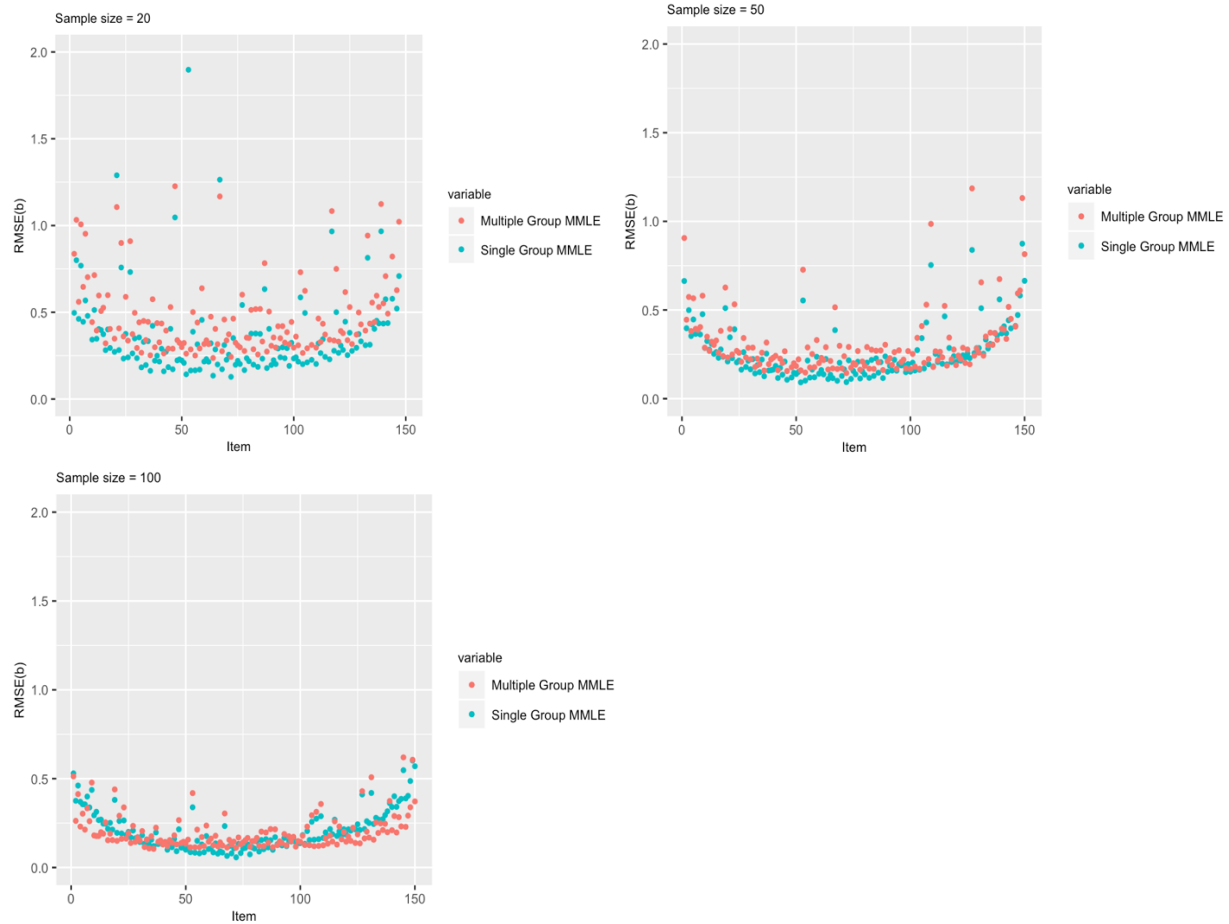


Figure 7.5. RMSE (from 0 to 2) for b Parameters under 5-Form Panel Missing Design across Different Sample Sizes

PANEL CHAINED DESIGN

Under panel chained design only 5-form design existed and only one factor, sample size, existed. Number of extreme estimation values, bias and RMSE obtained from single group MMLE and multiple group MMLE for a parameter estimation and b parameter estimation were compared across different sample sizes.

5-form panel chained design

Estimation of a parameters. For a parameter estimation, number of extreme estimation values, bias, and RMSE were presented.

Number of extreme estimation values. The number of extreme estimations values when estimating a parameters by single group MMLE and multiple group MMLE are presented in Figure 8.1 and Appendix F1. Figure 8.1 shows that through single group MMLE, more extreme estimation values were identified than multiple group MMLE when sample size was 20. When sample size was 50 or 100, no extreme estimation values were identified by using both MMLE approaches.

Bias. Figure 8.1 presents bias for a parameters obtained from multiple group MMLE and single group MMLE for each item across different sample sizes. It can be seen that a parameters were overestimated by both MMLE approaches. When sample size was 20, bias obtained from single group MMLE and multiple group MMLE were similar. But when sample size was 50 or 100, bias obtained from single group MMLE was closer to zero than bias obtained from multiple group MMLE. In addition, for both MMLE approaches, bias was closer to zero and the range of bias decreased as sample size increased. The five number summaries of bias of the estimated a parameters are summarized in Appendix F2.

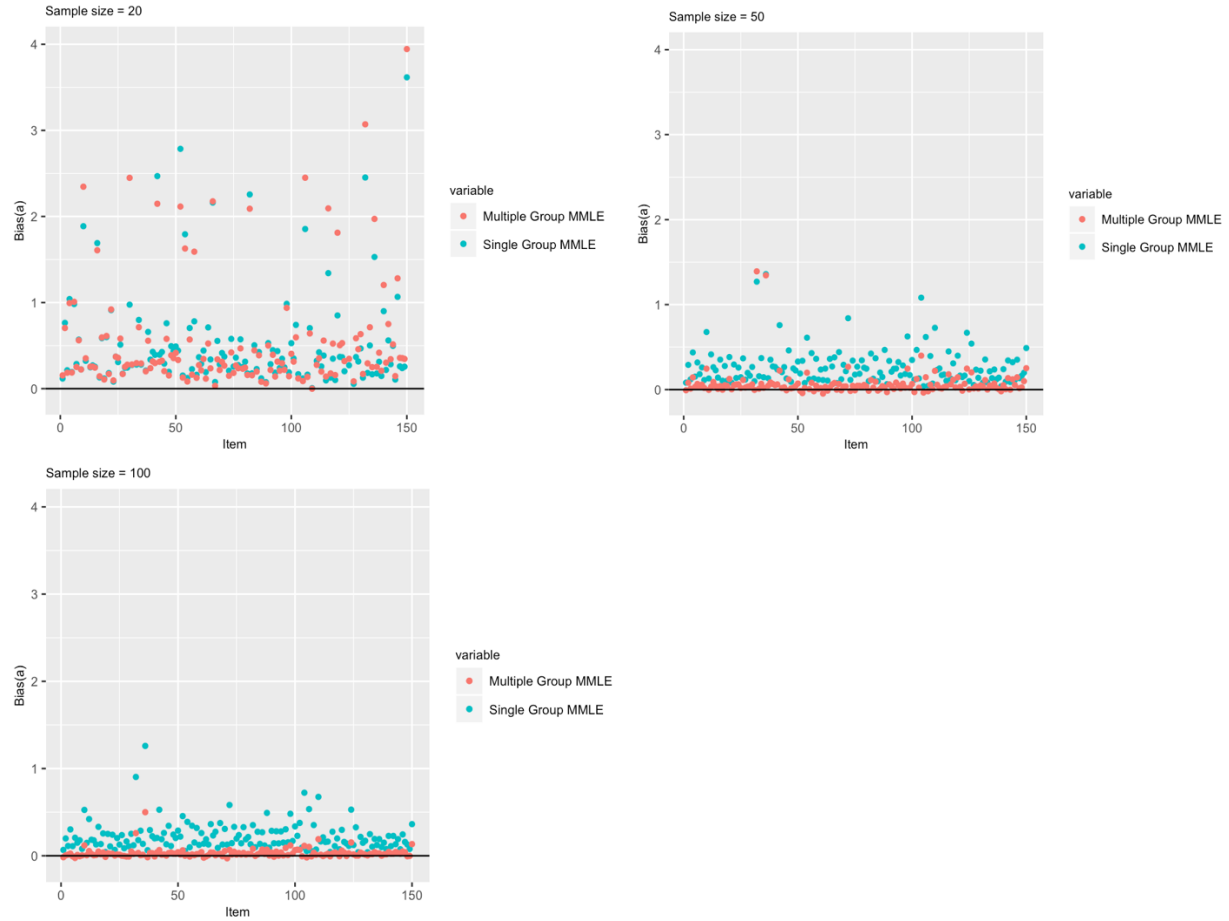


Figure 8.1. Bias for a Parameters under 5-Form Panel Chained Design across Different Sample Sizes

RMSE. Figure 8.2 shows RMSE for a parameters obtained from multiple group MMLE and single group MMLE for each item across different sample sizes. Figure 8.3 shows RMSE that was smaller than 2 to take a close look at the bottom of the plots in Figure 8.2. It was found that when sample size was 20, RMSE dots obtained from two MMLE approaches were mixed together and there were no consistent magnitude relationships existed. But when sample size was 50 or 100, RMSE obtained from single group MMLE was larger than RMSE obtained from multiple group MMLE. Besides, as sample size increased, RMSE was closer to zero and the range of RMSE decreased for both MMLE approaches. The five number summaries of RMSE for a parameters can be found in Appendix F3.

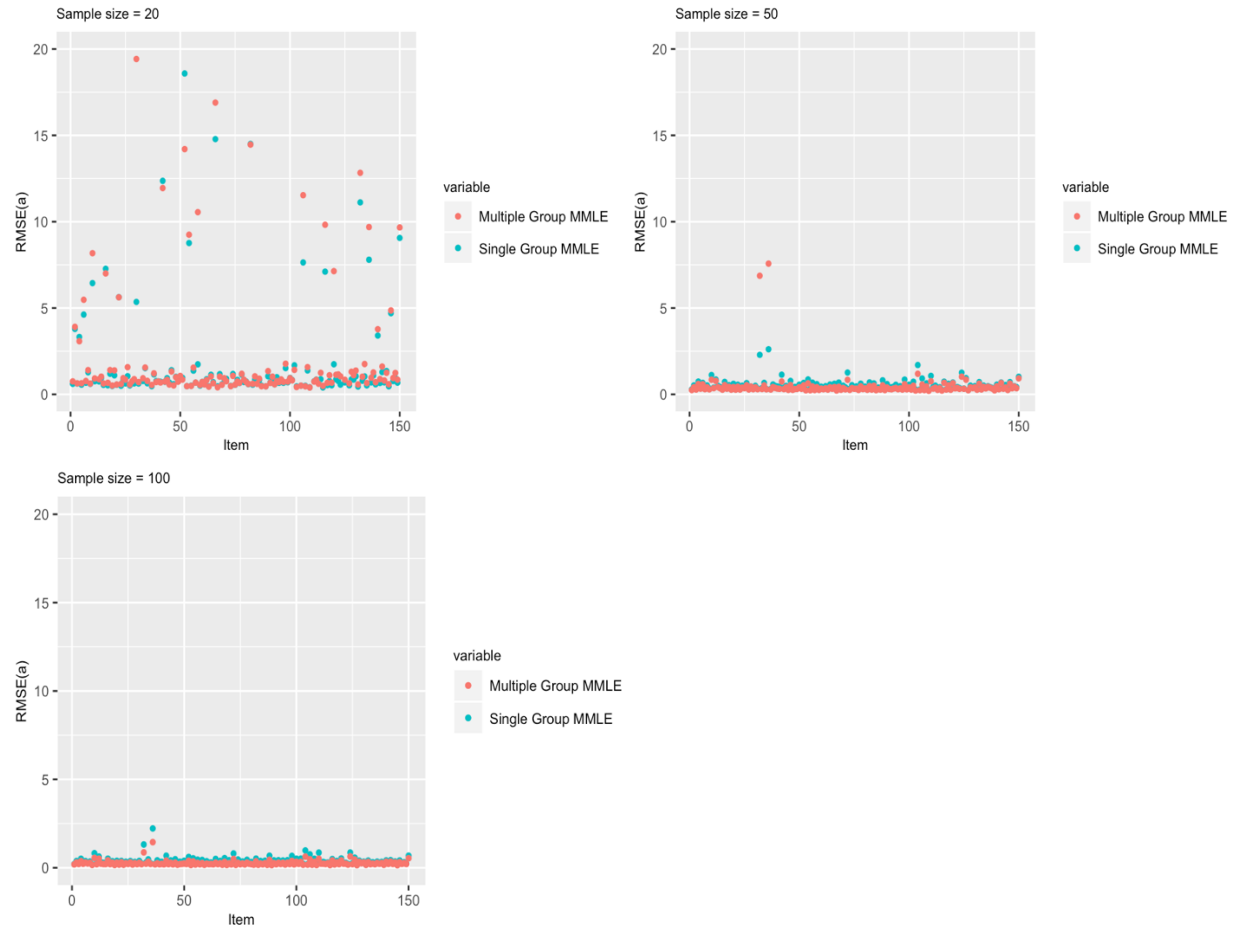


Figure 8.2. RMSE for a Parameters under 5-Form Panel Chained Design across Different Sample Sizes

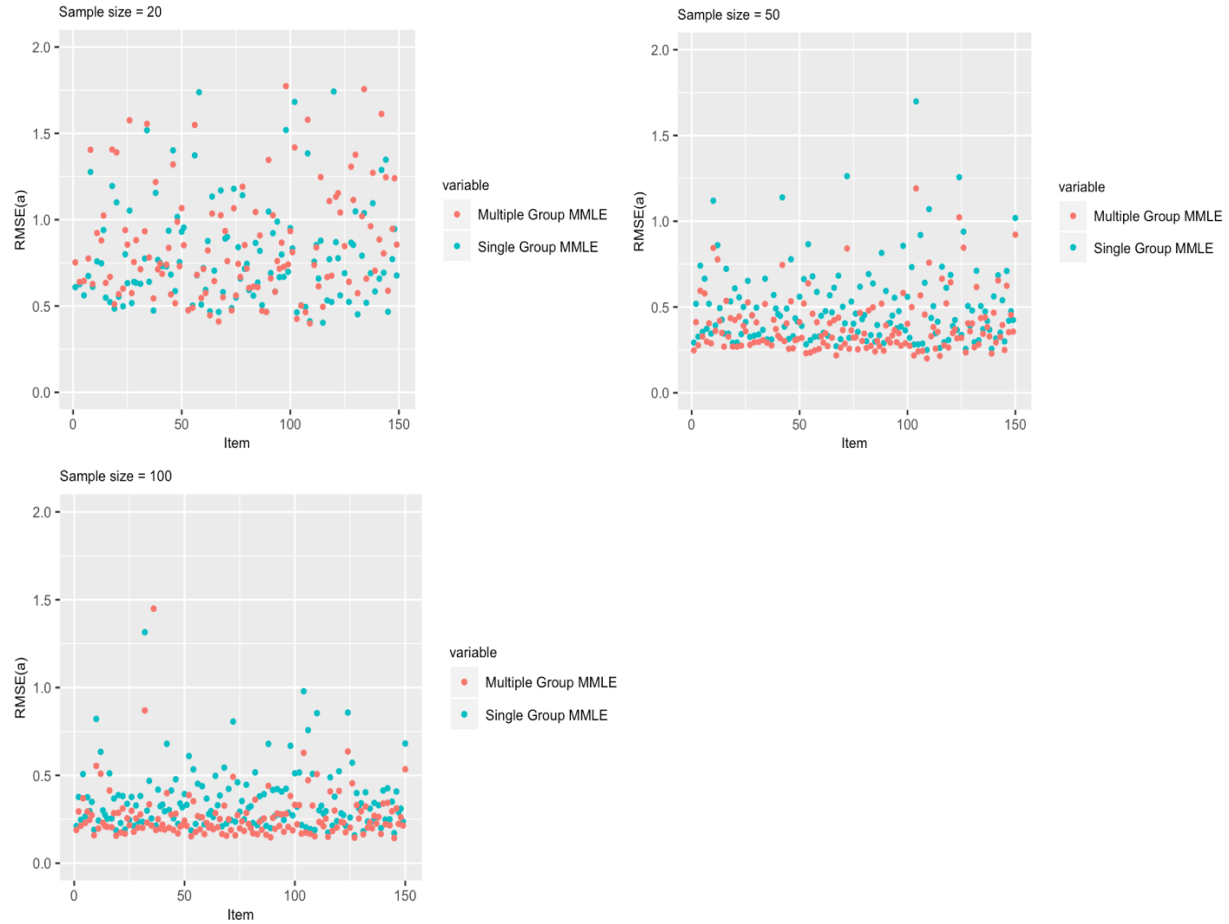


Figure 8.3. RMSE (from 0 to 2) for a Parameters under 5-Form Panel Chained Design across Different Sample Sizes

Estimation of b parameters. For b parameter estimation, number of extreme estimation values, bias, and RMSE were presented.

Number of extreme estimation values. When sample size was 20, single group MMLE identified one extreme estimation value while multiple group MMLE identified none. When sample size was 50, single group MMLE identified one extreme estimation value while multiple group MMLE identified none. When sample size was 100, multiple group MMLE identified one extreme estimation value while single group MMLE identified none. When sample size was 150, both MMLE approaches identified none. The table of the number can be found in Appendix F4. In general, multiple group MMLE was more stable than single group MMLE when sample size

was 20; single group MMLE was more stable when sample size was 50; two MMLE approaches have the same stabilities when sample size was 100.

Bias. Figure 8.4 presents bias for b parameters obtained from multiple group MMLE and single group MMLE for each item across different sample sizes. Difficulties of items 1-150 were increased. For single group MMLE, easy items were overestimated while hard items were underestimated; for multiple group MMLE, all items were almost unbiased. As sample size increased, bias dots obtained from both MMLE approaches were less scattered, and the range of bias decreased. The five number summaries of bias of the estimated b parameters are summarized in Appendix F5.

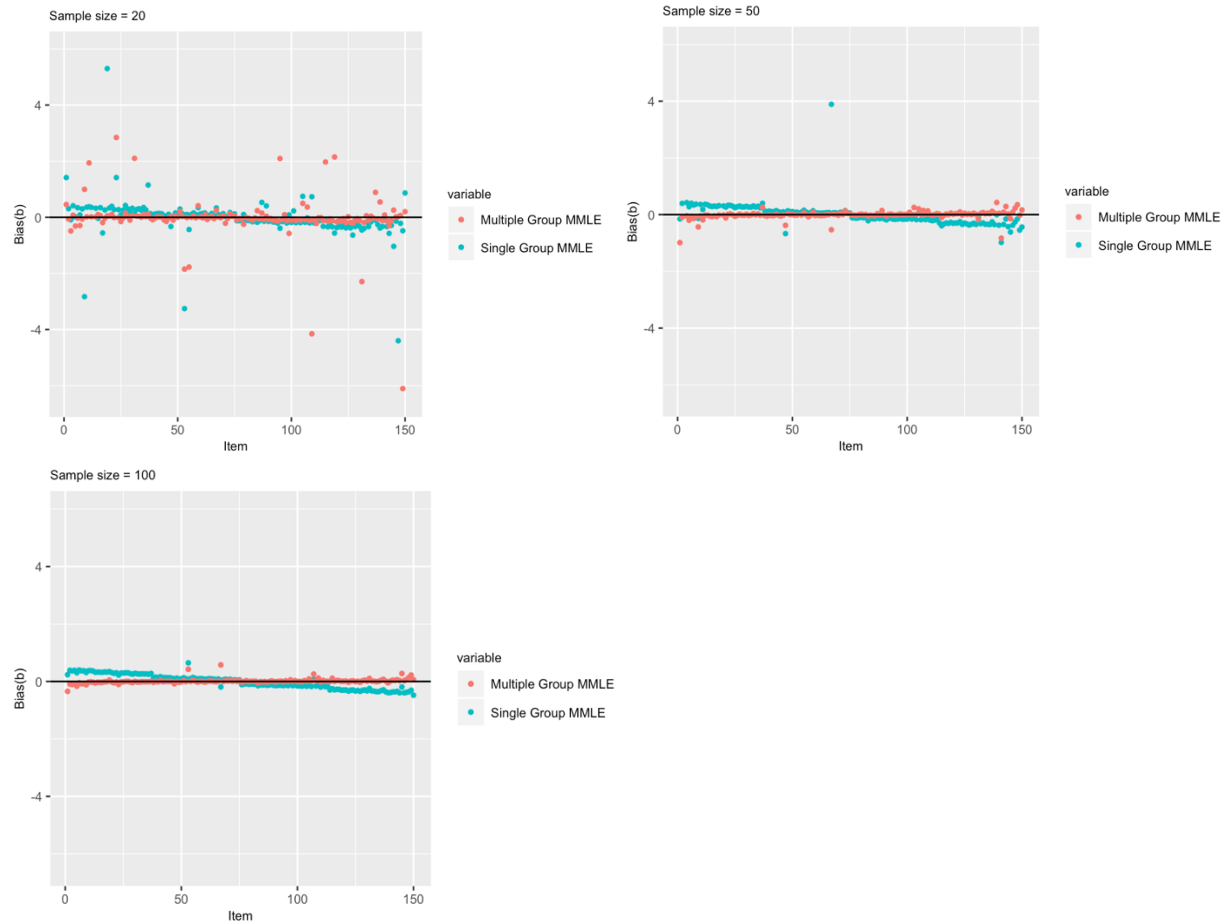


Figure 8.4. Bias for b Parameters under 5-Form Panel Chained Design across Different Sample Sizes

RMSE. Figure 8.5 presents RMSE for b parameters obtained from multiple group MMLE and single group MMLE for each item across different sample sizes. Figure 8.6 presents the RMSE that was smaller than 2 to take a close look at the bottom of the plots in Figure 8.5. It can be found that when sample size was 20, RMSE observed from multiple group MMLE was larger than that from single group MMLE. Items 1-150 were items from easy to hard. When sample size was 50 or 100, for easy and hard items, larger RMSE were observed for single group MMLE than for multiple group MMLE; and for middle difficult items, larger RMSE were observed for multiple group MMLE than for single group MMLE. For both MMLE approaches, as sample size increased, RMSE was closer to zero and the range of RMSE was decreased.

Besides, RMSE dots were less scattered as sample size increased. The five number summaries of RMSE of the estimated b parameters are summarized in Appendix F6.

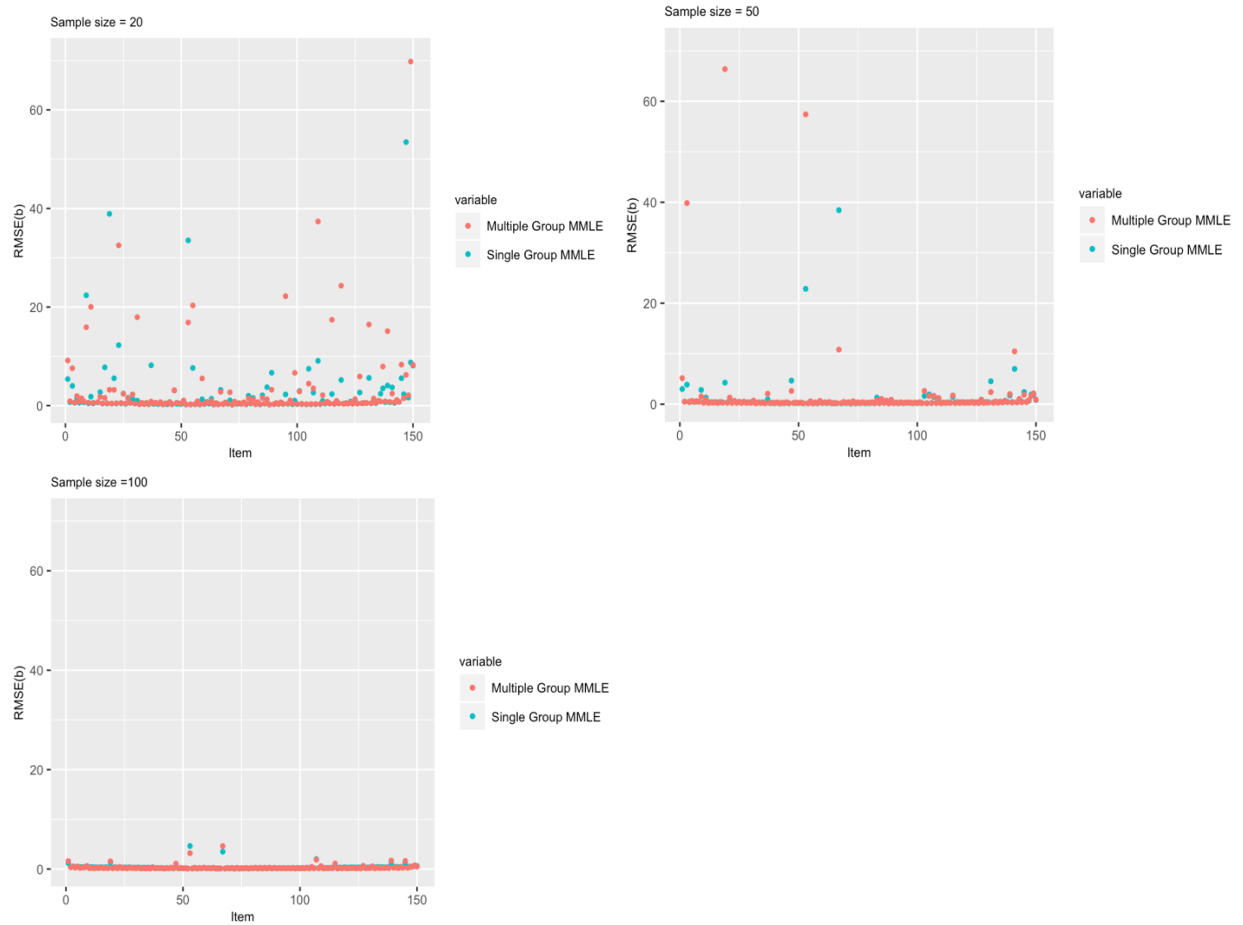


Figure 8.5. RMSE for b Parameters under 5-Form Panel Chained Design across Different Sample Sizes

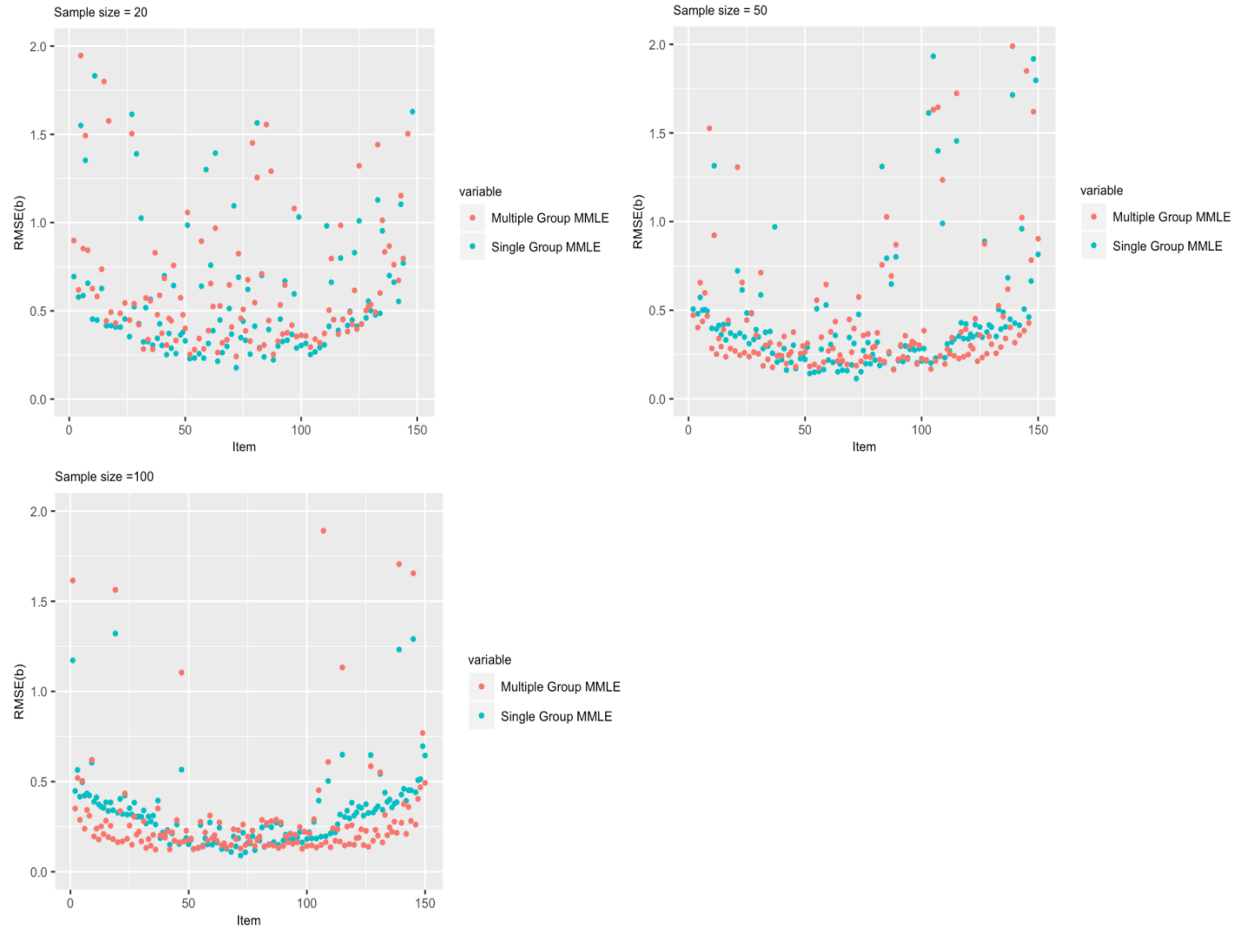


Figure 8.6. RMSE (from 0 to 2) for b Parameters under 5-Form Panel Chained Design across Different Sample Sizes

BIB DESIGN

Under BIB design only 5-form design exist and only one factor, sample size, exist.

Number of extreme estimation values, bias and RMSE obtained from single group MMLE and multiple group MMLE for a parameter estimation and b parameter estimation were compared across different sample sizes.

5-form BIB design

Estimation of a parameters. For a parameter estimation, number of extreme estimation values, bias, and RMSE were presented.

Number of extreme estimation values. When sample size was 20, single group MMLE and multiple group MMLE identified same number of extreme estimation values. When sample size was 50, single group MMLE identified one extreme estimation value while multiple group MMLE identified none. When sample size was 100, no extreme estimation value was identified by both MMLE approaches. Detailed numbers were shown in Appendix G1. In general, two MMLE approaches had the same stability when sample size was 20 or 100. When sample size was 50, multiple group MMLE was more stable than single group MMLE.

Bias. Figure 9.1 presents bias for a parameters obtained from multiple group MMLE and single group MMLE for each item across different sample sizes. Bias observed for multiple group MMLE was closer to bias observed for single group MMLE. Considering items 1-150 were items from easy to hard, for these two MMLE approaches, bias for easy and hard items were closer to zero than bias for middle difficult items. In addition, as sample size increased, bias dots were less scattered, and the range of bias decreased for both MMLE approaches. The five number summaries of bias of the estimated a parameters are summarized in Appendix G2.

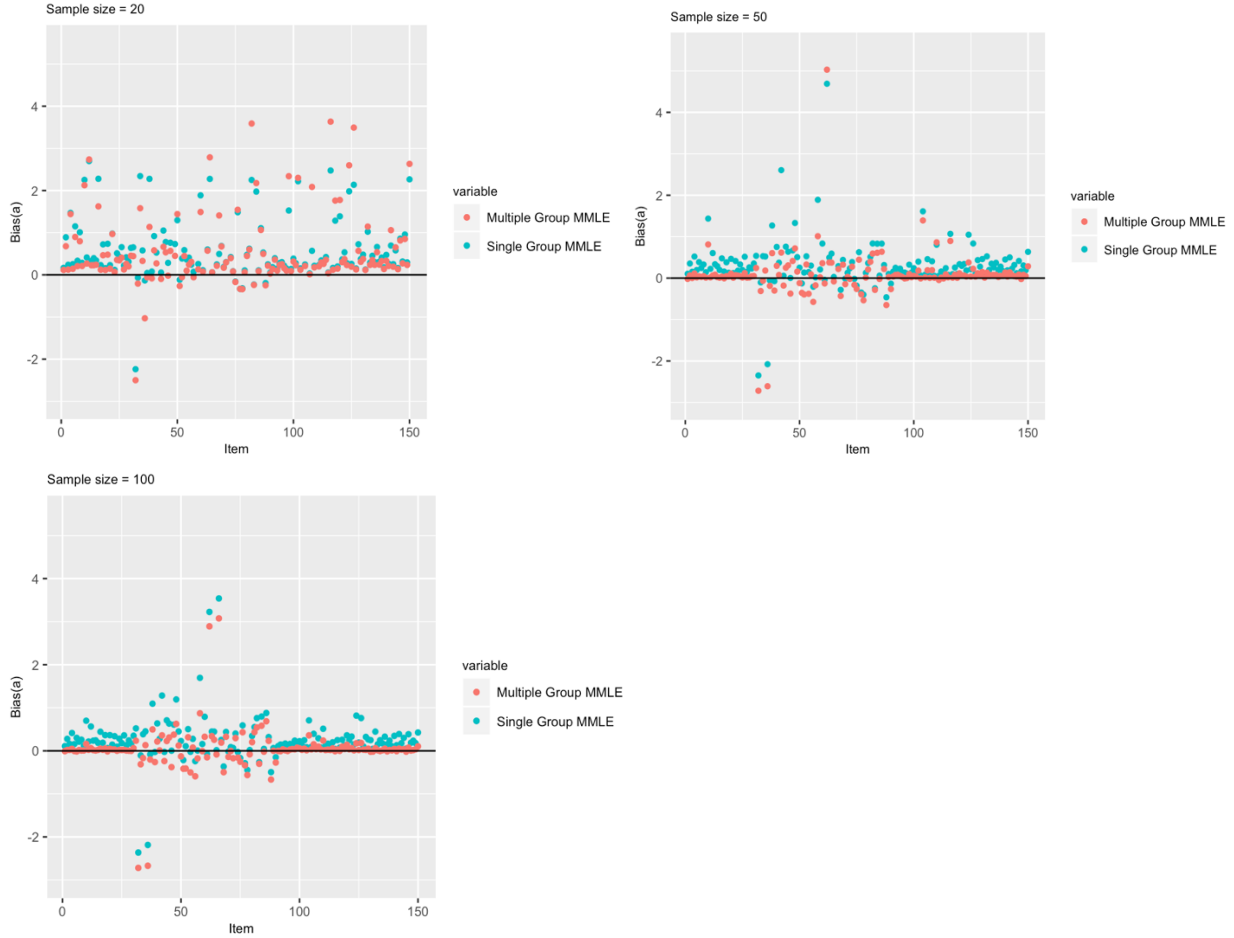


Figure 9.1. Bias for a Parameters under 5-Form BIB Design across Different Sample Sizes

RMSE. Figure 9.2 presents RMSE for a parameters obtained from multiple group MMLE and single group MMLE for each item across different sample sizes. Figure 9.3 shows the RMSE smaller than 2 to take a close look at the bottom of plots in Figure 9.2. It can be seen that when sample size was 20, RMSE dots obtained from two MMLE approaches were mixed together and there was no a consistent pattern. When sample size was 50 or 100, RMSE observed from single group MMLE were larger than RMSE observed from multiple group MMLE. In addition, as sample size increased, RMSE was closer to zero and the range of RMSE decreased for both MMLE approaches. RMSE dots were less scattered as sample size increased. The five number summaries of RMSE of the estimated a parameters are summarized in Appendix G3.

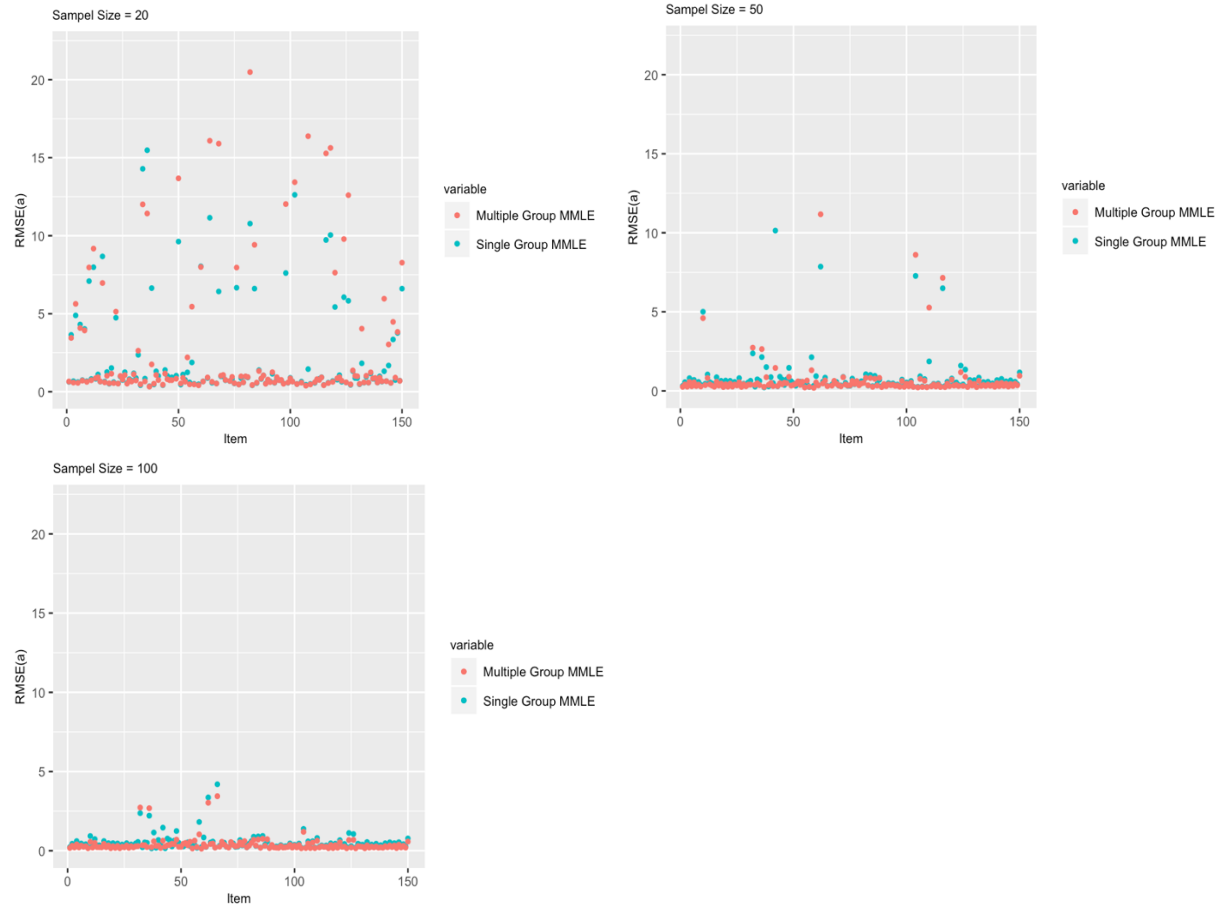


Figure 9.2. RMSE for a Parameters under 5-Form BIB Design across Different Sample Sizes

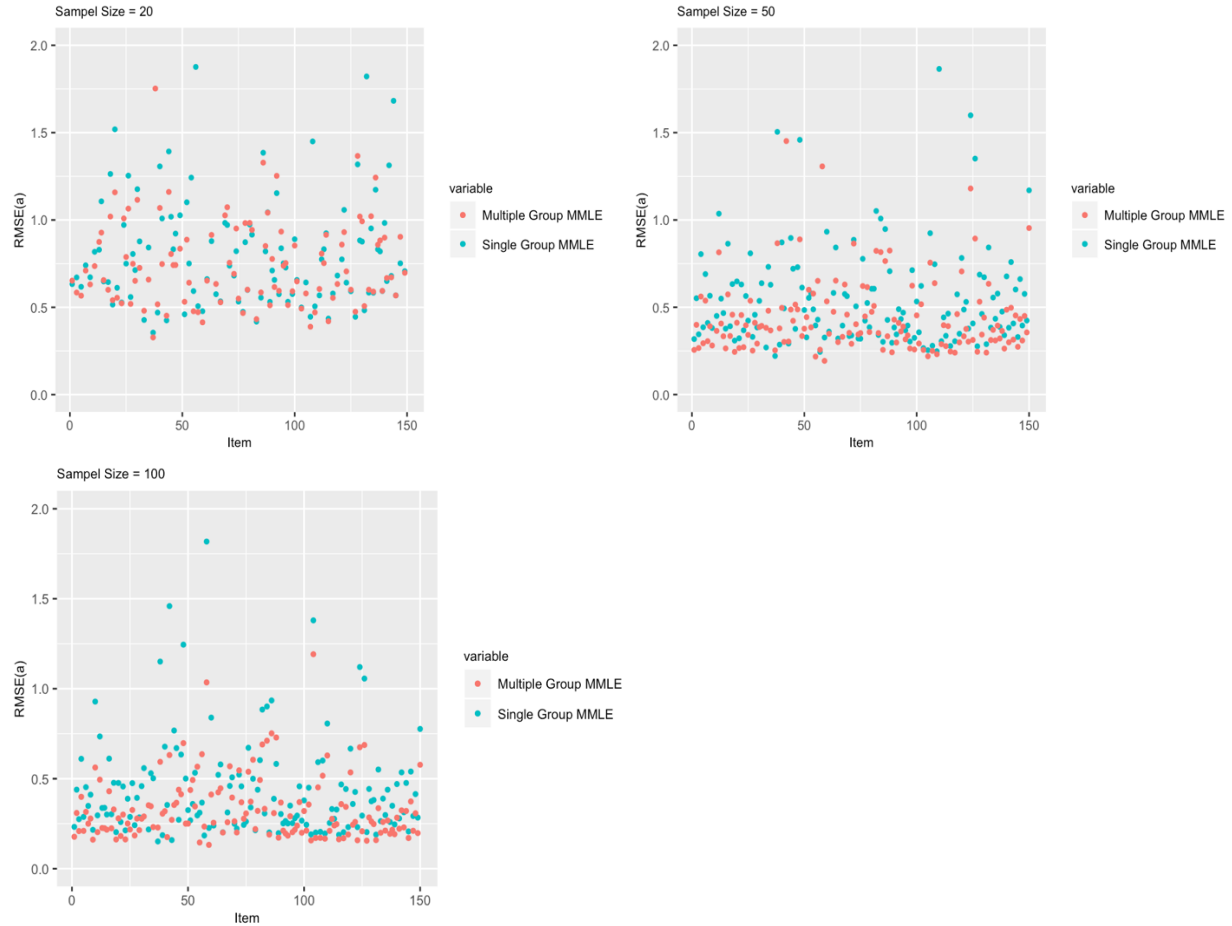


Figure 9.3. RMSE (from 0 to 2) for a Parameters under 5-Form BIB Design across Different Sample Sizes

Estimation of b parameters. For a parameter estimation, number of extreme estimation values, bias, and RMSE were presented.

Number of extreme estimation values. When sample size was 20, single group MMLE identified one extreme estimation values while multiple group MMLE identified none. When sample size was 50 or 100, both MMLE approaches identified no extreme estimation values. The table of the numbers can be found in Appendix G4. In general, multiple group MMLE was more stable than single group MMLE when sample size was 20, and two MMLE approaches had the same stabilities when sample size was 50 or 100.

Bias. Figure 9.4 presents bias for b parameters obtained from multiple group MMLE and single group MMLE for each item across different sample sizes. Items 1-150 were items from easy to hard. Regardless of sample size, for easy items, single group MMLE overestimated a parameters while multiple group MMLE almost unbiased them; for hard items, single group MMLE underestimated a parameters while multiple group MMLE almost unbiased them; for middle difficult items, bias were almost the same for single group MMLE and multiple group MMLE, but when items were easier, a parameters were overestimated by both MMLE approaches and when items were harder, a parameters were underestimated by both MMLE approaches. As sample size increased, bias dots were less scattered and bias range was decreased. The five number summaries of bias of the estimated a parameters are summarized in Appendix G5.

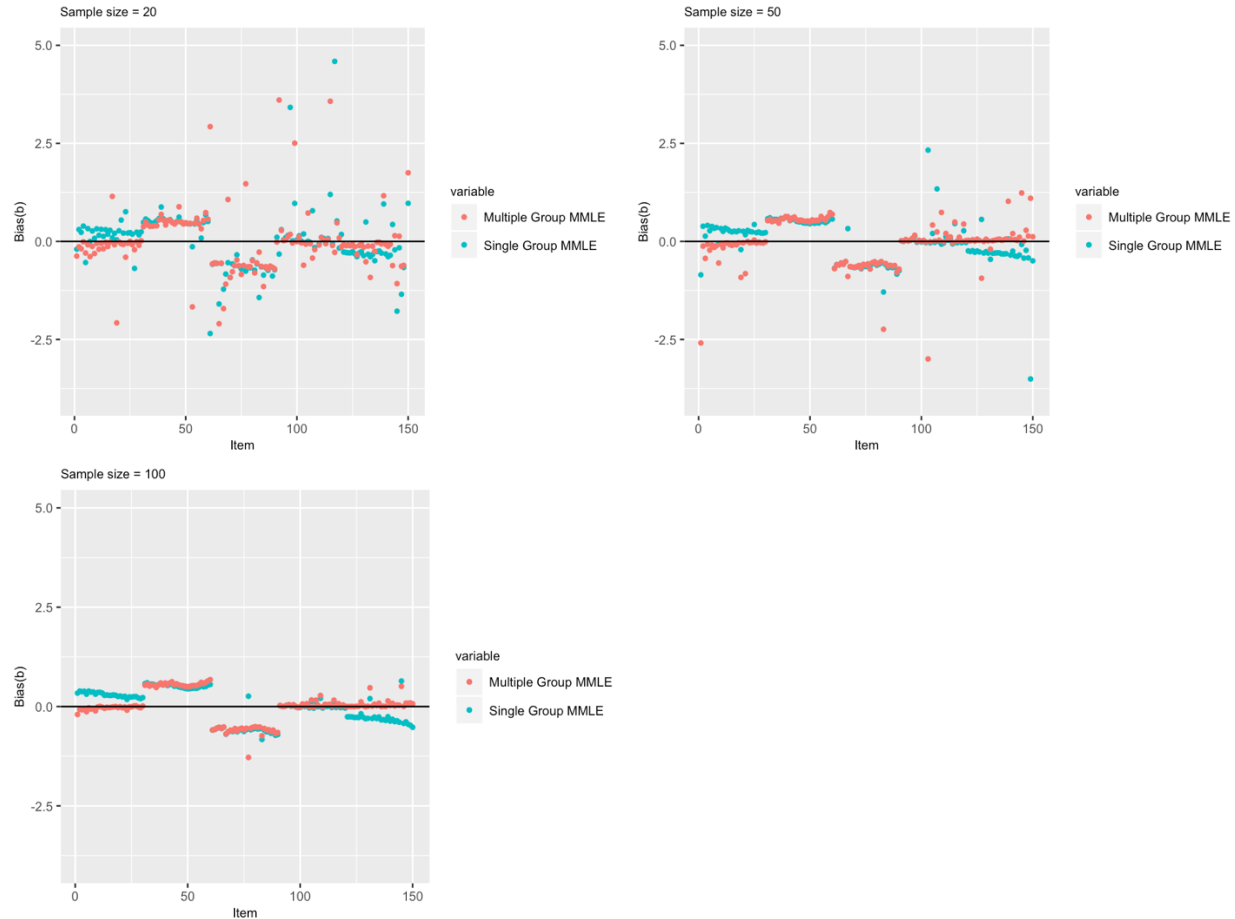


Figure 9.4. Bias for b Parameters under 5-Form BIB Design across Different Sample Size

RMSE. Figure 9.5 presents RMSE for b parameters obtained from multiple group MMLE and single group MMLE for each item across different sample sizes. Figure 9.6 presents the RMSE that was smaller than 2 to take a close look at the bottom of plots in Figure 9.5. It was found that when sample size was 20, RMSE obtained from single group MMLE was smaller than RMSE obtained from multiple group MMLE. Items from 1-150 were from easy to hard. For both MMLE approaches, RMSE for easy and hard items were closer to zero than RMSE for middle difficult items. When sample size was 20, RMSE obtained from single group MMLE was smaller than multiple group MMLE. When sample size was 50, RMSE obtained from single group MMLE was smaller than RMSE obtained from multiple group MMLE for middle difficult

items; while RMSE obtained from two MMLE approaches were the same for easy and hard items. When sample size was 100, RMSE obtained from single group MMLE was smaller than RMSE obtained from multiple group MMLE for moderate easy items; while RMSE obtained from single group MMLE was larger than RMSE obtained from multiple group MMLE for easy and hard items. In addition, as sample size increased, RMSE was closer to zero, RMSE range decreased, and RMSE dots were less scattered. The five number summaries of RMSE of the estimated b parameters are summarized in Appendix G6.

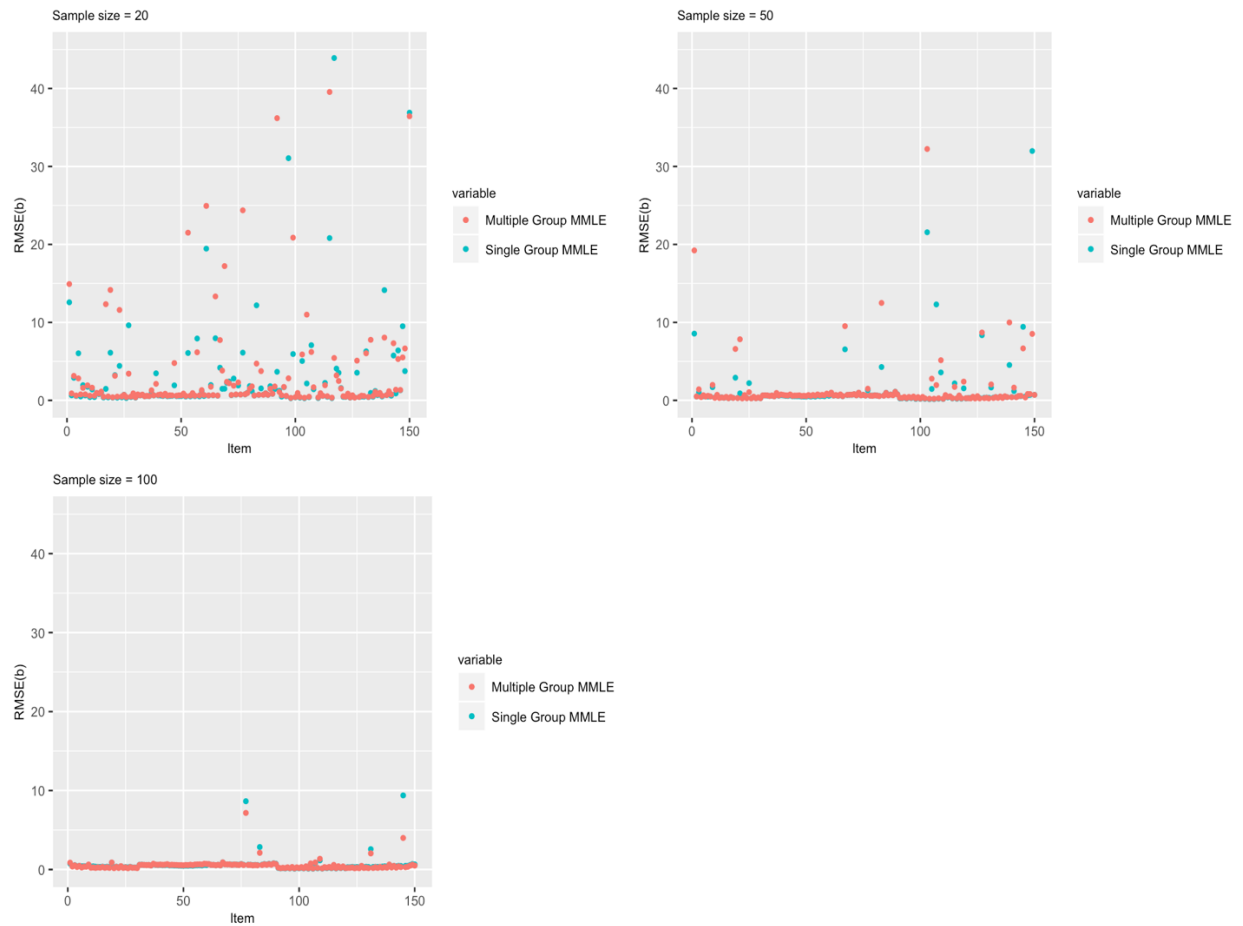


Figure 9.5. RMSE for b Parameters under 5-Form BIB Design across Different Sample Sizes

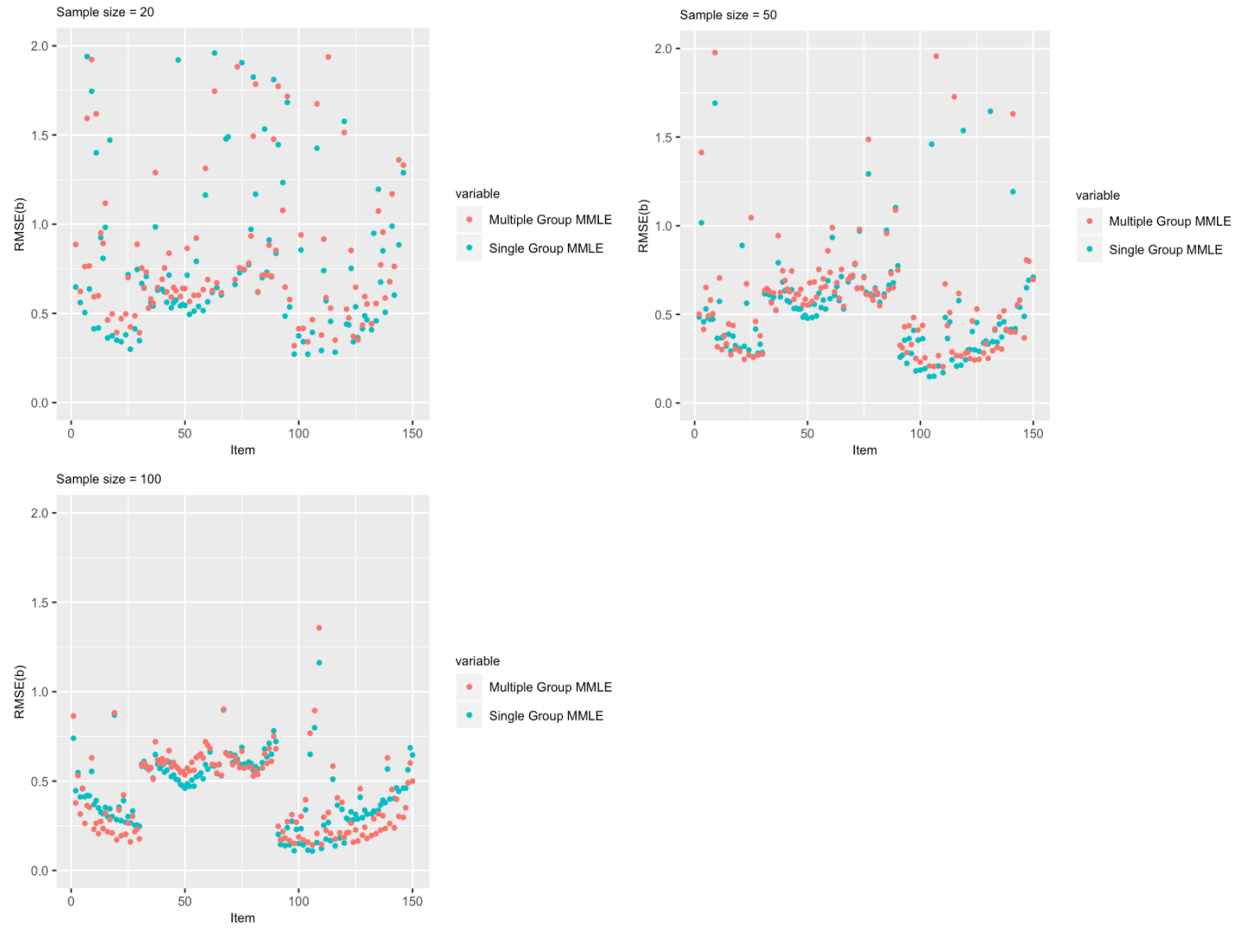


Figure 9.6. RMSE (from 0 to 2) for b Parameters under 5-Form BIB Design across Different Sample Sizes

COMPARISON BETWEEN 2-FORM AND 5-FORM DESIGNS

Because in this study, the NEAT design was the only design that had a 2-form design, when comparing the performance of single group MMLE and multiple group MMLE between 2-form and 5-form designs, results of 2-form NEAT design and 5-form NEAT design was compared.

In terms of stability of MMLE, 2-form NEAT design identified more extreme estimation values than 5-form NEAT design when using two MMLE approaches estimated a and b parameters (Figure 4.1, Figure 4.11, Figure 5.1, and Figure 5.4).

For bias and RMSE of a parameters, the trends of bias and RMSE obtained from 2-form NEAT design and 5-form NEAT design were similar (Figure 4.2, Figure 4.6, Figure 5.2, and Figure 5.9). Under 5-form NEAT design, when sample size was 20, two MMLE approaches performed the same; when sample size was 50 or 100, multiple group MMLE performed better than single group MMLE. For 2-form NEAT design, when sample size was 20 or 50, two MMLE approaches performed the same; when sample size was 100 or 200, multiple group MMLE performed better than single group MMLE.

For bias of b parameters, when sample size was 20, the trends of bias and RMSE obtained from 2-form NEAT design and 5-form NEAT design were similar (Figure 4.12-4.13, and Figure 5.5-5.6). Both single group MMLE and multiple group MMLE overestimated b parameters for easy items and underestimated b parameters for hard items. But when sample size was larger than 20, the difference appeared. For 5-form NEAT design, estimates obtained from multiple group MMLE were almost unbiased; but for 2-form NEAT design, estimates obtained from single group MMLE were more unbiased than those from multiple group MMLE.

For RMSE of b parameters, under 5-form NEAT design, RMSE obtained from single group MMLE was larger than RMSE obtained from multiple group MMLE for easy and hard items. But under 2-form NEAT design, RMSE obtained from single group MMLE was smaller than RMSE obtained from multiple group MMLE for all items (Figure 4.18 and Figure 5.11).

COMPARISON AMONG DIFFERENT PLANNED MISSING DATA DESIGNS

All comparisons were conducted between 5-form designs. Missing percentage is a considered factor when comparing the multiple form designs.

Comparison between NEAT design and panel missing design

We compared 5-form NEAT design and 5-form panel missing design because 5-form NEAT design (when 20 anchor items were used) had the largest missing percentage among all designs, 69.33% while 5-form panel missing design had the smallest missing percentage, 20%.

Comparing 5-form NEAT design (20 anchor items) and 5-form panel missing design, 5-form panel missing design identified much fewer extreme estimation values when estimating a and b parameters (Appendix B1, Appendix B4, Appendix E1, Appendix E4).

For bias of a parameters, 5-form panel missing design had a consistent trend that bias observed from multiple group MMLE was closer to zero than that from single group MMLE. But 5-form NEAT design (20 anchor items) didn't have a consistent pattern: when sample size was 20, bias obtained from single group MMLE was closer to zero; when sample size was 50 or 100, bias obtained from multiple group MMLE was closer to zero (Figure 4.3 and Figure 7.1).

For RMSE of a parameters, under both designs, RMSE obtained from single group MMLE was larger than RMSE obtained from multiple group MMLE. But the trend was more obvious under 5-form panel missing design (Figure 4.7-4.8, and Figure 7.2).

For bias of b parameters, under 5-form panel missing design, bias obtained from multiple group MMLE was almost unbiased. But under 5-form NEAT design (20 anchor items), when sample size was 20, multiple group MMLE overestimated b parameters for easy items and underestimated b parameters for hard items. The performances of single group MMLE under two

designs were the same: single group MMLE overestimated b parameters for easy items and underestimated b parameters for hard items (Figure 4.13 and Figure 7.3).

For RMSE for b parameters, the trends were similar when sample size was 50 or 100 for these two designs. That was RMSE obtained from single group MMLE was larger than that obtained from multiple group MMLE for easy and hard items, and single group MMLE was smaller for middle difficult items. But when sample size was 20, under 5-form panel missing design, RMSE observed for single group MMLE was smaller than that for multiple group MMLE; while under while 5-form NEAT design (20 anchor items), there didn't exist one clear pattern (Figure 4.17-4.18, Figure 7.4- 7.5).

Comparison between NEAT design and multiform design

We compared 5-form NEAT design and 5-form multiform design because they were the only two designs that including an anchor item set X. 5-form multiform design had more overlaps among test forms and less missing percentages than 5-form NEAT design. When there were 20, 30, 50 anchor items in item set X, 5-form NEAT design had 69.33%, 64%, 53.33% missing percentages respectively, and 5-form multiform design had 52%, 48%, 40% missing percentages respectively.

Comparing 5-form NEAT design and 5-form multiform design, 5-form multiform design identified fewer extreme estimation values when estimating a and b parameters (Appendix B1, Appendix B4, Appendix D1, Appendix D4).

For bias of a parameters, 5-form multiform design had a consistent trend that bias observed from multiple group MMLE was closer to zero than that from single group MMLE. But 5-form NEAT design didn't have a consistent pattern: when sample size was 20, bias obtained

from two MMLE approaches were very close; when sample size was 50 or 100, bias obtained from multiple group MMLE was closer to zero (Figure 4.2 and Figure 6.2).

For RMSE of a parameters, two designs had the same trend of RMSE obtained from single group MMLE and multiple group MMLE: when sample size was 20, RMSE obtained from two MMLE approaches were close; when sample size was 50 or 100, RMSE obtained from multiple group MMLE was smaller than that obtained from single group MMLE. But when sample size was 20, RMSE for 5-form multiform design was closer to zero than RMSE for 5-form NEAT design (Figure 4.6 and Figure 6.6).

For bias of b parameters, 5-form multiform design had a consistent trend that multiple group MMLE was almost unbiased while single group MMLE overestimated b parameters for easy items and underestimated b parameters for hard items. 5-form NEAT design had the same pattern for single group MMLE as multiform design but didn't had a consistent pattern for multiple group MMLE: when sample size was 20, multiple group MMLE overestimated b parameters for easy items and underestimated b parameters for hard items; when sample size was 50 or 100, multiple group MMLE was almost unbiased (Figure 4.13, and Figure 6.13).

For RMSE of b parameters, 5-form multiform design had a consistent trend that RMSE observed for single group MMLE was larger than RMSE observed for multiple group MMLE for easy and hard items, and RMSE observed for single group MMLE was smaller than that observed for multiple group MMLE for middle difficult items. That was the same as RMSE under 5-form NEAT design when sample size was 100. But for 5-form NEAT design, when sample size was 20, there didn't exist a clear pattern for RMSE obtained from two MMLE approaches; and when sample size was 50, RMSE observed for two MMLE approaches were close for easy and hard items, and RMSE observed for single group MMLE was smaller than

that observed for multiple group MMLE for middle difficult items (Figure 4.17-4.18, and Figure 6.17-6.18).

Comparison between 5-form NEAT design and panel chained design

5-form NEAT design and 5-form panel chained design had one similarity that in both of them, two adjacent item sets would be assigned to two adjacent test forms. For example, item sets A-B were assigned to test forms 1-2; item sets B-C were assigned to test forms 2-3... But 5-form NEAT design had one anchor item set X while 5-form panel chained design didn't have one. 5-form NEAT design had 69.33%, 64%, 53.33% missing percentages when there were 20, 30, or 50 anchor items. 5-form panel chained design had a missing percentage of 60.27%.

Comparing 5-form NEAT design and 5-form panel chained design, 5-form chained design identified fewer extreme estimation values when estimating a and b parameters (Appendix B1, Appendix B4, Appendix F1, Appendix F4).

For bias for a parameters, two designs had the same trend of bias that when sample size was 20, bias obtained from single group MMLE was close to bias obtained from multiple group MMLE; when sample size was 50 or 100, bias obtained from multiple group MMLE was closer to zero than bias obtained from single group MMLE (Figure 4.2-4.3, and Figure 8.1).

For RMSE for a parameters, two designs had the same trend of RMSE that when sample size was 20, RMSE obtained from two MMLE approaches were close; when sample size was 50 or 100, RMSE obtained from multiple group MMLE was smaller than that obtained from single group MMLE (Figure 4.6-4.8, and 8.2-8.3).

For bias for b parameters, 5-form panel chained design had a consistent pattern that multiple group MMLE was almost unbiased while single group MMLE overestimated b parameters for easy items and underestimated b parameters for hard items. However, for 5-form

NEAT design, both MMLE approaches overestimated easy items and overestimated hard items when sample size was 20. And when sample size was 50 or 100, the pattern was the same as the pattern under 5-form panel chained design (Figure 4.12-4.13, and Figure 8.4).

For RMSE for b parameters, when sample size was 20, under 5-form panel chained design, RMSE obtained from single group MMLE was smaller than that from multiple group MMLE; but under 5-form NEAT design, there didn't exist a clear pattern. When sample size was 50, under both designs, RMSE obtained from two MMLE approaches were the same for easy and hard items and RMSE obtained from single group MMLE was smaller for middle difficult items. When sample size was 100, under both designs, RMSE obtained from single group MMLE was larger for easy and hard items and was smaller for middle difficult items (Figure 4.16-4.18, and Figure 8.5-8.6).

Comparison between BIB design and the rest of designs

For 5-form NEAT design, 5-form multifactor design, and 5-form panel chained design, performances of single group MMLE and multiple group MMLE were similar when a and b parameters were estimated. However, 5-form BIB design had different patterns from those designs. The difference wasn't reflected in different signs of values or the inverse magnitude relationships as 5-form panel missing design but was reflected in totally different shapes of the result graphs:

For bias for a parameters, the shape of the graph of 5-form BIB design was like a spindle, which was round in the middle but flat at both ends (Figure 9.1). But in the rest of designs, the shape of the plots was a rectangle.

For bias and RMSE for b parameters, the graph was truncated into four segments under BIB design (Figure 9.4-9.6). But under the rest of designs, the graph was continuous.

CHAPTER 5

DISCUSSION

According to our results, it can be seen that the performances of single group MMLE and multiple group MMLE were different based on different multiple form designs and different group sample sizes. But the impacts of the number of anchor items and the anchor item selection strategies were tiny according to boxplots. They were even not obvious when shown in scatterplots.

Usually, for a parameter estimation, the multiple group MMLE performed better than single group MMLE when sample size was 50, 100, or 200. But when sample size was 20, two MMLE approaches performed very similar except for 5-form multiform design and 5-form panel missing design. Under 5-form multiform design and 5-form panel missing design, multiple group MMLE performed better.

For b parameter estimation, in most cases (under 5-form NEAT design, 5-form multiform design, 5-form panel missing design, and 5-form panel chained design), results of multiple group MMLE were almost unbiased while single group MMLE overestimated easy items and underestimated hard items regardless of sample size. But for 5-form NEAT design, when sample size was 20, multiple group MMLE overestimated easy items and underestimated hard items as well. Specifically, 2-form NEAT design and BIB design had very different patterns from other designs. Under 2-form NEAT design, when sample size was 20, multiple group MMLE overestimated easy items and underestimated hard items as single group MMLE; when sample size was 50, 100, or 200, results of single group MMLE were almost unbiased while multiple

group MMLE overestimated b parameters for items from all difficulty levels. Under 5-form BIB design, results of multiple group MMLE were almost unbiased for easy and hard items while single group MMLE overestimated easy items and underestimated hard items. Two MMLE approaches performed the same for middle difficult items. They overestimated moderate easy items and underestimated moderate hard items.

In terms of stabilities of two MMLE approaches, multiple group MMLE was more stable for a and b parameter estimation under 5-form multiform design, 5-form panel missing design, and 5-form BIB design. Under 5-form panel chained design, multiple group MMLE was more stable for a parameter estimation, but there was no consistent pattern for b parameter estimation. Under 5-form NEAT design and 2-form NEAT design, there was no consistent pattern existed for both a and b parameter estimation.

Overall, larger sample size (sample size larger than 20) was recommended because MMLE performed better as sample size increased. Besides, 5-form multiform design and 5-form panel missing design was recommended because MMLE under these two designs performed very well even when sample size was really small, for example, when sample size was 20. In these two designs, multiple group MMLE was recommended because it performed better than single group MMLE. When time is limited and students can only answer a few items, 5-form multiform design would be more proper because the length of each test form under this design was shorter.

This study takes various conditions into account. But more explorations can be made in the future. Firstly, more complicated IRT models can be used. For example, the three-parameter logistic (3-PL) model (Birnbbaum, 1968) which introduces a guessing parameter. Secondly, different sample sizes can be considered (Hulin, Lissak, and Drasgow, 1982). For example, Jia

and colleagues (2014) investigated sample sizes ranging from 60 to 300 under planned missing designs. Thirdly, anchor item selection strategies can be modified. For example, to select an anchor item set which is a miniature version of the test or which has less variance in item difficulty (Kolen & Brennan, 2004; Sinharay & Holland, 2006; Sinharay, Haberman, Holland, & Lewis, 2012). Number of anchor items can be changed as well. For example, 5, 10, or 25 (Kim, & Cohen, 1998). Besides, researchers can investigate planned missing data designs beyond the targeted testing design, for example, multistage testing designs (Eggen & Verhelst, 2011).

Reference

- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In Lord, F.M. & Novick, M.R., *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Campbell, B. F., Sengupta, S., Santos, C., & Lorig, K. R. (1995). Balanced incomplete block design: description, case study, and implications for practice. *Health education quarterly*, 22(2), 201-210.
- DeMars, C. (2002). Incomplete data and item parameter estimates under JMLE and MML estimation. *Applied Measurement in Education*, 15(1), 15-31
- Eggen, T.J.H.M. & Verhelst, N.D. (2011). Item calibration in incomplete testing designs. *Psicológica*, 32(1), 107-132.
- Graham, J. W., Taylor, B. J., & Cumsille, P. E. (2001). Planned missing-data designs in analysis of change. Pp. 335–353. In: *New Methods for the Analysis of Change*, edited by L.M. Collins and A.G. Sayer. Washington, DC: American Psychological Association.
- Hulin, C.L., Lissak, R.I., & Drasgow, F. (1982). Recovery of two-and three-parameter logistic item characteristic curves: A Monte Carlo study. *Applied psychological measurement*, 6(3), 249-260.
- Jia, F., Moore, E. W. G., Kinai, R., Crowe, K. S., Schoemann, A. M., & Little, T. D. (2014). Planned missing data designs with small sample sizes: How small is too small?. *International Journal of Behavioral Development*, 38(5), 435-452.
- Kim, S.-H., & Cohen, A.S. (1998). A comparison of linking and concurrent calibration under item response theory. *Applied Psychological Measurement*, 22(2), 131–143.

- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices (2nd ed.)*. New York, NY: Springer.
- Little, T. D., & Rhemtulla, M. (2013). Planned missing data designs for developmental researchers. *Child Development Perspectives*, 7(4), 199-204.
- Mislevy, R.J. & Sheenan, K.M. (1989). The role of collateral information about examinees in item parameter estimation. *Psychometrika*, 54, 661-680.
- Pokropek, A. (2011). Missing by design: Planned missing-data designs in social science. *ASK. Research & Methods*, (20), 81-105.
- Rhemtulla, M., & Hancock, G. R. (2016). Planned missing data designs in educational psychology research. *Educational Psychologist*, 51(3-4), 305-316.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen, Denmark: Danish Institute for Educational Research.
- Rubin, D.B. (1976). Inference and missing data. *Biometrika*, 63, 581-592.
- Sinharay, S., Haberman, S., Holland, P., & Lewis, C. (2012). A note on the choice of an anchor test in equating. *ETS Research Report Series*, 2012(2), i-9.
- Sinharay, S., & Holland, P. W. (2006). *The correlation between the scores of a test and an anchor test (ETS Research Report No. RR-06-04)*. Princeton, NJ: ETS.
- von Davier, A. (Ed.). (2010). *Statistical models for test equating, scaling, and linking*. Springer Science & Business Media.
- von Davier, A. A., & Wilson, C. (2007). IRT true-score test equating: A guide through assumptions and applications. *Educational and Psychological Measurement*, 67(6), 940-957.

Wang, S. (2018). The development of computerized adaptive testing with response revision to improve classroom instruction and assessment. Project funded by the 2018 Early Career Researcher Award. College of Education, University of Georgia.

Appendices

Appendix A1. *Anchor Item Selection for 5-Form NEAT Design*

anchor item set X		<i>item</i>		<i>item set</i>
	← 2 (1)	very low <i>b</i> low <i>a</i>	13 (14) →	A
	← 2 (1)	very low <i>b</i> high <i>a</i>	13 (14) →	
	← 2 (3)	low <i>b</i> low <i>a</i>	13 (12) →	B
	← 2 (2)	low <i>b</i> high <i>a</i>	13 (13) →	
	← 2 (3)	middle <i>b</i> low <i>a</i>	13 (12) →	C
	← 2 (3)	middle <i>b</i> high <i>a</i>	13 (13) →	
	← 2 (2)	high <i>b</i> low <i>a</i>	13 (13) →	D
	← 2 (3)	high <i>b</i> high <i>a</i>	13 (12) →	
	← 2 (1)	very high <i>b</i> low <i>a</i>	13 (14) →	E
	← 2 (1)	very high <i>b</i> high <i>a</i>	13 (14) →	

Appendix A2. *Item Parameters*

ID	b	a	ID	b	a
1	-2.3685212	0.60466409	76	0.18727677	1.18304899
2	-2.0044002	1.24949551	77	0.19051026	0.76387519
3	-1.9992891	0.68376112	78	0.21035584	1.71548933
4	-1.9149747	1.72500092	79	0.22444576	1.0450135
5	-1.8765044	0.79811737	80	0.22518475	1.12634592
6	-1.8286329	1.53927859	81	0.2379574	0.68908561
7	-1.8266932	0.95707592	82	0.25519487	1.55539969
8	-1.8028695	1.21444917	83	0.26928708	0.61604548
9	-1.7583307	0.55171323	84	0.29186774	1.49084582
10	-1.5134759	2.37198635	85	0.33305949	0.65673218
11	-1.4845677	0.9418445	86	0.33619	1.11876218
12	-1.4519178	2.36933591	87	0.34082847	0.58737493
13	-1.2829928	1.07281756	88	0.38858123	2.33568947
14	-1.20907	1.15711966	89	0.44547975	0.56504218
15	-1.1894537	0.83184133	90	0.46042165	1.43066068
16	-1.1511721	1.86206469	91	0.46709953	0.93999194
17	-1.1417098	0.80398691	92	0.47357907	1.33599868
18	-1.0641839	1.46209288	93	0.50349347	0.86655481
19	-1.0426581	0.39131118	94	0.50427051	1.42337512
20	-1.0061644	1.48513621	95	0.51862999	0.87783879
21	-0.9987152	0.5463284	96	0.51973938	1.18119867
22	-0.9824236	1.48337247	97	0.53482999	0.86926607
23	-0.968869	0.48386038	98	0.57839323	1.92627407
24	-0.8856832	1.26307569	99	0.57935817	0.84825685
25	-0.8614003	0.82280155	100	0.58935314	1.51259092
26	-0.840958	1.63285678	101	0.58982907	0.88049713
27	-0.8115438	0.64220894	102	0.61370787	1.74584887
28	-0.7602933	1.18579595	103	0.65959378	0.60692041
29	-0.7447795	0.84270749	104	0.7060618	3.41722552
30	-0.7028795	1.34269202	105	0.7162727	0.4760255
31	-0.7002809	0.70367425	106	0.73282897	2.30351409
32	-0.6942423	3.86268632	107	0.75973777	0.4007905
33	-0.6436342	1.00058477	108	0.77244723	1.81822308
34	-0.6428811	1.88776163	109	0.78113682	0.36707096
35	-0.6226594	0.89750198	110	0.78478978	2.61025768
36	-0.5903992	4.19908566	111	0.78863639	0.91252307
37	-0.5759164	0.61188446	112	0.79101298	1.26717439
38	-0.5725055	1.34301158	113	0.79868725	0.8822248

39	-0.5478961	1.0363807	114	0.8097397	1.25886538
40	-0.5387017	1.11917564	115	0.82337591	0.46919448
41	-0.4662867	0.87818895	116	0.82764804	2.12202478
42	-0.4509992	2.3439437	117	0.86262772	0.65005379
43	-0.4363446	0.94701369	118	0.8834651	1.47782056
44	-0.3900805	1.29676109	119	0.91391156	0.72483076
45	-0.3682271	0.73247233	120	0.93271577	2.0374759
46	-0.3443333	1.59417743	121	0.94404278	1.08121043
47	-0.3422707	0.40117938	122	0.95574339	1.33535425
48	-0.3291436	1.14412888	123	0.98066222	0.75107275
49	-0.2864598	0.95060076	124	1.04963474	2.88536571
50	-0.280212	1.28627023	125	1.07939673	0.90632952
51	-0.2605846	1.09327728	126	1.08870816	2.6035078
52	-0.259787	2.02036705	127	1.09045517	0.38385617
53	-0.2590082	0.30597233	128	1.11448179	1.49910738
54	-0.2545372	2.00716036	129	1.19571939	1.07556463
55	-0.2506168	0.59373092	130	1.20711113	1.39068564
56	-0.2468845	1.72349478	131	1.23635984	0.36801023
57	-0.236646	0.79204003	132	1.24337184	1.85554772
58	-0.2162799	1.60323167	133	1.27078361	0.74438298
59	-0.1967047	0.5623752	134	1.31117521	1.28677007
60	-0.1830838	1.13609556	135	1.39344194	0.84183885
61	-0.142639	0.88840905	136	1.40655841	1.35914243
62	-0.123768	1.09816256	137	1.42342348	0.87853099
63	-0.1198302	0.66646216	138	1.50474144	1.09993944
64	-0.1088835	1.69015014	139	1.66887165	0.52650399
65	-0.1013222	1.0092435	140	1.69661425	1.44255304
66	-0.0877451	1.48052224	141	1.7036697	0.74979293
67	0.00116919	0.40598675	142	1.72521942	1.61263856
68	0.01840208	1.82860899	143	1.72745545	0.76037156
69	0.02193052	0.72510375	144	1.91371938	1.29604199
70	0.0714691	1.30658256	145	1.91889789	0.41512925
71	0.08805961	0.88387267	146	1.92396651	1.50936338
72	0.14057685	2.61688128	147	2.11930331	0.93116435
73	0.14569152	0.70458911	148	2.45765794	1.11090825
74	0.15616236	1.47141366	149	2.70272576	0.7067197
75	0.17835973	1.0110256	150	2.8697336	1.57925905

Appendix B1. *Number of Extreme Estimation Values for a Parameters under 5-Form NEAT Design*

Sample size	Number of anchor items	Selection strategy of anchor items	Single-group MMLE	Multiple-group MMLE	Total
20	20	Uniformly	41	48	49
		Non-uniformly	33	36	36
	30	Uniformly	33	37	38
		Non-uniformly	34	33	35
	50	Uniformly	31	30	31
		Non-uniformly	27	26	28
50	20	Uniformly	2	2	2
		Non-uniformly	1	1	1
	30	Uniformly	3	3	3
		Non-uniformly	2	2	2
	50	Uniformly	2	1	2
		Non-uniformly	2	1	2
100	20	Uniformly	1	0	1
		Non-uniformly	0	0	0
	30	Uniformly	1	0	1
		Non-uniformly	1	0	1
	50	Uniformly	1	0	1
		Non-uniformly	1	0	1

Appendix B2. *Five Number Summaries for Bias for a Parameters under 5-Form NEAT Design*

Estimation method	sample size	number of anchor items	selection strategy of anchor items	min	Q1	median	Q3	max
Single-group MMLE	20	20	uniformly	0.059	0.413	0.974	2.174	4.744
			non-uniformly	0.072	0.403	1.093	2.337	5.260
		30	uniformly	0.084	0.344	0.753	1.812	4.848
			non-uniformly	-0.058	0.344	0.649	1.801	4.396
		50	uniformly	0.059	0.287	0.498	1.325	4.345
			non-uniformly	0.062	0.299	0.522	1.332	4.497
	50	20	uniformly	0.073	0.199	0.287	0.459	3.161
			non-uniformly	0.031	0.208	0.310	0.520	3.021
		30	uniformly	0.047	0.220	0.319	0.479	2.332
			non-uniformly	0.055	0.220	0.305	0.503	2.186
		50	uniformly	0.024	0.207	0.310	0.483	2.667
			non-uniformly	-0.013	0.211	0.333	0.502	2.674
	100	20	uniformly	0.055	0.155	0.238	0.333	1.110
			non-uniformly	0.007	0.173	0.256	0.383	1.904
		30	uniformly	0.045	0.179	0.251	0.364	1.555
			non-uniformly	0.028	0.184	0.250	0.386	1.155
		50	uniformly	0.029	0.176	0.268	0.373	2.126
			non-uniformly	0.039	0.178	0.274	0.391	1.234
Multiple-group MMLE	20	20	uniformly	0.105	0.556	1.357	2.531	4.792
			non-uniformly	-0.159	0.475	1.253	2.378	4.711
		30	uniformly	0.075	0.381	0.890	1.951	5.048
			non-uniformly	-0.061	0.295	0.651	1.940	5.091
		50	uniformly	0.057	0.203	0.422	1.347	4.619
			non-uniformly	0.018	0.231	0.422	1.508	4.474
	50	20	uniformly	-0.006	0.067	0.124	0.229	2.987
			non-uniformly	-0.050	0.055	0.106	0.198	2.111
		30	uniformly	-0.018	0.047	0.101	0.186	1.642
			non-uniformly	-0.026	0.028	0.081	0.167	1.928
		50	uniformly	-0.041	0.014	0.072	0.142	2.200
			non-uniformly	-0.064	0.008	0.055	0.133	2.334
	100	20	uniformly	-0.021	0.026	0.055	0.096	1.353
			non-uniformly	-0.049	0.020	0.048	0.095	0.840
		30	uniformly	-0.041	0.014	0.036	0.074	1.872
			non-uniformly	-0.024	0.008	0.034	0.071	1.066
		50	uniformly	-0.040	0.001	0.017	0.062	1.815
			non-uniformly	-0.031	0.008	0.026	0.070	1.592

Appendix B3. *Five Number Summaries for RMSE for a Parameters under 5-Form NEAT Design*

Estimation method	sample size	number of anchor items	selection strategy of anchor items	min	Q1	median	Q3	max
Single-group MMLE	20	20	uniformly	0.249	1.121	5.605	10.854	40.957
			non-uniformly	0.245	1.105	5.332	10.109	28.321
		30	uniformly	0.233	0.928	1.841	8.941	25.162
			non-uniformly	0.258	0.874	1.378	8.712	21.852
		50	uniformly	0.263	0.557	1.043	7.131	27.075
			non-uniformly	0.224	0.638	1.141	6.729	25.766
	50	20	uniformly	0.157	0.490	0.628	0.963	16.099
			non-uniformly	0.172	0.508	0.659	1.029	12.862
		30	uniformly	0.171	0.477	0.615	0.941	12.892
			non-uniformly	0.180	0.484	0.614	0.892	7.178
		50	uniformly	0.167	0.439	0.580	0.858	15.338
			non-uniformly	0.176	0.428	0.579	0.864	16.077
	100	20	uniformly	0.125	0.347	0.425	0.589	1.804
			non-uniformly	0.121	0.348	0.449	0.633	7.853
		30	uniformly	0.156	0.342	0.427	0.564	3.456
			non-uniformly	0.126	0.326	0.439	0.605	1.582
		50	uniformly	0.121	0.315	0.398	0.558	8.489
			non-uniformly	0.123	0.317	0.418	0.558	2.390
Multiple-group MMLE	20	20	uniformly	0.268	1.318	6.966	12.109	22.995
			non-uniformly	0.262	1.227	6.035	10.613	31.572
		30	uniformly	0.228	0.903	4.117	9.346	27.559
			non-uniformly	0.244	0.847	1.585	9.420	29.372
		50	uniformly	0.244	0.497	1.101	7.792	26.323
			non-uniformly	0.197	0.558	1.087	6.842	22.730
	50	20	uniformly	0.111	0.418	0.540	0.780	14.879
			non-uniformly	0.121	0.418	0.535	0.792	17.019
		30	uniformly	0.124	0.380	0.492	0.708	14.813
			non-uniformly	0.134	0.384	0.474	0.704	8.539
		50	uniformly	0.118	0.278	0.428	0.646	16.342
			non-uniformly	0.109	0.281	0.426	0.610	16.201
	100	20	uniformly	0.075	0.271	0.337	0.462	8.059
			non-uniformly	0.078	0.272	0.338	0.446	4.955
		30	uniformly	0.093	0.241	0.307	0.403	12.956
			non-uniformly	0.079	0.244	0.308	0.414	5.989
		50	uniformly	0.085	0.193	0.276	0.372	9.882
			non-uniformly	0.080	0.193	0.272	0.382	8.030

Appendix B4. *Number of Extreme Estimation Values for b Parameters under 5-Form NEAT Design*

Sample size	Number of anchor items	Selection strategy of anchor items	Single-group MMLE	Multiple-group MMLE	Total
20	20	Uniformly	3	4	7
		Non-uniformly	7	2	9
	30	Uniformly	2	3	5
		Non-uniformly	2	2	4
	50	Uniformly	4	2	6
		Non-uniformly	1	1	2
50	20	Uniformly	0	0	0
		Non-uniformly	1	1	1
	30	Uniformly	1	3	4
		Non-uniformly	3	2	4
	50	Uniformly	2	3	5
		Non-uniformly	0	1	1
100	20	Uniformly	1	0	1
		Non-uniformly	0	1	1
	30	Uniformly	1	0	1
		Non-uniformly	0	0	0
	50	Uniformly	1	0	1
		Non-uniformly	0	0	0

Appendix B5. *Five Number Summaries for Bias for b Parameters under 5-Form NEAT Design*

Estimation method	sample size	number of anchor items	selection strategy of anchor items	min	Q1	median	Q3	max
Single-group MMLE	20	20	uniformly	-5.189	-0.314	-0.067	0.192	4.251
			non-uniformly	-4.077	-0.266	-0.052	0.137	2.647
		30	uniformly	-6.581	-0.292	-0.100	0.153	3.356
			non-uniformly	-4.712	-0.288	-0.073	0.132	3.028
		50	uniformly	-3.498	-0.238	-0.065	0.120	2.734
			non-uniformly	-4.061	-0.207	-0.068	0.071	2.958
	50	20	uniformly	-1.189	-0.194	-0.029	0.128	2.052
			non-uniformly	-2.002	-0.188	-0.047	0.119	2.659
		30	uniformly	-1.669	-0.184	-0.030	0.125	2.074
			non-uniformly	-2.201	-0.178	-0.033	0.126	3.296
		50	uniformly	-1.772	-0.186	-0.042	0.107	0.499
			non-uniformly	-2.667	-0.160	-0.038	0.106	1.846
	100	20	uniformly	-2.019	-0.176	-0.014	0.144	2.857
			non-uniformly	-2.446	-0.164	-0.033	0.121	0.457
		30	uniformly	-1.046	-0.156	-0.026	0.126	2.728
			non-uniformly	-0.824	-0.158	-0.016	0.132	1.608
		50	uniformly	-3.850	-0.159	-0.028	0.113	0.472
			non-uniformly	-0.491	-0.141	-0.020	0.123	1.134
Multiple-group MMLE	20	20	uniformly	-4.712	-0.341	-0.083	0.178	2.694
			non-uniformly	-3.099	-0.279	-0.083	0.121	4.439
		30	uniformly	-2.698	-0.249	-0.087	0.110	6.160
			non-uniformly	-4.930	-0.248	-0.105	0.053	6.201
		50	uniformly	-3.006	-0.203	-0.072	0.044	3.036
			non-uniformly	-4.359	-0.178	-0.067	0.020	2.124
	50	20	uniformly	-5.042	-0.049	-0.016	0.021	1.456
			non-uniformly	-1.280	-0.039	-0.010	0.030	4.919
		30	uniformly	-2.159	-0.052	-0.022	0.007	2.764
			non-uniformly	-5.362	-0.050	-0.009	0.045	6.056
		50	uniformly	-1.523	-0.042	-0.008	0.029	5.296
			non-uniformly	-2.854	-0.046	-0.005	0.025	3.084
	100	20	uniformly	-4.803	-0.015	0.004	0.023	0.452
			non-uniformly	-2.857	-0.016	0.002	0.022	5.354
		30	uniformly	-1.276	-0.011	0.006	0.023	1.313
			non-uniformly	-0.995	-0.011	0.004	0.027	0.981
		50	uniformly	-0.951	-0.010	0.009	0.034	0.974
			non-uniformly	-0.743	-0.009	0.004	0.025	1.060

Appendix B6. *Five Number Summaries for RMSE for b Parameters under 5-Form NEAT Design*

Estimation method	sample size	number of anchor items	selection strategy of anchor items	min	Q1	median	Q3	max
Single-group MMLE	20	20	uniformly	0.157	0.578	1.710	4.375	54.445
			non-uniformly	0.116	0.457	1.361	3.287	36.099
		30	uniformly	0.124	0.419	1.306	3.510	62.797
			non-uniformly	0.136	0.403	1.166	3.846	46.281
		50	uniformly	0.126	0.307	0.846	2.320	36.926
			non-uniformly	0.119	0.313	0.758	2.385	37.813
	50	20	uniformly	0.104	0.277	0.409	0.919	19.290
			non-uniformly	0.078	0.264	0.411	0.771	25.358
		30	uniformly	0.085	0.238	0.349	0.721	26.991
			non-uniformly	0.087	0.224	0.362	0.806	27.627
		50	uniformly	0.082	0.205	0.332	0.577	17.626
			non-uniformly	0.078	0.203	0.325	0.587	20.022
	100	20	uniformly	0.081	0.215	0.321	0.426	31.395
			non-uniformly	0.067	0.209	0.301	0.439	15.441
		30	uniformly	0.064	0.183	0.277	0.398	27.477
			non-uniformly	0.072	0.178	0.269	0.409	16.247
		50	uniformly	0.059	0.166	0.252	0.380	35.693
			non-uniformly	0.063	0.157	0.254	0.370	10.928
Multiple-group MMLE	20	20	uniformly	0.170	0.531	1.269	3.605	50.251
			non-uniformly	0.139	0.446	1.092	3.216	43.181
		30	uniformly	0.160	0.430	1.082	3.608	62.274
			non-uniformly	0.140	0.410	0.976	3.431	58.871
		50	uniformly	0.160	0.348	0.990	2.879	28.318
			non-uniformly	0.148	0.338	0.791	2.607	46.417
	50	20	uniformly	0.161	0.305	0.411	1.030	37.731
			non-uniformly	0.157	0.292	0.425	0.907	48.137
		30	uniformly	0.149	0.271	0.366	0.854	33.290
			non-uniformly	0.156	0.268	0.386	1.055	167.917
		50	uniformly	0.147	0.242	0.348	0.656	54.356
			non-uniformly	0.148	0.255	0.342	0.633	32.309
	100	20	uniformly	0.134	0.205	0.285	0.397	43.125
			non-uniformly	0.124	0.206	0.273	0.384	64.242
		30	uniformly	0.123	0.188	0.257	0.389	22.742
			non-uniformly	0.117	0.187	0.255	0.388	14.564
		50	uniformly	0.109	0.175	0.225	0.371	13.517
			non-uniformly	0.116	0.168	0.226	0.345	11.449

Appendix C1. *Number of Extreme Estimation Values for a Parameters under 2-Form NEAT Design*

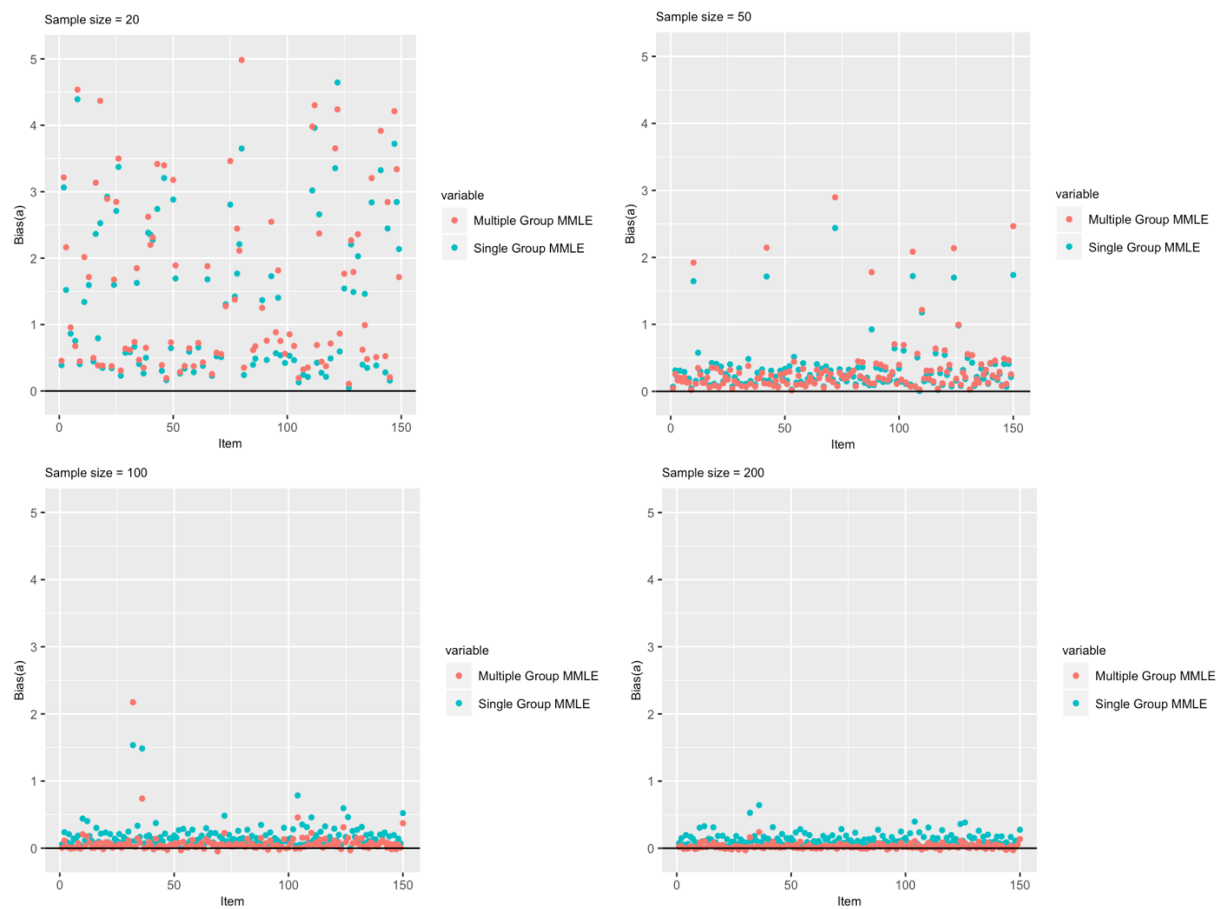
Sample size	Number of anchor items	Selection strategy of anchor items	Single-group MMLE	Multiple-group MMLE	Total
20	20	Uniformly	48	54	55
		Non-uniformly	50	50	52
	30	Uniformly	45	51	53
		Non-uniformly	41	47	47
	50	Uniformly	36	37	38
		Non-uniformly	31	32	32
50	20	Uniformly	3	3	3
		Non-uniformly	3	3	3
	30	Uniformly	3	4	4
		Non-uniformly	3	3	3
	50	Uniformly	3	4	4
		Non-uniformly	2	2	2
100	20	Uniformly	0	0	0
		Non-uniformly	0	1	1
	30	Uniformly	0	0	0
		Non-uniformly	0	1	1
	50	Uniformly	0	0	0
		Non-uniformly	0	1	1
200	20	Uniformly	0	0	0
		Non-uniformly	0	0	0
	30	Uniformly	0	0	0
		Non-uniformly	0	0	0
	50	Uniformly	0	0	0
		Non-uniformly	0	0	0

Appendix C2. *Five Number Summaries for Bias for a Parameters under 2-Form NEAT Design*

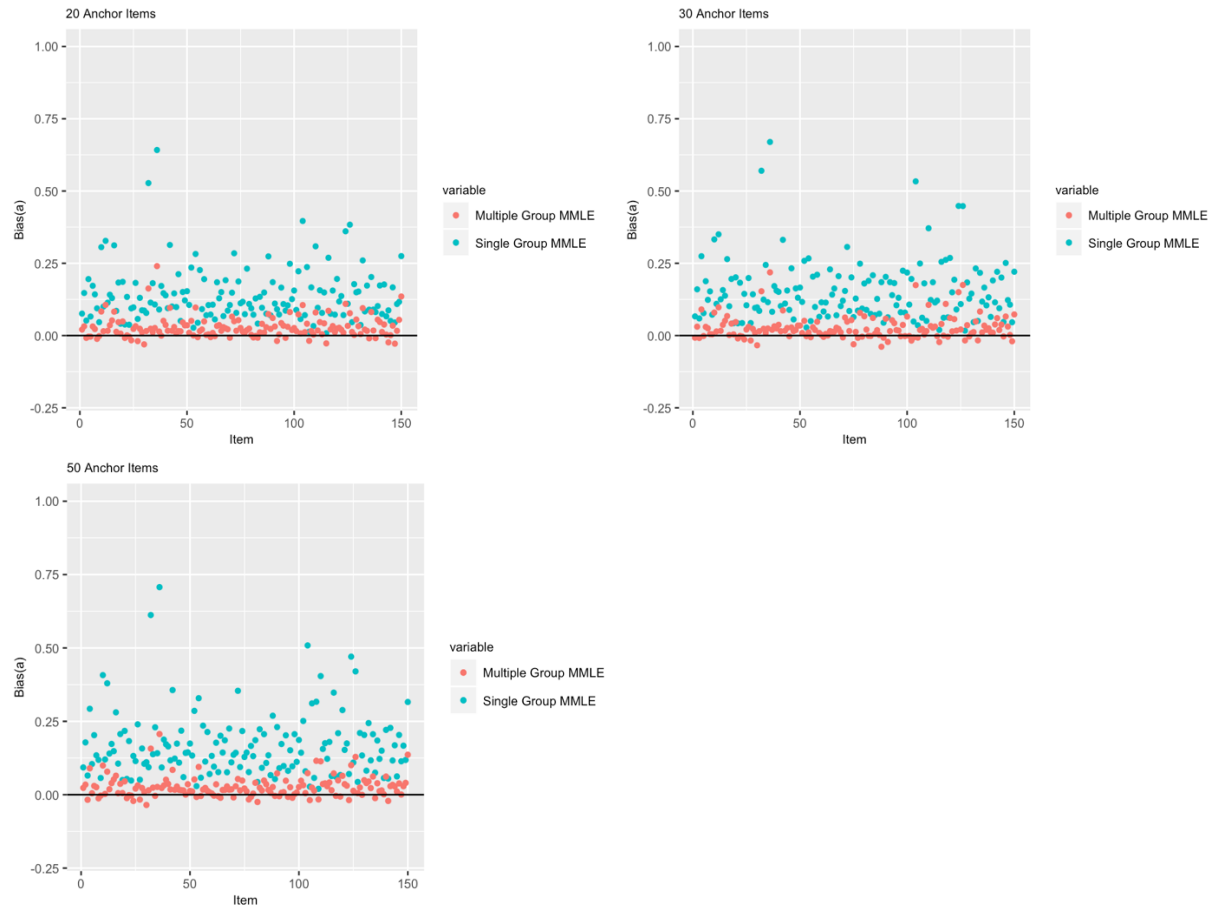
Estimation method	sample size	number of anchor items	selection strategy of anchor items	min	Q1	median	Q3	max
Single-group MMLE	20	20	uniformly	0.051	0.405	1.323	2.448	4.817
			non-uniformly	0.069	0.382	0.993	2.322	4.463
		30	uniformly	0.002	0.399	1.023	2.219	4.727
			non-uniformly	0.002	0.447	0.934	2.227	4.980
		50	uniformly	0.038	0.305	0.527	1.500	4.645
			non-uniformly	0.079	0.360	0.586	1.576	4.695
	50	20	uniformly	0.005	0.146	0.215	0.340	2.438
			non-uniformly	0.005	0.142	0.214	0.348	2.095
		30	uniformly	0.030	0.147	0.222	0.357	2.846
			non-uniformly	0.031	0.147	0.218	0.376	2.854
		50	uniformly	0.030	0.147	0.222	0.357	2.846
			non-uniformly	0.042	0.149	0.226	0.360	2.433
	100	20	uniformly	-0.005	0.094	0.144	0.217	1.535
			non-uniformly	-0.005	0.097	0.149	0.236	1.807
		30	uniformly	-0.005	0.096	0.152	0.240	1.516
			non-uniformly	-0.002	0.098	0.162	0.237	1.900
		50	uniformly	-0.005	0.096	0.152	0.240	1.516
			non-uniformly	0.014	0.106	0.169	0.239	1.662
	200	20	uniformly	0.008	0.079	0.113	0.173	0.642
			non-uniformly	0.007	0.082	0.118	0.186	0.674
		30	uniformly	0.017	0.085	0.128	0.199	0.670
			non-uniformly	0.015	0.086	0.129	0.205	0.707
		50	uniformly	0.017	0.085	0.128	0.199	0.670
			non-uniformly	0.021	0.096	0.142	0.208	0.715
Multiple-group MMLE	20	20	uniformly	0.106	0.489	0.874	2.496	4.984
			non-uniformly	0.133	0.469	1.066	2.554	4.777
		30	uniformly	0.033	0.473	1.010	2.323	4.892
			non-uniformly	0.055	0.499	1.019	2.367	4.604
		50	uniformly	0.093	0.368	0.652	1.801	4.668
			non-uniformly	0.115	0.409	0.649	1.746	4.897
	50	20	uniformly	0.016	0.120	0.201	0.320	2.897
			non-uniformly	0.015	0.123	0.209	0.332	2.666
		30	uniformly	0.011	0.117	0.177	0.289	2.887
			non-uniformly	0.017	0.120	0.185	0.321	2.785
		50	uniformly	0.011	0.117	0.177	0.289	2.887
			non-uniformly	0.007	0.090	0.131	0.240	3.157
	100	20	uniformly	-0.045	0.015	0.047	0.085	2.174
			non-uniformly	-0.030	0.018	0.048	0.087	2.115
		30	uniformly	-0.030	0.014	0.047	0.080	0.822

200	50	non-uniformly	-0.033	0.009	0.039	0.077	2.131
		uniformly	-0.030	0.014	0.047	0.080	0.822
	20	non-uniformly	-0.024	0.020	0.048	0.079	2.155
		uniformly	-0.030	0.005	0.020	0.041	0.240
	30	non-uniformly	-0.030	0.010	0.023	0.047	0.230
		uniformly	-0.039	0.000	0.018	0.037	0.219
	50	non-uniformly	-0.035	0.006	0.022	0.042	0.229
		uniformly	-0.039	0.000	0.018	0.037	0.219
		non-uniformly	-0.033	0.005	0.021	0.037	0.207

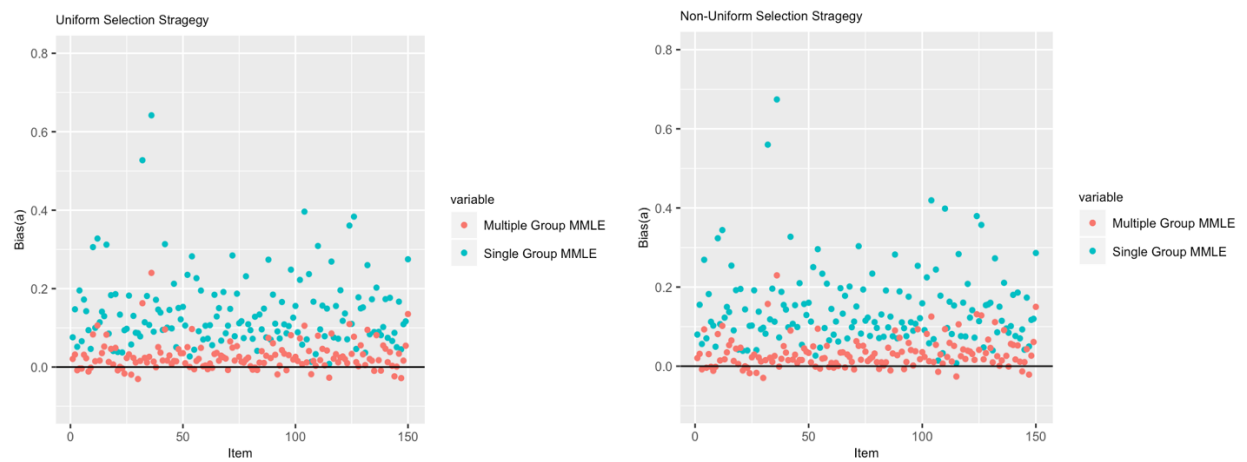
Appendix C3. Bias for a Parameters under 2-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes



Appendix C4. Bias for a Parameters under 2-Form NEAT Design with a Sample Size of 200 across Different Number of Uniformly Selected Anchor Items



Appendix C5. Bias for a Parameters under 2-Form NEAT Design with 20 Anchor Items and a Sample Size of 200 across Different Anchor Item Selection Strategies

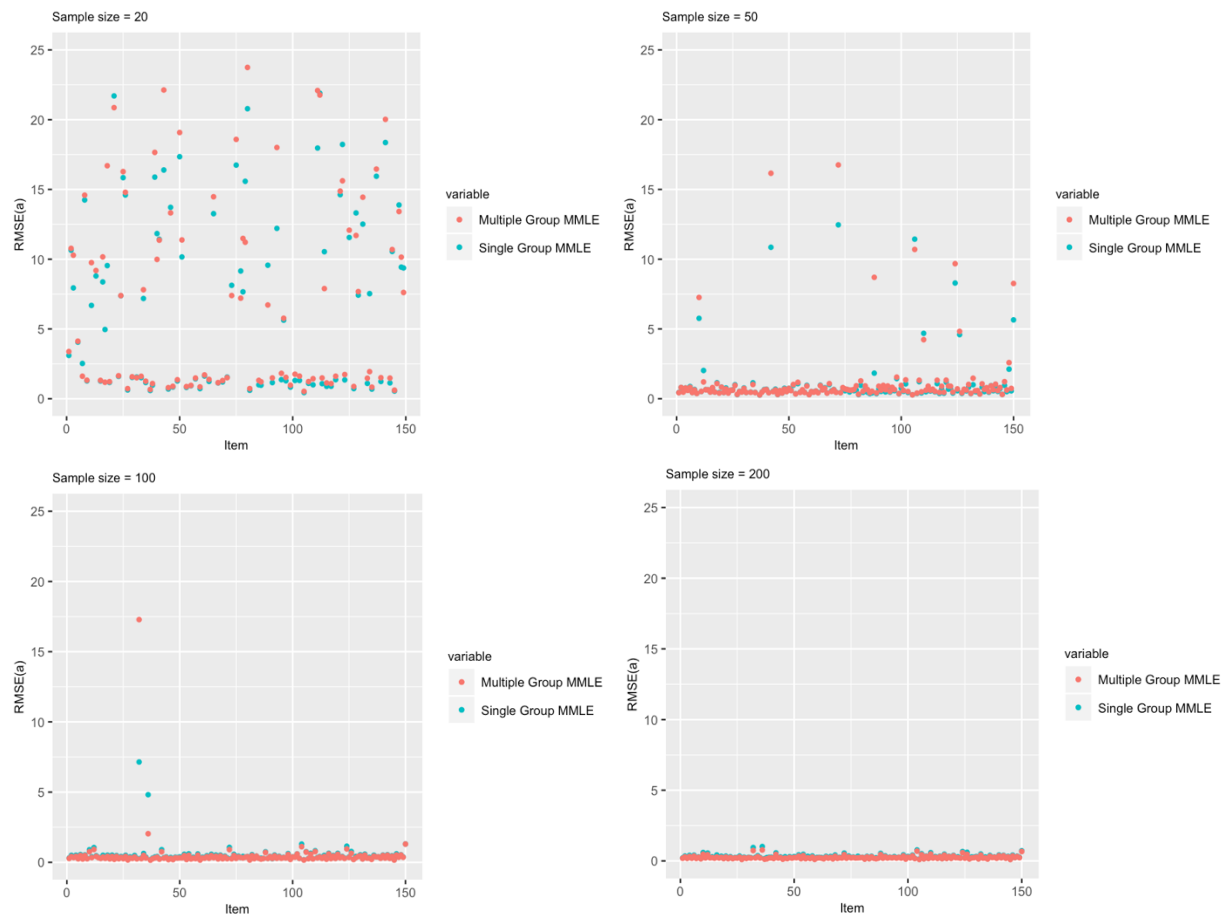


Appendix C6. *Five Number Summaries for RMSE for a Parameters under 2-Form NEAT Design*

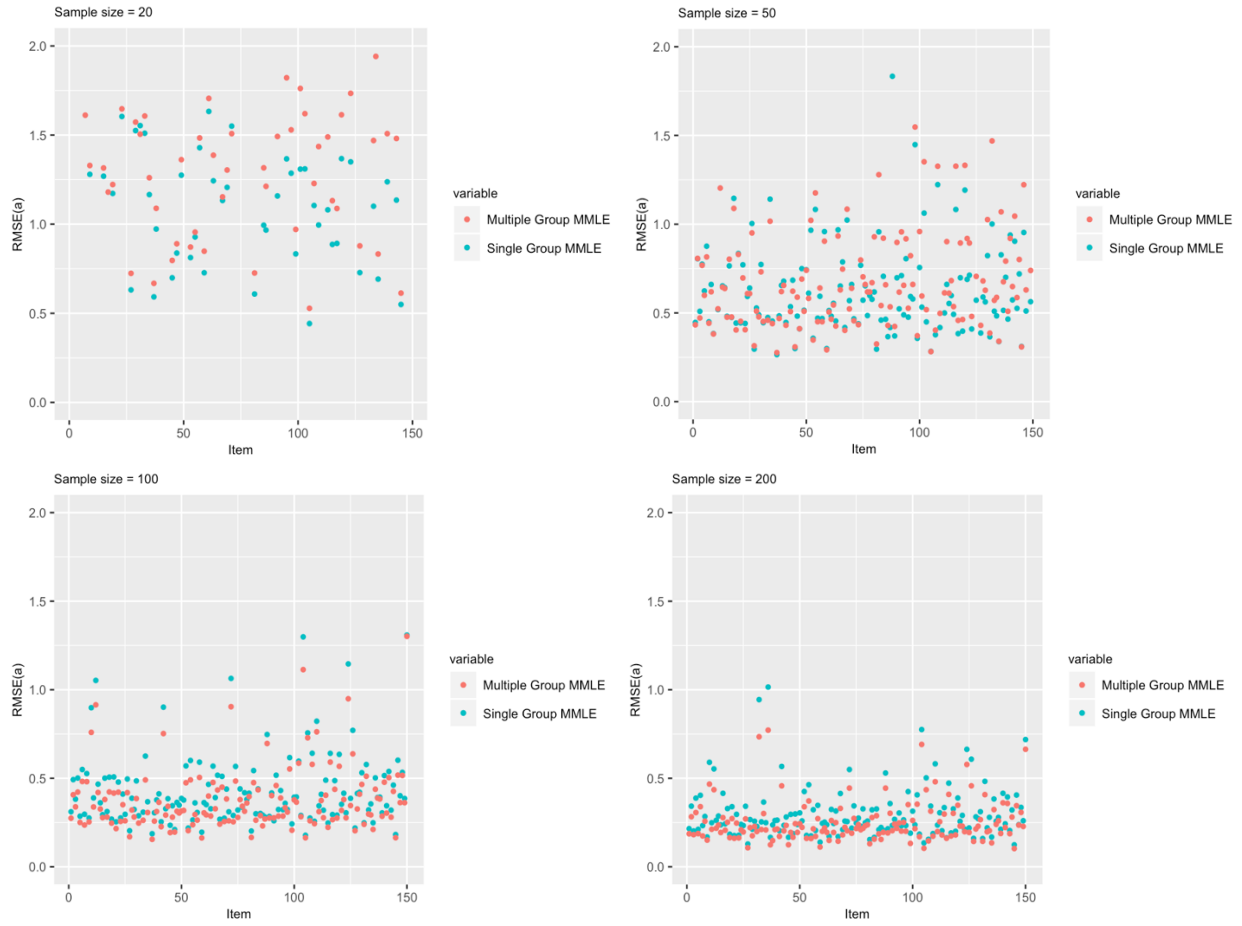
Estimation method	sample size	number of anchor items	selection strategy of anchor items	min	Q1	median	Q3	max
Single-group MMLE	20	20	uniformly	0.442	1.166	6.161	13.262	24.976
			non-uniformly	0.482	1.162	3.627	11.591	22.684
		30	uniformly	0.512	1.080	4.351	11.599	24.465
			non-uniformly	0.592	1.058	5.223	10.542	23.795
		50	uniformly	0.429	0.840	1.346	8.188	21.994
			non-uniformly	0.441	0.854	1.337	8.125	17.680
	50	20	uniformly	0.265	0.466	0.577	0.774	12.464
			non-uniformly	0.232	0.459	0.571	0.806	14.293
		30	uniformly	0.236	0.452	0.553	0.775	17.642
			non-uniformly	0.276	0.451	0.541	0.803	8.733
		50	uniformly	0.236	0.452	0.553	0.775	17.642
			non-uniformly	0.247	0.428	0.529	0.771	8.491
	100	20	uniformly	0.178	0.299	0.383	0.506	7.145
			non-uniformly	0.192	0.299	0.385	0.518	7.390
		30	uniformly	0.175	0.288	0.364	0.496	6.272
			non-uniformly	0.166	0.289	0.362	0.511	7.423
		50	uniformly	0.175	0.288	0.364	0.496	6.272
			non-uniformly	0.155	0.274	0.366	0.505	7.281
	200	20	uniformly	0.124	0.212	0.257	0.334	1.015
			non-uniformly	0.114	0.211	0.259	0.340	1.048
		30	uniformly	0.126	0.208	0.259	0.346	1.037
			non-uniformly	0.136	0.211	0.260	0.345	1.076
		50	uniformly	0.126	0.208	0.259	0.346	1.037
			non-uniformly	0.114	0.200	0.260	0.349	1.042
Multiple-group MMLE	20	20	uniformly	0.528	1.316	1.881	11.594	24.495
			non-uniformly	0.567	1.255	3.556	11.509	23.790
		30	uniformly	0.559	1.182	4.190	10.851	26.930
			non-uniformly	0.678	1.192	3.462	11.118	28.742
		50	uniformly	0.473	0.936	1.362	8.991	21.995
			non-uniformly	0.452	0.922	1.434	8.906	21.396
	50	20	uniformly	0.276	0.474	0.630	0.903	16.757
			non-uniformly	0.258	0.479	0.627	0.924	15.220
		30	uniformly	0.249	0.458	0.585	0.802	9.751
			non-uniformly	0.250	0.465	0.576	0.825	11.974
		50	uniformly	0.249	0.458	0.585	0.802	9.751
			non-uniformly	0.220	0.391	0.495	0.737	19.819
	100	20	uniformly	0.155	0.275	0.327	0.434	17.282
			non-uniformly	0.143	0.269	0.325	0.447	16.445
		30	uniformly	0.156	0.257	0.312	0.422	3.730

200	50	non-uniformly	0.140	0.249	0.308	0.423	16.708
		uniformly	0.156	0.257	0.312	0.422	3.730
	20	non-uniformly	0.137	0.235	0.311	0.414	16.739
		uniformly	0.103	0.186	0.217	0.278	0.772
	30	non-uniformly	0.086	0.182	0.217	0.282	0.771
		uniformly	0.098	0.172	0.210	0.269	0.778
	50	non-uniformly	0.107	0.176	0.213	0.269	0.773
		uniformly	0.098	0.172	0.210	0.269	0.778
		non-uniformly	0.091	0.163	0.202	0.271	0.728

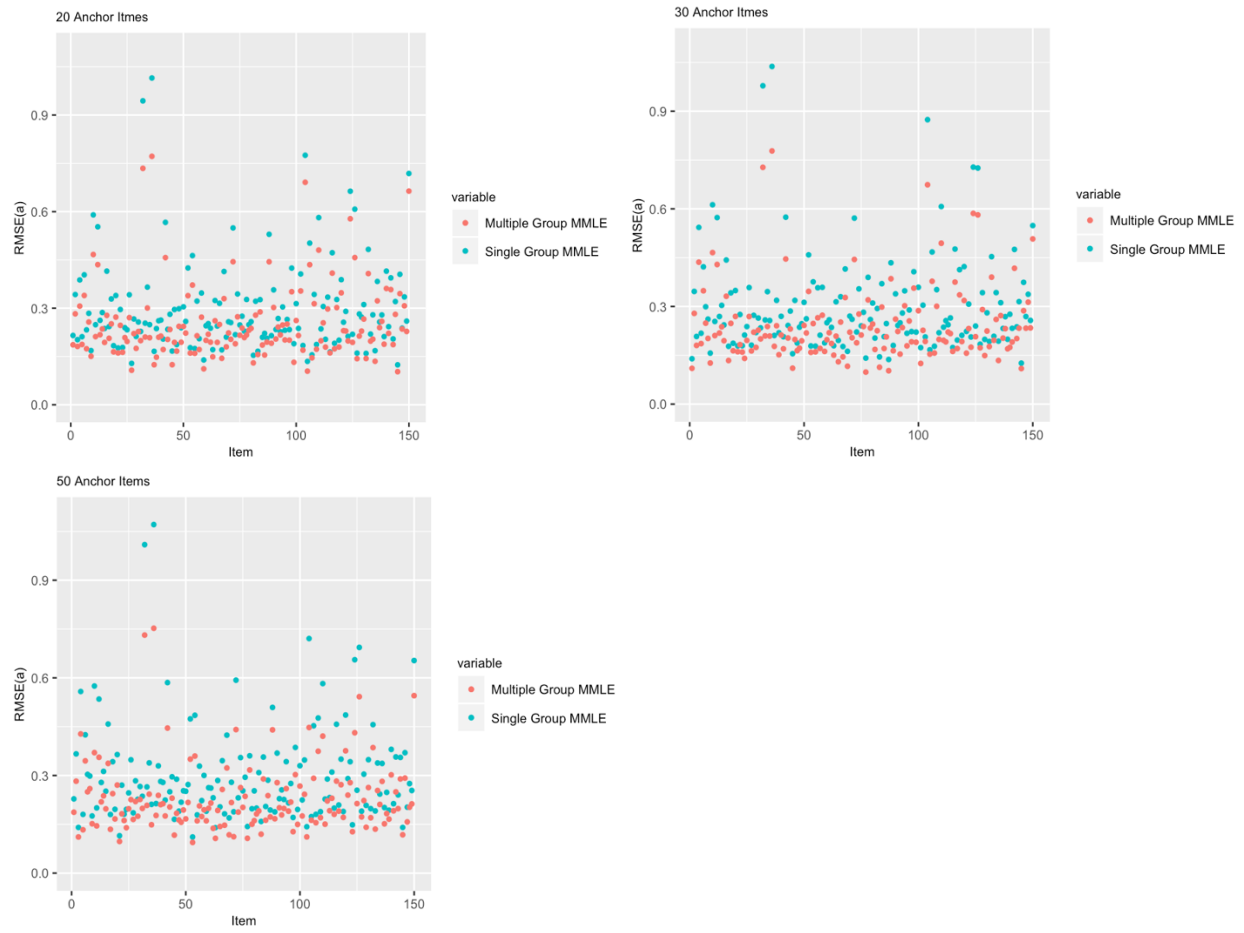
Appendix C7. RMSE for a Parameters under 5-Form NEAT Design with 20 Uniformly Selected Anchor Items across Different Sample Sizes



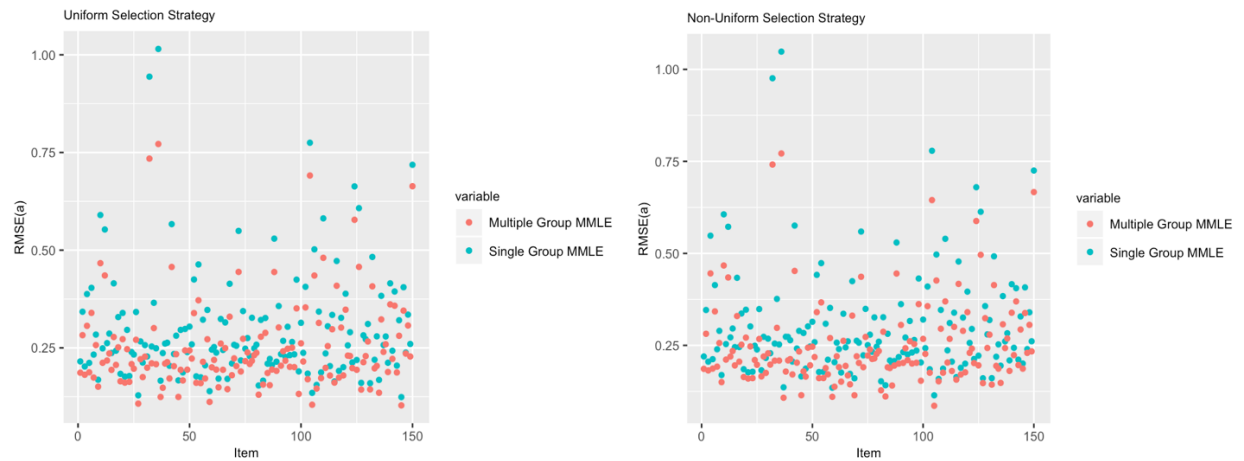
Appendix C8. *RMSE (from 0 to 2) for a Parameters under 5-Form NEAT with 20 Uniformly Selected Anchor Items across Different Sample Sizes*



Appendix C9. *RMSE for a Parameters under 2-Form NEAT Design NEAT with a Sample Size of 200 across Different Number of Uniformly Selected Anchor Items*



Appendix C10. *RMSE for a Parameters under 2-Form NEAT Design with 20 Anchor Items and a Sample Size of 200 across Different Anchor Item Selection Strategies*



Appendix C11. *Number of Extreme Estimation Values for b Parameters under 2-Form NEAT Design*

Sample size	Number of anchor items	Selection strategy of anchor items	Single-group MMLE	Multiple-group MMLE	Total
20	20	Uniformly	10	6	13
		Non-uniformly	8	7	13
	30	Uniformly	10	2	10
		Non-uniformly	3	5	7
	50	Uniformly	3	6	8
		Non-uniformly	7	4	8
50	20	Uniformly	2	4	4
		Non-uniformly	2	4	4
	30	Uniformly	3	3	5
		Non-uniformly	0	2	2
	50	Uniformly	3	3	5
		Non-uniformly	2	3	4
100	20	Uniformly	1	2	3
		Non-uniformly	0	1	1
	30	Uniformly	0	0	0
		Non-uniformly	1	1	1
	50	Uniformly	0	0	0
		Non-uniformly	1	1	1
200	20	Uniformly	0	1	1
		Non-uniformly	0	1	1
	30	Uniformly	0	0	0
		Non-uniformly	1	1	1
	50	Uniformly	0	0	0
		Non-uniformly	0	0	0

Appendix C12. *Five Number Summaries for Bias for b Parameters under 2-Form NEAT Design*

Estimation method	sample size	number of anchor items	selection strategy of anchor items	min	Q1	median	Q3	max
Single-group MMLE	20	20	uniformly	-4.716	-0.405	-0.129	0.165	3.671
			non-uniformly	-5.957	-0.384	-0.102	0.168	2.985
		30	uniformly	-2.188	-0.363	-0.146	0.087	4.688
			non-uniformly	-4.703	-0.373	-0.091	0.130	5.896
		50	uniformly	-5.084	-0.209	-0.044	0.164	2.771
			non-uniformly	-4.747	-0.192	-0.057	0.106	4.634
	50	20	uniformly	-3.669	-0.139	-0.006	0.116	1.876
			non-uniformly	-3.634	-0.121	0.004	0.122	3.548
		30	uniformly	-1.303	-0.129	-0.018	0.090	2.048
			non-uniformly	-4.364	-0.130	-0.017	0.079	3.957
		50	uniformly	-1.303	-0.129	-0.018	0.090	2.048
			non-uniformly	-2.901	-0.094	-0.003	0.071	3.421
	100	20	uniformly	-0.929	-0.127	0.013	0.113	6.758
			non-uniformly	-3.651	-0.127	0.012	0.110	1.649
		30	uniformly	-2.193	-0.118	-0.016	0.090	0.923
			non-uniformly	-1.752	-0.108	-0.012	0.089	1.225
		50	uniformly	-2.193	-0.118	-0.016	0.090	0.923
			non-uniformly	-1.580	-0.082	-0.012	0.063	4.777
	200	20	uniformly	-1.668	-0.128	0.008	0.120	2.628
			non-uniformly	-1.313	-0.122	0.012	0.109	0.509
		30	uniformly	-1.166	-0.122	0.003	0.095	0.299
			non-uniformly	-0.620	-0.119	-0.002	0.088	0.202
		50	uniformly	-1.166	-0.122	0.003	0.095	0.299
			non-uniformly	-0.465	-0.094	-0.012	0.076	0.847
Multiple-group MMLE	20	20	uniformly	-2.564	-0.375	-0.087	0.193	4.236
			non-uniformly	-6.107	-0.391	-0.112	0.229	5.017
		30	uniformly	-3.619	-0.343	-0.098	0.176	4.740
			non-uniformly	-4.777	-0.366	-0.099	0.143	4.026
		50	uniformly	-4.885	-0.228	-0.036	0.143	3.890
			non-uniformly	-2.794	-0.160	-0.003	0.159	3.649
	50	20	uniformly	-2.495	0.338	0.398	0.442	2.669
			non-uniformly	-5.790	0.329	0.389	0.430	1.443
		30	uniformly	-4.220	0.350	0.389	0.426	4.521
			non-uniformly	-1.269	0.364	0.398	0.437	3.612
		50	uniformly	-4.220	0.350	0.389	0.426	4.521
			non-uniformly	-3.449	0.402	0.432	0.463	2.202
	100	20	uniformly	-1.832	0.470	0.497	0.524	1.046
			non-uniformly	-1.223	0.475	0.494	0.521	3.000
		30	uniformly	-0.511	0.470	0.499	0.525	1.751

200	50	non-uniformly	-1.659	0.472	0.498	0.533	2.070
		uniformly	-0.511	0.470	0.499	0.525	1.751
	20	non-uniformly	-2.353	0.473	0.494	0.513	4.614
		uniformly	-0.018	0.495	0.511	0.530	0.688
	30	non-uniformly	-3.047	0.496	0.512	0.531	1.464
		uniformly	-0.367	0.494	0.512	0.530	1.015
	50	non-uniformly	-0.158	0.490	0.508	0.524	1.045
		uniformly	-0.367	-0.367	0.512	0.530	1.015
		non-uniformly	0.070	0.497	0.509	0.522	1.743

Appendix C13. *Five Number Summaries for RMSE for b Parameters under 2-Form NEAT Design*

Estimation method	sample size	number of anchor items	selection strategy of anchor items	min	Q1	median	Q3	max
Single-group MMLE	20	20	uniformly	0.235	0.645	1.851	4.117	40.926
			non-uniformly	0.220	0.726	2.263	5.885	60.228
		30	uniformly	0.175	0.514	1.682	4.087	33.309
			non-uniformly	0.212	0.507	1.615	4.740	72.091
		50	uniformly	0.208	0.433	1.516	4.651	89.040
			non-uniformly	0.179	0.425	1.580	5.627	64.060
	50	20	uniformly	0.154	0.309	0.453	1.492	35.809
			non-uniformly	0.163	0.310	0.450	1.532	92.327
		30	uniformly	0.155	0.283	0.412	1.128	29.867
			non-uniformly	0.129	0.263	0.408	1.191	37.623
		50	uniformly	0.155	0.283	0.412	1.128	29.867
			non-uniformly	0.153	0.251	0.362	0.988	134.713
	100	20	uniformly	0.105	0.211	0.281	0.430	71.741
			non-uniformly	0.108	0.210	0.273	0.431	34.869
		30	uniformly	0.097	0.199	0.266	0.424	25.780
			non-uniformly	0.106	0.193	0.260	0.408	13.191
		50	uniformly	0.097	0.199	0.266	0.424	25.780
			non-uniformly	0.100	0.174	0.235	0.415	48.296
	200	20	uniformly	0.084	0.173	0.212	0.283	28.873
			non-uniformly	0.085	0.165	0.206	0.283	13.643
		30	uniformly	0.075	0.156	0.202	0.276	11.282
			non-uniformly	0.077	0.149	0.195	0.268	5.630
		50	uniformly	0.075	0.156	0.202	0.276	11.282
			non-uniformly	0.072	0.129	0.185	0.269	6.260
Multiple-group MMLE	20	20	uniformly	0.223	0.612	1.834	4.966	53.846
			non-uniformly	0.218	0.609	2.449	7.502	57.896
		30	uniformly	0.175	0.534	1.956	4.542	49.426
			non-uniformly	0.203	0.501	1.461	4.069	41.033
		50	uniformly	0.204	0.435	1.216	3.723	89.049
			non-uniformly	0.199	0.436	1.565	5.324	37.867
	50	20	uniformly	0.461	0.558	0.675	1.412	24.582
			non-uniformly	0.455	0.549	0.688	1.954	94.084
		30	uniformly	0.446	0.534	0.638	1.434	62.769
			non-uniformly	0.449	0.523	0.620	1.317	53.272
		50	uniformly	0.446	0.534	0.638	1.434	62.769
			non-uniformly	0.448	0.521	0.620	1.047	134.717
	100	20	uniformly	0.498	0.543	0.581	0.679	24.852
			non-uniformly	0.502	0.540	0.574	0.690	69.436

200	30	uniformly	0.492	0.542	0.575	0.704	11.906
		non-uniformly	0.503	0.537	0.578	0.690	13.222
	50	uniformly	0.492	0.542	0.575	0.704	11.906
		non-uniformly	0.493	0.529	0.563	0.686	41.140
	20	uniformly	0.497	0.529	0.544	0.587	7.577
		non-uniformly	0.498	0.529	0.543	0.591	38.409
	30	uniformly	0.499	0.530	0.545	0.592	9.642
		non-uniformly	0.498	0.527	0.541	0.582	8.473
	50	uniformly	0.499	0.530	0.545	0.592	9.642
		non-uniformly	0.498	0.525	0.541	0.577	8.323

Appendix D1. *Number of Extreme Estimation Values for a Parameters under 5-Form Multiform Design*

Sample size	Number of anchor items	Selection strategy of anchor items	Single-group MMLE	Multiple-group MMLE	Total
20	20	Uniformly	7	8	10
		Non-uniformly	4	3	4
	30	Uniformly	5	4	5
		Non-uniformly	5	4	5
	50	Uniformly	5	4	5
		Non-uniformly	8	6	8
50	20	Uniformly	1	0	1
		Non-uniformly	1	0	1
	30	Uniformly	1	0	1
		Non-uniformly	1	0	1
	50	Uniformly	0	0	0
		Non-uniformly	1	0	1
100	20	Uniformly	0	0	0
		Non-uniformly	0	0	0
	30	Uniformly	0	0	0
		Non-uniformly	0	0	0
	50	Uniformly	0	0	0
		Non-uniformly	0	0	0

Appendix D2. *Five Number Summaries for Bias for a Parameters under 5-Form Multiform Design*

Estimation method	sample size	number of anchor items	selection strategy of anchor items	min	Q1	median	Q3	max
Single-group MMLE	20	20	uniformly	0.058	0.278	0.408	0.602	3.696
			non-uniformly	0.079	0.263	0.404	0.699	3.156
		30	uniformly	0.097	0.246	0.395	0.660	3.296
			non-uniformly	0.085	0.262	0.385	0.586	3.033
		50	uniformly	0.080	0.254	0.369	0.613	3.670
			non-uniformly	0.113	0.230	0.383	0.550	4.110
	50	20	uniformly	0.053	0.190	0.287	0.402	1.361
			non-uniformly	0.079	0.193	0.282	0.417	1.095
		30	uniformly	0.068	0.195	0.275	0.401	1.988
			non-uniformly	0.065	0.191	0.282	0.403	1.281
		50	uniformly	0.071	0.191	0.282	0.391	2.218
			non-uniformly	0.047	0.186	0.285	0.412	1.262
	100	20	uniformly	0.044	0.165	0.258	0.361	1.251
			non-uniformly	0.054	0.175	0.267	0.361	1.442
		30	uniformly	0.056	0.178	0.254	0.365	1.430
			non-uniformly	0.034	0.185	0.256	0.373	1.370
		50	uniformly	0.037	0.173	0.271	0.363	1.304
			non-uniformly	0.031	0.180	0.261	0.363	1.214
Multiple-group MMLE	20	20	uniformly	-0.010	0.116	0.185	0.314	4.149
			non-uniformly	0.001	0.118	0.211	0.443	2.980
		30	uniformly	-0.027	0.090	0.162	0.345	3.342
			non-uniformly	0.005	0.085	0.156	0.271	2.375
		50	uniformly	-0.009	0.075	0.140	0.254	3.700
			non-uniformly	0.014	0.097	0.162	0.277	4.072
	50	20	uniformly	-0.049	-0.001	0.020	0.050	0.771
			non-uniformly	-0.032	0.002	0.028	0.053	1.037
		30	uniformly	-0.066	-0.001	0.020	0.044	1.132
			non-uniformly	-0.056	-0.004	0.013	0.042	0.903
		50	uniformly	-0.089	-0.006	0.005	0.028	1.118
			non-uniformly	-0.041	-0.007	0.015	0.038	1.129
	100	20	uniformly	-0.039	0.004	0.022	0.053	0.252
			non-uniformly	-0.031	0.003	0.017	0.037	0.394
		30	uniformly	-0.030	0.003	0.017	0.038	0.376
			non-uniformly	-0.043	0.000	0.012	0.034	0.314
		50	uniformly	-0.026	0.001	0.014	0.035	0.317
			non-uniformly	-0.035	-0.002	0.011	0.029	0.175

Appendix D3. *Five Number Summaries for RMSE for a Parameters under 5-Form Multiform Design*

Estimation method	sample size	number of anchor items	selection strategy of anchor items	min	Q1	median	Q3	max
Single-group MMLE	20	20	uniformly	0.245	0.610	0.765	1.178	20.181
			non-uniformly	0.254	0.588	0.770	1.253	15.827
		30	uniformly	0.255	0.555	0.728	1.232	13.830
			non-uniformly	0.252	0.554	0.737	1.112	15.678
		50	uniformly	0.239	0.510	0.696	1.049	20.458
			non-uniformly	0.283	0.505	0.687	0.966	13.944
	50	20	uniformly	0.158	0.364	0.448	0.622	2.206
			non-uniformly	0.158	0.363	0.446	0.647	1.565
		30	uniformly	0.163	0.348	0.438	0.594	6.562
			non-uniformly	0.169	0.341	0.439	0.617	2.383
		50	uniformly	0.171	0.324	0.426	0.578	6.393
			non-uniformly	0.176	0.318	0.426	0.574	1.803
	100	20	uniformly	0.114	0.270	0.350	0.485	1.679
			non-uniformly	0.117	0.273	0.355	0.472	1.993
		30	uniformly	0.136	0.271	0.348	0.463	1.894
			non-uniformly	0.149	0.263	0.348	0.465	1.811
		50	uniformly	0.129	0.252	0.345	0.460	1.667
			non-uniformly	0.132	0.256	0.352	0.465	1.613
Multiple-group MMLE	20	20	uniformly	0.213	0.515	0.643	0.943	21.810
			non-uniformly	0.213	0.500	0.659	1.143	16.464
		30	uniformly	0.227	0.475	0.604	0.990	19.370
			non-uniformly	0.214	0.455	0.616	0.911	15.650
		50	uniformly	0.192	0.408	0.557	0.809	23.736
			non-uniformly	0.230	0.436	0.580	0.859	15.825
	50	20	uniformly	0.109	0.246	0.309	0.416	6.024
			non-uniformly	0.103	0.242	0.312	0.426	7.217
		30	uniformly	0.117	0.247	0.291	0.388	7.384
			non-uniformly	0.123	0.234	0.295	0.400	5.580
		50	uniformly	0.115	0.213	0.271	0.361	6.650
			non-uniformly	0.121	0.219	0.276	0.362	7.923
	100	20	uniformly	0.075	0.175	0.210	0.287	0.929
			non-uniformly	0.078	0.174	0.205	0.285	1.311
		30	uniformly	0.080	0.167	0.200	0.270	1.084
			non-uniformly	0.090	0.163	0.207	0.265	1.003
		50	uniformly	0.082	0.152	0.197	0.261	0.958
			non-uniformly	0.083	0.157	0.196	0.254	0.845

Appendix D4. *Number of Extreme Estimation Values for b Parameters under 5-Form Multiform Design*

Sample size	Number of anchor items	Selection strategy of anchor items	Single-group MMLE	Multiple-group MMLE	Total
20	20	Uniformly	2	1	3
		Non-uniformly	2	1	3
	30	Uniformly	0	1	1
		Non-uniformly	3	2	4
	50	Uniformly	1	0	1
		Non-uniformly	1	1	2
50	20	Uniformly	0	0	0
		Non-uniformly	1	0	1
	30	Uniformly	1	0	1
		Non-uniformly	1	0	1
	50	Uniformly	2	0	2
		Non-uniformly	0	0	0
100	20	Uniformly	0	0	0
		Non-uniformly	0	0	0
	30	Uniformly	0	0	0
		Non-uniformly	0	0	0
	50	Uniformly	0	0	0
		Non-uniformly	0	0	0

Appendix D5. *Five Number Summaries for Bias for b Parameters under 5-Form Multiform Design*

Estimation method	sample size	number of anchor items	selection strategy of anchor items	min	Q1	median	Q3	max
Single-group MMLE	20	20	uniformly	-3.373	-0.189	-0.023	0.105	1.842
			non-uniformly	-1.583	-0.187	-0.036	0.102	3.315
		30	uniformly	-2.929	-0.174	-0.036	0.101	2.070
			non-uniformly	-3.300	-0.200	-0.052	0.101	1.220
		50	uniformly	-5.140	-0.168	-0.026	0.110	4.053
			non-uniformly	-4.961	-0.180	-0.036	0.097	2.282
	50	20	uniformly	-0.465	-0.155	-0.038	0.121	1.280
			non-uniformly	-0.876	-0.173	-0.040	0.113	3.203
		30	uniformly	-1.735	-0.164	-0.030	0.115	2.607
			non-uniformly	-0.472	-0.168	-0.026	0.109	0.345
		50	uniformly	-0.847	-0.161	-0.045	0.103	0.328
			non-uniformly	-3.302	-0.153	-0.046	0.106	2.377
	100	20	uniformly	-0.485	-0.159	-0.035	0.117	0.339
			non-uniformly	-0.471	-0.163	-0.029	0.110	0.403
		30	uniformly	-0.748	-0.156	-0.032	0.112	0.340
			non-uniformly	-0.493	-0.160	-0.025	0.120	0.459
		50	uniformly	-0.453	-0.153	-0.029	0.104	0.395
			non-uniformly	-0.556	-0.153	-0.028	0.100	0.328
Multiple-group MMLE	20	20	uniformly	-3.531	-0.095	-0.047	0.003	3.399
			non-uniformly	-2.823	-0.097	-0.053	-0.005	2.346
		30	uniformly	-3.211	-0.060	-0.030	0.002	2.507
			non-uniformly	-2.933	-0.099	-0.062	-0.013	4.358
		50	uniformly	-2.388	-0.085	-0.040	-0.007	3.258
			non-uniformly	-3.000	-0.086	-0.049	-0.010	2.640
	50	20	uniformly	-0.923	-0.034	-0.002	0.034	5.781
			non-uniformly	-0.893	-0.031	-0.003	0.025	1.655
		30	uniformly	-1.185	-0.040	-0.006	0.022	2.975
			non-uniformly	-1.110	-0.037	0.004	0.032	2.613
		50	uniformly	-0.890	-0.042	-0.005	0.023	3.257
			non-uniformly	-3.071	-0.035	-0.004	0.028	1.048
	100	20	uniformly	-0.483	-0.007	0.004	0.019	0.286
			non-uniformly	-0.440	-0.010	0.006	0.023	0.315
		30	uniformly	-1.078	-0.011	0.004	0.018	0.542
			non-uniformly	-0.221	-0.009	0.007	0.021	0.644
		50	uniformly	-0.251	-0.004	0.006	0.020	0.329
			non-uniformly	-3.021	-0.011	0.001	0.015	0.473

Appendix D6. *Five Number Summaries for RMSE for b Parameters under 5-Form Multiform Design*

Estimation method	sample size	number of anchor items	selection strategy of anchor items	min	Q1	median	Q3	max
Single-group MMLE	20	20	uniformly	0.157	0.296	0.445	1.179	28.414
			non-uniformly	0.126	0.302	0.477	1.204	35.313
		30	uniformly	0.126	0.275	0.423	0.959	30.038
			non-uniformly	0.141	0.286	0.378	1.235	88.202
		50	uniformly	0.127	0.262	0.369	1.026	58.911
			non-uniformly	0.126	0.253	0.374	0.727	46.137
	50	20	uniformly	0.097	0.203	0.277	0.422	14.767
			non-uniformly	0.090	0.201	0.271	0.377	32.395
		30	uniformly	0.081	0.193	0.262	0.385	27.745
			non-uniformly	0.085	0.187	0.267	0.385	4.245
		50	uniformly	0.082	0.185	0.249	0.377	5.484
			non-uniformly	0.083	0.183	0.256	0.372	24.219
	100	20	uniformly	0.078	0.165	0.219	0.302	2.176
			non-uniformly	0.060	0.158	0.214	0.311	3.804
		30	uniformly	0.068	0.150	0.213	0.306	6.235
			non-uniformly	0.064	0.152	0.206	0.307	6.983
		50	uniformly	0.062	0.141	0.199	0.308	2.589
			non-uniformly	0.062	0.135	0.202	0.304	3.231
Multiple-group MMLE	20	20	uniformly	0.273	0.398	0.571	1.307	33.698
			non-uniformly	0.239	0.397	0.589	1.712	28.407
		30	uniformly	0.226	0.375	0.552	1.093	32.830
			non-uniformly	0.246	0.368	0.501	1.588	88.044
		50	uniformly	0.244	0.372	0.501	1.528	34.845
			non-uniformly	0.224	0.339	0.473	0.882	28.575
	50	20	uniformly	0.159	0.243	0.313	0.462	59.484
			non-uniformly	0.164	0.234	0.301	0.462	13.373
		30	uniformly	0.140	0.226	0.284	0.427	22.121
			non-uniformly	0.152	0.232	0.294	0.438	20.588
		50	uniformly	0.146	0.217	0.275	0.424	33.551
			non-uniformly	0.157	0.219	0.276	0.398	30.272
	100	20	uniformly	0.111	0.161	0.206	0.280	3.800
			non-uniformly	0.115	0.158	0.207	0.268	4.843
		30	uniformly	0.113	0.154	0.199	0.262	11.256
			non-uniformly	0.113	0.152	0.193	0.270	6.266
		50	uniformly	0.109	0.153	0.185	0.276	2.777
			non-uniformly	0.112	0.151	0.186	0.252	29.046

Appendix E1

Number of Extreme Estimation Values for a Parameters under 5-Form Panel Missing Design

	Single-group MMLE	Multiple-group MMLE	Total
20	2	1	2
50	0	0	0
100	0	0	0

Appendix E2

Five Number Summaries for bias for a Parameters under 5-Form Panel Missing Design

Estimation method	Sample size	min	Q1	median	Q3	max
Single-group MMLE	20	0.079	0.234	0.324	0.504	1.899
	50	0.068	0.182	0.268	0.383	1.499
	100	0.070	0.173	0.253	0.353	1.039
Multiple-group MMLE	20	-0.006	0.046	0.078	0.144	1.598
	50	-0.039	0.006	0.022	0.040	0.426
	100	-0.037	-0.006	0.001	0.012	0.144

Appendix E3

Five Number Summaries for RMSE for a Parameters under 5-Form Panel Missing Design

Estimation method	Sample size	min	Q1	median	Q3	max
Single-group MMLE	20	0.241	0.427	0.538	0.787	5.929
	50	0.168	0.272	0.365	0.505	2.215
	100	0.130	0.227	0.311	0.409	1.262
Multiple-group MMLE	20	0.192	0.336	0.421	0.611	8.824
	50	0.133	0.186	0.241	0.342	1.435
	100	0.087	0.126	0.154	0.208	0.689

Appendix E4

Number of Extreme Estimation Values for b Parameters under 5-Form Panel Missing Design

	Single-group MMLE	Multiple-group MMLE	Total
20	0	1	1
50	0	0	0
100	0	0	0

Appendix E5

Five Number Summaries for Bias for b Parameters under 5-Form Panel Missing Design

Estimation method	Sample size	min	Q1	median	Q3	max
Single-group MMLE	20	-3.413	-0.158	-0.026	0.068	4.607
	50	-0.427	-0.139	-0.040	0.103	0.312
	100	-0.498	-0.153	-0.038	0.101	0.367
Multiple-group MMLE	20	-1.138	-0.071	-0.028	0.010	1.759
	50	-0.550	-0.028	-0.007	0.019	0.572
	100	-0.121	-0.026	-0.009	0.007	0.158

Appendix E6

Five Number Summaries for RMSE for b Parameters under 5-Form Panel Missing Design

Estimation method	Sample size	min	Q1	median	Q3	max
Single-group MMLE	20	0.128	0.226	0.312	0.462	34.122
	50	0.092	0.159	0.215	0.308	5.008
	100	0.057	0.131	0.177	0.267	0.602
Multiple-group MMLE	20	0.236	0.325	0.434	0.616	24.370
	50	0.145	0.202	0.259	0.344	4.710
	100	0.105	0.140	0.168	0.229	0.620

Appendix F1

Number of Extreme Estimation Values for a Parameters under 5-Form Panel Chained Design

	Single-group MMLE	Multiple-group MMLE	Total
20	7	9	9
50	0	0	0
100	0	0	0

Appendix F2

Five Number Summaries for Bias for a Parameters under 5-Form Panel Chained Design

Estimation method	Sample size	min	Q1	median	Q3	max
Single-group MMLE	20	0.006	0.212	0.338	0.565	3.617
	50	0.039	0.130	0.224	0.343	1.359
	100	0.040	0.127	0.189	0.279	1.259
Multiple-group MMLE	20	-0.001	0.201	0.314	0.559	3.945
	50	-0.047	0.012	0.040	0.068	1.391
	100	-0.029	0.004	0.020	0.036	0.499

Appendix F3

Five Number Summaries for RMSE for a Parameters under 5-Form Panel Chained Design

Estimation method	Sample size	min	Q1	median	Q3	max
Single-group MMLE	20	0.403	0.601	0.771	1.148	18.583
	50	0.250	0.337	0.432	0.593	2.612
	100	0.159	0.246	0.313	0.417	2.223
Multiple-group MMLE	20	0.399	0.645	0.853	1.307	19.424
	50	0.200	0.280	0.347	0.444	7.571
	100	0.143	0.194	0.229	0.293	1.449

Appendix F4

Number of Extreme Estimation Values for b Parameters under 5-Form Panel Chained Design

	Single-group MMLE	Multiple-group MMLE	Total
20	2	0	2
50	0	3	3
100	0	0	0

Appendix F5

Five Number Summaries for Bias for b Parameters under 5-Form Panel Chained Design

Estimation method	Sample size	min	Q1	median	Q3	max
Single-group MMLE	20	-4.400	-0.202	-0.016	0.165	5.301
	50	-1.457	-0.192	-0.050	0.141	3.892
	100	-0.479	-0.191	-0.017	0.223	0.647
Multiple-group MMLE	20	-6.105	-0.101	-0.023	0.050	2.845
	50	-0.986	-0.027	0.000	0.031	0.429
	100	-0.344	-0.014	0.003	0.022	0.576

Appendix F6

Five Number Summaries for RMSE for b Parameters under 5-Form Panel Chained Design

Estimation method	Sample size	min	Q1	median	Q3	max
Single-group MMLE	20	0.178	0.367	0.572	1.730	53.454
	50	0.114	0.247	0.353	0.505	38.425
	100	0.090	0.186	0.304	0.388	4.648
Multiple-group MMLE	20	0.241	0.424	0.647	1.929	69.797
	50	0.164	0.243	0.313	0.469	10.809
	100	0.123	0.164	0.210	0.282	4.634

Appendix G1

Number of Extreme Estimation Values for a Parameters under 5-Form BIB Design

	Single-group MMLE	Multiple-group MMLE	Total
20	8	8	9
50	1	0	1
100	0	0	0

Appendix G2

Five Number Summaries for Bias for a Parameters under 5-Form BIB Design

Estimation method	Sample size	min	Q1	median	Q3	max
Single-group MMLE	20	-2.238	0.199	0.366	0.726	3.794
	50	-2.349	0.122	0.243	0.457	4.690
	100	-2.362	0.100	0.214	0.386	3.539
Multiple-group MMLE	20	-2.499	0.128	0.265	0.600	3.634
	50	-2.718	0.007	0.046	0.140	5.029
	100	-2.720	-0.008	0.021	0.076	3.076

Appendix G3

Five Number Summaries for RMSE for a Parameters under 5-Form BIB Design

Estimation method	Sample size	min	Q1	median	Q3	max
Single-group MMLE	20	0.355	0.612	0.832	1.392	16.418
	50	0.221	0.361	0.484	0.690	10.139
	100	0.151	0.260	0.357	0.530	4.196
Multiple-group MMLE	20	0.328	0.601	0.843	1.753	20.486
	50	0.193	0.303	0.390	0.532	11.173
	100	0.132	0.211	0.278	0.399	3.444

Appendix G4

Number of Extreme Estimation Values for b Parameters under 5-Form BIB Design

	Single-group MMLE	Multiple-group MMLE	Total
20	2	0	2
50	0	0	0
100	0	0	0

Appendix G5

Five Number Summaries for Bias for b Parameters under 5-Form BIB Design

Estimation method	Sample size	min	Q1	median	Q3	max
Single-group MMLE	20	-2.349	-0.390	0.057	0.446	4.590
	50	-3.508	-0.351	0.013	0.381	2.327
	100	-0.827	-0.357	0.016	0.358	0.642
Multiple-group MMLE	20	-2.099	-0.402	-0.068	0.388	4.391
	50	-2.997	-0.222	0.011	0.283	1.235
	100	-1.281	-0.061	0.008	0.118	0.676

Appendix G6

Five Number Summaries for RMSE for b Parameters under 5-Form BIB Design

Estimation method	Sample size	min	Q1	median	Q3	max
Single-group MMLE	20	0.271	0.538	0.745	2.061	43.910
	50	0.149	0.364	0.534	0.682	31.983
	100	0.108	0.301	0.461	0.592	9.382
Multiple-group MMLE	20	0.319	0.602	0.873	2.826	46.563
	50	0.205	0.377	0.594	0.721	32.241
	100	0.143	0.235	0.458	0.602	7.165