**Kastamonu Education Journal**, 2025, Vol. 33, No:1, 48-66
doi: 10.24106/kefdergi.1628232

| Research Article / Araştırma Makalesi |

# Comparison of Classification Accuracy and Parameters of DINA, DINO, HO-DINA and HO-DINO Models in the Framework of Cognitive Diagnosis in Health Education

# Sağlık Eğitiminde Bilişsel Tanı Çerçevesinde DINA, DINO, HO-DINA ve HO-DINO Modellerinin Sınıflama Doğruluğu ve Parametrelerinin Karşılaştırılması[1]

**Sena Gencan[2], Şeref Tan[3]**

**Keywords**

1. Cognitive Diagnostic Models
2. DINA model
3. DINO model
4. Q matrix
5. Health Education

**Anahtar Kelimeler**

1. Bilişsel Tanı Modelleri
2. DINA model
3. DINO model
4. Q matrisi
5. Sağlık Eğitimi

**Received/Başvuru Tarihi**
23.08.2023

**Accepted / Kabul Tarihi**
17.04.2024

## Abstract

*Purpose: This study aims to compare the parameters of DINA, DINO, HO-DINA and HO-DINO models according to different sample sizes (500, 2000, 5000) and different item numbers (60, 120) based on the Q matrices created for different attributes in health education based on simulation data.*

*Design/Methodology/Approach: In the simulation data, 50 replications were performed for each condition. In the study, two different Q-Matrixes were determined based on the learning domain determined by considering the 2018 TUS Spring Assessment Report and the taxonomy included in the Clinical assessment framework determined in Miller's 1990 study as the attributes dimension in the Q-Matrix in which matching of attribute and item is carried out. In the study, RMSEA, g and s parameters and classification accuracies were compared and under which conditions DINA, DINO, HO-DINA and HO-DINO models gave similar or different results were investigated.*

*Findings: According to the research findings, the Q-Matrix, in which Fields levels were used as the attribute dimension, was the matrix that gave the best parameter results in all models. In addition, it has been determined that the models that give the best RMSEA, g and s parameters and classification accuracies are DINO and HO-DINO models in the analysis.*

*Highlights: Based on the findings, when analyzing the results for the Basic Medical Sciences and Clinical Medical Sciences tests, it is evident that the Q matrix determined by Fields provides a better fit to the data, and moreover, it is advantageous for the Q matrix determined by Fields to be used for the TUS exam.*

## Öz

*Çalışmanın amacı: Bu çalışma simülasyon verilerine dayalı olarak sağlık eğitiminde farklı niteliklere göre oluşturulan Q matrislerini temel alarak DINA, DINO, HO-DINA ve HO-DINO modellerinin parametrelerini farklı örneklem büyüklüğü (500, 2000, 5000) ve farklı madde sayılarına (60, 120) göre karşılaştırmayı amaçlamaktadır.*

*Materyal ve Yöntem: Simülasyon veride her bir koşul için 50 replikasyon yapılmıştır. Çalışmada, nitelik ile madde eşleştirmesinin yapıldığı Q-Matrisinde nitelik boyutları olarak 2018 TUS İlkbahar Değerlendirme Raporu göz önüne alınarak belirlenen öğrenme alanları ve Miller'in 1990 yılındaki çalışmasında belirlediği Klinik değerlendirme çerçevesinde yer alan taksonomi temel alınarak iki farklı Q- Matrisi belirlenmiştir. Çalışmada RMSEA, g ve s parametreleri ve sınıflama doğrulukları karşılaştırılmış ve DINA, DINO, HO-DINA ve HO-DINO modellerinin hangi koşullar altında benzer ya da farklı sonuç verdikleri incelenmiştir.*

*Bulgular: Araştırmanın bulgularına göre nitelik boyutu olarak Alanlar düzeylerinin kullanıldığı Q-Matrisi tüm modellerde en iyi parametre sonuçları veren matris olmuştur. Ayrıca yapılan analizlerde en iyi RMSEA, g ve s parametreleri ve sınıflama doğruluklarını veren modellerin DINO ve HO-DINO modelleri olduğu tespit edilmiştir.*

*Önemli Vurgular: Bulgulara dayanarak Temel Tıp ve Klinik Tıp Bilimleri testi için sonuçlar incelendiğinde Alanlara göre belirlenen Q matrisinin veriye daha iyi uyum sağladığı görülmekle birlikte Alanlara göre belirlenen Q matrisinin TUS sınavı için kullanılmasının daha avantajlı olduğu ortaya konulmuştur.*

[1] This paper is adapted from a doctoral thesis carried out in Gazi University, Institute of Educational Sciences, Department of Measurement and Evaluation in Education.

[2] **Corresponded Author,** Gazi University, Department of Quality, TURKIYE; https://orcid.org/0009-0009-1836-4172

[3] * Prof. Dr., Ankara, TURKIYE; https://orcid.org/0000-0002-9892-3369

**Citation/Alıntı:** Gencan, S., & Tan, Ş. (2025). Comparison of Classification Accuracy and Parameters of DINA, DINO, HO-DINA and HO-DINO Models in the Framework of Cognitive Diagnosis in Health Education, *Kastamonu Education Journal, 33*(1), 48-66. doi: 10.24106/kefdergi.1628232

## INTRODUCTION

Health in the World Health Organization Charter (1948) is defined as “a state of physical, social and spiritual well-being and not merely the absence of disease or infirmity”. Health, one of the basic human rights, is not the purpose of people's life, but it is a resource for people to continue their daily lives. (Sağlık Bakanlığı, 2011).

On the other hand, health services are the services provided for the elimination of various factors that harm human health and the protection of society from the effects of these factors, the treatment of patients, and the rehabilitation of those with reduced physical and mental abilities and skills (Ministry of Health, 2001). Various health institutions operating in the private and public sectors provide health services to the community through doctors, nurses and other health personnel (Bakan et al., 2011). In the study by Doğan and Gencan (2014), in which doctors, nurses and other health personnel are considered the basic inputs in producing health services, it has been exposed that doctor are the most important input in providing health services.

In order to ensure quality assurance in health services, it is an important priority to equip the labor force of doctors, nurses and other health personnel with health services (Aydın & Demir, 2006). Medical faculties fulfill the task of training doctors, who are the most important personnel in delivering health services. Medical education aims to train doctors who will support and improve the health of all people (World Health Organization, 1988). When evaluated in this context, it is seen that it is important to create test designs that will make appropriate diagnoses in measurement and evaluation processes in health education.

Diagnostic assessments provide students, their families, and educators with more detailed information about scores. Diagnostic assessments that provide students and teachers with reliable feedback on student strengths and weaknesses impact education and training significantly. (Jang & Wagner, 2014).

Diagnostic tests measure a person’s competencies on components embedded in the theoretical learning model to aid instructional design. (Grégoire, 1997). Such diagnostic assessments identify specific deficiencies in students' essential prior skills or knowledge or permanent mislearning. Prior skills or knowledge include concepts or tasks required to complete the targeted tasks in the teaching field successfully and are often called ability in the cognitive model (Tatsuoka & Tatsuoka, 1997). Cognitive Diagnostic Models (CDMs) are one of the diagnostic assessment approaches that provide statistical classification of participants according to one or more abilities. (Rupp et al., 2010).

CDMs, which have gained increasing importance in the measurement literature in recent years, are essential for accurately classifying and ultimately identifying where and how respondents are missing from educational measurements to clinical assessments. (Rupp et al., 2010). In the literature, there are many studies in which CDMs are used in educational measurements. However, there are limited studies in which cognitive diagnosis models (CDM) are used in medical education (e.g. Collares, 2022).

Q matrix is used for analysis in all models of CDMs. The Q matrix defines the attributes required for a high probability of correctly answering each item. Each item in the Q matrix takes a value of 1 if the attribute is required and 0 if it is not (Henson et al., 2009). In some cases, a condition of having all the sub-skills measured by that item is required in the Q matrix, and possible attribute profiles are grouped under a particular hierarchy of skills. Thus, the attributes have a hierarchical structure (Aryadoust, 2018). Examining the skills hierarchy and applying different estimation models can improve the item fit and item parameters. Similarly, misidentifying or including an irrelevant attribute in the Q matrix can lead to classification problems in the model (DeCarlo, 2011; Su, 2013). In this respect, the correct specification of the Q matrix is one of the most critical steps in CDM analysis (Henson et al., 2009).

Due to the fact that DINA and DINO models (Templin & Bradshaw, 2014) are the most widely used CDM models in the literature, they were preferred in this study. In addition, HO-DINA and HO-DINO models were used to analyze the hierarchical structure between skills.

Based on this information, this study aims to compare the performance of different CDM models in determining the acquisition status of specific skills in medical education and to investigate the hierarchical structure between skills. For this purpose, model fit, item parameters and classification accuracies of the DINA and DINO models, the most commonly used CDM models, and the hierarchical cognitive diagnostic models HO-DINA and HO-DINO were compared based on simulation data. In addition, another aim of the study is to determine the Q-matrix that provides the best fit. In the study, Q matrices were created using the Turkish National Medical Specialty Exam (TUS) questions applied in the Spring 2018 semester and the simulation data sets produced for these Q matrices were analyzed. In this context, the general problem statement of the research is:

How do the parameters obtained from DINA, DINO, HO-DINA and HO-DINO change with sample size, number of items and Q-matrix in cognitive diagnostic models? In the context of this general problem statement, the following questions are to be answered:

1. In the Basic Medical Sciences Test, with the change of Q matrices and the number of items to 60 and 120 and the sample size to 500, 2000, and 5000, how do the
    a. RMSEA values,
    b. g and s item parameter estimates,
    c. classification accuracies for the models?
2. In the Clinical Medical Sciences Test, with the change of Q matrices and the number of items to 60 and 120 and the sample size to 500, 2000, and 5000, how do the

a. RMSEA values,

b. g and s item parameter estimates,

c. classification accuracies for the models?

DINA, DINO, HO-DINA and HO-DINO models are briefly explained in this study.

## DINA Model

The DINA model developed by Haertel (1989) is one of the simplest non-compensatory models (e.g., Haertel, E. H. 1989; Henson et al., 2009). Non-compensatory models assume that a respondent's lack of a particular attribute cannot be compensated for by a positive attribute in responding to an item (Rupp & Templin, 2007, 80-81). The interaction between the investigated attributes and the item properties defines the latent response variable, also known as the ideal response. The ideal response for the DINA model is defined as follows (Tatsuoka, 1995; De La Torre & Minchen, 2014; Rupp & Templin, 2008; Junker & Sijtsma, 2001):

$\xi_{ij} = \prod_{k=1}^{K} \alpha_{ik}^{q_{jk}})$ Equation 1

$q_{jk}$ is the Q matrix for attribute k of item j. The $\alpha_{ik}$, referred to as knowledge states by Tatsuoka (1995), is one if examinees i mastered attribute k and zero otherwise. If examinee i is mastered in all attributes for item j, $\xi_{ij}$=1; otherwise, $\xi_{ij}$=0. To account for the probabilistic nature of the observed response, the slip (s) and guess (g) estimation parameters associated with the ideal response are defined at the item level. The slip and guess parameters are given in Equations 2 and 3:

$s_j = P(X_{ij} = 0 | \xi_{ij} = 1)$

$g_j = P(X_{ij} = 1 | \xi_{ij} = 0)$ Equations 2 and 3

The guess probability (g), referred to as a false positive, represents the probability of responding correctly to the item when the examinees lack at least one required attribute. The slip probability (s) represents the probability that the examinees respond incorrectly when all required attributes are present and is also referred to as a false negative.

The response function for an item is given as follows:

$P(X_{ij} = 1 | \xi_{ij}) = (1 - s_j)^{\xi_{ij}} g_j^{(1-\xi_{ij})}$ Equation 4

The formula given in Equation 4 includes the estimated slip ($s_j$) and guess ($g_j$) parameters for each item.

## DINO Model

The DINO model, which is an alternative model to the DINA model, was developed by Templin and Henson (2006). The DINO model assumes that for an examinee to have a high probability of responding positively to an item, it is sufficient to master only one attribute, in contrast to the DINA model. (Rupp et al., 2010).

Similar to DINA, DINO models the probability of a correct response as a function of the slip parameter (sj) and the guess parameter (gj). However, instead of defining ξij, they use the parameters ωij. The latent variable ωij is defined as (Henson et al., 2009):

$\omega_{ij} = 1 - \prod_{k=1}^{K} (1 - \alpha_{ik})^{q_{jk}}$ Equation 5

Given ωij, the probability of a correct response is defined as:

$P(X_{ij} = 1 | \omega_{ij}) = (1 - s_j)^{\omega_{ij}} g_j^{(1-\omega_{ij})}$ Equation 6

In this case, if examinee i has at least one of the required attributes for item j, ωij=1, and if examinee i has none of the required attributes, ωij=0.

## The Higher-Order DINA Model and The Higher-Order DINO Model

The HO-DINA and HO-DINO models involve the hierarchical structure of cognitive abilities in the prediction process and are defined for situations where attributes are ordered hierarchically. HO-DINA and HO-DINO models have the same basic characteristics as traditional DINA and DINO models. The difference is that in the HO-DINA and HO-DINO models, possible attribute profiles are adapted under a specific attribute hierarchy. The number of attribute profiles can be determined for each hierarchical model based on the attribute hierarchy. The possible attribute profiles will differ for different attribute hierarchies (Su, 2013). HO-DINA is a model suggesting that the only necessary and sufficient condition for the response to a test item is to master all the sub-skills measured by that item (Aryadoust, 2018).

## METHOD/MATERIALS

### Research Design

This study aims to compare the model fit, classification accuracy and parameters of widely used cognitive diagnostic models (CDMs), DINA and DINO, with hierarchical cognitive diagnostic models, HO-DINA and HO-DINO, under different conditions. In this regard, the study is descriptive research.

### Data Generation

The data sets were generated using R programming according to the conditions investigated in the study. The GDINA package was used to generate the data. The simulation conditions to evaluate the model fit, item parameters and classification accuracy of the models are: a) number of items in the test, b) sample size, c) different Q-matrices.

**Table 1. Factors and Conditions Considered in The Study**

| Factors | Conditions | | |
|---|---|---|---|
| Sample Size | 500 | 2000 | 5000 |
| Number of Items | 60 | 120 | |
| Q-Matrices | Fields | Miller | |

The data was generated with sample sizes of 500, 2000, and 5000, number of items of 60 and 120, and different Q-matrices (Fields and Miller).

The Q-Matrix identified as "Fields" was determined based on the learning domains specified in the 2018 Spring Evaluation Report of the Turkish National Medical Specialty Exam (TUS) for the Basic Medical Sciences and Clinical Medical Sciences Tests. The attributes "Anatomy, Histology and Embryology, Physiology, Medical Biochemistry, Medical Microbiology, Medical Pathology, Medical Pharmacology" were identified as the attributes to be considered for the Basic Medical Sciences Test. In contrast, the "Internal Medicine, Pediatrics, Surgery, Obstetrics and Gynecology" attributes were identified for the Clinical Medical Sciences Test. The Q-matrix identified as "Miller" was determined for the Basic Medical Sciences and Clinical Medical Sciences Tests according to the taxonomy within the clinical assessment framework defined by Miller in his 1990 study. "Knows, knows how to do, knows how to show, knows what to do" were identified as the attributes to be addressed in the Basic Medical Sciences and Clinical Medical Sciences Tests according to the taxonomy established by Miller.

As a result, data sets are generated based on 12 simulation conditions consisting of 3 different sample sizes, 2 different numbers of items and 2 different Q matrices. There are 2 sub-tests as Basic Medical Sciences and Clinical Medical Sciences Tests. Accordingly, there are 24 simulation conditions, 12 for the Basic Medical Sciences Test and 12 for the Clinical Medical Sciences Test. There are 50 replications for each simulation condition.

The sample size of 1000 for the DINA model indicates that it is sufficient to provide accurate parameter estimates, and when the results are compared, the parameters show a clear improvement when the sample size is increased from 1000 to 4000. (de la Torre et al., 2010). For this reason, the sample size was determined as 500, which is below 1000 for the small sample, 2000 for the medium-sized sample, and 5000, above 4000 for the large sample.

Furthermore, a review of the literature revealed that there are studies with several replications (iteration) of 25 (De La Torre & Douglas, 2004), 50 (Ma et al., 2022), 100 (Kalkan & Başokçu, 2019; Sünbül & Kan, 2013; De La Torre, 2009; De La Torre et al., 2010) and 500 (Templin et al., 2014; Ma & Guo, 2019). Due to the long analysis time of the sets analyzed in this study, the number of replications was determined as 50.

### Measurement Instrument

The tests for which Q matrices are created in the study consisted of the 2nd Term Basic Medical Sciences Test in the Turkish National Medical Specialty Exam (TUS) and the 2nd Term Clinical Medical Sciences Test in the Turkish National Medical Specialty Exam (TUS).

The exam consists of two sections: Basic Medical Sciences Tests (TTBT) and Clinical Medical Sciences Tests (KTBT). Each test consists of 120 questions, and 150 minutes are given for each section. The exam consists of 11 fields, including Physiology, Medical Biochemistry, Medical Microbiology, Medical Pathology, Medical Pharmacology, Internal Medicine, Pediatrics, Surgery, Obstetrics and Gynecology and tests basic medical knowledge and the assessment of medical concepts and diseases. TTBT includes items related to Anatomy, Histology and Embryology, Physiology, Medical Biochemistry, Medical Microbiology, Medical Microbiology, Medical Pathology, Medical Pharmacology, and KTBT includes items related to Internal Medicine, Pediatrics, Surgery, Obstetrics and Gynecology. Based on the candidates' answers to these two tests, the Weighted Basic Medical Sciences Score (T Score), Weighted Clinical Medical Sciences Score (K Score), and Application Score for Contracted Family Physicians for Family Physicians Specialty Training (A Score) are calculated (Board of Higher Education (YÖK), https://dokuman.osym.gov.tr/pdfdokuman/2018/GENEL/tusilkbahardegraporweb13112018.pdf).

The TUS exam conducted in in Turkiye is also applied to diagnose and place individuals who have received medical education. With this characteristic, the TUS exam can be considered a diagnostic test. In this context, it is considered that approaching the TUS exam within the framework of cognitive diagnostic models will contribute to the field of health education.

## Identification of Attributes

Attributes are identified by forming two sets of attributes based on the literature, one according to the Learning Domains and the other according to Miller's taxonomy. The first Q-matrix is developed from the 2018 Spring TUS Assessment Report, which presents the learning domains, and the second is developed based on the taxonomy within the framework of clinical evaluation determined by Miller's 1990 study.

In the initial stage of attribute identification, the fields evaluated in the TTBT and KTBT in the 2018 TUS Spring Assessment Report published regarding the TUS exam were investigated. Anatomy, Histology and Embryology, Physiology, Medical Biochemistry, Medical Microbiology, Medical Pathology, Medical Pharmacology, Medical Pathology, Medical Biochemistry, Medical Microbiology, Medical Pathology, and Medical Pharmacology were identified as the attributes to be evaluated in the Basic Medical Sciences Test. In contrast, Internal Medicine, Pediatrics, Surgery, Obstetrics and Gynecology were identified as the attributes to be evaluated in the Clinical Medical Sciences Test. The 2018 Spring TUS Assessment Report, which presents the learning domains are given in Table 2.

**Table 2: Attributes Represented in The 2018 TUS Spring Assessment Report**

| Basic Medical Sciences Learning Domains | Clinical Medical Sciences Learning Domains |
|---|---|
| Anatomy | Medicine |
| Histology and Embryology | Pediatrics |
| Physiology | Surgery |
| Medical Biochemistry | Obstetrics and Gynecology |
| Medical Microbiology | |
| Medical Pathology | |
| Medical Pharmacology | |

In this study, the second set of attributes is identified based on Miller's Clinical Assessment Framework.

**Table 3: Four Levels in Miller's Clinical Assessment Framework**

| Levels | Description |
|---|---|
| Knows | Knowledge of basic knowledge and concepts |
| Knows How | Knowledge of normal-abnormal structure, mechanisms and functions and adaptation of known knowledge to new situations |
| Shows How | Demonstration of knowledge, skills and attitudes by applying them in an educational environment and under observation |
| Does | Practice of the profession in real-life conditions |

The attributes included in the test based on Miller's Clinical assessment framework are shown in Table 4.

**Table 4: Attributes Included in The Test Based on Miller's Clinical Assessment Framework**

| Basic Medical Sciences Learning domains | Clinical Medical Sciences Learning domains |
|---|---|
| Knows | Knows |
| knows how to do | knows how to do |
| knows how to show | knows how to show |
| knows what to do | knows what to do |

## Determination of Q-matrix

Considering the items in the test and their responses, the opinions of four experts in the field were sought as to which attribute or attributes each item was related to. The group of experts consisted of four people who graduated from the Faculty of Medicine, passed the TUS exam and were in postgraduate specialization (Research Assist. Doctor).

For both tests (the Basic Medical Sciences Test and the Clinical Medical Sciences Test) to be applied in this study, two different sets of attributes are defined and two different Q matrices, “Fields” and “Miller” were created. Q-matrices are constructed separately for tests with 60 items and 120 items. The Q-matrices for the 120-item test are created based on the exams conducted in the 2018 Spring Session of the Turkish National Medical Specialty Exam, consisting of 120 questions. Then, the attribute profiles for the Q-matrices with 120 items are extracted, and questions related to these profiles are determined. The questions related to the profiles are reduced by half, maintaining the distribution in the 120 items test to create a 60 items test.

The Q matrix for The Basic Medical Sciences Test created based on the Fields was investigated. Questions related to each attribute profile were identified and ranked. As there is no hierarchy in the Q-matrix created according to the Fields since the

attribute profiles are $2^k$ ve 7 attributes, $2^7$=128 attribute profiles were determined. Sixty items were formed by taking ((n-1)/2) number of items for those with the odd number of items, half (n/2) of the items with even number of items and at least one item for each attribute.

The Q matrix for The Clinical Medical Sciences Test created based on the Fields was investigated. Questions related to each attribute profile were identified and ranked. As there is no hierarchy in the Q-matrix created based on the fields since the attribute profiles are $2^k$ ve 4 attributes, $2^4$=16 attribute profiles were determined.

Since there is no hierarchy in the profiles created on the basis of Fields, the attribute profiles are determined as shown in Table 5.

**Table 5: Six-teen Attribute Profiles for The Clinical Medical Sciences Test Created Based on The Fields**

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| P1 | 0 | 0 | 0 | 0 | P9 | 0 | 1 | 1 | 0 |
| P2 | 1 | 0 | 0 | 0 | P10 | 0 | 1 | 0 | 1 |
| P3 | 0 | 1 | 0 | 0 | P11 | 0 | 0 | 1 | 1 |
| P4 | 0 | 0 | 1 | 0 | P12 | 1 | 1 | 1 | 0 |
| P5 | 0 | 0 | 0 | 1 | P13 | 1 | 1 | 0 | 1 |
| P6 | 1 | 1 | 0 | 0 | P14 | 1 | 0 | 1 | 1 |
| P7 | 1 | 0 | 1 | 0 | P15 | 0 | 1 | 1 | 1 |
| P8 | 1 | 0 | 0 | 1 | P16 | 1 | 1 | 1 | 1 |

White areas in Table 5 show the profiles included in the test. As shown in Table 5, there are questions related to profiles P2, P3, P4, P5, P7, and P9. Since 42 questions represent profile P2 and the number of questions has to be reduced by half, the first 21 questions of this group, the first 15 questions of the 30 questions representing profile P3, the first 15 questions of the 31 questions representing profile P4, the first 6 questions of the 12 questions representing profile P5, the first 2 questions of the 4 questions representing profile P7 and the single question representing profile P9 have been included, giving a total of 60 questions.

The Q matrix for The Basic Medical Sciences Test created based on Miller was investigated. Questions related to each attribute profile were identified and ranked. Sixty questions were found to represent the P2 profile (1,0,0,0,0), and the last 30 questions were taken from here since the number of questions would be halved. In addition, the first 19 of the 39 questions representing the P3 profile (1,1,0,0), the first 10 of the 20 questions representing the P4 profile (1,1,1,0) and the single question representing the P5 profile (1,1,1,1) were also included, giving a total of 60 questions.

Since the profiles created based on Miller are in a linear hierarchy, the attribute profiles are determined as in the Table 6.

**Table 6: Five Attribute Profiles for The Basic Medical Sciences Test Created Based on Miller**

| | | | | |
|---|---|---|---|---|
| P1 | 0 | 0 | 0 | 0 |
| P2 | 1 | 0 | 0 | 0 |
| P3 | 1 | 1 | 0 | 0 |
| P4 | 1 | 1 | 1 | 0 |
| P5 | 1 | 1 | 1 | 1 |

The Q matrix for The Clinical Medical Sciences Test created based on Miller was investigated. Questions related to each attribute profile were identified and ranked. Twenty-one questions were found to represent the P2 profile (1,0,0,0,0), and the last 11 questions were taken from here since the number of questions would be halved. In addition, the first 5 of the 11 questions representing the P3 profile (1,1,0,0), the last 36 of the 72 questions representing the P4 profile (1,1,1,0) and the first 8 of the 16 questions representing the P5 profile (1,1,1,1) were also included, giving a total of 60 questions.

The Fields and Miller Q matrices for the Basic Medical Sciences Test and Clinical Medical Sciences Test, created based on expert opinions, are not attached in the article to keep their length reasonable. For those interested in accessing the Q matrices, it is sufficient to email the author for further information.

## Data Analysis

In the data analysis, the error values, s and g parameters and classification accuracies of the high-order DINA model and the high-order DINO model, which have a higher-level structure between attributes, and the independent DINA and DINO models, which do not consider any hierarchical structure, were compared under different conditions. R 4.2.2 program and the G-DINA, openxlsx, doParallel, readxl packages were used for data analysis. In addition, Mixed Factorial ANOVA was used to compare the RMSEA values obtained for the models, the mean s and g parameters for the items and the classification accuracy of the test according to the simulation conditions. In addition, common effect plots were used to investigate the statistically significant four-way and three-way interactions identified by factorial ANOVA.

## FINDINGS

The assumption of normality was investigated before the analyzes were carried out. Kurtosis and skewness coefficients were taken into account in examining the condition of showing the normality assumption of the distribution of the parameters investigated at the levels of the simulation conditions. As a result of the examinations, it was determined that the values for most of the parameters remained in the range of -1 and 1. Parameter distributions showing small deviations from normality were ignored in the analysis due to the fact that the number of observations per cell in the analysis to be made is more than 50.

### 1. The Analysis Results of The Basic Medical Sciences Test

#### 1. a) The RMSEA values obtained for the model fit

The mean of the RMSEA values estimated according to the analysis models and simulation conditions for Basic Medical Sciences are presented in Table 7.

**Table 7: RMSEA for Model Fit Under Different Simulation Conditions for Basic Medical Sciences Test**

| | | RMSEA Values | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | DINA | | DINO | | HO-DINA | | HO-DINO | |
| | | Number of items | Number of items | Number of items | Number of items | Number of items | Number of items | Number of items | Number of items |
| Q matrix | N | 60 | 120 | 60 | 120 | 60 | 120 | 60 | 120 |
| Fields | 500 | .035 | .030 | .030 | .023 | .037 | .031 | .029 | .023 |
| | 2000 | .036 | .033 | .032 | .024 | .037 | .033 | .031 | .024 |
| | 5000 | .036 | .034 | .032 | .026 | .037 | .034 | .032 | .026 |
| Miller | 500 | .066 | .059 | .032 | .027 | .067 | .059 | .032 | .027 |
| | 2000 | .064 | .067 | .031 | .029 | .065 | .067 | .031 | .029 |
| | 5000 | .066 | .065 | .033 | .030 | .067 | .065 | .033 | .030 |

Table 7 shows that the RMSEA values obtained under different simulation conditions for the Basic Medical Sciences Test are less than 0.08. In other words, all models analyzed under different simulation conditions show a good fit. It was determined that the RMSEA values obtained for the attributes determined based on Fields in the Q matrix are lower than the RMSEA values obtained for the attributes determined based on Miller. Therefore, it can be said that determining the attributes based on Fields in the Q matrix provides better results than determining them based on Miller. When analyzing for sample size, it was found that the differentiation of the sample did not change the model fit. Regarding the number of items, it was found that the RMSEA values obtained for 120 items are lower than the RMSEA values obtained for 60 items, indicating that the models fit better for 120 items. Mixed factorial ANOVA was carried out to determine whether the differences showed statistically significant differences. The results are shown in Table 8.

**Table 8: Mixed Factorial ANOVA Results for RMSEA Values of Models in The Basic Medical Sciences Test**

| Source | Sum of Squares | df | Mean Square | F | p | $\eta_p^2$ |
|---|---|---|---|---|---|---|
| Method | 0.048 | 1 | 0.048 | 3489.603 | 0.000 | 0.856 |
| Method * Qmatrix | 0.023 | 1 | 0.023 | 1701.417 | 0.000 | 0.743 |
| Method * Sample | 0.000 | 2 | 0.000 | 2.146 | 0.118 | 0.007 |
| Method * Item | 0.000 | 1 | 0.000 | 11.745 | 0.001 | 0.020 |
| Method * Qmatrix * Sample | 0.000 | 2 | 0.000 | 0.755 | 0.471 | 0.003 |
| Method * Qmatrix * Item | 0.000 | 1 | 0.000 | 1.097 | 0.295 | 0.002 |
| Method * Sample * Item | 0.000 | 2 | 0.000 | 4.906 | 0.008 | 0.016 |
| Method * Qmatrix * Sample * Item | 0.000 | 2 | 0.000 | 1.875 | 0.154 | 0.006 |
| Error | 0.008 | 588 | 0.000 | | | |

*p<.01;

Table 8 shows that the interaction between “number of items, sample size and method” is statistically significant (p<0.01) for the Basic Medical Sciences Test. The common effect plots are presented in Figure 1 to investigate how the RMSEA values differ under different simulation conditions for the Basic Medical Sciences Test. Looking at Figure 1, it can be seen that the RMSEA values for 120 items are lower than 60 items in all models; furthermore, the RMSEA values obtained for DINA and HO-DINA are similar to each other, and the RMSEA values obtained for DINO and HO-DINO are similar to each other. It is also seen that the results are very similar for the samples. It is observed that the RMSEA values for DINA and HO-DINA are higher than those for DINO and HO-DINO. Therefore, it can be noted that DINO and HO-DINO models give better results. For the Q-matrix determined by Fields, the

RMSEA values are similar for all models, whereas, for the Q-matrix determined by Miller's, the RMSEA values for DINO and HO-DINO are lower than those for DINA and HO-DINA. In addition, it is revealed that the RMSEA values obtained within the models of DINO and HO-DINO for the Q matrix determined based on Miller are very similar to the RMSEA values obtained for the Q matrix determined based on the Fields. Considering that the RMSEA values given by the models for the Q matrix determined based on the Fields are more consistent with each other and lower than the RMSEA values obtained for the Q matrix determined based on Miller, it can be interpreted that it is more advantageous to use the Q matrix determined based on the Fields for the TUS exam.

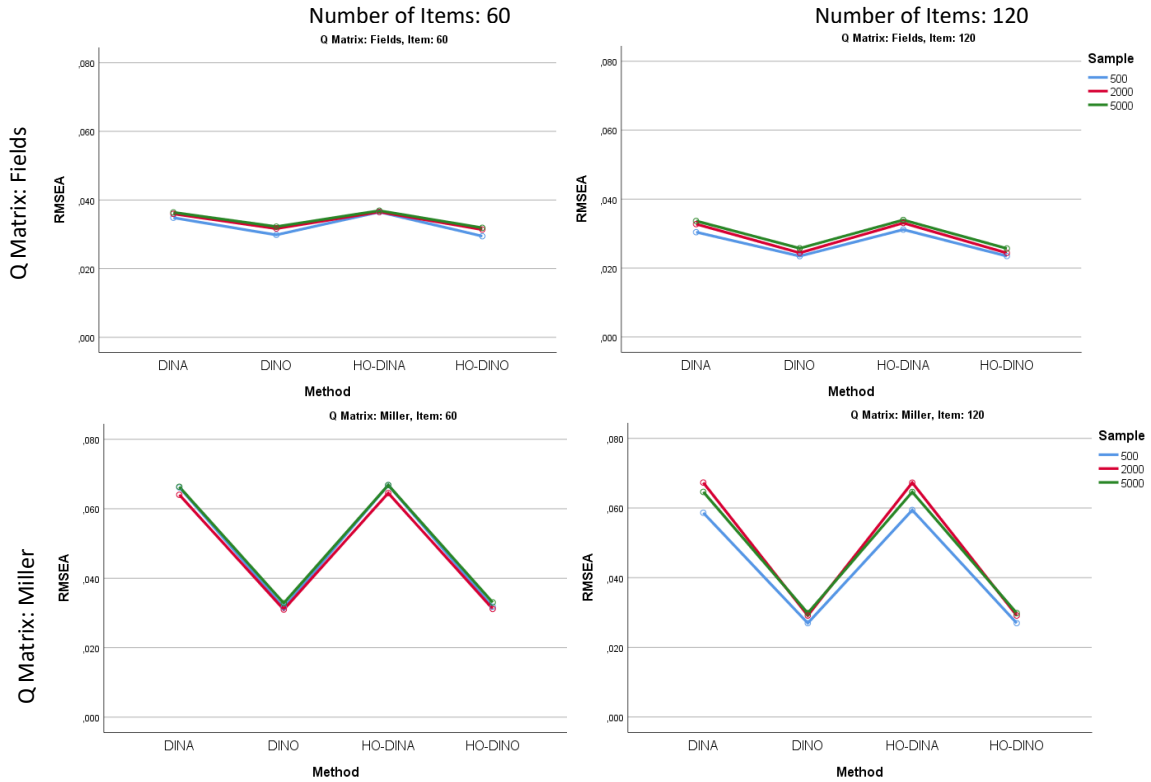


**Figure 1: RMSEA Values for The Basic Medical Sciences Test Under Simulation Conditions.**

**1. b) g and s item parameter estimates**

The mean values of the estimated g and s item parameter values for the Basic Medical Sciences Test according to the analysis models and simulation conditions are presented in Table 9.

**Table 9: The Mean Values of s and g Item Parameters for The Basic Medical Sciences Test Under Different Simulation Conditions.**

| | | | DINA | | DINO | | HO-DINA | | HO-DINO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Number of Items | | Number of Items | | Number of Items | | Number of Items | |
| | Q matrix | N | 60 | 120 | 60 | 120 | 60 | 120 | 60 | 120 |
| s parameters | Fields | 500 | .176 | .121 | .208 | .208 | .177 | .122 | .207 | .208 |
| | | 2000 | .170 | .124 | .204 | .207 | .168 | .124 | .204 | .207 |
| | | 5000 | .168 | .126 | .204 | .210 | .167 | .126 | .204 | .210 |
| | Miller | 500 | .158 | .149 | .234 | .250 | .158 | .150 | .234 | .250 |
| | | 2000 | .165 | .164 | .242 | .249 | .166 | .165 | .242 | .249 |
| | | 5000 | .163 | .163 | .237 | .250 | .163 | .163 | .237 | .250 |
| g parameters | Fields | 500 | .207 | .223 | .109 | .102 | .206 | .222 | .110 | .102 |
| | | 2000 | .211 | .220 | .111 | .102 | .212 | .220 | .111 | .102 |
| | | 5000 | .212 | .216 | .115 | .102 | .212 | .217 | .115 | .102 |
| | Miller | 500 | .216 | .210 | .121 | .108 | .216 | .209 | .121 | .109 |
| | | 2000 | .209 | .208 | .117 | .109 | .209 | .208 | .116 | .109 |
| | | 5000 | .211 | .207 | .120 | .107 | .211 | .207 | .120 | .107 |

If Table 9 is investigated, it can be seen that the g parameter values obtained for the Basic Medical Sciences Test under different simulation conditions are between 0.102 and 0.223, and the s parameters are between 0.121 and 0.250. In other words, since the g and s parameter values analyzed under different simulation conditions are low, all models show a good fit. It is found that the g and s parameter values obtained for the attributes determined based on the Fields in the Q matrix are lower than the g and s parameter values obtained for the attributes determined based on the Miller. Therefore, it can be said that determining attributes

based on the Fields in the Q-matrix gives better results than determining them based on Miller. Regarding the sample size, it is found that there is not much difference in the s and g parameters with the differentiation of the sample, so it does not affect the model fit. Concerning the number of items, it is noted that the g parameter values obtained with 60 items are similar to that obtained with 120 items. The g and s parameter values for the Fileds and Miller indicate that the model fit of the test is sufficient. Mixed Factorial ANOVA was conducted to investigate whether the differences showed statistically significant. The results are presented in Table 10.

**Table 10: Mixed Factorial ANOVA Results of g and s parameters for Models in The Basic Medical Sciences Test**

| | Source | Sum of Squares | df | Mean Square | F | p | $\eta_p^2$ |
|---|---|---|---|---|---|---|---|
| g parameters | Method | 1.250 | 1 | 1.250 | 20725.234 | 0.000 | 0.972 |
| | Method * Qmatrix | 0.004 | 1 | 0.004 | 67.712 | 0.000 | 0.103 |
| | Method * Sample | 0.000 | 2 | 0.000 | 1.438 | 0.238 | 0.005 |
| | Method * Item | 0.005 | 1 | 0.005 | 90.396 | 0.000 | 0.133 |
| | Method * Qmatrix* Sample | 0.000 | 2 | 0.000 | 0.471 | 0.625 | 0.002 |
| | Method * Qmatrix* Item | 0.001 | 1 | 0.001 | 19.524 | 0.000 | 0.032 |
| | Method * Sample * Item | 0.000 | 2 | 0.000 | 0.187 | 0.829 | 0.001 |
| | Method * Qmatrix* Sample * Item | 0.000 | 2 | 0.000 | 0.522 | 0.594 | 0.002 |
| | Error | 0.035 | 588 | 0.000 | | | |
| s parameters | Method | 0.612 | 1 | 0.612 | 4525.787 | 0.000 | 0.885 |
| | Method * Qmatrix | 0.018 | 1 | 0.018 | 135.088 | 0.000 | 0.187 |
| | Method * Sample | 0.001 | 2 | 0.000 | 1.912 | 0.149 | 0.006 |
| | Method * Item | 0.033 | 1 | 0.033 | 241.146 | 0.000 | 0.291 |
| | Method * Qmatrix * Sample | 0.000 | 2 | 0.000 | 1.331 | 0.265 | 0.005 |
| | Method * Qmatrix * Item | 0.010 | 1 | 0.010 | 70.374 | 0.000 | 0.107 |
| | Method * Sample * Item | 0.001 | 2 | 0.000 | 2.569 | 0.077 | 0.009 |
| | Method * Qmatrix * Sample * Item | 0.000 | 2 | 0.000 | 0.729 | 0.483 | 0.002 |
| | Error | 0.080 | 588 | 0.000 | | | |

*p<.01;

Table 10 shows that the interaction between the "number of items, Q matrix and method" for g parameters in the Basic Medical Sciences Test is statistically significant (p<0.01). The common effect plots are presented in Figure 2 to investigate how the g item parameters differ under different simulation conditions for the Basic Medical Sciences Test. Figure 2 shows that for all models, the g item parameters are similar for both 120 and 60 items; furthermore, g parameter values obtained for DINA and HO-DINA are similar to each other and g parameter values obtained for DINO and HO-DINO are similar to each other. In addition, the results for different samples are very similar. The g parameters obtained for the DINO and HO-DINO models are observed to be lower than those obtained for the DINA and HO-DINA models. Furthermore, while considering the Q matrix determined based on Fields and Miller, the g parameters obtained for all models are similar, and the g parameters obtained for DINO and HO-DINO models are lower than those obtained for DINA and HO-DINA models for the Q matrix determined based on Fields and Miller.

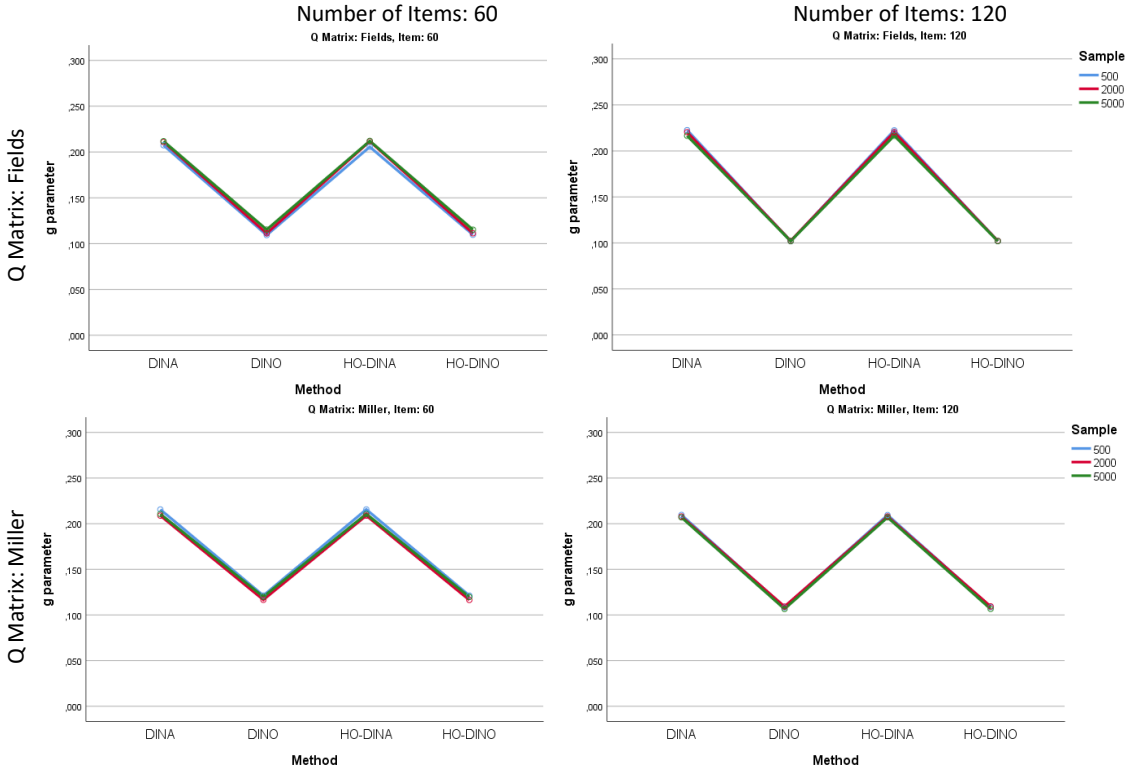


**Figure 2: g Parameter Values for The Basic Medical Sciences Test under Different Simulation Conditions**

Table 10 shows that the interaction between "number of items, Q-matrix and method" is statistically significant (p<0.01) for the s item parameters in the Basic Medical Sciences Test. The common effect plots are presented in Figure 3 to investigate how the s item parameters differ under different simulation conditions for the Basic Medical Sciences Test. Looking at Figure 3, it can be seen that for all models, the s item parameters are similar for both 120 items and 60 items. Additionally, the results for different samples are very similar. Moreover, when considering the Q matrix determined based on Fields and Miller, it is observed that the s item parameters obtained for all models are very similar.

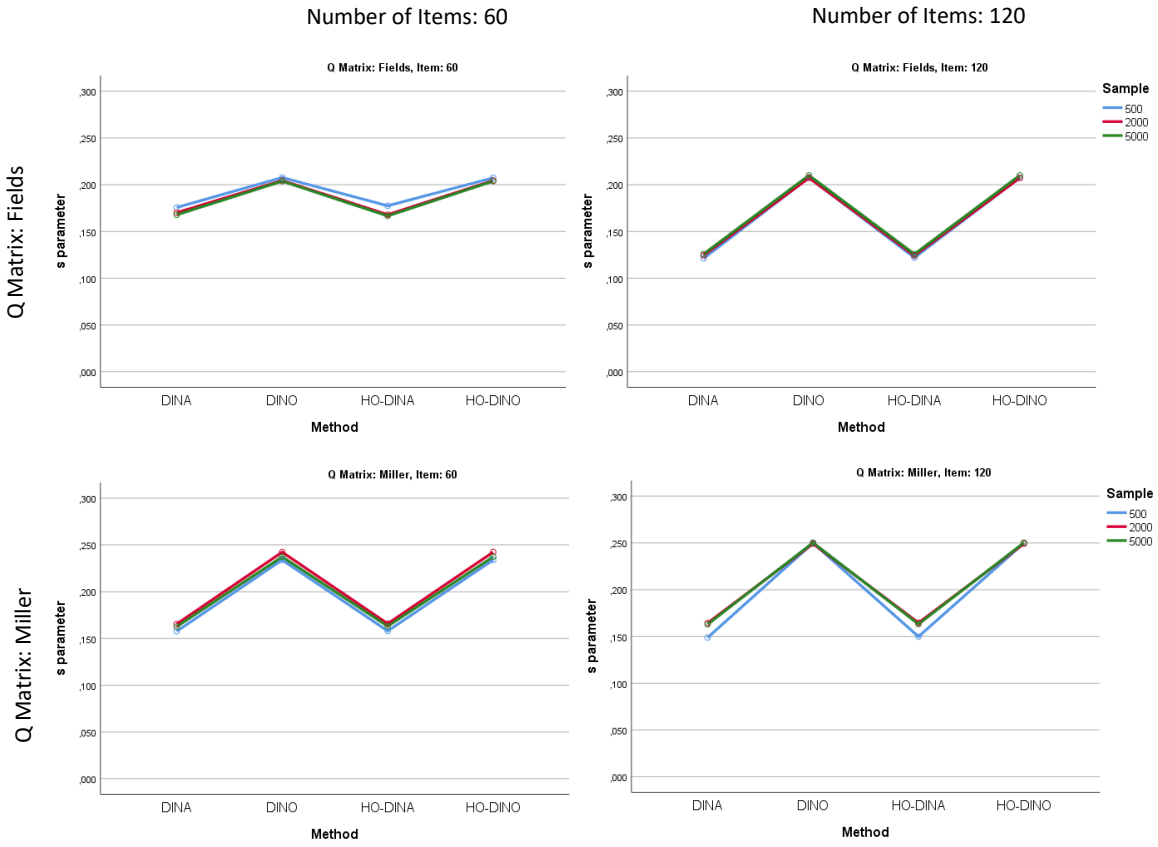


**Figure 3: s Parameter Values for the Basic Medical Sciences Test under Different Simulation Conditions**

**1. c) Classifying accuracy**

The mean values of the estimated classification accuracies for The Basic Medical Sciences Test according to the analysis models and simulation conditions are presented in Table 11.

**Table 11: The Mean Values of Correct Classification Rates for The Basic Medical Sciences Test Under Different Simulation Conditions**

| | | DINA classification rates | | DINO classification rates | | HO-DINA classification rates | | HO-DINO classification rates | |
|---|---|---|---|---|---|---|---|---|---|
| | | Number of Items | | Number of Items | | Number of Items | | Number of Items | |
| Q matrix | N | 60 | 120 | 60 | 120 | 60 | 120 | 60 | 120 |
| Fields | 500 | .837 | .921 | .914 | .970 | .806 | .907 | .898 | .965 |
| | 2000 | .815 | .903 | .893 | .966 | .800 | .896 | .888 | .964 |
| | 5000 | .806 | .904 | .886 | .964 | .798 | .900 | .883 | .963 |
| Miller | 500 | .282 | .280 | .528 | .551 | .404 | .400 | .562 | .565 |
| | 2000 | .277 | .279 | .533 | .561 | .386 | .386 | .572 | .577 |
| | 5000 | .276 | .289 | .538 | .557 | .384 | .385 | .569 | .570 |

Looking at Table 11, it can be seen that the values of the classification rates obtained for the attributes based on the Fields in the Q-matrix under different simulation conditions in The Basic Medical Sciences Test are between 0.798 and 0.970, and the values of the classification rates obtained for the attributes based on Miller are between 0.276 and 0.577. Accordingly, it can be interpreted that the attribute determination based on Fields in the Q-matrix gives better results than the attribute determination based on Miller. Regarding the sample size, it is noted that the variation in the sample did not affect the classification rates. On the other hand, in terms of the number of items, it was found that the classification rates obtained for 120 items were higher than the classification rates obtained for 60 items. Mixed factorial ANOVA was conducted to investigate whether the determined differences showed statistically significant. The results are presented in Table 12.

**Table 12: Mixed Factorial ANOVA Results of Classification Rates of Models in The Basic Medical Sciences Test**

| Source | Sum of Squares | df | Mean Square | F | p | $\eta_p^2$ |
|---|---|---|---|---|---|---|
| Method | 5.026 | 1 | 5.026 | 7586.985 | 0.000 | 0.928 |
| Method * Qmatrix | 2.747 | 1 | 2.747 | 4146.263 | 0.000 | 0.876 |
| Method * Sample | 0.006 | 2 | 0.003 | 4.209 | 0.015 | 0.014 |
| Method * Item | 0.004 | 1 | 0.004 | 5.949 | 0.015 | 0.010 |
| Method * Qmatrix * Sample | 0.005 | 2 | 0.003 | 4.142 | 0.016 | 0.014 |
| Method * Qmatrix * Item | 0.000 | 1 | 0.000 | 0.434 | 0.511 | 0.001 |
| Method * Sample * Item | 0.002 | 2 | 0.001 | 1.224 | 0.295 | 0.004 |
| Method * Qmatrix * Sample * Item | 0.000 | 2 | 0.000 | 0.076 | 0.927 | 0.000 |
| Error | 0.390 | 588 | 0.001 | | | |

*p<.01;

Table 12 indicates that there is a statistically significant interaction between the "Q-matrix and method" in the Basic Medical Sciences Test ($p < 0.01$). The common effect plots are presented in Figure 4 to investigate how the classification accuracies differ under different simulation conditions for The Basic Medical Sciences Test. Looking at Figure 4, it can be seen that for all models, the classification accuracies for 120 items are higher than those for 60 items based on the Fields in the Q-matrix. Moreover, when comparing the models based on Miller, the classification accuracies for both 120 and 60 items are similar. The classification accuracies obtained for the DINA and HO-DINA models are similar to those obtained for the DINO and HO-DINO models. In addition, the results for different samples are very similar. Furthermore, the classification accuracies obtained for the DINO and HO-DINO models are higher than those for the DINA and HO-DINA models. For the Q-matrix determined by Fields, the classification accuracy values are similar for all models, whereas, for the Q-matrix determined by Miller's, the classification accuracy values for the DINO and HO-DINO models are higher than those for the DINA and HO-DINA models. In addition, it can be seen that the classification accuracies obtained by all models for the Q-matrix determined by Fields are higher than the classification accuracies obtained for the Q-matrix determined by Miller. Considering that the classification accuracy values given by the models for the Q matrix determined based on the Fields are more consistent with each other and higher than the classification accuracy values obtained for the Q matrix determined based on Miller, it can be interpreted that it is more advantageous to use the Q matrix determined based on the Fields for the TUS exam.

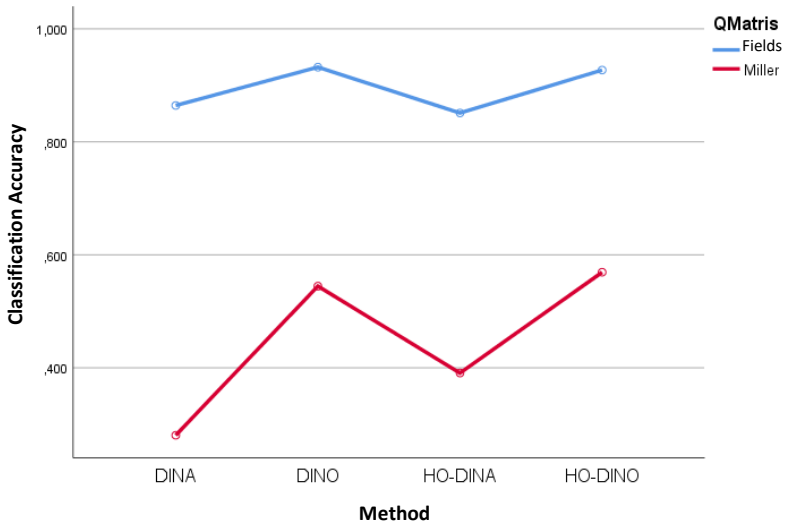


**Figure 4: Classification Accuracy Values for The Basic Medical Sciences Test under Different Simulation Conditions**

## 2. Analysis Results of The Clinical Medical Sciences Test

### 2. a) The RMSEA values obtained for the model fit

The mean of the RMSEA values estimated according to the analysis models and simulation conditions for Clinical Medical Sciences are presented in Table 13.

**Table 13: RMSEA for Model Fit Under Different Simulation Conditions for The Clinical Medical Sciences Test**

| | | RMSEA Values | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | DINA | | DINO | | HO-DINA | | HO-DINO | |
| | | Number of items | | Number of items | | Number of items | | Number of items | |
| Q matrix | N | 60 | 120 | 60 | 120 | 60 | 120 | 60 | 120 |
| Fields | 500 | .005 | .006 | .003 | .004 | .005 | .007 | .003 | .004 |
| | 2000 | .007 | .007 | .004 | .005 | .007 | .007 | .004 | .005 |
| | 5000 | .006 | .007 | .004 | .005 | .007 | .007 | .004 | .005 |
| Miller | 500 | .074 | .079 | .049 | .047 | .075 | .080 | .049 | .047 |
| | 2000 | .073 | .078 | .049 | .048 | .073 | .078 | .049 | .048 |
| | 5000 | .070 | .073 | .050 | .048 | .071 | .074 | .050 | .048 |

Table 13 shows that the RMSEA values obtained under different simulation conditions for The Clinical Medical Sciences Test are less than 0.08. In other words, all models analyzed under different simulation conditions show a good fit. It was determined that the RMSEA values obtained for the attributes determined based on Fields in the Q matrix are lower than the RMSEA values obtained for the attributes determined based on Miller. Therefore, it can be said that determining the attributes based on Fields in the Q matrix provides better results than determining them based on Miller. When analyzing for sample size, it was found that the differentiation of the sample did not change the model fit. Regarding the number of items, it was found that the RMSEA values obtained for 60 items are lower than the RMSEA values obtained for 120 items, indicating that the models fit better for 60 items. Mixed Factorial ANOVA was carried out to determine whether the differences showed statistically significant differences. The results are shown in Table 14.

**Table 14: Mixed Factorial ANOVA Results for RMSEA Values of Models in The Clinical Medical Sciences Test**

| Source | Sum of Squares | df | Mean Square | F | p | $\eta_p^2$ |
|---|---|---|---|---|---|---|
| Method | 0.023 | 1 | 0.023 | 2272.335 | 0.000 | 0.794 |
| Method * Qmatrix | 0.017 | 1 | 0.017 | 1662.000 | 0.000 | 0.739 |
| Method * Sample | 0.000 | 2 | 0.000 | 8.389 | 0.000 | 0.028 |
| Method * Item | 0.000 | 1 | 0.000 | 28.764 | 0.000 | 0.047 |
| Method * Qmatrix * Sample | 0.000 | 2 | 0.000 | 10.026 | 0.000 | 0.033 |
| Method * Qmatrix * Item | 0.000 | 1 | 0.000 | 22.164 | 0.000 | 0.036 |
| Method * Sample * Item | 0.000 | 2 | 0.000 | 0.385 | 0.681 | 0.001 |
| Method * Qmatrix * Sample * Item | 0.000 | 2 | 0.000 | 0.177 | 0.838 | 0.001 |
| Error | 0.006 | 588 | 0.000 | | | |

*p<.01;

Table 14 shows that the interaction between “number of items, Q matrix and method” and “sample size, Q matrix and method” are statistically significant (p<0.01) for The Clinical Medical Sciences Test. The common effect plots are presented in Figure 5 to investigate how the RMSEA values differ under different simulation conditions for The Clinical Medical Sciences Test. Looking at Figure 5, it can be seen that for all models, the RMSEA values are similar for both 120 and 60 items; furthermore, the RMSEA values obtained for DINA and HO-DINA are similar to each other, and the RMSEA values obtained for DINO and HO-DINO are similar to each other. It is also seen that the results are very similar for the samples. It is observed that the RMSEA values for DINA and HO-DINA are higher than those for DINO and HO-DINO. Therefore, it can be noted that DINO and HO-DINO models give better results. For the Q-matrix determined by Fields, the RMSEA values are similar for all models, whereas, for the Q-matrix determined by Miller's, the RMSEA values for DINO and HO-DINO are lower than those for DINA and HO-DINA. Considering the more consistent results between the models and lower RMSEA values obtained for the Q matrix determined by Fields, it can be concluded that the Q matrix based on Fields is more advantageous for the TUS exam.

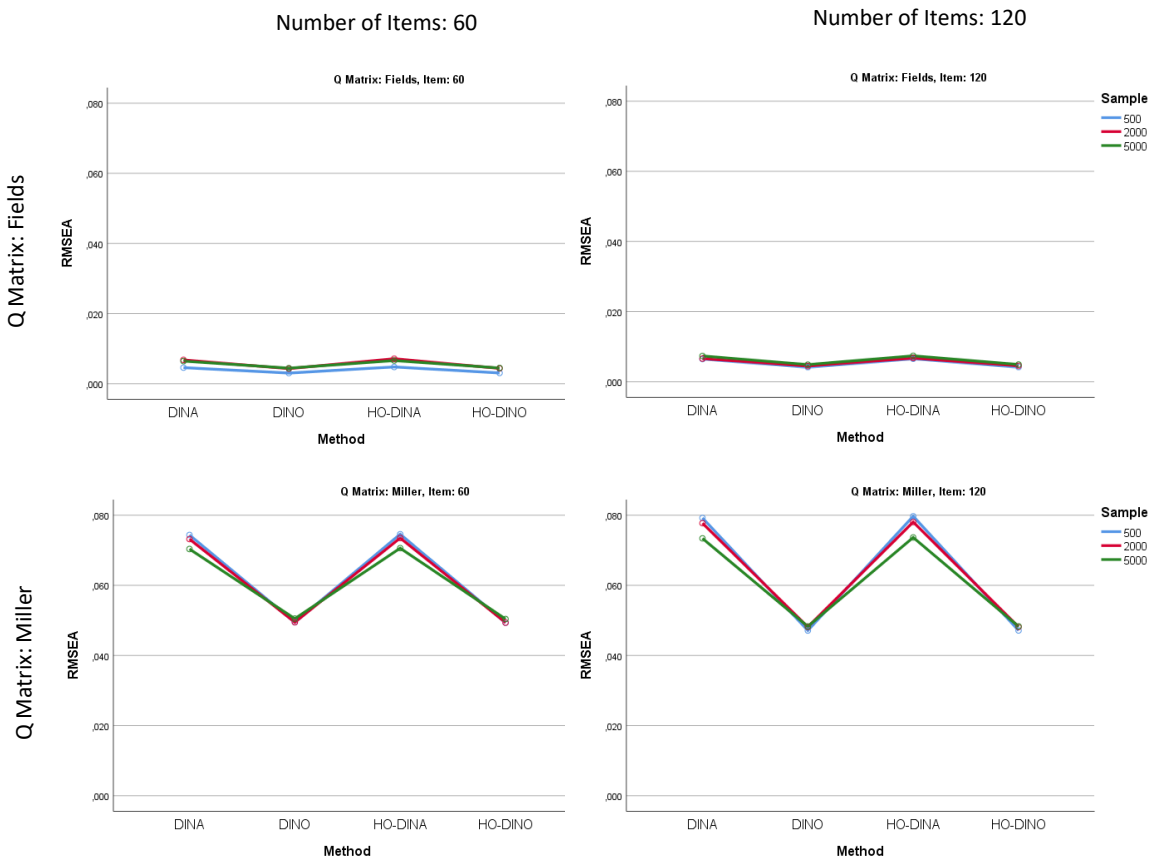


**Figure 5: RMSEA Values for The Clinical Medical Sciences Test under Simulation Conditions.**

**2. b) g and s item parameter estimates**

The mean values of the estimated g and s item parameter values for The Clinical Medical Sciences Test according to the analysis models and simulation conditions are presented in Table 15.

**Table 15: The Mean Values of s and g Item Parameters for The Clinical Medical Sciences Test under Different Simulation Conditions.**

| | | | DINA | | DINO | | HO-DINA | | HO-DINO | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Number of Items | | Number of Items | | Number of Items | | Number of Items | |
| | Q matrix | N | 60 | 120 | 60 | 120 | 60 | 120 | 60 | 120 |
| s parameters | Fields | 500 | .099 | .100 | .114 | .112 | .099 | .100 | .114 | .112 |
| | | 2000 | .100 | .100 | .114 | .112 | .100 | .100 | .114 | .112 |
| | | 5000 | .100 | .100 | .115 | .113 | .100 | .100 | .115 | .113 |
| | Miller | 500 | .270 | .166 | .298 | .301 | .270 | .166 | .298 | .301 |
| | | 2000 | .282 | .191 | .296 | .305 | .282 | .191 | .296 | .305 |
| | | 5000 | .293 | .220 | .297 | .307 | .293 | .220 | .297 | .307 |
| g parameters | Fields | 500 | .113 | .111 | .100 | .100 | .113 | .111 | .100 | .100 |
| | | 2000 | .114 | .111 | .100 | .100 | .114 | .111 | .100 | .100 |
| | | 5000 | .113 | .111 | .100 | .100 | .113 | .111 | .100 | .100 |
| | Miller | 500 | .311 | .399 | .137 | .136 | .311 | .399 | .137 | .136 |
| | | 2000 | .304 | .379 | .139 | .135 | .304 | .379 | .139 | .135 |
| | | 5000 | .303 | .355 | .138 | .138 | .302 | .355 | .138 | .138 |

If Table 15 is investigated, it can be seen that the g parameter values obtained for The Clinical Medical Sciences Test under different simulation conditions are between 0.100 and 0.399, and the s parameters are between 0.099 and 0.307. In other words, since the g and s parameter values analyzed under different simulation conditions are low, all models show a good fit. It is found that the g and s parameter values obtained for the attributes determined based on the Fields in the Q matrix are lower than the g and s parameter values obtained for the attributes determined based on the Miller. Therefore, it can be said that determining attributes based on the Fields in the Q-matrix gives better results than determining them based on Miller. Regarding the sample size, it is found that there is not much difference in the s and g parameters with the differentiation of the sample, so it does not affect the model fit. Concerning the number of items, it is noted that the g and s parameter values obtained with 60 items are similar to that obtained with 120 items. The g and s parameter values for the Fields and Miller indicate that the model fit of the

test is sufficient. Mixed Factorial ANOVA was conducted to investigate whether the determined differences showed statistically significant. The results are presented in Table 16.

**Table 16: Mixed Factorial ANOVA Results of g and s Parameters for Models in The Clinical Medical Sciences Test**

| | Source | Sum of Squares | df | Mean Square | F | p | $\eta_p^2$ |
|---|---|---|---|---|---|---|---|
| g parameters | Method | 1.403 | 1 | 1.403 | 5032.676 | 0.000 | 0.895 |
| | Method * Qmatrix | 1.114 | 1 | 1.114 | 3995.073 | 0.000 | 0.872 |
| | Method * Sample | 0.004 | 2 | 0.002 | 6.562 | 0.002 | 0.022 |
| | Method * Item | 0.038 | 1 | 0.038 | 135.127 | 0.000 | 0.187 |
| | Method * Qmatrix* Sample | 0.004 | 2 | 0.002 | 6.748 | 0.001 | 0.022 |
| | Method * Qmatrix* Item | 0.042 | 1 | 0.042 | 151.875 | 0.000 | 0.205 |
| | Method * Sample * Item | 0.002 | 2 | 0.001 | 3.303 | 0.037 | 0.011 |
| | Method * Qmatrix* Sample * Item | 0.002 | 2 | 0.001 | 3.258 | 0.039 | 0.011 |
| | Error | 0.164 | 588 | 0.000 | | | |
| s parameters | Method | 0.178 | 1 | 0.178 | 433.566 | 0.000 | 0.424 |
| | Method * Qmatrix | 0.075 | 1 | 0.075 | 181.597 | 0.000 | 0.236 |
| | Method * Sample | 0.006 | 2 | 0.003 | 7.454 | 0.001 | 0.025 |
| | Method * Item | 0.067 | 1 | 0.067 | 162.805 | 0.000 | 0.217 |
| | Method * Qmatrix * Sample | 0.006 | 2 | 0.003 | 7.543 | 0.001 | 0.025 |
| | Method * Qmatrix * Item | 0.073 | 1 | 0.073 | 177.461 | 0.000 | 0.232 |
| | Method * Sample * Item | 0.001 | 2 | 0.000 | 0.955 | 0.385 | 0.003 |
| | Method * Qmatrix * Sample * Item | 0.001 | 2 | 0.000 | 1.031 | 0.357 | 0.003 |
| | Error | 0.242 | 588 | 0.000 | | | |

*p<.01;

Table 16 shows that the interaction between “number of items, Q matrix and method” and “sample size, Q matrix and method” for g parameters in The Clinical Medical Sciences Test are statistically significant (p<0.01). To investigate how the g item parameters differ under different simulation conditions for The Clinical Medical Sciences Test, common effect plots are presented in Figure 6. Figure 6 shows that for all models, the g parameters are similar for both 120 and 60 items, furthermore g parameter values obtained for DINA and HO-DINA are similar to each other and g parameter values obtained for DINO and HO-DINO are similar to each other. In addition, the results for different samples are very similar. The g parameters obtained for the DINO and HO-DINO models are observed to be lower than those obtained for the DINA and HO-DINA models. Furthermore, while considering the Q matrix determined based on Fields, the g item parameters obtained for all models are similar; the g item parameters obtained for DINO and HO-DINO models are lower than those obtained for DINA and HO-DINA models for the Q matrix determined based on Miller.

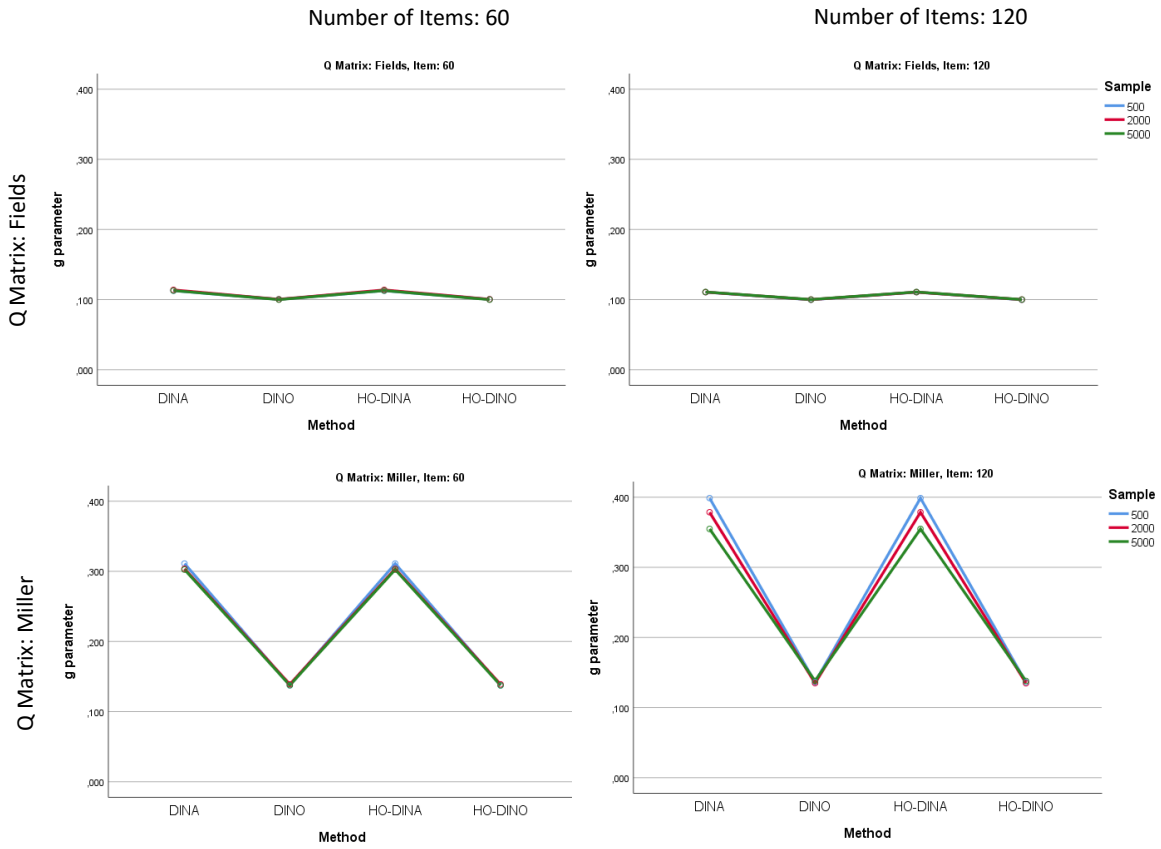


**Figure 6: g Parameter Values for The Clinical Medical Sciences Test under Different Simulation Conditions**

Table 16 shows that the interaction between "number of items, Q matrix and method" and "sample size, Q matrix and method" are statistically significant (p<0.01) for the s item parameters in The Clinical Medical Sciences Test. To investigate how the s item parameters differ under different simulation conditions for The Clinical Medical Sciences Test, common effect plots are presented in Figure 7. Looking at Figure 7 it can be seen that for all models considering the Q matrix determined based on Fields, the s parameters are similar for both 120 items and 60 items, furthermore the s item parameters for 120 items obtained for DINO and HO-DINO models are higher than those obtained for DINA and HO-DINA models in the Q matrix determined based on Miller. In addition, the results for different samples are very similar. In the Q matrix determined based on Miller, it was found that the s item parameters for 120 items obtained for DINA and HO-DINA models are lower than the s item parameters for 60 items obtained for DINA and HO-DINA models. In addition, it was determined that the s parameter values obtained for 60 and 120 items for the DINO and HO-DINO models are similar in the Q matrix determined based on Miller.

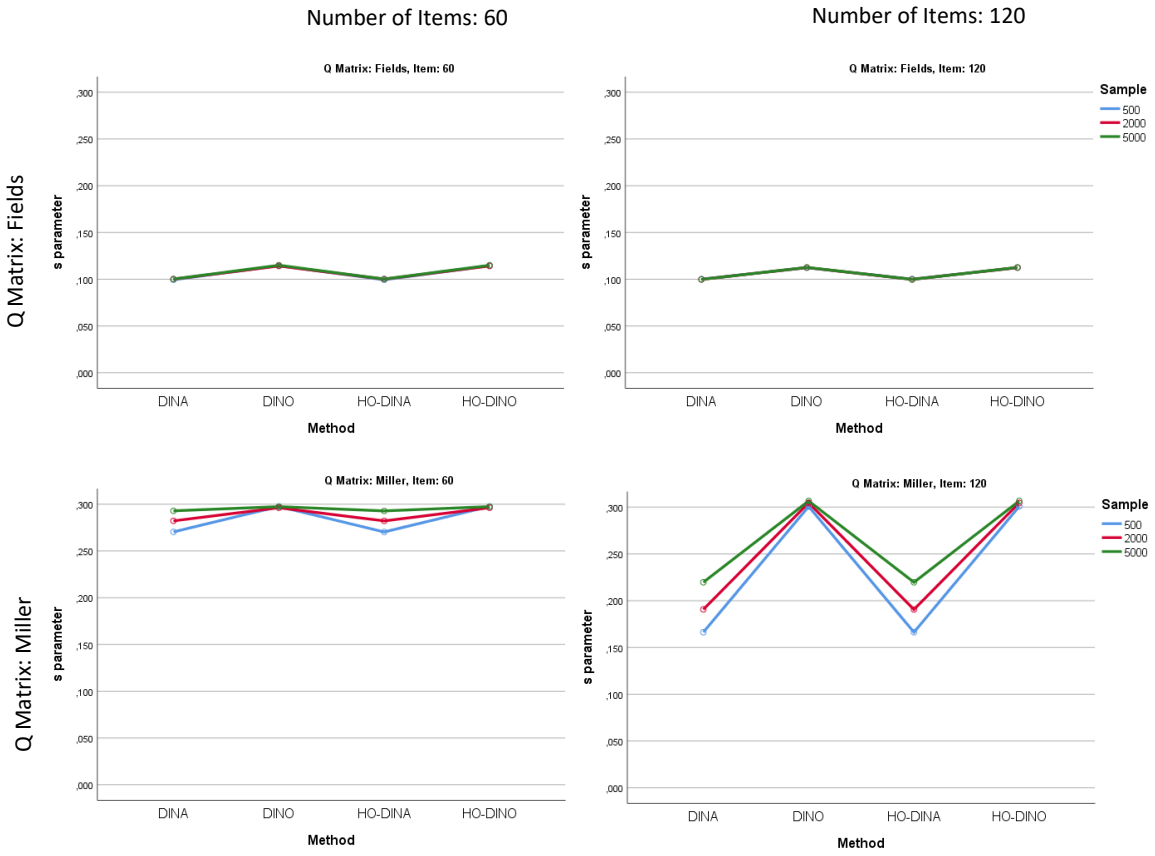


**Figure 7: s Parameter Values for The Clinical Medical Sciences Test under Different Simulation Conditions**

**2. c) Classifying accuracy**

The mean values of the estimated classification accuracies for The Clinical Medical Sciences Test according to the analysis models and simulation conditions are presented in Table 17.

**Table 17: The Mean Values of Correct Classification Rates for The Clinical Medical Sciences Test under Different Simulation Conditions**

| | | DINA classification rates | | DINO classification rates | | HO-DINA classification rates | | HO-DINO classification rates | |
|---|---|---|---|---|---|---|---|---|---|
| | | Number of Items | | Number of Items | | Number of Items | | Number of Items | |
| Q matrix | N | 60 | 120 | 60 | 120 | 60 | 120 | 60 | 120 |
| Fields | 500 | .996 | 1.000 | .996 | 1.000 | .996 | 1.000 | .996 | 1.000 |
| | 2000 | .996 | 1.000 | .996 | 1.000 | .996 | 1.000 | .996 | 1.000 |
| | 5000 | .996 | 1.000 | .996 | 1.000 | .996 | 1.000 | .996 | 1.000 |
| Miller | 500 | .538 | .492 | .735 | .807 | .651 | .579 | .671 | .688 |
| | 2000 | .559 | .499 | .715 | .809 | .660 | .583 | .648 | .687 |
| | 5000 | .569 | .519 | .712 | .812 | .664 | .602 | .646 | .679 |

Looking at Table 17, it can be seen that the values of the classification rates obtained for the attributes based on the Fields in the Q-matrix under different simulation conditions in The Clinical Medical Sciences Test are between 0.996 and 1.000, and the values of the classification rates obtained for the attributes based on Miller are between 0.519 and 0.812. Accordingly, it can be interpreted that the attribute determination based on Fields in the Q matrix gives better results than the attribute determination based on Miller. In other words, in the models analyzed under different simulation conditions, it was determined that the Fields were better classified than the Miller. Regarding the sample size, it is noted that the variation in the sample did not affect the classification rates. On the other hand, in terms of the number of items, it was found that for all models considering the Q matrix determined based on Fields, the classification rates obtained for 120 items were higher than the classification rates obtained for 60 items, furthermore the classification rates obtained for 60 items obtained for DINA and HO-DINA models were higher than the classification rates obtained for 120 items obtained for DINO and HO-DINO models in the Q matrix determined based on Miller. Mixed factorial ANOVA was conducted to investigate whether the determined differences showed statistically significant. The results are presented in Table 18.

**Table 18: Mixed Factorial ANOVA Results of Classification Rates of Models in The Clinical Medical Sciences Test**

| Source | Sum of Squares | df | Mean Square | F | p | $\eta_p^2$ |
|---|---|---|---|---|---|---|
| Method | 0.590 | 1 | 0.590 | 526.640 | 0.000 | 0.472 |
| Method * Qmatrix | 0.591 | 1 | 0.591 | 526.991 | 0.000 | 0.473 |
| Method * Sample | 0.015 | 2 | 0.008 | 6.834 | 0.001 | 0.023 |
| Method * Item | 0.014 | 1 | 0.014 | 12.569 | 0.000 | 0.021 |
| Method * Qmatrix * Sample | 0.015 | 2 | 0.008 | 6.871 | 0.001 | 0.023 |
| Method * Qmatrix * Item | 0.014 | 1 | 0.014 | 12.525 | 0.000 | 0.021 |
| Method * Sample * Item | 0.002 | 2 | 0.001 | 0.936 | 0.393 | 0.003 |
| Method * Qmatrix * Sample * Item | 0.002 | 2 | 0.001 | 0.944 | 0.390 | 0.003 |
| Error | 0.659 | 588 | 0.001 | | | |

*p<.01;

Table 18 indicates that there is a statistically significant interaction between the "number of items, Q matrix and method" and "sample size, Q matrix and method" in The Clinical Medical Sciences Test ($p < 0.01$). The common effect plots are presented in Figure 8 to investigate how the classification accuracies differ under different simulation conditions for The Clinical Medical Sciences Test. Looking at Figure 8, it can be seen that for all models, the classification accuracies for both 120 and 60 items are similar. The classification accuracy values for the Q matrix determined by Fields are similar for all models. In contrast, for the Q matrix determined by Miller, the classification accuracy values obtained for DINA and HO-DINA are similar, and those obtained for DINO and HO-DINO are similar. In addition, the results for different samples are very similar. For the Q-matrix determined by Fields, the classification accuracy values are similar for all models, whereas, for the Q-matrix determined by Miller's, the classification accuracy values for the DINO and HO-DINO models are higher than those for the DINA and HO-DINA models. In addition, it can be seen that the classification accuracies obtained by all models for the Q matrix determined by Fields are higher than the classification accuracies obtained for the Q matrix determined by Miller. Considering that the classification accuracy values given by the models for the Q matrix determined based on the Fields are more consistent with each other and higher than the classification accuracy values obtained for the Q matrix determined based on Miller, it can be interpreted that it is more advantageous to use the Q matrix determined based on the fields for the TUS exam.

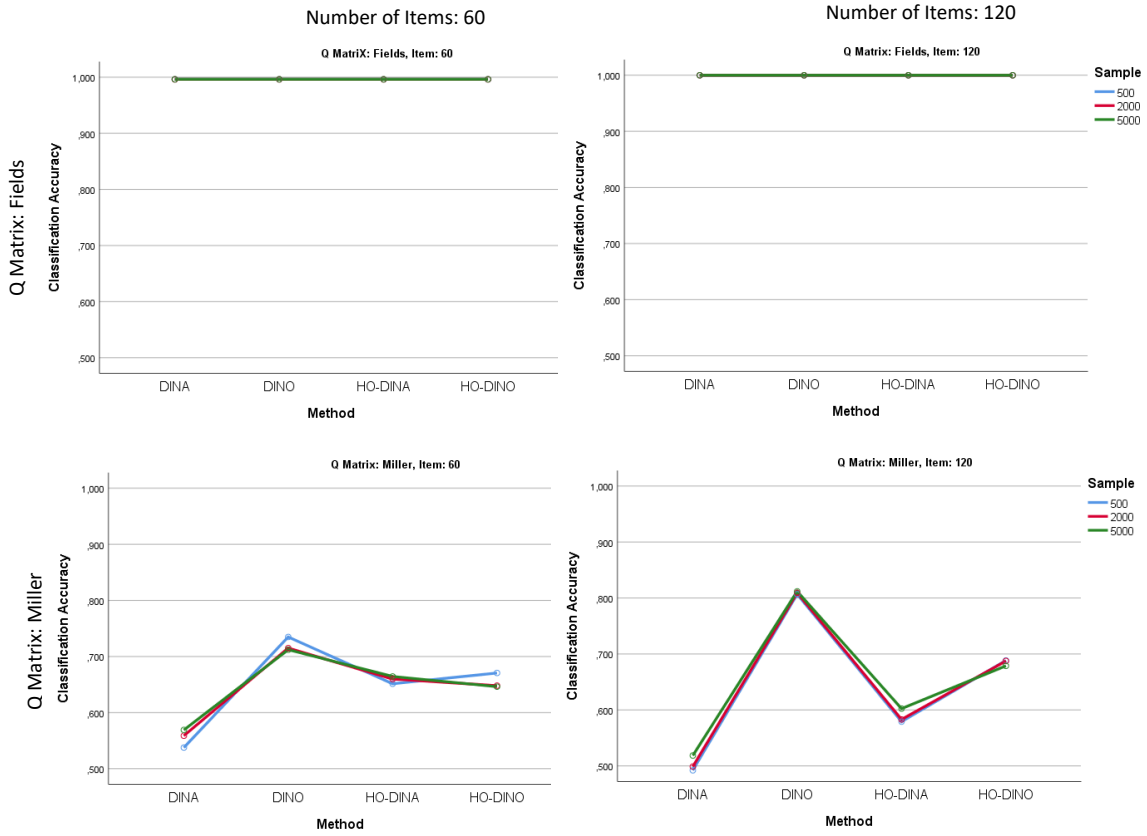


**Figure 8: Classification Accuracy Values for The Basic Clinical Sciences Test under Different Simulation Conditions**

## CONCLUSION, DISCUSSION AND RECOMMENDATIONS

In this study, the impact of sample size, number of items, and different Q matrices on the RMSEA, g and s parameters, and classification accuracy of the DINA, HO-DINA, DINO, and HO-DINO models was investigated for a test applied in health education. As stated by De la Torre et al. (2010), the accurate estimation of item parameters and the accurate classification of attributes are of paramount importance for obtaining valid inferences in cognitive diagnosis, hence their importance.

The findings of the study revealed that when examining the RMSEA values of the DINA, HO-DINA, DINO, and HO-DINO models for The Basic Medical Sciences and Clinical Medical Sciences tests under different simulation conditions; it can be observed that the DINO and HO-DINO models provide a better fit to the data in all conditions. Hu et al. (2016) noted that model fit is affected by the misidentification of the Q matrix. Therefore, based on the findings, when analyzing the RMSEA values for The Basic Medical Sciences and Clinical Medical Sciences tests, it is evident that the Q matrix determined by Fields provides a better fit to the data, and moreover, it is advantageous for the Q matrix determined by Fields to be used for the TUS exam. When all conditions were analysed, it was determined that variations in the sample size for Basic Medical and Clinical Medical Sciences did not change the model fit. In addition, although the RMSEA values were similar in terms of the number of items, it was determined that the RMSEA values obtained for 120 items in Basic Medical Sciences were lower, while the RMSEA values obtained for 60 items in Clinical Medical Sciences were lower for the Q matrix determined based on the Fields. Accordingly, it was observed that there was no significant difference in model fit with respect to the number of items. When analysed in terms of the number of items for the Q matrix determined based on Miller, although the RMSEA values were similar, it was determined that the RMSEA values obtained for 120 items in some cases and for 60 items in some cases in Basic Medical Sciences and Clinical Medical Sciences were lower. It can be said that this situation is due to the fact that the Q matrix determined based on Miller is not suitable for the measurement tool used.

When examining the s and g parameters obtained from the DINA, HO-DINA, DINO and HO-DINO models for The Basic Medical Sciences and Clinical Medical Sciences tests, it was observed that the values of the g and s parameters are low in all conditions, indicating that all models have a good fit. It was found that the g and s parameter values obtained for the attributes determined based on the Fields in the Q matrix were lower than the g and s parameter values obtained for the attributes determined based on the Miller. Accordingly, it can be concluded that determining attributes based on the Fields in the Q matrix gives better results than determining attributes based on Miller. When analyzing the impact of sample size, it was found that there is not much difference in the s and g parameters as a result of sample variation. Similarly, when considering the number of items, the g and s parameter values obtained for 60 items were similar to those obtained for 120 items.

In terms of the classification accuracy obtained from the DINA, HO-DINA, DINO and HO-DINO models for The Basic Medical Sciences and Clinical Medical Sciences tests, it was found that classification accuracy determined by Fields for the Q matrix was

higher than those determined by Miller. Therefore, it has been revealed that determining the attributes in the Q matrix based on the Fields gives better results than the determination based on Miller. Given that the baseline exam in this study is the TUS exam, it is suggested that the Q matrix determined based on the Fields gives better results. However, it should be noted that the results of Q matrices may vary depending on the content and extent of the examination. In future studies, the results of Q matrices determined by considering different exams related to medical specialty can be investigated. Analyzing the impact of sample size, it was found that variations in the sample did not change the classification accuracy. When considering the number of items, although the classification accuracy is similar, the classification accuracy obtained for 120 items in Basic Medical Sciences and Clinical Medical Sciences for the Q matrix determined based on the Fields are higher than the classification accuracy obtained for 60 items. When analysed in terms of the number of items for the Q matrix determined based on Miller, although the classification accuracy is similar, it is determined that the classification rates obtained for 120 items in some cases and 60 items in some cases in Basic Medical Sciences and Clinical Medical Sciences are high. It can be said that this situation is due to the fact that the Q matrix determined based on Miller is not suitable for the measurement tool used. De la Torre et al. (2010) suggested to investigate the ideal test lengths required to obtain classification accuracy, and it is found that the classification accuracies are similar according to the number of items, but when there is a suitable Q matrix, the classification accuracy increases as the number of items increases. However, in the following studies, the impact of the number of items on classification accuracy can be examined by increasing the conditions for the number of items (e.g. 15-30-60-120 items).

When the Q matrices created for Fields and Miller in the Basic Medical and Clinical Medical Sciences test is analysed according to the model fit, it is seen that the parameters related to DINO, HO-DINO models give better results than DINA, HO-DINA models. Reviewing the literature, it is found that a DINA model with statistically independent qualities gives better results than compensatory models. However, Delatorre stated in his study in 2004 that it would be more appropriate to use the compensatory model in psychiatric or medical diagnosis. In parallel with the literature, DINO, HO-DINO models gave better results than the compensatory models in the medical tests used in this study. Based on all these findings, the sample variation did not change the RMSEA, g and s parameters and classification rates for DINA, HO-DINA, DINO and HO-DINO models. Consistent with these findings, De la Torre et al. (2010) found that small sample sizes are sufficient to accurately estimate DINA model parameters. In this study, in parallel with this finding, it is determined that the difference in the number of samples does not change the model parameters. In the literature (e.g., Sünbül & Adnan, 2013, Chiu, 2013), regarding the number of items, an improvement in values was generally observed as the number of items increased. In this study, in parallel with this finding, it is noted that there is an improvement in the values as the number of items increases for the Q matrix determined based on Fields that better fit the data. When examining Q matrices, it was seen that the Q matrix determined based on Fields gave better results than the Q matrix determined based on Miller. Therefore, it can be interpreted that using the Q matrix determined based on the fields for the TUS exam is more advantageous. Research (e.g., Rupp & Templin, 2008) has indicated that the number of attributes measured by an item and the proportion of items measuring an attribute can impact estimation accuracy. In this study, since the number of items was reduced to 60 without changing the proportion of items related to the profiles, there may not have been a difference in classification accuracy. Therefore, when the number of items is reduced in the following studies, it can be taken as a random and equal proportion, and the results can be compared.


## Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) received no financial support for the research, author-ship, and/or publication of this article.


## Statements of publication ethics

I/We hereby declare that the study has not unethical issues and that research and publication ethics have been observed carefully.

## Researchers' contribution rate

The study was conducted and reported with equal collaboration of the researchers.

## REFERENCES


Aryadoust, V. (2018). A cognitive diagnostic assessment study of the listening test of the Singapore–Cambridge General Certificate of Education O-Level: Application of DINA, DINO, G-DINA, HO-DINA, and RRUM. *International Journal of Listening*, 35(1), 29-52.

Aydın, S., & Demir, M. (2006). *Sağlıkta performans yönetimi: Performansa dayalı ek ödeme sistemi*. Sağlık Bakanlığı.

Bakan, İ., Erşahan, B., Kefe, İ. & Bayat, M. (2011). Kamu ve özel hastanelerde tedavi gören hastaların sağlıkta hizmet kalitesine ilişkin algılamaları. *Kahramanmaraş Sütçü İmam Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 1(2), 1-26.

Bakanlığı, T. S. (2001). Sağlık Hizmetlerinin Yürütülmesi Hakkında Yönerge.

Bakanlığı, T. S., & Müdürlüğü, T. S. H. G. (2011). Sağlığın teşviki ve geliştirilmesi sözlüğü. *Bakanlık yayın*, 814(1), 23.

Chiu, C. Y. (2013). Statistical refinement of the Q-matrix in cognitive diagnosis. *Applied Psychological Measurement, 37(8),* 598-618.

Collares, C. F. (2022). Cognitive diagnostic modelling in healthcare professions education: an eye-opener. *Advances in Health Sciences Education*, 27(2), 427-440.

DeCarlo, L. T. (2011). On the analysis of fraction subtraction data: The DINA model, classification, latent class sizes, and the Q-matrix. *Applied Psychological Measurement*, 35(1), 8-26.

De La Torre, J. (2009). DINA model and parameter estimation: A didactic. *Journal of educational and behavioral statistics*, 34(1), 115-130.

De La Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69(3), 333-353.

De La Torre, J., Hong, Y., & Deng, W. (2010). Factors affecting the item parameter estimation and classification accuracy of the DINA model. *Journal of Educational Measurement*, 47(2), 227-249.

De La Torre, J., & Minchen, N. (2014). Cognitively diagnostic assessments and the cognitive diagnosis model framework. *Psicología Educativa*, 20(2), 89-97.

Doğan, N. Ö., & Gencan, S. (2014). VZA/AHP bütünleşik yöntemi ile performans ölçümü: Ankara'daki kamu hastaneleri üzerine bir uygulama. *Gazi Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi*, 16(2), 88-112.

Grégoire, J. (1997). Diagnostic assessment of learning disabilities: From assessment of performance to assessment of competence. *European Journal of Psychological Assessment*, 13(1), 10-20.

Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, 26(4), 301-321.

Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, 74(2), 191-210.

Hu, J., Miller, M. D., Huggins-Manley, A. C., & Chen, Y. H. (2016). Evaluation of model fit in cognitive diagnosis models. *International Journal of Testing*, 16(2), 119-141.

Jang, E. E., & Wagner, M. (2014). Diagnostic feedback in the classroom. *The companion to language assessment*, 2, 157-175.

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.

Kalkan, Ö. K., & Başokçu, T. O. (2019). The Effect of the Item–Attribute Relation on the DINA Model Estimations in the Presence of Missing Data. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 46(46), 290-306.

Ma, W., & Guo, W. (2019). Cognitive diagnosis models for multiple strategies. *British Journal of Mathematical and Statistical Psychology*, 72(2), 370-392.

Ma, C., Ouyang, J., & Xu, G. (2022). Learning latent and hierarchical structures in cognitive diagnosis models. *Psychometrika*, 1-33.

Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic medicine,* 65(9), S63-7.

Rupp, A. A., & Templin, J. (2007). The effects of Q-matrix misspecification on parameter estimates and classification accuracy in the DINA model. *Educational and Psychological Measurement*, 68(1), 78-96.

Rupp, A. A., & Templin, J. L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement*, 6(4), 219-262.

Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods and applications*. Guilford Press.

Su, Y. L. (2013). *Cognitive diagnostic analysis using hierarchically structured skills.* The University of Iowa.

Sünbül, S. Ö., & Adnan, K. A. N. (2013). Bilişsel Tanı Modellerinde Parametre Kestirimini ve Sınıflama Tutarlılığını Etkileyen Faktörlerin İncelenmesi Factors Affecting the Item Parameter Estimation and Classification Accuracy of the Cognitive Diagnostic Models.

Tatsuoka, K. K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. *Cognitively diagnostic assessment*, 327-359.

Tatsuoka, K. K., & Tatsuoka, M. M. (1997). Computerized cognitive diagnostic adaptive testing: Effect on remedial instruction as empirical validation. *Journal of Educational Measurement,* 34(1), 3-20.

Templin, J., & Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika*, 79(2), 317-339.

Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological methods*, 11(3), 287.

WORLD HEALTH ORGANIZATION, et al. World Conference on Medical Education: Edinburgh, 8-12 August 1988. In: *Executive Board Session*, 83. World Health Organization, 1988.