

# Bilişsel Tanı Modellerinde Madde Düzeyinde Tanımlanmış En Uygun İndirgenmiş Modelinin Belirlenmesi\*

Mahmut Sami KOYUNCU<sup>1</sup>  Şeref TAN<sup>2</sup> 

<sup>1</sup> Dr. Öğr. Üyesi, Afyon Kocatepe Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Afyon, Türkiye [ms\\_koyuncu@hotmail.com](mailto:ms_koyuncu@hotmail.com) (Sorumlu Yazar/ Corresponding Author)

<sup>2</sup> Prof. Dr., Gazi Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Ankara, Türkiye [sereftan4@yahoo.com](mailto:sereftan4@yahoo.com)

## Makale Bilgileri

## ÖZ

**Makale Geçmişi**  
Geliş: 28.08.2023  
Kabul: 08.10.2023  
Yayın: 29.10.2023

**Anahtar Kelimeler:**  
Bilişsel Tanı Modeli,  
G-DINA Model,  
Madde Düzeyinde Model,  
Model Uyumu.

Bu çalışmanın amacı, Q-matrisinde en az iki nitelik gerektiren maddeler için madde düzeyinde tanımlanan en uygun indirgenmiş modelin WALD, LR, 2LR ve LM test yöntemlerine göre belirlenmesidir. Ayrıca her bir madde için belirlenen indirgenmiş model kullanılarak tanımlanan modelin bağlı ve mutlak model uyum değerlerini G-DINA model uyum değerleriyle karşılaştırmak amaçlanmıştır. Araştırmanın çalışma grubu Ankara ilinde yedi farklı lisede 9.sınıfta öğrenim gören 712 öğrenciden oluşmaktadır. Çalışmada verileri araştırmacı tarafından geliştirilen beş niteliğin yoklandığı, çoktan seçmeli ve 26 maddeden oluşan tanılayıcı bir matematik testinden elde edilmiştir. Verilerin analizi RStudio yazılımında "GDINA" paket 2.9.4 versiyonu ile gerçekleştirilmiştir. Çalışma sonucunda, Wald, LR ve 2LR test yöntemlerinde en çok LLM modelinin, LM test yönteminde ise DINA modelin maddeler için en uygun indirgenmiş model olarak belirlendiği tespit edilmiştir. Tüm modellerde mutlak model uyum değerlerinin iyi olduğu, en düşük RMSEA değerinin G-DINA modelde, en düşük SRMSR değerinin ise  $TM_{LM}$  modelinde olduğu belirlenmiştir.  $TM_{wald}$  ve  $TM_{LM}$  modellerinin G-DINA model kadar iyi uyuma sahip olmadığı,  $TM_{LR}$  ve  $TM_{2LR}$  modellerinin G-DINA model kadar iyi model uyumuna sahip olduğu belirlenmiştir. Bu modellerde madde düzeyinde sadeleştirmeye gidilebileceği ve G-DINA model yerine sadeleştirilmiş bu modellerin kullanılabilirliği sonucuna ulaşılmıştır. Araştırmacılara madde düzeyinde hangi yöntem kullanılarak daha güvenilir ve geçerli bir şekilde model seçilebileceği konusunda çalışmalar yapması önerilmektedir.

## Determination of the Most Appropriate Reduced Model Defined at the Item Level in Cognitive Diagnostic Models

### Article Info

### ABSTRACT

**Article History**  
Received: 28.08.2023  
Accepted: 08.10.2023  
Published: 29.10.2023

**Keywords:**  
Cognitive Diagnostic Model,  
G-DINA Model,  
Item-Level Model,  
Model Fit.

This study aims to identify the most suitable reduced model at the item level for items requiring a minimum of two attributes in the Q-matrix utilizing the WALD, LR, 2LR, and LM test methods. Additionally, the study aims to compare the relative and absolute model fit values of the reduced model for each item with those of the G-DINA model. The research participants consists of 712 9<sup>th</sup> grade students attending seven different high schools in Ankara. Data were collected from a diagnostic mathematics test comprising 26 multiple-choice items developed by the researcher to assess five attributes. The analysis employed the "GDINA" package version 2.9.4 within RStudio software. The study's findings indicate that the LLM model serves as the most appropriate reduced model for items based on the Wald, LR, and 2LR test methods, while the DINA model provides the most suitable reduced model for items in the LM test method. All models exhibited good absolute model fit values; with the G-DINA model displaying the lowest RMSEA value, and the  $TM_{LM}$  model showing the lowest SRMSR value. It was determined that  $TM_{wald}$  and  $TM_{LM}$  models did not exhibit as good a fit as the G-DINA model, whereas  $TM_{LR}$  and  $TM_{2LR}$  models displayed a model fit as good as that of the G-DINA model. Consequently, the study concludes that these models can be simplified at the item level and used as alternatives to the G-DINA model. It is recommended that researchers undertake further investigations into the methods for selecting item-level models that are both reliably and validly.

**Atıf/Citation:** Koyuncu, M. S. & Tan, Ş. (2023). Bilişsel Tanı Modellerinde Madde Düzeyinde Tanımlanmış En Uygun İndirgenmiş Modelinin Belirlenmesi. *Ahmet Keleşoğlu Eğitim Fakültesi Dergisi (AKEF) Dergisi*, 5(3), 1526-1535.



"This article is licensed under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/) (CC BY-NC 4.0)"

\*Bu makale ilk yazarın "Bilişsel Tanı Modellerinde Yapısal Eşitlik Modeli ile Q-Matris Doğruluğunun Belirlenmesi" başlıklı doktora tezinin bir kısmından üretilmiştir.

## GİRİŞ

Ülkemizdeki eğitim sisteminde, geniş ölçekli merkezi sınavlar ve okullarda yapılan sınavlar olmak üzere birçok sınav yapılmaktadır. Ancak ne okulda gerçekleştiren sınavların sonuçları ne de merkezi sınav sonuçları, öğrencilere mevcut durumları hakkında geri bildirim vermek ya da öğretim faaliyetlerini izlemek için çoğunlukla kullanılmamaktadır. Bu sınavlardan elde edilen puanlar daha çok öğrencileri bir programa seçme veya yerleştirmek, ya da öğrencinin bir dersten başarılı olup olmayacağına karar vermek amacıyla kullanılmaktadır. De la Torre (2009)'nin de belirttiği gibi bu amaçlarla yapılan değerlendirmeler yeterli tanılayıcı bilgi vermemektedir.

Gereksinim duyulan tanılayıcı değerlendirme bilgilerine Bilişsel Tanı Modelleri (BTM) kullanılarak ulaşılabilir. Bu modellerde temel amaç öğrencilerin sahip olduğu nitelikler/beceriler hakkında “tanı” koymak ve öğrencilere ayrıntılı geri bildirim vermektedir (De Carlo, 2012). BTM, öğrencilerin zayıf ve güçlü oldukları yönler hakkında bilgi sağlamak amacıyla kullanılan psikometrik modellerdir. Bu modeller, öğrencilerin öğrenme ve gelişimlerinin etkili ölçümlerine, bireysel ve grup ihtiyaçlarına yönelik olası müdahalelere ve daha iyi öğretim tasarımlarına imkân veren puan profilleri hakkında özel bilgiler sunmaktadır (de la Torre, 2009).

Literatürde birçok farklı bilişsel tanı modeli vardır. Bir madde için en uygun BTM modelini seçmek zorlu bir süreçtir. Genelleştirilmiş BTM'ler daha iyi model-veri uyumu sağlasa da, daha basit olan spesifik BTM'ler daha basit yorumlara sahiptir, daha karardır ve doğru kullanıldığında daha doğru sınıflandırmalar sağlayabilmektedir (Ma vd., 2016).

G-DINA model yaygın olarak kullanılan birkaç BTM'yi kapsamaktadır. G-DINA çerçevesinde Wald testi ile indirgenmiş bir modelin (örneğin, DINA, DINO, A-CDM vb. gibi) doymuş bir modelin (G-DINA gibi) yerine kullanılıp kullanılmayacağını madde madde istatistiksel olarak belirlemek mümkündür (de la Torre ve Lee, 2013). Wald testi de la Torre (2011) tarafından önerilmiştir ve en az iki nitelik gerektiren maddeler için madde düzeyinde model seçimine imkân vermektedir. Ayrıca, G-DINA model çerçevesinde madde düzeyinde model seçimi için Olabilirlik Oran (Likelihood Ratio-LR) testi ve iki adımlı olabilirlik oran yaklaşımı (two-step LR test -2LR) testi (Ma, 2017; Ma ve de la Torre, 2019a; Sorrel vd., 2017a); sadece madde düzeyinden indirgenmiş DINA, DINO ve A-CDM modellerin inceleme yapmaya izin veren Lagrange Çarpanı (Lagrange Multiplier-LM) testi (Sorrel vd., 2017b) de kullanılabilir.

Literatürde BTM'leri içeren madde düzeyindeki uyum değerlendirmesine ilişkin araştırmalar oldukça azdır. Tutumluluk ilkesine göre, uyum iyiliği ile model karmaşıklığının dengelenmesi gereklidir. Genelleştirilmiş BTM'ler güvenilir bir şekilde kestirim yapabilmek için daha geniş örneklem boyutuna ihtiyaç duyar. Örneklem boyutu küçük olduğunda ve madde kalitesi düşük olduğunda, tipik olarak birçok ampirik uygulamada olduğu gibi, indirgenmiş modellere göre daha kötü nitelik sınıflama doğruluğuna yol açabilirler (Sorrel vd., 2017b).

Her bir madde için bağımsız bir modelin belirlenebilmesi (i) araştırmalarda BTM'lerin önceden belirlenmesine gerek olmamasına, (ii) istatistiksel olarak belirlenmiş birden fazla BTM'nin tek bir değerlendirme içinde kullanılabilmesine fırsat sunmaktadır. Bu sayede bilişsel tanı modellerinin uygulanması daha esnek hale gelmektedir (de la Torre & Lee, 2013). Bu açıdan makale, bilişsel tanı modellerinin pratik eğitim ortamlarında kullanımının geliştirilmesine yönelik önemli bir katkı sağlamaktadır.

Bu çalışmanın temel amacı, G-DINA model çerçevesinde Wald, LR, 2LR ve LM test yöntemlerine göre madde düzeyinde belirlenen en uygun indirgenmiş modellerin ne olduğunun belirlenmesidir. Ayrıca her bir madde için belirlenen indirgenmiş modeller kullanılarak tanımlanan modellerin G-DINA model ile karşılaştırmalı olarak model uyumlarının incelenmesi amaçlanmıştır. Bu kapsamda,

1. Wald, LR, 2LR ve LM test yöntemlerine göre madde düzeyinde belirlenen en uygun indirgenmiş modeller nedir?

2. Farklı test yöntemlerine göre madde düzeyinde belirlenen modeller kullanılarak tanımlanan modellerin mutlak ve bağıl model uyum değerleri nasıl değişmektedir?

### YÖNTEM

#### Araştırma Modeli

Çalışmada G-DINA model çerçevesinde Wald, LR, 2LR ve LM test yöntemlerine göre madde düzeyinde belirlenen en uygun indirgenmiş modellerin ne olduğunun belirlenmesi ve her bir madde için belirlenen indirgenmiş modeller kullanılarak tanımlanan modellerin G-DINA model ile karşılaştırmalı olarak model uyumlarının incelenmesi amaçlandığı için betimsel bir araştırmadır.

#### Çalışma Grubu

Araştırmanın çalışma grubu Ankara ilinde yedi farklı lisede 9.sınıfta öğrenim gören 712 öğrenciden oluşmaktadır. Öğrencilerin cinsiyete göre dağılımları Tablo1’de yer almaktadır.

**Tablo 1. Öğrencilerin Cinsiyete Göre Dağılımı**

Değişken	Kategori	N	%
Cinsiyet	Kadın	372	52.2
	Erkek	283	39.7
	Belirtilmemiş	57	8.0
<b>Toplam</b>		712	100.0

Tablo 1 incelendiğinde çalışma kapsamında yer alan öğrencilerin cinsiyete göre %53.2 (n=372)’si kadın, %39.7 (n=283)’sinin ise erkek öğrenci olduğu görülmektedir.

#### Veri Toplama Araçları ve Süreçleri

Çalışmada verileri araştırmacı tarafından geliştirilen 9.sınıf düzeyinde çoktan seçmeli ve 26 maddeden oluşan tanılayıcı bir matematik testinden elde edilmiştir. Öncelikte teste yer alacak nitelikler, yani test maddeleri ile ölçülmek istenilen bilgi ve beceriler belirlenmiştir. Test beş nitelikten oluşmaktadır. Bunlar: “N-1: Sayı kümelerini birbiriyle ilişkilendirir. N-2: Gerçek sayılar kümesinde aralık kavramını açıklar. N-3: Birinci dereceden bir bilinmeyenli denklem ve eşitsizliklerin çözüm kümelerini bulur. N-4: Mutlak değer içeren birinci dereceden bir bilinmeyenli denklem ve eşitsizliklerin çözüm kümelerini bulur. N-5: Birinci dereceden iki bilinmeyenli denklem ve eşitsizlik sistemlerinin çözüm kümelerini bulur.”

Nitelikler belirlendikten ve test maddeli yazıldıktan sonra Q-matrisi belirlenme aşamasına geçilmiştir. Q-matrisi nitelikler ile test maddesinin ilişkisini gösteren bir matristir. Q matrisi, bir niteliğin bir madde için gerekli olup olmadığını gösteren, 1 ve 0 girişleri olan bir matristir. Modellerin ve Q matrislerinin seçimi, maddelere ilişkin bilişsel süreçler ve işlemlerin altında yatan ilgili teorik varsayımların bir kümesini temsil etmektedir (Li ve Wang, 2015). Dört alan uzmanının salt çoğunluk görüşüne göre Q-matrisi belirlenmiştir. Uzman görüşüne göre belirlenen Q-matrisi Tablo 2’de yer almaktadır.

**Tablo 2. Beş Nitelikten Oluşan Q-Matrisi**

Madde	N-1	N-2	N-3	N-4	N-5
M-1	1	0	0	0	0
M-2	1	1	0	0	0
M-3	1	0	0	0	0
M-4	0	1	0	0	0
M-5	0	0	1	0	0
M-6	0	1	1	0	0
M-7	0	1	1	0	1
M-8	1	1	0	0	0
M-9	0	1	1	0	0
M-10	0	1	0	0	0
M-11	0	0	1	0	0
M-12	0	0	1	0	0
M-13	0	0	1	0	1
M-14	0	0	1	0	0

### Bilişsel Tanı Modellerinde Madde Düzeyinde Tanımlanmış En Uygun İndirgenmiş Modelinin Belirlenmesi

M-15	0	0	1	0	0
M-16	0	0	1	0	0
M-17	0	0	1	1	0
M-18	0	0	1	1	0
M-19	0	0	1	1	0
M-20	0	0	0	1	0
M-21	0	1	0	1	0
M-22	0	0	0	1	1
M-23	0	1	0	1	0
M-24	0	0	0	0	1
M-25	0	0	0	0	1
M-26	0	1	0	0	1
<b>Toplam</b>	<b>4</b>	<b>10</b>	<b>13</b>	<b>7</b>	<b>6</b>

Tablo 2’de görüldüğü üzere Nitelik-1’in 4; Nitelik-2’nin 10; Nitelik-3’ün 13; Nitelik-4’ün 7 ve Nitelik-5’in 6 madde ile yoklandığı görülmektedir. Hartz ve diğerleri (2002) güvenilir tanılayıcı bilgi elde etmek için her bir niteliğin en az 3 farklı madde ile yoklanması gerektiğini belirtmiştir.

#### Verilerin Analizi

Çalışma kapsamında ilk olarak G-DINA model çerçevesinde Wald, LR, 2LR, LM test yöntemlerine göre her bir madde için bağımsız olarak en uygun indirgenmiş model (DINA, DINO, A-CDM, LLM, R-RUM) belirlenmiştir. Ancak, LM test yöntemi maddeleri sadece DINA, DINO ve A-CDM modellerine göre değerlendirmeye izin vermektedir (Sorrel vd., 2017b). Madde düzeyinde hangi modelin veriye daha iyi uyum sağladığına karar vermek için 0.05 alfa düzeyinde (i) daha basit model ve (ii) en büyük  $p$  değeri kuralı birlikte kullanılarak karar verilmiştir. Farklı test yöntemlerine göre madde düzeyinden en uygun indirgenmiş modeller belirlendikten sonra, ikinci aşamada her bir madde için belirlenen en uygun modeller kullanılarak tanımlanan modellerin bağıl ve mutlak uyum değerleri incelenmiştir.

Bilişsel tanı modellerinin veri uyumu değerlendirilmesinde bağıl uyum indeksleri olarak -2LL, AIC, BIC, CAIC ve SABIC değerleri, mutlak uyum indeksi olarak  $M_2$ , RMSEA ve SRMSR değerleri kullanılmıştır. Bu değerlerin değerlendirilmesinde, daha küçük bağıl uyum indeksi değerleri modelin veriye daha iyi uyum sağladığının göstergesidir (Chen vd., 2013; De Ayala, 2009). Mutlak uyum indeksi SRMSR için 0.08’den ve RMSEA için 0.06’dan küçük değerler iyi uyum göstergesidir (Hu ve Bentler, 1999). Ayrıca, dört farklı test yöntemine göre madde düzeyinden tanımlanan her bir modelin G-DINA model ile -2LL değerleri olabilirlik oran testi yardımıyla ve ki-kare tablosu baz alınarak değerlendirilmiştir (Ma ve de la Torre, 2019b). Çalışmada RStudio yazılımı kullanılarak “GDINA” paket 2.9.4 versiyonu (Ma vd., 2023) ile analizler gerçekleştirilmiştir.

#### Etik

Bu araştırmada 2020 yılı öncesi araştırma verileri kullanılmış olup, doktora tezinden üretilmiştir. Ankara Valiliği Milli Eğitim Müdürlüğü’nden veri toplama aracına ait gerekli araştırma izni alınmıştır (Tarih:26/04/2018-Sayı:14588481-605.99-E.8338810).

#### BULGULAR

Çalışma kapsamında G-DINA model çerçevesinde en az iki nitelik gerektiren maddeler için Wald, LR, 2LR ve LM test yöntemlerine göre madde düzeyinde belirlenen en uygun indirgenmiş model sonucu Tablo 3’te sunulmuştur. Ayrıca bu test yöntemlerine göre belirlenen modeller ve  $p$  değerleri Ek-1’de sunulmuştur.

**Tablo 3.** Madde Düzeyinde Wald, LR, 2LR ve LM Test Yöntemlerine Göre Model Seçimi

MADDE	WALD	LR	2 LR	LM
Madde-1	GDINA	GDINA	GDINA	GDINA
Madde-2	RRUM	DINO	GDINA	ACDM
Madde-3	GDINA	GDINA	GDINA	GDINA
Madde-4	GDINA	GDINA	GDINA	GDINA

## Bilişsel Tanı Modellerinde Madde Düzeyinde Tanımlanmış En Uygun İndirgenmiş Modelinin Belirlenmesi

Madde-5	GDINA	GDINA	GDINA	GDINA
Madde-6	DINO	DINO	DINO	ACDM
Madde-7	LLM	GDINA	GDINA	ACDM
Madde-8	LLM	LLM	LLM	GDINA
Madde-9	DINO	DINO	DINO	DINO
Madde-10	GDINA	GDINA	GDINA	GDINA
Madde-11	GDINA	GDINA	GDINA	GDINA
Madde-12	GDINA	GDINA	GDINA	GDINA
Madde-13	LLM	LLM	LLM	ACDM
Madde-14	GDINA	GDINA	GDINA	GDINA
Madde-15	GDINA	GDINA	GDINA	GDINA
Madde-16	GDINA	GDINA	GDINA	GDINA
Madde-17	DINA	DINA	DINA	DINA
Madde-18	DINO	DINO	DINO	DINA
Madde-19	LLM	DINA	LLM	DINA
Madde-20	GDINA	GDINA	GDINA	GDINA
Madde-21	LLM	LLM	LLM	DINA
Madde-22	DINA	GDINA	GDINA	ACDM
Madde-23	ACDM	LLM	LLM	DINA
Madde-24	GDINA	GDINA	GDINA	GDINA
Madde-25	GDINA	GDINA	GDINA	GDINA
Madde-26	RRUM	RRUM	RRUM	DINA

*A-CDM*: Additive Cognitive diagnostic model, *DINA*: Deterministic, noisy and gate, *DINO*: Deterministic input, noisy or gate, *G-DINA*: Generalized deterministic, noisy and gate, *LLM*: Linear Logistic Model, *R-RUM*: Reduced Reparameterized Unified Model

Tablo 3 incelendiğinde sadece tek bir niteliği gerektiren maddeler de indirgenmiş model seçimi yapılmamış olup G-DINA model kullanılmıştır. Çalışma kapsamında incelenen Wald, LR, 2LR ve LM test sonuçlarına göre sadece iki madde de (Madde-9 ve Madde-17) belirlenen en uygun sadeleştirilmiş model aynıdır. Dört test yönteminde de Madde-9 için DINO, Madde-17 için ise DINA en uygun model olarak belirlenmiştir. Madde-2 için ise Wald testi RRUM, LR testi DINO, 2LR testi G-DINA ve LM testi A-CDM olmak üzere, dört yöntemde farklı sadeleştirilmiş modeli en uygun model olarak önermektedir. Dört farklı model seçiminde kullanılan test yöntemine göre, madde düzeyinde belirlenen en uygun sadeleştirilmiş modellerin maddelere göre dağılımı Tablo 4'te sunulmuştur. En az iki nitelik gerektiren sadece 13 maddeye ait sonuçlar Tablo 4'te yer almaktadır.

**Tablo 4.** Madde Düzeyinde Wald, LR, 2LR ve LM Test Yöntemlerine Göre İndirgenmiş Modellerin Maddelerde Dağılımı

Model	Yöntemlere Göre Belirlenen Maddeler			
	Wald	LR	2LR	LM
G-DINA	-	7,22	2,7,22	8
LLM	7,8,13,19,21	8,13,21,23	8,13,19,21,23	-
DINO	6,9,18	2,6,9,18	6,9,18	9
DINA	17,22	17,19	17	17,18,19,21,23,26
R-RUM	2,26	26	26	-
A-CDM	23	-	-	2,6,7,13,22

Tablo 4 incelendiğinde Wald, LR ve 2LR test yöntemlerinde en çok LLM modelinin, LM test yönteminde ise DINA modelin maddeler için en uygun indirgenmiş model olduğu belirlenmiştir. Wald testinde hiçbir maddede G-DINA model, LR ve 2LR test yöntemlerinde A-CDM modeli indirgenmiş en uygun model olarak belirlenmemiştir. Daha önceden belirtildiği üzere, LM test yöntemi maddeleri sadece DINA, DINO ve A-CDM modellerine göre değerlendirmektedir. Bu nedenle de doğal olarak hiçbir madde LLM ve R-RUM modelinde yer almamaktadır.

## Bilişsel Tanı Modellerinde Madde Düzeyinde Tanımlanmış En Uygun İndirgenmiş Modelinin Belirlenmesi

Tablo 3 ve Tablo 4'te görüldüğü gibi ilk aşamada farklı test yöntemlere göre her bir madde için en uygun model belirlenmiştir. İkinci aşama da ise farklı yöntemlere göre her bir madde için belirlenen en uygun indirgenmiş modeller tanımlanarak dört farklı özel tanımlanmış model elde edilmiştir. Wald testine göre madde düzeyinde belirlenen modellerle tanımlanan model “ $TM_{wald}$ ”, LR testine göre madde düzeyinde belirlenen modellerle tanımlanan model “ $TM_{LR}$ ”, 2LR testine göre madde düzeyinde belirlenen modellerle tanımlanan model “ $TM_{2LR}$ ” ve LM testine göre madde düzeyinde belirlenen modellerle tanımlanan model “ $TM_{LM}$ ” olarak adlandırılmıştır. Her bir madde için bağımsız olarak tanımlanan özel modellerle G-DINA modele için kestirilen mutlak uyum değerleri Tablo 5'te, bağıl uyum indeksi değerleri Tablo 6'da sunulmuştur.

**Tablo 5.** G-DINA ve Madde Düzeyinde Tanımlanmış Modellere Ait Mutlak Uyum Değerleri

Model	$M_2$	sd	p	RMSEA(% 90GA)	SRMSR
$TM_{wald}$	749.42	259	0.000	0.052 (0.047 - 0.056)	0.075
$TM_{LR}$	738.28	255	0.000	0.052 (0.047 - 0.056)	0.075
$TM_{2LR}$	705.68	252	0.000	0.050 (0.046 - 0.055)	0.075
$TM_{LM}$	888.55	260	0.000	0.058 (0.054 - 0.063)	0.065
G-DINA Model	669.14	238	0.000	0.050 (0.046 - 0.055)	0.075

*RMSEA: Root Mean Squared Error of Approximation, SRMSR: Standardized Root Mean Squared Residual, GA: Güven Aralığı*

Tablo 5 incelendiğinde madde düzeyinde her bir madde için özel tanımlanan dört model ve G-DINA model için  $M_2$  istatistiğinin anlamlı olduğu belirlenmiştir ( $p < 0,01$ ). Bu mutlak model veri uyumunun iyi olmadığı anlamına gelmektedir. Ancak  $M_2$  istatistiği model veri uyumsuzluklarına karşı çok duyarlı olduğu için genellikle model uyumunun olmadığı sonuçlar vermektedir. Bu sebeple  $M_2$  değerinin SRMSR değerleriyle beraber sunulması gerektiği belirtilmektedir (De Ayala, 2009; Maydeu-Olivares ve Joe, 2014). RMSEA değerleri incelendiğinde tüm modellerin 0.06'dan küçük değere sahip olduğu ve model veri uyumunun iyi olduğu görülmektedir. Tüm modellerde RMSEA değeri birbirine yakındır. En düşük RMSEA değeri G-DINA modelde, en yüksek ise  $TM_{LM}$ , yani LM test yöntemi sonuçlarına göre tanımlanan modelden elde edildiği görülmektedir. Benzer şekilde tüm modellerde SRMSR değeri 0.08 değerinden küçük olup iyi model uyumu göstermektedir. En küçük SRMSR değeri  $TM_{LM}$  modelinde kestirilmiş olup 0.065'tir. Diğer tüm modeller de ise SRMSR değeri eşit olup, 0.075'tir.

Veriye hangi modelin görece olarak daha iyi uyum sağladığını tespit etmek amacıyla bağıl uyum değerleri incelenmiş ve olabilirlik oran testi ile “ $H_0$ : İndirgenmiş model veri uyumunun doymuş (saturated) model kadar iyidir.” hipotezi test edilmiştir. Olabilirlik oran testinde  $H_1$  hipotezi ise “İki model eşdeğer değildir (basitleştirme gerekçeli değildir)” şeklindedir.

**Tablo 6.** G-DINA ve Madde Düzeyinde Tanımlanmış Modellere Ait Bağıl Uyum Değerleri

Model	-2LL	AIC	BIC	CAIC	SABIC	Parametre Sayısı	$\chi^2$	sd	p
G-DINA	19491.19	19717.19	20233.38	20346.38	19874.58	113	42.99	21	<0.01
$TM_{wald}$	19534.18	19718.18	<b>20138.45</b>	<b>20230.45</b>	19846.32	92			
G-DINA	19491.19	19717.19	20233.38	20346.38	19874.58	113	20.35	17	> 0.01
$TM_{LR}$	19511.53	<b>19703.53</b>	20142.07	20238.07	<b>19837.24</b>	96			
G-DINA	19491.19	19717.19	20233.38	20346.38	19874.58	113	15.55	14	>0.01
$TM_{2LR}$	19506.74	19704.73	20156.97	20255.97	19842.62	99			
G-DINA	19491.19	19717.19	20233.38	20346.38	19874.58	113	314.07	22	<0.01
$TM_{LM}$	19805.26	19987.27	20402.96	20493.96	20114.01	91			

*-2LL: -2 x log-likelihood, AIC: Akaike information criterion, BIC: Bayesian information criterion, CAIC: Consistent AIC, SABIC: Sample size adjusted BIC.*

Tablo 6 incelendiğinde madde düzeyinde özel tanımlanmış her bir modelin bağıl uyum değerlerinin G-DINA model bağıl uyum değerlerinin ikili olarak karşılaştırıldığı görülmektedir. Burada olabilirlik oran testi ile “İndirgenmiş model veri uyumu doymuş model kadar iyidir.” hipotezi test edilmiştir. Tablo 6

incelendiğinde G-DINA ile  $TM_{wald}$  modeli ve G-DINA ile  $TM_{LM}$  modelleri arasında gerçekleştirilen olabilirlik oran testinin anlamlı olduğu görülmektedir ( $p < 0,01$ ). Yani,  $TM_{wald}$  ve  $TM_{LM}$  modelleri G-DINA modele eşdeğer olmayıp indirgemeye (sadeleştirmeye) gerek yoktur. G-DINA ile  $TM_{LR}$  modeli ve G-DINA ile  $TM_{2LR}$  modelleri arasında gerçekleştirilen olabilirlik oran testi sonucu anlamlı değildir ( $p > 0,01$ ). Yani, indirgenmiş  $TM_{LR}$  ve  $TM_{2LR}$  model veri uyumu doymuş G-DINA model kadar iyidir. Kısaca, madde düzeyinde tanımlanan  $TM_{wald}$  ve  $TM_{LM}$  modelleri G-DINA model kadar iyi model veri uyumuna sahip değil iken,  $TM_{LR}$  ve  $TM_{2LR}$  modelleri G-DINA model kadar iyi model veri uyumuna sahiptir. Bu nedenle, madde düzeyinde sadeleştirme yapılarak, G-DINA model yerine indirgenmiş  $TM_{LR}$  ve  $TM_{2LR}$  modelleri kullanılabilir.

### TARTIŞMA, SONUÇ VE ÖNERİ

Çalışma kapsamında Q-matrisinde en az iki nitelik gerektiren maddeler için madde düzeyinde tanımlanan en uygun indirgenmiş modelin WALD, LR, 2LR ve LM test yöntemlerine göre belirlenmesi amaçlanmıştır. Daha sonra her bir madde için belirlenen indirgenmiş model tanımlanarak oluşturulan modelin bağıl ve mutlak model uyumları G-DINA model uyum değerleriyle karşılaştırılarak incelenmiştir.

Çalışma sonucunda, en az iki nitelik gerektiren Madde-9 ve Madde-17'nin tüm madde düzeyinde model seçim yöntemlerinde aynı indirgenmiş modelin en uygun model olduğunu belirttiği tespit edilmiştir. Sadece tek bir maddede (Madde-2) tüm test yöntemleri farklı modelleri en uygun yöntem olarak belirlediği sonucuna ulaşılmıştır. Wald, LR ve 2LR test yöntemlerinde en çok LLM modelinin, LM test yönteminde ise DINA modelin maddeler için en uygun indirgenmiş model olduğu tespit edilmiştir.

Mutlak model uyumları incelendiğinde, madde düzeyinde tanımlanmış  $TM_{wald}$ ,  $TM_{LR}$ ,  $TM_{2LR}$ ,  $TM_{LM}$  dört modelin ve G-DINA Modelin  $M_2$  istatistiğinin anlamlı olduğu ve mutlak veri uyumunun iyi olmadığını belirttiği tespit edilmiştir ( $p < 0,01$ ).  $M_2$  istatistiğinin genel olarak model uyumunun olmadığı sonuçlar verdiği bilinmekte olup ve tek başına değerlendirilmemesi gerektiği belirtilmektedir (De Ayala, 2009; Maydeu-Olivares ve Joe, 2014). Bununla birlikte, G-DINA ve tüm tanımlanmış modellerde RMSEA ve SRMSR değerlerinin iyi uyum sergilediği sonucuna ulaşılmıştır. En düşük RMSEA değeri G-DINA modelde, en düşük SRMSR değeri ise  $TM_{LM}$  modelinde olduğu belirlenmiştir. Genel olarak değerlendirildiğinde ise tüm modellerde mutlak model uyum değerlerinin iyi olduğu sonucuna ulaşılmıştır.

Bağıl uyum değerleri incelendiğinde, madde düzeyinde tanımlanan  $TM_{wald}$  ve  $TM_{LM}$  modelleri G-DINA model kadar iyi model veri uyumuna sahip olmadığı,  $TM_{LR}$  ve  $TM_{2LR}$  modellerinin ise G-DINA model kadar iyi bağıl model veri uyum değerlerine sahip olduğu belirlenmiştir. Yani,  $TM_{wald}$  ve  $TM_{LM}$  modelleri G-DINA modele eşdeğer olmayıp, bu modellerde madde düzeyinde modelin sadeleştirilmesine gerek yoktur. Ancak, indirgenmiş  $TM_{LR}$  ve  $TM_{2LR}$  modelleri doymuş G-DINA model kadar iyi model uyumuna sahip olduğu için, madde düzeyinde sadeleştirmeye gidilebileceği ve G-DINA model yerine sadeleştirilmiş bu modellerin kullanılabilmesi sonucuna ulaşılmıştır. Unutulmamalıdır ki, genel olarak gelişmiş modeller daha iyi model uyumuna sahiptir. Ancak gelişmiş modeller daha karmaşık ve yorumlanması daha zordur (Chen vd., 2013; de la Torre ve Lee, 2013). Bu nedenle gerçekleştirilen çalışmalarda bazı durumlarda model uyum değerleri kabul edilebilir düzeyde ise genelleştirilmiş modeller yerine basit modellerin kullanımı tercih edilebilir. Tabii ki bu kararı vermek güçtür. Ayrıca tutumluluk ilkesi, istatistiksel olarak ayırt edilemeyen bir dizi model mevcut olduğunda en basit modelin seçilmesini gerektirir (de la Torre & Lee, 2013). Ek olarak, Rojas ve diğerleri (2012) özellikle örneklem büyüklüğü küçük olduğunda doğru spesifik BTM modellerinin kullanılmasının genel bir BTM kullanılmasına göre daha yüksek doğru nitelik sınıflandırma oranına sahip olacağını belirtmiştir.

Araştırmacıları zorlayacak diğer bir durum ise madde düzeyinde model seçiminde hangi test yöntemini tercih edeceği konusudur. Sorrel ve diğerleri (2017b) çalışmasında I.tip hata ve güç karşılaştırmasına dayalı olarak LM testine göre LR ve Wald testinin tercih edilebileceğini belirtmiştir. Yapılan birçok araştırma madde bazlı bilgi matrisi kullanılarak hesaplanan Wald testinin I.tip hata oranını abartılı bir şekilde artırdığını ortaya koymuştur (Liu vd., 2018). Bu sorunu çözmek için Liu ve arkadaşları

(2018) Wald istatistiğinin orijinal versiyonundaki asimptotik kovaryans matrisini, geliştirilmiş hesaplama yöntemlerinden elde edilen bir matris ile değiştirmeyi önermiştir. Dolayısıyla, madde düzeyinde model seçiminde mükemmel bir yöntem yoktur, bu bakımdan araştırmacılara madde düzeyinde hangi yöntem kullanılarak daha güvenilir ve geçerli bir şekilde model seçilebileceği konusunda çalışmalar yapması önerilmektedir.

Çalışma kapsamında elde edilen sonuçlara göre, araştırmacıların geliştirdikleri ya da hazır olarak kullandıkları testleri BTM kapsamında değerlendirmek için tek bir modelle bağlı kalmalarına gerek olmadığı açıktır. Araştırmacılar her bir madde için bağımsız olarak tanımlanan daha basit modelleri kullanarak, tanımladıkları özel modellerle BTM değerlendirmelerini yapabilirler. Çünkü bazı maddeler farklı modellerde daha fazla bilgi sunabilmektedir. Araştırmacılara G-DINA model çerçevesinde madde düzeyinde model seçmeleri, eğer belirlenen model G-DINA model kadar iyi uyuma sahip ise madde düzeyinde sadeleştirilmiş modeli kullanmaları önerilmektedir.

### **Teşekkür**

Bu makale birinci yazarın doktora tezinin bir kısmından üretilmiştir. Doktora eğitimim süresince sağladığı maddi destekten ötürü Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK)'na teşekkürlerimi sunarım.

### **KAYNAKÇA**

- Chen, J., de la Torre, J., & Zhang, Z. (2013). Relative and absolute fit evaluation in cognitive diagnosis modeling. *Journal of Educational Measurement*, 50(2), 123-140. <https://doi.org/10.1111/j.1745-3984.2012.00185.x>
- De Ayala, R. J. (2009). *The theory and practice of item response theory. Methodology in the social sciences.* Guildford Publications.
- De Carlo, 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. <https://doi.org/10.1177/0146621610377081>
- de La Torre, J. (2009). A cognitive diagnosis model for cognitively based multiple-choice options. *Applied Psychological Measurement*, 33(3), 163-183. <https://doi.org/10.1177/0146621608320523>
- de la Torre, J., & Lee, Y. S. (2013). Evaluating the Wald test for item-level comparison of saturated and reduced models in cognitive diagnosis. *Journal of Educational Measurement*, 50(4), 355-373. <https://doi.org/10.1111/jedm.12022>
- Hartz, S., Roussos, L., & Stout, W. (2002). *Skills diagnosis: Theory and practice* [User Manual for Arpeggio software]. Educational Testing Service.
- Hu, L., & Bentler, P.M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6 (1), 1-55. <https://doi.org/10.1080/10705519909540118>
- Li, X., & Wang, W. C. (2015). Assessment of differential item functioning under cognitive diagnosis models: The DINA model example. *Journal of Educational Measurement*, 52(1), 28-54. <https://doi.org/10.1111/jedm.12061>
- Liu, Y., Andersson, B., Xin, T., Zhang, H., & Wang, L. (2019). Improved Wald statistics for item-level model comparison in diagnostic classification models. *Applied Psychological Measurement*, 43(5), 402-414. <https://doi.org/10.1177/0146621618798664>
- Ma, W., Iaconangelo, C., & de la Torre, J. (2016). Model similarity, model selection and attribute classification. *Applied Psychological Measurement*, 40, 200-217. <https://doi.org/10.1177/0146621615621717>



- Ma, W. (2017). *A Sequential cognitive diagnosis model for graded response: Model development, Q-matrix validation, and model comparison*. [Unpublished doctoral dissertation]. Rutgers University.
- Ma, W., & de la Torre, J. (2019a). Category-level model selection for the sequential G-DINA model. *Journal of Educational and Behavioral Statistics*, 44, 61-82. <https://doi.org/10.3102/1076998618792484>
- Ma, W. & de la Torre, J. (2019b). *GDINA: The generalized DINA model framework* (R package version 2.7.3.) <https://CRAN.R-project.org/package=GDINA>
- Ma, W., de la Torre, J., Sorrel, M., & Jiang, Z. (2023). *GDINA: The generalized DINA model framework* (R package version 2.9.4.) <https://cran.r-project.org/web/packages/GDINA/GDINA.pdf>
- Maydeu-Olivares, A., & Joe, H. (2014). Assessing approximate fit in categorical data analysis. *Multivariate Behavioral Research*, 49(4), 305–328. <https://doi.org/10.1080/00273171.2014.911075>
- Rojas G., de la Torre J., Olea J. (2012, April 14-16). *Choosing between general and specific cognitive diagnosis models when the sample size is small* [Paper presentation]. National Council on Measurement in Education Annual Meeting, Vancouver, British Columbia, Canada.
- Sorrel, M. A., de la Torre, J., Abad, F. J., & Olea, J. (2017a). Two-step likelihood ratio test for item-level model comparison in cognitive diagnosis models. *Methodology*, 13, 39-47. <https://doi.org/10.1027/1614-2241/a000131>
- Sorrel, M. A., Abad, F. J., Olea, J., de la Torre, J., & Barrada, J. R. (2017b). Inferential item-fit evaluation in cognitive diagnosis modeling. *Applied Psychological Measurement*, 41, 614-631. <https://doi.org/10.1177/0146621617707510>

## EXTENDED ABSTRACT

**Introduction:** The literature embodies various cognitive diagnosis models (CDM). Adopting the most appropriate CDM for an item is a challenging process. Although generalized CDMs provide better model-data fit, simpler specific CDMs have simpler interpretations, are more stable, and can provide more accurate classifications when used correctly (Ma et al., 2016). Within the G-DINA model framework, the Wald test allows item-level model selection for items that require at least two attributes (de la Torre (2011)). In addition, the Likelihood Ratio (LR) test and the two-step LR test (2LR) test (Ma, 2017; Ma & de la Torre, 2019a; Sorrel et al., 2017a) can also be used for item-level model selection, as well as the Lagrange Multiplier (LM) test (Sorrel et al., 2017b), which allows for the examination of DINA, DINO, and A-CDM models reduced only at the item level. The primary purpose of the current study is to find the most appropriate reduced models determined at item level according to Wald, LR, 2LR, and LM test methods within the framework of the G-DINA model. The study also sought to examine the model fit of the models defined using the reduced models determined for each item in comparison with the G-DINA model. In this context, the research questions include the followings:

1. What are the most appropriate reduced models determined at the item level according to Wald, LR, 2LR, and LM test methods?
2. How do the absolute and relative model fit values of the models defined using item-level models through different test methods differ?

**Materials and Methods:** The study has a descriptive design since it aims to determine the most appropriate reduced models determined at the item level according to Wald, LR, 2LR, and LM test methods within the framework of the G-DINA model and to examine the model fit of the models defined using the reduced models determined for each item in comparison with the G-DINA model. The study group of the research consists of 712 9th-grade students studying in seven different high schools in Ankara. The data were obtained from a diagnostic mathematics test developed by the researcher consisting of 26 multiple-choice items at the 9th-grade level. The attributes to be included in the test, i.e., the knowledge and skills to be measured by the test items, were determined. The test consists of five attributes.

Within the scope of the study, firstly, the most appropriate reduced model (DINA, DINO, A-CDM, LLM, R-RUM) was determined independently for each item according to Wald, LR, 2LR, and LM test methods within the framework of the G-DINA model. Following this step, the models' relative and absolute fit values were examined using the most appropriate models determined for each item in the second stage. The analyses were performed with the "GDINA" package version 2.9.4 (Ma et al., 2023) using RStudio software.

**Findings:** According to the Wald, LR, 2LR, and LM test results examined within the scope of the study, the

## Bilişsel Tanı Modellerinde Madde Düzeyinde Tanımlanmış En Uygun İndirgenmiş Modelinin Belirlenmesi

most appropriate simplified model determined is the same in only two items (Item-9 and Item-17). For Item-2, the Wald test RRUM, LR test DINO, 2LR test G-DINA, and LM test A-CDM suggest different simplified models as the most appropriate model in four methods. In the Wald, LR, and 2LR test methods, the LLM model was found to be the most appropriate reduced model, while in the LM test method, the DINA model was found to be the most appropriate one. The model defined by the item-level models according to the Wald test was named "TM<sub>wald</sub>," the model defined by the item-level models according to the LR test was named "TM<sub>LR</sub>," the model defined by the item-level models according to the 2LR test was named "TM<sub>2LR</sub>." The model defined by the item-level models according to the LM test was named "TM<sub>LM</sub>." RMSEA values were close to each other in all models. The lowest RMSEA value was obtained in the G-DINA model, and the smallest SRMSR value was obtained in the TMLM model. RMSEA and SRMSR values in all models showed a good fit. While TM<sub>wald</sub> and TM<sub>LM</sub> models defined at the item level did not have as good model-data fit as the G-DINA model, TM<sub>LR</sub> and TM<sub>2LR</sub> models had as good model-data fit as the G-DINA model. Therefore, the reduced TM<sub>LR</sub> and TM<sub>2LR</sub> models can be used by simplifying at the item level instead of the G-DINA model.

**Discussion:** The TM<sub>wald</sub> and TM<sub>LM</sub> models were not equivalent to the G-DINA model, so there was no need to simplify the model at the item level. However, since the reduced TM<sub>LR</sub> and TM<sub>2LR</sub> models had as good a model fit as the saturated G-DINA model, it was concluded that simplification at the item level could be made. These simplified models can be used instead of the G-DINA model. In general, advanced models have better model fit. However, advanced models are more complex and more challenging to interpret (Chen et al., 2013; de la Torre & Lee, 2013). Hence, in some studies, if the model fit values are acceptable, it may be preferable to use simple models instead of generalized ones. Of course, this is a difficult decision to make. Moreover, frugality requires selecting the simplest model when several statistically indistinguishable models are available (de la Torre & Lee, 2013). In addition, Rojas et al. (2012) noted that using accurate, specific CDM models will have a higher rate of correct attribute classification than using a general CDM, especially when the sample size is small

### Conclusion and Suggestions:

- Item-level reduced TM<sub>wald</sub> and TM<sub>LM</sub> models do not have as good a model-data fit as the G-DINA model.
- The reduced TM<sub>LR</sub> and TM<sub>2LR</sub> models have as good model fit as the saturated G-DINA model.
- Researchers do not need to stick to a single model to evaluate the tests they have developed or used as ready-made tests within the scope of the CDM.
- Using simpler models defined independently for each item, researchers can conduct CDM assessments with the specific models they define.
- Researchers are advised to select an item-level model within the framework of the G-DINA model. If the model fits as well as the G-DINA model, they should use the simplified model at the item level.
- Researchers are recommended to conduct studies on which method can be used to select item-level models more reliably and validly.

### Ek-1. Test Yöntemlerine Göre Madde Düzeyinde İngirgenmiş Modeller ve p Değeri

Madde	WALD		LR		TwostepLR		LM	
	Model	p değeri	Model	p değeri	Model	p değeri	Model	p değeri
M1	GDINA		GDINA		GDINA		GDINA	
M2	R-RUM	0.998	DINO	0.2793	GDINA		ACDM	0.9366
M3	GDINA		GDINA		GDINA		GDINA	
M4	GDINA		GDINA		GDINA		GDINA	
M5	GDINA		GDINA		GDINA		GDINA	
M6	DINO	0.656	DINO	0.6009	DINO	0.5814	ACDM	0.2781
M7	LLM	0.407	GDINA		GDINA		ACDM	0.9992
M8	LLM	0.999	LLM	0.1509	LLM	0.1458	GDINA	
M9	DINO	0.118	DINO	0.1181	DINO	0.1037	DINO	0.7713
M10	GDINA		GDINA		GDINA		GDINA	
M11	GDINA		GDINA		GDINA		GDINA	
M12	GDINA		GDINA		GDINA		GDINA	
M13	LLM	0.643	LLM	0.6203	LLM	0.6029	ACDM	0.9034
M14	GDINA		GDINA		GDINA		GDINA	
M15	GDINA		GDINA		GDINA		GDINA	
M16	GDINA		GDINA		GDINA		GDINA	
M17	DINA	0.210	DINA	0.1432	DINA	0.0999	DINA	0.3699
M18	DINO	0.123	DINO	0.3169	DINO	0.0896	DINA	0.1553
M19	LLM	0.958	DINA	0.2230	LLM	0.3249	DINA	0.1092
M20	GDINA		GDINA		GDINA		GDINA	
M21	LLM	0.413	LLM	0.3905	LLM	0.3721	DINA	0.4779
M22	DINA	0.197	GDINA		GDINA		ACDM	0.2497
M23	A-CDM	0.435	LLM	0.4192	LLM	0.3838	DINA	0.0642
M24	GDINA		GDINA		GDINA		GDINA	
M25	GDINA		GDINA		GDINA		GDINA	
M26	R-RUM	0.969	RRUM	0.9747	RRUM	0.9656	DINA	0.6228

\*Tüm modellerde düzeltilmiş p değeri 1'dir