

**Cognitive Diagnosis for Multiple-Choice Responses:
Nonparametric Classification Method, Q-Matrix Theory,
and Computerized Adaptive Testing**

A THESIS

**SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA**

BY

Yu Wang

**IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY IN EDUCATIONAL PSYCHOLOGY**

**Chia-Yi Chiu, Advisor
Nidhi Kohli, Co-advisor**

October, 2024

© Yu Wang 2024
ALL RIGHTS RESERVED

Acknowledgements

First and foremost, I wish to express my deepest gratitude to Professor Chia-yi Chiu, my advisor. This dissertation would not have been possible without her continuous guidance, mentoring, and encouragement over the past five years. I am also profoundly grateful to Professor Nidhi Kohli for her unwavering support, both before and after I joined the QME program. My sincere thanks extend to Professor Nana Kim and Professor Xiaou Li for serving on my dissertation committee and providing invaluable feedback.

I would like to acknowledge the faculty and staff members from the Department of Educational Psychology at the University of Minnesota, as well as the faculty at the Graduate School of Education at Rutgers University. My gratitude also goes to my supervisor Dr. William Muntean and my coworkers during my internship at the National Council of State Boards of Nursing.

Moreover, I want to extend my heartfelt thanks to my cohort and friends. A special thank you to Yuhang and Huaxin—I cannot imagine my life in Minneapolis without you. Thank you for taking care of my cats, my car, and me.

To the best kittens in the world—Hiou, Kiwi, Pepsi, and Sprite—thank you for your companionship. Although I may never understand the words you typed in my dissertation, your presence was a source of comfort during difficult times.

Lastly, I wish to express my deepest appreciation to my parents and grandparents for their unwavering love and support over the years. I am particularly grateful for their

respect for my decisions and their pride in my achievements. My heartfelt thanks also go to my husband, Huahao, and my parents-in-law for their encouragement, support, and companionship. My PhD journey was not always easy—it began just before the outbreak of COVID, and I moved to a new program in Minnesota after three years in New Jersey. Huahao and I faced these challenges together, and I am deeply thankful for his support, which made both the happy moments even brighter and the difficult ones more bearable.

Abstract

The multiple-choice (MC) item format has been widely used in educational assessments across diverse content domains. MC items purportedly allow for collecting richer diagnostic information. The effectiveness and economy of administering MC items may have further contributed to their popularity not just in educational assessment. The MC item format has also been adapted to the cognitive diagnosis (CD) framework. Early approaches simply dichotomized the responses and analyzed them with a CD model for binary responses. Obviously, this strategy cannot exploit the additional diagnostic information provided by MC items. De la Torre’s MC Deterministic Inputs, Noisy “And” Gate (MC-DINA) model was the first for the explicit analysis of items having MC response format. However, relying on the expectation-maximization algorithm for estimation, the MC-DINA model cannot guarantee stable and reliable estimates in small-sample settings —despite their effectiveness and efficiency when samples are large — due to computational feasibility issues caused by insufficient sample sizes. To address this issue, the nonparametric classification method for multiple-choice Items (MC-NPC) was proposed in Chapter 2. The mathematical justification supporting the legitimacy of this method was provided, and the classification performance was evaluated by simulation studies and a real-world application.

Computerized adaptive testing for cognitive diagnosis (CD-CAT) achieves remarkable estimation efficiency and accuracy by adaptively selecting and then administering items tailored to each examinee. The process of item selection stands as a pivotal component of a CD- CAT algorithm, with various methods having been developed for binary responses. However, the Jensen–Shannon divergence (JSD) index introduced by Yigit et al. is the only item selection method exclusively designed for MC items. As a parametric item selection method, the JSD index requires a large sample to calibrate item parameters by the MC-DINA model, which may be infeasible when there is only

a small or no calibration sample. To bridge this gap, a nonparametric item selection method for MC items (MC-NPS) was proposed in Chapter 3. The MC-NPS method implements novel discrimination power that measures an item's ability to effectively distinguish among different attribute profiles, as well as a Q-optimal procedure for MC items to select items during the initial phase of a CD- CAT algorithm. The effectiveness and efficiency of the MC-NPS method were confirmed by simulation studies.

Like most CD methods, the Q-matrix serves as the foundation for the CD methods for MC items. A crucial property of the Q-matrix is completeness, which refers to whether a Q-matrix can identify all possible attribute profiles. An incomplete Q-matrix may result in parameter identifiability issues and the misclassification of examinees, potentially distorting the interpretation of results and subsequent pedagogical interventions. Chapter 4 explored the sufficient and necessary conditions for a complete Q-matrix for MC items. Additionally, it addressed the closely related topic, the Q-optimal procedure for CD-CAT under the MC-DINA model, with formal establishment of its sufficient and necessary conditions. The proposed theorems, lemmas, and corollaries were rigorously proven through mathematical proofs, and simulation studies provided additional support for the completeness conditions of the Q-matrix for MC items and the Q-optimal procedure for CD-CAT under the MC-DINA model.

Contents

Acknowledgements	i
Abstract	iii
Contents	v
List of Tables	ix
List of Figures	xi
1 Introduction	1
2 Nonparametric Classification Method for Multiple-Choice Items in Cognitive Diagnosis	5
2.1 Introduction	5
2.2 Technical Background: Key Technical Concepts of Cognitive Diagnosis .	7
2.2.1 The DINA and MC-DINA Models	8
2.2.1.1 DINA Model	8
2.2.1.2 De la Torre's MC-DINA Model (2009)	9
2.2.2 The NPC Method	11
2.3 The MC-NPC Method	13
2.3.1 The Multiple-Choice Setting Modified for the MC-NPC Method	13

2.3.2	Inconsistencies of the Multiple-Choice Setting and Their Remedy	17
2.3.3	Theoretical Legitimization of the MC-NPC Method	20
2.4	Simulation Studies	23
2.4.1	Simulation Study I: Replication of de la Torre's (2009) Simulation	24
2.4.1.1	Design	24
2.4.1.2	Evaluation Criteria	26
2.4.1.3	Results	26
2.4.2	Simulation Study II: Assessment with Random Q-Matrices . . .	27
2.4.2.1	Design	27
2.4.2.2	Results	30
2.5	Real Data Analysis	32
2.5.1	Data	32
2.5.2	Results	35
2.6	Discussion	39

3 Nonparametric CD-CAT for Multiple-Choice Items:

	Item Selection Method and Q-Optimality	42
3.1	Introduction	42
3.2	Technical Background	44
3.2.1	The MC-NPC Method	46
3.2.2	Parametric CD-CAT for MC Responses: The JSD Index	47
3.2.3	Nonparametric CD-CAT for Dichotomized Responses: The GNPS Method	48
3.3	The Proposed MC-NPS Method and Q-Optimal Procedure for MC Items	49
3.3.1	The Proposed MC-NPS Method	49
3.3.2	Q-optimal Procedure for CD-CAT with MC Items	51
3.3.3	Algorithm	59
3.4	Simulation Studies	60

3.4.1	Simulation Study I: Evaluating the Effect of the Proposed MC Q-Optimal Procedure	61
3.4.1.1	Design	61
3.4.1.2	Measures	62
3.4.1.3	Results	62
3.4.2	Simulation Study II: Assessing the Performance of the Developed MC-NPS Algorithm	66
3.4.2.1	Design	66
3.4.2.2	Results	67
3.5	Discussion	74
4	Q-Matrix Completeness and Q-Optimal Design for Multiple-Choice Items in CD	79
4.1	Introduction	79
4.2	Technical Background	81
4.2.1	CD for Dichotomized Items	81
4.2.1.1	The DINA Model	82
4.2.2	CD for MC Items	82
4.2.2.1	The MC-DINA Model and Its Modifications	83
4.2.2.2	Useful and Proper Conditions for $\mathbf{Q}_{MC}$	85
4.3	Objective 1: Completeness of $\mathbf{Q}_{MC}$ for the MC-DINA Model	85
4.3.1	Review of Completeness of $\mathbf{Q}_B$ for the DINA Model	86
4.3.2	Level- x Design	86
4.3.2.1	Level-1 Design	89
4.3.2.2	Level-2 Design	90
4.3.2.3	Level-3 Design	90
4.3.2.4	Level-0 Design	91
4.3.3	Completeness of $\mathbf{Q}_{MC}$ for the MC-DINA Model	95

4.4	Objective 2: Identifications of Each Latent Class under the MC-DINA Model	100
4.4.1	Review of Identification of α_i under the binary responses	100
4.4.2	Identification of α_i under the MC-DINA model	100
4.5	Objective 3: Q-Optimal Design for CD-CAT with the MC-DINA Model	102
4.5.1	Review of the Q-Optimal Design for CD-CAT with the DINA Model	102
4.5.2	Minimum Test Length for Complete Q_{MC} under the MC-DINA Model	103
4.5.3	Q-Optimal Design for CD-CAT with the MC-DINA Model	104
4.6	Simulation Studies	107
4.6.1	Simulation Study I	108
4.6.1.1	Design	108
4.6.1.2	Evaluation Criteria	110
4.6.1.3	Results	110
4.6.2	Simulation Study II	113
4.6.2.1	Design	114
4.6.2.2	Evaluation Criteria	115
4.6.2.3	Results	115
4.7	Discussion	117
5	Conclusion and Discussion	120
	References	125
	Appendix A	131
	Appendix B	133

List of Tables

2.1	Fraction-subtraction Item	11
2.2	Example: Determining the Levels l of the Five Response Options of Item j	15
2.3	Example: Item Perturbation Parameters ϵ	16
2.4	Example: Item Perturbation Parameters $\epsilon_{jll'}$ When $H_j^* = 1$	19
2.5	Replication of de la Torre's (2009) Simulation Study: Mean PAR obtained for the MC-NPC, MC-DINA, NPC, and DINA Methods.	27
2.6	A Sample Q-Matrix Used in the Simulated Data ($K = 3$ and $J = 20$)	29
2.7	Mean PARs from the MC-NPC, MC-DINA, NPC, and DINA Methods When $H_j = 4$	30
2.8	Mean PARs from the MC-NPC, MC-DINA, NPC, and DINA Methods When $H_j = 5$	32
2.9	The Required Attributes of the Fraction Operations Assessment	33
2.10	Q-matrix of the Fraction Assessment	33
2.11	Items Removed in Class-level Analysis	35
2.12	PAR between the Total Examinees Analyzed by the Polytomous Methods and the Class Analyzed by Each Method	35
2.13	Distinct Estimated Attribute Profiles of Class5 With the MC-NPC Method and the MC-DINA Algorithm	38
2.14	Ideal Response Patterns With 2^K Attribute Profiles	41
3.1	Probabilities of Response Patterns (110) and (111)	53

3.2	Relationship Between N_e and Minimum H_j^*	55
3.3	The Number of Calibration Samples Generated in Each Condition	67
4.1	Classes Formulated by Distractor l and the Key of Item j	87
4.2	Ideal Responses to Four Items Assessing the Mastery of A2	88
4.3	Relationship Between the Level- x Design and Minimum H_j^*	95
4.4	Examples of Complete $\mathbf{Q}_{MC}$ with $K = 3$ and $H_j = 5$	98
4.5	$\varsigma(\boldsymbol{\alpha})$ of Items Shown in Table 4.4	99
4.6	Four Examples of $\mathbf{Q}_{MC}$	101
1	$E(\mathbf{I}_j \boldsymbol{\alpha})$ of Classes Formulated by Item j	136

List of Figures

2.1	An Item in the Fraction Assessment	34
3.1	PARs of Three Optimal Designs with Medium-Quality Items	63
3.2	PARs of Three Optimal Designs with High-Quality Items	65
3.3	PARs from the MC-NPS, JSD, GNPS, and Random Selection Methods with Medium-Quality Items When $H_j = 4$	68
3.4	PARs from the MC-NPS, JSD, GNPS, and Random Selection Methods with Medium-Quality Items When $H_j = 5$	70
3.5	PARs from the MC-NPS, JSD, GNPS, and Random Selection Methods with Medium-Quality Items When $H_j = 6$	71
3.6	PARs from the MC-NPS, JSD, GNPS, and Random Selection Methods with High-Quality Items When $H_j = 6$	73
3.7	PARs from the MC-NPS, JSD with True Parameters, and Random Se- lection Methods	76
4.1	An Example Set of Items to Distinguish between Mastery and Non- mastery of A2	92
4.2	Flowchart for Evaluating the Completeness of Q_{MC}	97
4.3	Evolution of $E(\hat{\alpha} \alpha)$ Across Items for the Latent Class $(0, 0, 0, 0)'$	111
4.4	Evolution of $E(\hat{\alpha} \alpha)$ Across Items for the Latent Class $(1, 0, 0, 0)'$	112
4.5	Evolution of $E(\hat{\alpha} \alpha)$ Across Items for the Latent Class $(1, 0, 1, 0)'$	113
4.6	Evolution of $E(\hat{\alpha} \alpha)$ Across Items for the Latent Class $(1, 1, 1, 1)'$	114

4.7	PARs of the MC-NPS and JSD Methods	116
A.1	PARs from the MC-NPS, JSD, GNPS, and Random Selection Methods with High-Quality Items When $H_j = 4$	131
A.2	PARs from the MC-NPS, JSD, GNPS, and Random Selection Methods with High-Quality Items When $H_j = 5$	132

Chapter 1

Introduction

Cognitive diagnosis (CD) are statistical instruments designed to assess the underlying cognitive skills, often referred to as attributes, manifested by examinees' responses to test items. Over the past decades, numerous CD models (CDMs) have been developed, each built upon distinct assumptions about the underlying cognitive mechanisms. For example, the deterministic inputs, noisy “and” gate (DINA; Junker & Sijtsma, 2001) model assumes that an examinee must possess all attributes required by an item to have a high probability of answering this item correctly. In contrast, the deterministic inputs, noisy “or” gate (DINO; Templin & Henson, 2006) model assumes that an examinee is highly likely to make a correct response by mastering at least one of the required attributes. The generalized-DINA (G-DINA; de la Torre, 2011) model relaxes the assumptions of the DINA/DINO models to accommodate more general yet complex underlying processes. The development of these CDMs has opened up diverse research and practical opportunities; however, these parametric models are not without problems. A pressing issue is that they do not perform adequately when samples are small, undermining the ultimate goal of CD - to improve learning and teaching when and where they occur. In response to this shortcoming, researchers have dedicated themselves to the development of nonparametric CD methods that elegantly avoid complex model

fitting, making them favorable in micro-educational settings. Typical examples include the nonparametric classification method (NPC; Chiu & Douglas, 2013) and the general NPC method (GNPC; Chiu et al., 2018; Chiu & Köhn, 2019). Both have demonstrated outstanding performance with small samples.

All the CDMs introduced above were developed for binary responses and thus, for certain test formats, data must be dichotomized so that they can be analyzed by these CDMs, which, consequently, may cause information loss. Multiple-choice (MC) items, a common format in educational assessment, are a case in point. With a careful design, MC items can offer diagnostic insights not only through the information provided from the correct answer (i.e., key) but also that embedded within the incorrect options (i.e., distractors). Unfortunately, when traditional CDMs are used, the responses to all the distractors are lumped together as the incorrect response without further analysis, wasting the diagnostic information underlying the distractors. To address this issue, several parametric CDMs have been developed, including the multiple-choice deterministic inputs, noisy “and” gate model (MC-DINA; de la Torre, 2009), structured MC-DINA models (Ozaki, 2015), and the generalized diagnostic classification models for MC option-based scoring (GDCM-MC; Dibello et al., 2015). These parametric CDMs for MC items have demonstrated superior performance in estimating examinees’ attributes compared to those developed for binary responses. However, for small-sample settings like classrooms, where they are presumably needed most, these parametric estimation methods in general often encounter computational feasibility issues and cannot guarantee stable and reliable estimates (e.g., Paulsen, 2019). In response to this challenge, the first study of this dissertation concerns the development of the nonparametric classification method for MC items (MC-NPC; Wang et al., 2023). The MC-NPC method overcomes the limitation the parametric methods have associated with sample size and presents a novel solution applicable across settings of any size. The paper describing the MC-NPC method has been published in *Journal of Educational and Statistical Measurement*.

Methods for computerized adaptive testing for CD (CD-CAT) have been developed to select items specifically tailored to each examinee’s ability, providing a more precise and reliable estimate of his/her attribute profile. However, to date, the only attempt to consider the information underlying distractors for item selection is the application of the Jensen-Shannon divergence (JSD) index to the MC responses conforming to the MC-DINA model (Yigit et al., 2019). It was demonstrated in their study that utilizing the JSD index to select MC items significantly improved classification accuracy in comparison to methods designed for binary responses. Nevertheless, the dependency on calibrated item parameters estimated by the MC-DINA model introduces limitations, hindering the applicability of the JSD index in CD-CAT scenarios with small calibration sample sizes or without any available calibration samples. These challenges are addressed in the second study of this dissertation, where a nonparametric item selection method for CD-CAT with MC items (MC-NPS) is proposed (Wang & Chiu, under review). The MC-NPS method utilizes the newly proposed discrimination power to select MC items and the MC-NPC method to estimate attribute profiles after administering each item. A remarkable feature of the MC-NPS method is that it is free of item calibration, which renders it applicable across calibration samples of any size. It has been shown that the MC-NPS method outperforms the JSD method when calibration samples are limited or unavailable. The paper describing the MC-NPS method has been published in *British Journal of Mathematical and Statistical Psychology*.

While advancements in estimation and item selection methods for MC items have capitalized on the diagnostic information embedded in distractors, the fundamental aspect of CDMs for MC items, the Q-matrix, remains largely unexplored. A Q-matrix specifying the relations between test items and cognitive skills they intend to measure (Tatsuoka, 1985) must be complete (Chiu et al., 2009) so that the identifications of all latent classes can be guaranteed (e.g., Chiu et al., 2009; Köhn & Chiu, 2016b; Köhn & Chiu, 2017). However, the completeness of the Q-matrix for MC items has yet to be investigated. An incomplete Q-matrix poses a risk of misclassifying examinees into groups

they do not belong to, thereby distorting the interpretation of results and subsequent pedagogical interventions. Therefore, it is crucial to investigate the conditions to determine the completeness of the Q-matrix for MC items. In addition to the exploration of completeness of the Q-matrix, another closely related topic, the Q-optimal design for CD-CAT with MC items will also be studied. A Q-optimal design holds the promise of providing optimal classification in the initial stage of a CD-CAT algorithm, with proven enhancements in estimation quality for binary responses (Xu et al., 2016). Thus, the third objective of this dissertation is to establish theorems outlining the completeness of the Q-matrix for MC items and defining the Q-optimal design for CD-CAT with MC items. This theoretical work aims to contribute significantly to the foundational understanding and effective application of CDMs for MC items.

To conclude, this dissertation comprises three interrelated studies in the CDMs for MC items. Chapter 2 introduces the MC-NPC method, a distance-based estimation approach specifically tailored for small-sample settings. The empirical evaluation of the MC-NPC method is conducted through both simulated and real-data studies. In Chapter 3, the focus shifts to the development of the MC-NPS method, a novel item selection technique for CD-CAT with MC items, notable for its exemption from the requirement of item parameter calibration. The performance of this method is systematically assessed through simulation studies. Finally, Chapter 4 contributes to the theoretical foundations of the Q-matrix by formulating two indispensable theorems. The first theorem addresses the completeness of the Q-matrix for the MC-DINA model, while the second establishes a Q-optimal design for CD-CAT with MC items. Additionally, relevant lemmas and corollaries are presented to augment the practical applicability of these theorems, collectively enriching the understanding and implementation of CDMs for MC items in educational assessments. All theorems, lemmas, and corollaries are rigorously proven through mathematical proofs, and simulation studies provide additional support for the two theorems.

Chapter 2

Nonparametric Classification Method for Multiple-Choice Items in Cognitive Diagnosis

2.1 Introduction

Despite mixed reviews of its specific merits (Wood, 2003), the multiple-choice (MC) item format has been widely used in educational assessments across diverse content domains. A typical MC item consists of the stem presenting the narrative and context that motivate the particular question the examinee is required to answer and a set of response options, which constitutes the “MC component” of the item. The correct answer, called the “key,” is embedded among several alternatives, the “distractors,” that differ in their degree of “(in-)correctness.” Compared with “traditional” items having a dichotomous response format, MC items purportedly allow for collecting potentially richer diagnostic information. In addition, different from essays and other constructed-response tasks that are also praised for their richer diagnostic information, MC items require examinees to spend less time on recording their answers and are less susceptible

to subjective scoring. More succinctly, the effectiveness and economy in administering MC items may account for their popularity not just in educational assessment.

The MC item format was also adapted to the cognitive diagnosis (CD) framework. CD seeks to provide a detailed evaluation of a student’s mastery of the instructional content in terms of skills learned and skills needing study. Early, less sophisticated approaches to analyzing MC items within the CD framework simply dichotomized the responses in scoring the key as 1, the distractors as 0, and applied one of the CD models available for binary responses. The disadvantage of such a strategy is obvious: the diagnostic potential of the MC response format is not exploited, the additional information that can be gleaned from the distractors is ignored. The first CD model for analyzing MC items in explicitly considering the distractors was de la Torre’s (2009) MC deterministic inputs, noisy “and” Gate (MC-DINA) model. Other CD models—henceforth, diagnostic classification models (DCMs)—for analyzing MC items were proposed by Ozaki (2015) and DiBello et al. (2015). The focus of de la Torre (2009) and Ozaki (2015) is on exploring how distractors can be used to improve examinee classification, whereas the generalized DCMs for MC option-based scoring (GDCM-MC) by DiBello et al. (2015) seeks to develop a system for identifying students’ misconceptions of the specific tasks posed by test items.

The DCMs developed by de la Torre (2009), Ozaki (2015) and DiBello et al. (2015) all rely on EM- or MCMC-based algorithms for the estimation of the model parameters. These approaches are effective and efficient if large samples are available. However, for small-sample settings like classrooms, where these DCMs are presumably needed most, parametric estimation methods in general often encounter computational feasibility issues and cannot guarantee stable and reliable estimates (e.g., Paulsen, 2019). In response to these difficulties, nonparametric methods for the diagnostic classification of examinees have been developed (Chang et al., 2019; Chiu & Chang, 2021; Chiu & Köhn, 2019; Chiu et al., 2018).

In this article, a nonparametric diagnostic classification algorithm for MC items

called MC-nonparametric classification (MC-NPC) method is presented that has been developed for the explicit use with small samples. The next section briefly reviews essential CD concepts. Technical details of the MC-NPC algorithm are described in Section 2.3, followed by a report on the results of simulation studies for evaluating the performance of the proposed MC-NPC algorithm including a comparison with de la Torre’s EM-based MC-DINA algorithm (implemented in the R package `GDINA`). Section 2.5 presents the side-by-side application of MC-NPC and EM-MC-DINA to a real-world data set. The discussion section concludes with a summary of the key insights and a discussion of some limitations and future research avenues.

2.2 Technical Background: Key Technical Concepts of Cognitive Diagnosis

Assume competence—or ability—in an instructional content area is characterized by a set of skills. The explicit purpose of CD is the assessment of mastery of these skills to provide immediate feedback to students on their strengths and weaknesses in terms of skills learned and skills needing study. CD-based tests consist of items that require for a correct response mastery of different skills. From the item responses, students’ skill mastery can be inferred. In CD idiom, skills, specific knowledge, aptitudes—any (cognitive) competence that is required to perform tasks—are collectively referred to as “attributes.” In formal terms, ability in a given knowledge domain is characterized as a composite of K attributes that an examinee may have mastered or failed, which is documented in a 0-1 profile having K entries denoted as $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_k, \dots, \alpha_K)'$ (DiBello et al., 2007; Haberman & von Davier, 2007; Leighton & Gierl, 2007; Nichols et al., 1995; Rupp et al., 2010; Sessoms & Henson, 2018; Tatsuoaka, 2009). Distinct binary vectors define attribute profiles of different classes of proficiency denoted by $\mathcal{C}_m$. (The terms profile and vector are used interchangeably here.) If attributes are not hierarchically organized, then the index m runs from 1 to $M = 2^K$. In reiterating, the

primary goal of cognitive diagnosis modeling is to assign examinees to one of these M classes based on their performance in a test that targets proficiency in the knowledge domain in question. Said differently, examinees' individual attribute profiles α_i must be estimated ($i = 1, 2, \dots, N$ is the examinee index).

Like examinees, so are the items of cognitively diagnostic tests characterized by individual profiles that determine which specific attributes are required for a correct response. The item-attribute profiles are K -dimensional binary vectors $\mathbf{q}_j$, $j = 1, 2, \dots, J$, where elements $q_{jk} = 1$ if a correct answer requires mastery of the k^{th} attribute α_k , and 0 otherwise. Notice that, given K binary attributes, at most $2^K - 1$ distinct item-attribute profiles can be constructed; item-attribute profiles comprising all zeroes are not legitimate.

The J item attribute profiles of a cognitively diagnostic assessment form the rows of its $J \times K$ Q-matrix, $\mathbf{Q} = \{q_{jk}\}_{(J \times K)}$ (Tatsuoka, 1985). As a key prerequisite to cognitive diagnosis, the Q-matrix of a test must be known and it must be complete. A Q-matrix is said to be complete if its specific composition can guarantee the identifiability of all realizable proficiency classes among examinees (Chiu et al., 2009; Köhn & Chiu, 2016a, 2017, 2019, 2021).

2.2.1 The DINA and MC-DINA Models

2.2.1.1 DINA Model

De la Torre's (2009) MC-DINA model is an extension of the Deterministic Inputs, Noisy "AND" Gate (DINA) model (Haertel, 1989; Junker & Sijtsma, 2001; Macready & Dayton, 1977). The item response function (IRF) of the DINA model is

$$P(Y_{ij} = 1 \mid \alpha_i) = (1 - s_j)^{\eta_{ij}} g_j^{(1 - \eta_{ij})},$$

where where $Y_{ij} = 1$ denotes a correct response to item j (otherwise, $Y_{ij} = 0$), s_j and g_j are the slipping and guessing parameters, respectively, for item j , subject to $0 \leq g_j <$

$1 - s_j \leq 1$. The ideal response (or conjunction parameter) $\eta_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}} \in \{0, 1\}$ indicates whether examinee i has mastered all attributes required by item j . The DINA is a conjunctive model, as the probability of a correct response is maximal only if an examinee has mastered all attributes required for a given item. Hence, each DINA item generates a bi-partition of the $M = 2^K$ proficiency classes of the latent attribute space into groups of examinees who have mastered the attributes required for said item as opposed to those who have not.

2.2.1.2 De la Torre's MC-DINA Model (2009)

Before the advent of de la Torre's (2009) MC-DINA model, according to common practice the polytomous responses of MC-items were dichotomized so that a student who had picked the key received a score of one whereas the choice of a distractor was scored as zero. Such a recoding scheme neglects the potential value of the diagnostic information from the distractors (e.g., partial knowledge as indicated by mastery of only a subset of the required attributes; students' misconceptions of the task required by an item; see, Haertel & Wiley, 1993; Nitko, 2001; Sadler, 1998). Thus, the motivation for the MC-DINA model (de la Torre, 2009) was to exploit the unused diagnostic information from the distractors to improve the classification of examinees—but how? Recall that each DINA item results in a bi-partition of the latent attribute space. In contrast, an MC item partitions the latent attribute space into as many proficiency classes as the number of coded response options plus one, thereby supposedly increasing the accuracy of examinee classification. The review of de la Torre's (2009) MC-DINA model follows in notation and terminology as closely as possible his original presentation in *Applied Psychological Measurement*.

The polytomous response to an MC item is denoted by the random variable X_{ij} taking on values $h = 1, 2, \dots, H_j$, with the total number of response options equal to H_j . MC-DINA jargon distinguishes between “non-coded” and “coded” or “cognitively based” response options. The total number of coded response options is H_j^* ; as not all

MC-DINA response options are coded, typically, $H_j^* \leq H_j$. Coded response options are linked to an item attribute vector $\mathbf{q}_{jh}$ that specifies the skills an examinee who endorses this option must have mastered. In contrast, “non-coded” response options like “none of these” or “all of the above” are not associated with a particular attribute vector. By convention, their item attribute vectors are written as a K -dimensional zero-vector, $\mathbf{q}_{j0} = (0, 0, \dots, 0)'$.

The relation between the different response options is governed by the following rules: the key always has the largest number of attributes; the attribute vectors of the coded response options must be nested within the q-vector of the key. As a further constraint, the attribute vectors of coded response options must be hierarchically nested within each other; said differently, they establish what Leighton, Gierl, and Hunka (2004) called a “linear hierarchy.” Hence, all response options form an ordinal scale with the key at the top, followed by the coded distractors, and the non-coded options last.

The ideal response of examinee i to MC item j is defined as

$$g_{ij} = \arg \max_{h'=0,1,2,\dots,H_j} \left\{ \alpha'_i \mathbf{q}_{jh'} \mid \alpha'_i \mathbf{q}_{jh'} = \alpha'_{jh'} \mathbf{q}_{jh'} \right\} \quad (2.1)$$

(the notation g_{ij} was introduced by de la Torre (2009), who also coined the term “latent group classification” for the ideal response). If an examinee has not mastered the attributes required for any coded option, then her ideal response g_{ij} is zero, and she is assumed to choose randomly among the H_j response options. In contrast, an examinee who has mastered all required attributes of at least one coded option has ideal response $g_{ij} = 1, 2, \dots, H_j$. Finally, the set $\mathcal{G}_j$ is defined by de la Torre (2009) as the set formed by the union of $\{0\}$ and a subset of $\{1, 2, \dots, H_j\}$.

Then, the probability that examinee i having ideal response $g_{ij} = g \in \mathcal{G}_j$ chooses option h of item j is written as

$$P(X_{ij} = h \mid g_{ij} = g) = P_j(h \mid g).$$

As illustration of these concepts, consider the following example presented in de la Torre (2009; p. 166) that concerns a test problem molded after Tatsuoka’s (1990; see also C. Tatsuoka, 2002) famous fraction-subtraction items:

Table 2.1: Fraction-subtraction Item

Item:	$2\frac{4}{12} - \frac{7}{12} = ?$	Required Attributes		
		α_1	α_2	α_3
Response Option 1	Distractor A. = $2\frac{3}{12}$		α_2	
Response Option 2	Distractor B. = $2\frac{1}{4}$		α_2	α_3
Response Option 3	Distractor C. = $1\frac{9}{12}$	α_1	α_2	
Response Option 4	Key = $1\frac{3}{4}$	α_1	α_2	α_3

The attributes are defined as follows:

α_1 Borrow one from whole number to fraction,

α_2 Basic fraction subtraction,

α_3 Reduce/Simplify.

In conclusion, as an important detail, one should notice that despite the MC response format, the MC-DINA model is still a conjunctive model—that is, the choice of any response category is assumed to be guided by the match of attributes required for a particular response option and those mastered by an examinee. More to the point, in case of the MC-DINA model, the more attributes an examinee has mastered the more likely she is to move up the hierarchy of response options as they match her attribute profile. Thus, the MC response format should not be confused with the concept of incremental probability of a correct answer as the result of mastering more items as it underlies general DCMs.

2.2.2 The NPC Method

The algorithm of the nonparametric classification (NPC) method (Chiu & Douglas, 2013) is adapted to form the basis of the MC-NPC method proposed in this article. The key idea underlying the NPC method is to estimate an examinee’s proficiency class, the attribute vector $\alpha_i = \alpha_m$, by comparing her observed item response vector

with each of the ideal response vectors of the $M = 2^K$ realizable proficiency classes. The NPC estimator $\hat{\alpha}_i$ of examinee i is defined as the attribute profile underlying that ideal response vector which among all ideal response vectors minimizes the Hamming distance to the manifest item response vector $\mathbf{Y}_i$. For $\mathcal{C}_m$, the ideal response vector is $\boldsymbol{\eta}^{(m)} = (\eta_1^{(m)}, \eta_2^{(m)}, \dots, \eta_J^{(m)})$; the Hamming distance is defined as

$$d_h(\mathbf{Y}_i, \boldsymbol{\eta}^{(m)}) = \sum_{j=1}^J |Y_{ij} - \eta_j^{(m)}|. \quad (2.2)$$

Depending on the specific data structure variations like the weighted Hamming distance may be more suitable (for further details, consult Chiu & Douglas, 2013). The estimated attribute profile of examinee i is identified by minimizing the distance $d_h(\mathbf{Y}_i, \boldsymbol{\eta}^{(m)})$ across all ideal response profiles $\boldsymbol{\eta}^{(1)}, \boldsymbol{\eta}^{(2)}, \dots, \boldsymbol{\eta}^{(M)}$ (recall $M = 2^K$) and observed response profile $\mathbf{Y}_i = \mathbf{y}_i$:

$$\hat{\alpha}_i = \arg \min_{\boldsymbol{\alpha}_m \in \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_M\}} d(\mathbf{y}_i, \boldsymbol{\eta}^{(m)}) = \arg \min_{\boldsymbol{\alpha}_m \in \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_M\}} \sum_{j=1}^J |y_{ij} - \eta_j^{(m)}|.$$

Given $P(Y_j = 1 \mid \eta_j = 0) < 0.5$ and $P(Y_j = 1 \mid \eta_j = 1) > 0.5$ for item j (Wang & Douglas, 2015) and completeness of the Q-matrix (Chiu et al., 2009; Köhn & Chiu, 2017), the estimator $\hat{\alpha}_i$ obtained by the NPC method is statistically consistent. Simulation studies (Chiu & Douglas, 20213; Chiu et al., 2018) showed that examinee classification by the NPC method is more accurate than that by parametric methods when samples are small. This remarkable feature also inspired the development of advanced applications of cognitive diagnosis tailored to small-scale educational settings such as CD-CAT for the classroom (Chang et al., 2019).

2.3 The MC-NPC Method

The development of the MC-NPC method was inspired by de la Torre’s (2009) MC-DINA model. Several features, however, distinguish the MC-NPC method and warrant a few remarks before presenting its technical details. First, different from MC-DINA, the q-vectors of the coded distractors do not need to be nested within each other and the q-vector of the key. But with these constraints removed, the linear ordering of the coded response options is lost. Consequently, for the MC-NPC method, the scale level of the response options is changed to nominal, which in turn calls for modifications of the ideal response as originally defined in Equation 2.1 for the MC-DINA model. Second, de la Torre (2009) does not provide the IRF of the MC-DINA model; however, the IRF underlying the MC-NPC method must be defined; otherwise, a comprehensive evaluation of its merits—theoretical and practical—is not possible.

2.3.1 The Multiple-Choice Setting Modified for the MC-NPC Method

To facilitate the implementation of these conceptual modifications and clarify the notation, in the MC-NPC method, the original MC-DINA response option index, $h = 1, 2, \dots, H_j$, is replaced by the index $l = 0, 1, 2, \dots, H_j^*$. As a convention, all non-coded response options are assigned $l = 0$; the key is indexed as $l = H_j^*$. The indices $l = 1, 2, \dots, H_j^* - 1$ are reserved for the remaining coded response options. The general notation for a q-vector having index l is $\mathbf{q}_j^{(l)} = (q_{j1}^{(l)}, q_{j2}^{(l)}, \dots, q_{jK}^{(l)})$. Thus, the q-vector of the key is denoted as $\mathbf{q}_j^{(H_j^*)}$, whereas $\mathbf{q}_j^{(0)}$ refers to non-coded options. But how should the indices “in-between” ranging from 1 to $H_j^* - 1$ be assigned to the q-vectors of the remaining coded responses? Recall that these response options are perceived as a nominal scale; hence, the associated indices serve merely as labels. Still, to avoid ambiguity and arbitrariness, a rationale is needed for assigning indices to q-vectors. Let $\mathbf{q}$ and $\mathbf{q}'$ denote two different q-vectors of the remaining coded responses. The first rule determines that if

(1) $\|\mathbf{q}_j^{(l)}\|_1 > \|\mathbf{q}_j^{(l')}\|_1$ then $l > l'$.

The second rule concerns the case where $\mathbf{q}$ and $\mathbf{q}'$ are of the same length. The response options are indexed based on their evaluation in lexicographic order—that is, $l > l'$ if the position of the first non-zero entry in $\mathbf{q}$ precedes that in $\mathbf{q}'$, or vice versa. Ties—both $\mathbf{q}$ -vectors share the position of the first non-zero entry—are ignored and the evaluation is based on the first position with distinct entries; such a position can always be identified because all coded response $\mathbf{q}$ -vectors must be distinct. Formally, define the set $\mathcal{L}(\mathbf{q}, \mathbf{q}') = \{k \mid q_k > q'_k, k = 1, 2, \dots, K\}$. Notice that $\mathcal{L}(\mathbf{q}, \mathbf{q}') \neq \mathcal{L}(\mathbf{q}', \mathbf{q})$ due to the evaluation of q_k and q'_k in lexicographic order. The second rule then determines that if

(2) $(\|\mathbf{q}_j^{(l)}\|_1 = \|\mathbf{q}_j^{(l')}\|_1) \wedge \left(\min \mathcal{L}(\mathbf{q}_j^{(l)}, \mathbf{q}_j^{(l')}) < \min \mathcal{L}(\mathbf{q}_j^{(l')}, \mathbf{q}_j^{(l)}) \right)$ then $l > l'$.

Here is an example using some arbitrary item j with five response options (i.e., $H_j = 5$). Response options 3 and 5 are non-coded; hence, their $\mathbf{q}$ -vectors receive indices $l = 0$ and are denoted as $\mathbf{q}_j^{(0)}$. Options 1 and 4 are (coded) distractors having $\mathbf{q}$ -vectors (110) and (101), respectively. Response option 2 with $\mathbf{q}$ -vector (011) is the key. There are two coded response options in addition to the key; hence, $H_j^* = 3$. The attribute vector of the key is indexed as $\mathbf{q}_j^{(H_j^*)} = \mathbf{q}_j^{(3)} = (011)$. But how to assign the two “non-used” indices $l = 1, 2$ to the two coded $\mathbf{q}$ -vectors? Comparing (110) and (101) leads to $\mathcal{L}((110), (101)) = \{2\}$ due to $q_{j2} > q'_{j2}$; then, $\mathcal{L}((101), (110)) = \{3\}$, as $q'_{j3} > q_{j3}$. Because $\min \mathcal{L}((110), (101)) < \min \mathcal{L}((101), (110))$ the $\mathbf{q}$ -vectors are indexed as $(101) = \mathbf{q}_j^{(1)}$ and $(110) = \mathbf{q}_j^{(2)}$. Table 2.2 summarizes these steps.

After the indices l of the item response options have been determined, the ideal response η_{ij} of examinee i to item j can be computed

$$\eta_{ij} = \max_{l=0,1,2,\dots,H_j^*} \left\{ l \prod_{k=1}^K I[\alpha_{ik} \geq q_{jk}^{(l)}] \right\}. \quad (2.3)$$

Notice that Equation 2.3 is equivalent to Equation 2.1 when all the coded options are

Table 2.2: Example: Determining the Levels l of the Five Response Options of Item j

Option h	q-Vector	Level l of X_{ij}	Comments
3, 5	(000)	0	Non-coded response options.
2	(011) = $\mathbf{q}_j^{(H_j^*)}$	3	Response Option 2 is the key; thus, $H_j^* = 3 = l$, as there are two additional coded response options.
1	(110) = $\mathbf{q}$	?	$q_{j2} > q'_{j2}$; hence, $\mathcal{L}((110), (101)) = \{2\}$
4	(101) = $\mathbf{q}'$	?	$q'_{j3} > q_{j3}$; hence, $\mathcal{L}((101), (110)) = \{3\}$
Therefore,			
1	(110) = $\mathbf{q}$	2	due to $\min \mathcal{L}((110), (101)) < \min \mathcal{L}((101), (110))$
4	(101) = $\mathbf{q}'$	1	$\Rightarrow 2 < 3$

nested within each other.—Here is an illustration how to use Equation 2.3. Suppose examinee i has attribute profile $\alpha_i = (110)$. Recall that response options 3 and 5 are non-coded; hence, their q-vectors receive indices $l = 0$ and are denoted as $\mathbf{q}_j^{(0)}$. Options 1 and 4 are (coded) distractors having q-vectors $(101) = \mathbf{q}_j^{(1)}$ and $(110) = \mathbf{q}_j^{(2)}$, respectively. Response option 2 with q-vector (011) is the key with attribute vector $\mathbf{q}_j^{(H_j^*)} = \mathbf{q}_j^{(3)} = (011)$. Thus, in using Equation 2.3, the ideal response of examinee i is computed as

$$\eta_{ij} = \max_{l \in \{0,1,2,3\}} \left\{ l \prod_{k=1}^5 I[\alpha_{ik} \geq q_{jk}^{(l)}] \right\} = \max\{0, 0, 2, 0\} = 2.$$

Given the different instances of η_{ij} for each of the α_m , the IRF of the variant of the MC-DINA model underlying the MC-NPC method can be explicitly formulated. A few conceptual clarifications are necessary. First, recall that in de la Torre's original concept of the MC-DINA model an examinee with $\eta_{ij} = 0$ is not disposed to choose a particular response option (in de la Torre's [2009] terminology, the examinee is not "attracted" to any of the coded response options). Instead, said examinee is supposed to choose one of the response options at random; this assumption is also maintained

here. Second, if for an examinee $(\eta_{ij} = l) \wedge (l \neq 0)$ is true, then she is supposed to choose the coded response option $X_{ij} = l$ with a high probability; however, with non-zero probability she may also pick one of the other response options. Third, because manifest and ideal item responses, X and η , now have more than two levels, addressing potential discrepancies between X and η as “slips” and “guesses” does not adequately account for the increased complexity of the MC setting with multiple item parameters. Instead, whenever observed and ideal responses disagree, the more general term “perturbation” should be preferred, which calls also for adjusting the notation. Define $\epsilon_{jll'}$ as the probability that the observed response level l disagrees with the ideal response level l' . The explicit form of the IRF of the MC-DINA model is then

$$P(X_{ij} = l \mid \alpha_i) = \begin{cases} \frac{1}{H_j} + \frac{H_j - H_j^* - 1}{H_j} I[l = 0] & \text{if } \eta_{ij} = 0 \\ \left(1 - \sum_{m \neq l} \epsilon_{jml}\right)^{I[\eta_{ij}=l, l>0]} \prod_{l' \neq l} \epsilon_{jll'}^{I[\eta_{ij}=l']} & \text{if } \eta_{ij} > 0. \end{cases} \quad (2.4)$$

Notice that in case of the DINA model, the two perturbations, slipping s_j and guessing g_j , are constrained to be less than 0.5 (otherwise, an individual mastering none of the attributes would have a probability greater than 0.5 to provide the correct answer). Of course, if there are more than two perturbation terms, then the desirable property is that $\sum_{m \neq l} \epsilon_{jml} < 0.5$. As an illustration, consider again the example used earlier. Table 2.3 reports the item parameters $\epsilon_{jll'}$ of the MC-DINA model.

Table 2.3: Example: Item Perturbation Parameters ϵ

Option h	Level l of X_{ij}	Level l' of η_{ij}			
		0	1	2	3
3, 5	0	2/5	ϵ_{j01}	ϵ_{j02}	ϵ_{j03}
4	1	1/5	$1 - \epsilon_{j01} - \epsilon_{j21} - \epsilon_{j31}$	ϵ_{j12}	ϵ_{j13}
1	2	1/5	ϵ_{j21}	$1 - \epsilon_{j02} - \epsilon_{j12} - \epsilon_{j32}$	ϵ_{j23}
2	3	1/5	ϵ_{j31}	ϵ_{j32}	$1 - \epsilon_{j03} - \epsilon_{j13} - \epsilon_{j23}$

Recall response options 3 and 5 are non-coded; hence, their q-vectors receive indices

$l = 0$ and are denoted as $\mathbf{q}_j^{(0)}$. Options 1 and 4 are (coded) distractors having q-vectors $(101) = \mathbf{q}_j^{(1)}$ and $(110) = \mathbf{q}_j^{(2)}$, respectively. Response option 2 with q-vector (011) is the key with attribute vector $\mathbf{q}_j^{(H_j^*)} = \mathbf{q}_j^{(3)} = (011)$. Hence, given $\alpha_i = (110)$ and $\eta_{ij} = 2$, Equation 2.4 of the IRF of the MC-DINA model provides the following probabilities for the response levels of X_{ij} :

$$\begin{aligned}
P(X_{ij} = 0 \mid \alpha_i) &= \epsilon_{j01}^{I[\eta_{ij}=1]} \epsilon_{j02}^{I[\eta_{ij}=2]} \epsilon_{j03}^{I[\eta_{ij}=3]} \\
&= \epsilon_{j02} \\
P(X_{ij} = 1 \mid \alpha_i) &= (1 - \epsilon_{j01} - \epsilon_{j21} - \epsilon_{j31}) \epsilon_{j12}^{I[\eta_{ij}=1]} \epsilon_{j13}^{I[\eta_{ij}=2]} \epsilon_{j13}^{I[\eta_{ij}=3]} \\
&= \epsilon_{j12} \\
P(X_{ij} = 2 \mid \alpha_i) &= \epsilon_{j21}^{I[\eta_{ij}=1]} (1 - \epsilon_{j02} - \epsilon_{j12} - \epsilon_{j32}) \epsilon_{j23}^{I[\eta_{ij}=2]} \epsilon_{j23}^{I[\eta_{ij}=3]} \\
&= 1 - \epsilon_{j02} - \epsilon_{j12} - \epsilon_{j32} \\
P(X_{ij} = 3 \mid \alpha_i) &= \epsilon_{j31}^{I[\eta_{ij}=1]} \epsilon_{j32}^{I[\eta_{ij}=2]} (1 - \epsilon_{j03} - \epsilon_{j13} - \epsilon_{j23}) \epsilon_{j23}^{I[\eta_{ij}=3]} \\
&= \epsilon_{j32}.
\end{aligned}$$

2.3.2 Inconsistencies of the Multiple-Choice Setting and Their Remedy

Recall that for the MC-DINA model as presented here the response options $h = 1, 2, \dots, H_j$ are mapped into the response levels $l = 0, 1, \dots, H_j^*$. If $\eta > 0$, then the probability of a response at level l is large provided l and l' agree, and it is small when l and l' disagree. For all proficiency classes having $\eta > 0$, the application of the MC-NPC method is a straightforward extension of the NPC method. However, if $\eta = 0$, then the MC setting could lead to the paradox that the probability $P(X_{ij} = 0 \mid \eta_{ij} = 0)$ is less than $P(X_{ij} > 0 \mid \eta_{ij} = 0)$. Verbally stated, an examinee with $\eta_j = 0$ is more likely to choose $X_{ij} > 0$ than $X_{ij} = 0$, which is a profound violation of a key assumption of the DINA model: examinees who fail mastery of the attributes required for an item (i.e., $\eta_j = 0$) should choose response $X_{ij} = 0$ with maximal and a (partially) correct response

option with minimal probability. As an illustration of such a violation, consider again the earlier example presented in Table 1. Clearly, $P(X_{ij} = 0 \mid \eta_{ij} = 0) = 2/5$, whereas $P(X_{ij} > 0 \mid \eta_{ij} = 0) = \sum_{l=1}^{H_j^*} P(X_{ij} = l \mid \eta_{ij} = 0) = 3/5$. For the MC-DINA model, this constellation is not a rare anomaly but occurs whenever the ideal response is zero and the ratio of coded to the total of response options exceeds 0.5 (i.e., $H_j^*/H_j > 0.5$). How does this affect the NPC method? Recall that the NPC method estimates an examinee's proficiency class in relying on the shortest distance between observed and ideal responses. If a response $X_{ij} > 0$, discrepant from an examinee's true ideal response $\eta_{ij} = 0$, is more likely to occur than the concordant response $X_{ij} = 0$, then the NPC method may result in the misclassification of an examinee because the more likely but discrepant response $X_{ij} > 0$ points at an incorrect proficiency class.

As a remedy to this issue, the distances used by the NPC method are adjusted to reduce the risk of potential misclassifications in case the ideal response $\eta_{ij} = 0$. A few more technical explanations seem warranted. First, notice that the Hamming distance between observed and ideal response as defined in Equation 2.2 is not suitable for the MC item format. However, as the Hamming distance concerns only binary entities, Equation 2.2 can be readily adapted to the MC item format in using the indicator function:

$$d_h(\mathbf{X}_i, \boldsymbol{\eta}^{(m)}) = \sum_{j=1}^J I[\eta_j^{(m)} \neq X_{ij}],$$

which can be written as

$$d_h(\mathbf{X}_i, \boldsymbol{\eta}^{(m)}) = \sum_{j=1}^J I[\eta_j^{(m)} > 0, X_{ij} \neq \eta_j^{(m)}] + \sum_{j=1}^J I[\eta_j^{(m)} = 0, X_{ij} \neq \eta_j^{(m)}],$$

in distinguishing explicitly between $\eta_j^{(m)} > 0$ and $\eta_j^{(m)} = 0$. In the latter case, if $X_{ij} \neq \eta_j^{(m)}$, then an item-specific penalty $w_j \in [0, 1]$ is applied to shrink the value of

the associated integer function resulting in the *penalized Hamming distance*

$$d_p(\mathbf{X}_i, \boldsymbol{\eta}^{(m)}) = \sum_{j=1}^J I[\eta_j^{(m)} > 0, X_{ij} \neq \eta_j^{(m)}] + \sum_{j=1}^J w_j I[\eta_j^{(m)} = 0, X_{ij} \neq \eta_j^{(m)}], \quad (2.5)$$

that reduces the impact of those instances possibly pointing at an incorrect $\eta_j^{(m)}$.—But how to determine w_j ?

Recall that the potential misclassification by the NPC method is related to the critical incident $P(X_{ij} = 0 \mid \eta_{ij} = 0) < P(X_{ij} > 0 \mid \eta_{ij} = 0)$. The difference between these two probabilities—in casual terms, the “severity” of the incident—can be shown to be a direct function of the number of coded response options of an item, H_j^* : the larger H_j^* , the more likely an undesirable impact on $d_p(\mathbf{X}_i, \boldsymbol{\eta}^{(m)})$. Specifically, the following scenarios can be distinguished. First, if there are no coded distractors—that is, $H_j^* = 1$ —then the MC-NPC method reduces to the familiar NPC method for binary items, as can be illustrated using the earlier example, where the 4×4 Table 2.3 is collapsed into the 2×2 Table 2.4 corresponding to a binary setting. Notice that now

Table 2.4: Example: Item Perturbation Parameters $\epsilon_{jll'}$ When $H_j^* = 1$

Level l of X_{ij}	Level l' of η_{ij}	
	0	1
0	4/5	ϵ_{j01}
1	1/5	$1 - \epsilon_{j01}$

$P(X_{ij} = 0 \mid \eta_{ij} = 0) = (H_j - 1)/H_j = 4/5$ exceeds $P(X_{ij} = 1 \mid \eta_{ij} = 0) = 1/H_j = 1/5$, as it should.

Next, consider $H_j^* > 1$. If H_j^* is increased by 1, then the probability $P(X_{ij} = 0 \mid \eta_{ij} = 0)$ can be shown to decrease by $1/H_j$. If $H_j^* = H_j - 1$, then $P(X_{ij} = 0 \mid \eta_{ij} = 0)$ is $1/H_j$. Finally, if $H_j^* = H_j$, then all distractors are coded and $X_j > 0$ must be true for all j . Said differently, the choice of $X_j = 0$ is impossible regardless of whether the ideal response η_j is zero or not. Thus, $P(X_{ij} = 0 \mid \eta_{ij} = 0) = 0$, which implies

$P(X_{ij} \neq 0 \mid \eta_{ij} = 0) = 1$. But this particular constellation is the ultimate violation of the conjunctive rule which invalidates the NPC method. Hence, the penalty w_j must be set to 0. In summary, these different scenarios suggest the following formal definition of the penalty term w_j :

$$w_j = \begin{cases} 1 - \frac{H_j^* - 1}{H_j} & H_j^* < H_j \\ 0 & H_j^* = H_j. \end{cases} \quad (2.6)$$

After w_j has been identified, the attribute profile of examinee i is estimated by minimizing the distance $d_p(\mathbf{X}_i, \boldsymbol{\eta}^{(m)})$ across all ideal response profiles $\boldsymbol{\eta}^{(1)}, \boldsymbol{\eta}^{(2)}, \dots, \boldsymbol{\eta}^{(M)}$ and observed response profile $\mathbf{X}_i = \mathbf{x}_i$:

$$\begin{aligned} \hat{\boldsymbol{\alpha}}_i &= \arg \min_{\boldsymbol{\alpha}_m \in \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_M\}} d_p(\mathbf{x}_i, \boldsymbol{\eta}^{(m)}) \\ &= \arg \min_{\boldsymbol{\alpha}_m \in \{\boldsymbol{\alpha}_1, \boldsymbol{\alpha}_2, \dots, \boldsymbol{\alpha}_M\}} \left(\sum_{j=1}^J I[\eta_j^{(m)} > 0, X_{ij} \neq \eta_j^{(m)}] + \sum_{j=1}^J w_j I[\eta_j^{(m)} = 0, X_{ij} \neq \eta_j^{(m)}] \right). \end{aligned}$$

2.3.3 Theoretical Legitimization of the MC-NPC Method

The legitimacy of the nonparametric MC-NPC method as a tool for the analysis of MC-DINA data is established in demonstrating that minimizing the distance between the observed and the ideal response in Equation 2.5 used by the MC-NPC method is equivalent to maximizing the likelihood of the MC-DINA model. Recall the MC-DINA IRF in Equation 2.4 (repeated here for convenience):

$$P(X_{ij} = l \mid \boldsymbol{\alpha}_i) = \begin{cases} \frac{1}{H_j} + \frac{H_j - H_j^* - 1}{H_j} I[l = 0] & \text{if } \eta_{ij} = 0 \\ \left(1 - \sum_{m \neq l} \epsilon_{jml}\right)^{I[\eta_{ij}=l, l>0]} \prod_{l' \neq l} \epsilon_{jl'l'}^{I[\eta_{ij}=l']} & \text{if } \eta_{ij} > 0. \end{cases} \quad (2.4 \text{ revisited})$$

For the MC-DINA model, the likelihood of examinee i is

$$L_i = \prod_{j=1}^J L_{ij},$$

where L_{ij} , the likelihood of item j of examinee i , given response x_{ij} , is defined as

$$L_{ij} = \begin{cases} \left(\frac{H_j - H_j^*}{H_j} \right)^{I[x_{ij}=\eta_{ij}]} \left(\frac{1}{H_j} \right)^{I[x_{ij} \neq \eta_{ij}]} & \text{if } \eta_{ij} = 0 \\ \left(1 - \sum_{m \neq l} \epsilon_{jml} \right)^{I[x_{ij}=\eta_{ij}=l]} \prod_{l' \neq l} \epsilon_{jll'}^{I[x_{ij}=l, \eta_{ij}=l']} & \text{if } \eta_{ij} > 0. \end{cases} \quad (2.7)$$

To emphasize the connection between the likelihood in Equation 2.7 and the distance used by the MC-NPC method defined in Equation 2.5, four cases are distinguished. (Recall that $\eta_{ij} = \eta_j^{(m)}$.)

Case 1: $\eta_{ij} = 0$ and $1 \leq H_j^* < H_j - 1$.

For this case, the likelihood L_{ij} reduces to

$$L_{ij} = L_{ij}^{(1)} = \left(\frac{H_j - H_j^*}{H_j} \right)^{I[x_{ij}=\eta_{ij}]} \left(\frac{1}{H_j} \right)^{I[x_{ij} \neq \eta_{ij}]}. \quad (2.8)$$

$H_j - H_j^* > 1$ implies $\frac{H_j - H_j^*}{H_j} > \frac{1}{H_j}$. Hence, $L_{ij}^{(1)}$ is maximized if $X_{ij} = \eta_{ij} = 0$. Also, if $\eta_j^{(m)} = 0$, then, according to Equation 2.5,

$$d_p(X_{ij}, \eta_j^{(m)}) = d_p^{(1)} = w_j I[\eta_j^{(m)} = 0, X_{ij} \neq \eta_j^{(m)}]. \quad (2.9)$$

Hence, $d_p(X_{ij}, \eta_j^{(m)})$ is minimized if $X_{ij} = \eta_j^{(m)} = 0$ —or, more succinctly, if $X_{ij} = 0$, then $d_p^{(1)}$ is minimized and $L_{ij}^{(1)}$ is maximized.

Case 2: $\eta_{ij} = 0$ and $H_j^* = H_j - 1$.

For this case, the likelihood L_{ij} equals

$$L_{ij} = \left(\frac{H_j - H_j^*}{H_j} \right)^{I[x_{ij}=\eta_{ij}]} \left(\frac{1}{H_j} \right)^{I[x_{ij} \neq \eta_{ij}]}. \quad (2.10)$$

Because $H_j^* = H_j - 1$, implying $H_j - H_j^* = 1$, L_{ij} reduces to

$$L_{ij} = L_{ij}^{(2)} = \left(\frac{1}{H_j}\right)^{I[x_{ij}=\eta_{ij}]} \left(\frac{1}{H_j}\right)^{I[x_{ij}\neq\eta_{ij}]} = \frac{1}{H_j}. \quad (2.10)$$

Now, based on Equation 2.6, the penalty w_j equals $w_j = \frac{2}{H_j}$, which leads to

$$d_p(X_{ij}, \eta_j^{(m)}) = d_p^{(2)} = \frac{2}{H_j} I[\eta_j^{(m)} = 0, X_{ij} \neq \eta_j^{(m)}]. \quad (2.11)$$

Thus, if $H_j^* = H_j - 1$, then $L_{ij}^{(2)}$ is a constant but $d_p^{(2)}$ is minimized when $X_{ij} = \eta_j^{(m)} = 0$.

Case 3: $\eta_{ij} = 0$ and $H_j^* = H_j$.

For this case, the likelihood L_{ij} equals

$$L_{ij} = \left(\frac{H_j - H_j^*}{H_j}\right)^{I[x_{ij}=\eta_{ij}]} \left(\frac{1}{H_j}\right)^{I[x_{ij}\neq\eta_{ij}]}.$$

Because X_{ij} cannot be 0 if $H_j^* = H_j$, $X_{ij} \neq \eta_{ij}$, and thus, $L_{ij} = \frac{1}{H_j}$, which is $L_{ij}^{(2)}$. In addition, based on Equation 2.6, the penalty w_j is 0, and thus, $d_p(X_{ij}, \eta_j^{(m)}) = 0$.

Case 4: $\eta_{ij} > 0$.

In this case, the likelihood L_{ij} is

$$L_{ij} = L_{ij}^{(3)} = \left(1 - \sum_{l' \neq l} \epsilon_{jl'l}\right)^{I[x_{ij}=\eta_{ij}=l>0]} \prod_{l' \neq l} \epsilon_{jll'}^{I[x_{ij}=l, \eta_{ij}=l'>0]}. \quad (2.12)$$

Given the assumption that an examinee whose $\eta_{ij} = l$ and $l \neq 0$ has a higher probability to be attracted by the coded option $X_{ij} = l$ than by any other option, it must be true that $1 - \sum_{l' \neq l} \epsilon_{jl'l} > \epsilon_{jll'}$. Hence, $L_{ij}^{(3)}$ is maximized when $X_{ij} = \eta_{ij}$. In addition, based on Equations 2.5 and 2.6, it can be shown that

$$d_p(X_{ij}, \eta_j^{(m)}) = d_p^{(3)} = I[\eta_j^{(m)} > 0, X_{ij} \neq \eta_j^{(m)}], \quad (2.13)$$

which is minimized when $X_{ij} = \eta_j^{(m)}$. In conclusion, $d_p^{(3)}$ is minimized when $X_{ij} = \eta_j^{(m)}$,

whereas $L_{ij}^{(3)}$ is maximized.

Based on the above cases, the likelihood of examinee i can be expressed as

$$\begin{aligned} L_i &= \prod_{j=1}^J L_{ij} \\ &= \prod_{1 \leq H_j^* < H_j - 1} L_{ij}^{(1)I[\eta_{ij}=0]} L_{ij}^{(3)I[\eta_{ij} \neq 0]} \prod_{H_j - 1 \leq H_j^* \leq H_j} L_{ij}^{(2)I[\eta_{ij}=0]} L_{ij}^{(3)I[\eta_{ij} \neq 0]}, \end{aligned} \quad (2.14)$$

where $L_{ij}^{(2)}$ is a constant. In addition, it can also be shown that

$$\begin{aligned} d_p(\mathbf{X}_i, \boldsymbol{\eta}^{(m)}) &= \sum_{j=1}^J d_p(X_{ij}, \eta_j^{(m)}) \\ &= \sum_{1 \leq H_j^* < H_j - 1} (d_p^{(1)} + d_p^{(3)}) \sum_{H_j^* = H_j - 1} (d_p^{(2)} + d_p^{(3)}) + \sum_{H_j^* = H_j} d_p^{(3)}. \end{aligned} \quad (2.15)$$

Based on Equation 2.15, when $d_p(\mathbf{X}_i, \boldsymbol{\eta}^{(m)})$ is minimized, $d_p^{(1)}$, $d_p^{(2)}$, and $d_p^{(3)}$ are minimized, which implies $L_{ij}^{(1)}$ and $L_{ij}^{(3)}$ are maximized based on the above case distinction, and thus, so is L_i .

In conclusion, the above justification supports the legitimacy of using the MC-NPC method to analyze the MC-DINA data.

2.4 Simulation Studies

Two simulation studies implementing different experimental conditions were conducted to evaluate the performance of the proposed MC-NPC method. Key research questions were as follows:

- (i) Can the purported advantage of the MC over the binary response format in improving the accuracy of examinee classification in CD be confirmed?
- (ii) Do parametric estimation methods underperform in comparison with nonparametric methods when sample sizes are small?

The technical details of the two studies and their results are reported in the subsequent sections.

2.4.1 Simulation Study I: Replication of de la Torre’s (2009) Simulation

Study I is essentially a replication of de la Torre’s (2009) original simulation study where he compared the performance of the DINA with that of his new MC-DINA model. In the present study, MC-NPC and NPC were used as additional methods for investigating the performance of the various methods also when sample sizes were small.

2.4.1.1 Design

De la Torre (2009) used only a sample size of $N = 1,000$, whereas here, further sample sizes of $N = 20, 30, 50$, and 100 were included. The remaining experimental factors were the number of items $J = 30$, of attributes $K = 5$, and of response options $H_j = 4$ for all j . The Q-matrix from de la Torre (2009, p. 13) was used, which had imposed the constraint that for each item, the q-vectors of the coded distractors must be nested within that of the key.

As a technical detail, one should notice that the version of the MC-DINA model used in de la Torre (2009) was a simpler version of the more general one that was presented in the earlier section “The DINA and MC-DINA Models.” Specifically, in de la Torre’s 2009-version, the perturbations $\epsilon_{jl'}$ for a given level l' of η_{ij} , with $l' > 0$, are constrained to equal probabilities for all levels l across different responses X_{ij} , with $l \neq l'$. Denote a particular perturbation shared by different responses as $\epsilon_{jl'}$ for level l' of η_{ij} ; then, Equation (2.4) is reduced to

$$P(X_{ij} = l | \alpha_i) = \begin{cases} \frac{1}{H_j} + \frac{H_j - H_j^* - 1}{H_j} I(l = 0) & \text{if } \eta_{ij} = 0 \\ (1 - (H_j - 1)\epsilon_{jl'})^{I[\eta_{ij}=l]} \epsilon_{jl'}^{I[\eta_{ij} \neq l]} (H_j - H_j^*)^{I[l=0]} & \text{if } \eta_{ij} > 0. \end{cases} \quad (2.16)$$

In reiterating, this simpler version from de la Torre's 2009-paper was used in Study I.

Like de la Torre (2009), the examinee attribute profiles were drawn from a discrete uniform distribution with equal probabilities. In addition, for this study, examinee attribute profiles were also generated based on the multivariate normal threshold model (Chiu, Douglas, & Li, 2009), with variances equal to 1 and covariances sampled from Unif(0.3, 0.5). The probability of an examinee in group $\eta_{ij} = l'$ choosing response level $X_{ij} = l$ of item j was specified as

$$P(X_{ij} = l | \eta_{ij} = l') = \begin{cases} \frac{1}{H_j} + \frac{H_j - H_j^* - 1}{H_j} I(l = 0) & \text{if } l' = 0 \\ 0.82 & \text{if } l' > 0 \text{ and } l = l' \\ \frac{0.18}{H_j - 1} & \text{if } l' > 0 \text{ and } l \neq l'. \end{cases}$$

The implication of this particular setting is that examinees who do not master any of the attributes required for a coded response option are not attracted to any response option. Hence, these examinees are assumed to choose a response randomly. In contrast, examinees who master the attributes required by at least one of the coded response options have a probability of 0.82 to pick this particular option; but they may also chose among the remaining options randomly.

The DINA and the MC-DINA model were fitted using the functions `GDINA` and `MCmodel`, respectively, in the R package `GDINA` that uses the EM algorithm for the marginal maximum likelihood estimation of the model parameters. Thus, the DINA and the MC-DINA model are henceforth referred as DINA and MC-DINA.

2.4.1.2 Evaluation Criteria

The performance of the methods used in Study I was assessed in terms of the pattern-wise agreement rate (PAR), which is essentially the percentage of correct examinee classifications defined as

$$\text{PAR} = \frac{\sum_{i=1}^N I[\hat{\alpha}_i = \alpha_i]}{N}.$$

The mean PAR of each set of 100 replications was computed. Notice that the MC-NPC, NPC, and DINA methods did not encounter any convergence issues; thus, the mean PAR is based on all 100 replications. However, for small samples, the MC-DINA algorithm in the `GDINA` package occasionally did not converge. Hence, in these instances, the mean PAR was computed based only on those replications where MC-DINA converged; these mean PAR are marked as “C.” For the sake of a fair comparison of performance, the mean PAR for the MC-NPC method was also computed for the replications where the MC-DINA algorithm converged. These additional mean PAR are identified by “MC-NPC*.”

2.4.1.3 Results

The results are reported in Table 2.5. The comparison of the MC-NPC method with the two methods for dichotomous responses, NPC and DINA, reveals that across all sample sizes the MC-NPC method resulted in substantially increased mean PAR—at least, for the Q-matrix with coded distractors as it was used by de la Torre (2009) and here. Comparing the mean PAR of the MC-NPC method (100 replications) with those obtained for the MC-DINA method (C replications only) confirms that the former outperformed the latter if samples were small. This was also true when only the C replications were compared (i.e., MC-NPC* versus MC-DINA column). However, as expected, the MC-NPC method performed slightly worse than the MC-DINA method if $N = 1,000$.

Notice that for small samples (i.e., $N = 20$ and 30) in 82% and 17%, respectively,

Table 2.5: Replication of de la Torre’s (2009) Simulation Study: Mean PAR obtained for the MC-NPC, MC-DINA, NPC, and DINA Methods.

α Structure	N	100 Replications			C Replications	
		MC-NPC	NPC	DINA	MC-NPC*	MC-DINA (C)
Unif	20	0.869	0.695	0.618	0.867	0.457 (21)
	30	0.895	0.738	0.658	0.894	0.789 (77)
	50	0.891	0.733	0.663	0.891	0.879 (100)
	100	0.884	0.718	0.683	0.884	0.868 (100)
	1,000	0.884	0.728	0.729	0.884	0.907 (100)
MVN	20	0.828	0.647	0.619	0.840	0.380 (15)
	30	0.801	0.607	0.588	0.803	0.753 (88)
	50	0.871	0.702	0.671	0.872	0.850 (99)
	100	0.835	0.666	0.679	0.835	0.852 (100)
	1,000	0.843	0.671	0.706	0.843	0.878 (100)

Note. “C” identifies the number of replications with successful convergence. The column MC-NPC* contains the mean PAR of the C replications only. PAR=patternwise agreement rate; MC=multiple-choice; DINA=deterministic inputs, noisy “and” gate; NPC=nonparametric classification.

of the replications the MC-DINA method did not converge. As a remarkable aside, Table 2.5 suggests that the MC-DINA method might lose its advantage of finer-grained examinee classification if samples are small: the PAR of the MC-DINA method are lower than those obtained from the DINA and NPC methods for dichotomous data.

2.4.2 Simulation Study II: Assessment with Random Q-Matrices

Recall that in Study I de la Torre’s (2009) Q-matrix was constrained such that the q-vectors of the coded response options formed a linear hierarchy. In contrast, Study II used randomly generated Q-matrices with the nestedness constraint removed.

2.4.2.1 Design

The experimental factors and evaluation criteria used in Study I were also used in Study II. However, the design of Study II was supplemented by several features that

considerably increased the complexity of the data generating process. The number of response options for the MC item, H_j , was set to 4 or 5. As already mentioned, the Q-matrices were constructed using a random multi-step procedure. First, the q-vectors of the keys were randomly generated so that the Q-matrix formed by the attribute profiles of the keys conformed to the completeness condition for the DINA model (Chiu, Douglas, & Li, 2009). Next, for each item key, the number of coded distractors needed was determined in randomly choosing from the set $\{0, 1, \dots, H_j^* - 1\}$. The q-vector of each coded distractor was selected from among the $2^K - 1$ admissible attribute patterns subject to two constraints: (i) they had to allow for the non-ambiguous identification of an examinee's ideal response; (ii) each of the distractors had to be non-redundant with regard to classifying the examinees. A redundant distractor is one that does not improve the classification of examinees beyond the response options already available for a given item. In a final step, the key and the coded distractors were randomly assigned to H_j^* positions among the total of H_j positions.

Table 2.6 provides an example of a Q-matrix as it was used to generate data with $K = 3$ attributes and $J = 20$ items. First notice that Items 1, 2, 3, 9, 12, 13, and 17 do not have coded distractors. For each item, the first row indicates the option ID and the q-vector of the key. The remaining rows of a particular item contain the q-vectors of the coded distractors. For example, the key of item 6 was assigned to the position of option 2; its q-vector is (1, 1, 0). A single coded distractor was generated and assigned to the position of option 3; its q-vector is (1, 0, 1).

The probability that an examinee in group $\eta_{ij} = l'$ was attracted by response level $X_{ij} = l$ of item j , $P(X_{ij} = l | \eta_{ij} = l')$, was controlled in first generating a parameter λ_j from $\mathcal{U}(0.74, 0.9)$ for each item j . The probability $P(X_{ij} = l | \eta_{ij} = l')$ was then

Table 2.6: A Sample Q-Matrix Used in the Simulated Data ($K = 3$ and $J = 20$)

Item	Option	Attribute			Item	Option	Attribute		
		1	2	3			1	2	3
1	3	1	0	0	11	1	1	0	0
2	1	0	1	0	12	2	1	0	0
3	3	0	0	1	13	2	1	1	1
4	4	1	1	1	14	2	1	0	1
4	1	0	1	1	14	1	0	0	1
4	2	0	1	0	14	3	1	0	0
5	1	1	1	1	14	4	0	1	0
5	3	0	1	0	15	1	1	1	1
6	2	1	1	0	15	4	0	1	1
6	3	1	0	1	16	3	1	1	1
7	1	0	1	1	16	2	1	0	0
7	2	0	0	1	16	4	0	0	1
7	4	1	0	1	17	1	0	0	1
8	4	1	1	1	18	1	1	1	0
8	2	0	1	1	18	2	1	0	0
8	1	0	0	1	19	2	1	0	1
9	1	1	1	1	19	4	1	0	0
10	4	1	0	1	20	3	1	1	1
10	1	1	0	0	20	2	0	0	1
11	4	1	0	1					

determined as

$$P(X_{ij} = l | \eta_{ij} = l') = \begin{cases} \frac{1}{H_j} + \frac{H_j - H_j^* - 1}{H_j} I(l = 0) & \text{if } l' = 0 \\ \lambda_j & \text{if } l' > 0 \text{ and } l = l' \\ \frac{1 - \lambda_j}{H_j - 1} & \text{if } l' > 0 \text{ and } l \neq l'. \end{cases}$$

The uniform distribution $\mathcal{U}(0.74, 0.9)$ was chosen to ensure that the mean of the parameter λ_j was about 0.82—the value used by de la Torre (2009).

2.4.2.2 Results

Table 2.7 reports the mean PAR from the four methods when each item has four options. Recall that when samples are small, the MC-DINA method may not converge in all 100 replicated data sets. Hence, like in Study I two mean PAR were computed for the MC-NPC method: the first was computed across the 100 replications and used to evaluate the performance of the MC-NPC method in comparison with NPC and DINA methods. The second mean PAR was computed only across those replications where MC-DINA converged (marked by “C”) and used to evaluate the performance of the two methods.

Table 2.7: Mean PARs from the MC-NPC, MC-DINA, NPC, and DINA Methods When $H_j = 4$

		N	100 Replications			C Replications	
α	Structure		MC-NPC	NPC	DINA	MC-NPC*	MC-DINA (C)
$K = 3$ $J = 20$	Unif	20	0.862	0.720	0.703	0.868	0.632 (42)
		30	0.934	0.841	0.798	0.935	0.868 (77)
		50	0.872	0.800	0.772	0.872	0.845 (99)
		100	0.915	0.764	0.767	0.915	0.915 (100)
	MVN	20	0.821	0.705	0.704	0.829	0.632 (28)
		30	0.910	0.803	0.779	0.908	0.829 (76)
		50	0.916	0.860	0.863	0.916	0.895 (97)
		100	0.879	0.757	0.753	0.879	0.887 (100)
$K = 5$ $J = 30$	Unif	20	0.771	0.488	0.418	0.650	0.050 (1)
		30	0.827	0.554	0.509	–	NA (0)
		50	0.836	0.612	0.577	–	NA (0)
		100	0.792	0.517	0.491	0.800	0.725 (2)
	MVN	20	0.775	0.547	0.521	0.775	0.700 (2)
		30	0.685	0.494	0.481	0.703	0.669 (12)
		50	0.789	0.600	0.590	0.763	0.757 (7)
		100	0.746	0.604	0.596	0.746	0.733 (100)

Note. “C” identifies the number of replications with successful convergence. The column MC-NPC* contains the mean PAR of the C replications only. PAR=patternwise agreement rate; MC=multiple-choice; DINA=deterministic inputs, noisy “and” gate; NPC=nonparametric classification.

The comparison between the first three columns with 100 replications shows that the MC-NPC method outperforms the NPC and the DINA methods across all conditions, which suggests that careful coding of the distractors substantially improves examinee classification in comparison with dichotomous responses. Observe that the differences in PAR increase when moving from condition $K = 3$ and $J = 20$ to $K = 5$ and $J = 30$.

The last two columns of Table 2.7 report the mean PAR computed for MC-NPC and MC-DINA across the C replications only. The results (labeled as MC-NPC*) confirm that the MC-NPC method is more effective than MC-DINA when samples are small. In reiterating, as observed earlier, the MC-DINA method was plagued by convergence issues that reduced its efficiency considerably. The results further show that the MC-DINA convergence issues were amplified when moving from the $K = 3$ and $J = 20$ conditions to those where $K = 5$ and $J = 30$. Notice if $K = 5$ and $J = 30$, with attribute profiles uniformly distributed, then MC-DINA barely ever converged, while the MC-NPC method remained efficient across all the 100 replications with PARs of at least 0.75.

In Table 2.8 the mean PAR are presented that were obtained from the four methods when each item had five response options. Comparing the results in Tables 2.8 and 2.7 reveals two similarities worth noting. First, in some conditions with small samples, fitting dichotomous responses with the NPC or DINA method led to higher PAR than fitting polytomous responses with the MC-DINA method (e.g., some conditions where $N = 20$ and 30). Second, Tables 2.7 and 2.8 suggest that a larger number of response options (like in Table 2.8) results in higher PAR. However, this finding should not be misinterpreted that a larger H_j automatically increases PAR. It simply means that more response options—if appropriate—may contribute to improved examinee classification. As a concluding remark, the standard deviations (not reported here) of the PAR obtained for MC-NPC and MC-DINA when the latter converged for a replicated data set were substantially smaller for the MC-NPC than for MC-DINA—an indication that the former produces more consistent PAR.

Table 2.8: Mean PARs from the MC-NPC, MC-DINA, NPC, and DINA Methods When $H_j = 5$

		N	100 Replications			C Replications	
α	Structure		MC-NPC	NPC	DINA	MC-NPC*	MC-DINA (C)
$K = 3$ $J = 20$	Unif	20	0.925	0.802	0.730	0.932	0.500 (14)
		30	0.940	0.858	0.831	0.936	0.724 (39)
		50	0.945	0.840	0.815	0.945	0.921 (90)
		100	0.940	0.867	0.861	0.940	0.931 (100)
	MVN	20	0.942	0.803	0.790	0.959	0.268 (17)
		30	0.935	0.827	0.804	0.932	0.708 (29)
		50	0.950	0.878	0.873	0.950	0.937 (87)
		100	0.937	0.845	0.848	0.937	0.933 (100)
$K = 5$ $J = 30$	Unif	20	0.876	0.663	0.611	–	NA (0)
		30	0.901	0.689	0.685	0.888	0.854 (16)
		50	0.853	0.591	0.523	0.875	0.810 (4)
		100	0.910	0.673	0.661	0.910	0.896 (100)
	MVN	20	0.744	0.606	0.604	–	NA (0)
		30	0.850	0.641	0.639	0.861	0.672 (6)
		50	0.845	0.631	0.613	0.847	0.800 (67)
		100	0.881	0.655	0.661	0.881	0.864 (100)

Note. “C” identifies the number of replications with successful convergence. The column MC-NPC* contains the mean PAR of the C replications only. PAR=patternwise agreement rate; MC=multiple-choice; DINA=deterministic inputs, noisy “and” gate; NPC=nonparametric classification.

2.5 Real Data Analysis

2.5.1 Data

Responses from 132 fourth-graders in Taiwan to 30 multiple-choice items on fraction operations were analyzed. Each item had four response options. The items required at most five attributes; their description is provided in Table 2.9.

The test included 15 single-attribute items; hence, completeness of the Q-matrix was guaranteed. It is presented in Table 2.10. For each item, a bundle of rows contains the q-vectors of the coded response options. The first row is the key, the remaining rows the coded distractors.

Table 2.9: The Required Attributes of the Fraction Operations Assessment

Attribute	Description
1	Write, read, do, and compare fractions
2	Identify and convert between proper, improper, and mixed fractions
3	Add or subtract fractions with the same denominator
4	Multiply fractions
5	Translate a word problem into a mathematical equation

Table 2.10: Q-matrix of the Fraction Assessment

Item	Option	α_1	α_2	α_3	α_4	α_5	Item	Option	α_1	α_2	α_3	α_4	α_5
1	1	1	0	0	0	0	18	2	0	0	1	0	0
2	3	1	0	0	0	0	19	1	1	1	0	0	0
3	4	0	1	0	0	0	20	2	0	0	0	1	1
4	2	0	1	0	0	0	21	1	0	1	1	0	1
5	1	0	1	0	0	0	21	4	0	0	1	0	0
6	4	0	1	0	0	0	22	4	0	1	0	1	1
7	1	1	0	0	0	0	22	2	0	1	0	0	0
8	2	1	0	0	0	0	23	3	0	0	1	0	1
9	4	0	0	1	0	0	23	1	0	0	1	0	0
10	1	0	0	0	1	0	23	4	0	0	0	0	1
10	3	0	0	1	0	0	24	1	0	0	0	1	1
11	2	0	0	0	0	1	24	2	0	0	0	0	1
12	3	0	0	0	0	1	25	1	0	1	0	1	1
13	3	0	0	0	0	1	25	4	0	0	0	1	1
14	1	0	0	0	0	1	26	1	1	1	1	0	1
15	2	1	1	0	0	0	27	3	1	1	1	0	1
16	4	1	1	0	0	0	27	1	0	0	1	0	0
16	1	1	0	0	0	0	28	4	0	0	0	0	1
17	1	0	1	1	0	1	29	4	1	1	0	0	1
17	3	0	0	1	0	0	30	1	1	1	1	0	1

Figure 2.1: An Item in the Fraction Assessment

Sarah's backpack weighs $2\frac{1}{9}$ kg and is $\frac{4}{9}$ kg lighter than Amy's. What is the weight of Amy's backpack? Please select the correct expression and the correct answer.

A. $2\frac{1}{9} - \frac{4}{9} = 1\frac{6}{9}$ kg B. $2\frac{1}{9} - \frac{4}{9} = 2\frac{3}{9}$ kg

C. $2\frac{1}{9} - \frac{4}{9} = 2\frac{5}{9}$ kg D. $2\frac{1}{9} - \frac{4}{9} = 2\frac{5}{18}$ kg

Figure 2.1 displays an example item: response option C is the key requiring Attributes 3 and 5. Options A and D are two coded distractors; the former requires Attribute 3, the latter Attribute 5. The 132 examinees belonged to six classes with 27, 24, 23, 23, 21, and 14 students. From the total of $132 \times 30 = 3,960$ responses 31 were missing. They were imputed with a response randomly chosen from among the four options of the item with missing responses. (Notice that listwise deletion of all cases with missing data would have led to a substantial loss of data, which would have derailed the MC-DINA estimation process.

Evaluating a new method with real-world data presents a challenge in so far as the true proficiency class of examinees, different from synthetic data, is never known. Any evaluation can only be based on relative standards. To develop such a relative standard, examinees' attribute profiles were estimated twice: first, at the class level using all four methods, MC-NPC, MC-DINA, NPC, and DINA. In a second step, the class-level data were merged into a single sample and re-analyzed with the MC-NPC and the MC-DINA method to assess their performance with a large sample. The large sample estimates—expected to result in more accurate examinee classification—were used as a benchmark for evaluating examinee classification using MC-NPC, MC-DINA, NPC, and DINA with small samples as they are common at the classroom level.

As a technical note, the MC-DINA method implemented in the **GDINA** package can only handle items where all response options have been chosen at least once, which

required to exclude several items from the cognitively diagnostic assessment. In case of the merged data set, Items 1 and 20 had to be removed. The analysis at the level of individual classes required the elimination of a larger number of items; they are listed in Table 2.11.

Table 2.11: Items Removed in Class-level Analysis

Class	Items Removed	
	MC-DINA	DINA
1	1, 3-11, 13-15, 17, 18-21, 23-24, 26, 28, 30	1, 6-7, 10, 24
2	1, 3, 5, 8, 11, 17, 20, 23, 26, 29	–
3	1, 4-11, 13, 15, 17-21, 23-25	10, 20, 24
4	1, 3, 8-10, 17, 20, 23-24	–
5	1, 3-11, 14-15, 17-22, 23-24, 26-28	19, 24
6	1-17, 19-20, 22-24, 26, 28, 30	6, 9, 24

2.5.2 Results

The PAR between the estimates of examinees' attribute profiles obtained by analyzing the merged data set and individual classes are reported in Table 2.12, which warrants some explanation. Table 2.12 is a three-way table reporting the results of

Table 2.12: PAR between the Total Examinees Analyzed by the Polytomous Methods and the Class Analyzed by Each Method

		Class1	Class2	Class3	Class4	Class5	Class6	Total
MC-NPC-ALL	MC-NPC	1.000	0.917	0.957	1.000	1.000	0.929	0.970
	MC-DINA	0.444	0.625	0.261	0.391	0.333	0.571	0.432
	NPC	0.963	0.708	0.870	0.739	0.857	0.643	0.811
	DINA	0.296	0.750	0.478	0.522	0.524	0.643	0.523
MC-DINA-ALL	MC-NPC	0.852	0.625	0.609	0.652	0.857	0.500	0.697
	MC-DINA	0.444	0.667	0.304	0.478	0.333	0.571	0.462
	NPC	0.852	0.458	0.696	0.652	0.857	0.571	0.689
	DINA	0.259	0.500	0.522	0.478	0.571	0.643	0.477

a complex consistency analysis of the various methods to evaluate the key hypothesis

that non-parametric methods for analyzing MC items outperform parametric methods if samples are small. The approach presented here is tailored to the specific dilemma of any real-world data analysis: the true proficiency classes of examinees are unknown. In assuming that larger samples result in more accurate examinee classification, the responses collected in the six classes were merged into a sample with 132 students and analyzed with the nonparametric MC-NPC method and the parametric MC-DINA algorithm. These two methods form the two levels of the first dimension of Table 2.12: “MC-NPC-ALL” versus MC-DINA-ALL”. Each of the two levels is completely crossed with the four methods MC-NPC, MC-DINA, NPC, and DINA that were used to classify examinees in the six classes. The third dimension of the table is formed by the six classes. The PAR values listed in the cells of the table were computed in comparing the classification results of the merged sample with those obtained for each of the four methods in the six classes. For example, the PAR in the first row result from comparing the classification of students obtained for each of the six classes with their classification when it was based on the merged sample. The hypothesis was that especially the parametric methods, MC-DINA and DINA, should produce varying classification depending on sample size; the nonparametric methods, MC-NPC and NPC, were supposed to be unaffected by changes in sample size. Consider, for example, the PAR in the first row of Table 2.12 that were all computed in comparing the MC-NPC classification of students based on the merged sample with those obtained in the six classes. As expected, the PAR all indicate perfect agreement (i.e., 1.000). In contrast, and as a further illustration of the principle underlying this specific analysis, consider the sixth row that contains the PAR computed in comparing the MC-DINA classification of the large sample with those realized in the six classes: here, indeed, moving from the large sample to the small class samples severely affected the agreement between the two classifications.

The major findings can be summarized as follows. First, sample size did not seem to affect the performance of the MC-NPC method significantly: it performed consistently well when compared to the nonparametric as well as the parametric methods (recall, for

example, the PAR of the MC-NPC-ALL versus MC-NPC when used in the six classes: all PAR were either 1 or at least 0.9). Second, in contrast, MC-DINA displayed much greater sensitivity to changes in sample size (see the earlier explanation concerning the sixth row of Table 2.12). Third, surprisingly, the two methods for binary responses, NPC and DINA, produced higher PAR than the MC-DINA method for almost all of the six classes. Finally, a brief comment concerning the last column of Table 2.12: it contains the PAR comparing the merged with the stacked classification as they were obtained separately for each of the six classes. Hence, the PAR reported in the last column can be viewed as a weighted mean PAR of the six classes.

Consistency or agreement of classification was also evaluated from a different point of view: does mastery of individual attributes agree with an examinee's estimated attribute profile; that is, her proficiency class? Specifically, for each student, attribute subscores were computed and compared with her estimated attribute profile. Attribute subscores of a given attribute are the percentage of items that require said attribute and that an examinee has answered correctly. In light of the conjunctive framework used by all four methods investigated here, extreme attribute subscores, say above 0.8 and below 0.1 are more informative than mid-range scores because the latter are typically distorted by the conjunctive framework: an examinee may have mastered a particular item, but failed to master all the attributes required for a specific item and did not answer the item correctly; hence, her mastery of that particular item is not included in the count. High and low attribute subscores are then compared with an examinee's estimated attribute profile: high and low attribute subscores should correspond to 1 and 0 entries, respectively, in the estimated attribute profile. Discrepancies in either direction—a 1 where there should be a zero, or vice versa—indicate an inconsistency between actual test performance and estimated proficiency. The students in Class 5 are used as examples here; their attribute profiles were estimated by the MC-NPC and MC-DINA method. Table 2.13 presents a few striking examples of grave discrepancies between attribute subscores and estimated proficiency class.

Table 2.13: Distinct Estimated Attribute Profiles of Class5 With the MC-NPC Method and the MC-DINA Algorithm

ID	MC-NPC		MC-DINA		Subscores				
	All	Class5	All	Class5	α_1	α_2	α_3	α_4	α_5
99	11111	11111	11111	01011	0.833	0.867	1.000	1.000	0.813
100	11011	11011	11011	11100	0.750	0.600	0.500	0.800	0.500
101	11111	11111	11111	00111	0.750	0.733	0.875	0.600	0.750
103	11111	11111	11111	11111	0.833	0.733	1.000	1.000	0.938
105	00010	00010	00010	11000	0.417	0.200	0.375	0.400	0.313
106	11111	11111	11111	10001	0.833	0.667	0.625	0.800	0.688
108	11111	11111	11111	00100	0.917	1.000	1.000	1.000	0.938
109	11111	11111	00000	11110	0.500	0.600	0.750	0.600	0.625
111	11111	11111	11111	10001	0.833	0.733	0.875	0.600	0.750
112	11010	11010	11110	01000	0.667	0.467	0.375	0.600	0.313
113	11111	11111	11111	00011	0.833	0.867	1.000	0.600	0.875
116	11111	11111	11111	00011	0.833	0.733	0.875	0.600	0.750
117	11111	11111	11111	01010	0.833	0.533	0.625	0.800	0.563
118	11111	11111	11111	00111	0.833	0.867	0.875	0.800	0.750

Consider, for example, Examinee 108. All proportions of correct responses to items associated with the five attributes were 100% or at least above 90%. Her attribute profile estimated by the MC-NPC method, (11111), was consistent with that of the subscores; however, the attribute profile estimated by the MC-DINA algorithm, (00100), was severely off. Two other extreme examples are Examinees 100 and 99. The former answered only 50% of the items requiring Attribute 3 correctly but was identified as a master of Attribute 3. In contrast, Examinee 99 answered 100% of these items correctly but was identified as a non-master of Attribute 3. The estimates of the attribute profiles of these two examinees obtained by using the MC-NPC method did not suffer from these grave inconsistencies. The MC-NPC method, however, is not without fault: for example, Examinee 106 only answered 62.5% of the items requiring Attribute 3 correctly but was identified as a master of this attribute. On the other hand, Examinee 113 answered 100% of these items correctly but was identified as a non-mastery of Attribute 3. Still, close inspection of all examinees in the fifth class revealed that overall, MC-DINA resulted in

considerably more inconsistencies than the MC-NPC method. A plausible explanation of this difference seems that many items needed to be deleted to meet the requirement of MC-DINA that each response option must be used at least once.

2.6 Discussion

A multiple-choice item consists of a list of response options comprising the key, the correct answer, and several distractors (i.e., incorrect responses). If designed carefully, additional diagnostic information can be gleaned from the responses to the distractors that improves the classification of examinees (de la Torre, 2009). The design of distractors according to de la Torre's (2009) MC-DINA model is guided by the assumption that distractors lack one of more attributes that are required for a correct response to the key. Said differently, the q-vectors of the distractors are restricted to be nested in that of the key. The implementation of de la Torre's (2009) MC-DINA model in the GDINA package in R, however, does not impose this restriction, as coded distractors are allowed to require attributes not required by the key.

In this article, the MC-NPC method is proposed as a nonparameteric alternative to de la Torre's (2009) MC-DINA model for use with small samples. Like its companion, the NPC method for dichotomous item responses, the MC-NPC method uses a loss function based on penalized Hamming distances (Chiu & Douglas, 2013). However, the loss function had to be re-designed to meet the specific challenges to accommodate the multiple-choice format of observed and ideal item responses. Three major findings from the simulation studies must be mentioned. First, the MC-NPC method can improve the examinee classification rate in comparison with the DINA model and the NPC method applied to dichotomized data. The results further suggest that MC-NPC outperforms MC-DINA in case of small samples. A specific drawback of MC-DINA is its susceptibility to fail convergence when samples are small. In contrast, the MC-NPC method does not rely on fitting a parametric model and is thus, immune to any convergence

issues, which makes it the method of choice for monitoring teaching and learning in small educational programs. In fact, the real-world application concerned the CD of assessment data collected in six classes of an elementary school. The key findings are: (1) sample size did not seem to affect the performance of the MC-NPC method significantly; (2) MC-DINA appeared to be more sensitive to changes in sample size; (3) For almost all of the six classes, NPC and DINA, the two methods for binary responses, produced higher PAR than the MC-DINA method. The MC-NPC method, however, should be used with caution for large samples, as it may be less effective compared to the parameter-based MC-DINA method, as the results of Simulation Study I suggest. As another limitation, the MC-NPC method does not provide item parameter estimates, as it does not rely on fitting the data with a parameterized statistical model.

As a general take home message from the simulation studies, using a valid Q-matrix is mandatory. Great efforts should be made to secure completeness of the Q-matrix to guarantee identifiability of the model (e.g., Gu & Xu, 2018) before any data are collected or analyzed. Thus, the development of a solid validation method for the q-vectors of the keys and the distractors should be prioritized. Within this context, recall that for the simulation studies a Q-matrix of the keys was generated that satisfied Lemma 1 in Chiu, Douglas, and Li (2009). Said lemma states that a Q-matrix is complete for the DINA model if and only if it contains a $K \times K$ identity submatrix formed by the q-vectors of the K distinct single-attribute items. The completeness condition described in Lemma 1 in Chiu et al. (2009) is sufficient and necessary for the NPC and the DINA methods. For MC-NPC and MC-DINA, however, the completeness condition of Lemma 1 is only sufficient. Said differently, the Q-matrix of assessment using MC items and the conjunctive DINA framework may be complete despite some single-attribute items missing from the q-vectors of the keys. As an example, consider a Q-matrix containing only two items with four response options, $H = 4$, and $K = 3$ attributes. Let $\mathbf{q}_1^3 = (110)$, $\mathbf{q}_1^2 = (100)$, and $\mathbf{q}_1^1 = (010)$ be q-vectors of three coded options of Item 1. Item 2 has only a single coded response option, the key, with q-vector $\mathbf{q}_2^1 = (001)$.

Table 2.14 presents the ideal response patterns for each of the 2^K distinct attribute profiles. Obviously, every attribute profile has its own, distinct ideal response pattern.

Table 2.14: Ideal Response Patterns With 2^K Attribute Profiles

Attribute Pattern	Item 1	Item 2
(000)	0	0
(100)	2	0
(010)	1	0
(001)	0	1
(110)	3	0
(101)	2	1
(011)	1	1
(111)	3	1

Therefore, the two items are sufficient to identify all possible attribute profiles. Thus, the Q-matrix formed by the q-vectors of coded response options of Items 1 and 2 is complete. In this example, the Q-matrix of the keys and coded distractors still includes all possible $K = 3$ single attribute q-vectors. However, since the coded distractors provide additional diagnostic information, the Q-matrix can be complete even if the test does not include all single-attribute items.

In conclusion, and closely related, one should also notice that additional (but not necessarily comprehensive) simulation studies (not reported here due to space limitations) suggest that MC-NPC in comparison with MC-DINA is surprisingly “insensitive” to the misspecification of Q-matrix entries—a remarkable finding that certainly deserves further systematic exploration.

Chapter 3

Nonparametric CD-CAT for Multiple-Choice Items: Item Selection Method and Q-Optimality

3.1 Introduction

Cognitive diagnosis (CD) is a psychometric methodology the aim of which to detect examinees' strengths and weaknesses in a particular set of knowledge or skills. This approach provides students, teachers, and other stakeholders with more detailed information about the examinees' current knowledge state, which can then be used to develop more effective interventions and support strategies. In the CD framework, examinees' responses are usually analyzed by CD models (CDMs). However, most applications with CDMs are limited to binary data. Thus, for certain test formats, data must be dichotomized so that they can be fitted with these CDMs, which, as a consequence, may cause information loss. Multiple-choice (MC) items, a common format in educational

assessment, are a case in point. With a careful design, MC items can offer diagnostic insights not only through the information provided from the correct answer (i.e., key) but also that embedded within the incorrect options (i.e., distractors). Unfortunately, when traditional CDMs are used, the responses to all the distractors are lumped together as the incorrect response without further analysis, wasting the diagnostic information underlying the distractors. In response to this issue, the deterministic inputs, noisy “and” gate model for MC items (MC-DINA; de la Torre, 2009) was the first genuine method proposed for analyzing MC items in CD. Research has shown that it can substantially improve the estimation of examinees’ attribute profiles compared to that obtained from CDMs for dichotomized data. However, the MC-DINA model exhibits difficulties when analyzing data from small samples. Not only does its performance decline, but it may also encounter convergence issues, thereby limiting its applicability in micro-educational settings. In this regard, the nonparametric classification method for MC items (MC-NPC; Wang et al., 2023) was developed as a nonparametric counterpart that estimates examinees’ attribute profiles without the need for model fitting and, thus, is more suitable for small samples.

To further enhance measurement efficiency and accuracy in assessments, methods for computerized adaptive testing for CD (CD-CAT) have been developed to select items specifically tailored to each examinee’s ability, providing a more precise and reliable estimate of his/her attribute profile. Numerous item selection methods for binary responses have been developed (e.g., Chang et al., 2019; Chiu & Chang, 2021; Kaplan et al., 2015). However, to date, the only attempt to consider the information underlying distractors for item selection is the application of the Jensen-Shannon divergence (JSD) index to the MC responses conforming to the MC-DINA model (Yigit et al., 2019). It was demonstrated in their study that utilizing the JSD index to select MC items significantly improved classification accuracy in comparison to methods designed for dichotomous responses. Nevertheless, the implementation of the JSD index in small-sample settings remains challenging. The main reason is that it relies on the calibrated

item parameters estimated by the MC-DINA model, and consequently, the limitations of the MC-DINA model discussed earlier will carry over to the context of CD-CAT with the JSD index. To address this challenge, the nonparametric item selection method for multiple-choice items (MC-NPS) that incorporates the MC-NPC method as the classification vehicle is proposed in this study. The MC-NPS method is characterized by the prominent feature that it does not require item calibration, making it applicable with samples of any size and particularly suitable when calibration samples are small or unavailable. Another significant contribution of this study is the development of a Q-optimal procedure exclusively for CD-CAT with MC items. This procedure can improve the classification accuracy during the initial cycles of any CD-CAT algorithm for MC items. In this study, the proposed MC Q-optimal procedure is integrated into the MC-NPS algorithm.

The remainder of this paper is laid out by first reviewing the methods employed in the study, followed by the elaboration of the proposed MC-NPS method and the Q-optimal procedure for MC items. Two simulation studies are conducted to assess the effectiveness and efficiency of the MC Q-optimal procedure and the performance of the MC-NPS method, in contrast to existing CD-CAT algorithms. The paper is wrapped up by discussing some insights, implications, and limitations of the proposed methods.

3.2 Technical Background

In the CD framework, the cognitive skills or knowledge is referred to as attributes. Suppose a test consists of J items requiring at most K attributes. Assuming that there is no hierarchy among the attributes, the K attributes constitute 2^K possible latent classes, denoted as C_m , where $m = 1, \dots, 2^K$. Each latent class is characterized by a K -dimensional binary attribute profile, denoted as $\boldsymbol{\alpha}_m = (\alpha_{m1}, \alpha_{m2}, \dots, \alpha_{mK})'$. Specifically, $\alpha_{mk} = 1$, indicating that examinees in latent class m have mastered the k th attribute and $\alpha_{mk} = 0$, otherwise. The attribute profile of examinee i is labeled as

α_i . Hence, if the examinee belongs to latent class C_m , then $\alpha_i = \alpha_m$.

The Q-matrix (Tatsuoka, 1985) plays a pivotal role in all CD analysis as it indicates the association between the attributes and items. Let $\mathbf{Q}_B$ be the Q-matrix for binary items. $\mathbf{Q}_B$ has dimensions $J \times K$, with $q_{jk} = 1$ if item j measures attribute k , and $q_{jk} = 0$, otherwise. In the context of MC items, not only is the key but also a subset of distractors are coded with their own q-vectors, and these options are referred to as coded options. Hence, the Q-matrix has to be augmented accordingly. Let H_j^* be the number of coded options of item j . Then the Q-matrix for a test of J MC items, denoted as $\mathbf{Q}_{MC}$, is a $(\sum_{j=1}^J H_j^*) \times K$ matrix and can be broken into J sub-matrices, expressed as

$$\mathbf{Q}_{MC} = \begin{pmatrix} \mathbf{Q}_1 \\ \vdots \\ \mathbf{Q}_J \end{pmatrix}.$$

Two issues may occur when constructing the Q-matrix for MC items. First, distractors may become redundant: they do not improve the classification of examinees beyond the response options already available for a given item. Second, undesirable diagnostic ambiguity can arise from distractors that are equally likely to be chosen by an examinee, but have distinct attribute profiles pointing at different diagnostic classifications. In response to these issues, Köhn et al. (2023) proposed two criteria, “useful” and “proper,” for detecting such problematic cases. An item is called “useful” if and only if the q-vector of its key is not nested within any of the q-vectors of the distractors. An item is called “proper” if and only if (i) the q-vector of the key is equal to or nested within the union of the q-vectors of two distractors, or (ii) the union of the q-vectors of two distractors is itself a distractor. In the studies reported in the manuscript, the items satisfy these two criteria; that is, they are assumed to be “proper” and “useful.”

3.2.1 The MC-NPC Method

The MC-NPC method (Wang et al., 2023) applies the penalized Hamming distance to classify examinees by incorporating information underlying coded distractors. Let $\mathbf{X}_i = (X_{i1}, X_{i2}, \dots, X_{iJ})'$ be the observed responses from examinee i to J items. Assuming that each item has H_j options, then response X_{ij} takes the values $h = 1, \dots, H_j$. Since the identification of options in a test, h , does not convey any information about the required attributes, the observed response X_{ij} is further coded as l , where $l = 0, 1, 2, \dots, H_j^*$. Let $\mathbf{q}_j^{(l)}$ be the q-vector of the coded option at level l . All noncoded options are assigned the value $l = 0$, and each of the coded options is assigned a unique value, with the key being coded at the highest level; namely, the q-vector of the key is denoted as $\mathbf{q}_j^{(H_j^*)}$. The ideal response to item j from examinee i , η_{ij} , is then formulated as follows:

$$\eta_{ij} = \max_{l=0,1,2,\dots,H_j^*} \{l \prod_{k=1}^K I[\alpha_{ik} \geq q_{jk}^{(l)}]\}.$$

After the examinees' ideal responses are determined, the penalized Hamming distance between the observed and ideal response vectors is used as the loss function, expressed as

$$d_p(\mathbf{X}_i, \boldsymbol{\eta}(\boldsymbol{\alpha}_m)) = \sum_{j=1}^J I[\eta_j(\boldsymbol{\alpha}_m) > 0, X_{ij} \neq \eta_j(\boldsymbol{\alpha}_m)] + \sum_{j=1}^J w_j I[\eta_j(\boldsymbol{\alpha}_m) = 0, X_{ij} \neq \eta_j(\boldsymbol{\alpha}_m)], \quad (3.1)$$

where w_j is the penalty weight, defined as:

$$w_j = \begin{cases} 1 - \frac{H_j^* - 1}{H_j} & H_j^* < H_j \\ 0 & H_j^* = H_j. \end{cases}$$

The attribute profile of examinee i is then obtained by minimizing $d_p(\mathbf{X}_i, \boldsymbol{\eta}(\boldsymbol{\alpha}_m))$ in Equation 3.1. Simulation studies have shown that the MC-NPC method results in more

accurate classifications in contrast to the MC-DINA model when the sample size is small and outperforms dichotomous CD methods across all sample sizes.

3.2.2 Parametric CD-CAT for MC Responses: The JSD Index

Yigit et al. (2019) introduced the JSD index, which measures the similarity between two probability distributions to select the best MC items for each examinee when data conform to the MC-DINA model. The item response function (IRF) of the MC-DINA model, $P(X_{ij} = h | \alpha_i)$, was elaborated in Wang et al. (2023) as follows:

$$P(X_{ij} = l | \alpha_i) = \begin{cases} \frac{1}{H_j} + \frac{H_j - H_j^* - 1}{H_j} I[l = 0] & \text{if } \eta_{ij} = 0 \\ \left(1 - \sum_{c \neq l} \epsilon_{jcl}\right)^{I[\eta_{ij}=l, l>0]} \prod_{l' \neq l} \epsilon_{jll'}^{I[\eta_{ij}=l']} & \text{if } \eta_{ij} > 0, \end{cases} \quad (3.2)$$

subject to $\sum_{c \neq l} \epsilon_{jcl} < .5$ for all items. Here, ϵ_{jcl} is the probability that the observed response level c disagrees with the ideal response level l . Furthermore, the JSD index for item j is computed as

$$\text{JSD}_j = S(\mathbf{P}_j \times \boldsymbol{\pi}') - \sum_m^{2^K} \pi_m S(\mathbf{P}_{jm}).$$

In the above equation, $\mathbf{P}_j$ denotes an $H_j \times 2^K$ matrix, where $\mathbf{P}_{jm} = (P_{j1m}, P_{j2m}, \dots, P_{jhm})'$ is the m th column of $\mathbf{P}_j$ and $P_{jhm} = P(X_{mj} = h | \alpha_m)$ is the h th entry of $\mathbf{P}_j$. In addition, $\boldsymbol{\pi} = (\pi_1, \pi_2, \dots, \pi_{2^K})'$ denotes the posterior distribution weights, where $\pi_m \equiv \pi(\alpha_m)$, and $S(\cdot)$ represents the Shannon entropy (SHE) computed as $S(\mathbf{P}_{jc}) = E[-\ln(\mathbf{P}_{jc})]$.

During the CAT procedure, the first item is randomly selected and administered to the examinee. Then the examinee's attribute profile can be estimated based on his/her response to the item. In each of the following cycles, the best item is selected by maximizing the JSD index, and the examinee's attribute profile is updated using the augmented response vector. The algorithm continues until it meets the stopping rule.

The simulation studies have demonstrated that by incorporating the information from the coded distractors, the JSD index significantly improves the classification accuracy in comparison to the generalized DINA model discrimination (GDI; Kaplan et al., 2015), and can reach high classification rates with very short tests.

3.2.3 Nonparametric CD-CAT for Dichotomized Responses: The GNPS Method

The general nonparametric item selection method (GNPS; Chiu & Chang, 2021) utilizes a nonparametric discrimination power to select binary items and then employs the general nonparametric classification (GNPC) method (Chiu et al., 2018; Chiu & Köhn, 2019) to estimate the examinee's attribute profile. The (estimated) discrimination power is formulated as

$$\hat{\delta}_j(\boldsymbol{\alpha}_m, \boldsymbol{\alpha}_{m'}) = d(\hat{\eta}_j^{(w)}(\boldsymbol{\alpha}_{m'}), \hat{\eta}_j^{(w)}(\boldsymbol{\alpha}_m)). \quad (3.3)$$

Here, $\eta^{(w)}$ is a weighted ideal response, defining the convex combination of the conjunctive and disjunctive ideal responses in the GNPC method, and $d(\hat{\eta}_j^{(w)}(\boldsymbol{\alpha}_{m'}), \hat{\eta}_j^{(w)}(\boldsymbol{\alpha}_m))$ is the Euclidean distance between the estimated ideal responses corresponding to the true attribute profile, $\boldsymbol{\alpha}_m$, and another attribute profile, $\boldsymbol{\alpha}_{m'}$.

The GNPS algorithm begins by randomly selecting the first item for an examinee and using his/her response to estimate the attribute profile. During each of the following cycles, the discrimination power of the remaining items in the item bank is calculated for the two most possible latent classes, and the item with the maximum discrimination power is selected and administered. After the new response is collected, the examinee's attribute profile is immediately updated. The algorithm iteratively proceeds until the stopping rule is met. The GNPS method can be used when the data conforms to any CDM for binary responses, and outperforms the major parametric methods when the calibration sample is small.

3.3 The Proposed MC-NPS Method and Q-Optimal Procedure for MC Items

The MC-NPS method is first proposed in this study to select the “best” MC items for each examinee in the context of CD-CAT by taking into account diagnostic information from both the key and distractors. In each CAT cycle, the best item is chosen by maximizing the proposed distance-based discrimination power, and the MC-NPC method is used to estimate the examinee’s attribute profile. The two phases (i.e., item selection and attribute profile estimation) alternate until the stopping criterion is met. The second goal of this study is to propose a Q-optimal procedure for MC items, which can be used to enhance the performance of a CD-CAT algorithm during the initial cycles. A CD-CAT algorithm usually begins by randomly selecting an item, which may result in sub-optimal performance during the initial phase. Xu et al. (2016) developed a non-asymptotic theorem for dichotomous items, outlining a procedure that ensures the optimal classification after administering the first K items of the CAT algorithm. In this study, a Q-optimal procedure for MC items that is inspired by theorems presented in Xu et al. (2016) but featuring a different yet novel design is proposed. In the last part of this section, a CD-CAT algorithm that integrates the Q-optimal procedure for MC items into the MC-NPS method is presented. Note that in this study, data are assumed to conform to the MC-DINA model.

3.3.1 The Proposed MC-NPS Method

The objective of the MC-NPS method, like other item selection methods for CD-CAT, is to choose items that can most effectively differentiate the estimated attribute profile from the others. The discrimination power proposed in this study, building upon the concept used in the GNPS method but customized to the context of MC items, is employed for item selection.

Within an MC item, the q-vectors of the coded options are partially ordered in

nature, and as a result, the options are treated as nominal rather than ordinal. Recall that the discrimination power for the GNPS method, as defined in Equation 3.3, utilizes the Euclidean distance between the estimated ideal responses of two latent classes. A similar concept is applied here; however, the Hamming distance is employed since each individual level of coded option is treated as a binary entry and independent of the other levels. Suppose $\eta_j(\alpha_m) = l_{mj}$ and $\eta_j(\alpha_{m'}) = l_{m'j}$, where $\alpha_m \neq \alpha_{m'}$. If $l_{mj} \neq l_{m'j}$ and both of them are greater than 0, according to the IRF of the MC-DINA model formulated in Equation 3.2, the observed response of an examinee is highly likely to match his/her nonzero ideal response. In other words, it is assumed that $P(X_{m'j} = l_{m'j} | \alpha_{m'}) > .5$ and $P(X_{mj} = l_{mj} | \alpha_m) > .5$, whereas $P(X_{m'j} = l_{mj} | \alpha_{m'}) < .5$ and $P(X_{mj} = l_{m'j} | \alpha_m) < .5$. Furthermore, $P(X_{mj} = l_{mj} | \alpha_m)$ is independent of the value of l_{mj} . Consequently, the discrimination power is not affected by the numeric difference between $\eta_j(\alpha_{m'})$ and $\eta_j(\alpha_m)$, and the indicator function, $I[\cdot]$, is used to calculate the discrimination power. To demonstrate, suppose for item j , $\eta_j(\alpha_{m'}) = 4$ and $\eta_j(\alpha_m) = 3$. For item j' , $\eta_{j'}(\alpha_{m'}) = 4$, and $\eta_{j'}(\alpha_m) = 2$. The discrimination power only takes into account the agreement/disagreement rather than the actual difference between $\eta_j(\alpha_{m'})$ and $\eta_j(\alpha_m)$. Hence, for $\alpha_{m'}$ and α_m , the discrimination power of item j is identical to that of item j' because $\delta_j(\alpha_m, \alpha_{m'}) = I[\eta_j(\alpha_m) \neq \eta_j(\alpha_{m'})] = 1$ and $\delta_{j'}(\alpha_m, \alpha_{m'}) = I[\eta_{j'}(\alpha_m) \neq \eta_{j'}(\alpha_{m'})] = 1$.

However, when one of the two distinct ideal responses is zero, a penalty needs to be applied to the discrimination power due to the random selection associated with a zero ideal response. Without loss of generality, assume that for item j' , $\eta_{j'}(\alpha_m) = l_{mj'} = 0$ and $\eta_{j'}(\alpha_{m'}) = l_{m'j'} > 0$. In this case, the probability that $l_{m'j'}$ is selected by an examinee whose attribute profile is α_m is $P(X_{mj'} = l_{m'j'} | \eta_{j'}(\alpha_m) = 0) = \frac{1}{H_{j'}}$. Notice that this probability is a constant and relatively large compared to $P(X_{mj} = l_{m'j} | \eta_j(\alpha_m) > 0) = \epsilon_{jl'l}$, indicating that the observed response of j' is more likely to fail in distinguishing between α_m and $\alpha_{m'}$. As a result, the discrimination power of

item j' is lower than that of item j , and, therefore, should be penalized. In the MC-NPS framework, the penalty is set to be $w_{j'}^{(s)} = 1 - \frac{1}{H_{j'}}$ with the notion that the largest possible difference between $\frac{1}{H_{j'}}$ and $\epsilon_{jl'l}$ is $\frac{1}{H_{j'}}$. Furthermore, when all options are coded, $\mathbf{q}_{j'}^{(0)}$ does not occur. Therefore, it is impossible for $\boldsymbol{\alpha}_m$, whose $\eta_{j'}(\boldsymbol{\alpha}_m) = 0$, to find a match. In other words, $P(X_{mj'} = 0 | \eta_{j'}(\boldsymbol{\alpha}_m) = 0) = 0$. Hence, the discrimination power of item j' should reduce to 0 by using $w_{j'}^{(s)} = 0$. Based on the above rationale, the discrimination power for MC items conforming to the MC-DINA model is defined as follows:

$$\delta_j(\boldsymbol{\alpha}_m, \boldsymbol{\alpha}_{m'}) = d_j(\eta(\boldsymbol{\alpha}_m), \eta(\boldsymbol{\alpha}_{m'})) = w_j^{(s)} I[\eta_j(\boldsymbol{\alpha}_{m'}) \neq \eta_j(\boldsymbol{\alpha}_m)],$$

where the $w_j^{(s)}$ is given by

$$w_j^{(s)} = \begin{cases} 1 & \text{if } (\eta_j(\boldsymbol{\alpha}_{m'}) \neq 0) \wedge (\eta_j(\boldsymbol{\alpha}_m) \neq 0) \\ (1 - \frac{1}{H_j}) I[H_j^* < H_j] & \text{if } (\eta_j(\boldsymbol{\alpha}_{m'}) \neq 0) \wedge (\eta_j(\boldsymbol{\alpha}_m) = 0) \vee (\eta_j(\boldsymbol{\alpha}_{m'}) = 0) \wedge (\eta_j(\boldsymbol{\alpha}_m) \neq 0) \\ 0 & \text{if } H_j^* = H_j. \end{cases}$$

Finally, when $l_{mj} = l_{m'j}$, the item is unable to differentiate between the two distinct attribute profiles. Hence, its discrimination power is assigned a value of zero. In conclusion, the discrimination power used in the MC-NPS method can also be expressed as follows:

$$\delta_j(\boldsymbol{\alpha}_m, \boldsymbol{\alpha}_{m'}) = \begin{cases} 1 & \text{if } (\eta_j(\boldsymbol{\alpha}_{m'}) \neq 0) \wedge (\eta_j(\boldsymbol{\alpha}_m) \neq 0) \wedge (\eta_j(\boldsymbol{\alpha}_{m'}) \neq \eta_j(\boldsymbol{\alpha}_m)) \\ (1 - \frac{1}{H_j}) I[H_j^* < H_j] & \text{if } (\eta_j(\boldsymbol{\alpha}_{m'}) \neq 0) \wedge (\eta_j(\boldsymbol{\alpha}_m) = 0) \vee (\eta_j(\boldsymbol{\alpha}_{m'}) = 0) \wedge (\eta_j(\boldsymbol{\alpha}_m) \neq 0) \\ 0 & \text{if } \eta_j(\boldsymbol{\alpha}_{m'}) = \eta_j(\boldsymbol{\alpha}_m) \text{ or } H_j^* = H_j. \end{cases}$$

As shown in the above equation, the novel aspect of the MC-NPS method lies in its distance-based discrimination power, which is free of item calibration, and thus, the MC-NPS method is well suited for settings with small or no calibration samples.

3.3.2 Q-optimal Procedure for CD-CAT with MC Items

The non-asymptotic Q-optimal algorithm proposed by Xu et al. (2016) enhances the classification accuracy of CD-CAT with dichotomous responses by optimizing the classification after administering the initial items. In their study, for a test consisting

of K binary items, the Q-matrix $\mathbf{Q}_B$ is Q-optimal if it satisfies the condition:

$$P(\hat{\alpha} = \alpha_m | \alpha_m, \mathbf{Q}_B) > \max_{\alpha_{m'} \neq \alpha_m} P(\hat{\alpha} = \alpha_{m'} | \alpha_m, \mathbf{Q}_B). \quad (3.4)$$

It was demonstrated that when data conform to the DINA model (Junker & Sijtsma, 2001) with the slipping and guessing parameters less than .5, a test consisting of K items is Q-optimal if and only if, for each examinee, the Q-matrix of the test items takes the following form after column swapping:

$$\mathbf{Q}_B = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ q_{2,1} & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ q_{K-1,1} & \dots & q_{K-1,K-2} & 1 & 0 \\ q_{K,1} & \dots & q_{K,K-2} & q_{K,K-1} & 1 \end{pmatrix}, \quad (3.5)$$

where the lower diagonal element, $q_{j,k}, j = k + 1, \dots, K$, is set to 0 if the examinee answers the k th item incorrectly and either 0 or 1 if the k th item is answered correctly. To demonstrate, consider the following example with $K = 3$. Suppose an examinee answers the first two items correctly. Then $\mathbf{Q}_B$ shown in the following equation is one of the possible Q-matrices that satisfy the Q-optimal conditions:

$$\mathbf{Q}_B = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}.$$

In this case, there are two possible response patterns, (110) and (111). The probabilities for the 2^K latent classes resulting in these two patterns are tabulated in Table 3.1. The results show that if an examinee answers the first two items correctly, then $\hat{\alpha} = (110)$ and $\hat{\alpha} = (111)$ are the two most likely attribute patterns among the 2^K possible attribute patterns, as they have the highest probability $(1 - s_1)(1 - s_2)$, given that $s, g < .5$.

Table 3.1: Probabilities of Response Patterns (110) and (111)

α	Response Patterns	
	(110)	(111)
(000)	$g_1 g_2 (1 - g_3)$	$g_1 g_2 g_3$
(100)	$(1 - s_1) g_2 (1 - g_3)$	$(1 - s_1) g_2 g_3$
(010)	$g_1 (1 - s_2) (1 - g_3)$	$g_1 (1 - s_2) g_3$
(001)	$g_1 g_2 (1 - g_3)$	$g_1 g_2 g_3$
(110)	$(1 - s_1) (1 - s_2) (1 - g_3)$	$(1 - s_1) (1 - s_2) g_3$
(101)	$(1 - s_1) g_2 (1 - g_3)$	$(1 - s_1) g_2 g_3$
(011)	$g_1 (1 - s_2) (1 - g_3)$	$g_1 (1 - s_2) g_3$
(111)	$(1 - s_1) (1 - s_2) s_3$	$(1 - s_1) (1 - s_2) (1 - s_3)$

(Recall that said examinee very likely must have mastered attributes α_1 and α_2 , as they have answered the two items correctly.) If an examinee also answers the third item correctly, then $\hat{\alpha} = (111)$ is the most likely attribute pattern having the highest probability $(1 - s_1)(1 - s_2)(1 - s_3)$; otherwise, $\hat{\alpha} = (110)$ is the most likely attribute pattern with the highest probability $(1 - s_1)(1 - s_2)(1 - g_3)$. Hence, $\mathbf{Q}_B$ is Q-optimal for $\alpha_m = (110)$ or (111) with no ambiguity. Given that binary items are diagnostically less informative than MC items, the Q-optimal procedure for binary items is no longer the optimal design for MC items. To bridge the gap, a Q-optimal procedure tailored for the CD-CAT with MC items is proposed in the study and elaborated in what follows.

The development of a Q-optimal procedure for MC items takes into account the following key factors: first, the completeness of a Q-matrix (Chiu et al., 2009; Köhn & Chiu, 2016) to guarantee the identifiability of all latent classes, with a particular focus on the class defined by $\mathbf{0}_K$; second, the Q-optimal procedure for binary items developed by Xu et al. (2016); and third, the proper and useful properties for MC items proposed by Köhn et al. (2023). Specifically, let $\mathbf{e}_k$ be a $1 \times K$ vector with the k th entry being 1 and all other entries 0. Including all $\mathbf{e}_k$, where $k = 1, \dots, K$, in a test is both sufficient and necessary to identify examinees whose attribute profile is $\mathbf{0}_K$ (Chiu et al., 2009; Köhn & Chiu, 2016; Xu et al., 2016). This concept implies that $\mathbf{e}_k$ is required to determine the k th entry of $\mathbf{0}_K$. Although other latent classes can also be identified by

including all e_k in the test, this condition is only sufficient but not necessary. Therefore, if a Q-matrix is sufficient and necessary to identify all possible latent classes, it must be sufficient and necessary to identify $\mathbf{0}_K$. Once the latent class $\mathbf{0}_K$ is identified, all the others can be identified as well. For this reason, the following discussion will focus on the determination of the entries of $\mathbf{0}_K$. In the case of dichotomous responses, one item can contain at most one e_k . Therefore, a set of K items defined by all possible e_k is sufficient to identify $\mathbf{0}_K$. In the case of MC items, multiple entries of $\mathbf{0}_k$ can be determined by one single MC item because an MC item can include more than one coded option, each requiring a distinct q-vector. Hence, an MC item can include more than one e_k to determine several entries of $\mathbf{0}_K$ concurrently. For example, to determine two entries of $\mathbf{0}_K$ by one MC item (and thus, requiring two e_k), the structure presented in the equation below requires a minimum of three coded options:

$$\begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \end{pmatrix}. \quad (3.6)$$

However, governed by the useful and proper criteria (Köhn et al., 2023), a constraint needs to be imposed to regulate the association between the number of e_k and H_j^* . Specifically, let the number of e_k be N_e . Because the union of each pair of coded distractors must exist in $\mathbf{Q}_j$ (i.e., the “proper” criteria), the total number of unions these e_k can produce equals $\binom{N_e}{2} + \binom{N_e}{3} + \dots + \binom{N_e}{N_e}$. Hence, the total number of e_k and their unions is $\sum_{r=1}^{N_e} \binom{N_e}{r} = 2^{N_e} - 1$. Because this number cannot exceed H_j^* , the constraint $2^{N_e} - 1 \leq H_j^*$ has to be imposed. Table 3.2 displays the minimum H_j^* required to include a certain number of e_k in an MC item. Based on Table 3.2, an item intended to concurrently measure three attributes of $\mathbf{0}_K$ must have at least seven coded options to guarantee that each coded option is useful and proper, which is less common in practical assessments. Hence, the current study will only focus on the case of $N_e = 2$ with its usefulness and applicability elaborated below.

Table 3.2: Relationship Between N_e and Minimum H_j^*

N_e	Minimum H_j^*
1	1
2	3
3	7
4	15

According to Equation 3.4, $P(X_{mj} = l | \eta_{mj} = l) > P(X_{mj} = l | \eta_{mj} = l')$ is met for a Q-optimal procedure, where $l \neq l'$. When $l > 0$, this condition is naturally satisfied due to the assumption of the MC-DINA model (i.e., $\sum_{c \neq l} \epsilon_{jcl} < .5$). However, when $l = 0$, the relationship between $P(X_{mj} = 0 | \eta_{mj} = 0) = \frac{H_j - H_j^*}{H_j}$ and $P(X_{mj} = 0 | \eta_{mj} = l')$ is undetermined because $\frac{H_j - H_j^*}{H_j}$ may not always be greater than or equal to .5. Therefore, to ensure $P(X_{mj} = 0 | \eta_{mj} = 0) > P(X_{mj} = 0 | \eta_{mj} = l')$ holds, it must be true that $\frac{H_j - 3}{H_j} \geq .5$, which is equivalent to $H_j \geq 6$, given $N_e = 2$. However, H_j is restricted to be below 14 because that is the minimal number of H_j required when $N_e = 3$, which is beyond the discussion of the current study. Furthermore, when $H_j = 5$, although $P(X_{mj} = 0 | \eta_{mj} = 0) = .4$, $P(X_{mj} = 0 | \eta_{mj} = l') = \epsilon_{j0l'}$ is always less than .4 with medium or high item quality. Hence, it is justified to assume that $P(X_{mj} = l | \eta_{mj} = l) > P(X_{mj} = l | \eta_{mj} = l')$ holds. In summary, the proposed Q-optimal procedure is applicable to items with $5 \leq H_j < 14$, a spectrum that is widely assumed in educational and psychological settings.

To simply and clearly illustrate the Q-optimal procedure for MC items with $5 \leq H_j < 14$, take a test requiring five attributes ($K = 5$) as an example. The optimal classification can be achieved if the first three items conform to the following design:

$$\mathbf{Q}_{\text{MC}} = \begin{pmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \\ \mathbf{Q}_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hline q_{2,1}^{(3)} & q_{2,2}^{(3)} & 1 & 1 & 0 \\ q_{2,1}^{(2)} & q_{2,2}^{(2)} & 1 & 0 & 0 \\ q_{2,1}^{(1)} & q_{2,2}^{(1)} & 0 & 1 & 0 \\ \hline q_{3,1}^{(H_3^*)} & q_{3,2}^{(H_3^*)} & q_{3,3}^{(H_3^*)} & q_{3,4}^{(H_3^*)} & 1 \end{pmatrix} \quad (3.7)$$

or

$$\mathbf{Q}_{\text{MC}} = \begin{pmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \\ \mathbf{Q}_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hline q_{2,1}^{(3)} & q_{2,2}^{(3)} & 1 & 1 & 0 \\ q_{2,1}^{(2)} & q_{2,2}^{(2)} & 1 & 0 & 0 \\ q_{2,1}^{(1)} & q_{2,2}^{(1)} & 0 & 1 & 0 \\ \hline * & * & * & * & 1 \\ q_{3,1}^{(l)} & q_{3,2}^{(l)} & q_{3,3}^{(l)} & q_{3,4}^{(l)} & 1 \end{pmatrix}. \quad (3.8)$$

In the above matrix, $q_{j,k}^{(l)}$, where $j = 2, 3$, $h = 1, \dots, 3$, and $k = 1, \dots, 4$, takes the value 0 if the option requiring the corresponding attribute k of item $\lceil \frac{k}{2} \rceil$ is not selected, and it takes the value 0 or 1 if that option is selected. In addition, $*$ can be either 0 or 1, as long as all coded options are useful and proper (Köhn et al., 2023). In matrices shown in Equations 3.7 and 3.8, the variation is made on $\mathbf{Q}_3$, while $\mathbf{Q}_1$ and $\mathbf{Q}_2$ remain the same. The levels of the key and coded distractor of item 3, as shown in Equations 3.7 and 3.8, respectively, are denoted in a general manner using H_j^* and l , because item 3 may contain additional coded distractors that make no contribution to the Q-optimal design and can be omitted. With this design, item 1 is used to distinguish between latent class $\mathbf{0}_5$, $\mathbf{e}_1$, and $\mathbf{e}_2$. Similarly, item 2 is used to distinguish between $\mathbf{0}_5$, $\mathbf{e}_3$, and

e_4 . Item 3 is used to distinguish between $\mathbf{0}_5$ and e_5 . As shown in the study by Chiu et al. (2009), these items will also distinguish between the other latent classes while distinguishing $\mathbf{0}_5$ from all the e_k .

Based on the design for $K = 5$, it can easily be inferred that the first two items in Equation 3.7 are sufficient to achieve the optimal design for a test with $K = 4$. As another example, consider $K = 6$. The following items are sufficient to reach the optimal classification with the same rationale:

$$\mathbf{Q}_{\text{MC}} = \begin{pmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \\ \mathbf{Q}_3 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ \hline q_{2,1}^{(3)} & q_{2,2}^{(3)} & 1 & 1 & 0 & 0 \\ q_{2,1}^{(2)} & q_{2,2}^{(2)} & 1 & 0 & 0 & 0 \\ q_{2,1}^{(1)} & q_{2,2}^{(1)} & 0 & 1 & 0 & 0 \\ \hline q_{3,1}^{(3)} & q_{3,2}^{(3)} & q_{3,3}^{(3)} & q_{3,4}^{(3)} & 1 & 1 \\ q_{3,1}^{(2)} & q_{3,2}^{(2)} & q_{3,3}^{(2)} & q_{3,4}^{(2)} & 1 & 0 \\ q_{3,1}^{(1)} & q_{3,2}^{(1)} & q_{3,3}^{(1)} & q_{3,4}^{(1)} & 0 & 1 \end{pmatrix}.$$

Notice that in the case of $K = 4$, only two items are required to reach Q-optimality; when $K = 5$ and 6, three items are needed. That said, with the proposed MC Q-optimal procedure, only $J^* = \lceil \frac{K}{2} \rceil$ items are needed to achieve optimal classification, a phenomenal feature that demonstrates the efficiency and effectiveness of the MC Q-optimal procedure, as opposed to that for binary items, where K items are required. This improvement is primarily owing to the utilization of the additional information embedded in well-designed coded distractors.

Beyond the general discussion above, the case of $H_j \leq 4$ requires special attention. Although a single MC item with three or four options can concurrently involve two e_k , it does not result in an optimal Q-matrix due to the assumption of random selection made by individuals whose ideal response is zero. When $H_j = 3$, the latent group

defined by $\mathbf{0}_K$ will never make an observed response that aligns with the ideal response. As a result, $\mathbf{0}_K$ cannot be distinguished from other latent classes, making the design sub-optimal. When $H_j = 4$, an example is used to illustrate the issue. Consider a test of $K = 4$, forming a total of 16 possible observed response patterns. Suppose that the MC Q-optimal design laid out in Equation 3.7 is adopted. Under the setup, there are three coded and one noncoded options for each item with the corresponding response levels being 3, 2, 1, and 0. Because each item includes only four options, it is true that $P(X_{ij} = 0|\mathbf{0}_K) = P(X_{ij} = 1|\mathbf{0}_K) = P(X_{ij} = 2|\mathbf{0}_K) = P(X_{ij} = 3|\mathbf{0}_K) = \frac{1}{4}$, and thus, examinees associated with $\mathbf{0}_K$ have an equal chance of producing any of the 16 observed response patterns. Hence, in this scenario,

$$P(\hat{\boldsymbol{\alpha}} = \mathbf{0}_K|\mathbf{0}_K, \mathbf{Q}_{MC}) = \max_{\boldsymbol{\alpha}_{m'} \neq \mathbf{0}_K} P(\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}_{m'}|\mathbf{0}_K, \mathbf{Q}_{MC}),$$

which violates the definition of the Q-optimal Q-matrix defined in Equation 3.4. Correspondingly, this Q-matrix, although aligning with Equation 3.7, is not optimal when $H_j = 4$ and consequently the level-one design with one $\mathbf{e}_k$ and minimum $H_j^* = 1$, as shown in Table 3.2, should be used. This design closely resembles the Q-optimal design for binary items and thus, requires K items to achieve Q-optimality. The only difference between these two Q-optimal designs is that the q-vectors specified in Equation 3.5 can define either the key or one of the coded distractors.

In summary, to fully leverage the advantages of MC items with the structure depicted in Equation 3.6, each item should comprise at least five options. Based on Table 3.2, the level-two design (i.e., two $\mathbf{e}_k$ included in one item with the minimum $H_j^* = 3$) does not lead to an optimal classification with $H_j \leq 4$. Accordingly, the more appropriate design would be level one, which closely resembles the Q-optimal procedure for binary items.

3.3.3 Algorithm

As previously discussed, when $H_j = 4$, the Q-optimal procedure for MC items is highly similar to the Q-optimal procedure for binary items, implying that K items are required to reach the optimal classification before selecting the best item based on the discrimination power. Then a crucial decision follows immediately: in the initial stage, is adopting a Q-optimal procedure devised for binary responses that involves administering K items more advantageous than administering the most discriminating items from the beginning of the process without implementing the Q-optimal procedure? Empirical evidence from Simulation Study I shows that administering the most discriminating items at the initial stage of the assessment results in a more precise classification than completing the Q-optimal procedure for binary data. Consequently, when $H_j = 4$, the MS-NPS algorithm will be used from the very beginning with no Q-optimal procedure implemented. When $H_j \geq 5$, the newly proposed Q-optimal procedure is incorporated into the MC-NPS algorithm. The complete algorithm is presented in detail below.

Step 0 Initialize the item pool $R^{(0)} = \{1, \dots, J\}$ and cycle $t = 0$.

Step 1 At cycle $t = 1$, if $H_j = 4$, randomly select one item from $R^{(0)}$; if $H_j \geq 5$, randomly select one item from those whose Q-matrix is $\mathbf{Q}_1$, as defined in Equation 3.7. Denote the item as item j and update the set $R^{(0)}$ by deleting this item; that is, $R^{(1)} = R^{(0)} \setminus \{j\}$.

Step 2 Administer item j to the examinee and initialize the response vector as $\mathbf{X}^{(1)} = (X^{(1)})$.

Step 3 If $H_j = 4$, let $J^* = 1$, and go to Step 6. If $H_j \geq 5$, at cycle $t = s$, where $s = 2, \dots, J^*$, select an item, denoted as item j' , from $R^{(s-1)}$ such that the Q-matrix of item j' is $\mathbf{Q}_s$, defined in Equation 3.7.

Step 4 Administer item j' to the examinee and update the response vector as $\mathbf{X}^{(s)} = (\mathbf{X}^{(s-1)}, X^{(s)})$.

Step 5 Let $s = s + 1$ and repeat Steps 3 and 4 until $s = J^*$.

Step 6 Estimate the examinee's attribute profile using the MC-NPC method with the observed responses $\mathbf{X}^{(J^*)}$ and label the estimate as $\hat{\boldsymbol{\alpha}}^{(J^*)}$. Arrange the distances between $\mathbf{X}^{(J^*)}$ and the other $2^K - 1$ J^* -dimensional ideal response vectors in ascending order, denoted as $d_2^{(J^*)}, \dots, d_{2^K}^{(J^*)}$. Denote the attribute profile corresponding to $d_2^{(J^*)}$ as $\tilde{\boldsymbol{\alpha}}^{(J^*)}$. In case multiple attribute profiles have the same distance $d_2^{(J^*)}$, the one mastering the fewest attributes will be designated $\tilde{\boldsymbol{\alpha}}^{(J^*)}$.

Step 7 At cycle $t = J^* + 1$, choose item j'' from $R^{(t-1)}$ that can most effectively discriminate $\hat{\boldsymbol{\alpha}}^{(t-1)}$ and $\tilde{\boldsymbol{\alpha}}^{(t-1)}$. In other words, choose an item that has the largest $\delta_j(\hat{\boldsymbol{\alpha}}^{(t-1)}, \tilde{\boldsymbol{\alpha}}^{(t-1)})$.

Step 8 Administer item j'' to the examinee and update $R^{(t-1)}$ and $\mathbf{X}^{(t-1)}$ by $R^{(t)} = R^{(t-1)} \setminus \{j''\}$ and $\mathbf{X}^{(t)} = (\mathbf{X}^{(t-1)}, X^{(t)})$.

Step 9 Estimate the examinee's attribute profile using the MC-NPC method with the observed responses $\mathbf{X}^{(t)}$ and label the estimate as $\hat{\boldsymbol{\alpha}}^{(t)}$. Arrange the distances between $\mathbf{X}^{(t)}$ and the other $2^K - 1$ t -dimensional ideal response vectors in ascending order, denoted as $d_2^{(t)}, \dots, d_{2^K}^{(t)}$. Denote the attribute profile corresponding to $d_2^{(t)}$ as $\tilde{\boldsymbol{\alpha}}^{(t)}$. In cases where multiple attribute profiles have the same distance $d_2^{(t)}$, the one mastering the fewest attributes will be designated $\tilde{\boldsymbol{\alpha}}^{(t)}$.

Step 10 Let $t = t + 1$ and repeat steps 7-9 until the stopping rule is met.

3.4 Simulation Studies

Two simulation studies were conducted to examine the performance of the proposed MC-NPS algorithm and the MC Q-optimal procedure. Simulation Study I was designed to evaluate the necessity of applying the proposed MC Q-optimal procedure to the initial phase of a CD-CAT algorithm and its impact on classification, relative to the Q-optimal

procedure devised for binary items (Xu et al., 2016) and the CD-CAT algorithm with no Q-optimal procedure implemented. Simulation Study II aimed to investigate the performance of the proposed MC-NPS algorithm by contrasting it with the JSD and random selection methods with a special focus on small calibration samples. The GNPS method was also used as a baseline to study the advantages of using the item selection methods for MC items in CD-CAT.

3.4.1 Simulation Study I: Evaluating the Effect of the Proposed MC Q-Optimal Procedure

In this simulation study, the effect of the MC Q-optimal algorithm was evaluated by comparing its classification rates and those obtained from the binary Q-optimal procedure and the MC-NPS algorithm without any Q-optimal design.

3.4.1.1 Design

The simulation conditions were created by considering different combinations of the number of options ($H_j = 4, 5$, and 6 across all j) and item quality (medium and high). While H_j was set constant across all items of a test, H_j^* was randomly generated for each j . The probability that examinee i in group $\eta_{ij} = l'$ was attracted by response level $X_{ij} = l$ of item j , $P(X_{ij} = l | \eta_{ij} = l')$, was generated following the MC-DINA framework:

$$P(X_{ij} = l | \eta_{ij} = l') = \begin{cases} \frac{1}{H_j} + \frac{H_j - H_j^* - 1}{H_j} I[l = 0] & \text{if } l' = 0 \\ \lambda_j & \text{if } l' > 0 \text{ and } l = l' \\ \frac{1 - \lambda_j}{H_j - 1} & \text{if } l' > 0 \text{ and } l \neq l', \end{cases} \quad (3.9)$$

where the parameter λ_j represented the probability that an examinee's observed and nonzero ideal responses agreed and was randomly sampled from a uniform distribution $\mathcal{U}(0.77, 0.88)$ and $\mathcal{U}(0.88, 1)$ for items of medium and high quality, respectively. Equation

3.9 implies that the examinee had an equal probability, $\frac{1-\lambda_j}{H_j-1}$, to select one of the other options if slipping occurred. It further shows that if the examinee's ideal response was zero, he/she would make a random selection from all the H_j options.

Additional simulation factors are specified as follows. An item bank consisting of 300 MC items, each requiring at most five attributes, was randomly generated with a fixed proportion of 20% of binary items. The three investigated algorithms, the Q-optimal procedures for MC and binary items and the MC-NPS method without any Q-optimal procedure, were implemented to select the initial five items for a sample of 1,000 examinees, as five items are required by the Q-optimal procedure for binary items. The attribute profiles of these examinees were generated from the multivariate normal threshold model (Chiu et al., 2009), with variances set to 1 and covariances sampled from $\mathcal{U}(0.3, 0.5)$.

3.4.1.2 Measures

The pattern-wise agreement rate (PAR) between the true and estimated attribute profiles was used to evaluate the effectiveness of the compared methods. PAR was calculated at each cycle for each condition using the results from the 1,000 examinees and was formulated as follows:

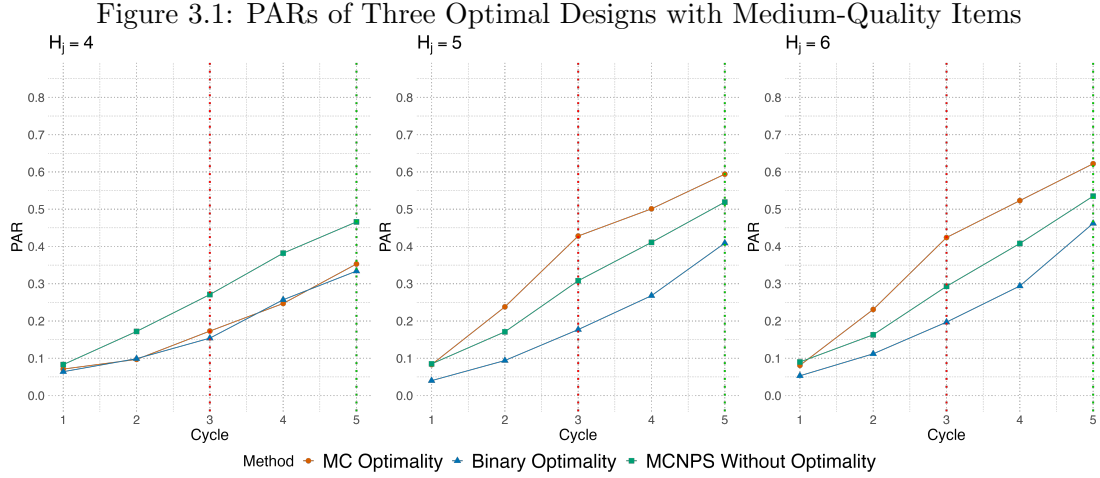
$$\text{PAR} = \frac{\sum_{i=1}^{1000} I[\hat{\alpha}_i = \alpha_i]}{1000}.$$

3.4.1.3 Results

As outlined earlier, when applying the Q-optimal procedure for binary items to MC items, $K = 5$ items were required, regardless of the value of H_j . In contrast, the MC Q-optimal procedure only required $\lceil \frac{K}{2} \rceil = 3$ items for instances with $H_j \geq 5$ to reach the optimal classification. Hence, when $H_j = 5$ or 6, the primary goal of this simulation study is to investigate to what extent the MC Q-optimal procedure outperforms that for binary items and the MC-NPS algorithm without any Q-optimal procedure. However,

recall that MC items with $H_j = 4$ were special cases in that the MC Q-optimal procedure was “reduced” from level 2 to 1. As a result, $K = 5$ items were required to achieve the optimal classification. Hence, in this case, the focus was shifted to exploring whether the Q-optimal procedure for binary items or the MC-NPS algorithm without any Q-optimal procedure resulted in higher PARs. When no Q-optimal procedure was applied, the initial item was selected randomly, and the sequential items were selected by maximizing the discrimination power.

Figure 3.1 displays the results for conditions with medium item quality. To facilitate the comparisons, two reference lines, $t = 3$ and $t = 5$, were drawn to indicate the cycles when the MC and binary Q-optimal procedures reached the optimal classifications, respectively. Recall that when $H_j = 4$, the Q-optimal procedure for MC



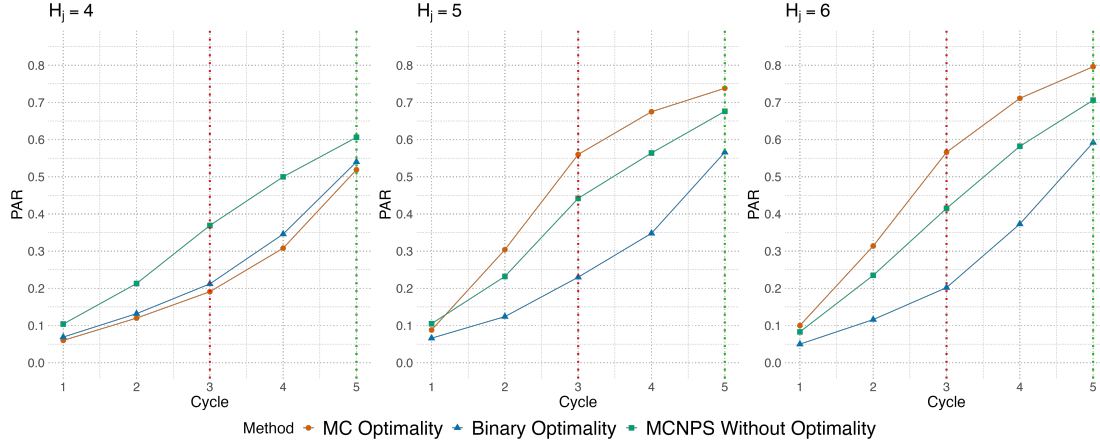
items was reduced to level 1 and, thus, required $K = 5$ items to achieve optimal classifications. Therefore, for the case of $H_j = 4$ as shown in the left-hand plot, the PAR trajectories for the MC and binary Q-optimal procedures (red line with filled circle and blue line with filled diamond, respectively) almost overlapped. The minor discrepancies between the two trajectories could be attributed to the different item selection criteria implemented in these two procedures. Intriguingly, the MC-NPS algorithm without any

Q-optimal procedure yielded the highest PARs among the three methods, an observation suggesting that the MC-NPS algorithm without Q-optimality benefited from the additional information provided by the coded distractors, enabling the selection of the most discriminating item at the earliest opportunity and ultimately improving classification accuracy. In contrast, while Q-optimal procedures can provide an optimal design, they compromised the classification accuracy by delaying the item selection until $t > K$, thus underutilizing the diagnostic information underlying the coded distractors.

Figure 3.1 further shows that for items with $H_j = 5$ and $H_j = 6$, the PARs of the binary Q-optimal procedure were consistently the lowest, a result that corresponded to the findings for the case of $H_j = 4$. In addition, the difference between the two Q-optimal procedures was significant. After two items were administered, the PAR obtained from the MC Q-optimal procedure was higher than that obtained from the binary procedure approximately by .15, and this difference increased to .2 at $t = 3$ and remained consistent at $t = 5$. It is worth noting that both Q-optimal procedures reached similar levels of PARs after completing their respective Q-optimal steps at $t = 3$ and $t = 5$ for the MC and binary Q-optimal procedures. These findings suggested that the newly proposed MC Q-optimal procedure was impressively effective in improving examinees' classifications in contrast to the binary Q-optimal procedure and the efficiency of each item was improved. In addition, the PAR of the MC Q-optimal procedure was .12 higher than that of the MC-NPS algorithm without Q-optimality at $t = 3$. The comparison of these three plots reveals that the inclusion of more options consistently improved the PARs of all three methods because when H_j was larger, there were more agreements between the ideal and observed responses for examinees whose ideal responses are zero.

Figure 3.2 presents the PARs from the three methods across different H_j when the items were high quality. Overall, the performance of all three procedures reaffirmed the conclusions drawn from the conditions for medium-quality items, with the only difference that the PARs were consistently higher due to the higher item quality. When $H_j = 4$, the MC Q-optimal procedure operated in a similar manner to the binary Q-optimal

Figure 3.2: PARs of Three Optimal Designs with High-Quality Items



procedure. In this case, the MC Q-optimal procedure did not integrate information from distractors to select MC items, which diminished the classification accuracy compared to the MC-NPS algorithm without Q-optimality. When $H_j = 5$ or 6, as shown in Figure 3.2, the MC Q-optimal procedure enhanced the measurement efficiency of individual MC items, reducing the number of required items in the Q-optimality design from 5 to 3. More importantly, the PARs of the MC Q-optimal procedure at $t = 3$ were comparable to those of the binary Q-optimal procedure at $t = 5$.

In conclusion, the newly proposed MC Q-optimal procedure takes advantage of the diagnostic information from the coded distractors, leading to higher PARs than the binary Q-optimal procedure. The comparison between the MC Q-optimal procedure and the MC-NPS algorithm without Q-optimality confirmed the immense advantage of integrating the MC Q-optimal procedure into the MC-NPS algorithm. After a thorough consideration of the trade-off between the Q-optimal procedure and administering the most discriminating items as early as possible, the MC-NPS method with $H_j = 4$ does not include the Q-optimal procedure for MC items. When $H_j = 4$, the first item is selected randomly for an examinee, and then his/her attribute profile is estimated by the MC-NPC method. Subsequently, steps 7-10, as outlined in the above algorithm, are

executed.

3.4.2 Simulation Study II: Assessing the Performance of the Developed MC-NPS Algorithm

This simulation study was conducted to assess the performance of the MC-NPS method in comparison to the JSD, GNPS, and random selection methods. Given that the GNPS method has been demonstrated to outperform the parametric methods when calibration samples are small with dichotomized responses (Chiu & Chang, 2021), no additional parametric item selection methods for dichotomized responses were employed in this study.

3.4.2.1 Design

This study adopted the same simulation design as that in Simulation Study I, with additional factors described in what follows. Recall that a calibration sample was required for the JSD method to calibrate item parameters using the MC-DINA model and for the GNPS method to estimate the weights in the GNPC method. In this regard, the MC-NPS method was free of calibration, making it a phenomenal CD-CAT algorithm particularly suitable for small samples. To evaluate the impact of small calibration samples on the performance of the compared methods, three sample sizes, $N_0 = 30$, 100, and 500, were implemented in the study. In addition, the attribute dependency was set to be independent or moderately dependent with the corresponding correlations generated from $\mathcal{U}(0.3, 0.5)$. When attributes were independent of each other, the examinee's attribute, α_{ik} , where $i = 1, \dots, N_0$ and $k = 1, \dots, 5$, was generated from $Ber(.5)$. When the attributes were moderately dependent, the examinees' attribute profiles were generated from the multivariate normal threshold model used in Simulation Study I. Furthermore, research has shown that the MC-DINA model suffers from convergence issues with small samples (Wang et al., 2023). Therefore, it was reasonable to assume that the JSD index might encounter calibration difficulties when calibration samples

were small. In this study, a new calibration sample was generated whenever the JSD method faced calibration issues until a successful calibration could be obtained or until a maximum of 10,000 calibration samples was reached.

The fixed-length stopping rule, with length $L = 4K$, served as the termination criterion for the study. The R packages `cdcatR` and `GDINA` were used to carry out the analysis for the GNPS method and the MC-DINA model, respectively.

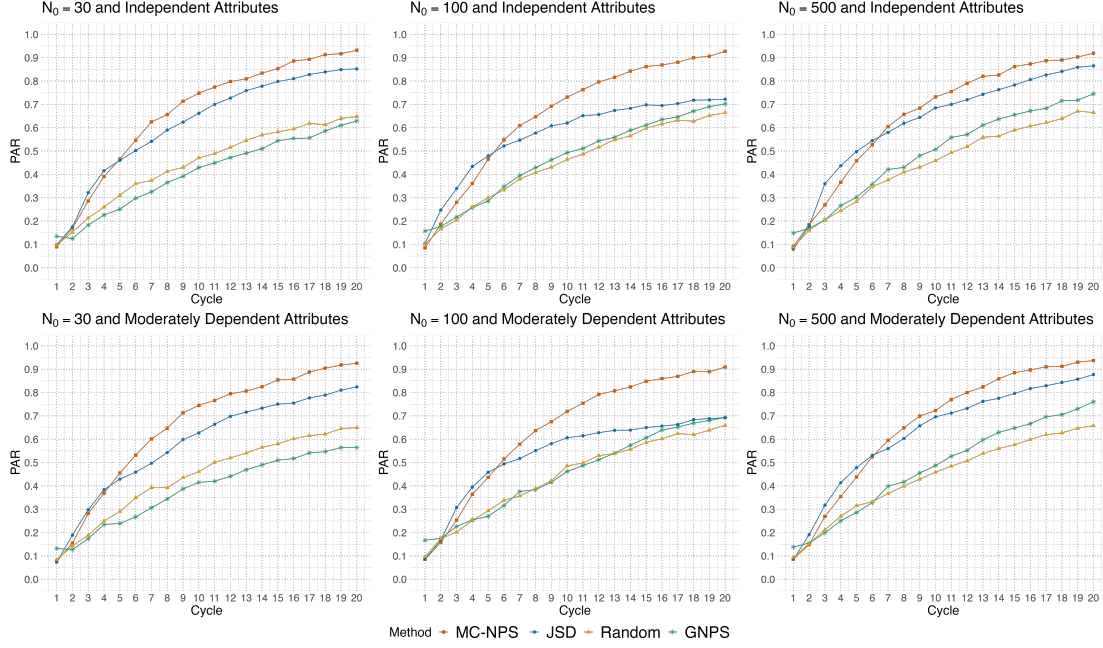
3.4.2.2 Results

Table 3.3 summarizes the number of calibration sample generations in each condition. Highlighted entries marked “10,000” indicate that the maximum of 10,000 calibration samples was reached but no calibration was obtained with the MC-DINA model. Thus, no results from the JSD index could be presented.

Table 3.3: The Number of Calibration Samples Generated in Each Condition

H_j	Attribute Structure	N_0	Number of Generations	
			Medium Item Quality	High Item Quality
4	Independent	30	839	44
		100	1	1
		500	1	1
	Moderately Dependent	30	661	38
		100	1	2
		500	1	1
5	Independent	30	10,000	80
		100	3	1
		500	1	1
	Moderately Dependent	30	10,000	14
		100	1	1
		500	1	1
6	Independent	30	10,000	89
		100	2	2
		500	1	1
	Moderately Dependent	30	10,000	56
		100	2	2
		500	1	1

Figure 3.3: PARs from the MC-NPS, JSD, GNPS, and Random Selection Methods with Medium-Quality Items When $H_j = 4$



The results show that when the item quality was medium, the MC-DINA model failed to calibrate item parameters when the calibration sample size was 30 and the items had either 5 or 6 options. When the item quality was high, fewer calibration sample generations were needed across all conditions. Furthermore, calibration samples with moderately dependent attributes required fewer generations than those with independent attributes, while keeping all other factors constant. It is important to note that although the maximum number of calibration sample generations allowed in the simulation was 10,000, in practice, it is usually the case that only one calibration sample can be obtained, implying a critical shortcoming of the JSD method to be used in small-scale settings.

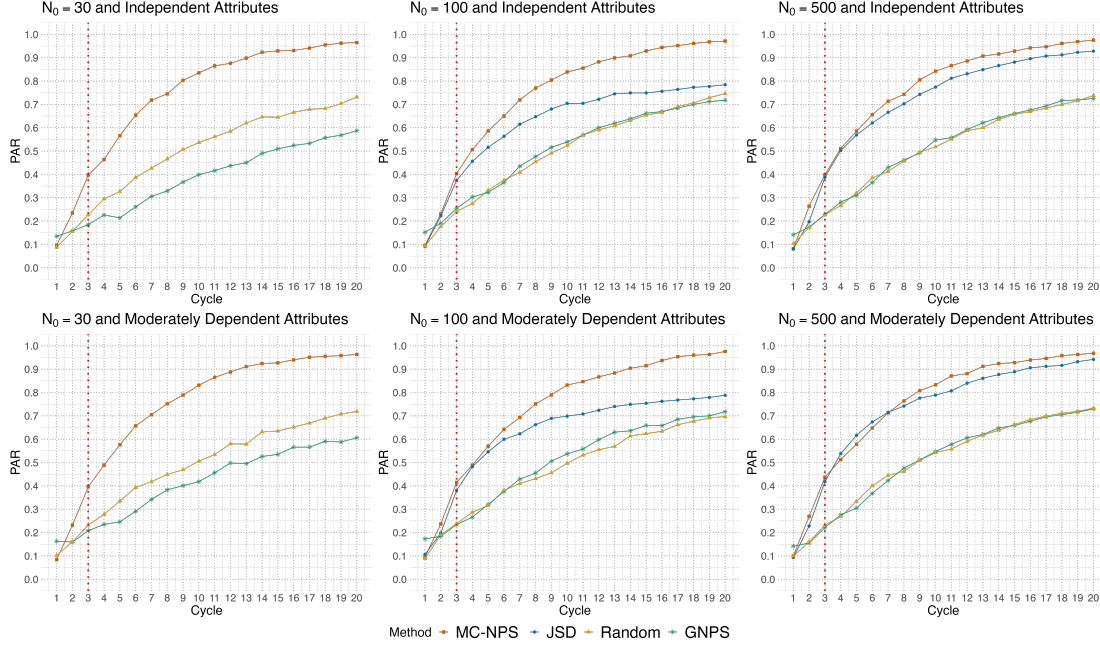
Condition I: Medium-Quality Items Figure 3.3 depicts the PARs of the four studied methods when items with four options were medium quality. Recall that when

$H_j = 4$, no Q-optimal procedure was implemented in the MC-NPS algorithm. Hence, it selected the first item randomly and proceeded to choose subsequent items by maximizing the proposed discrimination power. The same procedure was adopted for the JSD method except that items were selected by maximizing the JSD index. The results show that across all calibration sample sizes and attribute structures, the MC-NPS method exhibited comparable performance to the JSD index during the first six cycles, gradually surpassing it in subsequent cycles. However, the performance of the JSD index did not always improve along with the increase in the calibration sample sizes. For example, the PARs from conditions with $N_0 = 30$ were higher than those with $N_0 = 100$. It is important to note that the JSD index encountered severe calibration challenges when calibration samples were small. Table 3.3 showed that no calibration samples of size 30 could be successfully calibrated until the 840th and 662nd calibration samples with independent and moderately dependent attributes, respectively, were generated. These results implied that the MC-NPS method was more suitable for small samples than the JSD method.

The comparison between the GNPS method and the two item selection methods for MC items revealed a significant enhancement in examinee classification accuracy when considering the diagnostic information underlying coded distractors, especially when the calibration sample was small or large enough. Moreover, both item selection methods for MC items outperformed the random selection method, demonstrating their effectiveness in selecting discriminating MC items to improve the classification accuracy of CD-CAT. It is worth noting that the random selection method yielded superior results compared to the GNPS method when the calibration sample was as small as 30. In contrast, the GNPS method had better performance than the random selection method after administering seven items when the calibration sample size was 500. This finding aligned with the observation that the performance of the GNPS method is sensitive to the size of the calibration sample and item quality (Chiu & Chang, 2021).

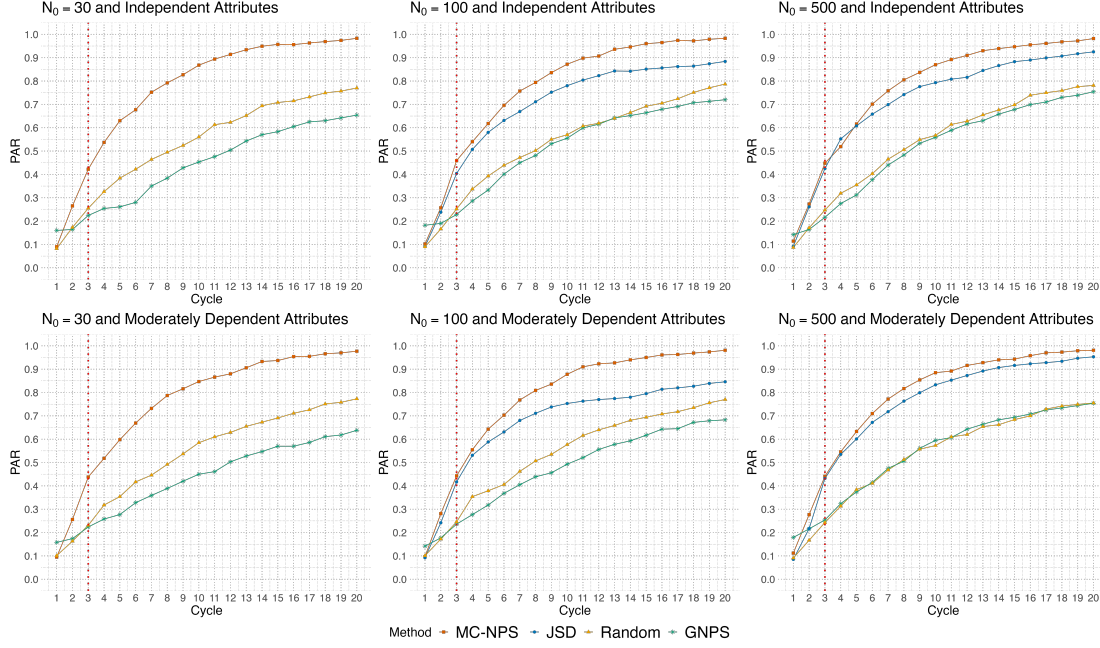
Figure 3.4 presents the results for $H_j = 5$. In the two plots in the left panel for

Figure 3.4: PARs from the MC-NPS, JSD, GNPS, and Random Selection Methods with Medium-Quality Items When $H_j = 5$



$N_0 = 30$, no trajectory describing the PARs of the JSD index was available. This absence was attributed to the MC-DINA model's inability to calibrate item parameters for the JSD index, even after generating 10,000 calibration samples, as indicated in Table 3.3. Nevertheless, the MC-NPS method quickly reached PARs = .9 and .95 after 13 and 18 items, respectively, were administered. Recall that the MC-NPS method employed the Q-optimal procedure for MC items when $H_j \geq 5$, and the vertical reference line at $t = 3$ indicated the completion of the Q-optimal procedure. The PARs of the MC-NPS method were approximately .4 at $t = 3$, which aligned with the results presented in Simulation Study I. When $N_0 = 100$ and 500, as shown in the middle and right panels of Figure 3.4, the MC-NPS method consistently outperformed the JSD index across both attribute structures. Furthermore, the JSD index exhibited superior performance with $N_0 = 500$ compared to $N_0 = 100$. Both the MC-NPS method and the JSD method, with successfully calibrated item parameters, outperformed the random selection and the

Figure 3.5: PARs from the MC-NPS, JSD, GNPS, and Random Selection Methods with Medium-Quality Items When $H_j = 6$



GNPS methods with significant differences in PARs across all conditions. In addition, the PARs of the GNPS method remained lower than those of the random selection method when the calibration sample was small. However, with calibration sample sizes of 100 and 500, the GNPS method exhibited comparable performance to the random selection method. The comparison between Figures 3.3 and 3.4 revealed that the number of options had little impact on the GNPS method, whereas the performance of the random selection method improved along with increasing number of options.

The PARs for the conditions with $H_j = 6$ are shown in Figure 3.5. Most patterns mirror those of the case of $H_j = 5$. When the calibration sample was 100, the disparity in PARs between the MC-NPS and JSD methods decreased further, but the difference remained around .1 at $t = 20$. The classification accuracy of the random selection was also improved along with the increased number of options. It outperformed the GNPS method at all cycles after $t = 3$, except for the cases where $N_0 = 100$ with the

independent attribute structure and of $N_0 = 500$.

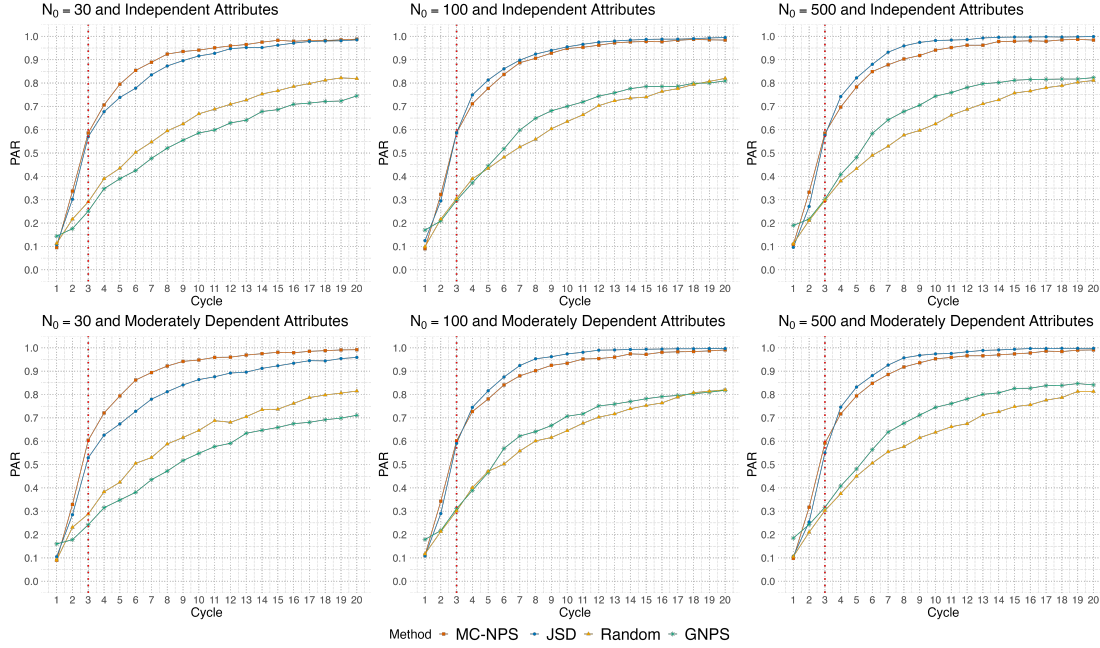
The comparison across three conditions with varying numbers of options reveals that the classification accuracy of the MC-NPS, JSD, and random selection methods improved as the number of options within each MC item increased. In contrast, the number of options had a limited effect on the guessing parameter when using the GNPS method. Consequently, its performance remained nearly constant across the three conditions with the same calibration sample size and attribute structure.

Condition II: High-Quality Items In summary, higher item quality led to improved classification accuracy across all four methods. In contrast to the medium item quality condition, the MC-DINA model could successfully calibrate the item parameters for the JSD index, enabling a comparison between the MC-NPS and JSD methods under $N_0 = 30$ and $H_j = 5$ or 6 . Consistent with the findings for medium-quality items, the MC-NPS method consistently outperformed the GNPS and random selection methods in all conditions. Unlike the medium-quality item condition, the performance of the JSD index improved substantially due to the higher item quality.

The results for $H_j = 4$ and $H_j = 5$ are shown in Figures A.1 and A.2 in Appendix A as they closely resemble the results for $H_j = 6$ shown in Figure 3.6. Under $N_0 = 30$, the MC-NPS method outperformed the JSD index across the two attribute structures. However, under $N_0 = 100$ and 500 , the JSD index exhibited higher PARs than the MC-NPS method across all conditions with minimal differences in PAR. Specifically, when the test concluded at $t = 20$, the PARs of both the JSD index and the MC-NPS method were close to 1 in all conditions. These comparisons suggest that the MC-NPS method is preferable when the calibration samples are small, even for high-quality items.

Regarding the comparison between the random selection and the GNPS methods, the random selection method displayed notably superior performance when the calibration sample was 30, while the performance gap decreased compared to medium-quality items. When $N_0 = 100$ and 500 , the GNPS method surpassed the random selection method

Figure 3.6: PARs from the MC-NPS, JSD, GNPS, and Random Selection Methods with High-Quality Items When $H_j = 6$



for $H_j = 4$, and their performances were comparable for $H_j = 5$ or 6. Although the performance of the GNPS method was enhanced due to the higher item quality, it still reached its ceiling at approximately $t = 15$, with the growth rate of PAR slowing significantly after this point. Conversely, the PAR for random selection continued to rise rapidly beyond 15 items, catching up to the GNPS method within a few items. Consequently, their PARs were comparable at $t = 20$. These results further support that dichotomizing MC items, which results in the loss of diagnostic information underlying the coded distractors in CD-CAT, impairs the classification, and the higher item quality cannot compensate for this deficiency.

3.5 Discussion

Numerous item selection methods have been developed for CD-CAT with a primary focus on the binary responses. The JSD index is currently the only item selection method for MC items but may encounter severe calibration difficulties when calibration samples are small. In response to these challenges, this study proposes the MC-NPS method, which relies on the proposed discrimination power that measures how effectively an item can differentiate between an examinee's two most likely attribute profiles to select items. After administering the selected item to the examinee and collecting the response, the estimated attribute profile of the examinee is updated by the MC-NPC method. The CD-CAT algorithm, guided by the discrimination power and attribute profile estimation, continues until the stopping criterion is met. It is important to note that when $H_j = 2$, the MC-NPS method reduces to the nonparametric item selection method (NPS; Chang et al., 2019). Similar to the NPS method, the MC-NPS method does not require any item calibration.

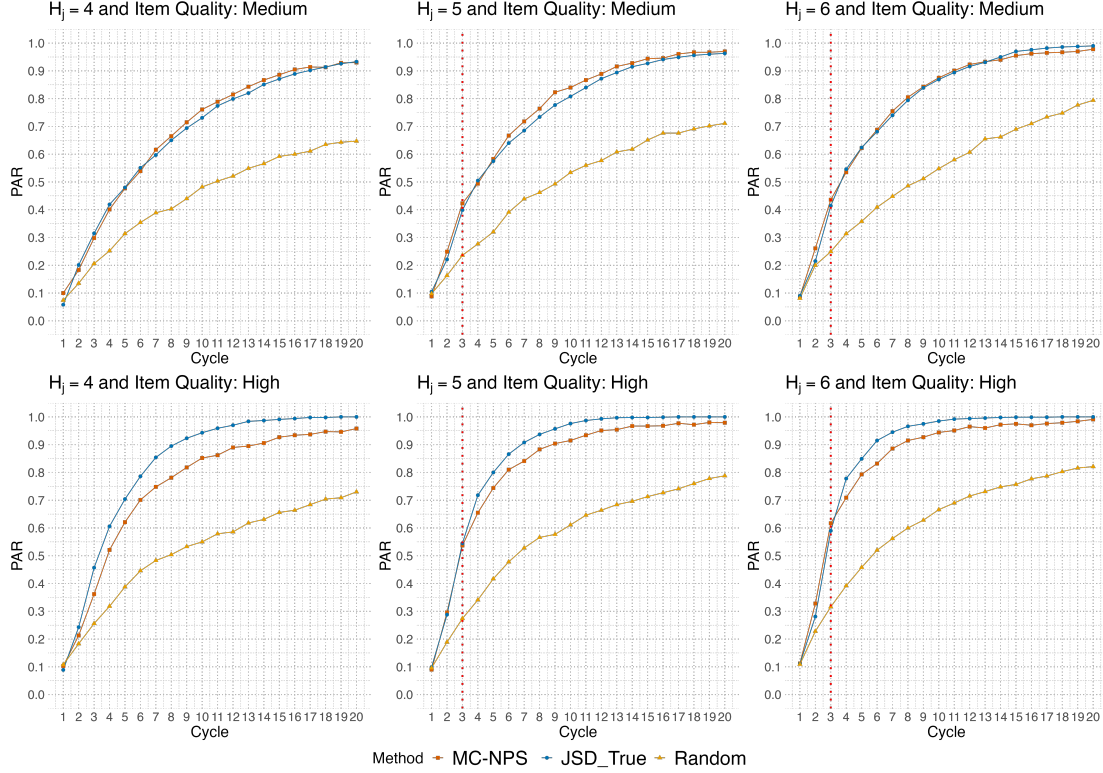
The MC Q-optimal procedure is also developed to select initial items for the MC-NPS method when $H_j \geq 5$. Simulation Study I demonstrates the MC Q-optimal procedure significantly improves classification accuracy during the initial phase of a CD-CAT algorithm, compared to the binary Q-optimal procedure and the MC-NPS method without any Q-optimal design when $H_j = 5$ or 6. However, the MC Q-optimal procedure closely resembles the binary Q-optimal procedure when $H_j = 4$, and in this case, its performance falls short of the MC-NPS method without any Q-optimal design. Although the MC Q-optimal procedure was applied to the MC-NPS method in the study, it can be used to select initial items for any method designed for CD-CAT with MC items.

Simulation Study II provides evidence that the MC-NPS method outperforms the JSD index when dealing with medium-quality items across all calibration samples and the differences are particularly substantial in conditions with five or six options. When the item quality was high, the JSD index exhibited a performance comparable to the

MC-NPS method, and Table 3.3 suggests that the convergence issue is less unlikely to occur with the MC-DINA model in these conditions and, thus, with the JSD method when $N_0 = 100$ and 500. However, the problem may be more serious than anticipated in practice. When the sample size is small, it is likely that there are only a few latent classes included in the data, especially when K is large. In such a situation, increasing the sample size may not remedy the issue with the MC-DINA model. A small-scale simulation was conducted to investigate whether an increased sample size can compensate for the consequence of homogeneous examinees. The first experiment included a calibration sample of size 100, and examinees were evenly assigned to one of the three randomly chosen latent classes, $\alpha_1 = (0, 0, 1, 0, 0)'$, $\alpha_2 = (1, 0, 0, 1, 1)'$, and $\alpha_3 = (0, 1, 0, 0, 0)'$. After generating the responses of these examinees, the MC-DINA model was then used to calibrate the items. The result revealed that no successful calibration could be obtained after generating 10,000 samples. Subsequently, the sample size was increased to $N_0 = 300$, and another set of latent classes, $\alpha_1 = (1, 0, 0, 1, 0)'$, $\alpha_2 = (0, 1, 0, 0, 0)'$, and $\alpha_3 = (1, 1, 1, 0, 1)'$, were randomly chosen. However, the outcome remained the same: no successful calibration could be obtained after 10,000 samples were generated. This preliminary simulation concluded that the executability of the JSD index heavily relied on not only a large calibration sample but also the presence of diverse latent classes in the data, which may restrict its applicability in small-scale settings.

Another interesting question is how the MC-NPS algorithm performs in comparison to the JSD method with the true item parameters being used to classify examinees. In this scenario, the results from the JSD method represent the best performance it could achieve, although it is not realistic. To investigate, a simulation study was conducted, and the results are shown in Figure 3.7. The results reveal that the MC-NPS method performed comparably with the JSD index when the items were medium quality, but slightly worse than the JSD index when the item quality was high. These findings indicate that MC-NPS, a method not requiring any calibration, is more effective than the JSD method with calibration samples of any size when the item quality is medium.

Figure 3.7: PARs from the MC-NPS, JSD with True Parameters, and Random Selection Methods



In addition, the MC-NPS method, as a nonparametric method, significantly outpaces the JSD index in terms of computational speed.

The comparison between the MC-NPS and GNPS methods in Simulation Study II suggests that incorporating the information underlying the coded distractors can improve the accuracy of estimation and increase the efficiency of item selection. However, it is important to acknowledge that the comparison between the GNPS and the other three methods might not be entirely equitable. The MC-NPS and random selection methods utilized the MC-NPC method to estimate attribute profiles, and the JSD index applied the MC-DINA method for attribute profile estimation. Both of these estimation methods are specifically designed for MC items. In contrast, the GNPS method

employed both item selection and estimation methods designed for the binary items. Furthermore, the estimation methods for binary items applied to the analysis of MC items are less effective than those designed for MC items (de la Torre, 2009; Wang et al., 2023). Therefore, the lower PARs of the GNPS method should not be interpreted as the consequence of the ineffectiveness or the inefficiency of the method.

In conclusion, this study makes at least three significant contributions. First, the proposed Q-optimal procedure for MC items improves the efficiency of the MC-NPS method and reduces the number of items required to attain a classification accuracy equivalent to that achieved by the Q-optimal procedure for binary items. Second, the proposed discrimination power boosts the item selection efficiency by leveraging the diagnostic information embedded within coded distractors. Third, the MC-NPS method produces high classification accuracy without the need for a calibration sample. In contrast to the JSD index that relies on a large calibration sample, the MC-NPS method is a more practical and effective approach.

This study does have certain limitations that warrant acknowledgment. First, the correct specification of the Q-matrix is mandatory for securing the validity of any CD analysis. It is of even greater importance within the context of MC items because a much larger number of q-vectors is needed to accommodate all distractors. However, currently, we are not aware of any Q-matrix refinement-validation methods for MC items as they are available for binary items (see Chiu, 2013; de la Torre & Chiu, 2016). In addition, model-data misfit is a realistic risk and limitation of the MC-DINA model and the MC-NPC method that, of course, extends to the JSD and MC-NPS methods. Unfortunately and to the best of our knowledge, no efficient procedures are available for detecting model misfit, presumably because, currently, most CDMs for MC items are variations of the DINA model. Hence, one of the most pressing future research tasks is the development of a general model for handling even highly complex data that should mitigate the issue of model misfit. In the meantime, given that JSD and MC-NPS are the only available CD-CAT algorithms for MC items, studies for exploring

their robustness to model misspecifications might be of high practical relevance. Finally, item exposure was not factored into the analysis, and future research should be done to investigate this important quality of the MC-NPS method. Lastly, this study only offers a justification of the sufficient condition for the Q-optimal procedure for MC items with $5 \leq H_j < 14$. Consequently, more theoretical work should be carried out to provide formal and solid support of the procedure.

Chapter 4

Q-Matrix Completeness and Q-Optimal Design for Multiple-Choice Items in CD

4.1 Introduction

Cognitive diagnosis (CD) aims to provide detailed insights into students' knowledge and skills, offering immediate feedback on their strengths and weaknesses in specific skills. However, most CD models (CDMs) were designed for binary responses and thus fall short in comprehensively analyzing incorrect responses. This limitation is particularly pronounced in the case of multiple-choice (MC) items, as some distractors also offer critical diagnostic insights. However, when analyzed by binary CDMs, all distractors are collapsed into the incorrect answer, preventing the extraction of additional diagnostic information from the distractors of MC items.

To addressing this limitation, various CD methods designed specifically for MC items have been developed (e.g., de la Torre, 2009; DiBello et al., 2015; Ozaki, 2015; Wang et al., 2023). Among these, the MC deterministic inputs, noisy “and” gate (MC-DINA)

model (de la Torre, 2009) stands out as the pioneer and has been extensively studied. The MC-DINA framework is characterized by a design in which, in addition to the key, a subset of distractors—referred to as coded distractors—is also linked to certain q-vectors. These q-vectors carry additional diagnostic information and are systematically analyzed to enhance the classifications of examinees. However, sharing the common challenge with many other parametric models, the MC-DINA model may encounter convergence issues, resulting in lower classification accuracy in small-scale settings. To overcome this difficulty, its nonparametric counterpart, the nonparametric classification method for MC items (MC-NPC; Wang et al., 2023), was developed. In addition to achieving more accurate classification, Wang et al. (2023) clearly introduced the structure of q-vectors of coded options, a crucial advancement for crafting MC items in CD.

Similar to CD methods for dichotomous responses, the Q-matrix (Tatsuoka, 1985) is an integral component of analyses for MC items. However, unlike CD methods for dichotomous responses, where extensive research has been conducted on the Q-matrix, some important statistical properties of the Q-matrix for MC items remain unexplored. One such property is completeness, which determines whether all possible attribute profiles in the data can be identified. An incomplete Q-matrix may lead to misclassification of examinees into incorrect groups, thereby distorting the interpretation of results and subsequent pedagogical interventions. Completeness is also a prerequisite for parameter identifiability (e.g., Gu & Xu, 2021; Lin & Xu, 2024). As an example, Lin & Xu (2024) outlined the conditions for parameter identifiability with two polytomous response models inspired by the DINA model framework: the GPDM model (Chen & de la Torre, 2018) and the sequential DINA model (Ma & de la Torre, 2016). However, rather than proposing and proving the conditions for a complete Q-matrix with these models, the Q-matrix was simply assumed to be complete, which may limit the value of the conditions for identifiability. Therefore, the primary goal of this study is to establish conditions for determining the completeness of the Q-matrix for MC items. One sufficient and necessary theorem is established to formally articulate these conditions,

based on legitimate evidence from mathematical proofs.

In addition to exploring the completeness of the Q-matrix, another highly related topic, the Q-optimal design for MC items, will also be studied. A Q-optimal design provides optimal classification during the initial stage of a computerized adaptive testing for CD (CD-CAT) algorithm and has been implemented in several CD-CAT algorithms to ensure high quality estimation (Xu et al., 2016). The MC Q-optimal procedure for MC items was first proposed by Wang et al. (2024) and implemented in the CD-CAT algorithm they proposed. The current study will go one step further to establish the theoretical foundation for the MC Q-optimal design and provide mathematical proofs to support its feasibility.

The remainder of this paper is structured as follows: First, a review of the deterministic inputs, noisy “and” gate (DINA) model (Junker & Sijtsma, 2001) and its completeness is presented, followed by an overview of the CD methods for MC items. Next, the completeness of the Q-matrix for MC items, identification conditions for each latent class, and the Q-optimal design for CD-CAT with MC items are introduced. Subsequently, two simulation studies are conducted to support the proposed theorems for the completeness of the Q-matrix for MC items and the Q-optimal design for CD-CAT with MC items. The discussion section concludes with a summary of the main findings and limitations. The proof of the main theorem and detailed derivations for the proposed algorithm are presented in Appendix B.

4.2 Technical Background

4.2.1 CD for Dichotomized Items

In CDMs, the cognitive skills assessed by an assessment are referred to as attributes. Assuming all attributes are binary and exhibit no hierarchical structure, if the assessment is used to measure K attributes, there will be 2^K latent classes. Each latent class C_m is defined by a K -dimensional attribute profile $\alpha_m = (\alpha_{m1}, \dots, \alpha_{mK})'$, where

$m = 1, \dots, 2^K$. Each entry α_{mk} in $\boldsymbol{\alpha}_m$ takes the value of 1 if the k th attribute has been mastered and 0 otherwise. For examinee i belonging to latent class C_m , his/her attribute profile, $\boldsymbol{\alpha}_i$, is equal to $\boldsymbol{\alpha}_m$.

For all CD analyses, the Q-matrix is an integral component, which is used to indicate relations between the test items and their required attributes. Suppose a test of J MC items measures at most K attributes. When the MC items are dichotomized to fit a binary CDM, only the q-vector of each key is specified. Thus, the Q-matrix, denoted as $\mathbf{Q}_B$, has the dimensions $J \times K$. Denote the q-vector of item j as $\mathbf{q}_j$. Similar to an examinee's attribute profile $\boldsymbol{\alpha}_i$, $\mathbf{q}_j$ is K -dimensional, with $q_{jk} = 1$ indicating that attribute k is required by item j and 0 otherwise.

4.2.1.1 The DINA Model

The DINA model is the most extensively researched dichotomous CDM. Let $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{iJ})'$ be the observed response vector of examinee i . The item response function (IRF) of the DINA model is defined as follows:

$$P(Y_{ij} = 1 | \boldsymbol{\alpha}_i) = (1 - s_j)^{\eta_{ij}} g_j^{1 - \eta_{ij}},$$

where s_j and g_j represent the slipping and guessing parameters of item j , respectively, subject to the constraint $0 < g_j < 1 - s_j < 1$. The parameter $\eta_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}}$ is the ideal response of examinee i to item j , indicating whether the examinee has mastered all the required attributes. Hence, the DINA model is a typical conjunctive model.

4.2.2 CD for MC Items

When traditional CDMs are used, the responses to all the distractors of a MC item are lumped together as the incorrect response. Consequently, the distractors remain unavailable for further analysis, resulting in the loss of valuable diagnostic information. A preferable approach is to employ CD methods specifically tailored for MC items, and

the current developments in CD for MC items are briefly presented below.

In the context of CD for MC items, not only the key but also all or a subset of the distractors are assigned specific q-vectors. In general, an option (which could be a key or a distractor) is said to be coded if it is assigned a q-vector. When the observed responses are analyzed using a CD method developed for MC items, the Q-matrix for MC items, denoted as $\mathbf{Q}_{\text{MC}}$, is augmented from $\mathbf{Q}_{\text{B}}$ to account for the specification of q-vectors for the coded distractors. In this case, each MC item corresponds to a sub-Q-matrix, $\mathbf{Q}_j$, with dimensions $H_j^* \times K$, where H_j^* is the number of coded options, expressed as follows:

$$\mathbf{Q}_j = \begin{pmatrix} \mathbf{q}_j^{(H_j^*)} \\ \vdots \\ \mathbf{q}_j^{(1)} \end{pmatrix}_{H_j^* \times K},$$

where $\mathbf{q}_j^{(H_j^*)}$ is the q-vector of the key, and the others represent the q-vectors of the coded distractors. $\mathbf{Q}_{\text{MC}}$ is essentially the collection of all $\mathbf{Q}_j$, having the dimensions $(\sum_{j=1}^J H_j^*) \times K$. Based on this setup, $\mathbf{Q}_{\text{MC}}$ can be expressed as follows:

$$\mathbf{Q}_{\text{MC}} = \begin{pmatrix} \mathbf{Q}_1 \\ \vdots \\ \mathbf{Q}_J \end{pmatrix}_{(\sum_{j=1}^J H_j^*) \times K}.$$

4.2.2.1 The MC-DINA Model and Its Modifications

The MC-DINA model (de la Torre, 2009), an extension of the DINA model, is the first CDM designed for MC items. In this study, a constraint that the q-vectors of the coded options were hierarchically nested was imposed, which limited the practicability of the model, and thus, all response options form ordinal options. Let $\mathbf{X}_i = (X_{i1}, \dots, X_{iJ})'$ be the observed response vector, and X_{ij} can take on values $h = 1, \dots, H_j$, where H_j is

the number of options. Then, the ideal response of examinee i to MC item j is defined as follows:

$$g_{ij} = \arg \max_{h'=0,1,2,\dots,H_j} \left\{ \boldsymbol{\alpha}'_i \mathbf{q}_j^{(h')} \mid \boldsymbol{\alpha}'_i \mathbf{q}_j^{(h')} = \mathbf{q}_j^{(h')'} \mathbf{q}_j^{(h')} \right\},$$

where h' is recoded from h based on the number of skills required by the coded options. More required skills result in a larger h' , and noncoded options have $h' = 0$. Then, the IRF of the MC-DINA model is formulated as follows:

$$P(X_{ij} = h \mid g_{ij} = g) = P_j(h \mid g),$$

without further specifying how to find $P_j(h \mid g)$.

The stringency of the nestedness constraint imposed on the MC-DINA model limits its applicability and was relaxed in the work by Wang et al. (2023). As a result, the coded options are no longer ordinal. To avoid ambiguity between notations, let l represent the level of responses without the nestedness constraint. Accordingly, the relaxed rule was proposed to level non-ordinal coded options: The key is assigned the highest level, H_j^* , noncoded options are collapsed to the lowest level of 0, and each remaining coded option is assigned a unique value from $\{1, \dots, H_j^* - 1\}$. The ideal response to MC item j from examinee i , η_{ij} , is then formulated as follows:

$$\eta_{ij} = \max_{l=0,1,2,\dots,H_j^*} \left\{ l \prod_{k=1}^K I[\alpha_{ik} \geq q_{jk}^{(l)}] \right\}. \quad (4.1)$$

Equation 4.1 shows that the inclusion of this ideal response makes the MC-DINA model conjunctive, akin to the DINA model. The IRF of the MC-DINA model can then be explicitly written as follows:

$$P(X_{ij} = l \mid \boldsymbol{\alpha}_i) = \begin{cases} \frac{1}{H_j} + \frac{H_j - H_j^* - 1}{H_j} I[l = 0] & \text{if } \eta_{ij} = 0 \\ \left(1 - \sum_{c \neq l} \epsilon_{jcl} \right)^{I[\eta_{ij}=l, l>0]} \prod_{l' \neq l} \epsilon_{jll'}^{I[\eta_{ij}=l']} & \text{if } \eta_{ij} > 0, \end{cases} \quad (4.2)$$

subject to $\sum_{c \neq l} \epsilon_{jcl} < .5$ for all items. In the above equation, ϵ_{jcl} is the probability that the observed response level c disagrees with the nonzero ideal response level l .

4.2.2.2 Useful and Proper Conditions for Q_{MC}

It is important to note that removing the nestedness constraint described above does not imply that the coded options can randomly take any q-vectors. Köhn et al. (2023) argued that the coded options of a MC item must be “proper” and “useful” to avoid ambiguity and redundancy, respectively. Chiu et al. (under review) further proposed two theorems to establish the conditions that can be used to determine whether MC items are proper and useful. Loosely speaking, item j is said to be useful if and only if the q-vector of the key is not included within any q-vector of coded distractors. Furthermore, item j is proper if and only if Q_j contains at most one coded distractor, or for each pair of the coded distractors, the union of their q-vectors must be included in Q_j , or the union must include $q_j^{(H_j^*)}$.

4.3 Objective 1: Completeness of Q_{MC} for the MC-DINA Model

As discussed earlier, the completeness of a Q-matrix guarantees that all attribute patterns can be identified and thus is a statistical property that must be investigated before CD analysis can be properly and accurately performed. Completeness is also a key assumption of some important statistical properties, such as parameter identifiabilities (e.g., Gu & Xu, 2021) and Q-optimality for CD-CAT (Xu et al., 2016). The second goal of this study is to establish the conditions for the completeness of a Q-matrix when data conform to the MC-DINA model, which will then be used in the next theorem to construct the Q-optimal procedure for CD-CAT with MC items. Several important propositions implied by the theorem are also discussed in this section.

4.3.1 Review of Completeness of $\mathbf{Q}_B$ for the DINA Model

The concept of a “complete” Q-matrix was first discussed and proven in the asymptotic classification theory for cognitive diagnosis (ACTCD; Chiu et al., 2009). A Q-matrix is said to be complete if it can identify all possible attribute profiles. Köhn and Chiu (2016b) revisited the ACTCD and offered a more general definition of Q-matrix completeness based on the conditional expectation of the observed response given an examinee’s attribute profile. Formally stated, defining $\mathbf{S}(\boldsymbol{\alpha}) = E(\mathbf{Y}|\boldsymbol{\alpha})$, then a test is complete if $\mathbf{S}(\boldsymbol{\alpha}) = \mathbf{S}(\boldsymbol{\alpha}^*)$ implies $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$ for all $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^*$.

Based on the definition of completeness of the Q-matrix, Chiu et al. (2009) and Köhn & Chiu (2016b) established the sufficient and necessary conditions for a complete Q-matrix for the DINA model. The theorem states that $\mathbf{Q}_B$ is complete for the DINA model if and only if it includes all the $\mathbf{e}_k$, where $k = 1, \dots, K$. Detailed proof can be found in Chiu et al. (2009) and Köhn and Chiu (2016b).

4.3.2 Level- x Design

Under the DINA model, $\mathbf{Q}_B$ is complete if and only if it includes all the K $\mathbf{e}_k$. Such a complete Q-matrix can then be used to identify all 2^K distinct attribute profiles, assuming no hierarchical structure among the attributes. Similarly, including all the $\mathbf{e}_k$ is essential for achieving the completeness of $\mathbf{Q}_{MC}$. However, under the MC-DINA model, a single MC item may include multiple $\mathbf{e}_k$ when the distractors are also coded. Therefore, the major challenge here is to ensure that the assignment of the x distinct $\mathbf{e}_k$, where $1 \leq x \leq K$, to the coded options within an MC item maximizes the effect of these $\mathbf{e}_k$, distinguishing 2^x attribute profiles. Specifically, these assignments cannot be random but governed by a) a particular arrangement such that every $\mathbf{e}_k$ can distinguish between examinees who have mastered attribute k and those who have not, and b) by the proper and useful rules. Hence, new theoretical developments on how to organize $\mathbf{e}_k$ in a $\mathbf{Q}_{MC}$ to guarantee the completeness are required.

Before the location of e_k can be identified, it is necessary to take a look how examinees are classified with specific relations between coded distractors and the key. Suppose item j contains a coded distractor, l . Given that the key is always selected when $\alpha \succeq q_j^{(H_j^*)}$, the latent classes have three different types of ideal responses: l' , where $l' \neq l$ and H_j^* , l , and H_j^* , thereby classifying examinees into three corresponding groups, as shown in Table 4.1. It is obvious that item j classifies examinees into two

Table 4.1: Classes Formulated by Distractor l and the Key of Item j

Ideal Response Class	$\alpha \not\succeq q_j^{(H_j^*)}$			$\alpha \succeq q_j^{(H_j^*)}$
	$\alpha \not\succeq q_j^{(l)}$		$\alpha \succeq q_j^{(l)}$	
	$l', \text{ where } l' \neq l \text{ and } l' \neq H_j^*$		l	
	G_1		G_2	G_3

latent classes based on whether $\alpha \succeq q_j^{(H_j^*)}$ (i.e., $G_1 + G_2$ vs. G_3). Nevertheless, item j may fail to identify all examinees who have mastered attributes required by distractor l . It is true that examinees in G_1 have not mastered the attributes required by distractor l , and those in G_2 have mastered them. If $q_j^{(l)} \prec q_j^{(H_j^*)}$, it is also true that G_3 consists of examinees who have mastered attributes required by distractor l . If $q_j^{(l)} \not\prec q_j^{(H_j^*)}$, however, G_3 may be composed of examinees who have and have not mastered attributes required by distractor l .

When e_k is the focus, there are three possible scenarios for e_k : a) $q_j^{(H_j^*)} = e_k$; b) $q_j^{(l)} = e_k$ and $q_j^{(H_j^*)} \succ e_k$; and c) $q_j^{(l)} = e_k$ and $e_k \not\prec q_j^{(H_j^*)}$. Based on the previous discussion, item j in cases a) and b) can divide examinees into two groups: examinees who have and have not mastered attribute k ; however, in case c), item j fails to do so. To demonstrate, take e_2 as an example. Suppose there are four items designed to measure the mastery of attribute 2 (A2) when $K = 3$, and their Q-matrices and ideal responses of latent classes are shown in Table 4.2. Further assume that these four items do not contain other coded distractors than those listed in Table 4.2. In the table, the four latent classes that have mastered A2 are listed first, followed by the other four that

Table 4.2: Ideal Responses to Four Items Assessing the Mastery of A2

		$\mathbf{Q}_1$	$\mathbf{Q}_2$	$\mathbf{Q}_3$	$\mathbf{Q}_4$
	$\mathbf{q}^{(H_j^*)}$	010	010	110	100
	$\mathbf{q}^{(l)}$	-	100	010	010
Mastery of A2	(010)	1	2	1	1
	(110)	1	2	2	2
	(011)	1	2	1	1
	(111)	1	2	2	2
Non-Mastery of A2	(000)	0	0	0	0
	(100)	0	1	0	2
	(001)	0	0	0	0
	(101)	0	1	0	2

have not mastered A2. $\mathbf{Q}_1$ and $\mathbf{Q}_2$ are used to demonstrate case a), where $\mathbf{q}_j^{(H_j^*)} = \mathbf{e}_k$; $\mathbf{Q}_3$ is used to exemplify case b), where $\mathbf{q}_j^{(l)} = \mathbf{e}_k$ and $\mathbf{q}_j^{(H_j^*)} \succ \mathbf{e}_k$; $\mathbf{Q}_4$ shows case c), where $\mathbf{q}_j^{(l)} = \mathbf{e}_k$ and $\mathbf{q}_j^{(H_j^*)} \neq \mathbf{e}_k$. Item 1 is equivalent to a binary item because it does not include any coded distractor. Therefore, latent classes who have mastered A2 can be differentiated from those who have not mastered. As for item 2 where $\mathbf{q}_2^{(2)} = \mathbf{e}_2$ and $\mathbf{q}_2^{(1)} = \mathbf{e}_1$, latent classes who have mastered A2 were classified into G_3 with the ideal response of 2. Those who have not mastered A2 were classified into either G_2 or G_1 with the ideal response being either 1 or 0. Therefore, a $\mathbf{Q}_j$ satisfying $\mathbf{q}_j^{(H_j^*)} = \mathbf{e}_2$ can effectively differentiate between examinees who have mastered A2 and those who have not. In $\mathbf{Q}_3$, $\mathbf{q}_3^{(2)} \succ \mathbf{e}_2$ and $\mathbf{q}_3^{(1)} = \mathbf{e}_2$. Consequently, examinees who have mastered A2 will select either the key or distractor, and those who have not mastered A2 will make a random selection among all options. As a result, $\mathbf{Q}_3$ can successfully distinguish between examinees who have mastered A2 and those who have not. With $\mathbf{Q}_4$, examinees who have mastered both A1 and A2 (i.e., (110) and (111)) and those who have not mastered A2 but mastered A1 (i.e., (100) and (101)) are classified into G_3 . Therefore, item 4 is unable to differentiate between examinees who have mastered A2 and those who have not. Hence, if $\mathbf{e}_k$ is linked to a distractor but the key does not require attribute k , this

item cannot fully distinguish between the mastery and non-mastery of attribute k .

The above discussion and examples show that the inclusion of e_k alone is not sufficient to constitute a complete Q_{MC} ; the assignment of e_k is also crucial to its completeness. Hence, to establish the completeness of Q_{MC} , the *level- x design* is first developed and defined to properly organize e_k in Q_{MC} .

Definition 1. A Q_j is said to have a level- x design for $x \geq 1$ if x distinct e_k , where $k = 1, 2, \dots, K$, are linked to its coded options and it can distinguish 2^x latent classes constituted by these x attributes. If Q_j only contains one e_k but item j fails to distinguish between examinees who have mastered attribute k and those who have not, Q_j is said to have a level-0 design.

Two common designs in practice, the level-1 and level-2 designs, are discussed in detail below. Additionally, the level-3 design is briefly introduced as an illustration for understanding the level- x design. The section is wrapped up with a discussion of the level-0 design, a crucial case where the assignment established by the level- x design breaks down. Lemma 1 is then presented as a remedy for the level-0 design to differentiate examinees who have mastered attribute k from those who have not. Because the level-0 design is a case requiring special treatment, it is not included when referring to the level- x design in this study.

4.3.2.1 Level-1 Design

Based on Definition 1, a Q_j is said to have a level-1 design if it can distinguish examinees who have mastered a particular attribute, say attribute k , from those who have not. Q_1 in Table 4.2 has a level-1 design. In contrast, although Q_2 contains two e_k (i.e., e_2 and e_1), it can only distinguish examinees who have mastered A2 from those who have not. Hence, Q_2 also has a level-1 design.

In conclusion, an MC item with a level-1 design can only determine the mastery of a single attribute and thus can be treated as an extension of a binary item under the

DINA model, allowing for the utilization of an e_k to define either the key or a coded distractor. It is important to note that a Q_j with a level-1 design can include other coded distractors as long as the item stays useful and proper.

4.3.2.2 Level-2 Design

A level-2 design allows a Q_j to effectively identify the mastery of two attributes, say k and k' , where $k \neq k'$. Consequently, Q_j can distinguish examinees into four groups (i.e., $2^2 = 4$): those who have mastered only attribute k , only attribute k' , both attributes k and k' , and neither attributes k nor k' . Q_3 , shown in Table 4.2, has a level-2 design. Although Q_2 and Q_4 contain two distinct e_k , they do not have a level-2 design because they fail to identify examinees who have mastered both A1 and A2. The only way to achieve a level-2 design for these items is to define a coded option as $e_1 + e_2$ subjective to $e_1 + e_2 \preceq q_j^{(H_j^*)}$. The corresponding Q-matrix must contain the following structure:

$$\begin{pmatrix} 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \end{pmatrix}. \quad (4.3)$$

Hence, it is important to note that an item conforming to the level-2 design should include not only e_k and $e_{k'}$ but also $e_k + e_{k'}$. A MC item whose Q-matrix has a level-2 design may include additional coded options as long as the item is proper and useful.

4.3.2.3 Level-3 Design

Likewise, a Q_j with the level-3 design can differentiate among examinees who have mastered only one of the three attributes, namely, k , k' , or k'' , those who have mastered any two of them, all of them, and none of them, where $k \neq k' \neq k''$. After column

swapping, a Q-matrix with a level-3 design has the following structure:

$$\begin{pmatrix} 1 & 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & 0 & \dots & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 \\ 0 & 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 1 & 0 & \dots & 0 \end{pmatrix}.$$

However, in practice, it is not common for a MC item to have seven or more options. Therefore, the level-3 design is not further discussed in the current paper.

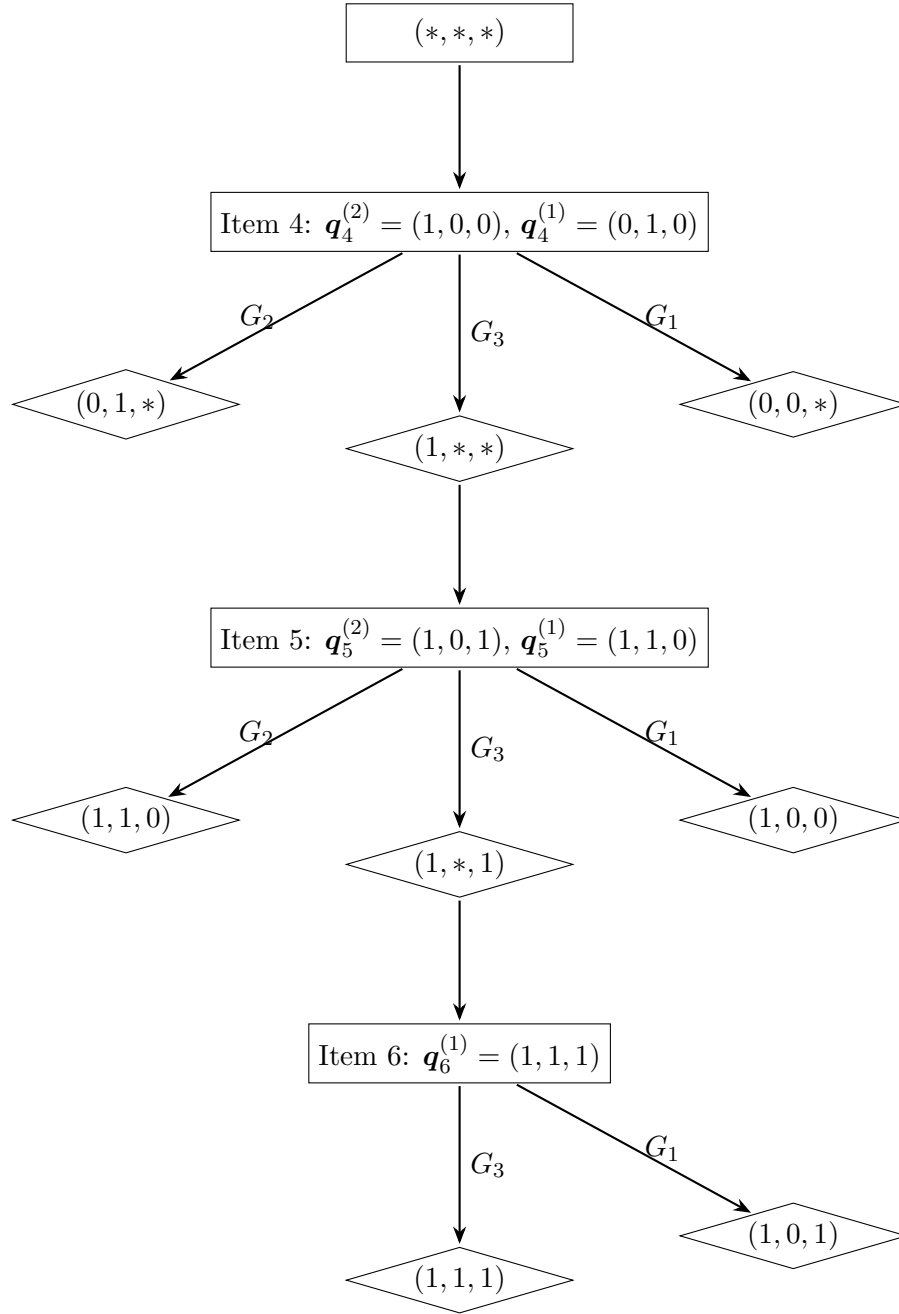
4.3.2.4 Level-0 Design

As introduced above, a $\mathbf{Q}_j$ with the level- x design, for $x \geq 1$, not only includes $\mathbf{e}_k$ but also distinguishes examinees who have mastered attribute k from those who have not. However, including an $\mathbf{e}_k$ does not necessarily guarantee the conformation of the level-1 design. A $\mathbf{Q}_j$ that contains an $\mathbf{e}_k$ but fails to make the distinction is said to have the level-0 design. Based on Definition 1, a $\mathbf{Q}_j$ has a level-0 design if distractor l is coded as $\mathbf{e}_k$ but $\mathbf{q}_j^{(H_{j*})} \neq \mathbf{e}_k$, as shown in the above case c). $\mathbf{Q}_4$ in Table 4.2, for example, is an item that has such a design.

Therefore, to distinguish examinees who have mastered attribute k from those who have not, it is recommended to begin with a $\mathbf{Q}_j$ that contains $\mathbf{e}_k$ and has a level- x design. However, if such a $\mathbf{Q}_j$ is not available and $\mathbf{e}_k$ is included in items that only have level-0 designs, the immediate question is that whether it is possible to separate examinees who have and have not mastered attribute k apart by attaching more items to a level-0-design item. If the answer is yes, what is the procedure one should follow?

Lemma 1 is developed to answer the above question. Before it is presented, an example, depicted in Figure 4.1, is used to illustrate and facilitate the reasoning.

Figure 4.1: An Example Set of Items to Distinguish between Mastery and Non-mastery of A2



Revisit Q_4 in Table 4.2. This item conforms to the level-0 design with the key and a coded distractor being defined by $q_4^{(2)} = (1, 0, 0)$ and $q_4^{(1)} = (0, 1, 0)$, respectively. After the item is administered, both $(1, 1, *)$ and $(1, 0, *)$, where $*$ denotes the entry can be either 0 or 1, are assigned to G_3 because item 4 cannot separate them apart based on A2. Notice that in this case, whether the next item requires A1 does not affect the differentiation between $(1, 1, *)$ and $(1, 0, *)$ because both have already mastered it. However, the requirement of A3 will mask the mastery of the A2 between $(1, 1, 0)$ and $(1, 0, 0)$. Hence, to distinguish between $(1, 1, *)$ and $(1, 0, *)$, an item with a coded option defined by $(*, 1, 0)$ is needed. If $(*, 1, 0)$ defines the key, then $(1, 1, *)$ will be assigned to G_3 , and $(1, 0, *)$ will be assigned to G_1 or G_2 . As a result, all the examinees can be separated apart based on A2. To illustrate a more complex scenario, assume that such an item is not available in the test. Rather, item 5 with a coded distractor and the key defined by $(1, 1, 0)$ and $(1, 0, 1)$, respectively, is presented to the examinees. After this item is administered, $(1, 1, 1)$ and $(1, 0, 1)$ are still mixed in G_3 and thus, more items are needed to separate them apart based on A2. Again, whether the next item requires A1 and A3 does not influence the differentiation between $(1, 1, 1)$ and $(1, 0, 1)$. The requirement is that this item should consist of a coded option defined by $(*, 1, *)$. As an example, item 6 in Figure 4.1 has the key defined by $(1, 1, 1)$, and thereby can be used to distinguish α from α^* . Differing from the level-1 and level-2 designs, where only 1 item is needed to distinguish the examinees who have and have not mastered A2, in this example, two more items are required when the item containing e_k has a level-0 design.

The example above can be generalized as follows. Suppose item j contains e_k with a level-0 design used to distinguish the mastery of attribute k . After this item is administered, examinees are categorized into three classes, as shown in Table 4.1: G_1 (not mastered attribute k), G_2 (mastered attribute k), and G_3 (a mix of $\alpha \succeq e_k + q_j^{(H_j^*)}$ and $\alpha^* \succeq q_j^{(H_j^*)}$ but $\alpha^* \not\succeq e_k$). Here, G_3 consists of examinees who have and have not mastered attribute k and thus, additional items are required to distinguish between α

and α^* in G_3 .

Let $\mathbf{q} = \mathbf{q}_j^{(H_j^*)}$ denote the attributes that examinees in G_3 have already mastered, which do not contribute to distinguishing between α and α^* . From the remaining items, select an item that has a coded option defined by $\mathbf{e}_k + \mathbf{q}_0$, where $\mathbf{q}_0 \preceq \mathbf{q}$. This item is required to differentiate α from α^* . Denote this item as j' . If $\mathbf{q}_{j'}^{(H_{j'}^*)} \succeq \mathbf{e}_k$, based on Table 4.1, this item can explicitly divide examinees in G_3 into two groups based on the mastery of attribute k . However, if $\mathbf{q}_{j'}^{(H_{j'}^*)} \not\succeq \mathbf{e}_k$, $\alpha \succeq \mathbf{e}_k + \mathbf{q}_0 + \mathbf{q}_{j'}^{(H_{j'}^*)}$ and $\alpha^* \succeq \mathbf{q}_0 + \mathbf{q}_{j'}^{(H_{j'}^*)}$ but $\alpha^* \not\succeq \mathbf{e}_k$ will be left in G_3 and remain un-separated. Update $\mathbf{q} = \mathbf{q} \vee \mathbf{q}_{j'}^{(H_{j'}^*)}$. Thus, an item who has a coded option defined by $\mathbf{e}_k + \mathbf{q}_0$, where $\mathbf{q}_0 \preceq \mathbf{q}$, is required to distinguish α from α^* . The process continues with updating $\mathbf{q}_0$ and finding additional items until the mastery of attribute k is determined.

The process detailed above outlines the need for additional items to assess the mastery of attribute k when dealing with an item characterized by a level-0 design. Lemma 1, shown below, not only summarizes this process, but makes it more comprehensive, as item j' might not be unique in practice. The proof of Lemma 1 is presented in Appendix B.

Lemma 1. *Suppose $\mathbf{Q}_{j_0}$ of item j_0 involving $\mathbf{e}_k$ conforms to the level-0 design; that is, $\mathbf{q}_{j_0}^{(H_{j_0}^*)} \not\succeq \mathbf{e}_k$. Consider the following algorithm:*

Step 0 *Let $\mathbf{q} = \mathbf{q}_{j_0}^{(H_{j_0}^*)}$ and initialize set $\mathcal{J} = \{j_0\}$.*

Step 1 *For the rest of test items, define set $\mathcal{J}_k$ as $\mathcal{J}_k = \{j | \mathbf{Q}_j \text{ contains } \mathbf{e}_k + \mathbf{q}_0, \text{ where } \mathbf{q}_0 \preceq \mathbf{q}\}$. If $\mathcal{J}_k = \emptyset$, stop; otherwise, update $\mathcal{J} = \mathcal{J} \cup \mathcal{J}_k$ and proceed to Step 2.*

Step 2 *If there exists a $j \in \mathcal{J}_k$ such that $\mathbf{q}_j^{(H_j^*)} \succ \mathbf{e}_k$, stop; otherwise, update $\mathbf{q} = \mathbf{q} \vee (\bigvee_{t=1}^{|\mathcal{J}_k|} \mathbf{q}_{j_t}^{(H_{j_t}^*)})$ and repeat Steps 1 and 2.*

Examinees who have and have not mastered attribute k can be separated apart by administering all the items in set $\mathcal{J}$ if and only if the above algorithm stops with $\mathcal{J}_k \neq \emptyset$.

Table 4.3: Relationship Between the Level- x Design and Minimum H_j^*

Level- x Design	Minimum H_j^*
1	1
2	3
3	7
4	15

Overall, a level- x design is characterized not only by the inclusion of x distinct $\mathbf{e}_k$ in $\mathbf{Q}_j$, but more importantly, by the capability of $\mathbf{Q}_j$ to differentiate among 2^x latent classes constituted by these x attributes. Under the MC-DINA model, when $\mathbf{Q}_j$ has a level- x design, where $x \geq 1$, the essential rule is to include all $\mathbf{e}_k$, as well as the sums of any two, three, and eventually all of $\mathbf{e}_k$. Table 4.3 displays the minimum H_j^* required to achieve a level- x design, calculated by $2^x - 1$. In addition, the inequality below can be constructed to describe the relationship between H_j and x :

$$2^x - 1 \leq H_j,$$

where $x \geq 1$. Further details about the relation between 2^x and H_j can be found in Wang et al. (2024).

4.3.3 Completeness of $\mathbf{Q}_{\text{MC}}$ for the MC-DINA Model

The definition of completeness established for $\mathbf{Q}_{\text{B}}$ under the DINA model is applicable to $\mathbf{Q}_{\text{MC}}$, provided specific setups and modifications. Notice that the q-vectors of the coded options in the MC context are only partially ordered, and as a result, $E(X|\boldsymbol{\alpha})$ calculated based on the original scale is no longer meaningful and needs to be recoded such that each individual level of coded option is treated as a random variable. Specifically, let I_{jl} be an indicator variable for response level l of item j . It can then be

expressed as follows:

$$I_{jl} = \begin{cases} 1 & \text{if } l \text{ is selected} \\ 0 & \text{if } l \text{ is not selected.} \end{cases}$$

Because I_{jl} is binary, it follows that

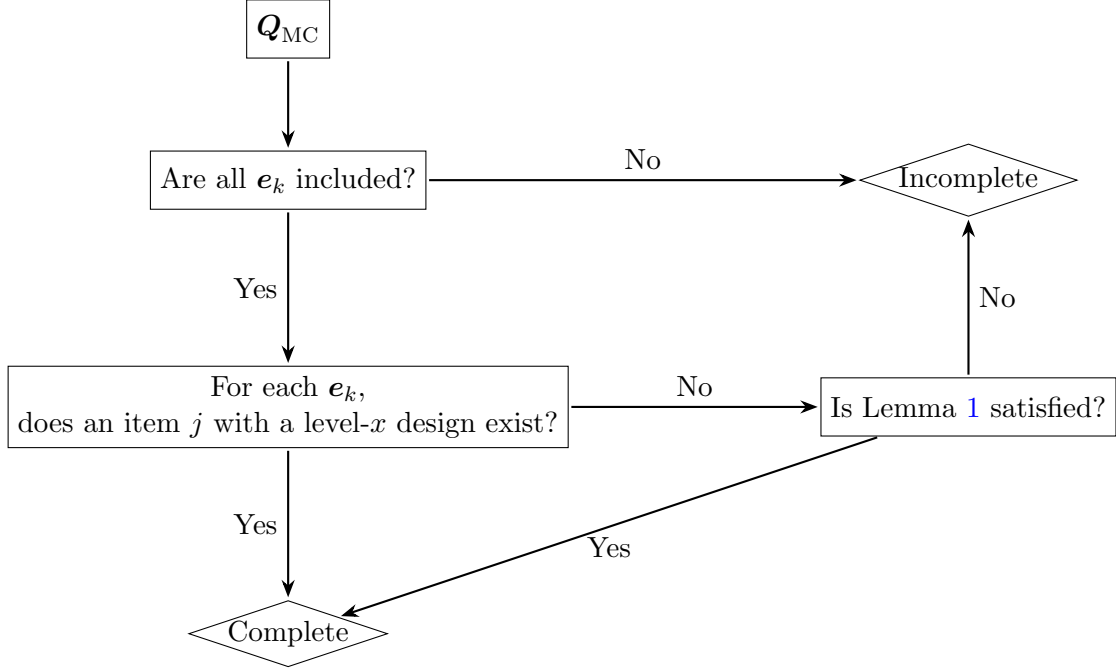
$$E(I_{jl}|\boldsymbol{\alpha}) = P(I_{jl} = 1|\boldsymbol{\alpha}) = P(X_j = l|\boldsymbol{\alpha}).$$

Let $\boldsymbol{\varsigma}(\boldsymbol{\alpha}) = E(\mathbf{I}|\boldsymbol{\alpha})$, where $\mathbf{I} = (I_{11}, \dots, I_{JH_J^*})'$. Here, $\boldsymbol{\varsigma}(\boldsymbol{\alpha})$ is no longer a $J \times 1$ as in the context of binary responses but rather a $\sum_{j=1}^J H_j^* \times 1$ vector in the MC context. Based on this set up, Definition 2 is developed to formally describe the completeness of $\mathbf{Q}_{MC}$ under the MC-DINA model.

Definition 2. A Q -matrix for MC items is said to be complete if it guarantees the identifiability of all possible latent classes. Specifically, let $\boldsymbol{\varsigma}(\boldsymbol{\alpha}) = E(\mathbf{I}|\boldsymbol{\alpha})$ be the vector of expected selections for each coded option, conditioning on attribute profile $\boldsymbol{\alpha}$. Then, $\mathbf{Q}_{MC}$ is complete if for any two $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^*$, $\boldsymbol{\varsigma}(\boldsymbol{\alpha}) = \boldsymbol{\varsigma}(\boldsymbol{\alpha}^*) \Rightarrow \boldsymbol{\alpha} = \boldsymbol{\alpha}^*$.

Definition 2 is then used to establish the sufficient and necessary conditions for the completeness of $\mathbf{Q}_{MC}$, as formally formulated in Theorem 1. This theorem is inspired by the condition for the completeness of $\mathbf{Q}_B$, reiterated here that $\mathbf{Q}_B$ is complete if and only if $\mathbf{Q}_B$ includes all the K $\mathbf{e}_k$ (for more information, please see the review subsection “Completeness of $\mathbf{Q}_B$ for the DINA Model”). However, differing from the binary case, where one item can only associate with at most one $\mathbf{e}_k$, an MC item may involve multiple $\mathbf{e}_k$, either through the key or coded distractors and thus, additional constraints, as depicted in Theorem 1, are required for organizing $\mathbf{e}_k$ within an MC item. The proof of Theorem 1 is presented in Appendix B.

Theorem 1. Suppose $\mathbf{Q}_{MC}$ is the Q -matrix of items conforming to the MC-DINA model. $\mathbf{Q}_{MC}$ is complete if and only if $\mathbf{Q}_{MC}$ includes all $\mathbf{e}_k$, where $k = 1, \dots, K$, and

Figure 4.2: Flowchart for Evaluating the Completeness of Q_{MC} 

for each k :

A1. At least one item j containing it has a level- x design.

A2. All items containing it have the level-0 design, and Lemma 1 is satisfied.

The flexibility in organizing e_k to achieve Q-matrix completeness, as described in Theorem 1, may present some challenges for practitioners. Hence, to improve its accessibility, a flowchart illustrating the procedure of how the theorem can be applied to evaluate the completeness of Q_{MC} is provided in Figure 4.2.

Theorem 1 further indicates that if the q-vectors of the item keys contain all possible e_k (i.e., Level-1 design), then Q_{MC} is complete, regardless of the q-vectors associated with the coded distractors. This result can serve as a first-line criterion for users to quickly assess the completeness of Q_{MC} before applying the full procedure outlined in Theorem 1. The following corollary 1 is established to outline this practically useful guideline.

Corollary 1. Suppose Q_B is complete for data conforming to the DINA model. Then

any useful and proper $\mathbf{Q}_{MC}$ augmented from $\mathbf{Q}_B$ with data conforming to the MC-DINA model is also complete.

Two examples of complete $\mathbf{Q}_{MC}$ are presented in Table 4.4. Both $\mathbf{Q}_{MC1}$ and $\mathbf{Q}_{MC2}$ include all possible $\mathbf{e}_k$ and they satisfy A1 and A2, respectively. Their corresponding

Table 4.4: Examples of Complete $\mathbf{Q}_{MC}$ with $K = 3$ and $H_j = 5$

$\mathbf{Q}_{MC1}$					$\mathbf{Q}_{MC2}$				
Item	Option	α_1	α_2	α_3	Item	Option	α_1	α_2	α_3
1	2	1	0	0	1	1	1	0	0
1	1	0	1	0	2	1	0	1	0
2	3	0	1	1	3	2	1	1	0
2	2	0	1	0	3	1	0	0	1
2	1	0	0	1	4	1	1	1	1

expected observed responses for each latent class, $\varsigma(\alpha)$, are presented in Table 4.5. The results show that the $\varsigma(\alpha)$ across all α are distinct under the MC-DINA assumption $1 - \sum_{c \neq l} \epsilon_{jcl} > \epsilon_{jcl}$.

Notice that $\mathbf{Q}_{MC1}$ requires fewer items than a complete $\mathbf{Q}_B$ with $K = 3$ items, while $\mathbf{Q}_{MC2}$ requires more items, which means the flexibility in organizing $\mathbf{e}_k$ may either reduce or increase the test length to achieve a complete $\mathbf{Q}_{MC1}$. If item j containing $\mathbf{e}_k$ satisfies A1, only one item is needed to differentiate the examinees who have mastered attribute k from those who have not. However, satisfying A2 requires more items than A1 to make the distinction. Therefore, the more $\mathbf{e}_k$ are contained in items with a level- x design, the shorter the test needed to achieve the completeness of the $\mathbf{Q}_{MC}$. Conversely, the more $\mathbf{e}_k$ contained in items with a level-0 design, the longer the test required to achieve the completeness of the $\mathbf{Q}_{MC}$.

Table 4.5: $\varsigma(\alpha)$ of Items Shown in Table 4.4

		α							
		(000)	(100)	(010)	(001)	(110)	(101)	(011)	(111)
Q_{MC1}	$E(I_{11} \alpha)$	$\frac{1}{5}$	ϵ_{112}	$1-\sum_{c\neq 1}\epsilon_{1c1}$	$\frac{1}{5}$	ϵ_{112}	ϵ_{112}	$1-\sum_{c\neq 1}\epsilon_{1c1}$	ϵ_{112}
	$E(I_{12} \alpha)$	$\frac{1}{5}$	$1-\sum_{c\neq 2}\epsilon_{1c2}$	ϵ_{121}	$\frac{1}{5}$	$1-\sum_{c\neq 2}\epsilon_{1c2}$	$1-\sum_{c\neq 2}\epsilon_{1c2}$	ϵ_{121}	$1-\sum_{c\neq 2}\epsilon_{1c2}$
	$E(I_{21} \alpha)$	$\frac{1}{5}$	$\frac{1}{5}$	ϵ_{212}	$1-\sum_{c\neq 1}\epsilon_{2c1}$	ϵ_{212}	$1-\sum_{c\neq 1}\epsilon_{2c1}$	ϵ_{213}	ϵ_{213}
	$E(I_{22} \alpha)$	$\frac{1}{5}$	$\frac{1}{5}$	$1-\sum_{c\neq 2}\epsilon_{2c2}$	ϵ_{221}	$1-\sum_{c\neq 2}\epsilon_{2c2}$	ϵ_{221}	ϵ_{223}	ϵ_{223}
	$E(I_{23} \alpha)$	$\frac{1}{5}$	$\frac{1}{5}$	ϵ_{232}	ϵ_{231}	ϵ_{232}	ϵ_{231}	$1-\sum_{c\neq 3}\epsilon_{2c3}$	$1-\sum_{c\neq 3}\epsilon_{2c3}$
Q_{MC2}	$E(I_{11} \alpha)$	$\frac{1}{5}$	$1-\sum_{c\neq 1}\epsilon_{1c1}$	$\frac{1}{5}$	$\frac{1}{5}$	$1-\sum_{c\neq 1}\epsilon_{1c1}$	$1-\sum_{c\neq 1}\epsilon_{1c1}$	$\frac{1}{5}$	$1-\sum_{c\neq 1}\epsilon_{1c1}$
	$E(I_{21} \alpha)$	$\frac{1}{5}$	$\frac{1}{5}$	$1-\sum_{c\neq 1}\epsilon_{2c1}$	$\frac{1}{5}$	$1-\sum_{c\neq 1}\epsilon_{2c1}$	$\frac{1}{5}$	$1-\sum_{c\neq 1}\epsilon_{2c1}$	$1-\sum_{c\neq 1}\epsilon_{2c1}$
	$E(I_{31} \alpha)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$1-\sum_{c\neq 1}\epsilon_{3c1}$	ϵ_{312}	$1-\sum_{c\neq 1}\epsilon_{3c1}$	$1-\sum_{c\neq 1}\epsilon_{3c1}$	ϵ_{312}
	$E(I_{32} \alpha)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	ϵ_{321}	$1-\sum_{c\neq 2}\epsilon_{3c2}$	ϵ_{321}	ϵ_{321}	$1-\sum_{c\neq 2}\epsilon_{3c2}$
	$E(I_{41} \alpha)$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$\frac{1}{5}$	$1-\sum_{c\neq 1}\epsilon_{4c1}$

4.4 Objective 2: Identifications of Each Latent Class under the MC-DINA Model

4.4.1 Review of Identification of α_i under the binary responses

Xu et al. (2016) proved that the test containing K items whose q-vectors are $\mathbf{e}_k$, where $k = 1, \dots, K$ is necessary and sufficient to identify $\mathbf{0}_K$ when responses are binary. Here, $\mathbf{e}_k$ is a $1 \times K$ unit vector, with the k th entry being 1 and all other entries being 0. Consider an examinee whose attribute profile $\alpha_i \neq \mathbf{0}_K$. Suppose k out of K elements in α_i equal to 1. Let $\mathbf{m}_k$ be a vector such that the first k elements are 1, and the rest are 0. Then via a column swapping of the attribute labels, α_i is equivalent to $\mathbf{m}_k$. To identify $\mathbf{m}_k$, it is necessary and necessary for there to exist a set of item(s) requiring all of the first k but none of the last $K - k$ attributes. Additionally, there must exist $K - k$ items in the test such that the corresponding Q-matrix takes the form below

$$\mathbf{Q} = \begin{pmatrix} \mathbf{O}_* & \mathbf{I}_{K-k} \end{pmatrix}_{(K-k) \times K},$$

where $\mathbf{O}_*$ denotes an unspecified $(K - k) \times k$ binary matrix and $\mathbf{I}_{K-k}$ is the identify matrix.

4.4.2 Identification of α_i under the MC-DINA model

When responses stem from MC items, the identification condition is determined at the option level rather than the item level. This is because of the rich diagnostic information contained within the coded distractors. Two lemmas regarding the identification of $\alpha = \mathbf{0}_K$ and $\alpha = \mathbf{m}_k$ are briefly reviewed in the following, and their proofs are shown in Appendix B.

Lemma 2. *Suppose $\mathbf{Q}_{MC}$ is the Q-matrix of a test of MC items conforming to the MC-DINA model. The attribute profile $\alpha = \mathbf{0}_K$ can be identified by $\mathbf{Q}_{MC}$ if and only if $\mathbf{Q}_{MC}$ includes all $\mathbf{e}_k$, regardless of whether they define the keys or coded distractors.*

Same as the binary case, to find the identification condition of $\alpha \neq \mathbf{0}_K$, it is equivalent to find the identification condition of $\mathbf{m}_k$.

Lemma 3. *Suppose $\mathbf{Q}_{MC}$ is the Q -matrix of a test of MC items conforming to the MC-DINA model. The attribute profile $\alpha = \mathbf{m}_k$, where $k = 1, \dots, K$, can be identified by $\mathbf{Q}_{MC}$ if and only if the test satisfies the following conditions:*

C1 There exists a set of k' items, where $1 \leq k' \leq k$, such that item j , $j = 1, 2, \dots, k'$, includes a coded option defined by $\mathbf{q}_j$, subject to $\bigcup_{j=1}^{k'} \mathbf{q}_j = \mathbf{m}_k$. Let O_1 be the set consisting of these k' q -vectors.

C2 There exists a set of k'' items, where $1 \leq k'' \leq K - k$, such that they consist of $K - k$ coded options with q -vectors taking the following form,

$$\begin{pmatrix} \mathbf{Q}_* & \mathbf{I}_{K-k} \end{pmatrix}$$

where $\mathbf{Q}_*$ is a $(K - k) \times k$ matrix with arbitrary entries. The keys of these items cannot be any $\mathbf{q} \in O_1$. Let O_2 be the set consisting of these $K - k$ q -vectors.

Table 4.6: Four Examples of $\mathbf{Q}_{MC}$

	$\mathbf{Q}_1$	$\mathbf{Q}_2$
$\mathbf{Q}_{MC1}$	$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$
$\mathbf{Q}_{MC2}$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$
$\mathbf{Q}_{MC3}$	$\begin{pmatrix} 1 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 0 & 1 \end{pmatrix}$
$\mathbf{Q}_{MC4}$	$\begin{pmatrix} 1 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 1 & 1 & 1 \end{pmatrix}$

Table 4.6 displays four examples of $\mathbf{Q}_{\text{MC}}$ to illustrate identification conditions of $\alpha_1 = \mathbf{0}_3$ and $\alpha_2 = (1, 1, 0)'$. According to Lemma 2, the identification of $\mathbf{0}_3$ requires $\mathbf{Q}_{\text{MC}}$ to include all e_k . Correspondingly, $\mathbf{Q}_{\text{MC1}}$ and $\mathbf{Q}_{\text{MC2}}$ can identify α_1 . Since e_1 and e_2 are missing from $\mathbf{Q}_{\text{MC3}}$ and $\mathbf{Q}_{\text{MC4}}$, they are unable to identify α_1 . In contrast, $\mathbf{Q}_{\text{MC1}}$, $\mathbf{Q}_{\text{MC3}}$, and $\mathbf{Q}_{\text{MC4}}$ can be used to identify α_2 . However, $\mathbf{Q}_{\text{MC2}}$ falls short since two coded options, measuring the mastered attributes (A1 and A2), are from the same item, preventing the determination of mastery of A2 through $\mathbf{Q}_1$.

4.5 Objective 3: Q-Optimal Design for CD-CAT with the MC-DINA Model

4.5.1 Review of the Q-Optimal Design for CD-CAT with the DINA Model

The first non-asymptotic theory of the optimal selection of initial items in CD-CAT was proposed by Xu et al. (2016), which achieves optimal classification results within finite items. In their study, the Q-matrix $\mathbf{Q}_B$ of a test of J dichotomous items is said to be Q-optimal if it satisfies the condition:

$$P(\hat{\alpha} = \alpha^* | \alpha^*, \mathbf{Q}_B) > \max_{a \neq \alpha^*} P(\hat{\alpha} = a | \alpha^*, \mathbf{Q}_B). \quad (4.4)$$

Based on Equation 4.4, Xu et al. (2016) proved that when data conform to the DINA model, a test consisting of K items is Q-optimal if and only if, for each examinee,

the Q -matrix of the test items takes the following form after column swapping:

$$\mathbf{Q}_B = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ q_{2,1} & 1 & 0 & \dots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ q_{K-1,1} & \dots & q_{K-1,K-2} & 1 & 0 \\ q_{K,1} & \dots & q_{K,K-2} & q_{K,K-1} & 1 \end{pmatrix}, \quad (4.5)$$

where the lower diagonal element, $q_{j,k}$, $j = k + 1, \dots, K$, is set to 0 if the examinee answers the k th item incorrectly and either 0 or 1 if the k th item is answered correctly.

4.5.2 Minimum Test Length for Complete $\mathbf{Q}_{MC}$ under the MC-DINA Model

In situations where minimizing the number of items in a test is desirable, such as tests with time constraints, two immediate questions arise: what is the minimal number of required items to achieve the completeness of $\mathbf{Q}_{MC}$, and is there a particular design that test developers must adopt? Corollary 2 is developed to answer these questions.

Corollary 2. *Suppose each MC item conforms to a level-2 design and $3 \leq H_j < 7$. Then, the minimum test length of a complete $\mathbf{Q}_{MC}$ is $\lceil \frac{K}{2} \rceil$, achieved by measuring $2 \times \lfloor \frac{K}{2} \rfloor$ attributes using $\lfloor \frac{K}{2} \rfloor$ MC items.*

The level- x design is specifically devised for MC items, allowing the incorporation of multiple e_k within a single MC item. To reduce the number of test items, it is optimal for MC items to utilize level-2 or higher level- x designs. Given that $3 \leq H_j < 7$ appears to be the commonly used range for H_j in practice, the maximum achievable level of x is 2, as indicated by Equation 4.3. Therefore, to achieve the minimum test length with $3 \leq H_j < 7$, $\lfloor \frac{K}{2} \rfloor$ of the MC items should employ level-2 designs. For instance, $\mathbf{Q}_{MC1}$ shown in Table 4.4 represents a complete Q -matrix with the minimum test length $\lceil \frac{3}{2} \rceil = 2$.

4.5.3 Q-Optimal Design for CD-CAT with the MC-DINA Model

The definition of a Q-optimal test for binary items, as specified in Equation 4.4, can be extended to a test with $\mathbf{Q}_{MC}$. It is tailored as follows to fit the context of MC items.

Definition 3. *A test for MC items with $\mathbf{Q}_{MC}$ is Q-optimal if it satisfies the following condition:*

$$P(\hat{\alpha} = \alpha^* | \alpha^*, \mathbf{Q}_{MC}) > \max_{\mathbf{a} \neq \alpha^*} P(\hat{\alpha} = \mathbf{a} | \alpha^*, \mathbf{Q}_{MC}). \quad (4.6)$$

The Q-optimal design for MC items is achieved by adopting the level- x design to guarantee the minimum test length. The study focuses on the case of level-2 design for its practical usefulness and further uses it to demonstrate the theoretical foundation of the Q-optimal condition. In establishing the Q-optimal condition for MC items, two forms based on the level-2 design are formulated in Equations 4.7 and 4.8, for an odd and even K , respectively. Differing from the binary case, where the Q-optimality is always achieved at $J = K$, the number of items required in the context of MC items varies and J^* is used to denote the required test length to achieve the optimal classification.

Form 1: When K is an odd number

$$\mathbf{Q}_{\text{MC}} = \begin{pmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \\ \vdots \\ \mathbf{Q}_{\frac{K-1}{2}} \\ \mathbf{Q}_{J^*} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 & 0 \\ \hline q_{2,1}^{(3)} & q_{2,2}^{(3)} & 1 & 1 & \dots & 0 & 0 & 0 \\ q_{2,1}^{(2)} & q_{2,2}^{(2)} & 1 & 0 & \dots & 0 & 0 & 0 \\ q_{2,1}^{(1)} & q_{2,2}^{(1)} & 0 & 1 & \dots & 0 & 0 & 0 \\ \hline \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \\ \hline q_{\frac{K-1}{2},1}^{(3)} & q_{\frac{K-1}{2},2}^{(3)} & q_{\frac{K-1}{2},3}^{(3)} & q_{\frac{K-1}{2},4}^{(3)} & \dots & 1 & 1 & 0 \\ q_{\frac{K-1}{2},1}^{(2)} & q_{\frac{K-1}{2},2}^{(2)} & q_{\frac{K-1}{2},3}^{(2)} & q_{\frac{K-1}{2},4}^{(2)} & \dots & 1 & 0 & 0 \\ q_{\frac{K-1}{2},1}^{(1)} & q_{\frac{K-1}{2},2}^{(1)} & q_{\frac{K-1}{2},3}^{(1)} & q_{\frac{K-1}{2},4}^{(1)} & \dots & 0 & 1 & 0 \\ \hline q_{J^*,1}^{(l)} & q_{J^*,2}^{(l)} & q_{J^*,3}^{(l)} & q_{J^*,4}^{(l)} & \dots & q_{J^*,K-2}^{(l)} & q_{J^*,K-1}^{(l)} & 1 \end{pmatrix}. \quad (4.7)$$

In the matrix shown in Equation 4.7, $q_{j,k}^{(l)}$, where $j = 2, \dots, J^*$, $l = 1, \dots, H_j^*$, and $k = 1, \dots, K - 1$, takes the value 0 if the option requiring the corresponding attribute k of item $\lceil \frac{k}{2} \rceil$ is not selected and takes the value 0 or 1 if that option is selected. The levels of the coded option of item J^* are denoted in a general manner using l , and $l = 1, \dots, H_{J^*}^*$, which means this coded option can be either the key or a coded distractor. If this coded option is a distractor (i.e., $l < H_{J^*}^*$), based on Lemma 3, $\mathbf{q}_{J^*}^{(H_{J^*}^*)}$ must satisfy $\mathbf{q}_{J^*}^{(H_{J^*}^*)} \succ \mathbf{e}_K$. In this case, $J^* = \frac{K+1}{2}$ items are required to achieve optimal classification.

Form 2: When K is an even number

$$\mathbf{Q}_{\text{MC}} = \begin{pmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \\ \vdots \\ \mathbf{Q}_{J^*} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 0 & 0 & \dots & 0 & 0 \\ 0 & 1 & 0 & 0 & \dots & 0 & 0 \\ \hline q_{2,1}^{(3)} & q_{2,2}^{(3)} & 1 & 1 & \dots & 0 & 0 \\ q_{2,1}^{(2)} & q_{2,2}^{(2)} & 1 & 0 & \dots & 0 & 0 \\ q_{2,1}^{(1)} & q_{2,2}^{(1)} & 0 & 1 & \dots & 0 & 0 \\ \hline \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \hline q_{J^*,1}^{(3)} & q_{J^*,2}^{(3)} & q_{J^*,3}^{(3)} & q_{J^*,4}^{(3)} & \dots & 1 & 1 \\ q_{J^*,1}^{(2)} & q_{J^*,2}^{(2)} & q_{J^*,3}^{(2)} & q_{J^*,4}^{(2)} & \dots & 1 & 0 \\ q_{J^*,1}^{(1)} & q_{J^*,2}^{(1)} & q_{J^*,3}^{(1)} & q_{J^*,4}^{(1)} & \dots & 0 & 1 \end{pmatrix}. \quad (4.8)$$

In the above matrix, $q_{j,k}^{(l)}$ follows the same rule as described for the matrix in Equation 4.7. In this case, $J^* = \frac{K}{2}$.

The above Q-matrices conform to a level-2 design, but each $\mathbf{Q}_j$ does not strictly have a level-2 design, as the lower diagonal entries are not fixed at zero in CD-CAT. However, H_j^* remains three, except for the last item when K is odd. The ratio between H_j and H_j^* influences $P(X_{ij} = 0|\boldsymbol{\alpha})$ because $P(X_{ij} = 0|\boldsymbol{\alpha}) = \frac{H_j - H_j^*}{H_j}$. For $P(X_{ij} = 0|\boldsymbol{\alpha}) > P(X_{ij} = 0|\boldsymbol{\alpha}^*)$ to always hold, $\frac{H_j - H_j^*}{H_j} > \epsilon_{jcl}$ must be satisfied. However, $\frac{H_j - H_j^*}{H_j} > \epsilon_{jcl}$ remains uncertain if both $\frac{H_j - H_j^*}{H_j}$ and ϵ_{jcl} are less than .5. Since ϵ_{jcl} is never known in reality except for the model assumption $\epsilon_{jcl} < 0.5$, a sufficient and practical assumption, $\frac{H_j - H_j^*}{H_j} \geq 0.5$, or equivalently, $H_j \geq 2H_j^*$, is added to guarantee that $\frac{H_j - H_j^*}{H_j} > \epsilon_{jcl}$ holds. Given that the level-2 design is used to demonstrate Theorem 2, the assumption $6 \leq H_j < 14$ is imposed based on the relation between H_j and H_j^* shown in Table 4.3. It is important to note that items other than the first one may contain extra coded distractors in addition to those shown in above equations, if the response to the previous item indicates mastery of the attribute(s). However, the

presence of these additional coded distractors does not affect the Q-optimal procedure. Further discussion on this matter will be provided after the proof in Appendix B.

Then, based on Corollary 2, which outlines the minimum number of MC items to construct a complete $\mathbf{Q}_{MC}$, and Lemmas 2 and 3, which specifies the conditions for identifying each latent class, Theorem 2 is developed and proves that the Q-optimal design laid out above is a sufficient and necessary condition for achieving optimal classification for CD-CAT with MC items when $6 \leq H_j < 14$. Theorem 2 is outlined as follows and its proof is shown in Appendix B.

Theorem 2. *For MC items under the MC-DINA model, with $6 \leq H_j < 14$, a set of $J^* = \lceil \frac{K}{2} \rceil$ MC items is Q-optimal if and only if for each α , the Q-matrices of these test items take the minimum form shown in Equations 4.7 or 4.8 after column swapping.*

Established in Theorem 2 are the conditions of Q-optimal design for MC items with $6 \leq H_j < 14$. To address the specific conditions for cases with $H_j = 4$ and $H_j = 5$, two additional corollaries are proposed.

Corollary 3. *For MC items with $H_j = 4$ conforming to the MC-DINA model, $J^* = K$ MC items are needed to achieve the Q-optimality.*

Corollary 4. *Theorem 2 holds for MC items with $H_j = 5$ if $\epsilon_{jcl} \leq \frac{2}{5}$.*

The condition $\frac{2}{5} < \epsilon_{jcl} < 0.5$ is uncommon in practice because it usually implies poor item quality. Therefore, Theorem 2 remains applicable to MC items with $H_j = 5$ in most practical situations.

4.6 Simulation Studies

Two simulation studies were conducted to support Theorems 1 and 2. In Simulation Study I, the effect of a complete $\mathbf{Q}_{MC}$ with a level-2 design is compared with that of a complete $\mathbf{Q}_{MC}$ augmenting from a complete $\mathbf{Q}_B$. In Simulation Study II, the

performance of two item selection methods for CAT, with and without the proposed MC Q-optimal procedure was evaluated.

4.6.1 Simulation Study I

This simulation study aims to examine the validity of the completeness conditions proposed for $\mathbf{Q}_{\text{MC}}$. A large sample was generated for each of the latent classes to simulate scenarios where one examinee took the same test multiple times. The observed proportions of each estimated latent class are supposed to converge to $E(\hat{\boldsymbol{\alpha}}|\boldsymbol{\alpha})$, where $\boldsymbol{\alpha}$ is the true attribute profile. Hence, if the expected values of the latent classes differ and $E(\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}|\boldsymbol{\alpha})$ is the largest, then this Q-matrix is deemed complete.

4.6.1.1 Design

Two tests designed to measure four attributes ($K = 4$), resulting in a total of $4^2 = 16$ latent classes, were implemented. Both tests contained four MC items ($J = 4$) with five options ($H_j = 5$). According to Corollary 1, a complete Q-matrix for the DINA model must be complete for the MC-DINA model. Therefore, $\mathbf{Q}_{\text{MC1}}$ was designed in such a way that it consisted of all the $\mathbf{e}_k$, where $k = 1, \dots, 4$, and $H_j^* = 1$ across all items. $\mathbf{Q}_{\text{MC2}}$ was constructed according to Theorem 1, where all distinct K $\mathbf{e}_k$ were included and each $\mathbf{Q}_j$ follows a level-2 design. While Corollary 2 indicates that two items are sufficient for a complete $\mathbf{Q}_{\text{MC}}$ with level-2 designs, the two items with this design were repeated to equalize the test length for a fair comparison between $\mathbf{Q}_{\text{MC1}}$ and $\mathbf{Q}_{\text{MC2}}$, given that $\mathbf{Q}_{\text{MC1}}$ requires four items for completeness. $\mathbf{Q}_{\text{MC2}}$ is provided below:

$$\mathbf{Q}_{\text{MC2}} = \begin{pmatrix} \mathbf{Q}_1 \\ \mathbf{Q}_2 \\ \mathbf{Q}_3 \\ \mathbf{Q}_4 \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ \hline 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \hline 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Here, $\mathbf{Q}_{\text{MC1}}$ served as a benchmark to assess the effectiveness of $\mathbf{Q}_{\text{MC2}}$ as a complete Q-matrix.

10,000 examinees ($N = 10,000$) were generated from each latent class, and their responses to the test items were generated using the MC-DINA model. Specifically, the likelihood that examinee i in group $\eta_{ij} = l'$ made a response $X_{ij} = l$ for item j , denoted as $P(X_{ij} = l | \eta_{ij} = l')$, is expressed as follows:

$$P(X_{ij} = l | \eta_{ij} = l') = \begin{cases} \frac{1}{H_j} + \frac{H_j - H_j^* - 1}{H_j} I[l = 0] & \text{if } l' = 0 \\ \lambda_j & \text{if } l' > 0 \text{ and } l = l' \\ \frac{1 - \lambda_j}{H_j - 1} & \text{if } l' > 0 \text{ and } l \neq l', \end{cases}$$

In this equation, the parameter λ_j signifies the probability that an examinee's observed response was consistent with the nonzero ideal response. This parameter was randomly sampled from a uniform distribution, $\mathcal{U}(0.74, 0.9)$. If slipping occurred, the examinee had an equal probability of $\frac{1 - \lambda_j}{H_j - 1}$ to select any of the remaining options. Furthermore,

examinees whose ideal response was zero made a random selection from all H_j options. A previous study (Wang et al., 2024) indicates that the MC-DINA model has severe convergence issues when examinees are homogeneous, regardless of the sample size. Therefore, the MC-NPC method is employed to estimate the attribute profiles.

4.6.1.2 Evaluation Criteria

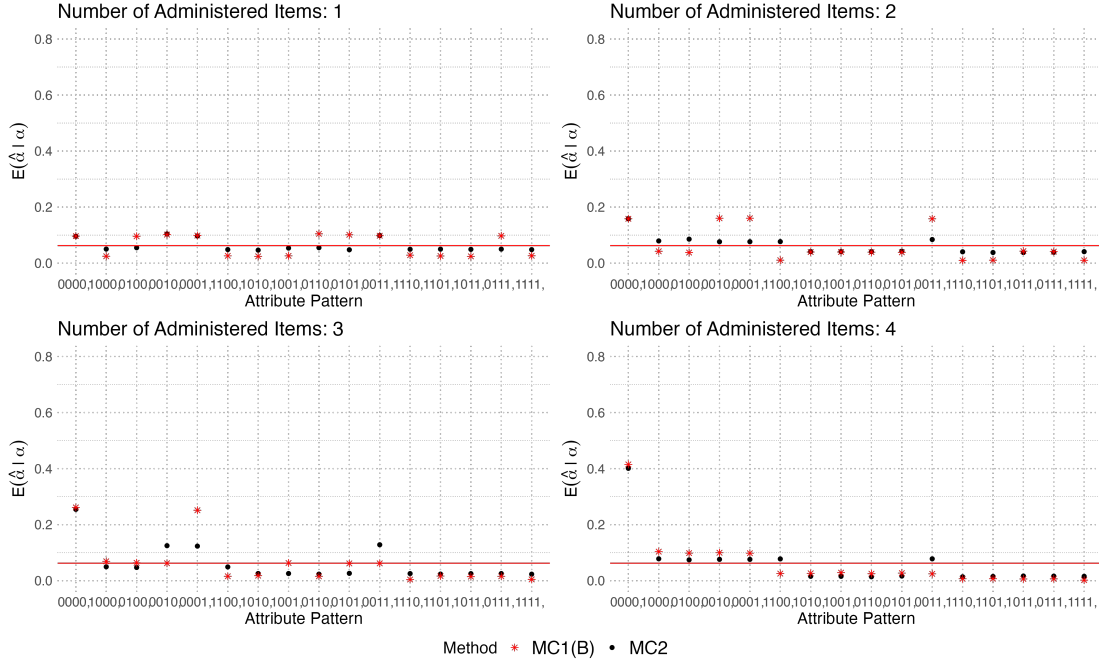
The proportion of examinees classified into the true latent class was calculated after each item was administered. In the current simulation study, the sample size ($N = 10,000$) was sufficiently large to infer that this observed proportion converged to $E(\hat{\alpha}|\alpha)$.

4.6.1.3 Results

The results from all latent classes are not presented here for two reasons: 1) latent classes that have mastered the same number of attributes are mathematically identical after swapping the attribute labels; 2) Based on the level-2 design, some latent classes (e.g., $(1, 0, 1, 0)'$ and $(1, 1, 1, 0)'$) are considered equivalent classes because they have nonzero ideal responses to four items. Consequently, only the results of latent classes $(0, 0, 0, 0)'$, $(1, 0, 0, 0)'$, $(1, 0, 1, 0)'$, and $(1, 1, 1, 1)'$ are presented in this section.

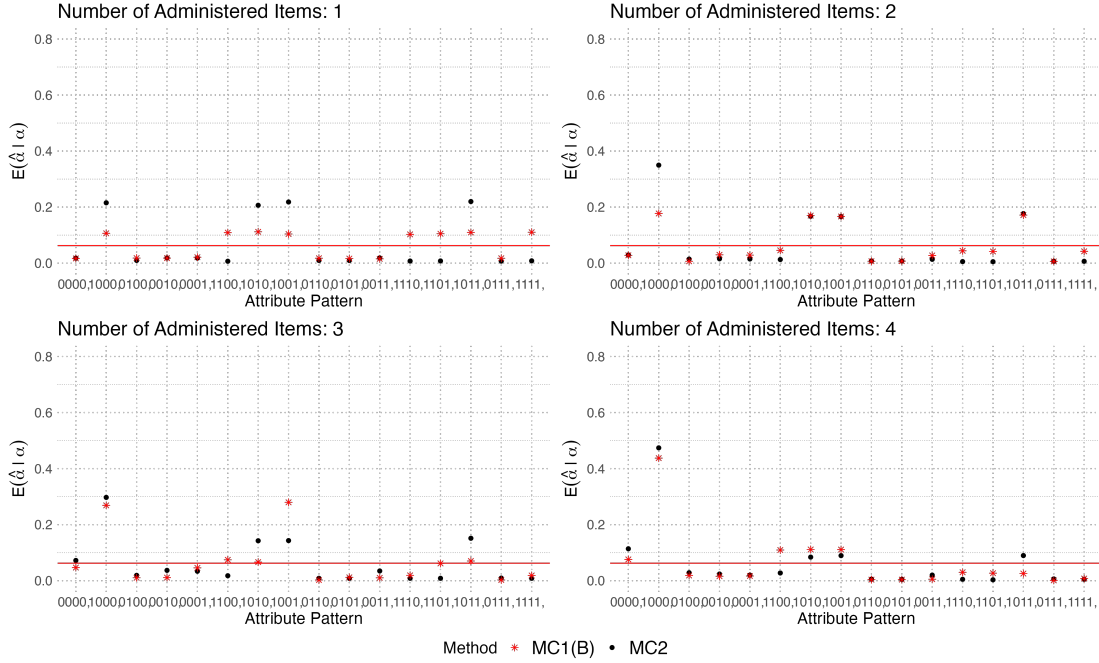
Figure 4.3 shows $E(\hat{\alpha}|\alpha)$, where $\alpha = (0, 0, 0, 0)'$, after each item was administered. Since the observed response might be different from the ideal response, the examinees were not always classified into their true latent class.

The red line represents the random classification level, indicating an equal chance of $\frac{1}{16}$ for examinees to be randomly classified into each latent class. After administering the first item, two distinct trends can be observed from the top-left plot. First, for both tests, the dots representing latent classes mastering A1 dropped below the random level. Second, for the test with $\mathbf{Q}_{MC2}$, the dots representing latent classes mastering A2 also fell below random level. These trends indicate that the first item defined by $\mathbf{e}_1$ in $\mathbf{Q}_{MC1}$ can determine the first entry of $(0, 0, 0, 0)'$, while the first item defined by $\mathbf{Q}_1$ in $\mathbf{Q}_{MC2}$

Figure 4.3: Evolution of $E(\hat{\alpha}|\alpha)$ Across Items for the Latent Class $(0, 0, 0, 0)'$ 

can determine the first two entries of $(0, 0, 0, 0)'$.

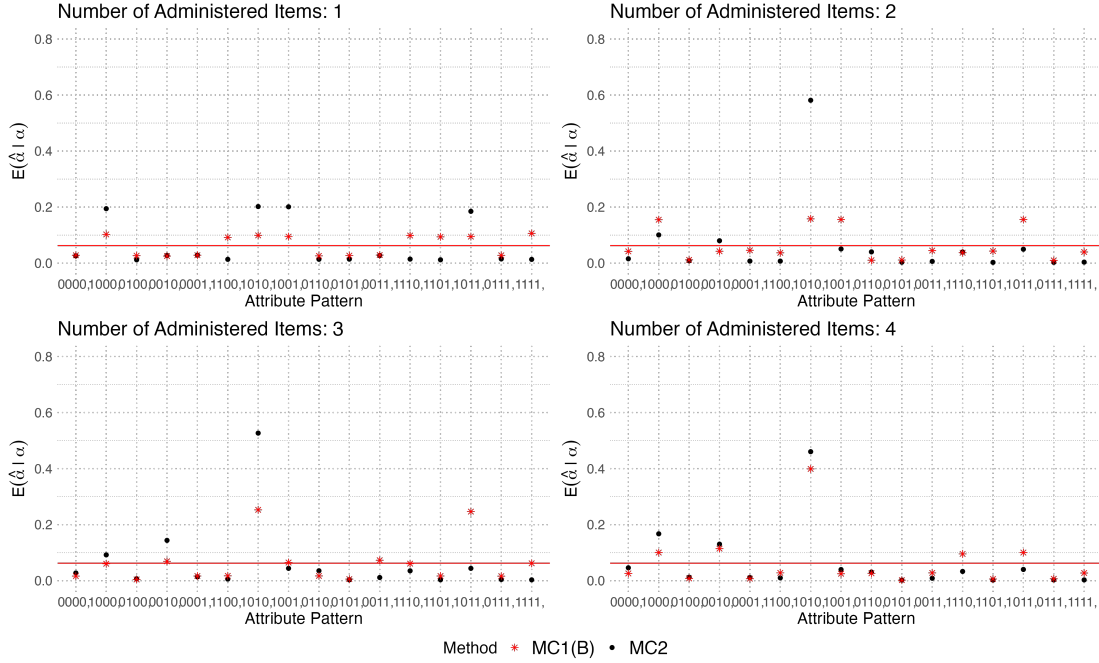
The top-right plot shows that after the first two items in $\mathbf{Q}_{MC2}$ were administered, $E(\hat{\alpha} = \alpha|\alpha)$ was the highest among all other latent classes. In contrast, after administering the first two items in $\mathbf{Q}_{MC1}$, examinees were classified into the latent classes defined by $(0, 0, 0, 0)'$, $(0, 0, 1, 0)'$, $(0, 0, 0, 1)'$, and $(0, 0, 1, 1)'$ with equal likelihood because administering the second item defined by $\mathbf{e}_2$ could only determine the second entry of $(0, 0, 0, 0)'$. Clearly, the true latent class, $\alpha = (0, 0, 0, 0)'$, was distinguishable after two items in $\mathbf{Q}_{MC2}$, while it remained not distinguishable after two items in $\mathbf{Q}_{MC1}$. Similarly, the third and fourth attributes were determined by the third item defined by $\mathbf{e}_3$ and the fourth item defined by $\mathbf{e}_4$, respectively. Although $E(\hat{\alpha} = \alpha|\alpha) > E(\hat{\alpha} = \mathbf{a}|\alpha)$ for $\mathbf{a} \neq \alpha$ is achieved after administering two items in $\mathbf{Q}_{MC2}$ and four items in $\mathbf{Q}_{MC1}$, respectively, the classification accuracy for these two matrices was similar after two items. This is because $P(X_j = \eta_j(\alpha)|\alpha, \mathbf{Q}_{MC2}) = \frac{5-3}{5} < P(X_j = \eta_j(\alpha)|\alpha, \mathbf{Q}_{MC1}) = \frac{5-1}{5}$,

Figure 4.4: Evolution of $E(\hat{\alpha}|\alpha)$ Across Items for the Latent Class $(1, 0, 0, 0)'$ 

where $\eta_j(\alpha) = 0$. The last two items in $\mathbf{Q}_{MC2}$ enhanced the classification accuracy, and these two tests exhibited comparable accuracy rates after these four items.

Figures 4.4, 4.5, and 4.6 display the evolution of $E(\hat{\alpha}|\alpha)$ across items for latent classes defined by $(1, 0, 0, 0)'$, $(1, 0, 1, 0)'$, and $(1, 1, 1, 1)'$, respectively. The overall patterns are similar to Figure 4.3. Items with $\mathbf{Q}_{MC1}$ could determine only one attribute per item, and the true latent class was identified after the completion of four items. However, items with $\mathbf{Q}_{MC2}$ determined two attributes per item, and the true latent class was identified after the completion of two items. In addition, the repetition of the set of two items enhanced the classification accuracy of the latent class defined by $(1, 0, 0, 0)'$, while it slightly impaired the classification accuracy of the latent classes defined by $(1, 0, 1, 0)'$ and $(1, 1, 1, 1)'$.

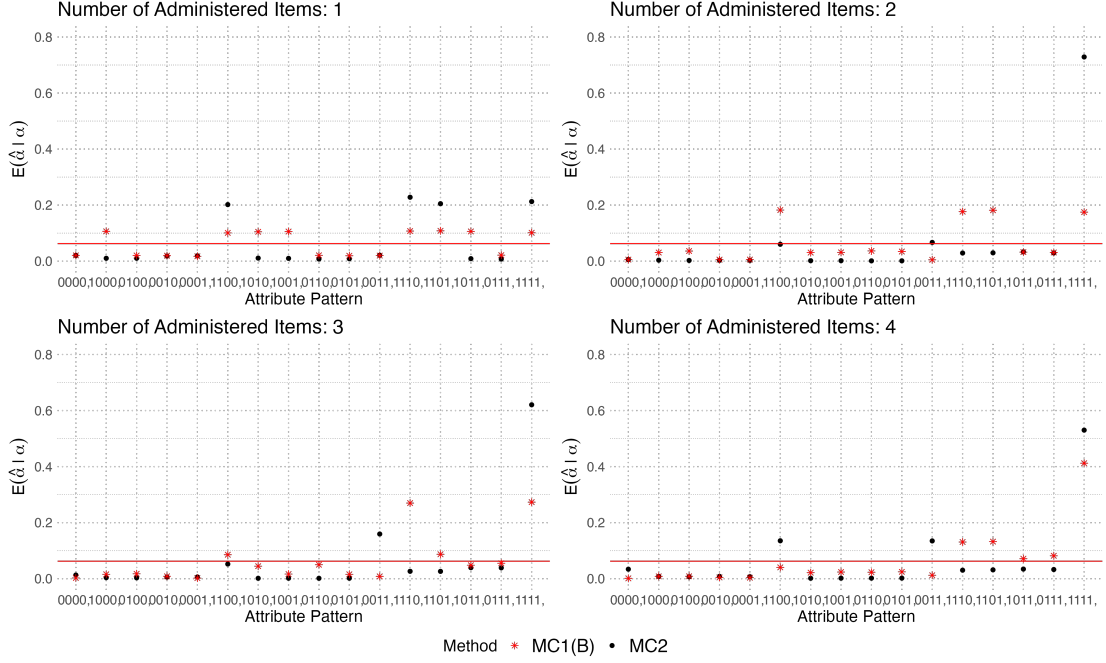
In conclusion, this simulation study examined the change of $E(\hat{\alpha}|\alpha)$ from items of two complete Q-matrices, $\mathbf{Q}_{MC1}$ and $\mathbf{Q}_{MC2}$. Notably, the sign of α became identifiable,

Figure 4.5: Evolution of $E(\hat{\alpha}|\alpha)$ Across Items for the Latent Class $(1, 0, 1, 0)'$ 

where $E(\hat{\alpha} = \alpha|\alpha) > E(\hat{\alpha} = a|\alpha)$ for $\alpha \neq a$, after two items in Q_{MC2} and four items in Q_{MC1} . Both Q-matrices successfully identified all latent classes, supporting the legitimacy of the proposed theorem for the completeness of Q_{MC} . Moreover, the level-2 design implemented in Q_{MC2} yielded higher efficiency in classification, determining two attributes per item, compared to the one-attribute-per-item determination in Q_{MC1} . This design reduces test length, underscoring its importance in Q-optimal design for CD-CAT.

4.6.2 Simulation Study II

This simulation study is designed to evaluate the performance of two CD-CAT algorithms for MC items, the MC-NPS (Wang et al., 2024) and Jensen–Shannon divergence (JSD) methods (Yigit et al., 2019), both with and without the Q-optimal design.

Figure 4.6: Evolution of $E(\hat{\alpha}|\alpha)$ Across Items for the Latent Class $(1, 1, 1, 1)'$ 

4.6.2.1 Design

Wang et al. (2024) demonstrated that a calibration sample of size 100 can effectively calibrate the item parameters for the JSD index, thus the current study also adopts the same calibration sample size ($N_0 = 100$). The attribute profiles of the examinees in the calibration sample were generated using the multivariate normal threshold model (Chiu et al., 2009), with variances set to 1 and covariances sampled from $\mathcal{U}(0.3, 0.5)$, demonstrating moderately dependent attributes. Each item had 4, 5, or 6 options ($H_j = 4, 5$ or 6 across all j), and H_j^* was randomly generated for each j . The item was either in medium-quality or in high-quality. $P(X_{ij} = l | \eta_{ij} = l')$ was the same as Simulation Study I, formulated in Equation 4.6.1.1, but λ_j was randomly sampled from a uniform distribution $\mathcal{U}(0.77, 0.88)$ and $\mathcal{U}(0.88, 1)$ for items of medium and high quality, respectively. An item bank consisting of 300 MC items, each requiring at most

five attributes ($K = 5$), was randomly generated with a fixed proportion of 20% of binary items.

4.6.2.2 Evaluation Criteria

The effectiveness of the compared methods was assessed by calculating the pattern-wise agreement rate (PAR) between the true and estimated attribute profiles, formulated as:

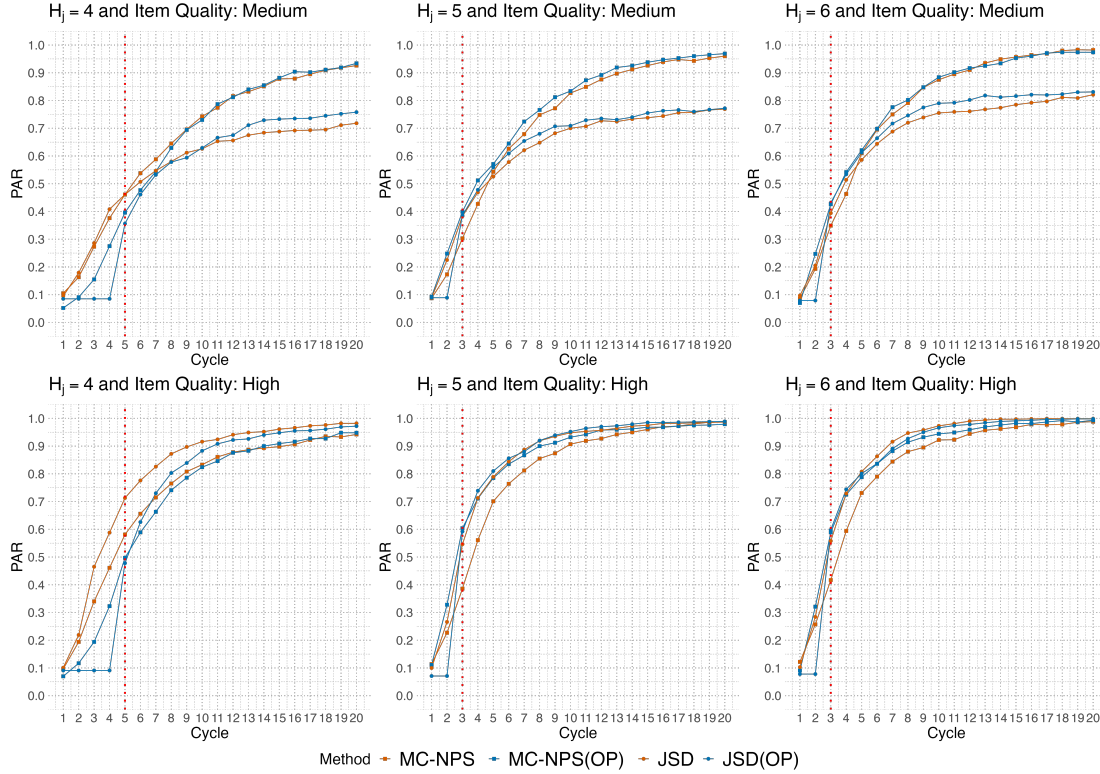
$$\text{PAR} = \frac{\sum_{i=1}^N I[\hat{\alpha}_i = \alpha_i]}{N}.$$

PAR was calculated across each cycle and condition, using data from the 1,000 examinees ($N = 1000$).

4.6.2.3 Results

Figure 4.7 presents the PARs of the MC-NPS and JSD methods with and without the newly proposed MC Q-optimal method. As detailed earlier, applying the MC Q-optimal procedure during the initial stage of the CD-CAT, requires $\lceil \frac{K}{2} \rceil = 3$ items for instances with $H_j = 6$ and most instances with $H_j = 5$ to achieve optimal classification, while $K = 5$ items are required for instances with $H_j = 4$, according to Corollary 3. For the cases without the MC Q-optimal procedure, the initial item was chosen randomly, and the subsequent items were selected by maximizing the discrimination power for the MC-NPS method or maximizing the JSD index for the JSD method. To facilitate the comparisons, two reference lines at fifth cycle (for tests with $H_j = 4$) and third cycle (for tests with $H_j = 5$ and 6) were provided to mark the cycles at which the optimal classifications produced by the MC Q-optimal procedures were reached. Different methods and the inclusion/exclusion of the Q-optimal procedure are represented by different colors and line types. The MC-NPS method is marked with square points, and the JSD index with circle points. Blue and orange lines indicate methods with and without the MC Q-optimal procedure, respectively. In addition, the

Figure 4.7: PARs of the MC-NPS and JSD Methods



MC-NPS and JSD methods with the MC Q-optimal procedure are labeled as MC-NPS (OP) and JSD (OP), respectively.

The upper part of the figure illustrates the results when items were of medium quality. For $H_j = 4$, both the MC-NPS (OP) and JSD (OP) methods initially performed worse than these methods without the Q-optimal procedure. However, as the test length increased, the MC-NPS (OP) method caught up with its non-Q-optimal counterpart. The JSD (OP) index eventually outperformed the version without the Q-optimal design. For $H_j = 5$ or 6, the MC-NPS (OP) method consistently outperformed the MC-NPS method at each cycle, although its advantage diminished as the test length increased. After 15 items were administered, the PARs of the MC-NPS (OP) and MC-NPS methods were nearly identical. In contrast, the JSD (OP) index

did not consistently outperform the JSD index before reaching the reference line, with PARs below 0.1 prior to the third cycle. Nevertheless, it slightly surpassed the JSD index at the third cycle. As the test length increased, the PARs for the JSD (OP) and JSD indices were about the same. These comparisons suggest that the newly proposed MC Q-optimal procedure improves classification accuracy of the MC-NPS method before $\lceil \frac{K}{2} \rceil$ items, particularly when $H_j \geq 5$, aligning with previous findings that the MC Q-optimal procedure enhances classification accuracy upon the completion of this procedure (Wang et al., 2024). When comparing the MC-NPS method to the JSD index, regardless of the Q-optimal design, the MC-NPS method consistently outperformed the JSD index across all conditions.

The lower panel of the figure illustrates the results for high-quality items. The results show that high item quality indeed improved classification accuracy across all five compared algorithms, particularly for the JSD index, with and without the Q-optimal procedure. The conclusions remain consistent: the MC Q-optimal procedure enhances classification accuracy of the MC-NPS method before $\lceil \frac{K}{2} \rceil$ items, especially with $H_j \geq 5$.

4.7 Discussion

In CD for MC items, several parametric and nonparametric classification methods (e.g., the MC-DINA model; de la Torre., 2009; the MC-NPC method; Wang et al., 2023) have been developed to analyze item responses to assessments involving MC items. These classification methods lead to the recent development of item selection methods for CD-CAT (e.g., the JSD index; Yigit, et al., 2019; the MC-NPS method; Wang et al., 2024). Although these methods exclusively tailored for MC items have demonstrated significant advantages over methods that can only be used by dichotomizing MC items, some fundamental issues related to test development remain untouched, imposing a risk of creating invalid test items for the currently available CD methods for MC items.

This study focuses on two key issues, *the completeness of the MC Q-matrix* and *the MC Q-optimal design for CD-CAT*, given the MC-DINA model, and establishes two primary theorems and several lemmas and corollaries, supported by rigorous mathematical proofs.

The sufficient and necessary conditions for a complete $\mathbf{Q}_{\text{MC}}$ are outlined in Theorem 1. It states that a complete $\mathbf{Q}_{\text{MC}}$ must include all possible $\mathbf{e}_k$. However, two additional conditions are introduced to ensure that $\mathbf{e}_k$ is contained by a $\mathbf{Q}_j$ with a level- x design or a level-0 design. Corollary 1 is further provided to make a connection between a complete $\mathbf{Q}_{\text{B}}$ to a complete $\mathbf{Q}_{\text{MC}}$, thereby facilitating the application of Theorem 1 by offering a more practical framework.

Lemmas 2 and 3 provide the sufficient and necessary conditions for how $\mathbf{0}_K$ and the other latent classes can be identified with MC items under the MC-DINA model and serve as foundation for Theorem 2. In addition, as a preparation of Theorem 2, the minimum test length for a complete $\mathbf{Q}_{\text{MC}}$ was determined. Based on Table 4.3, when $3 \leq H_j \leq 7$, the minimum test length of $\lceil \frac{K}{2} \rceil$ can be achieved if each $\mathbf{Q}_j$ of the $\lfloor \frac{K}{2} \rfloor$ items have level-2 designs. Therefore, the level-2 design is a foundation for Q-optimality. However, the Q-optimal design does not strictly adopt the level-2 design, as it does not require lower diagonal entries to be fixed as zero in the CD-CAT context. Hence, Theorem 2 was proposed to address the sufficient and necessary conditions for the Q-optimal design for CD-CAT under the MC-DINA model. It formalizes the aforementioned design and targets MC items with $6 \leq H_j < 14$. The range of H_j is different from that used above for the level-2 design because to achieve the Q-optimal design, $\frac{H_j - H_j^*}{H_j} \geq 0.5$ is adopted to ensure that $P(X_j = \eta_j(\boldsymbol{\alpha}) | \boldsymbol{\alpha}, \mathbf{Q}_{\text{MC}}) > P(X_j = \eta_j(\boldsymbol{\alpha}) | \boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}})$, where $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^*$, resulting in $6 \leq H_j < 14$. The Q-optimal design for MC items with another two common H_j in MC items, $H_j = 4$ and $H_j = 5$, are discussed in Corollaries 3 and 4, respectively.

The contribution of the proposed theoretical work was further supported by two simulation studies. Simulation Study I confirmed the completeness of $\mathbf{Q}_{\text{MC}}$ as proposed

in Theorem 1 by comparing a complete $\mathbf{Q}_{\text{MC}}$ constructed using a level-2 design to one augmented from a complete $\mathbf{Q}_{\text{B}}$. The results showed that $\mathbf{Q}_{\text{MC}}$ constructed based on Theorem 1 could successfully identify all latent classes, and $\mathbf{Q}_{\text{j}}$ with the level-2 design decreases the test length to identify all latent classes. Simulation Study II was conducted to evaluate two existing item selection methods for CD-CAT (the MC-NPS method and the JSD index), with and without the proposed MC Q-optimal procedure. For $H_j \geq 5$, the newly proposed Q-optimal design enhanced the classification accuracy of the MC-NPS method, particularly during the first $\lceil \frac{K}{2} \rceil$ items. However, when $H_j = 4$, the MC-NPS method with Q-optimal design underperformed compared to it without any Q-optimal design. In addition, the Q-optimal design had a greater impact on the MC-NPS method than on the JSD index.

Taken together, this study provides both theoretical and practical contributions to the development of a complete $\mathbf{Q}_{\text{MC}}$ and a Q-optimal design for CD-CAT under the MC-DINA model. Beyond theoretical advancements, the proposed theory also offers practical guidance for the development and construction of an MC assessment. For example, Theorem 1 can help with item bank design and construction in either a traditional paper-pencil test or CD-CAT settings. Theorem 2 provides a sophisticated algorithm to optimally select initial items for CD-CAT. Future research could investigate whether the proposed theorems hold for extensions of the MC-DINA model, such as the structured MC-DINA model (Ozaki, 2025). Additionally, further research on the parameter identifiability of the MC-DINA model could build upon the theoretical outcomes presented in this paper.

Chapter 5

Conclusion and Discussion

CD is a psychometric methodology designed to provide detailed insights into examinees' strengths and weaknesses within a specific knowledge or skill area. This approach offers students and education practitioners with more information about the examinees' current knowledge state, which can then be used to develop more effective interventions and support strategies. Within the CD framework, CDMs are usually used to analyze examinees' responses. However, many well-studied CDMs are compatible only with binary data. Consequently, for test formats such as MC items, the data must be dichotomized to fit these CDMs. This process can lead to the underutilization of information from distractors, as all responses to distractors are grouped together as the incorrect answer without further analysis.

In response to this issue, the MC-DINA model (de la Torre, 2009) was the first genuine method proposed for analyzing MC items in CD. It has been demonstrated that the examinees' attribute profiles are more accurately estimated, compared to estimates from CDMs for dichotomized data. However, the attribute profile estimation with the MC-DINA model relies on an expectation-maximization algorithm, which requires a large sample size to achieve accurate estimation. Therefore, the MC-DINA model is not feasible in micro-educational settings. To overcome the difficulty, the MC-NPC method

(Wang et al., 2023), which uses the penalized Hamming distance as the loss function to estimate the attribute profiles, was proposed in Chapter 2 as a counterpart of the MC-DINA model. The theoretical justification was provided to support the legitimacy of using the MC-NPC method to analyze the MC-DINA data. In addition, the simulation studies show that the MC-NPC method can improve the examinees' classifications in comparison with the DINA model and the NPC method applied to dichotomized data. The results further confirm that MC-NPC method outperforms the MC-DINA model when samples are small.

To improve measurement accuracy and efficiency, CD-CAT has been designed to adaptively select the next item based on an examinee's previous responses. While various item selection methods exist for binary responses (e.g., Chang et al., 2019; Chiu & Chang, 2021; Kaplan et al., 2015), the only approaches that incorporate distractor information for MC items is the JSD index (Yigit et al., 2019) and the MC-NPS method (Wang et al., 2024). However, the JSD index relies on the MC-DINA model for item parameter calibration, which presents challenges in small-sample settings. Developed in Chapter 3 is the MC-NPS method (Wang et al., 2024), which implements a novel discrimination power that measures an item's ability to effectively distinguish among different attribute profiles. Another newly proposed procedure in Chapter 3, Q-optimal procedure for MC items, is also empirically applied to select items during the initial phase of a CD-CAT algorithm to produce optimal classifications. The MC-NPS method is remarkable in that it does not require item calibration, making it applicable when the calibration sample is small or unavailable. The simulation studies demonstrate that incorporating the information underlying the coded distractors can improve the estimation accuracy and increase the efficiency of item selection. More importantly, the results show that the MC-NPS method outperforms the JSD index when items are of medium quality across all calibration samples, and have comparable performance to the JSD index when items are of high quality.

Like other CD methods, the Q-matrix plays a key role for the MC-NPC and MC-NPS methods. However, the requirements for the MC Q-matrix to be valid and useful are mandatory but yet remain unexplored. Chapter 4 aims to provide users with theoretically sound and practically valuable guidelines for developing feasible MC tests if they would like to analyze their test results with the MC-DINA model or the MC-NPC method. Specifically, the first focus of the chapter is on the development of the completeness condition of the MC Q-matrix. A Q-matrix is said to be complete if it can identify all possible attribute profiles. Theorem 1 in Chapter 4 provides the sufficient and necessary conditions for a complete Q-matrix under the MC-DINA model. The second focus of the chapter is the development of a closely related topic, the Q-optimal procedure for CD-CAT under the MC-DINA model. Although the Q-optimal procedure for MC items was applied to the MC-NPS method in Chapter 3, its sufficient and necessary conditions are formally established in Theorem 2. Both theorems are supported by rigorous mathematical proofs and well-designed simulation studies.

To sum up, this dissertation comprises three studies on CD methods for MC items, including the development of new methods particularly suited for small educational settings and the critical Q-matrix theory. However, certain limitations that warrant acknowledgment. First, the nonparametric classification methods, including the proposed MC-NPC and MC-NPS methods, offer the advantage of being “insensitive” to the variations of sample sizes, making them especially practical for smaller datasets. However, it should be recognized that, unlike parametric approaches, they do not provide item parameter estimates, as they do not rely on fitting a statistical model. Although estimating examinees’ attribute patterns is the ultimate goal of CD research, if the focus of analysis shifts to item parameter estimation, nonparametric methods would not be appropriate. Additionally, when sample sizes are large, parametric methods are generally recommended, as they tend to perform slightly better under such conditions than nonparametric approaches.

Second, while the studies show that incorporating diagnostic information from additional distractors promise more accurate attribute estimation, this has only been evaluated with a correctly specified Q-matrix. Hence, a critical question is: to what extent do the misspecifications of the Q-matrix jeopardize the classification accuracy, such that the effort of including distractors to the analysis is no longer practically meaningful. A preliminary simulation study, although not thorough, suggests that the MC-NPC method, in comparison to the MC-DINA model, demonstrates “insensitivity” to the Q-matrix misspecifications. Despite these misspecifications, the MC-NPC method remains effective in classifying examinees, achieving higher classification accuracy than the studied dichotomous CD methods. These findings underscore the need for further systematic investigation to fully understand the robustness of the MC-NPC method. Additionally, ensuring the correct specification of the Q-matrix is vital to uphold the validity of any CD analysis, particularly with MC items where additional q-vectors need to be specified for coded distractors. Currently, no refinement or validation methods specifically designed for correcting Q-matrices in MC items exist, indicating a mandatory direction for future research.

Third, another concern for the effectiveness of the proposed method is the presence of aberrant behaviors during assessment, as certain test-taking behaviors can distort CDM estimation and compromise the validity of assessments. For instance, rapid guessing, which occurs when students spend too little time on an item to read and comprehend its content before attempting an answer (Bulut et al., 2023). This behavior violates a central assumption of measurement models that examinees put in sufficient effort to reflect their true ability or skill levels. When rapid guessing occurs, estimates from a CD method may become less accurate, often underestimating students’ true proficiency levels (Hsu et al., 2020). More importantly, rapid guessing deserves particular attention in the context of MC items because it occurs more frequently than, say, in constructed-response items. Although extensive research has addressed rapid guessing in the context of item response theory (IRT) models (e.g., Wang & Xu, 2015; Wang et al., 2018),

this topic remains relatively underexplored in CDMs, with only one study to date (Hsu et al., 2020) incorporating rapid guessing into a CDM framework. Future studies could benefit from exploring the integration of rapid guessing adjustments in CD methods for MC items. For the present, however, a practical solution is to utilize response times – widely used in other studies (e.g., Wise, 2019) – to identify rapid guessing behavior. Response time data, readily available in computer-based assessments, can serve as a preliminary check before fitting data to the MC-NPC method or the MC-DINA model.

Building on this research, several additional directions for future study emerge. First, model-data misfit raises a realistic risk and limitation of the MC-NPC and MC-NPS methods. To date, efficient procedures for detecting model misfit are still underdeveloped, largely because most CDMs for MC items are variations of the DINA model and such misfit indices do not seem to be mandatory yet. Thus, one critical area for future research is to develop a general model for handling even highly complex data that can mitigate model misfit issues. Second, it would be valuable to investigate whether the proposed theorems for completeness and Q-optimality extend to variations of the MC-DINA model, such as the structured MC-DINA model (Ozaki, 2025). Last but not the least, the theoretical findings in this dissertation lay a foundation for exploring the parameter identifiability of the MC-DINA model, an area ripe for further theoretical development.

References

- Bulut, O., Gorgun, G., Wongvorachan, T., & Tan, B. (2023). Rapid guessing in low-stakes assessments: Finding the optimal response time threshold with random search and genetic algorithm. *Algorithms*, 16(2), 89.
- Chang, Y.-P., Chiu, C.-Y., & Tsai, R.-C. (2019). Nonparametric CAT for CD in educational settings with small samples. *Applied Psychological Measurement*, 43(7), 543-561.
- Chen, J., & de la Torre, J. (2018). Introducing the General Polytomous Diagnosis Modeling Framework. *Frontiers in Psychology*, 9, 1474.
- Chiu, C.-Y. (2013). Statistical refinement of the Q-matrix in cognitive diagnosis. *Applied Psychological Measurement*, 37(8), 598-618.
- Chiu, C.-Y., & Chang, Y.-P. (2021). Advances in CD-CAT: The general nonparametric item selection method. *Psychometrika*, 86(4), 1039-1057.
- Chiu, C.-Y., & Douglas, J. A. (2013). A nonparametric approach to cognitive diagnosis by proximity to ideal response profiles. *Journal of Classification*, 30(2), 225-250.
- Chiu, C.-Y., Douglas, J. A., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika*, 74(4), 633-665.

- Chiu, C.-Y., & Köhn, H. F. (2019). Consistency theory for the General Nonparametric Classification Method. *Psychometrika*, 84(3), 830–845.
- Chiu, C.-Y., Sun, Y., & Bian, Y. (2018). Cognitive diagnosis for small educational programs: The general nonparametric classification method. *Psychometrika*, 83(2), 355–375.
- Chiu, C.-Y., Köhn, H.-F., & Wang, Y. (under review). Useful and Proper Multiple-Choice Items for Diagnostic Classification.
- de la Torre, J. (2009). A cognitive diagnosis model for cognitively based multiple-choice options. *Applied Psychological Measurement*, 33(3), 163–183.
- DiBello, L. V., Henson, R. A., & Stout, W. F. (2015). A family of generalized diagnostic classification models for multiple choice option-based scoring. *Applied Psychological Measurement*, 39(1), 62–79.
- DiBello, L. V., Roussos, L. A., & Stout, W. F. (2007). Review of cognitively diagnostic assessment and summary of psychometric models. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Vol. 26, Psychometrics* (pp. 979–1030). Elsevier.
- Gu, Y. & Xu, G. (2018). The sufficient and necessary condition for the identifiability and estimability of the DINA model. *Psychometrika*, 84(2), 468–483.
- Haberman, S. J., & von Davier, M. (2007). Some notes on models for cognitively based skill diagnosis. In C. R. Rao & S. Sinharay (Eds.), *Handbook of statistics: Vol. 26, Psychometrics* (pp. 1031–1038). Elsevier.
- Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, 26(4), 333–352.
- Haertel, E. H., & Wiley, D. E. (1993). Representations of ability structures: Implications for testing. In N. Frederiksen, R. J. Mislevy, & I. Bejar (Eds.), *Test theory for a new generation of tests* (pp. 359–384). Hillsdale, NJ: Erlbaum.

- Hsu, C.-L., Jin, K.-Y., & Chiu, M. M. (2020). Cognitive diagnostic models for random guessing behaviors. *Frontiers in Psychology, 11*.
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement, 25*(3), 258-272.
- Kaplan, M., de La Torre, J., & Barrada, J. R. (2015). New Item Selection Methods for Cognitive Diagnosis Computerized Adaptive Testing. *Applied Psychological Measurement, 39*(3), 167–188.
- Köhn, H.-F., & Chiu, C.-Y. (2016a). Conditions of completeness of the Q-matrix of tests for cognitive diagnosis. In L. A. van der Ark, D. M. Bolt, W. C. Wang, J. A. Douglas, & M. Wiberg (Eds.), *Quantitative Psychology Research: The 80th Annual Meeting of the Psychometric Society* (pp. 255-264). New York: Springer.
- Köhn, H. F., & Chiu, C.-Y. (2016b). A proof of the duality of the DINA model and the DINO model. *Journal of Classification, 33*(2), 171-184.
- Köhn, H.-F., & Chiu, C.-Y. (2017). A procedure for assessing completeness of the Q-matrices of cognitively diagnostic tests. *Psychometrika, 82*(1), 112-132.
- Köhn, H.-F., & Chiu, C.-Y. (2019). Attribute hierarchy models in cognitive diagnosis: Identifiability of the latent attribute space and conditions for completeness of the Q-matrix. *Journal of Classification, 36*(3), 541-565.
- Köhn, H.-F., & Chiu, C.-Y. (2021). A unified theory of the completeness of Q-matrices for the DINA model. *Journal of Classification, 38*(3), 500–518.
- Köhn, H. F., Chiu, C.-Y. & Wang, Y. (2023). Proper and useful distractors in multiple-choice diagnostic classification models. In M. Wiberg, D. Molenaar, J. González, J.-S. Kim, & H. Hwang (Eds.), *Quantitative Psychology: The 87th Annual Meeting of the Psychometric Society* (pp. 97-106). New York: Springer.

- Lin, M., & Xu, G. (2024). Sufficient and Necessary Conditions for the Identifiability of DINA Models with Polytomous Responses. *Psychometrika* 89, 717–740.
- Leighton, J., & Gierl, M. (Eds.). (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge University Press.
- Leighton, J. P., Gierl, M. J., & Hunka, S. (2004). The attribute hierarchy method for cognitive assessment: A variation on Tatsuoaka's rule-space approach. *Journal of Educational Measurement*, 41(3), 205–236.
- Ma, W., & de la Torre, J. (2016). A sequential cognitive diagnosis model for polytomous responses. *British Journal of Mathematical and Statistical Psychology*, 69, 253–275.
- Macready, G. B., & Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics*, 2(2), 99–120.
- Nichols, P. D., Chipman, S. F., & Brennan, R. L. (Eds.). (1995). *Cognitively diagnostic assessment*. Lawrence Erlbaum.
- Nitko, A. J. (2001). *Educational assessment of students* (3rd ed.). Columbus, OH: Merrill Prentice Hall.
- Ozaki, K. (2015). DINA models for multiple-choice items with few parameters: Considering incorrect answers. *Applied Psychological Measurement*, 39(6), 431–447.
- Paulsen, J. (2019). *Examining cognitive diagnostic modeling in small sample contexts*. (Publication No. 22583956) [Doctoral Dissertation, University of Indiana, Bloomington, IN]. ProQuest Dissertations & Theses Global.
- Rupp, A. A., Templin, J. L., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford Press.

- Sadler, P. M. (1998). Psychometric models of student conceptions in science: Reconciling qualitative studies and distractor-driven assessment instruments. *Journal of Research in Science Teaching*, 35(3), 265-296.
- Sessoms, J., & Henson, R. A. (2018). Applications of diagnostic classification models: A literature review and critical commentary. *Measurement: Interdisciplinary Research and Perspectives*, 16(1), 1-17.
- Tatsuoka, C. (2002). Data-analytic methods for latent partially ordered classification models. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 51(3), 337-350.
- Tatsuoka, K. K. (1985). A probabilistic model for diagnosing misconception in the pattern classification approach. *Journal of Educational and Behavioral Statistics*, 10(1), 55-73.
- Tatsuoka, K. K. (1990). Toward an integration of item-response theory and cognitive error diagnosis. In N. Frederiksen, R. Glaser, A. Lesgold, & M. Safto (Eds.), *Monitoring skills and knowledge acquisition* (pp. 453-488). Erlbaum.
- Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction to the rule space method*. Routledge/Taylor & Francis Group.
- Wang, C., & Xu, G. (2015). A mixture hierarchical model for response times and response accuracy. *British Journal of Mathematical and Statistical Psychology*, 68(3), 456-477.
- Wang, C., Xu, G., Shang, Z., & Kuncel, N. (2018). Detecting Aberrant Behavior and Item Preknowledge: A Comparison of Mixture Modeling Method and Residual Method. *Journal of Educational and Behavioral Statistics*, 43(4), 469-501.
- Wang, S., & Douglas, J. (2015). Consistency of nonparametric classification in cognitive diagnosis. *Psychometrika*, 80(1), 85-100.

- Wang, Y., Chiu, C.-Y., & Köhn, H. F. (2024). Nonparametric CD-CAT for multiple-choice items: Item selection method and Q-optimality. *British Journal of Mathematical and Statistical Psychology*(00), 1-23.
- Wang, Y., Chiu, C.-Y., & Köhn, H. F. (2023). Nonparametric classification method for multiple-choice items in cognitive diagnosis. *Journal of Educational and Behavioral Statistics*, 48(2), 189–219.
- Wise, S. L. (2019). An Information-Based Approach to Identifying Rapid-Guessing Thresholds. *Applied Measurement in Education*, 32(4), 325–336.
- Wood, E. J. (2003). What are extended matching sets questions? *Bioscience Education*, 1(1), 1–8.
- Xu, G., Wang, C., & Shang, Z. (2016). On initial item selection in cognitive diagnostic computerized adaptive testing. *British Journal of Mathematical and Statistical Psychology*, 69(3), 291-315.
- Yigit, H. D., Sorrel, M. A., & de la Torre, J. (2018). Computerized adaptive testing for cognitively based multiple-choice data. *Applied Psychological Measurement*, 43(5), 388–401.

Appendix A

Figure A.1: PARs from the MC-NPS, JSD, GNPS, and Random Selection Methods with High-Quality Items When $H_j = 4$

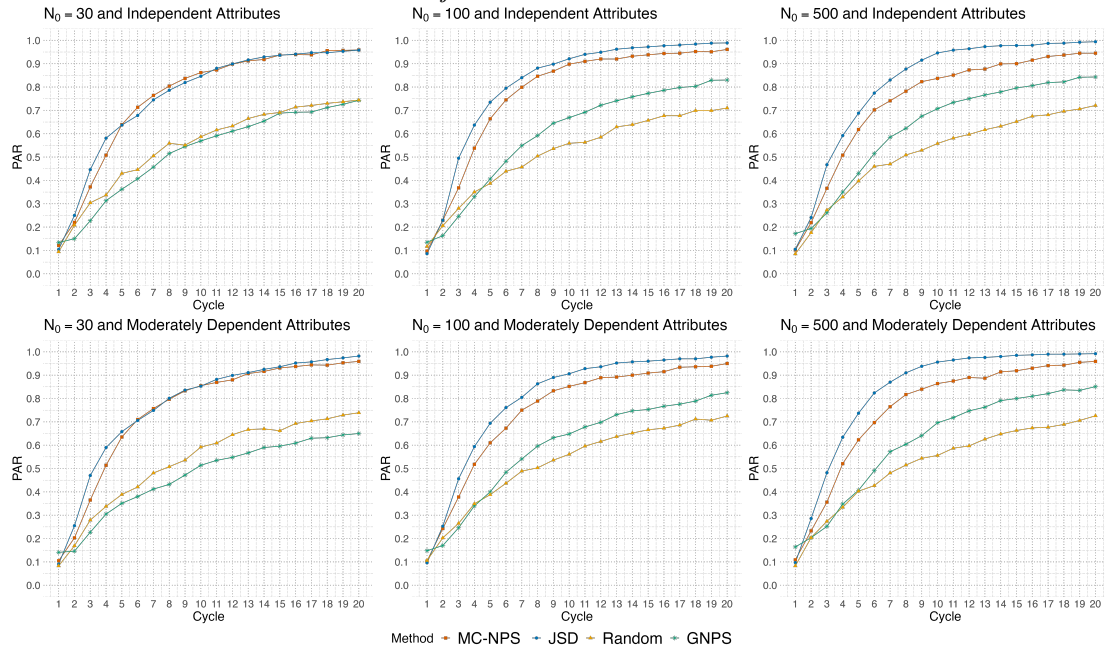
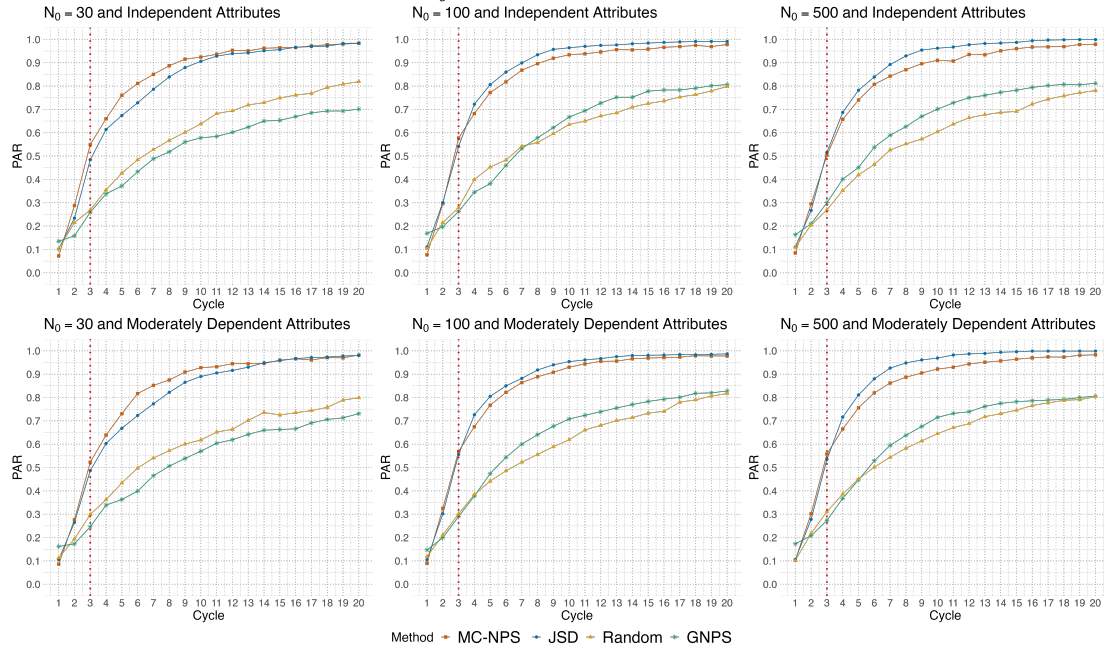


Figure A.2: PARs from the MC-NPS, JSD, GNPS, and Random Selection Methods with High-Quality Items When $H_j = 5$



Appendix B

Proof of Lemma 1

Lemma 1. *Suppose $\mathbf{Q}_{j_0}$ of item j_0 involving $\mathbf{e}_k$ conforms to the level-0 design; that is, $\mathbf{q}_{j_0}^{(H_{j_0}^*)} \neq \mathbf{e}_k$. Consider the following algorithm:*

Step 0 *Let $\mathbf{q} = \mathbf{q}_{j_0}^{(H_{j_0}^*)}$ and initialize set $\mathcal{J} = \{j_0\}$.*

Step 1 *For the rest of test items, define set $\mathcal{J}_k$ as $\mathcal{J}_k = \{j | \mathbf{Q}_j \text{ contains } \mathbf{e}_k + \mathbf{q}_0, \text{ where } \mathbf{q}_0 \preceq \mathbf{q}\}$. If $\mathcal{J}_k = \emptyset$, stop; otherwise, update $\mathcal{J} = \mathcal{J} \cup \mathcal{J}_k$ and proceed to Step 2.*

Step 2 *If there exists a $j \in \mathcal{J}_k$ such that $\mathbf{q}_j^{(H_j^*)} \succ \mathbf{e}_k$, stop; otherwise, update $\mathbf{q} = \mathbf{q} \vee (\bigvee_{t=1}^{|\mathcal{J}_k|} \mathbf{q}_{j_t}^{(H_{j_t}^*)})$ and repeat Steps 1 and 2.*

Examinees who have and have not mastered attribute k can be separated apart by administering all the items in set $\mathcal{J}$ if and only if the above algorithm stops with $\mathcal{J}_k \neq \emptyset$.

Proof. ($\Leftarrow$) Suppose that the algorithm stops with $\mathcal{J}_k \neq \emptyset$ and there must exist an item $j \in \mathcal{J}_k$ such that $\mathbf{q}_j^{(H_j^*)} \succ \mathbf{e}_k$.

Suppose item j_0 conforms to the level-0 design with respect to attribute k . Then, there exists a coded distractor l defined by $\mathbf{e}_k$ and $\mathbf{q}_{j_0}^{(H_{j_0}^*)} \neq \mathbf{e}_k$. According to Table 4.1, item j_0 classifies examinees into G_1 (i.e., not mastered attribute k), G_2 (i.e., mastered attribute k), and G_3 (i.e., mixed with examinees who may or may not have mastered attribute k , but mastered attributes required by $\mathbf{q}_{j_0}^{(H_{j_0}^*)}$). Let $\mathbf{q}$ denote the attributes

that have been mastered by the mixed latent class. In this iteration, $\mathbf{q} = \mathbf{q}_{j_0}^{(H_{j_0}^*)}$. The two latent classes, $\alpha \succeq \mathbf{e}_k + \mathbf{q}$ and $\alpha^* \succeq \mathbf{q}$ but $\alpha^* \not\succeq \mathbf{e}_k$, cannot be completely distinguished. Define $\mathcal{J}$ as the set that collects items that can differentiate examinees based on attribute k . In this step, $\mathcal{J} = \{j_0\}$.

Proceed to Step 1 and set $\mathcal{J}_k = \{j | \mathbf{Q}_j \text{ contains } \mathbf{e}_k + \mathbf{q}_0, \text{ where } \mathbf{q}_0 \preceq \mathbf{q}\}$, and then update $\mathcal{J} = \mathcal{J} \cup \mathcal{J}_k$. Every item in $\mathcal{J}_k$ can further classify α into G_2 or G_3 (i.e., mastered attributes required by $\mathbf{q}_j^{(H_j^*)}$). If there exists an item in $\mathcal{J}_k$ such that $\mathbf{q}_j^{(H_j^*)} \succ \mathbf{e}_k$, α^* will be classified into G_1 by this item. As a result, administering all items in $\mathcal{J}$ can differentiate examinees who have mastered attribute k from those who have not. However, if there does not exist such an item, α^* will be divided into G_1 or G_3 (i.e., mastered attributes required by $\mathbf{q}_j^{(H_j^*)}$). In this case, all latent classes mastering the attributes required by $\mathbf{q}_j^{(H_j^*)}$, where $j \in \mathcal{J}_k$, are classified into G_3 . The attributes that have been mastered by examinees in the mixed latent class is $\mathbf{q} \vee (\bigvee_{t=1}^{|\mathcal{J}_k|} \mathbf{q}_{j_t}^{(H_{j_t}^*)})$. Hence, update $\mathbf{q} = \mathbf{q} \vee (\bigvee_{t=1}^{|\mathcal{J}_k|} \mathbf{q}_{j_t}^{(H_{j_t}^*)})$. Similar to the above discussion, all items in $\mathcal{J}$ can differentiate examinees based on attribute k only when they have not mastered the attributes required by $\mathbf{q}$, but fails to distinguish between α and α^* .

Repeat Steps 1 and 2 until an item j that satisfies $\mathbf{q}_j^{(H_j^*)} \succ \mathbf{e}_k$ is found in $\mathcal{J}_k$. Assume that the termination occurs in the t th iteration. Then, up to the $(t-1)$ th iteration, all items in $\mathcal{J}$ can only differentiate examinees who have mastered attribute k from those who have not when the examinees have not mastered attributes required by $\mathbf{q}$, but fail to make the differentiation when the examinees have mastered attributes required by $\mathbf{q}$. In the t th iteration, item j satisfying $\mathbf{q}_j^{(H_j^*)} \succ \mathbf{e}_k$ can differentiate examinees who have mastered attribute k from those who have not when examinees have mastered attributes required by $\mathbf{q}$. Specifically, item j classifies α into G_2 and G_3 , and α^* into G_1 . Update $\mathcal{J} = \mathcal{J} \cup \mathcal{J}_k$. Consequently, after all the items in $\mathcal{J}$ are administered, examinees who have mastered attribute k can be differentiated from those who have not.

($\Rightarrow$) Suppose that the algorithm terminates at the t th iteration with $\mathcal{J}_k = \emptyset$. Further suppose all latent classes can be distinguished by the items in $\mathcal{J}$ based on whether

attribute k has been mastered. Now consider $\alpha = \mathbf{1}_K$ and $\alpha^* = \mathbf{1}_K - e_k$. After the $(t - 1)$ th iteration, since no items in $\mathcal{J}$ satisfy $\mathbf{q}_j^{(H_j^*)} \succ e_k$, every item in $\mathcal{J}$ classify both α and α^* to G_3 , implying that no items in $\mathcal{J}$ can differentiate between α and α^* . After the t th iteration, the algorithm terminates without adding any additional items to $\mathcal{J}$ because $\mathcal{J}_k = \emptyset$. Hence, α and α^* remain indistinguishable, which contradicts the assumption that items in $\mathcal{J}$ can distinguish the all latent classes based on attribute k . Therefore, $\mathcal{J}_k \neq \emptyset$ is necessary for successfully distinguishing these latent classes. $\square$

Proof of Theorem 1

Theorem 1. Suppose $\mathbf{Q}_{MC}$ is the Q -matrix of items conforming to the MC-DINA model. $\mathbf{Q}_{MC}$ is complete if and only if $\mathbf{Q}_{MC}$ includes all e_k , where $k = 1, \dots, K$, and for each e_k :

A1. At least one item j containing it has a level- x design.

A2. All items containing it have the level-0 design, and Lemma 1 is satisfied.

Proof. ($\Leftarrow$) Based on Definition 2, if $\varsigma(\alpha) \neq \varsigma(\alpha^*)$ for any $\alpha \neq \alpha^*$, then the proposed Q -matrix is complete for MC items. Recall that $\varsigma(\alpha) = (\varsigma_1(\alpha), \dots, \varsigma_J(\alpha))' = (E(\mathbf{I}_1|\alpha), \dots, E(\mathbf{I}_J|\alpha))'$, where $E(\mathbf{I}_j|\alpha) = (E(I_{j1}|\alpha), \dots, E(I_{jH_j^*}|\alpha))'$. To demonstrate that $\varsigma(\alpha) \neq \varsigma(\alpha^*)$ for any $\alpha \neq \alpha^*$, it suffices to show that there exists at least one j such that $E(\mathbf{I}_j|\alpha) \neq E(\mathbf{I}_j|\alpha^*)$. As noted earlier that item j can classify examinees into three groups based on the key and coded distractor l , as shown in Table 2. Their $E(\mathbf{I}_j|\alpha)$ are further presented in Table 1. Although item j may contain other coded distractor(s), besides coded distractor l , the inclusion of these additional coded distractors does not affect the proof due to the fact that an examinee can select only one option at a time and thus, only one coded distractor l is considered to analyze how examinees are classified by the key and this specific distractor. If additional coded

Table 1: $E(\mathbf{I}_j|\boldsymbol{\alpha})$ of Classes Formulated by Item j

Class	$\alpha \not\geq \mathbf{q}_j^{(H_j^*)}$		$\alpha \succeq \mathbf{q}_j^{(H_j^*)}$
	$\alpha \not\geq \mathbf{q}_j^{(l)}$		
	G_1	G_2	G_3
$E(I_{jH_j^*} \alpha)$	$\frac{1}{H_j}$ or $\epsilon_{jH_j^*l'}$	$\epsilon_{jH_j^*l}$	$1 - \sum_{c \neq H_j^*} \epsilon_{jcH_j^*}$
$E(I_{jl} \alpha)$	$\frac{1}{H_j}$ or $\epsilon_{jl'}$	$1 - \sum_{c \neq l} \epsilon_{jcl}$	$\epsilon_{jlH_j^*}$
$E(I_{j*} \alpha)$	$\frac{1}{H_j}$ or $1 - \sum_{c \neq l'} \epsilon_{jcl'}$	ϵ_{j*l}	$\epsilon_{j*H_j^*}$

distractors are included, the analysis would follow a similar pattern, simply shifting the focus to the new distractor. To maintain conciseness, the discussion is limited to the groups divided by the key and coded distractor l and thus, the potential other coded distractors are collectively denoted by $*$.

As shown in Table 1, examinees in G_3 have the ideal response H_j^* , so their $E(I_{jH_j^*}|\boldsymbol{\alpha})$, $E(I_{jl}|\boldsymbol{\alpha})$, and $E(I_{j*}|\boldsymbol{\alpha})$ are $1 - \sum_{c \neq H_j^*} \epsilon_{jcH_j^*}$, $\epsilon_{jlH_j^*}$, $\epsilon_{j*H_j^*}$, respectively. Examinees in G_2 have the ideal response l , so their $E(I_{jH_j^*}|\boldsymbol{\alpha})$, $E(I_{jl}|\boldsymbol{\alpha})$, and $E(I_{j*}|\boldsymbol{\alpha})$ are $\epsilon_{jH_j^*l}$, $1 - \sum_{c \neq l} \epsilon_{jcl}$, and ϵ_{j*l} , respectively. However, for examinees in G_1 , their $E(I_{jH_j^*}|\boldsymbol{\alpha})$, $E(I_{jl}|\boldsymbol{\alpha})$, and $E(I_{j*}|\boldsymbol{\alpha})$ depend on the presence of additional coded distractors. If there are no additional coded distractors, their $E(I_{jH_j^*}|\boldsymbol{\alpha}) = E(I_{jl}|\boldsymbol{\alpha}) = E(I_{j*}|\boldsymbol{\alpha}) = \frac{1}{H_j}$. If there are additional coded distractors, $E(I_{jH_j^*}|\boldsymbol{\alpha})$ is ϵ_{j*l} , $E(I_{jl}|\boldsymbol{\alpha})$ is ϵ_{j*l} , and $E(I_{j*}|\boldsymbol{\alpha})$ is $1 - \sum_{c \neq l'} \epsilon_{jcl'}$, where l' is a level of an other coded distractor. Another omitted case is that item j does not contain any coded distractor, the absence of coded distractors does not affect the result, because only G_3 and G_1 are relevant. In this case, for examinees in G_1 , $E(\mathbf{I}_j|\boldsymbol{\alpha}) = (\frac{1}{H_j})$, while for examinees in G_3 , their $E(\mathbf{I}_j|\boldsymbol{\alpha})$ is as shown in Table 1.

The key message Table 1 sends out is that for $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^*$ in different groups, $E(\mathbf{I}_j|\boldsymbol{\alpha}) \neq E(\mathbf{I}_j|\boldsymbol{\alpha}^*)$. Hence, to prove that $\varsigma(\boldsymbol{\alpha}) \neq \varsigma(\boldsymbol{\alpha}^*)$ for any $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^*$, it is equivalent to show that $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}^*$ are classified into different groups by at least one item. Now, suppose $\mathbf{Q}_{MC}$ includes all $\mathbf{e}_k$, where $k = 1, \dots, K$. For each $\mathbf{e}_k$, two cases

are to be discussed

Case 1: Suppose A1 is true. Suppose item j contains $\mathbf{e}_k$ and $\mathbf{e}_k \preceq \mathbf{q}_j^{(H_j^*)}$. Based on Table 1, examinees whose attribute profiles differ by attribute k are classified into different groups. Specifically, examinees who have mastered attribute k will be classified into G_2 or G_3 , while those who have not mastered attribute k will be classified into G_1 .

Case 2: Suppose A2 is true. According to Lemma 1, examinees who have mastered attribute k will be classified into G_2 or G_3 by any item in $\mathcal{J}$, while those who have not mastered attribute k will be classified into G_1 by at least one item in $\mathcal{J}$. Therefore, examinees whose attribute profiles differ by attribute k are classified into different groups by at least one item j , where $j \in \mathcal{J}$.

Consequently, for any $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^*$ that differ in the mastery of at least one attribute, there exists at least one item j that classifies $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^*$ into different groups. Therefore, there exists at least one item j such that $E(\mathbf{I}_j|\boldsymbol{\alpha})$ and $E(\mathbf{I}_j|\boldsymbol{\alpha}^*)$ differ, which leads to $\varsigma(\boldsymbol{\alpha}) \neq \varsigma(\boldsymbol{\alpha}^*)$. Hence, the Q-matrix is complete.

($\Rightarrow$) Suppose $\mathbf{Q}_{\text{MC}}$ is complete. Two cases are formed to facilitate the discussion:

Case 1: $\mathbf{Q}_{\text{MC}}$ does not include all $\mathbf{e}_k$. Assume that $\mathbf{e}_{k'}$ for some k' is absent from $\mathbf{Q}_{\text{MC}}$. Consider $\boldsymbol{\alpha} = \mathbf{0}_K$ and $\boldsymbol{\alpha}^* = \mathbf{e}_{k'}$, it can be derived that $\varsigma_j(\boldsymbol{\alpha}) = \varsigma_j(\boldsymbol{\alpha}^*) = \frac{1}{H_j^*}$ for all j . This result implies that $\mathbf{Q}_{\text{MC}}$ cannot distinguish all latent classes, which contradicts the assumption that $\mathbf{Q}_{\text{MC}}$ is complete. Hence, all $\mathbf{e}_k$ must be included in $\mathbf{Q}_{\text{MC}}$.

Case 2: $\mathbf{Q}_{\text{MC}}$ includes all $\mathbf{e}_k$, but for some attribute k , neither A1 nor A2 holds. Without loss of generality, assume that attribute k' is only contained in item j' but $\mathbf{e}_{k'} \not\preceq \mathbf{q}_{j'}^{(H_{j'}^*)}$. Further assume A2 is violated. Consider attribute profiles $\boldsymbol{\alpha} = \mathbf{1}_K$ and $\boldsymbol{\alpha}^* = \mathbf{1}_K - \mathbf{e}_{k'}$. In this case, they have the same ideal response H_j^* to all items, leading $\varsigma_j(\boldsymbol{\alpha}) = \varsigma_j(\boldsymbol{\alpha}^*)$ to all j . Therefore, $\mathbf{Q}_{\text{MC}}$ cannot differentiate all latent classes, which contradicts the assumption of completeness. Therefore, it is necessary for $\mathbf{Q}_{\text{MC}}$ to satisfy either condition A1 or A2.

□

Proof of Corollary 1

Corollary 1. *Suppose $\mathbf{Q}_B$ is complete for data conforming to the DINA model. Then any useful and proper $\mathbf{Q}_{MC}$ augmented from $\mathbf{Q}_B$ with data conforming to the MC-DINA model is also complete.*

Proof. Since $\mathbf{Q}_B$ is complete for data conforming to the DINA model, augmenting $\mathbf{Q}_B$ into $\mathbf{Q}_{MC}$ ensures that $\mathbf{q}_j^{(H_j^*)} = \mathbf{e}_k$ for each item j containing $\mathbf{e}_k$. Therefore, A1 of Theorem 1 is satisfied, and the augmented $\mathbf{Q}_{MC}$ is complete for data conforming to the MC-DINA model. $\square$

Proof of Lemma 2

Lemma 2. *Suppose $\mathbf{Q}_{MC}$ is the Q -matrix of a test of MC items conforming to the MC-DINA model. The attribute profile $\boldsymbol{\alpha} = \mathbf{0}_K$ can be identified by $\mathbf{Q}_{MC}$ if and only if $\mathbf{Q}_{MC}$ includes all $\mathbf{e}_k$, regardless of whether they define the keys or coded distractors.*

Proof. Recall that an attribute profile $\boldsymbol{\alpha}$ is identifiable if and only if $\varsigma(\boldsymbol{\alpha}) = \varsigma(\boldsymbol{\alpha}^*)$ implies $\boldsymbol{\alpha} = \boldsymbol{\alpha}^*$.

($\Leftarrow$) Suppose that $\mathbf{Q}_{MC}$ includes all $\mathbf{e}_k$. For $\boldsymbol{\alpha} = \mathbf{0}_K$, $\varsigma_j(\boldsymbol{\alpha}) = \frac{1}{H_j} \mathbf{1}_{H_j}$ for all j . Further suppose $\boldsymbol{\alpha}^* \neq \boldsymbol{\alpha}$. Then there must exist some k such that $\alpha_{i'k} = 1$. Assume that $\mathbf{e}_k$ defines coded option l of item j . Then, it must be true that $\varsigma_{jl}(\boldsymbol{\alpha}^*) = 1 - \sum_{c \neq l} \epsilon_{jcl}$. Because $1 - \sum_{c \neq l} \epsilon_{jcl} > .5$ but $\frac{1}{H_j} < .5$, it is true that $\varsigma_j(\boldsymbol{\alpha}) \neq \varsigma_j(\boldsymbol{\alpha}^*)$, which further implies that $\varsigma(\boldsymbol{\alpha}) \neq \varsigma(\boldsymbol{\alpha}^*)$. Hence, it can be concluded that $\boldsymbol{\alpha}^* \neq \boldsymbol{\alpha}$ implies $\varsigma(\boldsymbol{\alpha}) \neq \varsigma(\boldsymbol{\alpha}^*)$, which means $\boldsymbol{\alpha} = \mathbf{0}_K$ can be identified.

($\Rightarrow$) Assume that $\boldsymbol{\alpha} = \mathbf{0}_K$ can be identified by $\mathbf{Q}_{MC}$. Now, suppose that $\mathbf{e}_k$ is missing from $\mathbf{Q}_{MC}$. For $\boldsymbol{\alpha}^* = \mathbf{e}_k$, it can be derived that $\varsigma_{jl}(\boldsymbol{\alpha}^*) = \frac{1}{H_j}$ for all j and l . Hence, $\varsigma_j(\boldsymbol{\alpha}) = \varsigma_j(\boldsymbol{\alpha}^*) = \frac{1}{H_j} \mathbf{1}_{H_j}$ for all j , which implies that $\varsigma(\boldsymbol{\alpha}) = \varsigma(\boldsymbol{\alpha}^*)$. In

conclusion, $\alpha^* \neq \alpha$ but $\varsigma(\alpha) = \varsigma(\alpha^*)$. Hence, $\alpha = \mathbf{0}_K$ cannot be identified, which is a contradiction. Therefore, all e_k must be included in $\mathbf{Q}_{MC}$.

□

Proof of Lemma 3

Lemma 3. *Suppose $\mathbf{Q}_{MC}$ is the Q -matrix of a test of MC items conforming to the MC-DINA model. The attribute profile $\alpha = \mathbf{m}_k$, where $k = 1, \dots, K$, can be identified by $\mathbf{Q}_{MC}$ if and only if the test satisfies the following conditions:*

C1 There exists a set of k' items, where $1 \leq k' \leq k$, such that item j , $j = 1, 2, \dots, k'$, includes a coded option defined by $\mathbf{q}_j$, subject to $\bigcup_{j=1}^{k'} \mathbf{q}_j = \mathbf{m}_k$. Let O_1 be the set consisting of these k' q -vectors.

C2 There exists a set of k'' items, where $1 \leq k'' \leq K - k$, such that they consist of $K - k$ coded options with q -vectors taking the following form,

$$\begin{pmatrix} \mathbf{Q}_* & \mathbf{I}_{K-k} \end{pmatrix}$$

where $\mathbf{Q}_*$ is a $(K - k) \times k$ matrix with arbitrary entries. The keys of these items cannot be any $\mathbf{q} \in O_1$. Let O_2 be the set consisting of these $K - k$ q -vectors.

Proof. ($\Leftarrow$) Suppose a test with $\mathbf{Q}_{MC}$ satisfies C1 and C2. All items can be categorized into three groups:

- $\mathcal{J}_1 = \{j | \exists h \text{ s.t. } \mathbf{q}_j^{(h)} \in O_1\} \setminus \{j | \exists h \text{ s.t. } \mathbf{q}_j^{(h)} \in O_2\}$
- $\mathcal{J}_2 = \{j | \exists h \text{ s.t. } \mathbf{q}_j^{(h)} \in O_2\} \setminus \{j | \exists h \text{ s.t. } \mathbf{q}_j^{(h)} \in O_1\}$
- $\mathcal{J}_3 = \{j | \exists h \text{ and } h' \text{ s.t. } \mathbf{q}_j^{(h)} \in O_1 \text{ and } \mathbf{q}_j^{(h')} \in O_2\}$
- $\mathcal{J}_4 = \{j | \mathbf{q}_j^{(h)} \notin O_1 \cup O_2, \forall h\} = \{j | \mathbf{q}_j^{(h)} \succeq \mathbf{e}_s + \mathbf{e}_t, \text{ where } s \neq t, \text{ and } s, t \in \{k + 1, \dots, K\}, \forall h\}$

Consider another attribute profile $\alpha^* = \mathbf{m}_{k'}$. Suppose that $\eta_j(\alpha) = l$ and $\eta_j(\alpha^*) = l'$. Because C1 and C2 hold, sets $\mathcal{J}_1 \cup \mathcal{J}_3$ and $\mathcal{J}_2 \cup \mathcal{J}_3$ is non-empty.

If $k' > k$, it must be true that $\alpha_{k'}^* = 1$ but $\alpha_{k'} = 0$. Hence, according to C2, there exists an item $j \in \mathcal{J}_2 \cup \mathcal{J}_3$ such that $\varsigma_{jl'}(\alpha) = \frac{1}{H_j}$ or $\epsilon_{jl'l}$ and $\varsigma_{jl'}(\alpha^*) = 1 - \sum_{c \neq l'} \epsilon_{jcl'}$. Because $\frac{1}{H_j} < .5$, $\epsilon_{jl'l} < 0.5$, and $1 - \sum_{c \neq l'} \epsilon_{jcl'} > .5$, it leads to $\varsigma_{jl'}(\alpha) \neq \varsigma_{jl'}(\alpha^*)$, which implies $\varsigma_j(\alpha) \neq \varsigma_j(\alpha^*)$ and thus $\varsigma(\alpha) \neq \varsigma(\alpha^*)$. In conclusion, $\alpha \neq \alpha^*$ implies $\varsigma(\alpha) \neq \varsigma(\alpha^*)$.

If $k' < k$, it must be true that $\alpha_k^* = 0$ but $\alpha_k = 1$. According to C1, there exists an item $j \in \mathcal{J}_1 \cup \mathcal{J}_3$ such that $\varsigma_{jl}(\alpha^*) = \frac{1}{H_j}$ or $\epsilon_{jll'}$ and $\varsigma_{jl}(\alpha) = 1 - \sum_{c \neq l} \epsilon_{jcl}$. Similar to the reasoning applied above, it is true that $\varsigma(\alpha) \neq \varsigma(\alpha^*)$. In conclusion, it can be shown that $\alpha \neq \alpha^*$ implies $\varsigma(\alpha) \neq \varsigma(\alpha^*)$. Hence, $\alpha = \mathbf{m}_k$ can be identified by $\mathcal{Q}_{MC}$.

($\Rightarrow$) Assume that the attribute profile $\alpha = \mathbf{m}_k$, where $k = 1, \dots, K$, can be identified by $\mathcal{Q}_{MC}$.

i) Suppose that C1 does not hold. Without loss of generality, assume that the k th attribute is not required by any $\mathbf{q} \in O_1$; namely, the $\mathbf{q}$ -vectors in O_1 satisfy $\bigcup_{j=1}^{k'} \mathbf{q}_j = \mathbf{m}_{k-1}$. Further suppose that $\alpha^* = \mathbf{m}_{k-1}$. Then, for each $j \in \mathcal{J}_1 \cup \mathcal{J}_3$, $\eta_j(\alpha) = \eta_j(\alpha^*) = l$. Therefore, it can be derived that $\varsigma_{jl}(\alpha) = \varsigma_{jl}(\alpha^*) = 1 - \sum_{c \neq l} \epsilon_{jcl}$. Hence, $\varsigma_{j \in \mathcal{J}_1 \cup \mathcal{J}_3}(\alpha) = \varsigma_{j \in \mathcal{J}_1 \cup \mathcal{J}_3}(\alpha^*)$. Further more, as shown earlier, for each $j' \in \mathcal{J}_2$, $\eta_{j'}(\alpha) = \eta_{j'}(\alpha^*) = 0$, leading to $\varsigma_{j \in \mathcal{J}_2}(\alpha) = \varsigma_{j \in \mathcal{J}_2}(\alpha^*) = \frac{1}{H_j} \mathbf{1}_{H_j}$. Hence, $\varsigma_{j \in \mathcal{J}_2}(\alpha) = \varsigma_{j \in \mathcal{J}_2}(\alpha^*)$. Finally, for each $j'' \in \mathcal{J}_4$, $\eta_{j''}(\alpha) = \eta_{j''}(\alpha^*) = 0$. It can be derived that $\varsigma_{j''}(\alpha) = \varsigma_{j''}(\alpha^*) = \frac{1}{H_{j''}} \mathbf{1}_{H_{j''}}$, which implies that $\varsigma_{j \in \mathcal{J}_4}(\alpha) = \varsigma_{j \in \mathcal{J}_4}(\alpha^*)$. Based on the above reasoning and also because $\bigcup_{s=1}^4 \mathcal{J}_s = \{1, \dots, J\}$, it can be concluded that $\varsigma(\alpha) = \varsigma(\alpha^*)$, despite the fact that $\alpha \neq \alpha^*$, implying that α cannot be identified by $\mathcal{Q}_{MC}$, which is a contradiction. Hence, C1 has to hold.

ii) Suppose that C2 does not hold. Without loss of generality, assume that $\mathbf{q}_* + \mathbf{e}_{k+1}$, where $\mathbf{q}_* \prec \mathbf{m}_k$, is absent from O_2 (i.e., $\mathbf{e}_1$ is absent from $\mathbf{I}_{K-k}$). Further suppose $\alpha^* = \mathbf{m}_{k+1}$. There are two reasons for this absence. The first scenario is that $\mathbf{q}_* + \mathbf{e}_{k+1}$ is not present in $\mathcal{Q}_{MC}$. Then, for each $j \in \mathcal{J}_1 \cup \mathcal{J}_3$, $\eta_j(\alpha) = \eta_j(\alpha^*) = l$, and for each

$j' \in \mathcal{J}_2 \cup \mathcal{J}_4$, $\eta_{j'}(\boldsymbol{\alpha}) = \eta_{j'}(\boldsymbol{\alpha}^*) = 0$. The second possibility is that $\mathbf{q}_* + \mathbf{e}_{k+1}$ exists in $\mathbf{Q}_{MC}$, while the key of item j is from O_1 . Under this circumstance, it is true that $\eta_j(\boldsymbol{\alpha}) = \eta_j(\boldsymbol{\alpha}^*) = H_j^*$. Similar to the previous case, $\eta_{j'}(\boldsymbol{\alpha}) = \eta_{j'}(\boldsymbol{\alpha}^*) = l$ for each $j' \in \mathcal{J}_1 \cup \mathcal{J}_3 \setminus \{j\}$; $\eta_{j''}(\boldsymbol{\alpha}) = \eta_{j''}(\boldsymbol{\alpha}^*) = 0$ for each $j'' \in \mathcal{J}_2 \cup \mathcal{J}_4$. Using the arguments similar to those presented in (i), it can be established that $\boldsymbol{\varsigma}(\boldsymbol{\alpha}) = \boldsymbol{\varsigma}(\boldsymbol{\alpha}^*)$, despite the fact that $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}^*$. Hence, $\boldsymbol{\alpha}$ cannot be identified, which is a contradiction. Thus, C2 must hold.

□

Proof of Corollary 2

Corollary 2. *Suppose each MC item conforms to a level-2 design and $3 \leq H_j < 7$. Then, the minimum test length of a complete $\mathbf{Q}_{MC}$ is $\lceil \frac{K}{2} \rceil$, achieved by measuring $2 \times \lfloor \frac{K}{2} \rfloor$ attributes using $\lfloor \frac{K}{2} \rfloor$ MC items.*

Proof. If K is even, all $\mathbf{e}_k$ can be associated with $\frac{K}{2}$ items through a level-2 design. In contrast, if K is odd, $(K - 1)$ $\mathbf{e}_k$ can be associated with $\frac{K-1}{2}$ items through a level-2 design. To include the last $\mathbf{e}_k$, an additional MC item, which goes with either a level-1 or level-2 design, is required. Therefore, a total of $\frac{K+1}{2}$ MC items are required when K is odd. According to Theorem 1, a Q-matrix with this structure, where A1 is satisfied for each $\mathbf{e}_k$, is a complete Q-matrix.

In conclusion, the minimum test length needed to contain all $\mathbf{e}_k$ is $\lceil \frac{K}{2} \rceil$. This is under the condition that $\lfloor \frac{K}{2} \rfloor$ MC items, each designed with a level-2 design, collectively contain $2 \times \lfloor \frac{K}{2} \rfloor$ distinct $\mathbf{e}_k$.

□

Proof of Theorem 2

Theorem 2. *For MC items under the MC-DINA model, with $6 \leq H_j < 14$, a set of $J^* = \lceil \frac{K}{2} \rceil$ MC items is Q-optimal if and only if for each α , the Q-matrices of these test items take the minimum form shown in Equations 4.7 or 4.8 after column swapping.*

Proof. ($\Leftarrow$) Suppose that $\mathbf{Q}_{\text{MC}}$ take the minimum form shown in Equations 4.7 or 4.8 after column swapping. Further suppose that α^* is the true attribute profile. To show that $\mathbf{Q}_{\text{MC}}$ is Q-optimal, it is required to show that $P(\hat{\alpha} = \alpha^* | \alpha^*, \mathbf{Q}_{\text{MC}}) > \max_{\mathbf{a} \neq \alpha^*} P(\hat{\alpha} = \mathbf{a} | \alpha^*, \mathbf{Q}_{\text{MC}})$. The first step is to determine the estimate given the observed responses, $\mathbf{X} = (X_1, \dots, X_{J^*})'$. Subsequently, the second step is to show $P(\mathbf{X} = \eta(\alpha^*) | \alpha^*, \mathbf{Q}_{\text{MC}})$ is the largest among all other possible observed responses.

The first goal is to prove that given the observed responses of the first J^* items, $\mathbf{X} = (X_1, \dots, X_{J^*})'$, the posterior mode estimator is

$$\hat{\alpha} = \alpha_{\text{true}}, \quad (1)$$

where $\eta(\alpha_{\text{true}})$ satisfies $\eta(\alpha_{\text{true}}) = \mathbf{X}$. To prove 1, it is required to show that

$$P(\hat{\alpha} = \alpha_{\text{true}} | \mathbf{X}) > P(\hat{\alpha} = \mathbf{a} | \mathbf{X})$$

for all $\mathbf{a} \neq \alpha_{\text{true}}$, which is equivalent to

$$\frac{P(\hat{\alpha} = \alpha_{\text{true}} | \mathbf{X})}{P(\hat{\alpha} = \mathbf{a} | \mathbf{X})} > 1. \quad (2)$$

Recall that the posterior distribution of α , $P(\alpha | \mathbf{X})$, can be expressed as

$$P(\alpha | \mathbf{X}) = \frac{P(\alpha, \mathbf{X})}{P(\mathbf{X})} = \frac{P(\mathbf{X} | \alpha)P(\alpha)}{P(\mathbf{X})}.$$

Hence, it is true that

$$\frac{P(\hat{\alpha} = \alpha_{\text{true}}|\mathbf{X})}{P(\hat{\alpha} = \mathbf{a}|\mathbf{X})} = \frac{P(\mathbf{X}|\alpha_{\text{true}})P(\alpha_{\text{true}})}{P(\mathbf{X})} \frac{P(\mathbf{X})}{P(\mathbf{X}|\mathbf{a})P(\mathbf{a})} = \frac{P(\mathbf{X}|\alpha_{\text{true}})P(\alpha_{\text{true}})}{P(\mathbf{X}|\mathbf{a})P(\mathbf{a})}. \quad (3)$$

Given the non-informative prior, to prove Equation 2, it is equivalent to prove

$$\frac{P(\hat{\alpha} = \alpha_{\text{true}}|\mathbf{X})}{P(\hat{\alpha} = \mathbf{a}|\mathbf{X})} = \frac{P(\mathbf{X}|\alpha_{\text{true}})}{P(\mathbf{X}|\mathbf{a})} > 1. \quad (4)$$

To prove Equation 4, three cases are considered.

- *Case 1: $\mathbf{a} \succ \alpha_{\text{true}}$.*

Let $\mathcal{J}^* = \{1, 2, \dots, J^*\}$ be the set consisting of the items that achieve the Q-optimal condition. Each of these J^* items can be classified into one of the following three subsets:

- $\mathcal{J}_1 = \{j | \eta_j(\alpha_{\text{true}}) = \eta_j(\mathbf{a})\},$
- $\mathcal{J}_2 = \{j | \eta_j(\alpha_{\text{true}}) = 0, \quad \eta_j(\mathbf{a}) > 0\},$
- $\mathcal{J}_3 = \{j | \eta_j(\alpha_{\text{true}}) > 0, \quad \eta_j(\mathbf{a}) > 0, \text{ but } \eta_j(\alpha_{\text{true}}) \neq \eta_j(\mathbf{a})\}.$

Notice that $\mathcal{J}_1$, $\mathcal{J}_2$, and $\mathcal{J}_3$ are mutually exclusive, implying $\mathcal{J}^* = \mathcal{J}_1 \cup \mathcal{J}_2 \cup \mathcal{J}_3$, $\mathcal{J}_1 \cap \mathcal{J}_2 = \emptyset$, $\mathcal{J}_1 \cap \mathcal{J}_3 = \emptyset$, $\mathcal{J}_2 \cap \mathcal{J}_3 = \emptyset$. Hence, Equation 4 can be reformulated as follows:

$$\begin{aligned} \frac{P(\mathbf{X}|\alpha_{\text{true}})}{P(\mathbf{X}|\mathbf{a})} &= \frac{\prod_{j \in \mathcal{J}_1} P(X_j|\alpha_{\text{true}}) \prod_{j \in \mathcal{J}_2} P(X_j|\alpha_{\text{true}}) \prod_{j \in \mathcal{J}_3} P(X_j|\alpha_{\text{true}})}{\prod_{j \in \mathcal{J}_1} P(X_j|\mathbf{a}) \prod_{j \in \mathcal{J}_2} P(X_j|\mathbf{a}) \prod_{j \in \mathcal{J}_3} P(X_j|\mathbf{a})} \\ &= \frac{\prod_{j \in \mathcal{J}_2} P(X_j|\alpha_{\text{true}}) \prod_{j \in \mathcal{J}_3} P(X_j|\alpha_{\text{true}})}{\prod_{j \in \mathcal{J}_2} P(X_j|\mathbf{a}) \prod_{j \in \mathcal{J}_3} P(X_j|\mathbf{a})} \\ &= \frac{\prod_{j \in \mathcal{J}_2} \frac{H_j - H_j^*}{H_j} \prod_{j \in \mathcal{J}_3} (1 - \sum_{c \neq l} \epsilon_{jcl})}{\prod_{j \in \mathcal{J}_2} \epsilon_{j0l'} \prod_{j \in \mathcal{J}_3} \epsilon_{jl'}}. \end{aligned} \quad (5)$$

Given that $6 \leq H_j < 14$ and $H_j^* = 3$, it follows that $\frac{H_j^*}{H_j} \leq .5$, which implies that $\frac{H_j - H_j^*}{H_j} \geq .5$. Additionally, it is assumed that each $\epsilon_{jcl} < .5$ for every j . Hence,

$$\frac{\prod_{j \in \mathcal{J}_2} \frac{H_j - H_j^*}{H_j}}{\prod_{j \in \mathcal{J}_2} \epsilon_{j0l'}} > 1.$$

Furthermore, because $\sum_{c \neq l} \epsilon_{jcl} < .5$ for every j or equivalently, $1 - \sum_{c \neq l} \epsilon_{jcl} > .5$, it is true that

$$\frac{\prod_{j \in \mathcal{J}_3} (1 - \sum_{c \neq l} \epsilon_{jcl})}{\prod_{j \in \mathcal{J}_3} \epsilon_{jll'}} > 1.$$

Consequently, the value computed by Equation 5 must exceed 1, which implies that Equations 2 and thus, 1 hold. In conclusion, it is true that $\hat{\alpha} = \alpha_{\text{true}}$, where $\eta(\alpha_{\text{true}}) = \mathbf{X}$, is the posterior mode estimator when the observed response vector is $\mathbf{X}$.

- *Case 2: $\mathbf{a} \prec \alpha_{\text{true}}$.*

In this condition, $\mathcal{J}_1$ and $\mathcal{J}_3$ remain the same, but $\mathcal{J}_2$ changes to $\mathcal{J}_2 = \{j | \eta_j(\alpha_{\text{true}}) > 0, \eta_j(\mathbf{a}) = 0\}$. Then, Equation 4 can be reformulated as follows:

$$\begin{aligned} \frac{P(\mathbf{X} | \alpha_{\text{true}})}{P(\mathbf{X} | \mathbf{a})} &= \frac{\prod_{j \in \mathcal{J}_1} P(X_j | \alpha_{\text{true}}) \prod_{j \in \mathcal{J}_2} P(X_j | \alpha_{\text{true}}) \prod_{j \in \mathcal{J}_3} P(X_j | \alpha_{\text{true}})}{\prod_{j \in \mathcal{J}_1} P(X_j | \mathbf{a}) \prod_{j \in \mathcal{J}_2} P(X_j | \mathbf{a}) \prod_{j \in \mathcal{J}_3} P(X_j | \mathbf{a})} \\ &= \frac{\prod_{j \in \mathcal{J}_2} P(X_j | \alpha_{\text{true}}) \prod_{j \in \mathcal{J}_3} P(X_j | \alpha_{\text{true}})}{\prod_{j \in \mathcal{J}_2} P(X_j | \mathbf{a}) \prod_{j \in \mathcal{J}_3} P(X_j | \mathbf{a})} \\ &= \frac{\prod_{j \in \mathcal{J}_2} (1 - \sum_{c \neq l} \epsilon_{jcl}) \prod_{j \in \mathcal{J}_3} (1 - \sum_{c \neq l} \epsilon_{jcl})}{\prod_{j \in \mathcal{J}_2} \frac{1}{H_j} \prod_{j \in \mathcal{J}_3} \epsilon_{jll'}}. \end{aligned} \quad (6)$$

Again, because $\sum_{c \neq l} \epsilon_{jcl} < .5$ for every j , it is true that $1 - \sum_{c \neq l} \epsilon_{jcl} > .5$. Hence,

$$\frac{\prod_{j \in \mathcal{J}_2} (1 - \sum_{c \neq l} \epsilon_{jcl})}{\prod_{j \in \mathcal{J}_2} \frac{1}{H_j}} > 1,$$

and

$$\frac{\prod_{j \in \mathcal{J}_3} (1 - \sum_{c \neq l} \epsilon_{jcl})}{\prod_{j \in \mathcal{J}_3} \epsilon_{jll'}} > 1.$$

Consequently, the value derived from Equation 6 must be greater than 1, which implies that Equations 2 and thus, 1 hold. In conclusion, it is true that $\hat{\alpha} = \alpha_{\text{true}}$, where $\eta(\alpha_{\text{true}}) = \mathbf{X}$, is the posterior mode estimator when the observed response vector is $\mathbf{X}$.

- *Case 3: $\mathbf{a} \not\prec \alpha_{\text{true}}$ and $\mathbf{a} \not\prec \alpha_{\text{true}}$.*

In this condition, $\mathcal{J}_1$, $\mathcal{J}_2$, and $\mathcal{J}_3$ remain the same as in Case 1. However, $\mathcal{J}_4$ is constructed:

$$- \mathcal{J}_4 = \{j | \eta_j(\alpha_{\text{true}}) > 0, \quad \eta_j(\mathbf{a}) = 0\}.$$

Then, Equation 4 can be re-expressed as follows:

$$\begin{aligned} \frac{P(\mathbf{X} | \alpha_{\text{true}})}{P(\mathbf{X} | \mathbf{a})} &= \frac{\prod_{j \in \mathcal{J}_1} P(X_j | \alpha_{\text{true}}) \prod_{j \in \mathcal{J}_2} P(X_j | \alpha_{\text{true}}) \prod_{j \in \mathcal{J}_3} P(X_j | \alpha_{\text{true}}) \prod_{j \in \mathcal{J}_4} P(X_j | \alpha_{\text{true}})}{\prod_{j \in \mathcal{J}_1} P(X_j | \mathbf{a}) \prod_{j \in \mathcal{J}_2} P(X_j | \mathbf{a}) \prod_{j \in \mathcal{J}_3} P(X_j | \mathbf{a}) \prod_{j \in \mathcal{J}_4} P(X_j | \mathbf{a})} \\ &= \frac{\prod_{j \in \mathcal{J}_2} P(X_j | \alpha_{\text{true}}) \prod_{j \in \mathcal{J}_3} P(X_j | \alpha_{\text{true}}) \prod_{j \in \mathcal{J}_4} P(X_j | \alpha_{\text{true}})}{\prod_{j \in \mathcal{J}_2} P(X_j | \mathbf{a}) \prod_{j \in \mathcal{J}_3} P(X_j | \mathbf{a}) \prod_{j \in \mathcal{J}_4} P(X_j | \mathbf{a})} \\ &= \frac{\prod_{j \in \mathcal{J}_2} \frac{H_j - H_j^*}{H_j} \prod_{j \in \mathcal{J}_3} (1 - \sum_{c \neq l} \epsilon_{jcl}) \prod_{j \in \mathcal{J}_4} (1 - \sum_{c \neq l} \epsilon_{jcl})}{\prod_{j \in \mathcal{J}_2} \epsilon_{j0l'} \prod_{j \in \mathcal{J}_3} \epsilon_{jll'} \prod_{j \in \mathcal{J}_4} \frac{1}{H_j}}. \end{aligned} \quad (7)$$

Based on the analysis done in Cases 1 and 2, it is true that

$$\frac{\prod_{j \in \mathcal{J}_2} \frac{H_j - H_j^*}{H_j}}{\prod_{j \in \mathcal{J}_2} \epsilon_{j0l'}} > 1,$$

$$\frac{\prod_{j \in \mathcal{J}_3} (1 - \sum_{c \neq l} \epsilon_{jcl})}{\prod_{j \in \mathcal{J}_3} \epsilon_{jll'}} > 1,$$

and

$$\frac{\prod_{j \in \mathcal{J}_4} (1 - \sum_{c \neq l} \epsilon_{jcl})}{\prod_{j \in \mathcal{J}_4} \frac{1}{H_j}} > 1.$$

Hence, $\frac{P(\mathbf{X}|\boldsymbol{\alpha}_{\text{true}})}{P(\mathbf{X}|\mathbf{a})}$ in Equation 7 must be greater than 1, which implies that Equations 2 and thus, 1 hold. In conclusion, it is true that $\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}_{\text{true}}$, where $\boldsymbol{\eta}(\boldsymbol{\alpha}_{\text{true}}) = \mathbf{X}$, is the posterior mode estimator when the observed response vector is $\mathbf{X}$.

The above determines $P(\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}^*|\mathbf{X} = \boldsymbol{\eta}(\boldsymbol{\alpha}^*))$ is the largest, indicating that the estimate equals to the true attribute profile as long as the observed response vector is identical to the ideal response vector of the true attribute profile. Therefore, to further prove $P(\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}^*|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}}) > \max_{\mathbf{a} \neq \boldsymbol{\alpha}^*} P(\hat{\boldsymbol{\alpha}} = \mathbf{a}|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}})$, it is equivalent to prove the inequality $P(\mathbf{X} = \boldsymbol{\eta}(\boldsymbol{\alpha}^*)|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}}) > P(\mathbf{X} = \boldsymbol{\eta}(\mathbf{a})|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}})$ for all $\mathbf{a} \neq \boldsymbol{\alpha}^*$.

Let $\eta_j(\boldsymbol{\alpha}^*) = l_j^*$ and $\eta_j(\mathbf{a}) = l_j$. Formally, it can be shown that

$$\begin{aligned}
 P(\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}^*|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}}) &= P(\mathbf{X} = \boldsymbol{\eta}(\boldsymbol{\alpha}^*)|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}}) \\
 &= \prod_{j=1}^{J^*} P(X_j = \eta_j(\boldsymbol{\alpha}^*)|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}}) \\
 &= \prod_{j:l_j^* > 0} P(X_j = l_j^*|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}}) \prod_{j:l_j^* = 0} P(X_j = l_j^*|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}}) \\
 &= \prod_{j:l_j^* > 0} (1 - \sum_{c \neq l_j^*} \epsilon_{jcl_j^*}) \prod_{j:l_j^* = 0} \frac{H_j - H_j^*}{H_j}. \tag{8}
 \end{aligned}$$

Furthermore, for any $\mathbf{a} \neq \boldsymbol{\alpha}^*$,

$$\begin{aligned}
 P(\hat{\boldsymbol{\alpha}} = \mathbf{a}|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}}) &= P(\mathbf{X} = \boldsymbol{\eta}(\mathbf{a})|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}}) \\
 &= \prod_{j=1}^{J^*} P(X_j = \eta_j(\mathbf{a})|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}}) \\
 &= \prod_{j:l_j^* > 0, l_j = l_j^*} P(X_j = l_j|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}}) \prod_{j:l_j^* > 0, l_j \neq l_j^*} P(X_j = l_j|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}}) \\
 &\quad \prod_{j:l_j^* = 0, l_j = 0} P(X_j = l_j|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}}) \prod_{j:l_j^* = 0, l_j \neq 0} P(X_j = l_j|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}}) \\
 &= \prod_{j:l_j^* > 0, l_j = l_j^*} (1 - \sum_{c \neq l_j^*} \epsilon_{jcl_j^*}) \prod_{j:l_j^* > 0, l_j \neq l_j^*} \epsilon_{jl_j l_j^*} \prod_{j:l_j^* = 0, l_j = 0} \frac{H_j - H_j^*}{H_j} \prod_{j:l_j^* = 0, l_j \neq 0} \frac{1}{H_j} \tag{9}
 \end{aligned}$$

Because $1 - \sum_{c \neq l} \epsilon_{jcl} > .5 > \epsilon_{jcl}$, it is true that $\prod_{j:l_j^* > 0, l_j = l_j^*} (1 - \sum_{c \neq l_j^*} \epsilon_{jcl_j^*})$ $\prod_{j:l_j^* > 0, l_j \neq l_j^*} \epsilon_{jl_j l_j^*} < \prod_{j:l_j^* > 0} (1 - \sum_{c \neq l_j^*} \epsilon_{jcl_j^*})$. In addition, because $H_j - H_j^* > 1$, implying that $\frac{H_j - H_j^*}{H_j} > \frac{1}{H_j}$, it can be shown that $\prod_{j:l_j^* = 0, l_j = 0} \frac{H_j - H_j^*}{H_j} \prod_{j:l_j^* = 0, l_j \neq 0} \frac{1}{H_j} < \prod_{j:l_j^* = 0} \frac{H_j - H_j^*}{H_j}$. Consequently, Equation 9 must be smaller than Equation 8, which implies $P(\mathbf{X} = \boldsymbol{\eta}(\boldsymbol{\alpha}^*) | \boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}}) > P(\mathbf{X}_i = \boldsymbol{\eta}(\mathbf{a}) | \boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}})$, further implies $P(\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}^* | \boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}}) > P(\hat{\boldsymbol{\alpha}} = \mathbf{a} | \boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}})$. Hence, $\mathbf{Q}_{\text{MC}}$ is optimal.

($\Rightarrow$) Suppose $\mathbf{Q}_{\text{MC}}$ is optimal but does not conform to Equation 4.7 or 4.8. This implies that either not every item among these J^* items ($J^* - 1$ items when K is odd) evaluates the mastery of two attributes, or the unspecified entry $q_{j,k}^{(l)}$ is one if the option requiring attribute k of item $\lfloor \frac{k}{2} \rfloor$ is not selected.

To simplify the proof, assume that when not every item among these J^* items evaluates the mastery of two attributes, all unspecified entries are zero, regardless of the observed response. Without assuming $\mathbf{Q}_j$ of J^* items ($J^* - 1$ items when K is odd) take level-2 designs, the following three cases are considered.

- *Case 1.* At least one $\mathbf{Q}_j$ has a level-1 design, and the remaining $\mathbf{Q}_j$ have level-2 designs.

A $\mathbf{Q}_j$ with a level-1 design means item j assesses the mastery of only one attribute. The more $\mathbf{Q}_j$ with level-1 designs there are, the more items are needed to estimate the attribute profile. Therefore, consider the simplest condition: only one $\mathbf{Q}_j$ has a level-1 design, and the remaining $\mathbf{Q}_j$ have level-2 designs. Without loss of generality, consider that K is even and $\mathbf{Q}_1$ has a level-1 design for attribute 1, while the remaining $J^* - 1$ $\mathbf{Q}_j$ have level-2 designs measuring attributes 2 to $(K - 1)$. With this layout, attribute K is not evaluated by any item, leading to the situation where $\boldsymbol{\alpha}^* = \mathbf{1}_K$ and $\mathbf{a} = \mathbf{1}_K - \mathbf{e}_K$ have the same ideal responses to these J^* items; that is, $\eta_j(\boldsymbol{\alpha}^*) = \eta_j(\mathbf{a}) = H_j^*$ for every j . Based on Equation 6, $\mathcal{J}_2 = \emptyset$, and $\mathcal{J}_3 = \emptyset$, Equation 4 becomes

$$\frac{P(\hat{\alpha} = \mathbf{1}_K | \mathbf{X} = \boldsymbol{\eta}(\mathbf{1}_K))}{P(\hat{\alpha} = \mathbf{1}_K - \mathbf{e}_K | \mathbf{X} = \boldsymbol{\eta}(\mathbf{1}_K))} = \frac{\prod_{j \in \mathcal{J}_2} (1 - \sum_{c \neq l} \epsilon_{jcl}) \prod_{j \in \mathcal{J}_3} (1 - \sum_{c \neq l} \epsilon_{jcl})}{\prod_{j \in \mathcal{J}_2} \frac{1}{H_j} \prod_{j \in \mathcal{J}_3} \epsilon_{jll'}} = 1,$$

which means the posteriors of $\hat{\alpha} = \alpha^*$ and $\hat{\alpha} = \mathbf{a}$ are the same, given the observed responses are identical to the ideal responses of α^* . Therefore, $\frac{P(\hat{\alpha} = \alpha^* | \mathbf{X} = \boldsymbol{\eta}(\alpha^*))}{P(\hat{\alpha} = \mathbf{a} | \mathbf{X} = \boldsymbol{\eta}(\alpha^*))} > 1$ does not hold, showing this design distorts the posterior mode estimator. This implies that

$$P(\hat{\alpha} = \alpha^* | \alpha^*, \mathbf{Q}_{\text{MC}}) > \max_{\mathbf{a} \neq \alpha^*} P(\hat{\alpha} = \mathbf{a} | \alpha^*, \mathbf{Q}_{\text{MC}})$$

does not hold. Consequently, $\mathbf{Q}_{\text{MC}}$ is not optimal.

- *Case 2.* At least one $\mathbf{Q}_j$ has a level-3 or a higher-level design.

A $\mathbf{Q}_j$ with a higher-level design means item j can assess the mastery of more attributes due to including more $\mathbf{e}_k$, which increases H_j^* . Based on Table 4.3, achieving a level-3 design requires that $H_j^* \geq 7$. Given that $6 \leq H_j < 14$, it leads to $\frac{H_j^*}{H_j} > 0.5$ due to the inflated H_j^* . Consider the simplest condition: only one $\mathbf{Q}_j$ has a level-3 design, and the remaining $\mathbf{Q}_j$ have level-2 designs. Without loss of generality, consider that K is odd and item 1 has a level-3 design measuring attributes 1 to 3. Since K is odd, the remaining $\frac{K-3}{2}$ items are needed to measure attributes 4 to K . Suppose $\alpha^* = \mathbf{0}_K$ and $\mathbf{a} = \mathbf{e}_1$. Based on Equation 5, $\mathcal{J}_2 = \{j | \eta_j(\mathbf{0}_K) = 0, \eta_j(\mathbf{e}_1) > 0\}$, and $\mathcal{J}_3 = \emptyset$. As a result, Equation 4 becomes,

$$\begin{aligned} \frac{P(\hat{\alpha} = \mathbf{0}_K | \mathbf{X} = \boldsymbol{\eta}(\mathbf{0}_K))}{P(\hat{\alpha} = \mathbf{e}_1 | \mathbf{X} = \boldsymbol{\eta}(\mathbf{0}_K))} &= \frac{\prod_{j \in \mathcal{J}_2} \frac{H_j - H_j^*}{H_j} \prod_{j \in \mathcal{J}_3} (1 - \sum_{c \neq l} \epsilon_{jcl})}{\prod_{j \in \mathcal{J}_2} \epsilon_{j0l'} \prod_{j \in \mathcal{J}_3} \epsilon_{jll'}} \\ &= \frac{\prod_{j \in \mathcal{J}_2} \frac{H_j - H_j^*}{H_j}}{\prod_{j \in \mathcal{J}_2} \epsilon_{j0l'}}. \end{aligned}$$

Given $\frac{H_j^*}{H_j} > 0.5$, it follows that $\frac{H_j - H_j^*}{H_j} < 0.5$. Additionally, since $\epsilon_{j0l'} < 0.5$, the relationship between $\prod_{j \in \mathcal{J}_2} \frac{H_j - H_j^*}{H_j}$ and $\prod_{j \in \mathcal{J}_2} \epsilon_{j0l'}$ remains uncertain. Hence $\frac{P(\hat{\alpha} = \alpha^* | \mathbf{X} = \boldsymbol{\eta}(\alpha^*))}{P(\hat{\alpha} = \mathbf{a} | \mathbf{X} = \boldsymbol{\eta}(\alpha^*))} > 1$ does not always hold, indicating that

$$P(\hat{\alpha} = \alpha^* | \alpha^*, \mathbf{Q}_{\text{MC}}) > \max_{\mathbf{a} \neq \alpha^*} P(\hat{\alpha} = \mathbf{a} | \alpha^*, \mathbf{Q}_{\text{MC}})$$

is not always true. A higher-level design with a larger H_j^* increases the uncertainty between $P(\hat{\alpha} = \alpha^* | \alpha^*, \mathbf{Q}_{\text{MC}})$ and $P(\hat{\alpha} = \mathbf{a} | \alpha^*, \mathbf{Q}_{\text{MC}})$. Consequently, $\mathbf{Q}_{\text{MC}}$ is not always optimal when at least one $\mathbf{Q}_j$ has a level-3 or a higher-level design.

- *Case 3.* At least one $\mathbf{Q}_j$ does not take a design in any level.

In the context of MC items, a $\mathbf{Q}_j$ without any level design means it does not contain any $\mathbf{e}_k$. Without loss of generality, assume that K is even and item 1 has a Q-matrix without any level design, specified as follows,

$$\mathbf{Q}_1 = \begin{pmatrix} 1 & 1 & 1 & 0 & \dots & 0 & 0 \\ 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 0 & 1 & 0 & \dots & 0 & 0 \end{pmatrix},$$

and the remaining $J^* - 1$ items have level-2 designs for attributes 3 to K . For $\alpha^* = \mathbf{0}_K$ and $\mathbf{a} = \mathbf{e}_1$, $\eta_j(\alpha^*) = \eta_j(\mathbf{a}) = 0$ for every j . Based on Equation 5, $\mathcal{J}_2 = \emptyset$, and $\mathcal{J}_3 = \emptyset$, Equation 4 becomes,

$$\frac{P(\hat{\alpha} = \mathbf{0}_K | \mathbf{X} = \boldsymbol{\eta}(\mathbf{0}_K))}{P(\hat{\alpha} = \mathbf{e}_1 | \mathbf{X} = \boldsymbol{\eta}(\mathbf{0}_K))} = \frac{\prod_{j \in \mathcal{J}_2} \frac{H_j - H_j^*}{H_j} \prod_{j \in \mathcal{J}_3} (1 - \sum_{c \neq l} \epsilon_{jcl})}{\prod_{j \in \mathcal{J}_2} \epsilon_{j0l'} \prod_{j \in \mathcal{J}_3} \epsilon_{jll'}} = 1.$$

Same as Case 1, the posterior mode estimator is no longer α^* , given the observed responses are the same as ideal response of α^* . Therefore, $\frac{P(\hat{\alpha} = \alpha^* | \mathbf{X} = \boldsymbol{\eta}(\alpha^*))}{P(\hat{\alpha} = \mathbf{a} | \mathbf{X} = \boldsymbol{\eta}(\alpha^*))} > 1$

does not hold, which further indicates that

$$P(\hat{\alpha} = \alpha^* | \alpha^*, \mathbf{Q}_{\text{MC}}) > \max_{\mathbf{a} \neq \alpha^*} P(\hat{\alpha} = \mathbf{a} | \alpha^*, \mathbf{Q}_{\text{MC}})$$

is not met. Consequently, $\mathbf{Q}_{\text{MC}}$ is not optimal.

After establishing that every item out of these J^* items must evaluate the mastery of two attributes, the next step is to prove that the unspecified entry $q_{j,k}^{(l)}$ in Equations 4.7 and 4.8 must be 0 if the option requiring attribute k of item $\lceil \frac{k}{2} \rceil$ is not selected.

Unspecified entries exist in items after the first one. Based on the above proof, $\mathbf{Q}_1$ must conform to the level-2 design (if K is odd, it can be achieved by row swapping). Without loss of generality, suppose $\alpha^* = \mathbf{e}_3$. Since α^* will select a noncoded option of item 1, it shows the nonmastery of attributes 1 and 2. If $q_{2,1}^{(3)}$, $q_{2,2}^{(3)}$, $q_{2,1}^{(2)}$, $q_{2,2}^{(2)}$, $q_{2,1}^{(1)}$, and $q_{2,2}^{(1)}$ in $\mathbf{Q}_2$ are specified as 1, as shown below, it violates Equations 4.7 and 4.8.

$$\mathbf{Q}_2 = \begin{pmatrix} 1 & 1 & 1 & 1 & 0 & \dots & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & \dots & 0 & 0 \end{pmatrix}.$$

Further suppose $\mathbf{a} = \mathbf{0}_K$. Then, it can be shown that $\eta_j(\alpha^*) = \eta_j(\mathbf{a}) = 0$ for every j . As a result, $P(\hat{\alpha} = \alpha^* | \alpha^*, \mathbf{Q}_{\text{MC}}) = P(\hat{\alpha} = \mathbf{a} | \alpha^*, \mathbf{Q}_{\text{MC}})$. Based on Equation 6, $\mathcal{J}_2 = \emptyset$, and $\mathcal{J}_3 = \emptyset$, Equation 4 becomes,

$$\frac{P(\hat{\alpha} = \mathbf{e}_3 | \mathbf{X} = \boldsymbol{\eta}(\mathbf{e}_3))}{P(\hat{\alpha} = \mathbf{0}_K | \mathbf{X} = \boldsymbol{\eta}(\mathbf{e}_3))} = \frac{\prod_{j \in \mathcal{J}_2} (1 - \sum_{c \neq l} \epsilon_{jcl}) \prod_{j \in \mathcal{J}_3} (1 - \sum_{c \neq l} \epsilon_{jcl})}{\prod_{j \in \mathcal{J}_2} \frac{1}{H_j} \prod_{j \in \mathcal{J}_3} \epsilon_{jll'}} = 1,$$

which means α^* is no longer the sole posterior mode estimator. Thus, $\frac{P(\hat{\alpha} = \alpha^* | \mathbf{X} = \boldsymbol{\eta}(\alpha^*))}{P(\hat{\alpha} = \mathbf{a} | \mathbf{X} = \boldsymbol{\eta}(\alpha^*))} > 1$ does not hold, which further indicates that

$$P(\hat{\alpha} = \alpha^* | \alpha^*, \mathbf{Q}_{\text{MC}}) > \max_{\mathbf{a} \neq \alpha^*} P(\hat{\alpha} = \mathbf{a} | \alpha^*, \mathbf{Q}_{\text{MC}})$$

is not satisfied. Consequently, $\mathbf{Q}_{\text{MC}}$ is not optimal.

To sum up, the above proof demonstrates the necessity of Equations 4.7 and 4.8; otherwise, the posterior mode estimate will be distorted and will not equal to the true attribute profile. $\square$

Every item involved in the Q-optimal design, except for the initial one, is likely to include extra coded options additional to those shown in Equations 4.7 and 4.8, when some of the unspecified entries are 1. These unspecified entries being 1 indicate that the corresponding attributes have already been mastered by α_{true} . However, their presence does not affect the Q-optimal procedure.

Recall that in the proof of Theorem 2, it was first demonstrated that $P(\hat{\alpha} = \alpha_{\text{true}} | \mathbf{X}) > P(\hat{\alpha} = \mathbf{a} | \mathbf{X})$, where $\mathbf{X} = \boldsymbol{\eta}(\alpha_{\text{true}})$, followed by the verification of the inequality $P(\mathbf{X} = \boldsymbol{\eta}(\alpha^*) | \alpha^*, \mathbf{Q}_{\text{MC}}) > P(\mathbf{X} = \boldsymbol{\eta}(\mathbf{a}) | \alpha^*, \mathbf{Q}_{\text{MC}})$. To show that the additional coded options in $\mathbf{Q}_{\text{MC}}$ have no influence on $P(\hat{\alpha} = \alpha^* | \alpha^*, \mathbf{Q}_{\text{MC}}) > \max_{\mathbf{a} \neq \alpha^*} P(\hat{\alpha} = \mathbf{a} | \alpha^*, \mathbf{Q}_{\text{MC}})$, the discussion is divided into two parts. First, it will be demonstrated that the additional coded options does not affect $P(\hat{\alpha} = \alpha_{\text{true}} | \mathbf{X}) > P(\hat{\alpha} = \mathbf{a} | \mathbf{X})$. In the previous proof, this was demonstrated by showing $\frac{P(\hat{\alpha} = \alpha_{\text{true}} | \mathbf{X})}{P(\hat{\alpha} = \mathbf{a} | \mathbf{X})} > 1$, which was further analyzed in three cases.

Case 1: $\mathbf{a} \succ \alpha_{\text{true}}$.

$$\frac{P(\mathbf{X} | \alpha_{\text{true}})}{P(\mathbf{X} | \mathbf{a})} = \frac{\prod_{j \in \mathcal{J}_2} \frac{H_j - H_j^*}{H_j} \prod_{j \in \mathcal{J}_3} (1 - \sum_{c \neq l} \epsilon_{jcl})}{\prod_{j \in \mathcal{J}_2} \epsilon_{j0l'} \prod_{j \in \mathcal{J}_3} \epsilon_{jll'}}.$$

If additional coded options exist, $\eta_j(\alpha_{\text{true}})$ might change from 0 to $l \neq 0$ for some $j \in \mathcal{J}_2$, and thus these items are shifted from $\mathcal{J}_2$ to $\mathcal{J}_3$. However, other items in $\mathcal{J}_1$ and $\mathcal{J}_3$ remain the same. This does not alter the underlying logic of the proof.

Case 2: $\mathbf{a} \prec \alpha_{\text{true}}$.

$$\frac{P(\mathbf{X} | \alpha_{\text{true}})}{P(\mathbf{X} | \mathbf{a})} = \frac{\prod_{j \in \mathcal{J}_2} (1 - \sum_{c \neq l} \epsilon_{jcl}) \prod_{j \in \mathcal{J}_3} (1 - \sum_{c \neq l} \epsilon_{jcl})}{\prod_{j \in \mathcal{J}_2} \frac{1}{H_j} \prod_{j \in \mathcal{J}_3} \epsilon_{jll'}}.$$

Items in $\mathcal{J}_1$ and $\mathcal{J}_3$ remain unchanged, but $\eta_j(\mathbf{a})$ might change from 0 to $l' \neq 0$ for some $j \in \mathcal{J}_2$, and thus, these items will be moved from $\mathcal{J}_2$ to $\mathcal{J}_3$. Again, this does not affect the proof's validity.

Case 3: $\mathbf{a} \not\prec \boldsymbol{\alpha}_{\text{true}}$ and $\mathbf{a} \not\succ \boldsymbol{\alpha}_{\text{true}}$.

$$\frac{P(\mathbf{X}|\boldsymbol{\alpha}_{\text{true}})}{P(\mathbf{X}|\mathbf{a})} = \frac{\prod_{j \in \mathcal{J}_2} \frac{H_j - H_j^*}{H_j} \prod_{j \in \mathcal{J}_3} (1 - \sum_{c \neq l} \epsilon_{jcl}) \prod_{j \in \mathcal{J}_4} (1 - \sum_{c \neq l} \epsilon_{jcl})}{\prod_{j \in \mathcal{J}_2} \epsilon_{j0l'} \prod_{j \in \mathcal{J}_3} \epsilon_{jlw'} \prod_{j \in \mathcal{J}_4} \frac{1}{H_j}}.$$

In this case, the additional coded options do not affect the items in $\mathcal{J}_1$ and $\mathcal{J}_3$. Although some items may be moved to $\mathcal{J}_3$ because $\eta_j(\boldsymbol{\alpha}_{\text{true}}) = l \neq 0$ for some $j \in \mathcal{J}_2$ and $\eta_j(\mathbf{a}) = l' \neq 0$ for some $j \in \mathcal{J}_4$, this does not alter the reasoning underlying the proof.

In conclusion, the inclusion of additional coded options does not alter the condition $P(\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}_{\text{true}}|\mathbf{X}) > P(\hat{\boldsymbol{\alpha}} = \mathbf{a}|\mathbf{X})$, where $\mathbf{X} = \boldsymbol{\eta}(\boldsymbol{\alpha}_{\text{true}})$.

Second, it will be demonstrated that the relative magnitude of $P(\mathbf{X} = \boldsymbol{\eta}(\boldsymbol{\alpha}^*)|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}})$ and $P(\mathbf{X} = \boldsymbol{\eta}(\mathbf{a})|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}})$, as shown in Equations 8 and 9, remains unchanged. Assume additional coded options are added to item j , where the original value of $\eta_j(\boldsymbol{\alpha}^*)$ may be $l^* = 0$ or $l^* > 0$. If it is 0, $\eta_j(\boldsymbol{\alpha}^*)$ changes from 0 to $l^* > 0$ after additional coded options are introduced. If it is already $l^* > 0$, $\eta_j(\boldsymbol{\alpha}^*)$ remains unchanged. The additional coded options do not affect the model assumption, so $1 - \sum_{c \neq l} \epsilon_{jcl} > .5 > \epsilon_{jcl}$ continues to hold, keeping $\prod_{j: l_j^* > 0} (1 - \sum_{c \neq l_j^*} \epsilon_{jcl_j^*}) > \prod_{j: l_j^* > 0, l_j = l_j^*} (1 - \sum_{c \neq l_j^*} \epsilon_{jcl_j^*}) \prod_{j: l_j^* > 0, l_j \neq l_j^*} \epsilon_{jl_j l_j^*}$ valid. For the remaining part, $\prod_{j: l_j^* = 0} \frac{H_j - H_j^*}{H_j} > \prod_{j: l_j^* = 0, l_j = 0} \frac{H_j - H_j^*}{H_j} \prod_{j: l_j^* = 0, l_j \neq 0} \frac{1}{H_j}$ still holds, as $\frac{H_j - H_j^*}{H_j} > \frac{1}{H_j}$ remains true. Therefore, $P(\mathbf{X} = \boldsymbol{\eta}(\boldsymbol{\alpha}^*)|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}}) > P(\mathbf{X} = \boldsymbol{\eta}(\mathbf{a})|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}})$ continues to hold.

Based on the above discussion, $P(\hat{\boldsymbol{\alpha}} = \boldsymbol{\alpha}^*|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}}) > \max_{\mathbf{a} \neq \boldsymbol{\alpha}^*} P(\hat{\boldsymbol{\alpha}} = \mathbf{a}|\boldsymbol{\alpha}^*, \mathbf{Q}_{\text{MC}})$ remains true and thus, Theorem 2 still holds when additional coded options exist.

Proof of Corollary 3

Corollary 3. *For MC items with $H_j = 4$ conforming to the MC-DINA model, $J^* = K$ MC items are needed to achieve the Q-optimality.*

Proof. Based on Equations 8 and 9, to ensure that the inequality $P(\hat{\alpha} = \alpha^* | \alpha^*, \mathbf{Q}_{\text{MC}}) > P(\hat{\alpha} = \mathbf{a} | \alpha^*, \mathbf{Q}_{\text{MC}})$ holds, it is necessary that both $1 - \sum_{c \neq l} \epsilon_{jcl} > .5 > \epsilon_{jcl}$ and $H_j - H_j^* > 1$ are satisfied for each j . The former condition is automatically satisfied under the MC-DINA model assumption. However, for MC items with $H_j = 4$ to have a Q-matrix as in Equations 4.7 or 4.8, $H_j^* = 3$ is required, which violates $H_j - H_j^* > 1$. As a result, the Q-optimal process proposed in Theorem 2 does not apply in this case. Instead, using MC items with fewer coded options, such that $H_j^* = 1$ or 2, satisfying $H_j - H_j^* > 1$, is more appropriate. Therefore, the Q-optimal design becomes conceptually analogous to the Q-optimal design with binary items, as shown in Equation 4.5 after column swapping. This means each q-vector in Equation 4.5 is no longer required to define the key in $\mathbf{Q}_{\text{MC}}$ for the Q-optimal procedure but may instead define the coded distractor, as long as they are nested with their respective keys. $\square$

Proof of Corollary 4

Corollary 4. *Theorem 2 holds for MC items with $H_j = 5$ if $\epsilon_{jcl} \leq \frac{2}{5}$.*

Proof. Recall that in Theorem 2, the posterior mode estimator, $\hat{\alpha}$, is shown to satisfy $\hat{\alpha} = \alpha_{\text{true}}$, given $\mathbf{X} = (X_1, \dots, X_{J^*})'$, by demonstrating that $\frac{P(\hat{\alpha} = \alpha_{\text{true}} | \mathbf{X})}{P(\hat{\alpha} = \mathbf{a} | \mathbf{X})} > 1$. The proof is divided into three cases: Case 1 where $\mathbf{a} \succ \alpha_{\text{true}}$, Case 2 where $\mathbf{a} \prec \alpha_{\text{true}}$, and Case 3 where $\mathbf{a} \not\succ \alpha_{\text{true}}$ and $\mathbf{a} \not\prec \alpha_{\text{true}}$. When $H_j = 5$, given $\mathbf{X} = (X_1, \dots, X_{J^*})'$, $\hat{\alpha}$ may no longer equal α_{true} in Cases 1 and 3, which means $\frac{P(\hat{\alpha} = \alpha_{\text{true}} | \mathbf{X})}{P(\hat{\alpha} = \mathbf{a} | \mathbf{X})} > 1$ does not hold in these two cases.

As stated before, $\frac{P(\mathbf{X}|\boldsymbol{\alpha}_{\text{true}})}{P(\mathbf{X}|\mathbf{a})}$ in Case 1 is shown in Equation 5:

$$\frac{P(\mathbf{X}|\boldsymbol{\alpha}_{\text{true}})}{P(\mathbf{X}|\mathbf{a})} = \frac{\prod_{j \in \mathcal{J}_2} \frac{H_j - H_j^*}{H_j} \prod_{j \in \mathcal{J}_3} (1 - \sum_{c \neq l} \epsilon_{jcl})}{\prod_{j \in \mathcal{J}_2} \epsilon_{j0l'} \prod_{j \in \mathcal{J}_3} \epsilon_{jll'}}.$$

The RHS of this equation is greater than 1 if both $\frac{H_j - H_j^*}{H_j} > \epsilon_{jcl}$ and $1 - \sum_{c \neq l} \epsilon_{jcl} > \epsilon_{jll'}$ hold true. Although the latter is automatically satisfied under the MC-DINA model assumption, the former is not always true. Specifically, since $\epsilon_{jcl} < 0.5$ is assumed by the model, adopting a Q-matrix as in Equations 4.7 or 4.8 for Q-optimal procedure with $H_j = 5$, $\frac{H_j - H_j^*}{H_j} = \frac{2}{5} < 0.5$. As a result, $\frac{H_j - H_j^*}{H_j} > \epsilon_{jcl}$ may not hold, making it possible that $\frac{P(\mathbf{X}|\boldsymbol{\alpha}_{\text{true}})}{P(\mathbf{X}|\mathbf{a})} < 1$, which implies that $\hat{\boldsymbol{\alpha}}$ may not be $\boldsymbol{\alpha}_{\text{true}}$.

It was previously noted that in Case 3, $\frac{P(\mathbf{X}|\boldsymbol{\alpha}_{\text{true}})}{P(\mathbf{X}|\mathbf{a})}$ is expressed by Equation 7:

$$\frac{P(\mathbf{X}|\boldsymbol{\alpha}_{\text{true}})}{P(\mathbf{X}|\mathbf{a})} = \frac{\prod_{j \in \mathcal{J}_2} \frac{H_j - H_j^*}{H_j} \prod_{j \in \mathcal{J}_3} (1 - \sum_{c \neq l} \epsilon_{jcl}) \prod_{j \in \mathcal{J}_4} (1 - \sum_{c \neq l} \epsilon_{jcl})}{\prod_{j \in \mathcal{J}_2} \epsilon_{j0l'} \prod_{j \in \mathcal{J}_3} \epsilon_{jll'} \prod_{j \in \mathcal{J}_4} \frac{1}{H_j}}.$$

Similar to Case 1, because $\prod_{j \in \mathcal{J}_2} \frac{H_j - H_j^*}{H_j}$ is not always greater than $\prod_{j \in \mathcal{J}_2} \epsilon_{j0l'}$, it may happen that $\frac{P(\mathbf{X}|\boldsymbol{\alpha}_{\text{true}})}{P(\mathbf{X}|\mathbf{a})} < 1$, which implies that $\hat{\boldsymbol{\alpha}}$ may not be $\boldsymbol{\alpha}_{\text{true}}$.

To sum up, when adopting a Q-matrix as in Equations 4.7 or 4.8 for the Q-optimal procedure with $H_j = 5$, it can be derived that $\frac{H_j - H_j^*}{H_j} = \frac{2}{5}$. If $\epsilon_{jcl} \geq \frac{2}{5}$, the condition $\frac{H_j - H_j^*}{H_j} > \epsilon_{jcl}$ does not hold, violating the conditions of Theorem 2. However, if $\epsilon_{jcl} < \frac{2}{5}$, the condition $\frac{H_j - H_j^*}{H_j} > \epsilon_{jcl}$ remains valid, and Theorem 2 holds.

□