



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

Seçil ÖMÜR SÜNBUİL**, Adnan KAN***

• Geliş Tarihi: 22 Aralık 2014 • Kabul Tarihi: 22 Aralık 2015 • Yayın Tarihi: 31 Ekim 2016

ÖZ: Bu çalışmanın amacı, Bilişsel Tanı Modelleri'nde madde parametre kestirimini, madde uyumunu ve sınıflama tutarlılığını etkileyen faktörlerin neler olduğunun incelenmesidir. Bu amaç doğrultusunda, tamamlayıcı olmayan model (DINA) kullanılarak çeşitli faktörlere (örneklem büyüklüğü, özellikler arası korelasyon, özellik sayısı, madde sayısı, s ve g parametre düzeyleri) göre veri üretilmiştir. Üretilen veri, DINA analiz modeline göre analiz edilmiş, R 3.0 programı ve CDM paketi kullanılmış ve her bir durum için 100 yinleme (replikasyon) yapılmıştır. Parametre kestirimi, madde uyumu ve sınıflama tutarlılığı için, elde edilen çıktı dosyaları, değişimleme ölçütlerine göre hem temel etkiler hem de ortak etkiler bazında düzenlenmiştir. g parametre kestirimi için elde edilen "Mutlak Ortalama Yanlılık (MOY)" ortalamalarına, örneklem büyüklüğünün, madde sayısının ve s ve g parametre düzeylerinin anlamlı bir etkisinin olduğu gözlenmiştir. s parametre kestiriminden elde edilen MOY ortalamalarına ise, örneklem büyüklüğünün, özellikler arası korelasyon düzeyinin ve s ve g parametre düzeylerinin anlamlı bir etkisinin olduğu gözlenmiştir. Madde uyumu için elde edilen RMSEA ortalama değerlerine, örneklem büyüklüğünün, özellik sayısının, madde sayısının ve s ve g parametre düzeylerinin anlamlı bir etkisinin olduğu gözlenmiştir. Sınıflama tutarlılığı için elde edilen "Doğru Sınıflama Oranları (DSO) sonuçları incelendiğinde ise, özellik sayısının, madde sayısının ve s ve g parametre düzeylerinin anlamlı bir etkisinin olduğu gözlenmiştir.

Anahtar sözcükler: Bilişsel Tanı Modelleri, DINA model, parametre kestirimi, sınıflama tutarlılığı

ABSTRACT: The purpose of this study is to investigate factors affecting the item parameter estimation, item fit and classification accuracy of the Cognitive Diagnostic Models (CDM). For this purpose, the data is generated by using noncompensatory model (DINA) according to various factors (sample size, correlation between attributes, the number of attributes, the number of item, s and g parameters levels). The simulated data were analyzed by using DINA models. Data simulation and analyses were conducted by using R 3.0 with CDM package. The output files were organized for parameter estimation, item fit and classification accuracy for both main and interaction effects. By using DINA analysis model obtained from mean values of "Absolute Mean Bias" (MOY) to estimate g parameter, sample size, number of items and levels of g and s parameters of significant effect were observed. s parameter estimation obtained from the mean values MOY, sample size, level of the correlation between attributes and level of the s and g parameters of a significant effect were observed. By using DINA analysis model obtained from mean values of RMSEA, sample size, number of item, number of attribute and levels of g and s parameters of significant effect was observed. By using DINA analysis model obtained from mean values of "Correct Classification Rate (CCR)", number of item, number of attribute and levels of g and s parameters of significant effect was observed.

Keywords: Cognitive Diagnostic Models, DINA Model, parameter estimation, classification accuracy

1. GİRİŞ

Eğitim ve psikolojide ilgilenilen yapıyı ortaya çıkarabilmek, daha iyi incelemek ve açıklamak için, çeşitli kuramlar geliştirilmiştir ve geliştirilmeye devam edilmektedir. Bu açıdan bakıldığında ilk olarak Klasik Test Kuramı (KTK), sonra Madde Tepki Kuramı (MTK) geliştirilmiştir. KTK'da elde edilen gözlenen puanlar ve MTK'da elde edilen yetenek puanları

*Bu çalışma Mersin Üniversitesi, Eğitim Bilimleri Enstitüsü'nde yapılan doktora tezinin bir kısmından üretilmiştir.

**Yrd. Doç. Dr., Mersin Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri Bölümü, Eğitimde Ölçme ve Değerlendirme Ana Bilim Dalı, secilomur@gmail.com

***Prof. Dr., Gazi Üniversitesi, Eğitim Fakültesi, Eğitim Bilimleri, Rehberlik ve Psikolojik Danışmanlık Bölümü, adnankan@gazi.edu.tr

bireylerin görelî olarak sıralanmasında kullanılır ve bu kuramların uygulanması sonucunda tek bir puan rapor edilir. Bu sonuca bakılarak, bireyler hakkında önemli kararlar verilmektedir. Bireylerin ilgilenilen özellik açısından güçlü ya da zayıf yönlerinin ortaya çıkarılmasında bu kuramların yetersiz kalması nedeniyle, daha zengin bilgi verebilecek kuramlara ihtiyaç duyulmuştur. Bu tür eksiklikler göz önünde bulundurularak Tanı Sınıflama Modelleri geliştirilmiştir. Amerika’da eğitim alanında yapılan araştırmalarda, öğretmenlerin ve ailelerin, öğrencilerin bilgi, beceri ve yeteneklerinin güçlü ve zayıf yönleri hakkında daha detaylı bilgiler elde etmeye çalıştıkları ve bu konuya önem verdikleri ortaya çıkmıştır (Huff ve Goodman, 2007). Özellikle 2001 yılında, “No Child Left Behind” hareketinin başlamasıyla öğrencilerin, öğretmenlerin ve ailelerin sosyal talepleri daha belirgin hale gelmeye başlamış ve bilişsel tanı yaklaşımı yaygınlık kazanmıştır (Kato, 2009). Bu hareketle birlikte, anne babalar, öğretmenler ve okul müdürleri için, öğrencilerin neyi öğrendiklerini ya da öğrenme eksikliklerinin neler olduğunu gösteren tanı puan raporları hazırlanmıştır.

Tanı Sınıflama Modelleri, özellikle eğitim ve psikolojide, bireylerin zihinsel tepki süreçleri hakkında daha zengin, anlamlı bilgi elde edilmesi istenildiğinde sıklıkla kullanılan önemli modellerdendir. Bu modeller, bireylerin tepkilerinin altında yatan zihinsel becerilerine bağlı olarak, ölçülmek istenilen değişkenin tüm bileşenlerinin profillerinin çıkarılmasını ve bunların çok değişkenli olarak sınıflanmasını sağlar (Rupp ve Templin, 2008b). Bu durum, ölçülmek istenilen değişkenlerin bütün bileşenlerine ilişkin olarak, bireylerin üstün ve zayıf yönlerinin belirlenmesine yardım eder. Bu modellerde, her bir bireyin her bir beceriye sahip olup olmadığı belirlenirken, özellikler ve maddeler arasındaki ilişki anlaşılmasına çalışılır. Literatür incelendiğinde, Tanı Sınıflama Modelleri şeklinde ifade edilen modellere farklı isimler verildiği görülmektedir (Haertel, 1989; Macready ve Dayton, 1976; Maris, 1995; 1999; Mislevy, 2007; Nichols, Chipman ve Brennan, 1995; Rupp, 2007; Rupp ve Mislevy, 2007). Bu çalışma kapsamında Bilişsel Tanı Modelleri (BTM) ismi kullanılacaktır.

BTM’yle, bireyin bir maddeye doğru cevap verebilmesi için, sahip olduğu örtük özellik/ler ya da beceri/ler ortaya çıkarılmaya çalışılır (Fu ve Li, 2007, Rupp ve Templin, 2008). Test edilen bireylerin yeteneklerini, maddeleri doğru cevaplamak için gerekli olan özelliklere sahip olup olmama bakımından kategorik olarak belirleyen modellerdir (de la Torre, 2009). BTM, bireyin önceden belirlenmiş beceri ya da özelliklere sahip olup olmadığına ilişkin profil sağlar. Böylelikle bireyler hakkında daha zengin, anlamlı, yönlendirici bilgiler sağlanabilir. Ayrıca BTM, bireyin bir maddeyi çözmeye neden başarılı olmadığını da açıklar (Henson, Templin ve Willse, 2009). BTM, bilişsel beceriler ya da örtük özellikler açısından genellikle ikili ya da ayrık olarak bölümlenen örtük uzayı daha çok ve iyi tanımlanmış şekilde bölümler ve bireyleri her bir özellikteki yetkinlik düzeylerine göre değerlendirir (Hartz, 2002). Bu durumda, BTM yapısı çok boyutlu test yapılarına benzetilebilir.

Çeşitli BTM, ölçülen özelliklerin birbirleriyle etkileşimlerinin nasıl olacağına bağlı olarak farklı sayıtları içerir ve bireylerin sahip oldukları becerilerin onların test performanslarını nasıl etkileyeceğini göz önünde bulundurur. İlgili literatür tarandığında, BTM’nin içerdikleri sayıtlar açısından farklı şekillerde sınıflandırıldığı görülmektedir (DiBello, Roussos ve Stout, 2007; Rupp ve Templin, 2008). BTM’nde de tamamlayıcı model ya da tamamlayıcı olmayan model olarak iki model tipi bulunmaktadır. Ayrıca, örtük ya da tepki değişkenlerinin puanlanma (ikili ya da çoklu) şekline göre çeşitli BTM bulunmaktadır. Örneğin, DINA, DINO, NIDO gibi modellerler ikili puanlamada kullanılırken, C-RUM, BIN, GDM gibi modeller çoklu puanlamada kullanılan modellerdir. Tamamlayıcı ve tamamlayıcı olmayan modeller arasındaki fark, örtük yordayıcı değişkenlerin, açık tepki değişkenlerini ortaya çıkaran farklı beceriler karşısında nasıl bağlanacağını belirlenmesinden kaynaklanmaktadır (Rupp ve Templin, 2008b). Tamamlayıcı modellerde, bir becerideki eksiklik diğer kalan beceriler tarafından giderilebilirken; tamamlayıcı olmayan modellerde ise, her bir beceri, doğru cevabın verilmesi için gereklidir. Tamamlayıcı olmayan BTM’nde, birey tepkilerindeki becerilerin ortak etkisi

modellenirken, bireyin testteki bir maddeye doğru cevap vermesi için gerekli olan bütün becerilere sahip olması gerekmektedir. Başka bir ifadeyle, bireyin gerekli olan becerilerden herhangi birine sahip olmaması durumu, o bireyin maddeye doğru cevap verme olasılığını düşürecektir. Henson, Templin ve Willse (2009), bağlayıcı modeli, gerekli olan özelliklerden bir tanesinin eksik olması, sahip olunan diğer özellikler tarafından giderilemeyen model olarak tanımlamışlardır. Bu çalışma kapsamında tamamlayıcı olmayan modellerden DINA modeli kullanılmıştır.

1.1. Q Matrisi

BTM’de, cevaplama özeliğinin belirlendiği $j \times k$ şeklinde oluşturulan ve 1-0 şeklinde kodlanan bir Q matrisi üzerinden hesaplama yapılır (Embretson, 1984; Tatsuoka, 1985). Matriste q_{jk} hücresi j maddesini doğru cevaplamak için, k özeliğine sahip olunması gerekip gerekmediğine işaret eder. q_{jk} , k özeliği j maddesinde bulunuyorsa 1, bulunmuyorsa 0 değerini alır. Satırlarında maddeler, sütunlarında da özellikler yer alan Q matrisindeki hücrelerde yer alan 1’ler ve 0’lar aşağıda belirtilen şekilde yazılabilir.

$$q_{jk} = \begin{cases} 1, & k \text{ özeliği } j \text{ maddesinde bulunuyorsa} \\ 0, & k \text{ özeliği } j \text{ maddesinde bulunmuyorsa} \end{cases}$$

Belirli bir görev üzerindeki başarılı performans, o görev için belirli özelliklerin bir dizi başarılı uygulamalarını gerektirir (de la Torre, 2009). Bu açıdan bakıldığında, Q matrisi, herhangi bir maddede başarılı bir performans gösterilebilmesi için belirli özelliklerden hangisine sahip olunması gerektiğini belirten bir matristir.

Tablo 2: Q matrisi olası tepkiler örneği

Madde	α_1	α_2	α_3
1	1	0	0
2	1	1	0
3	1	1	1
4	0	1	1
5	0	0	1

Tablo 2 incelendiğinde, satırlarında maddeler ve sütunlarında özellikler yer alan bir matris gösterilmektedir. Bu matrisin hücrelerinde ise, ilgili maddenin doğru cevaplanabilmesi için, belirtilen özelliğin gerekip gerekmediği belirtilmektedir. Örneğin, 1. maddeyi doğru cevaplayabilmek için sadece 1. özeliğe sahip olunması gerektiği görülmektedir. Ayrıca, 2. maddeyi doğru cevaplayabilmek için hem 1. özeliğe hem de 2. özeliğe; 3. maddeyi doğru cevaplayabilmek için ise, 3 özeliğe de sahip olunması gerektiği görülmektedir.

Q matrisi, test hazırlama sürecinde, bilişsel özelliklerin kavram haritasını temsil eder (Leighton, Gierl ve Hunka, 2004; Junker, 1999). Bu açıdan bakıldığında, Q matrisi, test geliştirme sürecinde önemli bir role sahiptir. Q matrisi, ölçme aracıyla elde edilen tanısal bilginin kalitesini belirleyen temel unsurdur (Tatsuoka, 1983; Rupp ve Templin, 2008). Ayrıca, Q matrisinin doğru belirlenmiş olması, elde edilen sonuçların geçerliği açısından büyük önem taşımaktadır. Her bir bireyin özellik profili, kullanılan Q matrisi ve bireylerin tepki verileri kullanılarak kestirilebilmektedir.

1.2. DINA (Deterministic Input Noisy “And”) Model

Bu model Haertel (1989) tarafından geliştirilmiş olup, cevaplayıcıları her özellik için iki boyutta sınıflar. Bunlar; “*Yokluk sınıfı*”: belirlenen özeliğe sahip olmayanların oluşturduğu sınıftır. Birey, maddeyi doğru cevaplamak için gerekli özelliklerden sadece bir tanesine bile sahip

değilse buraya atanır. “*Tam sınıf*”: belirlenen özeliğe sahip olanların oluşturduğu sınıftır. Birey, maddeyi doğru cevaplamak için gerekli özelliklerin tamamına sahipse buraya atanır. DINA model matematiksel olarak aşağıdaki eşitlikte belirtildiği şekilde ifade edilmektedir.

$$\xi_{ij} = \prod_{a=1}^A \alpha_{ca}^{q_{ia}}$$

DINA model için geliştirilen yukarıdaki formül incelendiğinde üç temel elemanın gerekli olduğu görülmektedir. Bunlardan ilki, örtük değişken olarak tanımlanan ξ_{ic} , c örtük sınıfında yer alan bireyin i maddesine vermiş olduğu ve α tarafından belirlenen tepkidir. Bu, DINA modelin belirleyici (deterministic) kısmını oluşturmaktadır. Fonksiyon bireyde bulunan özellikler ile maddenin doğru cevaplanabilmesi için gerekli olan özelliklerin eşleşmesi durumunda “1”, eşleşmemesi durumunda ise “0” değerini alır. İkinci gerekli olan eleman ise q_{ia} olarak gösterilmiş olup, i maddesine doğru cevap vermek için a özeliğinin gerekli olup olmadığını belirtir. Özellik gerekli olduğunda “1” gerekli olmadığında “0” değerini alır. Üçüncü eleman olarak belirtilen α_{ia} ise, ilgili özellik maddeyi doğru cevaplamak için gerekli değilse $q_{ia} = 0$ değerini ve $\alpha_{ca}^0 = 1$ değerini alır. İlgili özellik maddeyi doğru cevaplamak için gerekli ise $q_{ia} = 1$ değerini ve c örtük sınıfında yer alan cevaplayıcının ölçülen özeliğe sahip olup olmaması durumuna göre $\alpha_{ca} = 1$ ya da $\alpha_{ca} = 0$ değerini alır (Rupp, Templin ve Henson, 2010).

DINA modelde, i maddesi için c örtük sınıfında yer alan bireyin maddeyi doğru cevaplama olasılığı aşağıdaki eşitlikte gösterilmektedir:

$$\pi_{ic} = P(X_{ic} = 1 / \xi_{ic}) = (1 - s_i)^{\xi_{ic}} g_i^{1 - \xi_{ic}}$$

P; maddeyi çözmek için gerekli bütün becerilere sahip olan bireyin maddeyi doğru cevaplama olasılığıdır. *Slip (s)* “*kaydırma*” ve *guess (g)* “*tahmin, şans*” parametreleridir ve Rupp, Templin ve Henson (2010), s ve g parametrelerini Tablo 2’de belirtildiği şekilde göstermişlerdir.

Tablo 2: DINA modeldeki tepki olasılıkları

	$X_{ic} = 1$ (Doğru tepki)	$X_{ic} = 0$ (Yanlış tepki)
$\xi_{ic} = 1$ (Ölçülen bütün özelliklere sahip)	$1 - s_i$	s_i
$\xi_{ic} = 0$ (Ölçülen özelliklerden en az birine sahip değil)	g_i	$1 - g_i$

Tablo 2 incelendiğinde, ölçülen bütün özelliklere sahip iken maddeye yanlış cevap verme olasılığı “*kaydırma (slip) parametresi*” olarak adlandırılmaktadır. Ölçülen özelliklerden en az birine sahip değilken maddeye doğru cevap verme olasılığı ise “*tahmin (guess) parametresi*” olarak adlandırılmaktadır. Kaydırma (slip) ve tahmin (guess) parametreleri aşağıdaki eşitliklerde gösterilmektedir.

$$s_i = P(X_{ic} = 0 / \xi_{ic} = 1)$$

$$g_i = P(X_{ic} = 1 / \xi_{ic} = 0)$$

1.3. Araştırmanın Amacı ve Önemi

Eğitimde ve psikolojide KTK ve tek boyutlu MTK, ölçülmek istenilen yapının tek boyutlu olması durumunda kullanılmaktadır. KTK’de elde edilen gözlenen puanlar ve tek

boyutlu MTK'de elde edilen yetenek puanları bireylerin görel olarak sıralanmasında kullanılır. Ayrıca tek boyutlu MTK'de süreçten çok maddeye doğru cevap verme olasılıkları üzerine odaklanılmaktadır, İlgilenilen yapı çok boyutlu olduğunda ise, tek boyutluluk sayılıtsı gerektirmeyen çok boyutlu MTK geliştirilmiş olup, bu modelde birden fazla örtük özeliği ölçen maddelere ilişkin performansın gözlenebilmesine olanak tanınmaktadır.

Bireylerin ilgilenilen özellik açısından güçlü ya da zayıf yönlerinin ortaya çıkarılmasında, önceden belirlenmiş beceri ya da özelliklere sahip olup olmadığına ilişkin bilgiler elde edilmesinde ya da bireyler hakkında daha zengin, anlamlı, yönlendirici bilgiler sağlanmasında bu kuramların eksiklikleri bulunmaktadır. Özellikle hiyerarşinin olmadığı becerilerin yoklanmasında eksiklikler söz konusudur. Bu tür eksiklikler göz önünde bulundurularak Bilişsel Tanı Modelleri geliştirilmiştir. Bu modeller sayesinde, ailelere, öğretmenlere, eğitim yöneticilerine ve öğrencilere, öğrencilerin neyi öğrendiklerine, öğrenme eksikliklerinin neler olduğuna ya da istenilen becerilerden hangilerine sahip olup olmadıklarına ilişkin daha anlamlı ve zengin geri dönüt verilebilmektedir.

BTM uygulamalarında, birey sınıflamasının güvenilirliği ve model uyumunun değerlendirilmesi, yapılacak olan çıkarımlar açısından büyük önem taşımaktadır. Bu nedenle, bu çalışma kapsamında, Bilişsel Tanı Modelleri'nde parametre kestirimini ve sınıflama tutarlılığını etkileyen faktörlerin, bazı sınırlılıklar çerçevesinde, neler olduğunun belirlenmesi amaçlanmıştır. Bu amaçla, üretilen veriye ilişkin parametre kestirimlerinin hangi koşullarda daha yansız ve sınıflama tutarlılığının hangi koşullarda daha tutarlı sonuçlar verdiğine bakılmıştır.

İlgili literatür incelendiğinde, yurtiçinde BTM'nde parametre kestirimini ya da sınıflama tutarlılığını etkileyen faktörlerin araştırılmasına yönelik herhangi bir çalışmanın yapılmadığı görülmüştür. Yurt dışında ise, bu çalışma kapsamına benzer simülasyon çalışmalarının yapıldığı görülmüştür ancak, bu çalışmada, değişimlenen faktörler ve bu faktörlerin düzeyleri çeşitlendirilerek çalışmaya özgül değer katılmıştır. Bu çalışmanın, ülkemizde Bilişsel Tanı Modelleri'nin kullanılmasının yaygınlaşmasına ışık tutacağı düşünülmektedir.

2. YÖNTEM

2.1. Araştırmanın Türü

Bu çalışmada, Bilişsel Tanı Modelleri'nden DINA Modelde madde parametre kestirimini, madde uyumunu ve sınıflama tutarlılığını etkileyen faktörlerin neler olduğunun incelenmesi, hangi durumlarda daha yansız kestirimlerin ve daha doğru sınıflamaların elde edildiğinin belirlenmesi amaçlanmıştır. Bu nedenle, bu çalışma temel araştırma olarak değerlendirilebilir.

2.2. Veri Üretim ve Veri Analiz Koşulları

Veri üretim çalışmasında, örneklem büyüklüğü (200, 500, 1000, 5000), özellikler arası korelasyon (0.2, 0.5, 0.8), özellik sayısı (3 ve 4), madde sayısı (15, 30, 45), s ve g parametre düzeyleri (düşük düzey için, s ve g \approx U(.05 - .10) ve yüksek düzey için, s ve g \approx U(.20 - .30)) değişimlenmiş olup, her bir durum için 100 yinleme (replikasyon) yapılmıştır. Araştırma kapsamında kullanılan özellik sayılarına ve madde sayılarına uygun Q matrisleri belirlenmiştir. 3 ve 4 özellik ve 15, 30, 45 madde için ayrı ayrı Q matrisleri oluşturulmuştur. Bu çalışma kapsamında, özelliklerin güçlük düzeyleri sabit tutulmuştur. Tüm özellikler için özellik güçlükleri aynı olarak belirlenmiş ve 3 özellik için ortalama vektörü $\mu=(0.00, 0.00, 0.00)$, 4 özellik için ortalama vektörü $\mu=(0.00, 0.00, 0.00, 0.00)$ olarak ayarlanmıştır. Veri üretimi, R 3.0 programında CDM paketi kullanılarak gerçekleştirilmiştir. Araştırma kapsamında değişimlenen faktörler ve bu faktörlere ilişkin düzeyler Tablo 3'te gösterilmektedir. Bu çalışma kapsamında, $4 \times 3 \times 2 \times 3 \times 4 = 288$ tane farklı deneysel yapı ve her biri için 100 replikasyon yapılmıştır.

2.3. Verilerin Analizi

Veri, Tamamlayıcı Olmayan Modele (DINA) göre, belirtilen koşulların çaprazlanması sonucu oluşan her duruma uygun olarak üretilmiştir. Üretilen veri DINA modele göre analiz edilmiştir. Veri üretimi ve verinin modellere göre analizi R 3.0 programında CDM paketi kullanılarak gerçekleştirilmiştir. Parametre kestirimi (MOY), madde uyumunu (RMSEA) ve sınıflama tutarlılığı (DSO) için elde edilen 28800 tane çıktı dosyası, değişimleme ölçütlerine göre hem temel etkiler hem de ortak etkiler bazında düzenlenmiştir. Analiz sonuçlarından elde edilen bulguların organize edilmesi ise, araştırmacı tarafından R 3.0 programında kod yazılarak gerçekleştirilmiştir.

Parametre kestirimleri için gerçek değerler ile kestirilen değerler arasındaki farktan elde edilen “yanlılık değerleri”, parametre kestirim sürecinin sağlıklı işleyip işlemediğinin değerlendirilmesinde bir iyilik ölçütü olarak kullanılabilir (de la Torre, Hong ve Deng, 2010). Bu açıdan bu çalışma kapsamında, parametre kestiriminde, gerçek ile kestirilen değerler arasında tutarlı sonuçların elde edilip edilmediğinin belirlenmesi için “mutlak ortalama yanlılık (absolute mean bias)” hesaplaması yapılmıştır. Madde-model uyumunun bir göstergesi olarak “madde uyum RMSEA” değerleri hesaplanmıştır. Bireylerin doğru sınıflanıp sınıflanmadığını belirlemek üzere “Doğru Sınıflama Oranları” hesaplanmıştır. Burada, üretilen veriye ilişkin olarak elde edilen bireylere ait özellik örüntüleri ile Ençok Olabilirlik Kestirimi yaklaşımı kullanılarak kestirilen bireylerin özellik örüntüleri arasındaki doğru sınıflama oranları hesaplanmıştır.

Tablo 3: Yapılan değişimlemelere ilişkin düzenek

Faktör	Düzye Sayısı	Düzye Değerleri
Örnekleme Büyüklüğü (N)	4	200 (N_1)
		500 (N_2)
		1000 (N_3)
		5000 (N_4)
Özellikler Arası Korelasyon (c)	3	0.2 (c_1)
		0.5 (c_2)
		0.8 (c_3)
Özellik Sayısı (a)	2	3 (a_3)
		4 (a_4)
Madde Sayısı (MS)	3	15 (MS_1)
		30 (MS_2)
		45 (MS_3)
s ve g düzeyleri	4	s ve g düşük ($s_d g_d$)
		s düşük - g yüksek ($s_d g_y$)
		s yüksek - g düşük ($s_y g_d$)
		s ve g yüksek ($s_y g_y$)
Analiz Modeli	1	DINA
Replikasyon sayısı	100	

3. BULGULAR

3.1.1. Çeşitli faktörlerin g parametre kestirimleri için elde edilen MOY ortalamalarına temel etkileri nasıldır?

Çeşitli faktörlerin, g parametre kestirimleri için elde edilen MOY ortalamalarına temel etki sonuçları Tablo 4'te gösterilmektedir. Çeşitli örneklem büyüklükleri açısından Tablo 4 incelendiğinde, g parametresi kestirimlerinden elde edilen MOY ortalama değerlerinin 0.006 ile 0.028 arasında değiştiği görülmektedir. g parametresi kestiriminde, örneklem büyüklüğü arttıkça

MOY ortalama değerlerinin azaldığı görülmektedir. Bu bulgulara göre, örneklem büyüklüğü arttıkça g parametrelerinin daha yansız kestirildiği söylenebilir. Özellikler arasındaki korelasyon düzeyleri ve özellik sayısı değişimlendiğinde, elde edilen MOY ortalama değerlerinin minimum ve maksimum değerleri arasındaki değişiminin yok denecek kadar az olduğunu söylemek mümkündür. Başka bir ifadeyle, g parametreleri kestirimi için elde edilen MOY ortalamalarının, özellikler arası korelasyon düzeylerinin ve özellik sayısının değişimlenmesiyle birlikte çok fazla değişmediği görülmektedir. Tablo 4 madde sayıları açısından incelendiğinde, madde sayısının artırılmasıyla, g parametrelerinin daha yansız kestirilebileceği söylenebilir. s ve g parametre düzeyleri açısından Tablo 4 incelendiğinde, düşük s ve g parametre düzeylerinde g parametre kestirimlerinden elde edilen MOY ortalama değerlerinin daha düşük olduğu görülmektedir. Bu bulgulara göre, düşük s ve g parametrelerinin düzeylerinde, g parametresinin daha yansız kestirildiği söylenebilir.

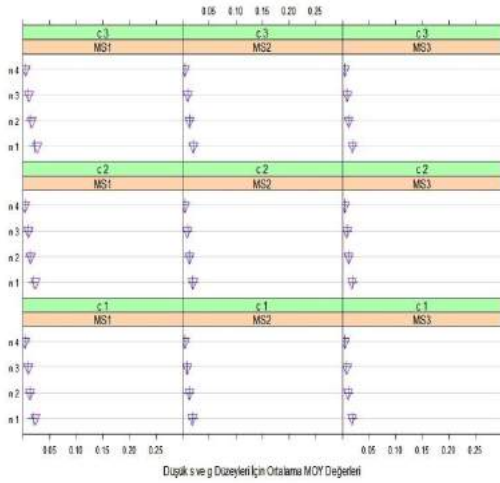
Tablo 4: g parametre kestiriminden elde edilen MOY değerlerine ilişkin faktörlerin temel etki değerleri

	Düzeyley	Yanlılık Değerleri
Örneklem Büyüklüğü	200	0.028
	500	0.018
	1000	0.012
	5000	0.006
Özellikler Arası Korelasyon	0.2	0.015
	0.5	0.016
	0.8	0.016
Özellik Sayısı	3	0.015
	4	0.016
Madde Sayısı	15	0.018
	30	0.015
	45	0.015
s ve g	s ve g düşük	0.011
	s düşük - g yüksek	0.019
	s yüksek - g düşük	0.012
	s ve g yüksek	0.021

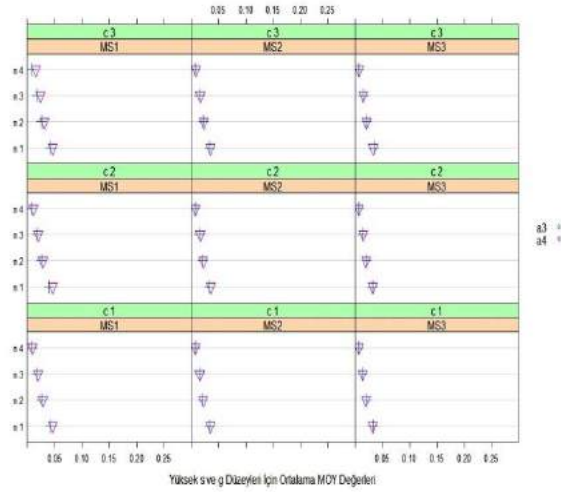
3.1.2. Çeşitli faktörlerin g parametre kestirimleri için elde edilen MOY ortalamalarına ortak etkileri nasıldır?

Çeşitli s ve g parametre düzeylerinde, g parametre kestirimleri için ortalama MOY değerleri elde edilmiştir. Çeşitli faktörlerin ortak etkileşimlerine ilişkin saçılım Grafik 1, 2, 3 ve 4'te sunulmuştur. Grafiklerde yer alan koşulların ortak etkileri açısından ortalama MOY dağılımları incelendiğinde, anlamlı yapılanmaların olduğu göze çarpmaktadır. Özellikle örneklem büyüklüğü arttıkça ortalama MOY değerlerinin azaldığı ve bunun kısmen de olsa madde sayısı ile ilişkili olduğu görülmektedir. Madde sayısı arttıkça ortalama MOY değerlerinin çok az da olsa azaldığı görülmektedir.

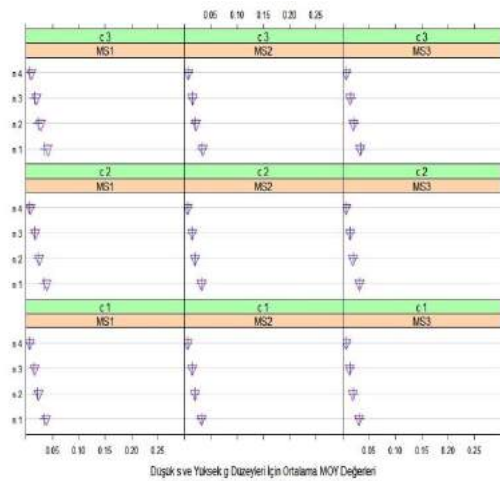
Dikkat edilmesi gereken bir nokta ise, dört grafikte de özellikler arasındaki korelasyon düzeylerinin ve özellik sayılarının g parametre kestirimindeki önemsizliğidir. Bütün özellikler arası korelasyon düzeylerinde ve özellik sayılarında yaklaşık aynı sonuçlar elde edilmiştir. Grafik 1 ve Grafik 2, ortak etkiler açısından incelendiğinde, genel olarak Grafik 1'den elde edilen sonuçların (düşük s ve g düzeyi), Grafik 2'den elde edilen sonuçlara (yüksek s ve g düzeyi) göre daha düşük olduğu görülmektedir. Grafik 3 ve Grafik 4 ortak etkiler açısından incelendiğinde, faktörlerin tüm düzeylerinde, Grafik 4'ten (g parametresi düşük düzeyde) elde edilen MOY ortalamalarının, Grafik 3'ten (g parametresi yüksek düzeyde) elde edilen MOY ortalamalarına göre daha düşük olduğu görülmektedir. Bu durumda, g parametre kestirimi için MOY ortalamaları elde edildiğinde, s ve g parametre düzeylerinin etkili bir faktör olduğu söylenebilir.



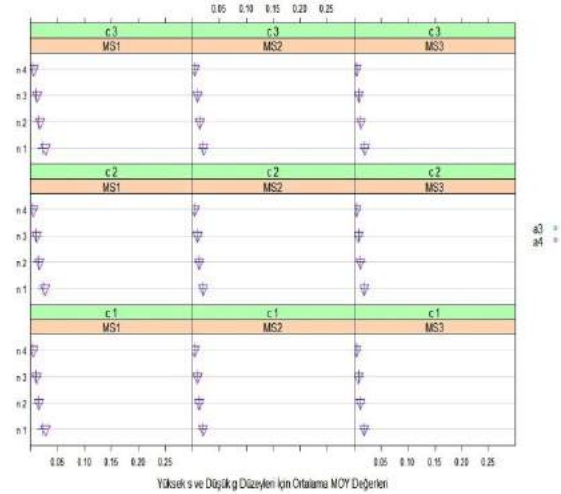
Grafik 1: Düşük s ve g parametre düzeyleri kullanıldığında, g parametre kestiriminden elde edilen ortalama MOY değerleri



Grafik 2: Yüksek s ve g düzeyleri kullanıldığında, g parametre kestiriminden elde edilen ortalama MOY değerleri



Grafik 3: Düşük s ve yüksek g parametre düzeyleri kullanıldığında, g parametre kestiriminden elde edilen ortalama MOY değerleri



Grafik 4: Yüksek s ve düşük g parametre düzeyleri kullanıldığında, g parametre kestiriminden elde edilen ortalama MOY değerleri

3.2.1. Çeşitli faktörlerin s parametre kestirimleri için elde edilen MOY ortalamalarına temel etkileri nasıldır?

Çeşitli faktörlerin s parametre kestirimleri için elde edilen MOY ortalamalarına temel etki sonuçları Tablo 5'te gösterilmektedir. Tablo 5'e göre örneklem büyüklükleri değiştiğinde, s parametresi kestirimlerinden elde edilen ortalama MOY değerlerinin 0.007 ile 0.036 arasında değiştiği görülmektedir. s parametre kestiriminde, örneklem büyüklüğü arttıkça MOY ortalama değerlerinin azaldığı ve bu durumda s parametrelerinin daha yansız kestirildiği söylenebilir. Özellikler arasındaki korelasyon düzeyleri ve madde sayıları değiştiğinde, elde edilen MOY ortalama değerleri birbirine yakın olmasına rağmen, özellikler arasındaki korelasyon düzeyi ve madde sayısı yüksek olduğunda s parametrelerinin daha yansız kestirildiği

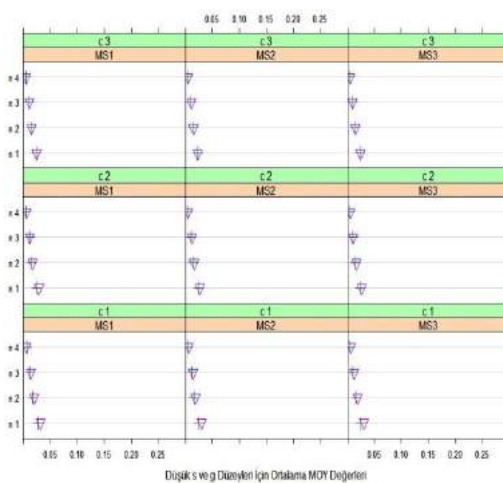
söylenbilir. Özelik sayıları açısından Tablo 5 incelendiğinde, elde edilen ortalama MOY değerleri özelik sayısının artışıyla birlikte MOY ortalamalarının da arttığı görülmektedir. Bu durumda düşük özelik sayısında s parametrelerinin daha yansız kestirildiği söylenebilir. s ve g parametre düzeyleri açısından incelendiğinde ise, elde edilen MOY ortalama değerlerinin, 0.015 ile 0.027 arasında değiştiği görülmektedir. Bu bulgulara göre, s parametreleri kestirildiğinde, özellikle s ve g parametrelerinin düşük düzeyde olmasının daha yansız kestirimlerin elde edilmesi açısından önemli olduğu söylenebilir.

Tablo 5: s Parametre kestiriminde DINA analiz modeli kullanıldığında elde edilen MOY değerlerine ilişkin faktörlerin temel etki değerleri

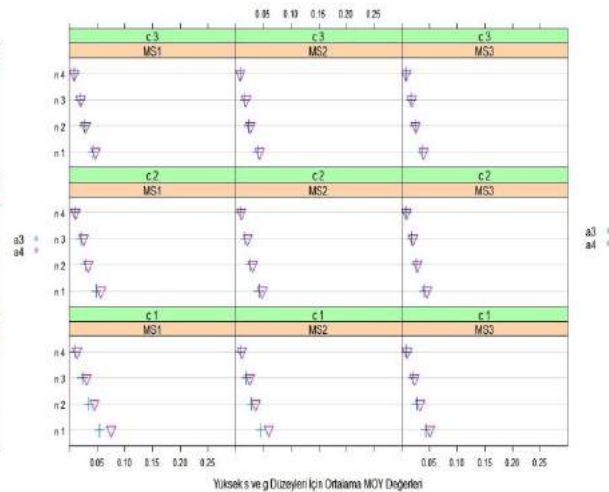
	Düzyer	Yanlılık Değerleri
Örnekleme Büyüklüğü	200	0.036
	500	0.023
	1000	0.016
	5000	0.007
Özelikler Arası Korelasyon	0.2	0.023
	0.5	0.020
	0.8	0.018
Özelik Sayısı	3	0.019
	4	0.022
Madde Sayısı	15	0.022
	30	0.020
	45	0.019
	s ve g	s ve g düşük
s düşük - g yüksek		0.015
s yüksek - g düşük		0.024
s ve g yüksek		0.027

3.2.2. Çeşitli faktörlerin s parametre kestirimleri için elde edilen MOY ortalamalarına ortak etkileri nasıldır?

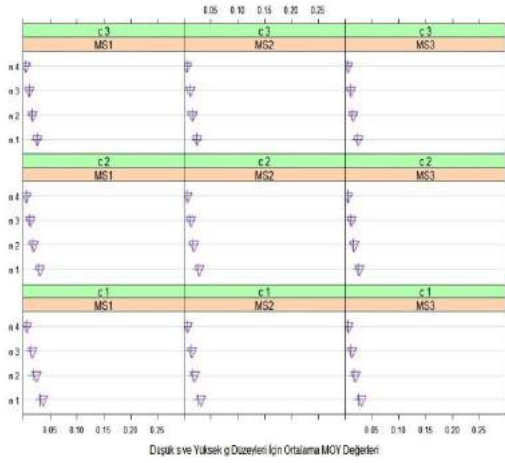
Çeşitli s ve g parametre düzeylerinde, s parametre kestirimleri için ortalama MOY değerleri elde edilmiştir. Çeşitli faktörlerin ortak etkileşimlerine ilişkin saçılım Grafik 5, 6, 7 ve 8'de sunulmuştur. Örneklem büyüklüğü arttıkça ortalama MOY değerlerinin azaldığı görülmektedir. Başka bir ifadeyle, diğer faktörlerin her düzeyi için örneklem büyüklüğü arttıkça yanlılığın azaldığı söylenebilir. Buradan hareketle, s parametre kestirimi için MOY ortalamalarının elde edilmesinde örneklem büyüklüğünün etkili bir faktör olduğu görülmektedir.



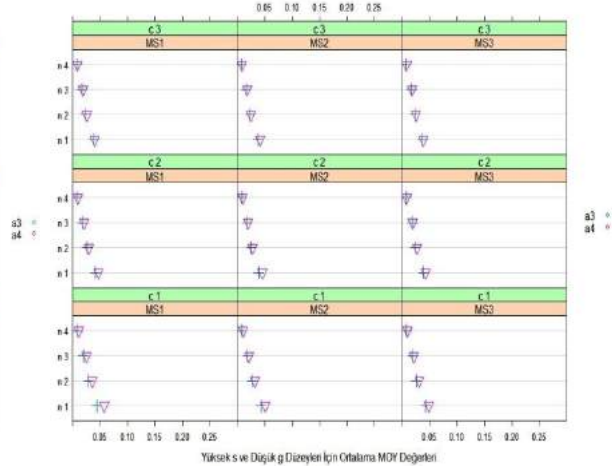
Grafik 5: DINA modelde ve düşük s ve g parametre düzeylerinde, s parametre kestiriminden elde edilen ortalama MOY değerleri



Grafik 6: DINA modelde, yüksek s ve g düzeylerinde s parametre kestiriminden elde edilen ortalama MOY değerleri



Grafik 7: DINA modelde, düşük s ve yüksek g parametre düzeylerinde, s parametre kestiriminden elde edilen ortalama MOY değerleri



Grafik 8: DINA modelde ve yüksek s ve düşük g parametre düzeylerinde, s parametre kestiriminden elde edilen ortalama MOY değerleri

Grafik 5 ve Grafik 6, ortak etkiler açısından incelendiğinde, genel olarak Grafik 5'ten elde edilen sonuçların, Grafik 6'dan elde edilen sonuçlara göre daha düşük olduğu görülmektedir. Ayrıca, Grafik 7 ve Grafik 8 ortak etkiler açısından incelendiğinde, s parametresinin daha düşük olduğu düzeylerden elde edilen MOY ortalamalarının, s parametresinin daha yüksek olduğu düzeylerden elde edilen MOY ortalamalarına göre daha düşük olduğu görülmektedir. Bu durumda, s parametre kestirimi için MOY ortalamalarının elde edilmesinde, s ve g parametrelerinin düzeylerinin etkili olduğu söylenebilir. Dört grafik ayrı ayrı incelendiğinde, madde sayılarının ve özellik sayılarının s parametre kestirimi için MOY ortalamalarının elde edilmesinde önemsiz olduğu söylenebilir. Bütün madde ve özellik sayıları koşullarında yaklaşık aynı sonuçların elde edildiği görülmektedir.

3.3.1. Çeşitli faktörlerin madde uyumu için elde edilen RMSEA ortalamalarına temel etkileri nasıldır?

Çeşitli faktörlerin madde uyumu için elde edilen RMSEA ortalamalarına ilişkin temel etki sonuçları Tablo 6'da gösterilmektedir.

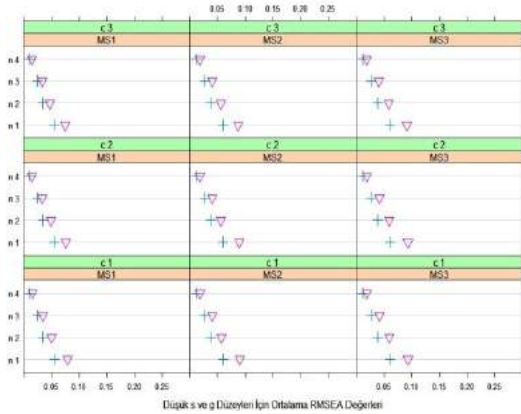
Tablo 6: Faktörlerin RMSEA ortalamalarına ilişkin temel etki değerleri

	Düzyerler	RMSEA değerleri
Örneklem Büyüklüğü	200	0.084
	500	0.053
	1000	0.038
	5000	0.017
Özellikler Arası Korelasyon	0.2	0.049
	0.5	0.048
	0.8	0.047
Özelik Sayısı	3	0.041
	4	0.055
Madde Sayısı	15	0.039
	30	0.050
	45	0.054
s ve g	s ve g düşük	0.041
	s düşük - g yüksek	0.056
	s yüksek - g düşük	0.046
	s ve g yüksek	0.049

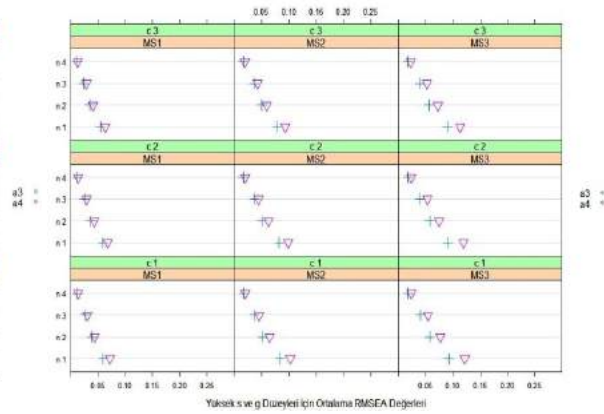
Tablo 6 çeşitli örneklem büyüklükleri açısından incelendiğinde, DINA modelinden elde edilen RMSEA ortalama değerlerinin 0.017 ile 0.084 arasında değiştiği görülmektedir. DINA modeli madde uyumunda, örneklem büyüklüğü arttıkça RMSEA ortalama değerlerinin azaldığı görülmektedir. Başka bir ifadeyle, örneklem büyüklüğü arttıkça maddelerin daha fazla uyum gösterdiği söylenebilir. Özellikler arasındaki çeşitli korelasyon düzeyleri açısından Tablo 6 incelendiğinde, DINA modelinde bu değerlerinin minimum ve maksimum değerleri arasındaki değişikliğin yok denecek kadar az olduğunu söylemek mümkündür. Başka bir ifadeyle, DINA modeli kullanıldığında, RMSEA ortalama değerlerinin özellikler arası korelasyon düzeylerinin değişimlenmesiyle birlikte çok fazla değişmediği görülmektedir. Özellik sayısı ve madde sayısı açısından Tablo 6 incelendiğinde, özellik sayısı ve madde sayısı arttıkça elde edilen RMSEA ortalama değerlerinin de arttığı görülmektedir. Özellik sayısının ve madde sayısının düşük olmasının madde uyumu açısından önemli olduğu görülmektedir. Başka bir ifadeyle, özellik sayısı ve madde sayısı azaldıkça maddelerin daha fazla uyum gösterdiği söylenebilir. Tablo 6 s ve g parametre düzeyleri açısından incelendiğinde, RMSEA ortalama değerlerin 0.041 ile 0.056 arasında değiştiği görülmektedir. Bu sonuçlar ışığında, madde uyumu incelendiğinde, s ve g parametrelerinin düşük düzeyde olmasının madde uyumu açısından önemli olduğu görülmektedir.

3.3.2. Çeşitli faktörlerin madde uyumu için elde edilen ortalama RMSEA ortalamalarına ortak etkileri nasıldır?

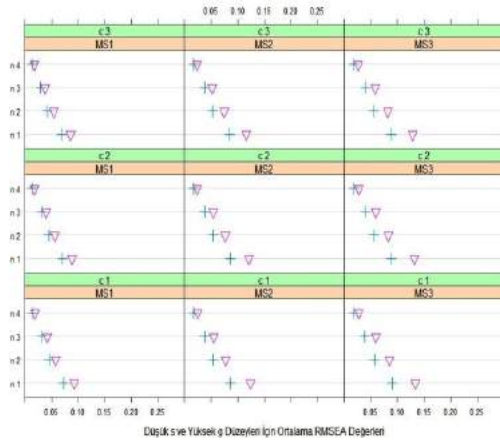
Çeşitli s ve g parametre düzeylerinde, madde uyumu için ortalama RMSEA ortalama değerleri elde edilmiştir. Çeşitli faktörlerin ortak etkileşimlerine ilişkin saçılım Grafik 9, 10, 11 ve 12’de sunulmuştur. Grafiklerde yer alan koşulların ortak etkileri açısından RMSEA ortalama dağılımları incelendiğinde, anlamlı bir yapılanmanın olduğu göze çarpmaktadır. Özellikle örneklem büyüklüğü arttıkça RMSEA ortalama değerlerinin azaldığı ve bunun madde sayısı ile ilişkili olduğu görülmektedir. Madde sayısı arttıkça, ortalama RMSEA ortalama değerlerinin de arttığı görülmektedir. Dört grafik ayrı ayrı incelendiğinde, özellik sayısının 3 olduğu durumdan elde edilen RMSEA ortalama değerlerinin daha düşük olduğu görülmektedir. Başka bir ifadeyle, özellik sayısı 3 olduğunda madde uyumunun daha fazla olduğu göze çarpmaktadır. Özellik sayısı ve madde sayısının arttığı durumlarda, RMSEA ortalamalarının arttığı görülmektedir. Bu durumun modelin kompleksleşmesinden kaynaklandığı söylenebilir.



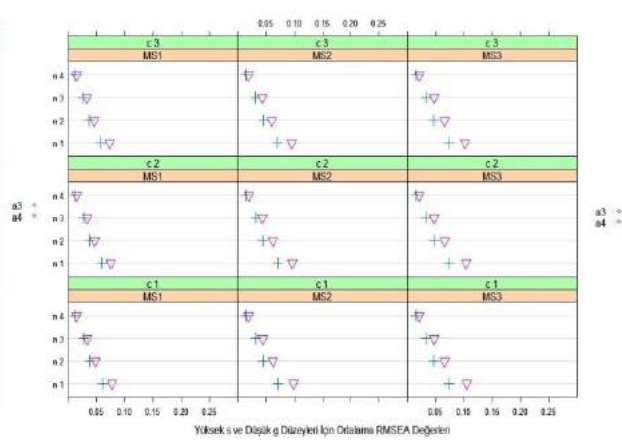
Grafik 9: DINA modelde ve düşük s ve g parametre düzeylerinde, madde uyumu için elde edilen RMSEA ortalama değerleri



Grafik 10: DINA modelde, yüksek s ve g düzeylerinde madde uyumu için elde edilen RMSEA ortalama değerleri



Grafik 11: DINA modelde, düşük s ve yüksek g parametre düzeylerinde, madde uyumu için elde edilen RMSEA ortalama değerleri



Grafik 12: DINA modelde ve yüksek s ve düşük g parametre düzeylerinde, madde uyumu için elde edilen RMSEA ortalama değerleri

Ayrıca özellikler arası korelasyon düzeylerinin, madde uyumu için hesaplanan RMSEA ortalama değerleri için önemsiz olduğu söylenebilir. Bütün özellikler arası korelasyon düzeylerinde yaklaşık aynı sonuçların elde edildiği görülmektedir. Grafik 9 ve Grafik 10 ortak etkiler açısından incelendiğinde, genel olarak Grafik 9'dan elde edilen sonuçların (düşük s ve g düzeyi), Grafik 10'dan elde edilen sonuçlara (yüksek s ve g düzeyi) göre daha düşük olduğu görülmektedir. Bu durumda, madde uyumu için elde edilen RMSEA ortalama değerleri için, s ve g parametre düzeylerinin etkili bir faktör olduğu söylenebilir.

3.4.1. Çeşitli faktörlerin Doğru Sınıflama Oranları (DSO) ortalamalarına temel etkileri nasıldır?

Çeşitli faktörlerin, sınıflama tutarlılığı için elde edilen Doğru Sınıflama Oranları (DSO) ortalamalarına ilişkin temel etki sonuçları Tablo 7'de gösterilmektedir.

Tablo 7: DINA model kullanıldığında faktörlerin doğru sınıflama oranları ortalamalarına ilişkin temel etki değerleri

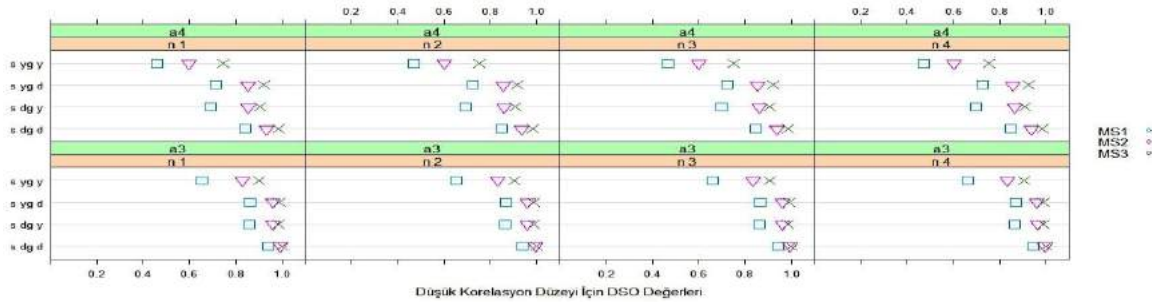
	Düzyerler	DSO Değerleri
Örneklem Büyüklüğü	200	0.851
	500	0.852
	1000	0.853
	5000	0.853
Özellikler Arası Korelasyon	0.2	0.854
	0.5	0.853
	0.8	0.851
Özellik Sayısı	3	0.911
	4	0.794
Madde Sayısı	15	0.764
	30	0.869
	45	0.925
s ve g	s ve g düşük	0.948
	s düşük - g yüksek	0.872
	s yüksek - g düşük	0.884
	s ve g yüksek	0.706

Çeşitli örneklem büyüklüklerinden elde edilen DSO ortalama değerlerinin 0.851 ile 0.853 arasında ve özellikler arasındaki çeşitli korelasyon düzeylerinden elde edilen DSO ortalama değerlerinin ise 0.854 ile 0.851 arasında değiştiği görülmektedir. Bu sonuçlar ışığında, örneklem büyüklüğünün ve özellikler arası korelasyon düzeylerinin değişimlenmesinin DSO ortalamalarını

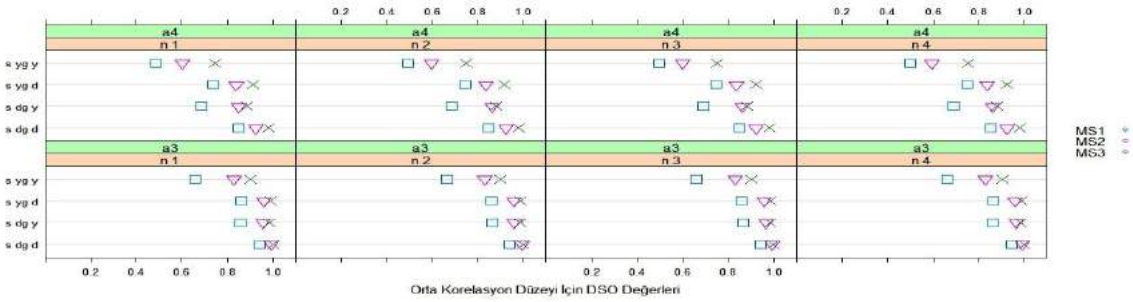
çok fazla değiştirmedeği söylenebilir. Özelik sayıları açısından Tablo 7 incelendiğinde, elde edilen DSO ortalama değerlerinin 0.794 ve 0.911 olduğu görülmektedir. Özelik sayısı arttıkça DSO ortalama değerlerinin azaldığı ve doğru sınıflamanın daha az yapıldığı söylenebilir. Madde sayıları açısından incelendiğinde, elde edilen DSO ortalama değerlerinin 0.764 ile 0.925 arasında değiştiği görülmektedir. DINA modelinde, madde sayısı arttıkça DSO ortalama değerlerinin arttığı görülmektedir. Bu sonuçlar ışığında, madde sayısı artışıyla birlikte DSO ortalamalarının da arttığı görülmektedir. s ve g parametre düzeyleri açısından Tablo 7 incelendiğinde, elde edilen DSO ortalama değerlerinin 0.706 ile 0.948 arasında değiştiği görülmektedir. Bu bulgulara göre, düşük s ve g parametre düzeyinde daha yüksek DSO ortalamalarının elde edildiği görülmektedir.

3.4.2. Çeşitli faktörlerin sınıflama tutarlılığı için elde edilen DSO ortalamalarına ortak etkileri nasıldır?

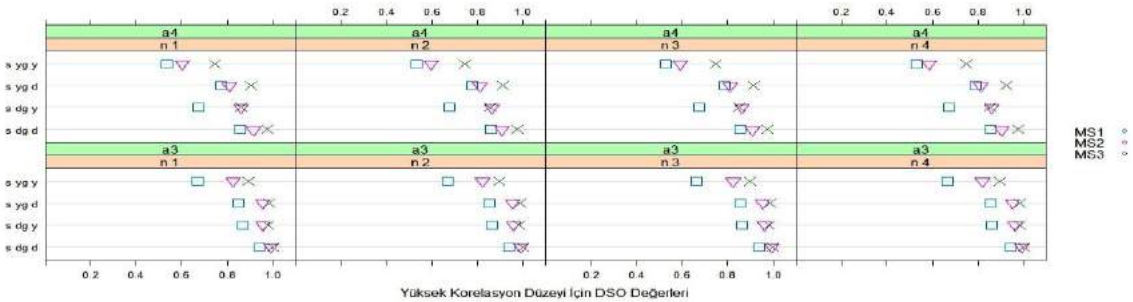
Çeşitli özellikler arası korelasyon düzeylerinde, sınıflama tutarlılığı için DSO ortalama değerleri elde edilmiştir. Çeşitli faktörlerin ortak etkileşimlerine ilişkin saçılım Grafik 13, 14 ve 15’de sunulmuştur.



Grafik 13: DINA modelde ve özellikler arası korelasyonun düşük düzeyinde elde edilen DSO ortalama değerleri



Grafik 14: DINA modelde ve özellikler arası korelasyonun orta düzeyinde, elde edilen DSO ortalama Değerleri



Grafik 15: DINA modelde ve özellikler arası korelasyonun yüksek düzeyinde, elde edilen DSO ortalama değerler

Grafik 13, 14 ve 15’de yer alan DSO ortalama dağılımları incelendiğinde, anlamlı bir yapılanmanın olduğu göze çarpmaktadır. Madde sayısı arttıkça, s ve g parametre düzeyleri ve

özelik sayısı azaldıkça DSO değerlerinde artma görülmektedir. Başka bir ifadeyle, madde sayısının, s ve g düzeylerinin ve özellik sayısının, sınıflama tutarlılığı açısından önemli olduğu söylenebilir. Ayrıca tüm grafikler ayrı ayrı incelendiğinde, örneklem büyüklüğünün ve özellikler arası korelasyon düzeylerinin sınıflama tutarlılığı açısından önemsiz olduğu göze çarpmaktadır

4. TARTIŞMA ve YORUM

4.1. g ve s Parametre Kestirimi İçin Elde Edilen MOY Sonuçlarının Değerlendirilmesi

Temel Etkiler Açısından: Örneklem büyüklüğü arttıkça, g ve s parametre kestirimi için elde edilen MOY ortalama değerlerin azaldığı görülmektedir. Başka bir ifadeyle, örneklem büyüklüğü arttıkça g ve s parametrelerin daha yansız kestirildiği söylenebilir. İlgili literatürde yapılan çalışmalar incelendiğinde, örneklem büyüklüğünün parametre kestiriminde önemli bir faktör olduğu ve elde edilen sonuçların, bu çalışmanın sonuçlarıyla örtüştüğü görülmektedir (de la Torre, Hong ve Deng, 2010; Choi ve diğerleri, 2010; Habenicht, Rupp ve Wilhelm, 2012). Bu çalışma kapsamında, diğer çalışmalardan farklı olarak daha düşük (200, 500) örneklem büyüklükleri de değişimlenmiştir. Bu örneklem büyüklükleri kullanıldığında, g ve s parametre kestirimi için elde edilen MOY değerlerinin arttığı, dolayısıyla daha yanlış kestirimlerin yapıldığı görülmektedir. Özellikler arasındaki korelasyon düzeyleri değişimlendiğinde, g ve s parametreleri kestiriminden elde edilen MOY değerlerinin çok az değiştiği görülmektedir. Bu bulgular, Kim'in (2011) yapmış olduğu çalışmanın bulgularıyla paralellik göstermektedir. Özellik ve madde sayıları açısından bakıldığında ise, g ve s parametre kestiriminden elde edilen MOY değerlerinin çok az değiştiği görülmektedir. g ve s parametre kestirimlerinde, s ve g parametre düzeyleri düşük olduğunda daha yansız kestirimlerin yapıldığı görülmektedir. Bu sonuç, de la Torre, Hong ve Deng'in (2010), yaptıkları çalışmadan elde ettikleri bulgularla paralellik göstermektedir.

Ortak Etkiler Açısından: g parametresi kestiriminde faktörlerin ortak etkileri incelendiğinde, örneklem büyüklüğü ve madde sayısı arttıkça, s ve g parametre düzeyleri azaldıkça, g parametrelerin daha yansız kestirildiği söylenebilir. Ayrıca, özellikler arasındaki korelasyon düzeylerinin ve özellik sayılarının, ortak etkiler açısından g parametre kestiriminde önemli bir etkisinin olmadığı yorumu yapılabilir. s parametresi kestiriminde ise, örneklem büyüklüğü arttıkça, s ve g parametre düzeyleri azaldıkça daha yansız kestirildiği söylenebilir. Ayrıca, madde sayılarının ve özellik sayılarının s parametre kestiriminde önemsiz olduğu söylenebilir. Bütün madde ve özellik sayıları koşullarında yaklaşık aynı sonuçların elde edildiği görülmektedir.

4.2. Madde Uyumu İçin Elde Edilen RMSEA Sonuçlarının Değerlendirilmesi

Temel Etkiler Açısından: Örneklem büyüklüğü arttıkça ortalama RMSEA değerlerinin azaldığı görülmektedir. Başka bir ifadeyle, örneklem büyüklüğü arttıkça maddelerin daha fazla uyum gösterdiği görülmektedir. Özellikler arasındaki çeşitli korelasyon düzeyleri açısından ortalama RMSEA değerleri incelendiğinde, DINA modelinde bu faktörün önemsiz olduğu söylenebilir. Özellik sayıları açısından bakıldığında, DINA modelinde parametre kestiriminden elde edilen ortalama RMSEA değerlerinde çok az bir değişikliğin meydana geldiği, ancak madde sayıları açısından bakıldığında, madde sayısı azaldıkça madde uyumunun daha fazla olduğu görülmektedir. s ve g parametre düzeyleri açısından bakıldığında ise, DINA modelinde, s ve g parametre düzeyleri azaldıkça madde uyumunun daha fazla olduğu görülmektedir.

Ortak Etkiler Açısından: DINA modeli kullanılarak faktörlerin ortak etkileri incelendiğinde, örneklem büyüklüğü arttıkça, madde sayısı, özellik sayısı, s ve g parametre düzeyleri azaldıkça madde uyumunun daha fazla olduğu göze çarpmaktadır. Ayrıca, özellikler arasındaki korelasyon düzeylerinin madde uyumu açısından önemli bir etkisinin olmadığı yorumu yapılabilir.

4.3. Sınıflama Tutarlılığı İçin Elde Edilen DSO Sonuçlarının Değerlendirilmesi

Temel Etkiler Açısından: Çeşitli örneklem büyüklükleri açısından bakıldığında, DINA modelde örneklem büyüklüğünün doğru sınıflama oranlarına pek bir etkisinin olmadığı görülmektedir. Bu bulgular, Habenicht, Rupp ve Wilhelm'in (2012) yapmış olduğu çalışmadan elde ettikleri bulgularla paralellik göstermektedir. Özellikler arasındaki çeşitli korelasyon düzeyleri açısından bakıldığında, DINA modelinde bu faktörün etkisinin yok denecek kadar az olduğunu söylemek mümkündür. Bu bulgular, Kim'in (2011) ve Henson ve diğerlerinin (2008) yapmış oldukları çalışmaların bulgularıyla paralellik göstermektedir. Özellik sayıları açısından bakıldığında, DINA analiz modelinde, özellik sayısı arttıkça DSO ortalama değerlerinin azaldığı görülmektedir. Habenicht, Rupp ve Wilhelm'in (2012), yapmış oldukları çalışmada özellik sayısı arttıkça DSO'nun azaldığı görülmektedir. Ayrıca, Henson ve Douglas'ın (2005) yapmış oldukları çalışmada da hem DINA hem de RUM modellerinde DSO değerlerinin, düşük özellik sayısında daha yüksek bulunduğu görülmektedir. Başka bir ifadeyle, özellik sayısı arttıkça daha az doğru sınıflamanın yapıldığı söylenebilir. DINA modelinde, madde sayısı arttıkça DSO ortalama değerlerinin arttığı görülmektedir. Başka bir ifadeyle, madde sayısı arttıkça daha fazla doğru sınıflamanın yapıldığı söylenebilir. s ve g parametre düzeyleri açısından bakıldığında, s ve g parametre düzeyleri arttıkça DSO ortalama değerlerinin azaldığı görülmektedir. Başka bir ifadeyle, s ve g parametre düzeyleri arttıkça yapılan doğru sınıflamanın azaldığı söylenebilir.

Ortak Etkiler Açısından: DINA modeli kullanılarak faktörlerin ortak etkileri incelendiğinde, madde sayısı arttıkça, s ve g parametre düzeyleri ve özellik sayısı azaldıkça sınıfla tutarlılığının arttığı görülmektedir. Ayrıca, örneklem büyüklüğünün ve özellikler arası korelasyon düzeylerinin sınıflama tutarlılığı açısından önemsiz olduğu göze çarpmaktadır.

Bu çalışmada, veri tamamlayıcı olmayan modele (DINA) göre üretilmiştir. Aynı çalışma, veri üretimi aşamasında tamamlayıcı olmayan diğer modeller ya da tamamlayıcı modeller kullanılarak yapılabilir. Bu çalışmada, DINA analiz modeli kullanılmıştır. Aynı çalışma, analiz aşamasında başka BTM kullanılarak yapılabilir. Benzer çalışma gerçek veriler üzerinde yapılabilir.

5. KAYNAKLAR

- Ackerman, T. (1996). Graphical representation of multidimensional item response theory analyses. *Applied Psychological Measurement*, 20, 311-329.
- Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: ÖSYM Yayınları.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. USA: Rinehart and Winston Inc.
- 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, 343-362.
- de la Torre, J. (2009). A cognitive diagnosis model for cognitively-based multiple-choice options. *Applied Psychological Measurement*, 33, 163-183.
- de la Torre, J., Hong, Y. & Deng, W. (2010). Factors affecting the item parameter estimation and classification accuracy of the DINA model. *Applied Psychological Measurement*, 47, 227-249.
- DiBello, L. V., Stout, W. F., & Roussos, L. A. (1995). *Unified cognitive/psychometric diagnostic assessment likelihood-based classification techniques*. In P.D. Nichols, S.
- Embretson, S. E. & Reise, S. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum Associates.
- Gelbal, S. (1994). *p madde güçlük indeksi ile rasch modelinin b parametresi ve bunlara dayalı yetenek ölçüleri üzerine bir karşılaştırma*. Yayınlanmamış doktora tezi, Hacettepe Üniversitesi, Ankara.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory: principles and applications*. Boston: Academic Publishers Group.
- Hambleton, R. K. & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement*, 12, 38-47.

- Haertel, E.H. (1984). An application of latent class models to assessment data. *Applied Psychological Measurement* 8, 333-346.
- Haertel, E. H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, 26, 333-352.
- Haertel, E.H. (1990). Continuous and discrete latent structure models of item response data. *Psychometrika* 55, 477-494.
- Henson, R. & Douglas, J. (2005) Test construction for cognitive diagnosis, *Applied Psychological Measurement*, 29(4): 262-277.
- Henson, R. A., Roussos, L., Douglas, J., & He, X. (2008). Cognitive diagnostic attribute level discrimination indices. *Applied Psychological Measurement*, 32, 275-288.
- 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, 191-210.
- Huff, K. & Goodman, D. P. (2007). The demand for cognitive diagnostic assessment. J. P. Leighton ve M. J. Gierl (Eds.), *Cognitive assessment for education: Theory and applications* (19-60). New York: Cambridge University Press.
- Huyn, S. K. (2011). *diagnosing examinees' attributes-mastery using the bayesian inference for binomial proportion: a new method for cognitive diagnostic assessment*. Yayınlanmamış doktora tezi, Georgia Institute of Technology.
- Junker, B. (1999). *Some statistical models and computational methods that may be useful for cognitively relevant assessment*. Prepared for the Committee on the Foundations of Assessment, National Research Council, November 30, 1999.
- 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.
- Lazarsfeld P. F. & Henry N. W. (1968). *Latent Structure Analysis*. Boston: Houghton Mifflin.
- Lord, F. M. & Novick M. R. (1968). *Statistical theories of mental test scores*. New York: Addison- Wesley Publishing Company.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, New Jersey: Lawrence Erlbaum Associates.
- Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer Dordrecht Heidelberg.
- Rupp, A. & Templin, J. (2008). Unique characteristics of diagnostic models: a review of the current state-of-the-art. *Measurement*, 6, 219-262.
- Rupp, A. A., Templin, J., & Henson, R. J. (2010). *Diagnostic measurement: Theory, methods, and applications*. New York: Guilford Press
- Sünbül, Ö. (2011). *Çeşitli boyutluluk özelliklerine sahip yapılarda, madde parametrelerinin değişmezliğinin klasik test teorisi, tek boyutlu madde tepki kuramı ve çok boyutlu madde tepki kuramı çerçevesinde incelenmesi*. Yayınlanmamış doktora tezi, Hacettepe Üniversitesi, Ankara.
- Tatsuoka, K. (1985). A probabilistic model for diagnosing misconceptions in the pattern classification approach. *Journal of Educational Statistics*, 12, 55-73.
- Vermunt, J. K. & Madigson, J. (2004). Local Independence. In A. B. M.S. Lewis Beck, *The Sage Encyclopedia of Social Sciences Research Methods* (ss. 732-733). Thousand Oaks: Sage Publications.
- Wang, Y. C. (2009). *Factor Analytic Models and Cognitive Diagnostic Models: How Comparable Are They?—A Comparison of R-RUM and Compensatory MIRT Model with Respect to Cognitive Feedback*. Yayınlanmamış doktora tezi, The University of North Carolina, Greensboro.
- von Davier, M. (2005). *A general diagnostic model applied to language testing data*. ETS Araştırma Raporu: RR-05-16, Educational Testing Service, Princeton, NJ.

Extended Abstract

Testing industries have focused on constructing measures to assess a single dimension. The test is assumed to measure only one latent or unobserved ability. Each examinee is rank ordered based on the total item scores or a single continuous latent ability. Only a single score is reported. Most psychological and educational tests measure multiple skills and the unidimensionality assumption cannot be met under these circumstances (Hambleton & Swaminathan, 1985). Cognitive diagnosis models (CDMs), as alternative approaches to unidimensional item response models, have received increasing attention in recent years.

Cognitive diagnostic model is designed to measure the specific knowledge structures and the processing skills that examinees possess so as to provide information about the cognitive strengths and weaknesses of examinees (Leighton & Gierl, 2007). Cognitive diagnosis models are developed primarily for the purpose of identifying students' mastery or nonmastery of fine-grained attributes. In the CDM literature, skills have been generically referred to as attributes, and are represented by the binary vector α . CDMs aim to assess whether or not an examinee has mastered a set of several discretely defined skills, or attributes.

A key element in all models of cognitive diagnosis is the so-called Q-matrix. The cognitive diagnosis models (CDMs) use to relate the latent skills to observed behavior (tasks) require a Q-matrix (Tatsuoka, 1983). Henson and Douglas (2005) have defined this Q-matrix as a matrix having elements q_{jk} for J items and K attributes indicating whether mastery of attribute k is required by item j such that, $q_{jk} = 1$, if item j requires attribute k , and 0 otherwise. Obviously, examinees' true skill masteries are unknown, and it is desirable for them to be classified correctly, that is, assigned the correct skill pattern.

Each attribute vector or pattern represents a specific combination of the attribute mastery and nonmastery, and therefore, can be viewed as a unique latent class. The latent classes are considered unstructured if the attributes are independent, in that mastery of one attribute is not a prerequisite to the mastery of another attribute. In addition, because the attributes are latent, a domain with K attributes has 2^K associated latent classes.

Several cognitive diagnostic models have been developed to evaluate examinees' status relative to mastery or non-mastery on each of a set of attributes. One of the commonly used Cognitive Diagnosis Models is the *deterministic, inputs, noisy "and" gate* (DINA; Junker & Sijtsma, 2001) model. The DINA model is a discrete latent variable model that allows inferences about both the cognitive information of the items and the cognitive attributes of the examinees, and has been the foundation of several approaches to cognitive diagnosis and assessment (Tatsuoka, 1995). The DINA model is a parsimonious and interpretable model that requires only two parameters for each item (i.e., g_j and s_j). The noisy component of the model allows the possibility for students who possess all the required attributes for item j to slip and answer the item incorrectly, and students who do not possess all the required attributes to guess and answer item j correctly. Like many CDMs, implementation of the DINA model requires construction of a Q matrix. The DINA model assumes that the deficiency of the student in one attribute cannot be compensated by the mastery of other attributes. This assumption is referred to as *noncompensatory*.

The purpose of this study is to investigate factors affecting the item parameter estimation and classification accuracy of the DINA model one of the Cognitive Diagnostic Models

For this purpose, the data is generated by using noncompensatory model (DINA) according to various factors (sample size, correlation between attributes, the number of attributes, the number of item, s and g parameters levels). For this study 288 experimental cells designed for data simulation and 100 replications were conducted for each cell. The simulated data were analyzed by using DINA models. Data simulation and analyses were conducted by using R 3.0 with CDM package. The output files were organized for parameter estimation, item fit and classification accuracy for both main and interaction effects.

By using DINA analysis model obtained from mean values of "Absolute Mean Bias" (MOY) to estimate g parameter, sample size, number of items and levels of g and s parameters of significant effect were observed. s parameter estimation obtained from the mean values MOY, sample size, level of the correlation between attributes and level of the s and g parameters of a significant effect were observed. By

using DINA analysis model obtained from mean values of RMSEA, sample size, number of item, number of attribute and levels of g and s parameters of significant effect was observed.

Correct Classification Rate (CCR) was computed for examinees' attribute mastery patterns as well as for each attribute marginally. By using DINA analysis model obtained from mean values of Correct Classification Rate, number of item, number of attribute and levels of g and s parameters of significant effect was observed.