

**Bilişsel Tanı Çerçevesinde GDINA, DINA, DINO Model Uyumlarının Madde ve Test Düzeyinde Gerçek Veriye Dayalı Olarak Karşılaştırılması\***

**Comparison of GDINA, DINA, DINO Model Fit in Cognitive Diagnosis at Item and Test Level Based on Real Data**

Seyhan SARITAŞ AKYOL<sup>1</sup>, Mehtap ÇAKAN<sup>2</sup>

<sup>1</sup> Gazi Üniversitesi, Gazi Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı. seyhansaritasakyol@gmail.com

<sup>2</sup> Gazi Üniversitesi, Gazi Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı. cakanmehtap@hotmail.com

*Makale Türü/Article Types: Araştırma Makalesi/Research Article*

*Makalenin Geliş Tarihi: 07.10.2021*

*Yayına Kabul Tarihi: 07.02.2022*

**ÖZ**

Bilişsel tanı modelleri belirli kategorik gizil özellikler aracılığıyla bireylerin nitelik profillerinin oluşturulmasına imkân vermektedir. Bu modellerden elde edilen çıkarımların geçerli olup olmadığına yönelik kanıt toplamak için model veri uyumunun ortaya konması ve uygun modellerin belirlenmesi büyük önem taşımaktadır. Bu çalışmada, bilişsel tanı yaklaşımı kapsamında belirlenen doymuş GDINA modeli yerine indirgenmiş (reduced) DINA ve DINO modellerinin çalışma verisi için kullanılıp kullanılmayacağı her madde düzeyi için Wald istatistiğinin kullanılarak belirlenmesi ve veriye en iyi uyum sağlayan modelin madde ve test düzeyinde karşılaştırılarak belirlenmesi amaçlanmıştır. Veri seti olarak 478 öğrenciden elde edilen, 36 maddeden oluşan Sayı Karşılaştırma testine ait cevap örüntüsü kullanılmıştır. Analizler R 3.5.3 programında yürütülmüş R programındaki GDINA paketinden yararlanılmıştır. Çalışmada ilk olarak, temele doymuş GDINA modeli alınarak mutlak ve göreceli uyum indeksleri birlikte değerlendirilmiş ve doymuş modelin en iyi uyumu verdiği görülmüştür. Devamında Wald istatistiği kullanılarak doymuş GDINA modeli yerine her madde düzeyi için model veri uyumunda manidar bir kayıp olmadan indirgenmiş DINA ve DINO modellerinin kullanılabilirliği incelenmiştir.

---

\***Alıntı:** Sarıtaş Akyol, S. ve Çakan, M. (2022). Bilişsel tanı çerçevesinde GDINA, DINA, DINO model uyumlarının madde ve test düzeyinde gerçek veriye dayalı olarak karşılaştırılması. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 42(2), 1083-1108.

**Anahtar Sözcükler:** *Wald testi, Model karşılaştırma, Model uyumu, BTM*

**ABSTRACT**

*Cognitive diagnostic models enable the creation of individuals' attribute profiles through particular categorical latent variables. In order to collect evidence of whether the inferences drawn from these models are valid or not, it is of great importance to reveal the model-data fit and to identify the models that fit. This study aims to determine whether reduced DINA and DINO models can be used instead of the saturated GDINA model determined within cognitive diagnostic modelling framework by using Wald statistics for each item level and to specify the model that best fits the data by comparing it at the item and test level. The data set of the study was obtained from 478 students and the answer pattern of the Symbolic Number Comparison test comprising 36 items. Analyses were performed in R 3.5.3 program and the GDINA package in the R program was used. In the first stage, the saturated GDINA model was taken as the basis, and the absolute and relative fit indices were evaluated together and it was seen that the saturated model gave the best fit. Afterward, using Wald statistics, instead of the saturated GDINA model, the availability of reduced DINA and DINO models for each item-level with no significant loss in model-data fit was examined.*

**Keywords:** *Wald test, Model comparison, Model fit, CDM*

## GİRİŞ

Son yirmi yıl sürecinde eğitimdeki değerlendirme yaklaşımı, eğitimin paydaşlarına tek bir genel test puanı raporlamaktan ziyade öğrencilerin belirli beceriler üzerindeki performansları ile ilgili ayrıntılı bilgi sağlamak şeklinde değişmektedir (Sen ve Bradshaw, 2017). Özellikle 2001 yılında Amerika’da çıkarılan “No Child Left Behind” yasası eğitimde yapılacak olan reformlarda belirleyici rol oynamıştır (Jiao, 2009; Leighton ve Gierl, 2007). Bu yasa, öğrencilerin yıllık olarak yeterli bir ilerleme kaydetmelerini ve sonunda yetkin bir seviyeye ulaşmalarını gerektirmektedir (Jiao, 2009). Dolayısıyla öğrenci, öğretmen ve ailelerin eğitimde yapılan değerlendirmelere ilişkin derinlemesine bilgilendirilme beklentisi, Amerika başta olmak üzere giderek daha fazla benimsenmiştir (Leighton ve Gierl, 2007). Bilişsel tanı modelleri, öğrenme ve öğretime yardımcı olmada zengin tanılayıcı bilgi sağlayarak bu bağlamda ümit olmuştur (Rupp ve Templin, 2008). Zira bilişsel tanı modelleri, temel olarak bireylerin iyi tanımlanmış niteliklere sahip olup olmama durumlarına ilişkin birey hakkında çıkarımlar yapılmasına imkân tanıyabilmektedir. Eğitimsel ölçmeden doğan bu modellere özellikle bireylerin güçlü ve geliştirmesi gereken yönleri ortaya çıkarılmak istendiğinde veya detaylı geri bildirim beklediği durumlarda başvurulmaktadır.

Bilişsel tanıya dayalı modellemenin altında yatan bilişsel yapı Tatsuoka (1983) tarafından geliştirilen Q-matrisi ile tanımlanmaktadır. İstatistiksel bir yaklaşımla Q-matrisi, hangi maddenin hangi gizil değişkenle ilişkili olduğunu gösteren yük matrisi veya örüntü matrisidir. Q-matrisi doğrusal faktör analizi modelleri ve doğrulayıcı madde tepki teorisi de dâhil olmak üzere, tüm doğrulayıcı istatistiksel modellerde kullanılır (Rupp, Templin ve Henson, 2010, s.55). Q matrisinde her bir madde tepkisinin altında yatan nitelik ya da nitelikler arasındaki ilişki alan uzmanları tarafından belirtilerek matrisin girdileri oluşturulur. Tipik bir Q matrisinde satırlarda maddeler, sütunlarda nitelikler temsil edilir.  $J, K \in \mathbb{Z}^+$  olmak üzere Q matrisi  $J \times K$  tipinde  $Q=(q_{j,k})$  şeklinde tanımlandığı varsayımında, j. maddenin cevaplanması k. niteliği gerektiriyorsa 1, j. maddenin cevaplanması k. niteliği gerektirmiyorsa 0 olarak

kodlanır. Dolayısıyla K tane niteliğin bulunduğu Q matrisinde en fazla  $2^k$  tane olası gizil nitelik sınıfı oluşabilir. Bu durum sadece niteliklerin ikili kodlanması durumunda gerçekliğini korumaktadır.

Alan yazında çok sayıda bilişsel tanı modeli önerilmesine karşın sadece sınırlı sayıda uygulama gerçekleştirilebilmiştir (Jiao, 2009). DINA modeli en yaygın bilinen modellerden olmasına karşın Demir ve Koç (2018) DINA modelin gelişimi ve uygulanabilirliğine ilişkin gerçek verilere dayalı yapılan çalışmaların sınırlı olduğunu belirtmiştir. Bununla birlikte, gerçek verileri analiz etmek için bilişsel tanı modellerinin (BTM) kullanıldığı bazı çalışmalar bulunmaktadır (Başokçu, 2012; Başokçu, Öğretmen ve Kelecioğlu, 2013; Bradshaw, Izsak, Templin ve Jacobson, 2014; Demir ve Koç, 2018; Koyuncu, 2020, Lee, Park ve Taylan, 2011; Templin ve Henson, 2006; von Davier, 2008) ancak bunlar henüz pratikte yaygın olarak kullanılmamaktadır.

Bu durumun, BTM'de model uyum literatüründeki eksikliklerden kaynaklanıyor olabileceği vurgulanmaktadır (Hu, Miller, Huggins-Manley ve Chen, 2016). Bilişsel tanı modellerinden elde edilen yorum ve çıkarımların geçerliğine ilişkin kanıt toplamak adına araştırmacının uygun modeli belirlemesi ve veri uyumunun ortaya koyması kullanılan modellerin yararlılığı açısından büyük önem taşımaktadır. Ancak de la Torre ve Lee (2013) araştırmacının belirlediği tercih dışında, kullanılacak bilişsel tanı modellerinden en uygun modeli seçmenin en doğru yolunun net olmadığını ifade etmiştir. Han ve Johnson (2019) çalışmalarında modelin yanlış tanımlanmasının veya model uyumsuzluğunun türü ne olursa olsun, BTM kullanıcılarının model-veri uyumsuzluğunu çeşitli açılardan araştırmak için araçlara ihtiyacı olduğunu belirtmişlerdir. Bahsedilen durumlardan dolayı eğitimde test uygulayıcıları, çok sayıda modelden hangi modeli nasıl seçeceği hususunda zorluk yaşayabilmektedir (Jiao, 2009). Bu durum, bilişsel tanı yaklaşımında alan yazındaki gelişmelerin eğitimdeki uygulamalara verimli olarak yansıtılmaması ve pratikteki yararlılığının sınırlı kalması sonucunu doğurmaktadır. Bunun yanı sıra birçok farklı bilişsel tanı modelleri formülasyonun farklı model sınıflarını temsil edip etmediği veya modellerin birbirleriyle ne ölçüde ilişkili olduğu tam olarak açık olmamaktadır. Bu farklı

formülasyonlarda model parametrelerinin birlikte kestirilmesi ve bunların madde düzeyinde görelî uyumlarını karşılaştırmak zor görevler olarak kalmıştır (de la Torre, 2011).

Uygulamadan gelen gerçek verilere dayalı bilişsel tanı analizine ilişkin araştırma sayısının artması şüphesiz ki bu konuda çalışan araştırmacılara ve eğitim uygulayıcılarına ışık tutacaktır. von Davier ve Lee'ye göre (2019) son zamanlardaki gerçek veri analizleri dikkate alındığında farklı maddeler hem indirgenmiş (reduced) hem de doymuş (saturated) olmak üzere farklı BTM türlerini gerektirebilmektedir. Bu çalışmada uygun modelin hem test düzeyinde hem de madde düzeyinde belirlenebilmesi adına GDINA modeli kullanılmıştır. Çünkü genel bir BTM çerçevesi olarak GDINA, model formülasyonuna ek olarak her bir madde için tasarım (design) ve ağırlık matrislerine dayalı model kestirimi için bir bileşen (component) ve her bir madde için Wald testine dayalı model karşılaştırma için bir bileşen içermektedir (de la Torre, 2011). Bu çalışma, GDINA çerçevesinde gerçek veri üzerinde tek bir testte DINA ve DINO modelleri gibi birden fazla BTM parametresi kestirmeye, kısıtlanmış (constrained) ve kısıtlanmamış (unconstrained) BTM'nin test düzeyinde mutlak ile görelî uyumlarının değerlendirilmesine ve her madde düzeyinde Wald testine dayalı uygun modelin belirlenmesine imkân sağlamaktadır. Bu çalışmada bilişsel tanı yaklaşımı ile belirlenen doymuş GDINA modeli yerine indirgenmiş DINA ve DINO modellerinin kullanılıp kullanılmayacağı ve veriye en iyi uyum sağlayan modelin madde ve test düzeyinde karşılaştırılarak belirlenmesi amaçlanmıştır.

### **Problem İfadesi**

Bilişsel tanı çerçevesinde GDINA, DINA ve DINO modellerinin, gerçek veriye dayalı olarak model uyumları test düzeyinde nasıl değişiklik göstermektedir ve madde düzeyinde veriye en iyi uyum sağlayan model ya da modeller hangileridir?

### **Araştırma Soruları**

Bu çalışmada aşağıdaki sorulara yanıt aranmıştır:

1. GDINA, DINA ve DINO modelleri için test düzeyinde mutlak uyum istatistikleri nelerdir?
2. GDINA, DINA ve DINO modellerinden elde edilen görelî uyum indeksleri incelendiğinde hangi model veriye daha iyi uyum sağlamaktadır?
3. Wald istatistiği kullanılarak bakıldığında madde düzeyinde hangi indirgenmiş model veya modeller veriye daha iyi uyum göstermiştir?

Çalışma kapsamında kullanılan DINA, GDINA, DINO modellerine kısaca yer verilmiştir.

### DINA Model

DINA (Deterministic Input; Noisy “And” Gate) modeli hem maddelerin bilişsel bilgisi hem de bireylerin bilişsel özellikleri hakkında çıkarımlara izin veren, bilişsel tanı ve değerlendirmeye yönelik çeşitli yaklaşımlara temel olmuş bir modeldir (de la Torre, 2008). Bunun yanı sıra DINA model olasılıksal bir bağlayıcı modeldir (de la Torre ve Douglas, 2004). Bireyin bir maddeyi doğru cevaplayabilmesi için Q matrisinde o madde için belirtilen tüm niteliklere sahip olması gerekli olurken, niteliklerden sadece birkaçına sahip olan veya niteliklerden hiçbirine sahip olmayan bireylerin soruyu doğru yanıtlama olasılıkları birbirine eşit olmaktadır. Modelin olasılıksal yönü ise maddeyi doğru yanıtlamak için gerekli tüm niteliklere sahip olduğunda doğru cevabı garanti etmemesi ve niteliklerden tamamen yoksun olduğunda da yanlış cevabı garanti etmemesi ile açıklanabilir. Modelde örtük özellik ile gözlenen cevap örüntüsü arasındaki bağlantı şu şekildedir (Haertel, 1989; Junker ve Sijstma, 2001; Macready ve Dayton, 1977):

$$\xi_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}} \quad (\text{Eşitlik 1})$$

Eşitlik 1’de  $i \in I = (1, \dots, I)$  olmak üzere bireyi,  $j \in J = (1, \dots, J)$  olmak üzere maddeyi,  $k \in K = (1, \dots, K)$  niteliği belirtmektedir.

$\xi_{ij}$  = i. bireyinin j. maddesine ilişkin örtük cevap vektörü

$q_{jk}$  = j. maddenin k. niteliği için Q matrisi girişi

Eğer i. birey k. niteliğe sahipse  $\alpha_{ik} = 1$ , sahip değilse  $\alpha_{ik} = 0$  olacaktır.

DINA model için gerekli olan bir diğer eşitlik, maddenin doğru yanıtlanabilmesi için gerekli niteliklere bağlı olan doğru cevaplama olasılığı, aşağıdaki koşullu olasılık formülüdür:

$$P(X_{ij} = 1 | \xi_{ij}) = (1 - s_j) \xi_{ij} g_j (1 - \xi_{ij}) \quad (\text{Eşitlik 2})$$

Eşitlik 2’de yer alan formül her madde için kestirilen kaydırma (slip,  $s_j$ ) ve tahmin (guess,  $g_j$ ) parametrelerini içermektedir. Genel olarak kaydırma parametresi bireyin maddeyi doğru yanıtlamak için gerekli tüm niteliklere sahipken yanlış yanıt verme olasılığını, tahmin parametresi ise bireyin maddeyi doğru yanıtlamak için gerekli nitelikler arasından en az birine sahip değilken maddeyi doğru yanıtlama olasılığını temsil etmektedir.

### G-DINA Model

DINA modelde kullanılan iki olasılık sınırlaması -maddenin doğru yanıtlanması için o maddeye ilişkin belirtilen tüm niteliklere sahip olunması gerekliliği ve niteliklerin hiçbirine sahip olmayan bireyler ile birkaçına sahip olan bireylerin soruyu doğru yanıtlama olasılıklarının birbirine eşit olma durumu- gereksiz yere kısıtlayıcı olabilir (Rupp, vd. 2010). Çünkü model maddenin doğru cevaplanması için Q matrisinde gerekli niteliklerden an az birinin sağlanmaması ile hiçbirinin sağlanmaması durumunda aynı olasılığı kestirmektedir. Dolayısıyla DINA model genişletilerek bu durumun üstesinden gelen G-DINA (Generalized Deterministic Input; Noisy “And” Gate) model oluşturulmuştur. G-DINA modelde temel alınan formül doğası gereği bireyin maddeyi doğru yanıtlama olasılığını maddenin doğru yanıtlanması için gerekli olan her bir niteliğe sahip olma durumuna göre ayrı kestirmekte ve sahip olunan niteliklerin birbirleriyle etkileşimi bireyin doğru yanıtlama olasılığının farklılaşmasına izin vermektedir (de la Torre, 2011). Eşitlik 3’te G-DINA modelde kullanılan ana formüle yer verilmiştir:

$$P(\alpha_{ij}^*) = \delta_{j0} \sum_{k=1}^{K_j^*} \delta_{jk} \alpha_{ik} + \sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'} \alpha_{ik} \alpha_{ik'} \dots + \delta_{j12\dots K^*j} \prod_{k=1}^{K_j^*} \alpha_{ik}$$

(Eşitlik 3)

$\delta_{j0}$  = j maddesinin kesişimi

$\delta_{jk}$  =  $\alpha_k$  'dan kaynaklanan temel etki

$\delta_{jkk'}$  =  $\alpha_k$  ve  $\alpha_{k'}$  'dan kaynaklanan etkileşim etkisi

$\delta_{j12...K*j}$  =  $\alpha_1, \dots, \alpha_{K*j}$  'dan kaynaklanan etkileşim etkisi

### DINO Model

DINO (Deterministic Input; Noisy “Or” Gate) model bağlayıcı olmayan (disjunctive) özelliği dolayısıyla DINA modelindeki durumun aksine bireyin soruyu doğru yanıtlama olasılığı, Q matrisinde belirlenen niteliklerin tamamına veya en az birine sahipse yüksek olmaktadır. Birey ancak soruya doğru cevap vermek için gerekli hiçbir niteliğe sahip değilse bireyin soruyu doğru yanıtlama olasılığı düşük kestirilmektedir.

DINA modeldeki  $\xi_{ij}$  gizil değişkeni DINO model için  $\omega_{ij}$  notasyonu kullanılarak;

$$\omega_{ij} = 1 - \prod_{k=1}^K (1 - \alpha_{ik}) \quad \text{qjk}$$

DINO model için doğru yanıt olasılığı için eşitlik 4'te oluşturulmuştur (Templin ve Henson, 2006):

$$P(X_{ij} = 1 | \omega_{ij}) = (1 - s_j) \omega_{ij} g_j (1 - \omega_{ij}) \quad \text{(Eşitlik 4)}$$

Bu durumda bireyler, soruyu doğru yanıtlamak için Q matrisinde tanımlanan niteliklerden en az birine sahip olan bireyler  $\omega_{ij} = 1$  grubuna ait olurken Q matrisinde tanımlanan niteliklerden hiçbirine sahip olmayan bireyler  $\omega_{ij} = 0$  grubuna ait olarak iki gruba ayrılmaktadır. DINO modelde de DINA model de olduğu gibi her madde için kaydırma ve tahmin parametreleri olmak üzere iki parametre elde edilmektedir.



## YÖNTEM

### Araştırma Modeli

Çalışma bilişsel tanı yaklaşımı ile belirlenen doymuş GDINA modeli yerine indirgenmiş DINA ve DINO modellerinin kullanılıp kullanılmayacağı ve veriye en iyi uyum sağlayan modelin madde ve test düzeyinde karşılaştırılarak belirlenmesini amaçlayan betimsel bir araştırmadır.

### Çalışma Grubu

Çalışma grubunu TÜBİTAK kapsamında 15.03.2012 ile 15.03.2015 tarihleri arasında yürütülen “6 -11 Yaş Türk Çocukları Örnekleminde Diskalkuliye Yatkınlığı Ayırt Etmede Kullanılacak Bir Ölçme Aracı Geliştirme Çalışması” isimli projede katılımcı olan 1., 2., 3. ve 4. sınıflarda normal öğrenim gören toplam 478 öğrenci oluşturmaktadır. Çalışma grubunun cinsiyet ve sınıf düzeyine göre dağılımı Tablo 1’de verilmiştir.

**Tablo 1.** Çalışma Grubunun Cinsiyet ve Sınıf Düzeyine Göre Dağılımı

Sınıf Düzeyleri	Cinsiyet					
	Kız		Erkek		Toplam	
	N	%	N	%	N	%
1-sınıf	59	12.34	60	12.55	119	24.89
2-sınıf	60	12.55	60	12.55	120	25.1
3-sınıf	61	12.76	59	12.34	120	25.1
4-sınıf	59	12.34	60	12.55	119	24.89
<b>Toplam</b>	239	50	239	50	478	100

**Etik Kurallara Uygunluk**

Bu araştırmada kullanılan veri, TÜBİTAK kapsamında 111K545 proje numarası ile gerçekleştirilen “6-11 Yaş Türk Çocukları Örnekleminde Diskalkuliye Yatkınlığı Ayırt Etmede Kullanılacak Bir Ölçme Aracı Geliştirme Çalışması” başlıklı proje çalışmasından sağlanmıştır. Proje kapsamında yürütülen araştırmanın amacına uygun olarak belirlenen örneklem için il milli eğitim müdürlüğünden gerekli uygulama izinleri alınmıştır (Olkun, 2015). Dolayısıyla bu araştırma için etik kurul başvurusuna gerek olmamıştır. Araştırma sürecinde kullanılan kaynaklara atıf gösterme kurallarına dikkat edilerek atıf yapılmış ve ilgili kaynaklar eksiksiz olarak sunulmuştur.

**Veri Toplama Aracı ve Verilerin Analizi**

Araştırmada Olkun'un (2015) çalışmasında geliştirilen temel sayı işleme performansını ölçmek amacıyla geliştirilen test kapsamındaki Sayı Karşılaştırma (symbolic number comparison, SNC) alt testine ait cevap örüntüsü kullanılmıştır. Kullanılan ölçme aracı tablet üzerinden uygulanmış olup, toplam 36 maddeyi içermektedir. Maddeler doğru (1) ile yanlış (0) olmak üzere ikili puanlanmıştır. Sayısal Stroop paradigmasına göre tasarlanan görevlerden oluşturulan bu sembolik sayı karşılaştırma (SNC) testinde, öğrencilere 3 ile 9 arasında değişen sayı çiftleri rasgele biçimde sunularak öğrencilerden ekran üzerinde sayısal olarak büyük olan sayıya dokunmaları istenmiştir. Burada öğrencilerden fiziksel bir karşılaştırma görevi beklenmemiştir. Görevlerdeki sayılar 3 farklı şekilde sunulmuştur; (1) uyumlu yani sayısal olarak büyük olan sayının fiziksel olarak da büyük sunulması, (2) nötr yani sayıların aynı fiziksel büyüklükte sunulması (3) uyumsuz ise sayısal olarak büyük olan sayının fiziksel olarak daha küçük sunulması şeklindedir. Testte 12 uyumlu, 12 nötr ve 12 uyumsuz olmak üzere toplam 36 madde yer almıştır. Bilişsel tanı analizi için gerekli olan Q matrisinin tanımlanması için gerekli olan niteliklerin belirlenmesi amacıyla konu alan uzmanlarından görüşler alınarak bu testte yer alan nitelikler aşağıdaki şekilde sıralanmıştır:

1. Sayı değeri ve sembol büyüklükleri uyumlu iken bireyin sayıları karşılaştırabilmesi, (N1)
2. İki sayı değerinde de semboller eşit büyüklükte iken bireyin sayıları karşılaştırabilmesi, (N2)
3. Sayı değeri ve sembol büyüklükleri uyumsuz iken bireyin sayıları karşılaştırabilmesi, (N3)
4. Sayılar arası uzaklık 1'e eşit iken bireyin sayıları karşılaştırabilmesi, (N4)
5. Sayılar arası uzaklık 1'den büyük iken bireyin sayıları karşılaştırabilmesi, (N5).

Bir bilişsel tanı modelinde Q matrisinin yetersiz tanımlanması, analizde tanılayıcı ve sınıflandırıcı değeri düşük olan model parametrelerinin elde edilmesiyle sonuçlanabilir (Templin ve Henson, 2006). Dolayısıyla, Q matrisinin geçerliğine yönelik kanıtlar elde etmek bilişsel tanı uygulamalarında yapılacak olan çıkarımların geçerliği için bir ön koşul niteliğindedir. Bu çalışmada oluşturulan Q matrisinin geçerliğine yönelik kanıt elde etmek için uzman görüşüne başvurulmuştur. Birbirinden bağımsız uzman grubundan her bir maddenin ölçtüğü nitelik ya da niteliklerin saptanması beklenmiştir. Uzman grubu üyeleri, hem matematik öğretmenliği niteliği taşıyan hem de eğitimde ölçme ve değerlendirme alanında uzman ve doktor düzeyinde olan beş kişiden oluşturulmuştur. Uzmanların görüş birliği sonucunda Q matrisi oluşturulmuş olup Tablo 2' de sayı karşılaştırma alt testine ilişkin Q matrisine yer verilmiştir. Tablo 2' ye göre her maddeye karşılık gelen satırda, öğrencinin maddeyi doğru cevaplaması için gerekli olan nitelik ya da nitelikler "1", gerekli olmayan nitelikler ise "0" olarak kodlanmıştır.

**Tablo 2.** Sayı Karşılaştırma Alt Testine İlişkin Q Matrisi

Maddeler	Nitelikler					Maddeler	Nitelikler				
	N1	N2	N3	N4	N5		N1	N2	N3	N4	N5
1	0	0	1	0	1	19	1	0	0	1	0
2	0	0	1	0	1	20	0	0	1	1	0
3	0	1	0	0	1	21	0	1	0	0	1
4	0	1	0	0	1	22	0	1	0	0	1
5	1	0	0	1	0	23	1	0	0	0	1
6	0	0	1	0	1	24	0	1	0	1	0
7	1	0	0	0	1	25	1	0	0	1	0
8	1	0	0	0	1	26	0	1	0	0	1
9	0	0	1	0	1	27	1	0	0	1	0
10	0	0	1	1	0	28	0	0	1	0	1
11	1	0	0	0	1	29	1	0	0	0	1
12	0	0	1	0	1	30	0	1	0	1	0
13	0	1	0	1	0	31	0	0	1	1	0
14	0	1	0	1	0	32	0	1	0	0	1
15	0	0	1	0	1	33	0	1	0	0	1
16	0	0	1	1	0	34	1	0	0	0	1
17	0	1	0	0	1	35	1	0	0	0	1
18	0	1	0	0	1	36	0	0	1	0	1

Tablo 3'te Sayı Karşılaştırma alt testinde yer alan maddelerin niteliklerine ilişkin dağılım yer almaktadır. Hartz, Roussos ve Stout (2002)'e göre tanılayıcı olarak güvenilir bilgi elde edebilmek için bir niteliğin en az üç maddenin doğru cevaplanabilmesi için gerekli olması durumu söz konusudur. Tablo 3'e bakıldığında her bir niteliğin en az 11 en çok 24 madde ile ölçüldüğü görülmektedir.

**Tablo 3.** Sayı Karşılaştırma Alt Test Maddelerinin Niteliklere İlişkin Dağılımı

Nitelikler	Niteliği Ölçen Maddeler	N
N1	5, 7, 8, 11, 19, 23, 25, 27, 29, 34, 35	11
N2	3, 4, 13, 14, 17, 18, 21, 22, 24, 26, 30, 32, 33	13
N3	1, 2, 6, 9, 10, 12, 15, 16, 20, 28, 31, 36	12
N4	5, 10, 13, 14, 16, 19, 20, 24, 25, 27, 30, 31	12
N5	1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 15, 17, 18, 21, 22, 23, 26, 28, 29, 32, 33, 34, 34, 36	24

Model Uyumu için çalışmada GDINA, DINA ve DINO modellerine ilişkin kestirilen ve model veri uyumsuzluğunu belirlemede güçlü bilgi sağlayan (Liu, Tian ve Xin, 2016)  $M_2$  istatistikleri ile modellerin serbestlik dereceleri ve  $p$  değerleri sunulmuştur. Maydeu-

Olivares ve Joe (2005, 2006) artık momentlerde sadece quadratik formlar olan test istatistiği ailesi  $M_2$ 'yi önermişlerdir. Çalışmada kullanılan modeller  $M_2$  istatistiğine dayalı kestirilen  $RMSEA_2$  (limited information root mean square error of approximation) değerleri %90 güven aralığı ile birlikte raporlanmıştır. GDINA, DINA ve DINO modelleri mutlak uyum değerlendirmesi için  $RMSEA_2$  değerlerinin .05'ten küçük olması dikkate alınmıştır (Kunina-Habenicht, Rupp ve Wilhelm, 2009). Modellerin görelî karşılaştırılmasına ilişkin en çok kullanılan ve programın sağladığı Akaike Bilgi Kriteri (Akaike's Information Criterion, AIC), Bayes Bilgi Kriteri (Bayesian Information Criterion, BIC), Tutarlı Akaike Bilgi Kriteri (Consistent AIC, CAIC) ile SABIC, (Sample size Adjusted BIC) değerleri kullanılmıştır. Loglikelihood değerlerine dayalı hesaplanan görelî model uyum istatistiklerinden (Hu vd., 2016) daha düşük olanın daha iyi model uyumu sağladığı düşünülür. GDINA modelinin DINA ve DINO modelleri ile görelî karşılaştırılması için de olabilirlik oran testi ile tek yönlü varyans analizi yapılmıştır. Model uyumunun madde düzeyinde ele almak için çalışmada her madde düzeyi için Wald istatistiği hesaplanarak doymuş GDINA modeli yerine model veri uyumunda manidar bir kayıp olmadan indirgenmiş DINA ve DINO modellerinin kullanılabilme durumu incelenmiştir. Çalışmada her madde için hesaplanan Wald istatistiğine ilişkin  $p$  değerleri raporlanmıştır. Model seçme kuralı olarak "daha basit" uygulanmıştır. Bu durumda model uyumunu sağlayan daha basit modellerden  $p$  değeri daha yüksek olan model veriye daha iyi uyum sağlamaktadır (Ma ve de la Torre 2020).

Araştırma verilerinin analizi R Studio arayüzü kullanılarak R (version 3.5.3) programı ile yürütülmüş ve R programındaki GDINA (Ma ve de la Torre, 2020) paketinden faydalanılmıştır. Pakette kullanılan ana fonksiyonlar `autogdina` ve `gdina` fonksiyonu olarak özetlenebilir. G-DINA model parametreleri marjinal en çok olabilirlik kestirimi kullanılarak elde edilmiştir.



## BULGULAR

Çalışmada ilk araştırma sorusu olarak “GDINA, DINA ve DINO modelleri için test düzeyinde mutlak uyum istatistikleri nelerdir?” sorusuna yanıt aranmıştır. Tablo 4’ te sırasıyla GDINA, DINA ve DINO modellerine ilişkin kestirilen ve model veri uyumsuzluğunu belirlemede güçlü bilgi sağlayan (Liu, Tian ve Xin, 2016)  $M_2$  istatistikleri ile modellerin serbestlik dereceleri ve  $p$  değerleri sunulmuştur. Çalışmadaki modeller için  $M_2$  istatistiğine dayalı kestirilen  $RMSEA_2$  değerleri de %90 güven aralığı ile birlikte raporlanmıştır. GDINA, DINA ve DINO modelleri için sırasıyla  $RMSEA_2$  değerleri 0.029, 0.032 ve 0.033 olup bu değerlerin tümü .05 değerinin altında kalmaktadır. Bu durum belirtilen üç modelin de veriye iyi uyum sağladığına işaret etmektedir (Kunina-Habenicht, Rupp ve Wilhelm, 2009).

**Tablo 4.** Modellere İlişkin Mutlak Uyum Değerleri

	$M_2$	$sd$	$p$	$RMSEA_2$	90 % CI
<b>GDINA</b>	683.11	491	0.00*	0.029	[0.023, 0.034]
<b>DINA</b>	834.42	563	0.00*	0.032	[0.027, 0.036]
<b>DINO</b>	863.35	563	0.00*	0.033	[0.029, 0.038]

\* $p < 0.05$  Not:  $sd$  = serbestlik derecesi CI= Confidence Intervals.

Çalışmanın ikinci araştırma sorusunda “GDINA, DINA ve DINO modellerinden elde edilen görel uyum indeksleri incelendiğinde hangi model veriye daha iyi uyum sağlamaktadır?” sorusuna yanıt aranmıştır. Buna göre belirtilen modellere ilişkin elde edilen AIC, BIC, CAIC ile SABIC değerleri Tablo 5’ te sunulmuştur.

**Tablo 5.** Modellere İlişkin Görel Uyum İndeksleri

	Loglikelihood	AIC	BIC	CAIC	SABIC
<b>GDINA</b>	-3739.99	7829.97	8559.66	8734.66	8004.23
<b>DINA</b>	-3835.32	7876.64	8306.11	8409.11	7979.20
<b>DINO</b>	-3837.58	7881.15	8310.62	8413.62	7983.71

Loglikelihood değerlerine dayalı hesaplanan görel model uyum istatistiklerinden (Hu, vd. 2016) daha düşük olanın daha iyi model uyumu sağladığı düşünülür. Buna göre modeller arasında AIC indeksi için GDINA modelinin daha iyi uyum gösterdiği görülürken; BIC, CAIC ve SABIC indeksleri için DINA modelinin daha iyi uyum

gösterdiği görülmektedir. GDINA modelinin DINA ve DINO modelleri ile karşılaştırılması için olabilirlik oran testi kullanılarak yapılan tek yönlü varyans analizine ilişkin bulgular da Tablo 6' da sunulmuştur.

**Tablo 6.** GDINA Modelinin DINA ve DINO Modelleri ile Karşılaştırılması

	-2Loglikelihood	Ki-kare	sd	p değeri
<b>GDINA</b>	7479.97			
<b>DINA</b>	7670.64	190.66	72	<0.001*
<b>DINO</b>	7675.15	195.18	72	<0.001*

\*p<0.05

Burada yapılan iki karşılaştırma için de  $H_0$  hipotezi, doymuş modelden indirgenmiş modele geçişin manidar bir farklılık oluşturmadığı anlamını taşımaktadır. Ancak tutumluluk ilkesi gereği bu durumda, indirgenmiş modelin doymuş modelden istatistiksel olarak daha iyi olduğu anlamı taşıdığı için daha basit model tercih edilmektedir. Tablo 6' da yer alan  $p$  değerleri <0.05 olduğundan iki karşılaştırma için de  $H_0$  hipotezi reddedilmiştir. Yani olabilirlik oran testi kullanılarak yapılan tek yönlü varyans analizi dikkate alındığında GDINA modeli DINA ve DINO modellerine göre veriye daha iyi uyum sağlamaktadır.

Çalışmanın üçüncü araştırma sorusunda “Wald istatistiği kullanılarak madde düzeyinde hangi indirgenmiş model veya modeller veriye daha iyi uyum göstermiştir?” sorusuna yanıt aranmıştır. Burada her madde düzeyi için Wald istatistiği kullanılarak doymuş GDINA modeli yerine model veri uyumunda manidar bir kayıp olmadan indirgenmiş DINA ve DINO modellerinin kullanılabilme durumu belirlenmiştir. Tablo 7' de her madde düzeyi için hesaplanan Wald istatistiğine ilişkin  $p$  değerleri sunulmuştur. Model seçme kuralı olarak “daha basit” uygulanmıştır. Dolayısıyla model uyumunu sağlayan daha basit modellerden  $p$  değeri daha yüksek olan modelin veriye daha iyi uyum sağladığı değerlendirilir.



**Tablo 7.** Madde Düzeyinde Wald İstatistiğine İlişkin Hesaplanan  $p$  Değerleri ve Seçilen Modeller

Maddeler	Seçilen model	$p$ değeri	Maddeler	Seçilen model	$p$ değeri
1	DINA	0.9132	19	DINO	0.9993
2	DINA	0.8515	20	DINA	0.6411
3	DINO	0.9638	21	DINA	0.9714
4	DINA	0.9862	22	DINA	0.9854
5	DINA	0.9815	23	DINA	0.9730
6	DINA	0.8297	24	DINA	0.7718
7	DINA	0.9799	25	DINA	0.9902
8	DINA	0.9852	26	DINA	0.9621
9	DINA	0.8255	27	DINA	0.8380
10	DINA	0.8479	28	DINA	0.9799
11	DINO	0.9521	29	DINA	0.9791
12	DINA	0.8394	30	DINO	0.9876
13	DINA	0.4904	31	DINA	0.6999
14	DINA	0.6161	32	DINO	0.9956
15	DINA	0.9238	33	DINA	0.9035
16	DINA	0.9319	34	DINO	0.9970
17	DINA	0.9812	35	DINO	1.0000
18	DINA	0.9388	36	DINO	0.7344

Maddelerden 3., 11., 19., 30., 32., 34., 35. ile 36. maddeler DINO modeli ile daha uyumlu iken diğer 28 maddenin DINA modelle ile daha iyi uyum gösterdiği görülmektedir.

## TARTIŞMA ve SONUÇ

Çalışmanın bulgularından elde edilen sonuca göre GDINA, DINA ve DINO modelleri için test düzeyinde mutlak uyum istatistikleri dikkate alındığında kullanılan üç model de veriye iyi uyum sağlamaktadır. Ancak Hu ve arkadaşları (2016) çalışmasında mutlak uyum istatistiğinin özellikle  $q$  matrisinin yanlış tanımlanma durumunu saptamak için kullanışlı olduğunu, BTM seçimi için uygun olmadığını belirtmiştir. Çünkü doğru bilişsel tanı modelinden bağımsız olarak, aşırı tanımlanmış  $q$  matrislere sahip doymuş modelin verilere uygun olduğu durumlar dışında,  $q$  matrislerde yanlış tanımlanmış herhangi bir modeli reddedecektir.

GDINA, DINA ve DINO modellerinden elde edilen görelî uyum indeksleri incelendiğinde, AIC indeksi için modeller arasında veriye GDINA modelinin daha iyi

uyum gösterdiği; BIC, CAIC ve SABIC indeksleri için veriye DINA modelinin daha iyi uyum gösterdiği görülmüştür. GDINA modelinin DINA ve DINO modelleri ile karşılaştırılması için olabilirlik oran testi kullanılarak yapılan tek yönlü varyans analizi sonucu doymuş GDINA modelinin veriye daha iyi uyum gösterdiği görülmüştür. Yani görece uyum indeksleri dikkate alındığında olabilirlik oran testi ve AIC indeksine göre GDINA modeli, DINA ve DINO modellerine göre veriye daha iyi uyum sağlamaktadır. BIC, CAIC ve SABIC indeksleri için DINA modelinin veriye daha iyi uyum gösterdiği görülmektedir. de la Torre (2011), ilgili niteliklerin tanımlandığı ve Q matrisinin doğru bir şekilde belirlendiği varsayımında doymuş modelin en iyi uyumu vereceğini belirtmiştir. de la Torre ve Lee (2013) genel bilişsel tanı modellerinin belirli (specific) bilişsel tanı modellerine kıyasla daha iyi model veri uyumu sağladığını ancak aynı zamanda daha karmaşık olduğunu ve bu nedenle doğru bir kestirim yapılabilmesi için daha büyük örneklem gerektirdiğini belirtmiştir. Başokçu, Öğretmen ve Kelecioğlu (2013) çalışmasında DINA ile GDINA modellerini büyük örnekleme model veri uyumu açısından karşılaştırmış ve GDINA modelinin DINA modeline göre veriye daha iyi uyum sağladığı görülmüştür. Rojas, de la Torre ve Olea (2012), özellikle örneklem büyüklüğü küçük olduğunda, genel bir BTM ile karşılaştırıldığında doğru belirli BTM'leri kullanarak, daha yüksek doğru sınıflandırma oranları elde edilebileceğini göstermiştir. Mevcut çalışmada çalışma grubunun 476 kişiden oluşması görece uyum indekslerinin farklılaşmasını etkilemiş olabileceği düşünülmektedir.

Alan yazında BTM' de model uyumlarının değerlendirilmesine ilişkin yapılan birtakım çalışmalar bulunmaktadır (de la Torre ve Lee, 2013; Hu vd., 2016; Ma, Iaconangelo ve de la Torre, 2016). Hu, Miller, Huggins-Manley ve Chen (2016) araştırmalarında belirledikleri çeşitli koşullarda görece uyum indekslerinin çoğu koşulda doğru DINA ve GDINA modellerini seçebildiğini göstermişlerdir. Buna karşın Q matrisinin yetersiz tanımlanması veya yanlış tanımlanması durumlarında görece uyum indeksleri doğru DINA ve GDINA modellerini reddetmiştir. Li, Cohen, Kim ve Cho (2009) simülasyon çalışmalarında dikkate aldıkları modeller için BIC değeri çoğu koşulda doğru modeli seçerken, AIC değeri bazı koşullar altında daha karmaşık modeli seçme eğiliminde

olmuştur. Kalkan (2016) gerçek veri ve simülasyon veri ile yürüttüğü çalışmasında, yapılan BTM karşılaştırmalarında BIC değerinin AIC değerine göre daha tutarlı sonuçlar verdiğini göstermiştir.

Ancak Sen ve Bradshaw (2017) çalışmalarında, görelî uyum indekslerinin BTM arasında seçim yapılması ve BTM ile madde tepki kuramı modellerinin karşılaştırılması konusunda kullanılmasının çok güvenilir olmadığı bir dizi koşulu göstermiştir. Aynı çalışmada AIC, BIC, SABIC model uyum indekslerinin öneminin abartılmaması ile birlikte model veri uyumu değerlendirmelerinin birden fazla kanıtla dayandırılması önerilmiştir. Chen, de la Torre ve Zhang (2013) çalışmalarında, yanlış tanımlamayı etkili bir şekilde tespit edebilmek için temele doymuş bir model olarak görelî ve mutlak uyum değerlendirmelerinin entegre edilmesini önermiştir. Bu çalışmada da temele doymuş GDINA modeli alınarak mutlak ve görelî uyum indeksleri birlikte değerlendirilmiştir.

von Davier ve Lee'ye göre (2019) son zamanlardaki gerçek veri analizleri, farklı maddelerin hem indirgenmiş hem de doymuş olmak üzere farklı BTM türlerini gerektirebileceğini göstermektedir. Bu çalışmada son araştırma bulgusuna ilişkin her madde için en uygun BTM'yi ampirik olarak belirlemek için la Torre (2011)'nin önermiş olduğu madde düzeyinde model seçme yöntemi kullanılmıştır. Her madde düzeyi için Wald istatistiği kullanılarak doymuş GDINA modeli yerine model veri uyumunda manidar bir kayıp olmadan indirgenmiş DINA ve DINO modellerinin kullanılabilme durumunun incelenmesi sonucunda madde 28 DINA modeliyle daha iyi uyum gösterirken, kalan sekiz madde DINO modeliyle daha iyi uyum sağlamıştır. Mevcut çalışmanın bulgularından çıkarılan sonuç von Davier ve Lee'nin (2019) belirtilen çıkarımını desteklemektedir.

Bu çalışmadan elde edilen bulgu ve sonuçlar çalışmada kullanılan veri seti ile sınırlı olup diğer BTM çalışmalarında olduğu gibi çalışmada kullanılan q matrisinin doğru olduğu varsayımı altında yorumlanmıştır. Ayrıca çalışmanın q matrisinde her maddenin iki niteliği ölçüyor olması da çalışma bulgularını sınırlayan bir etkidir. Bu çalışmanın sonucunda hem test düzeyinde hem de madde düzeyinde değerlendirilen model

uyumları test düzeyinde ve madde düzeyinde farklılaşma göstermiştir. Dolayısıyla test uygulayıcı ve değerlendiricilerine model uyumları ele alınırken tek bir modele dayalı sadece test düzeyinde uyum istatistiklerini değerlendirmeyi benimsemekten ziyade birden fazla modelin ele alınması, test ve madde düzeyinde model uyumu çoklu kanıtlara dayandırılarak değerlendirilmesi önerilmektedir.

Uygulamaya yönelik bir diğer öneri, farklı çalışma grupları, çeşitli nitelik düzeylerinde farklı konu alanlarındaki becerilerin ölçülmesi ile planlanan veri setleri ile doymuş ve indirgenmiş modeller ele alınarak gerçekleştirilecek olan bilişsel tanı analizlerinin model uyumları test ve madde düzeyinde ele alınabilir. Bu çalışma kapsamında madde düzeyi model uyumu için sadece Wald istatistiğine dayalı yorumlama ile sınırlı olup gelecekte olan çalışmalarda madde düzeyi uyum değerlendirmesi için farklı istatistiksel hesaplamalar bir arada kullanılabilir.


**KAYNAKLAR**


- Başokçu, O. T. (2012). DINA model parametreleri kullanılarak tahminlenen madde ayırıcılık indekslerinin incelenmesi. *Eğitim ve Bilim*, 37(163).
- Basokcu, T. O., Ogretmen, T., & Kelecioğlu, H. (2013). Model data fit comparison between DINA and G-DINA in cognitive diagnostic models. *Education Journal*, 2(6), 256-262. <http://doi.org/10.11648/j.edu.20130206.18>
- Bradshaw, L., Izsak, A., Templin, J., & Jacobson, E. (2014). Diagnosing teachers' understandings of rational numbers: Building a multidimensional test within the diagnostic Classification framework. *Educational Measurement: Issues and Practice*, 33, 2-14.
- Chen, J., de la Torre, J., & Zhang, Z. (2013). Relative and absolute fit evaluation in cognitive diagnosis modeling. *Journal of Educational Measurement*, 50, 123-140.
- de la Torre, J. (2008). An empirically based method of Q-matrix validation for the DINA model: Development and applications. *Journal of educational measurement*, 45(4), 343-362.
- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76, 179-199.
- de la Torre, J., & Douglas, J. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69(3), 333-353.
- 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.
- Demir, E. K., & Koç, N. (2018). DINA model ile geliştirilen bir testin psikometrik özelliklerinin belirlenmesi. *Abant İzzet Baysal Üniversitesi Eğitim Fakültesi Dergisi*, 18(1), 130-156.
- Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, 26, 333-352.
- Han, Z., & Johnson, M. S. (2019). Global-and item-level model fit indices. *In Handbook of Diagnostic Classification Models (pp. 265-285)*. Springer, Cham.
- Hartz, S., Roussos, L., & Stout, W. (2002). *Skills diagnosis: Theory and Practice. User Manual for Arpeggio software [Computer software manual]*. Princeton, NJ: Educational Testing Service.
- 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, DOI: 10.1080/15305058.2015.1133627

- Jiao, H. (2009). Diagnostic classification models: Which one should I use? *Measurement: Interdisciplinary Research & Perspective*, 7(1), 65–67.
- Junker, B. W., & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 258–272. DOI: 10.1177/01466210122032064
- Kalkan, Ö. K. (2016). *Bilişsel tanı modellerinin değişen koşullar altında karşılaştırılması: DINA, RDINA, HODINA ve HORDINA Modelleri*. Doktora Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Koyuncu, M. S. (2020). *Bilişsel tanı modellerinde yapısal eşitlik modeli ile q-matris doğruluğunun belirlenmesi*. Doktora Tezi, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Kunina-Habenicht, O., Rupp, A. A., & Wilhelm, O. (2009). A practical illustration of multidimensional diagnostic skills profiling: Comparing results from confirmatory factor analysis and diagnostic classification models. *Studies in Educational Evaluation*, 35(2), 64-70.
- Lee, Y. S., Park, Y. S., & Taylan, D. (2011). A cognitive diagnostic modeling of attribute mastery in Massachusetts, Minnesota, and the US national sample using the TIMSS 2007. *International Journal of Testing*, 11, 144–177.
- Leighton, J. P., & Gierl, M. J. (2007). Defining and evaluating models of cognition used in educational measurement to make inferences about examinees' thinking processes. *Educational Measurement: Issues and Practice*, 26, 3-16.
- Li, F., Cohen, A. S., Kim, S. H., & Cho, S. J. (2009). Model selection methods for mixture dichotomous IRT models. *Applied Psychological Measurement*, 33(5), 353-373.
- Liu, Y., Tian, W., & Xin, T. (2016). An Application of M2 Statistic to Evaluate the Fit of Cognitive Diagnostic Models. *Journal of Educational and Behavioral Statistics*, 41, 3-26.
- 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., & de la Torre, J. (2020). GDINA: an R package for cognitive diagnosis modeling. *Journal of Statistical Software*, 93(14), 1-26.
- Macready, G. B., & Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics*, 2, 99–120.
- Maydeu-Olivares, A., & Joe, H. (2005). Limited and full information estimation and goodness of-fit testing in 2n contingency tables: A unified framework. *Journal of the American Statistical Association*, 100, 1009–1020.

- Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika*, *71*, 713–732.
- Olkun, S. (2015). 6-11 Yaş Türk Çocukları Örnekleminde Diskalkuliye Yatkınlığı Ayırt Etmede Kullanılacak Bir Ölçme Aracı Geliştirme Çalışması. 111K545 Nolu TÜBİTAK Projesi, Ankara, Türkiye.
- Rojas, G., de la Torre, J., & Olea, J. (2012, April). *Choosing between general and specific cognitive diagnosis models when the sample size is small*. Paper presented at the meeting of the National Council on Measurement in Education, Vancouver, Canada.
- Rupp, A., Templin, J., & Henson, R. (2010). *Diagnostic Measurement: Theory, Methods, and Applications*. New York: The Guildford.
- Rupp, A. A., & Templin, J. L. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement: Interdisciplinary Research and Perspectives*, *6*(4), 219–262.
- Sen, S., & Bradshaw, L. (2017). Comparison of relative fit indices for diagnostic model selection. *Applied psychological measurement*, *41*(6), 422–438.  
<https://doi.org/10.1177/0146621617695521>
- Tatsuoka, K. K. (1983). Rule Space: An Approach for Dealing with Misconceptions Based on Item Response Theory. *Journal of Educational Measurement*, *20*(4), 345–354. Retrieved from <http://www.jstor.org/stable/1434951>
- Templin, J. & Henson, R. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological Methods*, *11* (3), 287–305.
- von Davier, M. (2008). A general diagnostic model applied to language testing data. *British Journal of Mathematical and Statistical Psychology*, *61*, 287–307.
- von Davier, M., & Lee, Y. S. (2019). *Handbook of Diagnostic Classification Models*. New York, NY: Springer.

## ORCID

Seyhan SARITAŞ AKYOL  <https://orcid.org/0000-0002-6172-207X>

Mehtap ÇAKAN  <http://orcid.org/0000-0001-6602-6180>

## **SUMMARY**

### **Introduction**

*Cognitive diagnostic models can enable inferences about the individual to be made, regardless of whether individuals have well-defined attributes. These models, which arise from educational assessment, are especially used when it is desired to reveal the strengths and aspects of individuals that need to be developed, or when detailed feedback is expected. Although many cognitive diagnosis models have been proposed in the literature, only a limited number of applications have been performed (Jiao, 2009). This situation results in the fact that the developments in the literature in the cognitive diagnosis approach cannot be reflected efficiently to the applications in education and its usefulness in practice is limited. In addition, it is not entirely clear whether the formulations of many different cognitive diagnosis models (CDM) represent different classes of models or to what extent the models are related to each other. Estimating the model parameters together and comparing their relative fit at the item level in these different formulations remained hard tasks (de la Torre, 2011).*

*In order to collect evidence for the validity of interpretations and inferences drawn from cognitive diagnostic models, it is of great importance for the researcher to determine the appropriate model and to reveal the data fit in terms of the usefulness of the models used. However, practitioners in education may have difficulties in choosing the right model among these models (Jiao, 2009). It is emphasized that this may be because of the deficiencies in the model fit literature in CDM (Hu, Miller, Huggins-Manley & Chen, 2016).*

*This study allows the estimation of multiple CDM parameters such as DINA and DINO models in a single test on real data within the framework of GDINA, the evaluation of the absolute and relative fit of constrained and unconstrained CDMs at the test level, and the determination of the model that fits based on the Wald test at each item level. In this study, it is aimed to determine whether reduced DINA and DINO models can be used instead of the saturated GDINA model determined by the cognitive diagnosis approach and to compare models to find the model that best fits the data at the item and test level.*

### **Method**

*The data set of the study was obtained from 478 students and the answer pattern of the Symbolic Number Comparison test comprising 36 items. Items were scored in binary. In order to determine the attributes required for the definition of the  $Q$  matrix required for cognitive diagnostic analysis, the opinions of domain experts were taken, and five attributes were determined. In the first stage, the saturated GDINA model was taken as the basis, and the absolute and relative fit indices were evaluated together, and it was seen that the saturated model gave the best fit. Afterwards, using Wald statistics, instead of the saturated GDINA model, the availability of reduced DINA and DINO models for each item-level with no significant loss in model-data fit was examined. Analysis of the research data was carried out using the GDINA package in the R Studio program. G-DINA model parameters were obtained using marginalized maximum likelihood estimation (MMLE).*



### Results

Obtained the absolute fit statistics for the GDINA, DINA, and DINO models at the test level indicated that the three models mentioned fit well with the data. According to relative fit indices obtained from the GDINA, DINA and DINO although there is no big difference between the models, it is seen that the GDINA model fits better for the AIC index, while the DINA model fits better for the BIC, CAIC and SABIC indices. The findings regarding the one-way analysis of variance (One-Way ANOVA) using the likelihood ratio test to compare the GDINA model with the DINA and DINO models indicate that  $H_0$  hypothesis was rejected for both comparisons. In other words, when the relative fit indices are considered, the GDINA model fits the data better than the DINA and DINO models. Finally, using Wald statistics for each item level, the availability of reduced DINA and DINO models without a significant loss in model data fit was determined instead of the saturated GDINA model. P values for the Wald statistics calculated for each item level are reported. "Simpler" was applied as the model selection rule. In this case, the model with a higher p-value, among the simpler models that provide model fit, fits the data better. It was seen that the 3rd, 11th, 19th, 30th, 32nd, 34th, 35th and 36th items of the items were more compatible with the DINO model, while the remaining 28 items were better compatible with the DINA model.

### Discussion and Conclusion

According to the result obtained from the findings of the study, considering the absolute fit statistics at the test level for the GDINA, DINA and DINO models, all three models used fit the data well. When the relative fit indices got from the GDINA, DINA and DINO models were examined, it was seen that the saturated GDINA model fitted the data better as a result of the one-way analysis of variance using the likelihood ratio test to compare the GDINA model with the DINA and DINO models. In the literature, there are several studies on the evaluation of model fit in CDM. Sen and Bradshaw (2017) have shown in their study a number of conditions in which the use of relative fit indices for choosing between CDMs and comparing CDMs with item response theory models is not very reliable. In the same study, it was suggested that model data fit assessments should be based on over one evidence, although the importance of AIC, BIC, SABIC model fit indices should not be exaggerated. Hu, Miller, Huggins-Manley, and Chen (2016) showed that the relative fit indices in various conditions they determined in their research could select the correct DINA and GDINA models for most conditions. On the other hand, the relative fit indices rejected the correct DINA and GDINA models in cases of under-definition or misidentification of the Q-matrix. In their study, Chen, de la Torre and Zhang (2013) suggested the integration of relative and absolute fit assessments by taking a saturated model to detect misidentification effectively. In this study, absolute and relative fit indices were evaluated together by taking the saturated GDINA model. de la Torre (2011) stated in his study that the saturated model will give the best fit, assuming that the relevant attributes are defined, and the Q-matrix is correctly determined. The results of the study support this finding.

According to von Davier and Lee (2019) recent real data analyses show that different items may require different types of CDM, both reduced and saturated. In this study, the item-level model selection method suggested by la Torre (2011) was used to empirically determine the most appropriate CDM for each item related to the latest research finding. Because of examining the

*availability of reduced DINA and DINO models without a significant loss in model data fit, instead of the saturated GDINA model using Wald statistics for each item level, eight items showed better fit with the DINO model, while the remaining 28 items fit better with the DINA model. The conclusions drawn from the findings of the present study supports the stated conclusion of von Davier and Lee (2019). Rather than choosing a single model and applying it directly to the data, it is recommended to increase the number of models used and to examine their suitability in terms of items and attribute structure. Based on this, the validity of the interpretations and inferences to be made will be more robust, with the model-data fit being ensured.*