

ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

EĞİTİMDE ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME PROGRAMI

İKİ FAKTÖR MODELDE (BİFACTOR) DİKLİK VARSAYIMININ FARKLI
KOŞULLAR ALTINDA SINANMASI

DOKTORA TEZİ

Fulya BARIŞ PEKMEZCİ

Ankara, Ocak, 2018

ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

EĞİTİMDE ÖLÇME VE DEĞERLENDİRME ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME PROGRAMI

İKİ FAKTÖR MODELDE (BİFACTOR) DİKLİK VARSAYIMININ FARKLI
KOŞULLAR ALTINDA SINANMASI

DOKTORA TEZİ

Fulya BARIŞ PEKMEZCİ

DANIŞMAN:
Yrd. Doç. Dr. H. Deniz GÜLLEROĞLU

Ankara, Ocak, 2018

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼'ne

Fulya BARIŞ PEKMEZCİ'nin hazırladığı "İki Faktör Modelde (Bifactor) Diklik Varsayımının Farklı Koşullar Altında Sınanması" başlıklı bu çalışma jürimiz tarafından Ölçme ve Deđerlendirme Anabilim Dalı Ölçme Deđerlendirme Programı'nda DOKTORA TEZİ olarak kabul edilmiştir.

İmza

Başkan Doç. Dr. Nuri DOĐAN

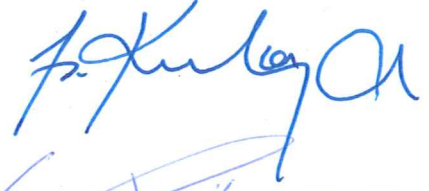


Üye Yrd. Doç. Dr. H. Deniz G¼LLEROĐLU (Danışman)



Üye Doç. Dr. Ömay ÇOKLUK BÖKEOĐLU

Üye Doç. Dr. İsmail KARAKAYA



Üye Yrd. Doç. Dr. Deha DOĐAN



ONAY

Bu tez Ankara Üniversitesi Lisansüstü Eđitim-Öđretim ve Sınav Yönetmeliđi'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından/...../2018 tarihinde uygun gör¼lmüş ve Enstitü Yönetim Kurulunca/...../2018 tarihinde kabul edilmiştir.

Prof. Dr. İsmail G¼VEN
Eđitim Bilimleri Enstitü M¼d¼r¼

ETİK BİLDİRİM

Tez içindeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edilerek sunulduğunu, ayrıca tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.



Fulya BARIŞ PEKMEZCİ

ÖZET

İKİ FAKTÖR MODELDE (BİFACTOR) DİKLİK VARSAYIMININ FARKLI KOŞULLAR ALTINDA SINANMASI

Bariş Pekmezci, Fulya

Doktora, Eğitim Bilimleri Anabilim Dalı, Ölçme Değerlendirme Bölümü

Tez Danışmanı: Yrd. Doç. Dr. H. Deniz GÜLLEROĞLU

Ocak, 2018, xiv + 94 sayfa

Bu araştırmada, Çok Boyutlu Madde Tepki Kuramı'nın modellerinden biri olan İki Faktör Kuramı'nın en önemli varsayımı olan diklik varsayımının, farklı koşullar altında incelemesi amaçlanmıştır. Deneysel araştırma modelindeki bu araştırmanın verileri simülasyon yolu ile üretilmiştir. Simülasyon yolu ile üretilen veriler iki faktör (bifactor) model'e uygun üretilmiştir. Üretilen veriler bağımsız ve bağımlı değişkenler olarak iki gruba ayrılmaktadır. Bağımsız değişkenler; iki spesifik faktör arasında ilişki olan "Model 1", tüm spesifik faktörler arasında ilişki olan "Model 2", diklik ihlali düzeyleri ve test uzunluklarıdır. Bağımlı değişkenler ise madde ve birey parametrelerinin kestirim yanlılıklarıdır.

Verilerin simülasyonunda ve modellerin analizinde "R GUI 3.4.0" yazılımı kullanılmıştır. İki faktör (bifactor) model kestirimleri "R GUI 3.4.0" yazılımında bulunan "mirt" paketi ile, betimsel istatistikler "psych" paketi ile, grafiksel gösterimler ise "ggplot2" ile yapılmıştır. Araştırma kapsamında 200 replikasyon yapılmış, toplam 18 koşul için 3600 (18x200) veri dosyası analiz edilmiştir.

Araştırma sonucunda iki spesifik faktörün ilişkili olma durumunun (Model 1) tüm spesifik faktörlerin ilişkili olma durumu ile (Model 2) hem birey hem de madde parametre kestirim doğruluğunda aynı etkiye sahip olduğu görülmüştür. Ayırt edicilik parametrelerinin kestiriminde diklik ihlali arttıkça yanlılığın arttığı görülmüştür. İki spesifik faktörün ilişkili olma durumunda test uzunluğuna bağlı olarak ayırt edicilik parametre kestiriminde iyileşme görülürken, tüm spesifik faktörlerin ilişkili olması durumunda bu iyileşme görülememiştir. Test uzunluğu arttıkça, ayırt edicilik ve güçlük parametrelerinin kestirim kesinliği yani güvenilirliği

düşmüştür. Güçlük parametreleri, kestirim doğruluğu en düşük parametreler olarak bulunmuştur. Madde sayısını arttırmak, birey parametrelerinin kestirim kesinliğini yani güvenilirliğini arttırmıştır. Birey parametrelerinin kestiriminde, güvenilirliği en düşük parametre kestirimlerinin her iki model için de (Model 1 ve Model 2) en küçük test uzunluğunda olduğu saptanmıştır. Test uzunluğunu arttırmak, birey parametrelerinin kestirim güvenilirliğini de arttırmıştır. Buna rağmen tüm test uzunluklarında ve diklik ihlal düzeylerinde kestirim güvenilirliği en düşük parametreler birey parametreleridir.



SUMMARY

INVESTIGATION OF ORTHOGONALITY ASSUMPTION IN BIFACTOR MODEL UNDER DIFFERENT CONDITIONS

Bariş Pekmezci, Fulya

Ph.D., Department of Measurement and Evaluation

Supervisor: Yrd. Doç. Dr. H. Deniz GÜLLEROĞLU

January 2018, xiv + 94 pages

In this research, it is aimed to examine the assumption of orthogonality, which is the most important assumption of Bifactor Item Response Theory, which is one of the models of Multidimensional Item Response Theory, under different conditions. The data of this research, that is experimental model, is generated by simulation method in accordance with Bifactor Model. Generated data are grouped as dependent and independent variables. Independent variables of this research are; Model 1 which presents correlation between two specific factors, Model 2 which offers correlation between all specific factors, violation levels of orthogonality and test lengths. Dependent variables are estimate bias for item and person parameters.

Data simulation and model analysis was conducted by “R GUI 3.4.0” software. Bifactor model estimates were handled with “mirt” package while the descriptive statistics with the “psych” package. And the graphical representations with the “ggplot2”. Within the scope of the research, 200 replications were made and 36000 (18x200) model analyzes were performed under 18 conditions.

The findings of this research revealed that Model 1 in which two specific factors were correlated and Model 2 in which all specific factors were correlated had the same effect on accuracy of person and item parameter estimations. It was observed that as orthogonality violation is increased bias is increased in estimation of discrimination parameter. There was recovery in parameter estimation depending on test length when two specific factors were correlated, but this recovery was not observed when all the specific factors were correlated. As the test length increases, estimation accuracy and reliability of discrimination and difficulty parameter is decreased. Difficulty parameters’ estimation was found as the least accurate

parameters. Increasing the number of items has increased the estimation accuracy that is reliability of person parameters. In estimation of person parameters, for both models (Model 1 and Model 2), least reliable parameter estimation was acquired at minimum test length. Increasing the test length also increased the estimation accuracy of person parameters. However, the parameters with the lowest estimation reliability are the person parameters in all test lengths and orthogonality violation levels.





Bilime adanmış ömürlere...

İÇİNDEKİLER

	Sayfa
ONAY.....	ii
BİLDİRİM.....	iii
ÖZET.....	iv
SUMMARY.....	vi
İÇİNDEKİLER.....	ix
ÇİZELGELER DİZİNİ.....	xii
ŞEKİLLER DİZİNİ.....	xiii
KISALTMALAR.....	xiv
BÖLÜM 1.....	1
GİRİŞ.....	1
1.1 PROBLEM.....	1
1.1.1 Klasik Test Kuramı.....	2
1.1.2 Madde Tepki Kuramı.....	3
1.1.2.1 Madde Tepki Kuramının Varsayımları.....	4
1.1.2.1.1 Tek Boyutluluk.....	4
1.1.2.1.2 Yerel Bağımsızlık.....	5
1.1.3 Madde Tepki Kuramı Modelleri.....	6
1.1.3.1 İki Faktör (bifactor) Kuramı.....	10
1.1.3.1.1 İki faktör (bifactor) kuramın avantajları ve dezavantajları.....	13
1.1.3.1.2 İki faktör (bifactor) kuramı ile ilgili araştırmalar.....	15
1.2. AMAÇ.....	21
1.3. ÖNEM.....	22
1.4. SINIRLILIKLAR.....	23
BÖLÜM 2.....	24
YÖNTEM.....	24
2.1 Araştırma Modeli.....	24
2.2 Veriler.....	25
2.2.1. Verilerin Üretilmesi.....	25

2.3. Verilerin Analizi.....	31
BÖLÜM 3.....	34
BULGULAR VE YORUM.....	34
3.1. Birinci Araştırma Sorusuna Yönelik Bulgular.....	34
3.1.1. Model 1 İçin Madde Parametrelerinin Kestirim Yanlılıkları...34	
3.1.1.1 Test Uzunluğu 12 Madde İken Madde Parametrelerinin Kestirim Yanlılıkları.....	34
3.1.1.2 Test Uzunluğu 40 Madde İken Madde Parametrelerinin Kestirim Yanlılıkları.....	37
3.1.1.3 Test Uzunluğu 100 Madde İken Madde Parametrelerinin Kestirim Yanlılıkları.....	40
3.1.2. Model 1 İçin Birey Parametrelerinin Kestirim Yanlılıkları..42	
3.1.2.1. Test Uzunluğu 12 Madde İken Birey Parametrelerinin Kestirim Yanlılıkları.....	43
3.1.2.2. Test Uzunluğu 40 Madde İken Birey Parametrelerinin Kestirim Yanlılıkları.....	44
3.1.2.3. Test Uzunluğu 100 Madde İken Birey Parametrelerinin Kestirim Yanlılıkları.....	46
3.2. İkinci Araştırma Sorusuna Yönelik Bulgular.....	47
3.2.1. Model 2 İçin Madde Parametrelerinin Kestirim Yanlılıkları...47	
3.2.1.1 Test Uzunluğu 12 Madde İken Madde Parametrelerinin Kestirim Yanlılıkları.....	48
3.2.1.2 Test Uzunluğu 40 Madde İken Madde Parametrelerinin Kestirim Yanlılıkları.....	50
3.2.1.3 Test Uzunluğu 100 Madde İken Madde Parametrelerinin Kestirim Yanlılıkları.....	51
3.2.2. Model 2 İçin Birey Parametrelerinin Kestirim Yanlılıkları..54	
3.2.2.1. Test Uzunluğu 12 Madde İken Birey Parametrelerinin Kestirim Yanlılıkları.....	54
3.2.2.2. Test Uzunluğu 40 Madde İken Birey Parametrelerinin Kestirim Yanlılıkları.....	55
3.2.2.3. Test Uzunluğu 100 Madde İken Birey Parametrelerinin Kestirim Yanlılıkları.....	56

3.3. Üçüncü Araştırma Sorusuna Yönelik Bulgular.....	57
3.3.1. Ayırt edicilik parametrelerinin model türüne ve test uzunluğuna göre ortalama yanlılık ve RMSE değerler.....	57
3.3.2. Güçlük parametrelerinin model türüne ve test uzunluğuna göre ortalama yanlılık ve RMSE değerleri.....	59
3.3.3. Birey parametrelerinin model türüne ve test uzunluğuna göre ortalama yanlılık ve RMSE değerleri.....	61
BÖLÜM 4.....	64
SONUÇLAR VE ÖNERİLER.....	64
4.1. Sonuçlar.....	64
4.2. Öneriler.....	66
KAYNAKÇA.....	67
EKLER.....	75
EK-A. Simülasyon için R programında yazılan kod.....	76
EK-B. 12 maddelik test uzunluğu için spesifik modellerdeki madde parametreleri kestirim yanlılıkları.....	77
EK-C. 40 maddelik test uzunluğu için spesifik modellerdeki madde parametreleri kestirim yanlılıkları.....	80
EK-D. 100 maddelik test uzunluğu için spesifik modellerdeki madde parametreleri kestirim yanlılıkları.....	83
EK-E. 12 maddelik test uzunluğu için spesifik modellerdeki birey parametrelerinin kestirim yanlılıkları.....	86
EK-F. 40 maddelik test uzunluğu için spesifik modellerdeki birey parametrelerinin kestirim yanlılıkları.....	89
EK-G. 100 maddelik test uzunluğu için spesifik modellerdeki birey parametrelerinin kestirim yanlılıkları.....	92

ÇİZELGELER DİZİNİ

	Sayfa
ÇİZELGE 1. Araştırma Deseni Özeti.....	28
ÇİZELGE 2. Simülasyon Desenindeki Madde ve Birey Parametrelerinin Dağılımları	29
ÇİZELGE 3. Ayırt edicilik ve güçlük katsayılarının 12 madde için kestirim yanlılıkları.....	35
ÇİZELGE 4. Ayırt edicilik ve güçlük katsayılarının 40 madde için kestirim yanlılıkları.....	37
ÇİZELGE 5. Ayırt edicilik ve güçlük katsayılarının 100 madde için kestirim yanlılıkları.....	40
ÇİZELGE 6. Birey parametrelerinin 12 madde için kestirim yanlılıkları.....	43
ÇİZELGE 7. Birey parametrelerinin 40 madde için kestirim yanlılıkları.....	45
ÇİZELGE 8. Birey parametrelerinin 100 madde için kestirim yanlılıkları.....	46
ÇİZELGE 9. Model 2.1 İçin Korelasyon Değerleri.....	48
ÇİZELGE 10. Model 2.2 İçin Korelasyon Değerleri.....	48
ÇİZELGE 11. Model 2.3 İçin Korelasyon Değerleri.....	48
ÇİZELGE 12. Ayırt edicilik ve güçlük katsayılarının 12 madde için kestirim yanlılıkları.....	49
ÇİZELGE 13. Ayırt edicilik ve güçlük katsayılarının 40 madde için kestirim yanlılıkları.....	50
ÇİZELGE 14. Ayırt edicilik ve güçlük katsayılarının 100 madde için kestirim yanlılıkları.....	52
ÇİZELGE 15. Birey parametrelerinin 12 madde için kestirim yanlılıkları.....	55
ÇİZELGE 16. Birey parametrelerinin 40 madde için kestirim yanlılıkları.....	56
ÇİZELGE 17. Birey parametrelerinin 100 madde için kestirim yanlılıkları.....	57
ÇİZELGE 18. Model 1 ve Model 2 için Standart Hata ve ortalama RMSE değerleri.....	58
ÇİZELGE 19. Model 1 ve Model 2 için Standart Hata ve ortalama RMSE değerleri.....	60
ÇİZELGE 20. Model 1 ve Model 2 için Standart Hata ve ortalama RMSE değerleri.....	62

ŞEKİLLER DİZİNİ

	Sayfa
ŞEKİL 1. İki Faktör Model Gösterimi.....	11
ŞEKİL 2. İki spesifik faktörün ilişkili olma durumu.....	26
ŞEKİL 3. Tüm spesifik faktörlerin ilişkili olma durumu.....	26



KISALTMALAR

KTK: Klasik Test Kuramı

MTK :Madde Tepki Kuramı

ÇBMK: Çok Boyutlu Madde Tepki Kuramı

EAP Posteriori İçin Beklenen Değer/Sonsal Beklenti Kestirimi (Expected A Posteriori)

Model 1: İki Faktörün İlişkili Olma Durumu

Model 2: Tüm Faktörlerin İlişkili Olma Durumu

RMSE: Hataların Ortalama Karekökü (Root Mean Square Error of Approximation)

Max: Maksimum

Min: Minimum

SE: Kestirimin Standart Hatası

Bias: Ortalama Yanlılık

\bar{X} : Ortalama

σ : Standart sapma

a : Ayırt Edicilik Parametresi

d: Madde Güçlük Parametresi

θ :Yetenek Parametresi

BÖLÜM 1

GİRİŞ

Bu bölümde araştırmanın problemi açıklanmış, araştırmanın amacı, önemi, sınırlılıklar ve tanımlar belirtilmiş, “Klasik Test Kuramı” ve “Madde Tepki Kuramı” hakkında genel bilgiler verilmiş, “Çok Boyutlu Madde Tepki Kuramı” modellerinden olan “İki Faktör Madde Tepki Kuramı” tanıtılmıştır.

1.1. Problem

İnsan özelliklerini ölçme girişimleri Çin’e kadar uzanmaktadır. Fakat ilk sistematik girişimler başlayalı yüzyıldan fazla zaman olmuştur. Bu durum psikolojinin bağımsız bir bilim dalı olmasıyla eş zamanlıdır. Psikoloji biliminin bağımsız bir bilim dalı olarak ortaya çıkması, insan davranışlarını ölçme girişimleriyle paralellik gösterir (Erkuş, 2003). Bugün psikolojinin araştırma konuları olan tutum, duygu, düşünce, yetenek gibi değişkenlerin ölçülmesi yoluyla insan davranışlarına anlam ve yol verilmektedir. Bu değişkenlerin ölçülmesi oldukça meşakkatlidir çünkü bu değişkenler doğrudan gözlenemezler; ancak bunların tümü dolaylı olarak varlığı bilinen ve kanıtlanan birer olgudur (Karasar, 2010). Doğrudan gözlenemeyen bu özellikler ise dolaylı ölçme yöntemiyle ölçülebilmektedir. Doğrudan ölçmede ölçülmek istenen kavram ve onu temsil eden değişken birbiriyle doğrudan ilgilidir; doğal olarak ya da tanımlama yoluyla üzerinde anlaşmaya varılmış (metre, kilogram vb.) standart birimleri vardır (Karasar, 2010). Dolaylı ölçmelerde bu türden birebir eşleme yapılması ya da standart birimler geliştirilmesi çok güç ya da olanaksızdır. Bu durumda ölçme dolaylı göstergeler üzerinden yapılır. Doğrudan gözlenemeyen ve ölçülmek istenen bu psikolojik özellikler literatürde “yapı (construct)” olarak ifade edilmekte olup hipotetik kavramlardır (Crocker ve Algina, 1986). Bu psikolojik yapıların en doğru şekilde ortaya konması amacıyla çeşitli kuramlar geliştirilmiştir. Bu kuramlardan ilki Klasik Test Kuramı (KTK)’dir.

1.1.1.Klasik Test Kuramı (KTK)

Klasik test kuramı (KTK)'nın temellerini 1904 yılında Charles Spearman'ın fikirleri oluşturmuştur (Smith, 2009). KTK, gerçek puanın gözlenen puandan kestirilmeye çalışıldığı bir kuramdır. Bu yüzden KTK'ya gerçek puan modeli de denilmektedir. Kuramda gerçek puanı kestirmek amacıyla, ölçme yoluyla elde edilen puana bireyin gözlenen puanı denir. Bu kurama göre, bir bireyin bir ölçme aracına ilişkin gözlenen puanı (X); gerçek puan (T) ve hata puanı (E)'nin toplamıdır (Crocker ve Algina, 1986; Wainer ve Thissen, 2001).

KTK'da, ölçülmek istenen özelliğe ilişkin puan bireyin ölçekten aldığı ham puana karşılık gelirken, bireyin yetenek ölçüsü, ham puanların standartlaştırılmış değerleri olmaktadır (Doğan, 2002). Her kuramın olduğu gibi Klasik Test Kuramı'nın da sayıltıları vardır.

Klasik Test Kuramı'na ilişkin sayıltılar:

- 1) Hata puanlarının ortalaması sıfırdır.
- 2) Gerçek puan ile hata puanı arasındaki korelasyon sıfırdır.
- 3) İki paralel ölçmeye ilişkin hata puanları arasındaki korelasyon sıfırdır.
- 4) İki farklı ölçmeye ait gerçek puanlar ve hata puanları arasındaki korelasyon sıfırdır. (Gulliksen, 1950; Baykul, 2000).

KTK'nın dört temel sayıltısı var olarak bilinse de, ölçülmesi amaçlanan değişkenin tek boyutlu olması ve bireylere uygulanan testlerin paralel olması gibi sayıltıları da vardır (Sünbül, 2011). KTK'nın en büyük avantajı, varsayımlarının test edilebilmesinin kolay olmasıdır. Böylelikle kuramın varsayımları çoğu durumda kolaylıkla test edilebilmekte ve kuram, farklı uygulamalara uyarlanabilmektedir (Hambleton ve Jones, 1993). KTK uygulamaları diğer ölçme kuramlarına göre daha kolay ve karışık olmayan işlemler gerektirmesi açısından yıllarca kullanılmıştır. Bu avantajlarının yanında, KTK'nın testten elde edilen madde ve test istatistiklerinin, testin uygulandığı gruba ve uygulanan teste bağımlı olması, bütün yetenek ranjı için tek bir hata kestirimi yapılması gibi sınırlılıkları vardır (Crocker ve Algina, 1986; Hambleton ve Swaminathan, 1985; Embretson ve Reise, 2000; Lord, 1980; Lord ve Novick, 1968). KTK'nın bu temel sınırlılıklarının giderilebileceği iddiasıyla, 1930'lu yılların sonunda "Örtük Özellikler Kuramı" olarak da bilinen "Madde Tepki

Kuramı” ortaya çıkmıştır (Crocker ve Algina, 1986).

1.1.2. Madde Tepki Kuramı (MTK)

Klasik test kuramının sınırlılıklarının giderebilmesi amacıyla 1930’lu yılların sonunda ortaya çıkan Madde Tepki Kuramı (MTK) aynı zamanda örtük özellikler kuramı olarak da adlandırılır (Crocker ve Algina, 1986). Tucker 1940 yılında örtük özellik kuramının en önemli kavramlarından olan "madde karakteristik eğrisi" kavramını ilk defa kullanmıştır. MTK’nın istatistiksel temellerinde ise Lord, Novick ve Birnbaum’un 1968 yılındaki çalışmaları yatar. Fakat, Lord’un 1950’li yılların başındaki çalışmaları örtük özellikler kuramının başlangıcı olarak kabul edilir. Lord, ilk kez bu kuram doğrultusunda "normal ogive" modelini geliştirip, başarı ve yetenek testlerinde kullanmıştır. Birnbaum ise 1950’li yılların sonunda "normal ogive" modelinin yerine, bu alanda çalışmaları kolaylaştıran "lojistik" modeli geliştirmiştir (Baker, 2001; Embretson ve Reise, 2000; Hambleton ve Swaminathan, 1985)

Lord (1953), KTK’da kullanılan, gözlenen ve gerçek puanların MTK’da kullanılan yetenek puanlarıyla aynı anlama gelmediklerini, gözlenen ve gerçek puanlar teste bağımlı iken, yetenek puanlarının testten bağımsız olduğunu ve bu nedenle de yetenek puanlarının daha temel olduğunu vurgulamıştır (akt: Hambleton ve Jones, 1993). MTK’na göre, bireyler zor testlerde daha düşük gerçek puana sahip olurken, kolay testlerde daha yüksek gerçek puana sahip olabilmektedirler; ancak bireylerin yetenekleri sabit kalmaktadır (Hambleton ve Jones, 1993). KTK’da gözlenen puanlar aracılığıyla gerçek puanlar hakkında çıkarımlar yapılırken, MTK’da örtük özellik ile yanıtlama davranışı arasındaki ilişki, olasılıklı bir model ile tanımlanmaya çalışılır (Embretson ve Reise, 2000).

MTK’nın en önemli avantajlarından birisi, madde güçlük değerlerinin ve yetenek düzeylerinin aynı ölçek düzeyinde yer almasıdır (Hambleton, Swaminathan ve Rogers, 1991; Spencer, 2004). Yani MTK, bireyin örtük özellikteki konumunu model parametrelerini (madde güçlüğü, madde ayırt ediciliği ve şans parametresi) göz önüne alarak tahmin etmektedir. MTK’nın diğer avantajları ise; yanıtlayıcıların yeteneklerinin daha kesin (precision of measurement) ve daha küçük ölçme hataları ile ortaya konabilmesi, kestirilen parametrelerin örneklem arası değişmezliği, bireylerin karşılaştırılmasında yetenek düzeylerini kullanarak daha kesin sonuçlara varılmasıdır (Embretson ve Reise, 2000; Hambleton, Swaminathan ve Rogers, 1991).

1.1.2.1. Madde Tepki Kuramının Varsayımları

Madde tepki kuramı, güçlü matematiksel ve istatistiksel varsayımlara dayanmaktadır. Bu varsayımların karşılanması koşulu ile bu kuramın avantajlarından faydalanılabilmektedir.

1.1.2.1.1. Tek boyutluluk (unidimensionality)

Tek boyutluluk varsayımı, madde yanıtları arasındaki ortak varyansı tek bir örtük değişkenin açıklaması demektir (Crocker ve Algina, 1986). Diğer bir deyişle, bireyin bir madde üzerindeki performansını etkileyen sadece bir yetenek türü olması durumudur (Lord ve Novick, 1968; Lord, 1980; Hambleton ve Swaminathan, 1985; Crocker ve Algina, 1986; Embretson ve Reise, 2000). Yani bir ölçme aracındaki maddeleri, bir arada tutan tek bir ortak özellik olmalıdır. Eğer bireyler, yapıların hiçbirinde farklılaşmıyor, tüm bireysel farklılıklar diğer yapılardan kaynaklanıyorsa, tepkiler matematiksel olarak tek boyutlu olacaktır. Farklı maddeler, yapıların farklı kombinasyonlarından istifade ediyorsa ve bireyler her iki yapıda da çeşitleniyorsa tepkiler çok boyutlu olacaktır. Örneğin, bir testin amacı bilgi düzeyini ölçmek olmasına rağmen, bu testin sınırlı zamanda yapılması istendiğinde, testin sonuna doğru maddeleri hızlı cevaplama yapısı, bilgiyi ölçme yapısının önüne geçecek ve çok boyutluluk oluşacaktır. Bireylerin tümü maddelere hızlı cevap verebilecek yetenek koşulunda olduğu takdirde ise tepkiler tek boyutlu olacaktır (Erkuş, Sünbül, Sünbül, Yormaz ve Aşiret 2017). Sonuç olarak testi oluşturan maddelerin yanıtlanmasında birden fazla özellik etkin olursa, tek boyutluluk varsayımının karşılanması mümkün olmayacaktır. Bireylerin maddeleri yanıtlarken tek bir boyut doğrultusunda hareket etmeleri beklenemez; bununla beraber, ölçülmek istenen değişken kompleks bir yapıya sahip olabilir. Böyle bir yapı tek boyutluluk sayılına dayanan MTK modelleri ile ölçülemez. Bu gibi nedenlerden dolayı, tek boyutlu MTK modelleri, bu sayılıntının pratikte karşılanma güçlüğünden kaynaklı olarak oldukça eleştirilmiştir (Embretson ve Reise, 2000; Kirisci, Hsu ve Yu, 2001).

Aslında MTK modellerinde bahsedilen bu varsayım “esas boyutluluktur” ve çoğunlukla tek boyutluluk olarak ifade edilir (Embretson & Reise, 2000). Fakat her test, her yanıt grubu aslında bir derece de çok boyutludur (Harrison, 1986). Pratikte tamamen tek boyutlu veri elde edilmesi yüksek olasılıkla mümkün değildir. Bundan

dolayı, Drasgow ve Parsons (1983) “yeterli derecede tek boyutlu” ismini, Stout (1990) ise “temel olarak tek boyutlu” ismini bir baskın örtük boyutun dikkate alındığı ve alt boyutların ihmal edildiği durumlar için kullanmıştır. Sonuç olarak boyutluluk; teorik tek boyutluluk, temel olarak tek boyutluluk ve çok boyutluluğa doğru uzanan bir süreç olarak tanımlanabilir (Stout, 1990)

Geçmişten günümüze boyutluluğun değerlendirilmesi ile ilgili bir çok ölçüt kullanılmıştır. Örneğin özdeğerlerin 1.0 den büyük olması (Kaiser, 1960), scree plot/saçılım grafiği’ndeki eğrinin kırılma noktası/dirsek (Cattell, 1966), omega değeri (Heise ve Bohrnstedt, 1970; McDonald, 1970), paralel analiz vb gibi. Fakat bununla birlikte şimdiye kadar tek boyutluluğun istatistiksel olarak yeterli kabul edilebilecek tek bir ölçüsü bulunmamaktadır (Quinn, 2014). Sıklıkla birkaç standarda bağlı indeksler, boyutluluğu ölçmede kullanılmaktadır fakat bunlar da verinin yapısının değerlendirilmesinde çelişkili sonuçlara neden olmaktadır. Bu boşluk araştırmacılara verinin çok boyutlu olduğu düşünüldüğü zaman toplam puanının yanında alt boyut puanları rapor etme özgürlüğü vermiştir (Quinn, 2014). Veri setinin temel olarak tek boyutlu olduğuna karar verildiği zaman bunu toplam puan ile ifade etmek yeterli olacaktır. Bu tür verilerde küçük (alt boyutlar) boyutlar tespit edilebilir fakat bunlar ihmal edilebilir. Bu durumun aksine, verinin çok boyutlu olduğuna karar verildiğinde, alt boyutların varyansa olan katkıları açık bir şekilde verilmelidir. Tek boyutluluğun standart bir ölçüsü olmamasıyla beraber Quinn (2014)’in yaptığı çalışmaya göre iki etkenli (bifactor) madde tepki kuramı da boyutluluk hakkında bilgi vermede kullanılabilir. Ayrıca bir çok araştırmacı da ilgilendikleri ölçeklerin boyutluluk analizlerini iki faktör (bifactor) modeline göre yapmışlardır (Boduszek, Dhingra, Shevlin ve Egan 2014; Brouwer, Meijer, Weekers ve Baneke, 2008; Golay ve Lecerf, 2011; Reise, Ventura ve diğ., 2011; Martel, Von Eye ve Nigg, 2010).

1.1.2.1.2. Yerel bağımsızlık (local dependency)

MTK’nın ikinci varsayımı yerel bağımsızlıktır. Yerel bağımsızlık, test performansını etkileyen yetenek sabit tutulduğunda, bireylerin maddelere vereceği tepkilerin istatistiksel olarak bağımsız veya ilişkisiz olmasıdır (Hambleton ve Swaminathan, 1985; Lord ve Novick, 1968; Lord, 1980; Reckase, 2009).

Yerel bağımsızlık varsayımının karşılanabilmesi için önce tek boyutluluk varsayımının karşılanması gerekmektedir. Eğer bir maddenin yanıtlanmasında birden fazla yetenek etkili ise, maddeleri bağımsız yapmak için süreçte etkili olan yeteneklerden birini sabitlemek yeterli olmayacaktır ve sabitlenmemiş olan diğer yetenek, maddeler arasında bağımlılık veya ilişki oluşturacaktır. Yerel bağımsızlığın sağlanamaması, özellikle madde ve birey parametrelerinin kestirilmesinde kullanılan maksimum olabilirlik (maximum likelihood) fonksiyonunun temelini oluşturan, olabilirlik (likelihood) fonksiyonlarından hesaplanacak olan olasılıkları tehlikeye sokacaktır. Çünkü olabilirlik fonksiyonları, belirli bir yetenek koşulu altında, maddelere bağımsızmış gibi davranarak olasılık sonuçları üretmektedir (Sünbül, 2011). Uygulamada yerel bağımsızlık sayıltısı bir çok faktörden etkilenmektedir. Örneğin; ortak köke dayalı maddelerde, bir sonraki maddeye verilecek tepkiler kendisinden önceki maddeye verilen tepkilerden etkilenmektedir. Diğer faktörler ise; hız testi olması, öğrenciler pratik yapma oranlarının farklılaşması ve maddelerin birbirlerine zincirlendiği yani maddeleri cevaplamanın diğer maddeleri cevaplamayı etkilediği durumlardır (Embretson ve Reise, 2000).

Yerel bağımsızlık varsayımının karşılanıp karşılanmadığını değerlendirmek için son yıllarda durumsal kovaryans hesaplamalarından yola çıkan çeşitli indeksler geliştirilmiştir. Bununla birlikte; genel olarak tek boyutluluk varsayımının karşılanması, yerel bağımsızlık varsayımının da yerine getirilmesi için yeterli görülmektedir (Sünbül, 2011). Lord (1980), yerel bağımsızlığı ek bir varsayım olarak değil, tek boyutluluğun bir getirisi olarak görmektedir.

Bu varsayımların karşılanması sonucunda, son olarak ise Madde Tepki Kuramının avantajlarından yararlanmak için model-veri uyumu incelenmelidir. Model-veri uyumu üç yolla incelenir; Verilerin modelin varsayımlarını sağlayıp sağlamadığı test edilir, kullanılan modelin MTK'nın sağladığı avantajları yerine getirip getirmediği incelenir ve gözlenen, kestirilen sonuçların uyumuna bakılır (Hambleton ve diğ., 1991).

1.1.3. Madde Tepki Kuramı Modelleri

MTK ile ilgili olarak yürütülen ilk çalışmada bir bireyin bir testte yer alan maddelere verdiği yanıtların doğru veya yanlış olarak sınıflandığı modeller üzerine

durulmuştur (McKinley ve Reckase, 1982). Bir, iki ve üç parametrelili bu MTK modelleri, bireyin yeteneğinin yalnızca varlığını ve yokluğunu dikkate alan iki kategorili modeller olarak bilinirler. Ancak tüm cevaplayıcı-madde ilişkilerinin iki kategorili modellerle incelenmesi uygun olmayabilir. Örneğin kısmi olarak doğru yanıtlanmış bir maddeye puan vermek için ikiden fazla kategori içeren bir modele ihtiyaç duyulur (De Ayala, 2009). Aynı zamanda tek boyutlu iki kategorili MTK modelleri iki ve daha fazla örtük özellikten etkilenen maddeler için de uygun değildir. Bireyler düzeyinde düşünüldüğünde farklı strateji kullanan, farklı bilgi yapılarına, seviyelerine ve yorumlarına sahip bireylere uygulanan maddeler için de bu model uygun olmayacaktır. Bu tür maddeler için çok boyutlu MTK modelleri daha uygundur (Emberson ve Reise, 2000). Bu durum MTK modellerinde farklı sınıflama gereksinimine neden olmuştur (Jang, 2014).

Madde tepki Kuramı modellerinin sınıflandırılmasının gerekliliği ile bu sınıflandırmalar çeşitli ölçütler kullanılarak yapılmıştır. McDonald (1982) madde tepki kuramı modellerini;

Testle ölçülen baskın bir özelliğin olup olmamasına göre;

1. Tek boyutlu modeller (Unidimensional models)
2. Çok boyutlu modeller (Multidimensional models)

Testte yer alan maddelere verilen tepkilerin sayısına göre;

1. İkili puanlanan modeller (Dichotomous models)
2. Çoklu puanlanan modeller (Polythomous models)

Doğrusallık yapısına göre;

1. Doğrusal modeller (Linear models)
2. Doğrusal olmayan modeller (Nonlinear models) olarak sınıflandırmıştır.

Ayrıca çok boyutlu MTK'da ikili puanlanan maddeler için modeller, genel olarak örtük özelliklerden birinin diğerini telafi edip edememesine göre, telafisel (compensatory) ve telafisel olmayan (non-compensatory) olmak üzere iki şekilde sınıflandırılmaktadır (Ackerman, 1996; Reckase, 2009; Sijtsma ve Junker, 2006). Bu sınıflamalarda yer almayan fakat son zamanlarda eğitim ve psikolojide yaygın olarak kullanılan bir diğer madde tepki kuramı modeli ise iki faktör (bifactor) modelidir.

İki faktör (bifactor) modeli, çok boyutlu MTK'nın bir uzantısıdır. Çok boyutlu MTK modelleri kendi içinde basit yapı ve karmaşık yapı olarak da ikiye ayrılmaktadır. Basit yapı çok boyutlu modeller, tüm maddelerin sadece bir boyutta yüklenmesi ile olurken, karmaşık yapı çok boyutlu modellerde ise bazı maddeler birden fazla boyutta yüklenir. Bundan dolayı iki etkenli (bifactor) model karmaşık yapı çok boyutlu bir modeldir (McDonald, 1982).

İki faktör modelinin anlaşılması için öncelikle tek boyutlu ve çok boyutlu modellerin anlaşılması gerekmektedir. Çok boyutlu MTK modelleri, tek boyutlu modellerin geliştirilmiş halidir. İki faktör modeli (örn; Chen, West ve Sousa, 2006; Ebesutani vd., 2011; Patrick, Hicks, Nichol ve Krueger, 2007, Reise, Morizot ve Hays, 2007) ise çok boyutlu MTK modelleri içinde yer almaktadır.

Tek boyutlu MTK modelleri, tüm maddelerin baskın tek bir özelliği ölçtüğüne dair bir varsayıma sahiptir. Yani, kurama göre ölçülen yapı tek boyutlu bir yapıdır ve bu yapı maddelerin birbiriyle olan ilişkisini açıklamaktadır. Tek boyutlu MTK modellerinin uygulanması matematiksel hesaplamalarının kolaylığı açısından basittir ve analizleri BiLOG-MG, Testfact, WinBUGS, OpenBUGS, Mplus, and IRTpro gibi programlarla yapılabilir. Fakat eğitim ve psikolojide birçok yapı çok boyutludur ve tek boyutlu madde tepki kuramının gerektirdiği tek boyutluluk varsayımı çoğu kez ihlal edilir (Zheng, 2013).

Tek boyutluluk, sağlanması zor bir varsayım olmasına rağmen tek boyutlu MTK, uygulanma kolaylığı açısından çok boyutlu madde tepki kuramına göre daha sık kullanılmıştır. Fakat, son yıllarda tek boyutluluktan sapma gösteren modeller için Çok Boyutlu Madde Tepki Kuramı (ÇBMTK) kestirim sürecinde ilerlemeler yaşanmıştır (Edwards ve Edelen, 2009; Reckase, 2009). ÇBMTK modelleri geliştirilmiş olsa da, tüm MTK modelleri için öncelikle “esas boyutluluk” varsayımı karşılanmalıdır. Bundan dolayı önce boyutluluğun değerlendirilmesi, sonrasında ise modelin seçilmesi gerekmektedir. Değerlendirilen veri eğer “temel olarak tek boyutlu” ise veri tek boyutlu MTK için uygun, değil ise, yani veri çok boyutluluk eğilimi gösteriyor ise, ÇBMTK modellerinin kullanılması uygun olacaktır.

Çok boyutlu MTK, faktör analizi ya da yapısal eşitlik modelinin özel bir hali veya tek boyutlu MTK'nın bir uzantısı olarak (McKinley ve Reckase, 1982) kullanılmaktadır. MTK, olasılık modelleri yoluyla madde karakteristikleri ile cevaplayıcıların örtük özellikleri arasındaki ilişkiyi açıklar. Çok boyutlu MTK, test maddelerinin karakteristikleri ve yanıtlayıcı parametreleri tarafından tanımlanan

uzaydaki yeriyle maddeyi doğru yanıtlama olasılığını birbirine bağlayan bir ilişkinin fonksiyonel formunu kullanarak test maddelerinin karakteristiklerini tanımlamaktadır (Reckase, 1997).

Çok boyutlu MTK model kestirimleri, tek boyutlu modellerden farklılaşmaktadır. Örneğin, tek boyutlu MTK'da doğru yanıtlama olasılığı tek bir özellikten etkilendiği için, doğru yanıtlama olasılıkları, madde karakteristik eğrisi grafikleriyle gösterilir. Çok boyutlu MTK'da ise doğru yanıtlama olasılığı artık iki veya daha fazla özellikten etkilenir ve madde karakteristik yüzeyi olarak belirtilir. Benzer şekilde çok boyutlu MTK'da test karakteristik eğrisi de test karakteristik yüzeyi formunu alır. Çok boyutlu MTK'da, birden fazla örtük özellik ile ilgilenildiği için, artık yüzeyle (facet) çalışılmaya başlanır ve bu durum da madde ayırt edicilik indeksi, madde güçlük indeksi gibi özel madde karakteristiklerinin tanımlarının değiştirilmesini zorunlu kılmaktadır (Ackerman, 1994).

Çok boyutlu MTK modelleri bilindiği üzere verilerin çok boyutlu özellik gösterdiği durumlarda kullanılmaktadır. Çok boyutlu veriler tek boyutlu modeller ile kestirildiği zaman sadece birincil faktör dikkate alınmaktadır çünkü Tek Boyutlu MTK modelleri sadece birincil faktöre ilişkin örtük yeteneği kestirir. Yani çok yüzeyli (multifaceted) alt alanlar bu model ile dikkate alınmaz ve bu durumda, farklı alt boyutlardaki maddeler arasında yüksek korelasyonlar yerel bağımsızlık varsayımının ihlal edilmesine neden olur (Zheng, 2013). Demars (2006), yerel bağımsızlık varsayımının ihlal edilmesinin, güvenilirliği şişirdiğini, yetenek ve madde parametrelerinin kestirimine ilişkin standart hataları etkilediğini ve buna bağlı olarak yetenek ve madde parametrelerinin hatalı kestirildiğini belirtmektedir. Bu nedenlerden dolayı çok boyutlu verilerde yetenek ve madde parametrelerinin, çok boyutlu madde tepki kuramı modelleri kullanılarak kestirilmesi daha uygun ve doğru olacaktır.

Çok boyutlu MTK modelleri alan yazında zamanla daha sık kullanılmaya başlanmıştır. Bunun nedeni çok boyutlu MTK modellerinin eğitimsel ve psikolojik yapıların karmaşıklığını dikkate almasıdır. Bu modeller, genel faktör dışında grup (spesifik) faktörlerine sahip karmaşık yapıların çok boyutluluğunu doğru bir şekilde modellemektedir (Zheng, 2013). Ayrıca bu boyutlardan elde edilen puanlar tanısal amaçlar için kullanılmaktadır. Örneğin; bir öğrencinin matematik testinin alt boyutlarındaki düşük puanları, öğrencinin o alana henüz yeterince hakim olmadığını göstermektedir. Bununla birlikte çok boyutlu madde tepki modelleri genel (birincil)

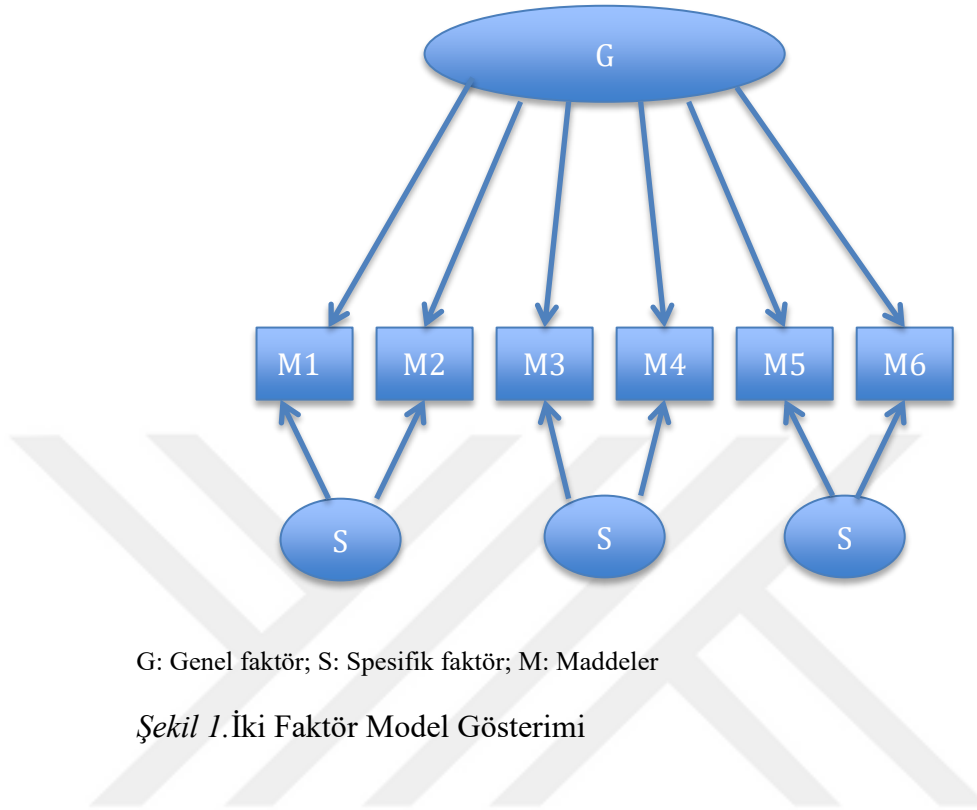
faktörün değil sadece alt faktörlerin parametrelerini kestirebilmektedir. Reise, Moriziot ve Hays (2007), eğer maddeler genel faktörde düşük; fakat grup (spesifik) faktörlerinde yüksek yüklere sahip ise çok boyutlu madde tepki kuramının kullanılmasını önermiştir. Bundan dolayı çok boyutlu madde tepki kuramı modelleri eğer maddeler genel (birincil) faktörde daha az yüklenmiş fakat grup faktörlerinde (spesifik faktörlerde) daha yüksek yüklere sahip ise uygundur. Çok boyutlu MTK'nın sadece alt faktörlere yönelik kestirimler yapması ve genel faktöre yönelik bir çıkarım yapmaması, iki faktör (bifactor) model uygulamalarını artırmıştır.

Çok boyutlu MTK modellerinin aksine iki faktör (bifactor) modeli diğer madde tepki kuramı modellerinden de farklı olarak genel (birincil) faktör çıkarır ve genel faktör ile grup (spesifik) faktörlerin birbiriyle ilişkisi yoktur. Ayrıca iki faktör (bifactor) modeli birden fazla örtük yapıyı barındırır ve çok boyutlu madde tepki kuramı alt faktörlere yönelik kestirimler yaparken iki faktör (bifactor) madde tepki kuramı hem genel faktör hem de alt faktörler için kestirimler yapar (Canviez 2016; Zheng, 2013)

1.1.3.1. İki Faktör (Bifactor) Kuramı

İki faktör (bifactor) modeli, adından da anlaşıldığı üzere, Spearman'ın iki-faktör kuramının bir uzantısı olarak Holzinger ve Swineford (1937) tarafından geliştirilmiştir (Zheng, 2013). Spearman'ın iki faktör kuramına göre, insanlar g faktörü olarak nitelenebilecek tek bir bilişsel kapasiteyle doludur, kalımsal olan ve yaşam boyu değişmeyen "g" bir bireyin soyut düşünme ve problem çözme becerisi, karmaşık zihinsel işlemleri yapma yeteneğini ifade etmektedir. İkinci faktör olan "s" ise, bireyin matematiksel ya da sözel yetenekleriyle ilgili kendine özgü zihinsel yeteneklerini temsil etmektedir. Özet olarak Spearman'ın bilişsel yeteneklerin/becerilerin kavramsallaştırılmasına göre, tüm değişkenler genel bir faktör ile ilişkilidir ve ayrıca hepsi özel bir faktör de içermektedir. İki faktör (bifactor) modeli ise, faktörler tarafından açıklanan (ortak varyans/commonality) bir genel faktörün ve birden fazla spesifik (özgül) faktörün olduğunu ve bu özgül (unique) etkilerin genel faktör üzerinde etkisinin olduğunu varsaymaktadır. Ayrıca en önemli varsayımı ise diklik varsayımıdır. Diklik varsayımı, spesifik faktörlerin birbiri ile ve genel faktör ile ilişkisiz olmasını gerektirir. Yani iki faktör (bifactor)

modelinde, faktörler diktir. İki faktör (bifactor) modelinin şematik gösterimi Şekil 1’de verilmiştir.



Şekil 1. İki Faktör Model Gösterimi

İki faktör (bifactor) MTK modeli ile ilgili olarak ilk Holzinger ve Swineford (1937) çalışma yapmış ve çalışmalarında bu yöntemi kısaca tanıtmakla birlikte bu yöntemin daha karmaşık değişkenlerle nasıl uygulanacağını göstermiştir. Gibbons ve diğ. (2007) ise Holzinger ve Swineford (1937)’un çalışmalarını geliştirerek iki etkenli (bifactor) modeli çoklu puanlanan maddeler için (çoklu puanlanan dereceleme ölçekleri gibi) uyarlamışlardır.

Holzinger ve Swineford (1937)’un çalışmalarında da bahsedildiği gibi iki faktör (bifactor) madde tepki kuramı modelleri hem ikili hem de çoklu puanlanan veriler için uygundur. İkili puanlanan veriler için kestirim iki parametrelili lojistik iki etkenli (bifactor) model ya da üç parametrelili lojistik iki etkenli (bifactor) model ile yapılabilir. İki parametrelili lojistik iki etkenli (bifactor) modelde ayırt edicilik (a) ve güçlük (b) parametreleri kestirilirken, üç parametrelili lojistik iki etkenli (bifactor) modelde bunlara ek olarak şans (c) parametresi de kestirilmektedir. Madde ayırt ediciliği faktör yükleri ile aynı işleve sahiptir ve örtük faktörde yanıtlayıcılardan yüksek puanlara/yeterliğe sahip olanlar ile düşük puanlara/yeterliğe sahip olanları

maddenin ne kadar iyi ayırdığıdır. Madde güçlüğü, madde eşik değeri ya da madde yer parametresi olarak da bilinir, yanıtlayıcıların doğru ya da daha pozitif yanıt verme olasılıkları ile ilgilidir. Şans parametresi ise sadece tahmin ile doğru cevap verme olasılığıdır. Çoklu puanlanan verilerde, sadece iki parametrelili lojistik iki faktör (bifactor) modeli uygundur çünkü madde yanıtları, tahmin ile cevap verilmesine uygun değildir ve cevaplarda doğru ya da yanlış kaygısı yoktur (Örneğin; likert tipi ölçekler). Bundan dolayı çoklu puanlanan verilerde tahmin söz konusu değildir ve şans parametresi genellikle kestirilmez (Zheng, 2013).

Tüm MTK modellerinde olduğu gibi iki faktör (bifactor) modelinin de kendine özgü varsayımları vardır. İki faktör modelinin en önemli varsayımlarından biri verinin hem genel faktörü hem de diğer (spesifik) faktörleri ölçmesidir. Bu varsayım karşılanması zor bir varsayım olmamakla birlikte çok boyutlu veriyi gerektirmektedir. Diğer varsayım olan faktörlerin dik ve birbirinden bağımsız (ilişkisiz) olması varsayımı ise pratikte karşılanması çok da mümkün görünmemektedir. İki faktör (bifactor) modelinin gerektirdiği diklik varsayımı, modelin sınırlandırılmış (Reise, Moore ve Haviland, 2010), geleneksel (Martel, Von Eye ve Nigg, 2010) veya iki etkenli modelin kanonik versiyonu (Chen, West ve Sousa, 2006) şeklinde çeşitli adlandırmalara neden olmuştur. Reise Moore ve Haviland, (2010)'e göre "sınırlandırılmış (restricted) iki faktör (bifactor) modeli sadece verinin çok boyutlu olmasını değil aynı zamanda çok iyi yapılandırılmış (her maddenin genel özelliği ve sadece bir alt özelliği ölçmesi) olmasını gerektirmektedir. Diğer bir deyişle test geliştiriciler sadece birincil faktörü ve ayrıca bir alt alanı ölçen maddeler yazmalıdır. Asıl problem ise pratikte böyle maddeler yazmanın çok zor olduğudur. Örneğin, sayılar, cebir ve geometri gibi üç alt alandan oluşan iyi bir şekilde tasarlanmış matematik testi hazırlandığında, diklik varsayımı gereği bu alt alanların paketlenmiş üniteler gibi birbirinden bağımsız ve ilişkisiz olması gerekmektedir. Cebir ve geometri sorularının sayılar bilgisi olmadan doğru şekilde cevaplanabileceğini hayal etmek bile oldukça güçtür. Bu maddeler arasında çapraz yüklenmeler olması kaçınılmazdır. Bu örtük faktörler (cebir ve geometri) teorik olarak mümkün olabilir ama gerçekte bazı çakışmalar olacaktır. İlişkili faktörleri dik olmaya zorlamak, ölçülen yapı ile ilgili olarak bilgi kaybına neden olacak ve güvenilir olmayan parametre kestirimleri ile sonuçlanacaktır. Özetlemek gerekirse, diklik varsayımı ihlalinde sonuçların yorumlanması zorlaşacaktır. Reise, Moore, ve Haviland, (2010), dik olmayan verilere diklik varsayımı gerektiren bir modeli

uygulamanın madde ve birey parametrelerinin kestiriminde yanlılığa neden olacağını belirtmişlerdir.

1.1.3.1.1. İki faktör (bifactor) modelin avantajları ve dezavantajları

İki faktör modeli, psikolojik yapıların ölçülmesinde sıklıkla kullanılmaktadır ve psikolojik yapılar genellikle birkaç ilişkili yüzey (facet) ile nitelendirilir. Bu duruma psikolojinin ilgilendiği birçok konudan örnekler (depresyon, kişisel özellikler/bireysel özellikler) verilebilir. Çok yüzeyli (multifaceted) yapılar ile ilgilenen araştırmacılar iki sorunsalla karşılaşmaktadır; çok yüzeyli (multifaceted) yapılar, her alt ölçeğin bir yapıyı ölçtüğü birkaç alt ölçekle mi, yoksa tek bir ölçeğin genel yapıyı ölçtüğü ölçekle mi değerlendirilmeli. Bu sorun beraberinde “toplam puan yaklaşımı (total score approach) mı?” yoksa “bireysel puan yaklaşımı (individual score approach) mı?” tartışmasını getirmektedir.

Toplam puan yaklaşımı, çok yüzeyli yapılarda yüzeylerin toplamını ya da ortalamasını alır ve her yüzeyi eşit ağırlıklandırılarak bu yüzeylerin bir bileşimini (composite) oluşturur. Elde edilen bu bileşim puanı, ilgilenilen dışsal değişkenlerin yordayıcısı olarak kullanılır. Toplam puan yaklaşımının üç avantajı vardır: 1) kavramsal olarak anlaşılması ve veri çözümü açısından basittir; 2) herhangi bir yüzeydeki madde sayısında daha çok maddeye sahiptir ve madde sayısı arttıkça güvenilirlik artacaktır; 3) bireysel olarak yüzeylerin sahip olduğundan daha fazla kapsama sahiptir (Carmines ve Zeller, 1979). Bu avantajlarının yanısıra toplam puan yaklaşımının en önemli eksikliği, bireysel olarak her yüzeyin yordanan değişken ile olan ilişkisi hakkında bilgi sağlamamasıdır. Yani, toplam puan yordayıcı değişken olarak düşünülürse, her yüzeyin yordanan değişken ile eş ağırlıkla mı ilişkili olduğu yoksa farklı büyüklüklerde mi ilişkili olduğu bilinemez. Bu belirsizliğin en önemli dezavantajı ise, çok yüzeyli yapılarda bazı yüzeylerin gözlenen etkileri açıklaması durumunda, toplam puan yaklaşımının zayıf kalacak olmasıdır. Carver (1989) ve Hull vd. (1991) yordayıcı olmayan yüzeylerin, araştırmalarda yordayıcı puanlara eklenmesi ile yanlış teorilerin geliştirilmesine yol açtığı konusunda uyarmıştır. Fakat hala toplam puan yaklaşımı sıklıkla kullanılmaktadır.

Bireysel puan yaklaşımına göre ise her yüzeyin yordanan değişken ile ilişkisi bireysel olarak kurulur. Bu yaklaşım, toplam puan yaklaşımının dezavantajlarını elimine etmektedir. Fakat bireysel puan yaklaşımının bu avantajının yanısıra

dezavantajı da vardır. Bireysel puan yaklaşımı da kavramsal olarak belirsizdir çünkü yüzeylerin özgül katkılarını yüzeyler tarafından paylaşılan varyansdan (ortak varyans) ayıramamaktadır. Kavramsal olarak, yüzey ile yordanan değişken arasındaki ilişkinin iki kaynağı oluşmaktadır: yüzeylere ait özgül varyans ve tüm yüzeylerin paylaştığı ortak varyans. Bu durumda bazı yüzeylerin özgül varyansı, yordanan değişken ile ilişkili iken bazı yüzeylerin ortak varyansı yordanan değişken ile ilişkili olacaktır.

Her iki yaklaşım da (bireysel puan ve toplam puan) belirsizlikle sonuçlanmaktadır. Bahsedilen bu yaklaşımlardan hangisinin kullanılması gerektiği hala ihtilafıdır. Bu iki yaklaşımın da dezavantajları araştırmacıları alternatif bir yol bulmaya yöneltmiştir. Chen, Hayes, Laurenceau ve Zhang (2012) çalışmalarında var olan bu belirsizliği çözmeye yönelik alternatif bir yaklaşım olarak iki faktör (bifactor) modeli sunmuşlardır. Bu model psikolojinin birçok alanında/dalında kullanılmakla birlikte, toplam puan ve bireysel puan yaklaşımlarının avantajlarını birlikte içerirken aynı zamanda eksikliklerini de gidermektedir.

Toplam puan yaklaşımında olduğu gibi iki faktör (bifactor) modeli de bir genel örtük değişkeni içermektedir. Ayrıca bireysel puan yaklaşımına göre de bu genel örtük değişkeni açıklayan özgül yüzeyler (facetler) vardır. İki etkenli (bifactor) modelin iki temel avantajı vardır; (1) aynı anda yordanan değişken ile genel faktör ve özgül faktörlerin genel faktörden arınık bireysel katkılarını test etmesi, (2) ortak varyansı kontrol altına alındığından, manidar yordayıcı olmayan yüzeylerin belirlenmesine olanak sağlamasıdır.

Canviez (2016)'e göre iki faktör (bifactor) modelin temel avantajları şunlardır; (1) genel faktörün her madde ve madde grupları üzerindeki etkisi kolayca yorumlanabilmektedir. Bu durum ikinci dereceden (second-order) modeller, ilişkili özellikler (correlated trait) modeli ve tek boyutlu modeller (Chen, West ve Sousa, 2006; Immekus ve Imbrie, 2008) ile sağlanamamaktadır; (2) hem genel hem de spesifik faktörlerin maddeler üzerindeki etkisi eş zamanlı olarak kestirilebilmektedir (Reise, 2012; Reise, Moore ve Haviland, 2010); (3) genel ve spesifik faktörlerinin puanlaması ve yorumlaması için gerekli psikometrik özellikler iki faktör model aracılığıyla elde edilebilmektedir (DeMars, 2013); (4) genel ve spesifik özelliklerin (trait) diğer değişkenleri açıklamadaki özgül etkileri daha hatasız bir şekilde elde edilmektedir; (5) madde ve birey parametrelerinin kestirilmesinde, iki faktör (bifactor) modeli ortak köklü madde (testlet-effect) modelinden daha kesin (accurate)

ve güvenilir kestirimler sağlamaktadır (Wainer & Wang, 2000).

Daha önce de bahsedildiği gibi iki faktör (bifactor) model psikolojik özelliklerin ölçeklenmesinde oldukça yaygındır ve yüzeylerin genel faktöre olan özgül katkılarını çok iyi ayırtmaktadır. Bu nedenle iki faktör (bifactor) modeli, ölçek geliştirme ve değerlendirme çalışmalarına oldukça uygundur. Genel yapıyı ve belirli yüzeyleri ölçmeyi amaçlayan yeni bir çok yüzeyli (multifaceted) ölçek geliştirirken ya da değerlendirirken, faktör yüklerinin genel faktör ve spesifik faktörlerdeki gücü, madde seçmede ve değerlendirmede yol gösterici olacaktır. İdeal olarak maddeler genel faktörde daha yüksek yüke ya da en azından spesifik faktörden daha büyük bir yüke sahip olacaklardır. Eğer maddeler genel yapıda yüzeylerden daha büyük bir yüke sahipse bu maddeler seçilecek, fakat spesifik faktörlerde genel faktöre göre daha büyük yüklere sahipse bu maddeler ölçekten çıkarılacaktır. Bunun nedeni ise bu maddelerin genel yapıya önemli bir derecede katkılarının olmamasıdır. Ayrıca iki faktör (bifactor) model, çok boyutlu ölçekten tek boyutlu ölçek ya da kısa tek boyutlu ölçek oluşturmak amacıyla da kullanılmaktadır (Stucky, Thissen, & Edelen, 2013; Stucky, Edelen, Vaughan, Tucker, & Butler, 2014; Stucky & Edelen, 2015).

Bu avantajlarının yanında iki faktör (bifactor) modelinde bazı sınırlılık/dezavantaja sahiptir. En büyük sınırlılığı ise iki faktör (bifactor) madde tepki kuramı modelinin diklik varsayımının karşılanma güçlüğüdür (Chen, West ve Sousa, 2006; Reise, Horan ve Blanchard, 2011; Simms, Grös, Watson ve O'Hara., 2007). Yapısal eşitlik modelinde olduğu gibi, iki faktör (bifactor) modeli de toplam puan ve bireysel puan yaklaşımına göre, oldukça büyük örnekleme ihtiyaç duyar. Ayrıca iki faktör (bifactor) model yorumları spesifik faktörler arası ilişkiye izin verildiğinde oldukça karmaşılaşır (Rindskopf & Rose, 1988) ve model sıklıkla tanımlanamaz. Bunlara ek olarak diğer faktöriyel modellerde olduğu gibi zayıf/küçük faktör yüklenmelerinde zayıf model uyumu vermektedir (MacCallum, Widaman, Zhang ve Hong, 1999; Jennrich ve Bentler, 2012).

1.1.3.1.2. İki faktör (bifactor) kuramı ile ilgili araştırmalar

İki faktör (bifactor) modeli ile ilgili literatür incelendiğinde, şimdiye kadar eğitim ve psikoloji alanındaki ölçmelerde, boyutluluğun tespiti ve madde

performansının incelenmesi gibi konularda yoğunlaştığı görülmüştür. Brouwer, Meijer, Weekers ve Baneke (2008), Umut Eğilimi ölçeği (The Dispositional Hope Scale, DHS)'nin boyutluluğunu belirlemede iki faktör (bifactor) modeli kullanmıştır. Sonuç olarak ölçeğin tek boyutlu yapıda olduğuna karar vermiştir. Brown, Finney ve France (2011), araştırmalarında Hong psikolojik tepkisellik ölçeğinin (The Hong Psychological Reactance Scale-HPRS) faktör yapısını belirlemeye çalışmışlardır. Faktör yapısını belirlemek amacıyla tek-faktör modeli, dört faktör modeli, ikinci derece (second-order) modeli ve iki faktör (bifactor) model analizleri yapılmış, sonuç olarak yapıya en iyi uyumu iki faktör (bifactor) model göstermiştir.

Garn (2017), beden eğitimi dersinde durumsal ilgilerin çok boyutlu ölçümlerinin yapısının araştırıldığı çalışmasında açıklayıcı iki faktör yapısal eşitlik modelini (ESEM) kullanmıştır. Araştırma sonucunda, durumsal ilgilerin çok boyutlu yapıların doğasına iki faktör modelinin uygun olduğu bulunmuştur.

İki faktör (bifactor) modelin eğitim alanındaki uygulamaları bu modelin, alt ölçek puanlarının kullanılması gerektiğinde, alt ölçeklerin puanlanması ve güvenilirliğinin değerlendirilmesi açısından kullanışlı olduğu belirtilmektedir. Örneğin DeMars (2013), iki faktör (bifactor) modeli TIMSS 4. Sınıf fen testindeki çoktan seçmeli ve yapılandırılmış yanıtlar (constructed response) verisi için kullanmıştır. DeMars (2013) bu çalışmasında, alt ölçek puanlarının genel faktör kontrol altına alındıktan sonra güvenilirlik yeterli ise kullanılabilirliğini yoksa alt ölçek puanlarının tavsiye edilmeyeceği sonucuna ulaşmıştır. Bunlara ek olarak, iki faktör (bifactor) model Watkins and Beaujean (2014) tarafından, Wechsler okulöncesi ve ilkökul zeka testi- 4. Baskı (Wechsler Preschool and Primary Scale of Intelligence-Fourth Edition)'da kullanılmış ve Wechsler (2012)'in kullandığı üst düzey (higher order) modelleri ile kıyaslanmıştır. Sonuçlar ise, iki faktör modelin, sadece genel zeka boyutunun klinik değerlendirme ve yorumlanmasında, kesin (precise) sonuçlar verdiği ve sadece bu boyut için kullanılması konusunda yeterince güçlü (robust) bulunmuştur.

Martel, Von Eye ve Nigg (2010), dikkat bozukluğu/hiperaktivite rahatsızlığının (Attention-deficit/hyperactivity disorder (ADHD) heterojen bir yapı olarak tanımlanmakta olduğunu fakat bu tanımın istatistiksel olarak test edilmediği için belirsiz kaldığını belirtmiştir. Bundan dolayı bu yapı tek faktör modeli, iki faktör modeli, üç faktör modeli, ikinci derece (second-order) faktör modeli ve iki faktör (bifactor) modeline göre test etmişlerdir. Araştırma kapsamında 548 çocuktan veri

toplanmıştır. Çocukların yaşları 6-18 yaş aralığındadır. Sonuç olarak ise iki faktör (bifactor) modelin diğer modellere göre daha yüksek uyum verdiği görülmüştür.

Reise, Ventura ve diğ. (2011), şizofren hastalarının bilişsel işlevselliğini (cognitive functioning) ölçmeyi amaçlayan iki ölçme aracının psikometrik özelliklerini belirlemeye çalıştıkları araştırmalarında; tek boyutlu, çok boyutlu ve iki faktör (bifactor) modeli kullanmışlardır. Sonuç olarak verinin çok boyutlu olduğu ve boyutlar arası korelasyonların yüksek olduğu görülmüştür. Ayrıca iki faktör (bifactor) model analizlerinde genel faktördeki madde yüklerinin tek boyutlu modelden çok farklı olmadığı sonucuna ulaşılmıştır.

Golay ve Lecerf (2011), araştırmalarında Wechsler zeka ölçeğininin (French Wechsler Intelligence Scale for Adults –WAIS, French), Cattell–Horn–Carroll (CHC) modeline, iki faktör (bifactor) modeline ve geleneksel üst düzey yapıya (the traditional higher order structure) göre analiz etmişlerdir. Sonuç olarak iki faktör (bifactor) model, geleneksel üst düzey yapıdan (the traditional higher order structure) daha iyi uyum göstermiştir. Ayrıca üst düzey CHC modelininin, geleneksel 4-faktör modelden daha iyi uyum verdiği görülmüştür. Sonuç olarak araştırmacılar WAIS iki faktör modeli en uygun yapı olarak bulmuşlardır.

Cucina ve Byle (2017), zeka testlerinin faktör yapısının belirlenmesinde kullanılan üst düzey yapı (the traditional higher order structure) ile doğrulayıcı iki faktör modelin uyum düzeylerini karşılaştırdıkları çalışmalarında, 31 test bataryası, 58 veri seti ve 1,712,509 kişiden oluşan arşiv verisi kullanmışlardır. Sonuç olarak iki faktör (bifactor) modelin, üst düzey modelden daha güçlü olduğu bulunmuştur.

Boduszek, Dhingra, Shevlin ve Egan (2014), çalışmalarında, Rosenberg Özsaygı Ölçeği (the Rosenberg Self-Esteem Scale- RSES) kullanarak iki faktör (bifactor) model yaklaşımını kullanmışlardır. Daha önce RSES ölçeği, özsaygının tek boyutluluk ölçüsü olarak görülmüştür. Deneysel bulgular ise tek-faktör sonuçlarını (ilişkili ölçme hataları) ve çok boyutlu gösterimleri doğrulamaktadır. Bu nedenlerden dolayı iki faktör modeli mevcut bu tutarsızlığa, teorik ve metodolojik olarak tatmin edici çözüm olarak görülmüştür. Üç alternatif faktör modeli (tek boyutlu, iki faktörlü ve iki faktör-bifactor) bu araştırma kapsamında test edilmiştir. İki faktör (bifactor) model içlerinden en iyi uyum veren model olarak ortaya çıkmıştır.

İki faktör (bifactor) modeli ile ilgili yapılan araştırmalar farklı madde tepki kuramı modellerinin karşılaştırılması üzerine de olmuştur. Bu modeller, tek boyutlu modeller, çok faktörlü çok boyutlu modeller, hiyerarşik modeller ya da ikinci derece

(second-order) modellerdir. Model uyumu, log-likelihood, chi-square, ve p değerleri ile değerlendirilerek ölçme aracının boyutluluğu hakkında sonuçlara ulaşılmıştır.

Chen, West ve Sousa (2006), çalışmalarında iki faktör (bifactor) ve ikinci dereceden (second-order) modeli karşılaştırmışlardır. Araştırmacılar, çalışmalarında yaşam kalitesi (Quality of Life) olarak adlandırdıkları veri setini kullanmışlardır. Bu veri seti toplamda 403 gözlemden oluşmaktadır. Sonuç olarak iki faktör (bifactor) modelin manidar düzeyde ikinci dereceden (second-order) modelden daha iyi uyum sağladığı sonucuna ulaşılmıştır.

Chen, Hayes, Carver, Laurenceau ve Zhang (2012), iki faktör (bifactor) model ve alternatif modellerin çok yüzeyli yapılarda uygulanmasına yönelik çalışmalarında iki faktör (bifactor) modeli ikinci derece (second order) modellerle kıyaslamışlardır. Sonuç olarak, iki faktör modelinin göreceli olarak ikinci derece (second order) modele göre daha iyi uyum sağladığı sonucuna ulaşılmıştır.

Thomas (2012), psikiyatrik rahatsızlıkların altında yatan genel yapı ile spesifik faktörlerin, hiyerarşik ilişkilere sahip olduğunu düşündüğü araştırmasında, standart tek boyutlu ölçme modellerinin bu yapıyı keşfetmekte sıkıntı yaşadığını düşünerek çok boyutlu madde tepki kuramı modellerinden iki faktör (bifactor) modeli, tek boyutlu madde tepki kuramı modeli ve basit yapı (simple structure) ile karşılaştırmıştır. Veri olarak daha önceden Kısa/öz Semptom Envanteri (Brief Symptom Inventory-BSI)'den elde edinmiş arşiv verisi kullanılmıştır. Veri seti toplamda 688 bireyden oluşmaktadır. Sonuç olarak ise iki faktör (bifactor) modelin diğer modellere göre daha iyi uyum sağladığı sonucuna ulaşılmıştır.

İki faktör (bifactor) modeli, ortak köklü madde modeli temelli (testlet-based) değerlendirmelerle de karşılaştırılmıştır. Demars (2006), araştırmasında iki faktör (bifactor) modeli, ortak köklü madde modeli (testlet-based), çoklu puanlanan model ve bağımsız maddeler modeli ile simülatif veri kullanarak karşılaştırmıştır. Sonuç olarak ise; (1) maddelerin ortak kök (testlet) modelinde olduğu zaman; bağımsız maddeler modeli, madde güçlük değerlerinde büyük RMSE (Root Mean Square) değerine sahip olmakta ve ayırt edicilik katsayıları olduğundan küçük kestirmektedir, (2) maddeler bağımsız olarak (ortak kök modeli yerine) üretildiğinde ise; iki faktör (bifactor) modeli, güçlük ve ayırt edicilik parametrelerinde biraz daha yüksek RMSE değerlerine sahip olmuştur. Simülatif veri ile elde edilen bu bulgular, gerçek veri ile elde edilen bulgular ile benzerlik gösterdiği sonucuna ulaşılmıştır.

Rijmen (2009) ise araştırmasında, iki etkenli (bifactor) model, ortak kök (testlet) modeli ve ikinci derece modellerini tanıtmış ve ikinci derece modellerinin ortak kök (testlet) modellerle eşit olduğu ve bu ikisinin de iki faktör (bifactor) modelin özel durumları olduğu sonucuna ulaşmıştır. Lafond (2014), iki faktör (bifactor) modeli ve ortak kök (testlet) teori (TRT) modelleri için karar tutarlılığı ve kesinlik indekslerini (accuracy indices) kestirmede yeni yöntemleri geliştirmeye çalıştığı araştırmasında, Monte Carlo yaklaşımı ile yeni bir metod (M method) geliştirilmiştir. Bu metodun sıklıkla kullanılan diğer metotlar kadar kesin kestirimler yaptığı sonucuna ulaşmıştır.

İki etkenli (bifactor) modeller ayrıca Bayesian tekniği ile farklılaşan madde fonksiyonunu (DIF) belirlemede de kullanılmıştır. Fukuhara (2009), Bayesian Markov Chain Monte Carlo kestirim metodunun uygulandığı iki etkenli (bifactor) modelin geliştirilmesi ile yeni bir DIF belirleme modeli tespit etmiştir. Bu simülasyon çalışmasının sonuçlarına göre iki etkenli (bifactor) DIF modelleri, DIF ve DIF içermeyen maddeleri doğru olarak tespit ettiği, ayrıca diğer madde tepki kuramlarına göre, DIF olmayan maddelerin belirlenmesinde daha kesin sonuçlar verdiği sonucuna ulaşılmıştır. Yani iki etkenli (bifactor) DIF modelleri daha küçük hata payları ile DIF kestirimi yapabilmektedir. Wang ve Kim (2017), yaptıkları simülasyon çalışmasında, çok düzeyli iki faktör modelin yanlış oluşturulmasında (misspecification) bayes yaklaşımına dayanan kestirimlerin etkililiğini araştırmışlardır. Modelin yanlış oluşturulmasının yapısal katsayılar üzerindeki etkisi, yanlılık (bias) ve güç (power) kapsamında incelenmiştir. Sonuçlar bayes yönteminin, modelin yanlış oluşturulma düzeyinin artmasına rağmen, yapısal katsayıların kestiriminde iyi performans gösterdiği görülmüştür.

Bazı araştırmacılar ise iki etkenli (bifactor) modelin algoritması ile ilgilenmişlerdir. Yang, Song, and Xu (2002), simülatif veri ile çalışarak ilişkili gözlemlerin güçlü kestiricisini, eş ağırlıklı iki faktör (bifactor) model ile betimlemiştir. Gibbons ve diğ. (2007), ikili madde tepki kuramı (binary item response theory) modelinde kullanılan tam bilgi madde faktör analizini (full information item factor analysis), dereceli madde tepki kuramı modeline uyarlamıştır. Sonuç olarak ise çok boyutlu madde tepki kuramı modellerinin; iki faktör (bifactor) modeli ve ortak kök (testlet) madde tepki modeli, iki kademeli tam bilgi (full information) çatısı altında olduğu görülmüştür. Rodriguez, Reise ve Haviland (2016), çalışmalarında psikolojik değişkenlerin değerlendirilmesinde

kullanılan iki faktör model indekslerinde düzenlemeler yapmıştır.

İki faktör modelin avantajları üzerine de arařtırmalar yapılmıřtır. Reise, Moore, ve Haviland (2010)'ın iki faktör (bifactor) madde tepki kuramının avantajlarını incelediđi arařtırmalarında ulařtıđı sonuçlarından ilki, iki faktör (bifactor) model için genel faktörün parametre kestiriminde, madde kümelerini dikkate alarak verinin çok boyutlu yapısının bozulmasını kontrol altına aldıđı ve tek boyutlu madde tepki kuramının ise bu bozulmayı dikkate almadıđı ve bundan dolayı verinin çok boyutlu yapısının bozulduđudur. İkincisi ise, iki faktör (bifactor) modelin alt alanlara yönelik ayırt edicilik indeksi ve yetenek puanları gibi betimleyici istatistikler verdiđi ve bu bilginin madde ve birey performanslarının geliřtirilmesinde tanısıl amaçla kullanılabilirdiđi ve iki etkenli modellerin parametre kestiriminde yařanacak bozulmaları çok boyutlu veriyi tek boyutlu madde tepki kuramı ile çalışırken de destekleyebildiđidir. Reise (2012) daha sonra iki faktör (bifactor) modellerin yeniden keřfedilmesi adlı çalışmasında, iki faktör modelinin sınırlılıklarını ele almıř ve her psikolojik yapıda kullanılamayacađını belirtmiřtir. Wiesner ve Schanding (2013), iki faktör (bifactor) modelin avantajlarını incelediđi arařtırmasında, iki faktör (bifactor) modeli tek boyutlu (unidimensional) model, üst düzey (higher-order) model ve iliřkisel model ile karřılařtırmıřtır. Çalışmasında, iki faktör (bifactor) model, tek boyutlu modellerde, model uyumsuzluklarının üstesinden gelmek adına global risk puanı (global risk score) üretilmiřtir. Ayrıca bir spesifik faktördeki puanların bile uygulayıcılara yapıyı yorumlamada büyük avantaj getireceđi sonucuna ulařmıřlardır.

Alan yazındaki arařtırmalar özetlenecek olursa, genellikle iki faktör (bifactor) modelin tek boyutlu, çok boyutlu, ortak kök (testlet) ve ikinci dereceden (second order) vb. modeller ile model uyumu ve parametre kestirim kesinliđi açısında karřılařtırılmasına ve geliřtirilen ölçeklerin boyutluluđun belirlenmesine odaklanılmıřtır. İki faktör (bifactor) modelinin kullanılmasını kısıtlayan durumun diklik varsayımı olmasına karřın řimdiye kadar diklik varsayımının deđiřik kořullarda sınanması ile ilgili olarak alan yazında yürütölen sadece bir çalışmaya (Zheng, 2013) rastlanmıřtır. Bu çalışma (Zheng, 2013) da her simölasyon çalışması gibi sınırlı kořullarda gerçekeřtirilmiřtir. Ayrıca bu çalışmanın dıřında Rindskopf ve Rose (1988) çalışmalarında, iki faktör (bifactor) modelde spesifik faktörler arası iliřkiye izin verildiđinde model parametrelerinin yorumlanmasının karmařıklařtıđı sonucuna ulařmıřtır. Faktörler arası yüklenmeler aynı zamanda faktörler arası

korelasyonlara da imkan tanıyacağından bu durum bir nevi faktörler arası korelasyon gibi düşünülebilir. Rindskopf & Rose, (1988) çalışmalarında bu durumun hangi düzeylerdeki yüklenmelerde (ilişkilerde) meydana geldiği hakkında bir bilgiye ulaşamamıştır.

İki faktör kuramı, gerektirdiği varsayımdan (diklik) dolayı kullanımı sınırlanan bir kuramdır. Bu sınırlılığının yanısıra psikolojik ve eğitimsel yapıların modellenmesinde ve ölçek geliştirme çalışmalarında varsayım göz ardı edilerek sıklıkla kullanılmaktadır. Diklik varsayımının sağlanmadığı koşullarda psikolojik ve eğitimsel yapıların doğru modellenmesi, geliştirilen ölçeğin doğru faktör yapısına ulaşılması ve parametre kestirimlerinin doğru olması mümkün olmayacaktır. Parametre kestirimlerinin kesinliği ve doğruluğu ise yapılan her ölçme işleminde önemli bir durumdur. Çünkü parametre kestirimleri, madde performansı ve yanıtlayıcı yetenek düzeyinin belirlenmesinde önemli bir unsurdur. Belirtilen bu gerekçelerden kaynaklı, iki faktör kuramının, spesifik faktörler arası farklı ilişki düzeylerine olanak tanıyarak incelenmesi yani hangi diklik ihlal düzeylerinin kuram tarafından tolere edilip, diklik ihlaline rağmen kararlı, kesin ve doğru kestirimler yapılabildiğinin belirlenmesi bu araştırmanın problemini oluşturmaktadır.

1.2. Amaç

Bu çalışmanın amacı, farklı düzeylerdeki diklik varsayımı ihlalinin parametre kestirimlerine etkisini araştırmaktır. Veriler simülasyon yolu ile elde edilen ve ikili puanlanan madde yapıları şeklindedir. Bu amaç doğrultusunda aşağıdaki araştırma sorularına cevap aranmaktadır;

1. İki spesifik faktör arasındaki farklı diklik ihlal düzeylerinde (Model-1), madde ve birey parametrelerinin kestirimi ideal duruma (diklik ihlalinin olmadığı) göre nasıldır?

2. Tüm spesifik faktörler arasındaki farklı diklik ihlal düzeylerinde (Model-2), madde ve birey parametrelerinin kestirimi, ideal duruma (diklik ihlalinin olmadığı) göre nasıldır?

3. Model türüne ve test uzunluğuna göre, madde ve birey parametrelerinin yanlılık puanları nasıldır?

1.3. Önem

Eğitim ve psikolojide yapılar genellikle çok boyutludur ve alt alanlara sahiptir. Ayrıca insan faktörünün tam da içinde olduğu, insanın araştırıldığı çalışmalarda, incelenen yapıların tek boyutlu olduğunu düşünmek ve bu verileri tek boyutluymuş gibi değerlendirmek gerçekçi bir yaklaşım değildir. Çok boyutlu olan örtük yapıların incelenmesinde kullanılacak Madde Tepki Kuramı modellerinden biri ise iki faktör (bifactor) modelidir. Çünkü iki faktörlü (bifactor) Madde Tepki Kuramı modelleri varolan çok boyutlu madde tepki kuramı modelleri arasında oldukça yüksek açıklayıcılığa ve uyuma sahip modellerdir. İki faktör (bifactor) modelinin uygulanmasıyla ilgili olarak önemli bir sınırlılık diklik varsayımının karşılanmasıdır. Dik olmayan verilere diklik varsayımı gerektiren bir modeli uygulamak madde ve birey parametrelerinin kestiriminde hatalı kestirime yani yanlışlığa neden olabilmektedir. Diklik varsayımının ihlali durumunda parametre kestirimlerinde ortaya çıkacak hatalar ile karşılaşmamak için, iki faktör modelinin diklik ihlalini ne düzeye kadar tolere edebileceğinin araştırılması çalışmanın odak noktasıdır ve çalışmanın esas önemini oluşturmaktadır. Ulaşılabilecek sonuç ile iki faktör (bifactor) modelinin kullanım sınırları belirli koşullar altında belirlenmiş olacak ve kullanımı daha da yaygınlaşabilecektir. İki faktör model daha önce de bahsedildiği gibi toplam puan ve bireysel puan yaklaşımına göre en avantajlı modeldir. Çünkü iki faktör modelde, hem toplam (genel faktör) hem de bireysel (spesifik faktörler) puan hesaplanabilir. Bu durum eğitim alanında başarının değerlendirilmesine yönelik yapılan ölçmelerde, öğrencilerin sahip olduğu bilgilerin hem genel hem de spesifik faktör puanlarının hesaplanması açısından önemlidir. Örneğin; bir öğrencinin matematik testinin alt boyutlarındaki düşük puanları, öğrencinin o alana henüz yeterince hakim olmadığını göstermektedir. Bunun yanı sıra iki faktör modeli diklik varsayımı gereği faktörler arası ilişkiye izin vermemektedir. Fakat eğitim ve psikoloji alanında ilişkisiz faktörlere sahip olabilecek ölçme araçları geliştirmek neredeyse imkansızdır. Örneğin, sayılar, cebir ve geometri gibi üç alt alandan oluşan iyi bir şekilde tasarlanmış matematik testi hazırlandığında, diklik varsayımı gereği bu alt alanların paketlenmiş üniteler gibi birbirinden bağımsız ve ilişkisiz olması gerekmektedir. Cebir ve geometri sorularının sayılar bilgisi olmadan doğru şekilde cevaplanabileceğini hayal etmek bile oldukça güçtür. Bu maddeler arasında illa ki çapraz yüklenmeler olacaktır. Bu örtük faktörler

(cebir ve geometri) teorik olarak mümkün olabilir ama gerçekte bazı çakışmalar olacaktır. Bu mevcut ilişkili faktörleri dik olmaya zorlamak ise ölçülen yapı ile ilgili olarak bilgi kaybına neden olacak ve güvenilir olmayan parametre kestirimleri ile sonuçlanacaktır. Bu çalışma aracılığıyla iki faktör (bifactor) modelin kullanımını kısıtlayan varsayımın incelenmesi ve belli kriterler ışığında değerlendirilmesi ile elde edilecek sonuçların alan yazına hem teorik anlamda hem de modelin daha doğru uygulanabilirliği açısından çok önemli katkılar sağlayacağı düşünülmektedir.

1.4. Sınırlılıklar

Bu araştırmadan elde edilen sonuçlar, araştırmanın amacı doğrultusunda ele alınan değişkenler ve oluşturulan simülasyon koşulları ile sınırlıdır.

BÖLÜM 2

YÖNTEM

Bu bölümde, araştırmanın modeli, verilerin üretilmesi, verilerin analizi ile ilgili açıklamalara yer verilmiştir.

2.1. Araştırma Modeli

Bu araştırmada, farklı düzeylerde ve test uzunluklarındaki diklik ihlalinin parametre kestirimlerindeki yanlılığı incelenmiştir. Araştırmanın amacına uygun olarak Monte Carlo simülasyon tekniği kullanılmıştır.

Monte Carlo Yöntemi, rastgele sayılar kullanılarak sonuca ulaşmayı amaçlayan, deneysel bir yöntemdir. Bu şekilde matematik ve fizik alanındaki problemlerin çözümü amaçlanmaktadır (Krykova, 2003). Ayrıca Monte Carlo yöntemi, gerçek test koşullarının modellenemesine ve ampirik veriler yoluyla elde edilemeyen istatistiklerin hesaplanmasına olanak sağlamaktadır (Boomsma, 2013).

Simülasyonlar, bilgisayar yazılımları aracılığıyla yapılabileceği gibi bilgisayar desteği olmadan da yapılabilir. Simülasyon sürecinde çok sayıda işlem yapma gereksinimi varsa ve çok uzun bir zaman aralığı için sistem simüle edilecekse, bilgisayar kullanmak daha uygun olacaktır (Kavcar, 2004). Bilgisayarlar deterministik bir yaklaşımla çalışmaktadırlar. Deterministik yaklaşım, sistemin gelecek durumlarının belirlenmesinde hiçbir rastgeleliğin olmaması durumudur. Yani iyi modellenmiş bir deterministik sistem için, aynı şartlar altında ve aynı başlangıç durumları için sistem her zaman aynı sonucu vermektedir.

Bilgisayarlar deterministik (rastgele olmayan) bir yapıda çalıştıkları için gerçek anlamda rastgele sayı üretmezler. Ancak, uygun algoritmalarla, bir bilgisayarın düzgün bir dağılımdan seçilen ve genellikle $[0,1]$ arasında gerçek değerler alan rastgele sayı üretmesi sağlanabilmektedir. Bilgisayarların ürettiği bu rastgele sayılar *yalancı rastgele sayı* (pseudo-random numbers) olarak adlandırılmaktadır. Rastgele sayı üreten bu algoritmalara ise *rastgele sayı üretici*

(random number generator) denir (Niederreiter, 1992). Rastgele sayılar, özel bir dağılımdan seçilen düzensiz (caotic) bir sayıdır (L'Ecuyer, 2014). Bazı durumlarda, bilgisayar yazılımlarının ürettiği rastgele sayılar kullanılarak belirsiz, yani tahmini sistemler modellenenmektedir.

Bu araştırma, farklı düzeylerde ve test uzunluklarındaki diklik varsayımı ihlalinin madde ve birey parametrelerinin kestirimlerine etkisini araştırmak amacı ile Monte Carlo yöntemi ile elde edilecek veriler aracılığıyla yürütüldüğü için simülatif araştırma modelindedir.

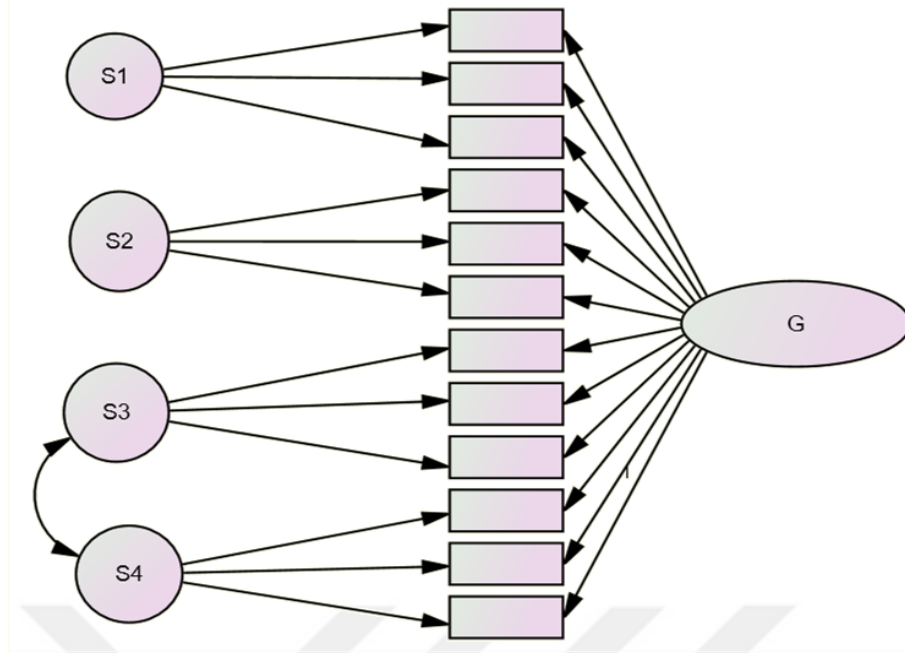
2.2. Veriler

Bu araştırmanın amacı doğrultusunda veriler iki faktör (bifactor) modele göre üretilmiştir. Üretilen veriler bağımsız ve bağımlı değişkenler olarak iki gruba ayrılmaktadır. Bağımsız değişkenler; iki spesifik faktör arasında ilişki olan Model 1 (Şekil 1), tüm spesifik faktörler arasında ilişki olan Model 2 (Şekil 2), diklik ihlali düzeyleri ve test uzunluklarıdır. Bağımlı değişkenler ise madde ve birey parametrelerinin kestirim değerleri.

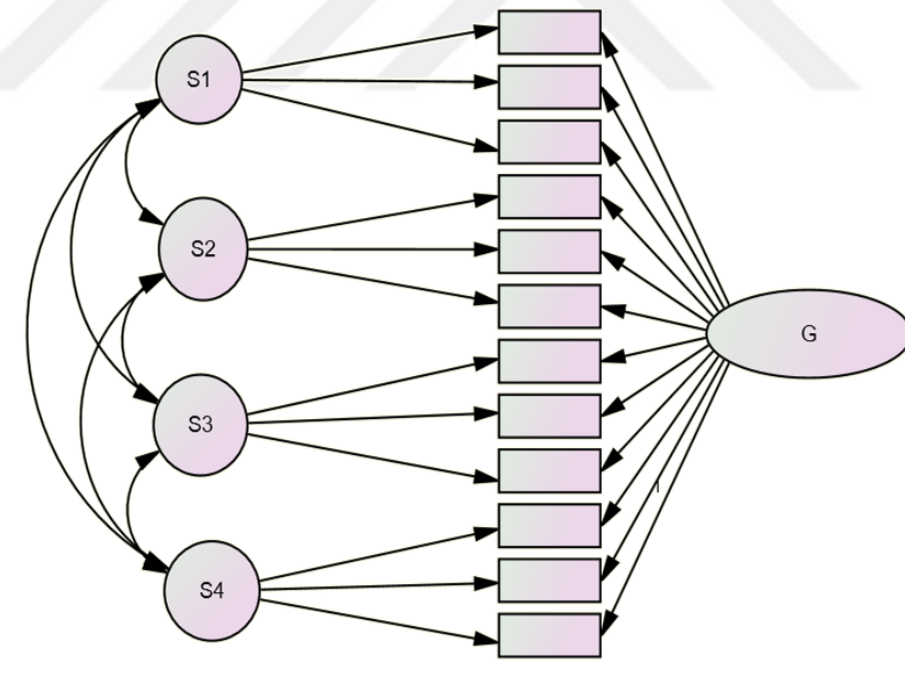
2.2.1. Verilerin Üretilmesi

Bu araştırma için veriler iki adet iki faktör (bifactor) iki parametrelili modele göre üretilmiştir. İki parametrelili modelin ve ikili puanlanan verinin seçilmesinin nedeni; yazılımlar ile uygulanabilirliğinin ve anlaşılabilirliğinin kolay olmasıdır. Simülatif yolla veriler iki modele (Model-1 ve Model-2) göre üretilmiştir. Şekil-2'de verilen iki spesifik faktör arasında çapraz yüklenmelerden dolayı oluşan diklik ihlalini gösteren Model-1'dir. Burada incelenen nokta iki spesifik faktör arasındaki diklik ihlalinin tüm faktörlerdeki parametre kestirimlerine olan etkisidir. Şekil-3'de verilen Model-2 ise, tüm spesifik faktörler arasındaki ilişkiyi göstermektedir.

Kurulan iki faktör modeller bir genel faktör ve dört spesifik faktörden oluşmaktadır. Spesifik faktör sayısının dört tutulmasının nedeni; iki spesifik faktörün ilişkili olma durumunda, bu ilişkinin diğer faktörlerin (genel ve spesifik) parametre kestirimlerinde ne tür bir etki yaratacağının incelenmek istenmesidir.



Şekil 2. İki spesifik faktörün ilişkili olma durumu



Şekil 3. Tüm spesifik faktörlerin ilişkili olma durumu

Spesifik modellerde manipüle edilen değişkenler faktörler arası korelasyon düzeyleri ve test uzunluklarıdır. Kurulan modeller için (Model 1 ve Model 2)

korelasyon kabul düzeyleri 0.10 (çok düşük), 0.40 (orta), 0.70 (yüksek) olarak ele alınmıştır (Cohen, 1988). Spesifik faktörlerdeki madde sayılarına karar vermek amacıyla yapılan alan yazın incelemesi sonucunda, çalışmalarda 25 ve 50 (Demars, 2006), 12 ve 24 (Köse, 2010), 10, 25 ve 50 (Li, 2011), 10, 20 ve 30 (Şahin, 2012), 40, 60 ve 80 (Zheng, 2013) maddelik test uzunluklarının kullanıldığı görülmüştür. Bu araştırma kapsamında ise bir faktörün faktör olabilmesi için sahip olması gereken madde sayısının en az üç olması gerektiği referans alınarak (Kline, 1994) en küçük test uzunluğu 12 madde olarak karar verilmiştir. Alan yazından farklı olarak bu çalışmada her faktördeki madde sayısının eşit olması göz önünde bulundurularak diğer test uzunlukları 40 ve 100 olarak belirlenmiştir.

Araştırma boyunca sabit tutulacak (manipüle edilmeyecek) değişken ise örneklem büyüklüğüdür. Replikasyon sayısı ise bir değişkenlik kaynağı olarak görülmemekle birlikte, tüm modellerde sabit tutulmuştur.

Örneklem büyüklüğüne karar vermek için yapılan alan yazın incelemesi sonucunda; ağırlıklandırılmamış en küçük kareler (Unweighted Least Squares-ULS) ve robust ağırlıklandırılmış en küçük kareler (Robust Weighted Least Squares-RWLS) kestirim yöntemleri kullanıldığında, kestirim keskinliğinin (precision) örneklem büyüklüğü ile doğru orantılı bir şekilde artış gösterdiği görülmüştür. Finch (2010) örneklem büyüklüğü ile ilgili çalışmasında, örneklem büyüklüğünün 250'den fazla olduğu durumlarda ise kestirim sonuçlarına örneklem büyüklüğünün manidar bir etkisi olmadığı sonucuna ulaşmıştır. Li ve Lissitz (2012), iki etkenli (bifactor) modelde dikey ölçeklemeyi inceledikleri araştırmalarında 1000, 2000 ve 4000'lik örneklem büyüklüklerini kullanmışlardır. Araştırmaları sonucunda, kestirim doğruluğu (accuracy) ve kestirim tutarlılığının (stability) örneklem büyüklüğü ile doğru orantılı bir şekilde artış gösterdiği sonucuna ulaşmışlardır. Özellikle araştırmalarında, 2000'lik ve 4000'lik örneklem büyüklüklerinde, 1000'lik örneklem büyüklüğüne kıyasla çok daha düşük RMSE ve SE'ler elde etmişlerdir.

Khojasteh (2012) ise uyum iyiliği indekslerinin metrik değişmezliğini iki faktör (bifactor) model ile incelediği araştırmasında, 400, 800 ve 1200'lük örneklem kullanmıştır. Araştırma sonucunda küçük örneklemelerde (400) testin gücünün düştüğü sonucuna ulaşmıştır.

Desa (2012) telafisel ve telafisel olmayan iki faktör (bifactor) modellerde, alt puan kestirimlerini incelediği çalışmasında, 1500'lük örneklem büyüklüğü kullanmıştır. Araştırmada kullanılan örneklem büyüklüğünün araştırma kapsamında

incelenen modellerde parametre kestirimleri için yeterli olduğu belirtilmiştir.

Liu (2011) iki parametrelili iki faktör (bifactor) modelde test istatistiklerine dayalı olarak yerel bağımlılığı belirlemeye çalıştığı araştırmasında örneklem büyüklüğü için 200, 500 ve 1000'lik örneklem kullanmıştır. LaFond (2014) ise iki etkenli (bifactor) ve ortak kök (testlet) modelleri için karar tutarlığı (decision consistency) ve doğruluğu (accuracy) için yeni indeksler belirlemede yeni yöntemleri bulmaya çalıştığı araştırmasında, örneklem büyüklüğü 5000 olarak belirlenmiştir.

Yapılan alan yazın taraması sonucunda MTK modellerinin büyük örneklemde daha iyi kestirime olanak verdiği görülmüştür. Yapılan iki etkenli (bifactor) model kestirimlerinde kestirim kesinliği (precision) ve doğruluğu (accuracy)'nun örneklem büyüklüğünden etkilenmemesi bu araştırmanın önemli noktalarındandır. Çünkü örneklem büyüklüğü, bu çalışmada manipüle edilen değişkenlerden biri değildir. Bu çalışmada, örneklem büyüklüğünden kaynaklanacak yanlışlığı engellemek amacıyla mevcut çalışmalarda kullanılan en büyük örneklem büyüklüğü (5000) simülasyon örnekleme olarak belirlenmiştir. Araştırma deseninin özeti Çizelge 1'de verilmiştir.

Çizelge 1

Araştırma Deseni Özeti

	Korelasyon düzeyleri	Madde sayısı	Örneklem Büyüklüğü
Model 1			
Model 1.1	$r_{3,4}=0.10$	12-40-100	5000
Model 1.2	$r_{3,4}=0.40$	12-40-100	5000
Model 1.3	$r_{3,4}=0.70$	12-40-100	5000
Model 2			
Model 2.1	$r_{2,4} \leq 0.10$	12-40-100	5000
Model 2.2	$0.10 < r_{2,4} \leq 0.40$	12-40-100	5000
Model 2.3	$0.40 < r_{2,4} \leq 0.70$	12-40-100	5000

Replikasyon sayısı daha önce de belirtildiği üzere simülasyon süreci boyunca sabit tutulmuştur. Replikasyon sayısının sabit tutulmasının nedeni, bu araştırmanın deneysel temelli olması ve manipüle edilen değişkenlik kaynakları dışındaki faktörlerin etkisinin sıfırlanmasıdır. Replikasyon sayısına karar vermek için yapılan alan yazın incelemesi sonucunda, replikasyon sayısına karar vermede araştırmacıların tam olarak kullandığı ölçütlerin olmadığı ve replikasyon sayısına

karar verirken belirli bir gerekçe sunmadıkları görülmüştür. Alanyazın incelemesi sonucunda iki faktör (bifactor) modellerde kullanılan replikasyon sayılarının genelde 100 (Demars, 2006; Zang, 2008), 200 (Zheng, 2013) 500 (Cai, Yang ve Hansen, 2011) olduğu görülmüştür. Bu çalışmada ise replikasyon sayısını belirlemek amacıyla araştırma öncesinde 50, 100, 150 ve 200 replikasyon yapılarak ideal model (diklik varsayımının olmadığı model) üzerinden ortalama yanlılık kestirilmiştir. Ortalama yanlılığın 200 replikasyonda diğerlerine oranla sifıra yakın olduğu görülmüş ve pratiklik açısından replikasyon sayısı 200 olarak belirlenmiştir.

Belirlenen replikasyon sayısı ile iki parametrelilik iki faktörlük (bifactor) veri setini üretmek için, ayırt edicilik parametresi (a), güçlük parametresi (b) ve birey parametresi (θ) dağılımının belirlenmesi gerekmektedir. Ayırt edicilik parametresi (a), tek biçimli (uniform) 0.2 ile 2.0 ranj aralığında bir dağılım, güçlük parametresi (b) ve birey parametresi (θ) random normal dağılım gösterecek şekilde simülasyon modeli oluşturulmuştur. İki etkenli (bifactor) iki parametrelilik modelin matematiksel olarak ifade edilişii aşağıdaki gibidir:

$$p(y=1|\theta_g, \theta_s) = \frac{1}{1 + \exp\{-(d+a_g\theta_g+a_s\theta_s)\}}$$

Ayırt edicilik, güçlük ve birey parametrelerinin dağılım özellikleri Çizelge 1’de de verilen 18 (2x3x3) koşul için de aynıdır. İlk koşul için üretilen gerçek parametrelerde random sabit (seed) atanmış ve diğer koşullarda bu sabit (seed) sayı ile gerçek parametrelerin modeller arası değişmezliğı sağlanmıştır. Madde ve birey parametreleri ile ilgili bilgiler Çizelge 2’de verilmiştir. Üretilen güçlük katsayısı aşağıda verilen formül ile çok boyutlu güçlük katsayısına çevrilmiştir:

$$d = -b \sqrt{a_g^2 + a_s^2}$$

Çizelge 2

Simülasyon Deseniindeki Madde ve Birey Parametrelerinin Dağılımları

Dağılımlar	
Madde ve Birey Parametreleri	a Tek biçimli (uniform) dağılım (0.2, 2.0)
	b Random normal dağılım (0, 1)
	θ Random normal dağılım (0, 1)

Verilerin simülasyonunda Monte Carlo benzetim yöntemi kullanılmıştır. Monte Carlo yöntemi, Nicholas Constantine Metropolis tarafından bulunmuştur ve atom bombasının geliştirildiği Los Alamos Ulusal Labratuvarında, bombanın patlamasından sonra dağılan nötronlara karşı kalkan modellemek için Stanislaw Ulam tarafından günümüze taşınmıştır. Metropolis algoritması olarak da bilinir. Algoritma, kesin çözüm yapmanın zor olduğu problemlerde tahmini çözümlere gitmeyi amaçlar yani olasılık teorisi üzerine kurulmuştur (Krykova, 2003).

Monte Carlo yöntemi bilgisayar yazılımlarını kullanarak istatistiksel örnekleme deneyleri ile çeşitli matematik problemlerine yaklaşık çözümler sağlar. Bu yöntem yıllardır kullanılmaktadır fakat son yıllarda daha da popüler olmuştur. Monte Carlo yöntemi bilgisayar yazılımları tarafından bir rutin ile üretilen random sayı ve vektörleri kullanmaktadır. Bilgisayar algoritmalarının random sayıları üretmesi rastgele (deterministic) bir süreçtir ve sayı dizilimleri seçkisizdirler yani hiç bir örüntü içermemektedirler. Seçkisizliği (randomness) tanımlamak zor olsa da çoğu zaman tahmin edilemezlik (unpredictability) ile ilişkilendirilir. Gerçek seçkisizliğin diğer bir özelliği ise tekrar edilemez olmasıdır. Bilgisayar sistemleri aracılığıyla üretilen bu random sayı dizilimlerine *yalancı rastgele sayı* (pseudo-random numbers) denmekle birlikte en önemli özellikleri yeniden üretilebilir ve tahmin edilebilir olmalarıdır. Monte Carlo yöntemleri, bu rastgele sayıları (pseudorandom numbers) kullanarak tahmini sistemleri modelleyen algoritmalarıdır (örn; Hücre Simülasyonu, Borsa Modelleri, Dağılım Fonksiyonları, Sayısal Analiz, Doğal olayların simülasyonu, Atom ve Molekül Fiziği, Nükleer Fizik ve Yüksek Enerji Fiziği modellerini test eden simülasyonlar, deneylerde kullanılan aletlerin simülasyonu) (Krykova, 2003).

Monte Carlo yönteminde, üretilecek değişkenlerin dağılımı belliyse (örneğin: Normal dağılım) ve aynı zamanda dağılım özellikleri de biliniyorsa (μ_x : Ortalama ve σ_x : Standart sapma) X değerleri için defalarca rastgele sayılar üretilerek f fonksiyonuyla Y değerleri hesaplanabilir. Buradan da Y ile ilgili istatistiksel bilgilere ulaşılabilir. Ne kadar çok örnekleme yapılırsa o kadar iyi sonuçlar elde edilir (Krykova, 2003; Niederreiter, 1992). Simülasyon çalışmaları istatistiksel problemlere çözüm olarak 1900'lü yıllardan beri kullanılmaktadır (Harwell, Stone, Hsu ve Kristici; 1996). Bu çalışma, Monte Carlo yöntemine dayalı, belirlenen koşullara uygun verilerin simülasyonu ve iki faktör (bifactor) model parametrelerinin üretilmesi amacıyla yazılan (Ek.A) syntax (Zheng, 2013) ile R 3.4.0 GUI yazılımı

kullanılarak yapılmıştır. Düzenlenen syntax'ın ve üretilen veri dosyalarının doğruluğu diklik ihlali içermeyen bir model üzerine yanlılık ortalamaları hesaplanmış ve yanlılık ortalamalarının sıfıra yakın olduğu görülmüştür.

2.3. Verilerin Analizi

Simülasyon sonucunda üretilen verilerden iki faktör (bifactor) model kestirimleri yapılmıştır. Her koşul için 200, toplamda ise 36000 (18x200) model analiz edilmiştir. İki faktör (bifactor) model kestirimleri R 3.4.0 GUI yazılımında bulunan “mirt” (Chalmers, 2016) paketi ile, betimsel istatistikler “psych” (Revelle, 2017) paketi ile, grafiksel gösterimler “ggplot2” (Wickham, 2008) ile yapılmıştır.

R yazılımı aracılığı ile yapılan iki faktör (bifactor) model analizlerinde Quasi Monte Carlo (QMC) yöntemi kullanılmıştır. Matematik ve fizikte sıklıkla analitik olarak çözülemeyen problemlerle karşılaşmaktadır. Özellikle çok boyutlu integral alma durumu söz konusu olunca hesaplamalar gittikçe karmaşıklaşmakta ve maliyet yükselmektedir. Monte Carlo yönteminde yüksek boyutlu (higher dimensional) integrallerin çözümü daha kolaylaşmakta ve hata oranı işlem yapılmadan önce yaklaşık olarak rahatlıkla kestirilebilmektedir. QMC yönteminde, Monte Carlo ile aynı hesaplama yöntemi kullanılır fakat Monte Carlo'nun aksine yarı/sözde rasgele sayılar (quasi-random numbers) dizilişi kullanılmaktadır. Yarı/sözde rasgele sayılar (quasi-random numbers) bilgisayar tarafından üretilir ve *yalancı rastgele sayılar* (pseudo-random numbers) ile benzerdir fakat ek olarak hataları azaltmak için eşit dizilmiş sıraları seçer (Ueberhuber 1997; akt. Kyrkova, 2003). QMC metodu ilk olarak 1950 yılında önerilmiştir. High dimensional yapıların kestiriminde MC yerine daha doğru (accurate) kestirimler sağlamaktadır (Niederreiter, 1992; L'Ecuyer, 2014). QMC metodu high-dimensional (integration) bütünleştirme sorunlarında sıklıkla kullanılmaktadır.

Bu araştırma kapsamında, parametrelerin kestiriminde EAP (Expected a Posteriori) yöntemi kullanılmıştır. MTK'ya göre parametre kestirimlerinde çeşitli teknikler vardır. MTK parametrelerinin kestirimlerinin arkasındaki temel, MTK model karakteristiklerinin bireyin maddeleri doğru yanıtladığı zaman elde edilebilmesidir. Puan örüntülerinde yanlışlar ve doğrular olduğu durumlarda, bireyin tüm madde yanıtlarındaki örüntüsü üzerindeki puan örüntüsünün birleşim olasılığı, bireyin düşük ve yüksek olasılıkla yanıtlayacağı olasılıkların birleşik olasılığını betimler. Bu yöntemde Maksimum olabilirlik (Maksimum Likelihood

Estimation/MLE) ya da birleşik maksimum olabilirlik (Joint Maksimum Likelihood Estimation /JMLE) kestirimi denir. Bununla birlikte eğer çok sayıda birey ve/veya testte çok sayıda madde varsa bu yöntemler çok fazla hesaplama içermektedir. Öte yandan birincil ve/veya spesifik kapsam alanlarındaki maddelere verilen yanıtların tümü doğru (örn; Kusursuz puan/perfect score) ya da tümü yanlış olduğu durumlarda (sıfır puan/zero score) JMLE kullanılamaz. Tamamı doğru yanıt örüntüleri için olasılıkların çarpımı pozitif sonsuz yetenek değerleri, ve tamamının yanlış yanıtladığı durumlar için ise sonsuz negatif yetenek değerleri vermektedir (Baker, 1985; Baker & Kim, 2004; Hambleton & Swaminathan, 1985; Thissen & Orlando, 2001).

MLE ve JMLE metodlarından farklı olarak EAP (Expected a Posteriori) kestirim yöntemi, veri setindeki yanıt örüntüsündeki tüm durumların (case) doğru (perfect score) ya da tüm durumların yanlış (zero score) olduğu koşullarda da kullanılabilir. MLE ve JMLE bu durumlarda sonsuza giden kestirimler yaparken EAP alt puan (subscore) kestirimleri sonsuz değerler almayacaktır. EAP kestirim yöntemi alt puanları doğrusal olarak evren ortalamasına çekmekte ve onları kendi standart hataları ile orantılamaktadır. Temel olarak bu metod, madde puanlarının olabilirlik fonksiyonu ile birey parametrelerinden elde edilen önsel dağılımı ile birleştirerek, birey parametrelerinin sonsal dağılımın tüm elementlerini kestirmek için kullanılır. Bu nedenle birey yeterlikleri ile ilgili hem özel hem de spesifik kapsam için sonsal dağılımlarda dayalı olarak yapılabilmektedir (Baker & Kim, 2004; Hambleton & Swaminathan, 1985).

Bu araştırma kapsamında iç geçerliği sağlamak amacıyla 200 replikasyon yapılmış ve her durum için 1-0 yanıt örüntüsünü içeren 200 veri seti elde edilmiştir. Toplamda (200x18) 3600 adet veri dosyası analiz edilmiştir. Replikasyon dosyalarında yanıt örüntüleri içinde, neredeyse tüm durumların yanlış (zero score) ya da tüm durumların doğru (perfect score) olduğu veri dosyaları bulunduğu için EAP yönteminin parametre kestiriminde daha uygun olduğu görülmüştür. Daha önce başka bir çalışmada (Zheng, 2013) kullanılmış olan syntax, bu çalışmaya uygun şekilde düzenlenerek, öncelikle madde ve birey parametreleri üretilmiş ve daha sonra buna uygun 200 adet (1-0) yanıt örüntüsü oluşturulmuştur. İlk anda üretilen ve kaydedilen bu parametreler gerçek (true) parametre olarak adlandırılmıştır. Bu gerçek parametrelerin modeller arası değişmezliği, üretimi aşamasında daha önce de belirtildiği üzere bir sabit (seed) atanarak sağlanmıştır. Üretilen 200 adet yanıt

örüntüsünün her birinden tekrar R yazılımında “mirt” (Chalmers, 2016) paketi ile birey ve parametre kestirimleri yapılmıştır.

Parametre kestirimlerinin replikasyonlar boyunca doğruluğunun (accuracy) değerlendirilmesi; ortalama yanlılık (bias), RMSE (Root Mean Square Error) ve kestirimlerin standart hatası (Standart Error/Standart Hata) ile yapılmıştır. Değerlendirme kriterleri olan ortalama yanlılık, RMSE, ve standart hata 18 durum için ayrı ayrı hesaplanmıştır. Ortalama yanlılık (Bias/ME) kestirilen parametrelerin gerçek parametrelerden farkının ortalamasıdır. Bu değerler parametre kestirimlerinin gerçek parametrelerden farkını vermektedir; mutlak yanlılık ne kadar küçükse parametre kestirimlerinin o kadar yansız olduğu söylenir. RMSE ise toplam/bütün parametre kestirimlerin doğruluğunu (accuracy) vermektedir. RMSE değeri ne kadar küçükse parametre kestirimleri o kadar doğru olacaktır. Standart hata (SE) ise parametre kestirimlerindeki kararlılığı (stability) vermektedir. SE ne kadar düşüğe o kadar kararlı parametre kestirimleri yapılabilir. Değerlendirme kriterleri aşağıda verilen formüller aracılığıyla hesaplanmıştır;

$$\text{Ortalama yanlılık } (\hat{\beta}) = \frac{\sum_{r=1}^R (\hat{\beta}_r - \beta)}{R}$$

$$\text{RMSE}(\hat{\beta}) = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\beta}_r - \beta)^2}$$

$$\text{SE}(\hat{\beta}) = \sqrt{\frac{1}{R} \sum_{r=1}^R (\hat{\beta}_r - \frac{\sum_{r=1}^R \hat{\beta}_r}{R})^2}$$

Yukarıdaki formüllerde verilen;

β : gerçek birey parametresi ya da madde parametreleri

$\hat{\beta}$: r'inci replikasyonda kestirilen birey ve madde parametreleridir (Li ve Rupp, 2011).

BÖLÜM 3

BULGULAR VE YORUMLAR

Bu bölümde, araştırmanın amacı doğrultusunda yapılan istatistiksel analizlerden elde edilen bulgular ve bulgulara ilişkin yorumlar yer almaktadır.

3.1. İki spesifik faktör arasındaki farklı diklik ihlal düzeylerinde (Model-1), madde ve birey parametrelerinin kestirimi ideal duruma (diklik ihlalinin olmadığı) göre nasıldır?

3.1.1. Model 1 İçin Madde Parametrelerinin Kestirim Yanlılıkları

Bu bölümde üçüncü ve dördüncü spesifik faktörler arasındaki diklik ihlali durumunda madde parametre kestirim yanılıkları verilmektedir. Diklik ihlal düzeylerinin hiyerarşik sıralaması Model 1.1 ($r_{3,4} = 0.10$), Model 1.2 ($r_{3,4} = 0.40$), Model 1.3 ($r_{3,4} = 0.70$) şeklindedir. Kestirilen ve gerçek parametreler arasındaki yanılık test uzunluğu ve model bazında tablolatırılmıştır. Yanlılık puanlarının ortalaması, standart sapması ve çeyrekler gibi betimsel istatistikler, farklı koşullarda gerçek parametreler ile kestirilen parametreleri karşılaştırarak parametreler arasındaki iyileşme (recovery) bakımından incelenmiştir.

3.1.1.1. Test Uzunluğu 12 Madde İken Madde Parametrelerinin Kestirim Yanlılıkları

Madde parametre kestirimlerindeki iyileşmeyi (recovery) incelemek için, gerçek parametrelerin, kestirilen parametrelerden farkının ortalaması alınarak yanılık hesaplanmıştır. Çizelge 3’de 12 madde için kestirim yanılıkları verilmiştir. Çizelge 3’de verilen güçlük (intercept/d) katsayılarına dair yanılıkların betimsel istatistikleri incelendiğinde, ortalama yanılıkların ($\bar{X}_{Model11} = 0.260$, $\bar{X}_{Model12} = 0.240$, $\bar{X}_{Model13} = 0.260$), tüm modellerde yüksek olup ve modeller arası (diklik

ihlal düzeylerine göre) fazla deęişkenlik göstermedięi görülmüştür. Benzer şekilde güçlük katsayılarının standart sapmaları ($\sigma_{Model11} = 0.330$, $\sigma_{Model12} = 0.340$, $\sigma_{Model13} = 0.330$) yani yanlılık puanları arasındaki deęişkenlik üç model için de hemen hemen aynıdır. Standart sapmanın yüksek olması, kestirilen parametrelerin gerçek parametrelerden deęişkenliğinin fazla olduğunu göstermektedir.

Çizelge 3

Ayrırt edicilik ve güçlük katsayılarının 12 madde için kestirim yanlılıkları

		\bar{X}	σ	Min	Q1	Q2	Q3	Max
MODEL 1.1	d katsayısı	0.260	0.330	-0.331	0.023	0.225	0.522	0.756
	G	0.007	0.040	-0.033	-0.020	-0.001	0.019	0.101
	S1	0.007	0.030	-0.014	-0.007	0.021	0.025	0.043
	S2	0.016	0.020	-0.002	0.002	0.007	0.053	0.070
	S3	-0.008	0.030	-0.032	-0.025	-0.017	0.004	0.026
	S4	0.017	0.130	-0.095	-0.055	-0.015	0.073	0.163
MODEL 1.2	d katsayısı	0.240	0.340	-0.346	-0.013	0.227	0.507	0.724
	G	0.022	0.120	-0.163	-0.057	0.020	0.094	0.274
	S1	0.055	0.020	0.033	0.044	0.055	0.066	0.077
	S2	0.064	0.020	0.050	0.050	0.050	0.071	0.092
	S3	-0.106	0.050	-0.162	-0.123	-0.083	-0.030	-0.073
	S4	-0.124	0.190	-0.332	-0.206	-0.081	-0.020	0.041
MODEL 1.3	d katsayısı	0.260	0.330	0.330	-0.300	0.018	0.216	0.520
	G	0.085	0.260	-0.300	0.018	0.216	0.520	0.757
	S1	0.100	0.250	-0.251	-0.106	0.044	0.223	0.511
	S2	0.090	0.060	0.036	0.072	0.109	0.130	0.152
	S3	-0.250	0.030	0.074	0.074	0.075	0.099	0.121
	S4	-0.441	0.090	-0.350	-0.282	-0.215	-0.196	-0.176

Üç modelde de uçdeęerler bulunmaktadır ($Max_{Model13} = 0.757$, $Min_{Model12} = -0.346$). Ayrıca Ek-B' de verilen grafikler incelendiğinde üç model için de güçlük parametresinin yanlılık puan ortalamasının sıfırın üzerinde olduğu görülmüştür. Yani güçlük parametresi her model için gerçek deęerden daha büyük kestirilmiştir. Ortalama ve standart sapmaların modeller arası deęişkenliğinin az olması nedeniyle diklik varsayımı ihlalinin güçlük parametre kestirimlerinde etkisi olmadığı söylenebilir. Yani güçlük parametresi en düşük diklik ihlali düzeyi için bile yanlı kestirilmiş ve bu yanlılık dięer ihlal düzeylerinde de aynı düzeyde görülmüştür.

Çizelge 3'de görülebileceęi gibi, ayrırt edicilik parametrelerinin üçüncü ve dördüncü spesifik faktörlerdeki diklik ihlalinin birinci ve ikinci spesifik faktörlerdeki parametre kestirimlerine çok fazla etkisi olmamıştır. Ayrırt edicilik parametrelerinin

ortalama yanlılığı üçüncü spesifik modelde ilk ve ikinci spesifik modele göre daha yüksektir ($\bar{X}_{S1\&Model11} = 0.007$, $\bar{X}_{S1\&Model12} = 0.055$, $\bar{X}_{S1\&Model13} = 0.100$; $\bar{X}_{S2\&Model11} = 0.016$, $\bar{X}_{S2\&Model12} = 0.064$, $\bar{X}_{S2\&Model13} = 0.090$).

Yanlılık puanlarının standart sapmaları incelendiğinde, değişkenliğin düşük olduğu ve diklik varsayımı ihlalinin yanlılık puanlarının değişkenliğine neredeyse hiç etkisinin olmadığı söylenebilir (örn; $\sigma_{S1\&Model11} = 0.030$, $\sigma_{S1\&Model12} = 0.020$, $\sigma_{S1\&Model13} = 0.060$; $\sigma_{S2\&Model11} = 0.020$; $\sigma_{S2\&Model12} = 0.020$, $\sigma_{S2\&Model13} = 0.030$). Bu bulgulara göre üçüncü ve dördüncü spesifik faktör arasındaki diklik varsayımı ihlalinin diğer spesifik faktörlerdeki parametre kestirimlerine çok fazla etkisi olmadığı görülmektedir.

Bu durumun aksine, üçüncü ve dördüncü spesifik faktörlerdeki diklik varsayımı ihlalinin aynı spesifik faktörlerdeki parametre kestirimlerine etkisi olduğu görülmüştür. Bu spesifik faktörlerdeki ayırt edicilik parametreleri incelendiğinde, diklik ihlali arttıkça kestirilen parametreler ile gerçek parametreler arasındaki farklılık da artmaktadır ($\bar{X}_{S3\&Model11} = -0.008$, $\bar{X}_{S3\&Model12} = -0.106$, $\bar{X}_{S3\&Model13} = -0.250$; $\bar{X}_{S4\&Model11} = 0.017$, $\bar{X}_{S4\&Model12} = -0.124$, $\bar{X}_{S4\&Model13} = -0.441$). Diklik ihlal düzeylerindeki artış, parametre kestirimlerindeki yanlılık puan ortalamalarında da artışa neden olmuştur. Standart sapma değerleri de diklik ihlal düzeylerindeki artışın yanlılık puanlarındaki değişkenliği artırdığını göstermektedir ($\sigma_{S4\&Model11} = 0.130$, $\sigma_{S4\&Model12} = 0.190$, $\sigma_{S4\&Model13} = 0.520$). Standart sapmaların yüksek olması kestirilen parametrelerin gerçek parametrelerden değişkenliğinin fazla olduğunu göstermektedir. Ayrıca standart sapmalardaki spesifik modeller arası artış, kestirimlerin güvenilirliğinin düştüğünü gösterir. Yani diklik ihlal düzeyi arttıkça güvenilir kestirimler yapılamamaktadır.

Bu sonuçlar, ilgili spesifik faktörlerdeki diklik ihlalinin, yine aynı spesifik faktörlerdeki parametre kestirim yanlılığını artırdığını göstermektedir. İki spesifik faktördeki diklik ihlali genel faktördeki parametre kestirimlerini de etkilemiştir. Kestirilen ve gerçek parametreler arasındaki yanlılık puanlarındaki artış birinci ve ikinci model arasında az olmakla birlikte ikinci ve üçüncü model arasında daha fazladır ($\bar{X}_{G\&Model11} = 0.007$, $\bar{X}_{G\&Model12} = 0.022$, $\bar{X}_{G\&Model13} = 0.085$). Yanlılık puanlarının standart sapmaları incelendiğinde yine diklik ihlali arttıkça

değişkenliğin arttığı görülmektedir ($\sigma_{G\&Model11} = 0.040$, $\sigma_{G\&Model12} = 0.120$, $\sigma_{G\&Model13} = 0.250$). Bu sonuçlar ışığında diklik varsayımı ihlalinin genel faktörde de etkili olduğu söylenebilir.

Ek-B’de verilen grafikler incelendiğinde sadece Model 1.1’de uç değerler bulunmaktadır. Yanlılık puanlarının ranjı 1.744’dür ($Max_{Model13} = 0.757$, $Min_{Model13} = -0.987$). Yanlılık puanlarının ortalamaları sıfırın altında olmasından dolayı, üçüncü ve dördüncü spesifik faktörlerdeki diklik ihlal düzeyindeki artışın, parametre değerlerinin gerçek değerlerden daha düşük kestirilmesine neden olduğu söylenebilir.

3.1.1.2. Test Uzunluğu 40 Madde İken Madde Parametrelerinin Kestirim Yanlılıkları

12 maddelik test uzunluğu için yapılan yanlılık hesaplamaları 40 maddelik test uzunluğu için de yapılmıştır. Madde parametre kestirimlerindeki iyileşmeyi (recovery) incelemek için, gerçek parametrelerin, kestirilen parametrelerden farkının ortalaması alınarak yanlılık hesaplanmıştır. Çizelge 4’de 40 madde için kestirim yanlılıkları verilmiştir.

Çizelge 4

		\bar{X}	σ	Min	Q1	Q2	Q3	Max
Model 1.1	d katsayısı	0.049	1.780	-3.845	-0.783	0.085	1.302	3.430
	G	-0.037	0.640	-1.194	-0.390	0.005	0.083	1.292
	S1	-0.004	0.010	-0.018	-0.006	-0.003	0.000	0.006
	S2	-0.154	0.560	-1.020	-0.617	-0.005	0.207	0.562
	S3	-0.020	0.880	-1.600	-0.507	-0.262	0.723	1.094
	S4	-0.010	0.860	-1.228	-0.650	-0.044	0.515	1.691
Model 1.2	d katsayısı	0.083	1.780	-3.808	-0.786	0.112	1.347	3.454
	G	0.002	0.650	-1.179	-0.308	-0.002	0.205	1.414
	S1	0.030	0.020	-0.013	0.019	0.037	0.043	0.051
	S2	-0.150	0.560	-1.024	-0.609	0.000	0.221	0.567
	S3	-0.095	0.880	-1.714	-0.582	-0.353	0.660	0.991
	S4	-0.050	0.860	-1.249	-0.700	-0.080	0.468	1.703
Model 1.3	d katsayısı	0.060	1.780	-3.838	-0.777	0.083	1.323	3.430
	G	0.061	0.660	-1.186	-0.210	-0.016	0.338	1.627
	S1	0.066	0.050	-0.043	0.055	0.080	0.101	0.113

Ayrt edicilik ve güçlük katsayılarının 40 madde için kestirim yanlılıkları

Çizelge 4 Devam

		\bar{X}	σ	Min	Q1	Q2	Q3	Max
Model	S2	-0.134	0.570	-1.007	-0.601	0.008	0.242	0.613
1.3	S3	-0.213	0.880	-1.894	-0.700	-0.484	0.572	0.812
	S4	-0.180	0.860	-1.293	-0.741	-0.263	0.256	1.647

Çizelge 4’de verilen güçlük (intercept/d) katsayılarına dair yanlılık puanlarının betimsel istatistikleri incelendiğinde, yanlılık ortalamalarının 12 maddelik test uzunluğuna göre göreceli olarak daha düşük iken bu durumun aksine standart sapma değerlerinin 12 maddelik test uzunluğuna göre oldukça yüksek olduğu görülmüştür. Standart sapma değerlerinin yüksek olması kestirilen parametrelerin gerçek parametrelerden oldukça farklılaştığını ve kestirim güvenilirliğinin düştüğünü göstermektedir. Modeller arası (Model1.1, Model 1.2, Model 1.3) yanlılık ortalamaları ve standart sapmalar çok fazla değişmemesinden dolayı spesifik modellerin (Model1.1, Model 1.2, Model 1.3) güçlük katsayısının kestirimine etkisinin fazla olmadığı söylenebilir. Bunun yanı sıra 40 maddelik test uzunluğunda güçlük katsayılarının kestirim güvenilirliği, standart sapmanın büyümesinden dolayı düşmüştür.

Çizelge 4’de verildiği üzere ayırt edicilik parametreleri incelendiğinde, 12 maddelik test uzunluğunda olduğu gibi, üçüncü ve dördüncü spesifik faktörlerdeki diklik ihlalinin birinci spesifik faktördeki parametre kestirimlerine çok fazla etkisi olmazken, aksine ikinci spesifik faktördeki parametre kestirimlerine göreceli olarak daha fazla etkisi olduğu görülmüştür. İkinci spesifik faktördeki yanlılık ortalamaları incelendiğinde, ortalamaların modeller arası nerdeyse aynı olduğu görülmüştür. Standart sapma değerleri incelendiğinde ikinci spesifik faktördeki yanlılık puanlarının değişkenliği ilk spesifik faktördekine göre daha yüksektir fakat modeller arası bu değişkenlik her iki faktör için de yok denecek kadar azdır ($\sigma_{S1\&Model11} = 0.010$, $\sigma_{S1\&Model12} = 0.020$, $\sigma_{S1\&Model13} = 0.050$; $\sigma_{S2\&Model11} = 0.560$, $\sigma_{S2\&Model12} = 0.560$, $\sigma_{S2\&Model13} = 0.570$). Bu sonuçlara göre, yanlılık puanlarının ortalama ve standart sapmaları incelendiğinde, 12 maddelik test uzunluğunun aksine, ikinci spesifik faktörde yanlılık puanlarının değişkenliği artmıştır. 12 maddelik test uzunluğuna kıyasla, 40 maddelik test uzunluğundaki standart sapmalardaki artış, kestirim güvenilirliğinin düştüğünü gösterir. Yani diklik ihlal düzeyi arttıkça gerçek parametrelerden

kestirilen parametrelerin farklılığı artmaktadır ve güvenilir kestirimler yapılamamaktadır.

Üçüncü ve dördüncü spesifik faktörlerdeki ayırt edicilik parametreleri incelendiğinde 12 maddelik test uzunluğunun aksine yanlılık puan ortalamaları 40 maddelik test uzunluğunda daha düşüktür. Diklik ihlalini içeren faktörlerdeki ayırt edicilik parametreleri incelendiğinde, diklik ihlali arttıkça kestirilen parametreler ile gerçek parametreler arasındaki farklılık da artmaktadır ($\bar{X}_{S3\&Model11} = -0.020$, $\bar{X}_{S3\&Model12} = -0.095$, $\bar{X}_{S3\&Model13} = -0.213$; $\bar{X}_{S4\&Model11} = -0.010$, $\bar{X}_{S4\&Model12} = -0.050$, $\bar{X}_{S4\&Model13} = -0.178$). Diklik ihlal düzeylerindeki artış parametre kestirim yanlılıklarında da artışa neden olmuştur. Fakat standart sapma değerleri diklik ihlal düzeylerindeki artışa rağmen üçüncü ve dördüncü spesifik faktörde üç model için de aynı kalmıştır.

12 maddelik test uzunluğuna kıyasla 40 maddelik test uzunluğunda yanlılık ortalamalarının düşük olduğu görülmüştür. Bununla birlikte, 40 maddelik test uzunluğunda yanlılık puanlarının değişkenliği fazla olmakla birlikte bu değişkenlik modeller arası (Model1.1, Model 1.2, Model 1.3) neredeyse aynıdır. Standart sapmalardaki bu artış kestirimlerin güvenilirliğinin düştüğünü göstermektedir. Ayırt edicilik parametrelerinin tüm spesifik modeller için kestirim güvenilirliği düşüktür.

Tüm bu sonuçlar ilgili spesifik faktörlerdeki diklik ihlalinin yine aynı spesifik faktörlerdeki parametre kestirimlerindeki yanlılığı artırdığını göstermiştir. İki spesifik faktördeki diklik ihlali genel faktördeki parametre kestirimlerini çok fazla etkilememiştir. Kestirilen ve gerçek parametre arasındaki farklılık birinci ve ikinci model arasında az olmakla birlikte ikinci ve üçüncü model arasında göreceli olarak daha fazladır ($\bar{X}_{G\&Model11} = -0.037$, $\bar{X}_{G\&Model12} = 0.002$, $\bar{X}_{G\&Model13} = 0.061$). Yanlılık puanlarının standart sapmaları incelendiğinde yanlılık puanlarındaki değişkenlik modeller arası neredeyse aynıdır. 12 maddelik test uzunluğu ile kıyaslandığında, 40 maddelik test uzunluğunda genel boyuttaki yanlılık ortalamaları düşük fakat değişkenlik daha fazladır. Standart sapmaların yüksek olması kestirim güvenilirliğinin 40 madde için 12 maddeye kıyasla daha düşük olduğunu göstermektedir.

Ek-C'de verilen grafikler incelendiğinde genel faktörde ve birinci spesifik faktörlerde uç değerler olduğu görülmektedir. Yanlılık puanlarının ranjı "7.299"

($Max_{Model13} = 3.454$, $Min_{Model13} = -3.845$) olarak bulunmuştur. Yanlılık puanlarının ortalamaları sıfırın altında olmasından dolayı, tüm modeller için, ikinci, üçüncü ve dördüncü spesifik faktörlerdeki parametrelerin gerçek değerlerden daha düşük kestirildiği söylenebilir.

3.1.1.3. Test Uzunluğu 100 Madde İken Madde Parametrelerinin Kestirim Yanlılıkları

12 ve 40 maddelik test uzunluğu için yapılan yanlılık hesaplamaları 100 maddelik test uzunluğu için de yapılmıştır. Madde parametre kestirimlerindeki iyileşmeyi (recovery) incelemek için, gerçek parametrelerin, kestirilen parametrelerden farkının ortalaması alınarak yanlılık hesaplanmıştır. Çizelge 5’de 100 madde için kestirim yanlılıkları verilmiştir.

Çizelge 5

Ayırt edicilik ve güçlük katsayılarının 100 madde için kestirim yanlılıkları

		\bar{X}	σ	Min	Q1	Q2	Q3	Max
Model 1.1	d katsayısı	-0.078	1.830	-4.310	-1.084	-0.029	1.247	4.061
	G	0.030	0.670	-1.672	-0.407	0.000	0.437	1.649
	S1	-0.070	0.600	-1.389	-0.018	-0.007	0.037	1.260
	S2	-0.021	0.830	-1.990	-0.626	-0.017	0.664	1.189
	S3	-0.017	0.870	-1.715	-0.367	0.045	0.602	1.800
	S4	0.015	0.550	-0.982	-0.367	-0.121	0.472	1.158
Model 1.2	d katsayısı	-0.066	1.830	-4.290	-1.061	-0.031	1.270	4.051
	G	0.004	0.660	-1.690	-0.445	-0.044	0.421	1.577
	S1	-0.051	0.830	-1.389	-0.002	0.007	0.060	1.284
	S2	0.011	0.830	-1.990	-0.381	0.002	0.561	1.755
	S3	-0.051	0.860	-1.715	-0.414	0.015	0.555	1.978
	S4	-0.003	0.570	-1.035	-0.401	-0.133	0.441	1.203
Model 1.3	d katsayısı	-0.027	1.830	-4.399	-1.022	0.000	1.310	4.106
	G	0.070	0.670	-1.710	-0.374	-0.011	0.517	1.572
	S1	-0.042	0.600	-1.389	-0.004	0.020	0.084	1.286
	S2	0.010	0.820	-1.990	-0.600	-0.003	0.701	1.212
	S3	-0.130	0.850	-1.715	-0.436	-0.086	0.457	1.668
	S4	-0.100	0.580	-1.160	-0.523	-0.204	0.326	1.153

Çizelge 5’e göre güçlük (intercept) katsayıları incelendiğinde, katsayılarının ortalama yanlılıkları tüm modellerde çok yüksek olmamakla birlikte diklik ihlal

düzeylerine göre fazla değişkenlik göstermemiştir. 12 maddelik test uzunluğuna göre, yanlılık ortalamaları daha düşüktür. Standart sapma değerleri göre yüksek olmakla birlikte modeller arası (Model 1.1, Model 1.2, Model 1.3) aynı kalmıştır. Üç test uzunluğu koşulunda (12, 40, 100) güçlük katsayıları ele alınacak olursa, bu koşullar arasında test uzunluğu arttıkça yanlılık ortalamalarının düştüğü gözlemlenmiştir. Bu durumun aksine ise test uzunluğu arttıkça değişkenlik artmıştır. Değişkenliğin artması kestirim güvenilirliğini düşürmektedir. En küçük diklik ihlal düzeyinde bile, tüm test uzunluklarında güçlük katsayısının kestirim güvenilirliği düşüktür.

Çizelge 5’de verildiği üzere, ayırt edicilik parametreleri incelendiğinde, 12 ve 40 maddelik test uzunluklarında olduğu gibi, üçüncü ve dördüncü faktörlerdeki diklik ihlalinin birinci ve ikinci faktörlerdeki yanlılık ortalamalarına çok fazla etkisi olmadığı görülmüştür. Standart sapmalar incelendiğinde ikinci spesifik faktördeki yanlılık puanlarının değişkenliğinin, ilk spesifik faktöre göre daha fazla olduğu görülmüştür. Fakat, modeller arası bu değişkenlik ikinci spesifik faktörde neredeyse aynı kalmıştır. Sonuç olarak 12 ve 40 maddelik test uzunluğuna kıyasla 100 maddelik test uzunluğunda, yanlılık puanlarının standart sapmalarının oldukça büyük olduğu görülmüştür.

Üçüncü ve dördüncü spesifik faktörlerdeki ayırt edicilik parametreleri incelendiğinde, 40 maddelik test uzunluğunun aksine, yanlılık puan ortalamalarının 100 maddelik test uzunluğunda daha düşük olduğu görülmüştür. Tüm test uzunlukları için, diklik ihlali içeren spesifik faktörlerdeki ayırt edicilik parametreleri incelendiğinde, diklik ihlali arttıkça kestirilen parametreler ile gerçek parametreler arasındaki farklılık artmaktadır. Diklik ihlal düzeylerindeki artışın, parametre kestirimlerindeki yanlılık ortalamalarına etkisi olduğu söylenebilir. Bu bulgu Zheng (2013)’ün çalışmasıyla benzerlik göstermektedir. Genel olarak modeller arası değişkenlik her iki faktörde de neredeyse aynıdır. 12 maddelik test uzunluğuna kıyasla standart sapma değerleri oldukça yüksektir. Bu durum, faktörlerdeki binişik madde sayısının artması ile açıklanabilir. Madde sayısı arttıkça, spesifik faktörlerdeki değişkenlik daha da artmıştır.

Son olarak, genel faktördeki yanlılık ortalamaları incelendiğinde, kestirilen ve gerçek parametreler arasındaki farklılığın 40 maddelik test uzunluğuna göre daha az olduğu görülmüştür. Yanlılık puanlarının standart sapmaları incelendiğinde yanlılık puanlarındaki değişkenlik 40 maddelik test uzunluğu ile benzerlik göstermekte ve

modeller arası neredeyse aynıdır.

Ek-D'de verilen grafikler incelendiğinde birinci spesifik faktör için uç değerler olduğu görülmüştür. Yanlılık puan ortalamaları sıfırın altında olduğu için, tüm modellerde birinci ve üçüncü spesifik faktör parametrelerinin gerçek değerden daha düşük kestirildiği söylenebilir. İkinci ve dördüncü spesifik faktörde ise diklik ihlal düzeyi arttıkça parametreler gerçek değerden daha büyük kestirilmiştir. Yanlılık puanlarının ranjı “8.505” ($Max_{Model13} = 4.106$, $Min_{Model12} = -4.399$) olarak bulunmuştur. Ranjın bu kadar geniş olması kestirim güvenilirliğinin düşmesine neden olmaktadır. Aynı zamanda tüm modellerde genel faktör ve spesifik faktörlerde standart sapmanın da yüksek olması kestirimlerinin güvenilirliğini düşürmektedir.

Madde parametre kestirimlerinde tüm test uzunlukları için görülen örüntü şu şekildedir; üç test uzunluğunda (12, 40, 100) da sıfıra yakın yanlılık ortalamalarına sahip faktörler genel, birinci ve ikinci faktördür. Yani diklik ihlali içermeyen faktörlerdir. Bu durumun aksine diklik ihlali içeren faktörlerde test uzunluğuna bağlı olarak yanlılık puan ortalamalarında düşüş gözlenmiştir. Model 1.3 ele alınacak olursa, diklik ihlali içeren faktörlerde (üçüncü ve dördüncü) test uzunluğu arttıkça yanlılık ortalamalarının da düştüğü gözlenmiştir (örn; $\bar{X}_{S3\&12} = -0.250$, $\bar{X}_{S4\&12} = -0.441$; $\bar{X}_{S3\&40} = -0.213$, $\bar{X}_{S4\&40} = -0.180$; $\bar{X}_{S3\&100} = -0.130$, $\bar{X}_{S4\&100} = -0.100$). Test uzunluğunun parametre kestirim yanlılığına etkisi olabilir. Bu durumun aksine ise test uzunluğu arttıkça standart sapma büyümüş, ranj genişlemiştir. 12 madde için en büyük uç değerler 0.757 ve -0.987 iken 40 madde için 3.454 ve -3.845, 100 madde için ise 4.106 ve -4.399. Ranjın geniş, standart sapmanın yüksek olması kestirim güvenilirliğini düşürmektedir. Yani madde sayısındaki artış kestirim kesinliğinde yani güvenilirliğinde düşüşe neden olmuştur. Bu durum yanlı madde miktarındaki artış ile açıklanabilir. Yani modele ne kadar korelatif madde eklenirse değişkenlik o kadar artmıştır.

3.1.2. Model 1 İçin Birey Parametrelerinin Kestirim Yanlılıkları

Bu bölümde üçüncü ve dördüncü spesifik faktörler arasındaki diklik ihlali durumunda birey parametrelerinin kestiriminde oluşan yanlılıklar verilmektedir. Diklik ihlal düzeylerinin hiyerarşik sıralaması Model 1.1 için ($r_{3,4} = 0.1$), Model 1.2 ($r_{3,4} = 0.4$), Model 1.3 ($r_{3,4} = 0.7$) şeklindedir. Yanlılık puanları, test uzunluğu ve

model bazında tablolaştırılmıştır. Yanlılık puanlarının ortalaması, standart sapması ve çeyrekler gibi betimsel istatistikler, farklı koşullar için gerçek parametreler ile kestirilen parametreleri karşılaştırarak parametreler arasındaki iyileşme (recovery) bakımından incelenmiştir.

3.1.2.1. Test Uzunluğu 12 Madde İken Birey Parametrelerinin Kestirim Yanlılıkları

Madde parametre kestirimlerindeki iyileşmeyi (recovery) incelemek için, gerçek parametrelerin, kestirilen parametrelerden farkının ortalaması alınarak yanlılık hesaplanmıştır. Çizelge 6'de 12 madde için kestirim yanlılıkları verilmiştir.

Çizelge 6

Birey parametrelerinin 12 madde için kestirim yanlılıkları

		\bar{X}	σ	Min.	Q1	Q2	Q3	Max.
Model 1.1	G	0.017	0.500	-2.517	-0.300	0.007	0.325	3.835
	S1	-0.017	0.740	-2.439	-0.517	-0.032	0.483	2.880
	S2	-0.026	0.650	-2.884	-0.447	-0.073	0.363	2.850
	S3	-0.020	0.650	-3.061	-0.436	-0.042	0.380	2.966
	S4	-0.023	0.780	-2.867	-0.551	-0.038	0.487	3.114
Model 1.2	G	0.009	0.510	-1.841	-0.306	0.006	0.326	2.280
	S1	0.016	0.750	-2.551	-0.476	0.008	0.500	3.511
	S2	-0.040	0.650	-2.936	-0.455	-0.064	0.357	3.353
	S3	0.010	0.670	-3.109	-0.420	0.002	0.441	2.944
	S4	0.020	0.820	-3.620	-0.533	0.016	0.547	3.190
Model 1.3	G	0.011	0.550	-2.390	-0.350	-0.000	0.364	2.252
	S1	0.003	0.730	-2.596	-0.488	0.006	0.483	2.686
	S2	-0.034	0.650	-2.311	-0.452	-0.060	0.354	2.940
	S3	-0.020	0.710	-3.523	-0.458	-0.036	0.421	3.004
	S4	-0.018	0.870	-4.092	-0.587	-0.040	0.528	3.004

Çizelge 6'da 12 maddelik test uzunluğu için birey parametrelerindeki kestirim yanlılıkları verilmektedir. Tüm spesifik modeller için birey (theta) parametreleri gerçek parametrelerden çok farklıdır çünkü yanlılık puanları geniş bir ranja sahiptir. Yanlılık puanlarının %50'si -0.658 ile 0.616 arasında (örn., $Q1_{G\&Model11} = -0.300$, $Q2_{G\&Model11} = 0.007$; $Q3_{G\&Model11} = 0.325$; $Q1_{S1\&Model11} = -0.517$, $Q2_{S1\&Model11} = -0.032$; $Q3_{S1\&Model11} = 0.483$) ve yanlılık puanlarının diğer %50'lik kısmı ise -4.092 ile 3.511 arasındadır. Yanlılık

puanlarının dağılımı genellikle simetriktir çünkü yanlılık puanlarının ortalaması tüm spesifik modeller için sıfır civarındadır (örn.; $\bar{X}_{G\&Model11} = 0.017$, $\bar{X}_{S1\&Model11} = -0.017$, $\bar{X}_{S2\&Model11} = -0.026$, $\bar{X}_{S3\&Model11} = -0.020$, $\bar{X}_{S4\&Model11} = -0.023$).

Tüm spesifik modeller için, genel faktördeki yanlılık puanlarının standart sapması en düşük, dördüncü spesifik faktördeki için ise en yüksek (örn; $\sigma_{G\&Model11} = 0.500$, $\sigma_{S4\&Model11} = 0.780$) olmakla birlikte genel olarak tüm faktörlerde değişkenlik oldukça yüksektir. Değişkenliğin yüksek ve ranjin geniş olması kestirim güvenilirliğini düşürmektedir. Tüm spesifik modeller için, genel faktördeki yanlılık puanlarının değişkenliği diğer faktörlerden daha azdır. Bu nedenle genel faktördeki parametre kestirimlerinin diğer faktörlere göre daha güvenilir olduğu söylenebilir. Buradan yola çıkarak spesifik modellerin (diklik ihlal düzeyleri) birey (theta) parametrelerinin kestirim yanlılığına etkisi olmadığı söylenebilir.

EK-E'de birey (theta) parametrelerinin 12 madde için yanlılık puanlarının grafikleri verilmiştir. Grafikler incelendiğinde tüm spesifik modellerde tüm faktörlerin (genel ve spesifik) ortalamaları yaklaşık olarak aynı ve sıfır civarındadır. Bununla birlikte yanlılık puanları geniş bir ranja sahiptir. Tüm faktörler (genel ve spesifik) arasında dördüncü spesifik faktördeki parametrelerin kestirim kesinliği en düşük parametreler, genel faktördeki parametreler ise göreceli olarak kestirim güvenilirliği en yüksek parametrelerdir. Genel olarak kestirilen birey parametreleri gerçek birey parametrelerinden oldukça farklı, kestirim güvenilirliği ve kesinliği düşüktür.

3.1.2.2. Test Uzunluğu 40 Madde İken Birey Parametrelerinin Kestirim Yanlılıkları

12 maddelik test uzunluğunda olduğu gibi, madde parametre kestirimlerindeki iyileşmeyi (recovery) incelemek için, gerçek parametrelerin, kestirilen parametrelerden farkının ortalaması alınarak yanlılık hesaplanmıştır. Çizelge 7'de 40 madde için kestirim yanlılıkları verilmiştir.

Çizelge 7'de verildiği üzere 40 maddelik test uzunluğunda yanlılık puanları 12 maddelik test uzunluğunda olduğu gibi geniş bir ranja (Min=2.878, Max= 3.041) sahiptir fakat puanlarının ranjı 12 maddelik test uzunluğuna göre 40 madde için daha dardır. Yanlılık puan dağılımları tüm modellerde simetriktir ve yanlılık ortalamaları

0 civarındadır. Tüm modellerde, yanlılık puanlarının standart sapma değerleri genel faktörde en düşük iken, üçüncü spesifik faktörde en yüksektir. Tüm faktörlerde (genel ve spesifik), son modelde (Model 1.3) yanlılık puanlarındaki değişkenlik diğer modellerden daha fazladır.

Çizelge 7

Birey parametrelerinin 40 madde için kestirim yanlılıkları

		\bar{X}	σ	Min	Q1	Q2	Q3	Max
Model 1.1	G	-0.010	0.290	-1.575	-0.193	-0.011	0.170	1.622
	S1	0.010	0.450	-2.400	-0.267	0.004	0.283	2.107
	S2	-0.010	0.440	-2.261	-0.283	-0.019	0.253	2.341
	S3	0.030	0.440	-2.532	-0.222	0.039	0.290	2.751
	S4	-0.004	0.430	-2.141	-0.218	0.013	0.241	2.527
Model 1.2	G	-0.029	0.310	-2.128	-0.226	-0.027	0.169	1.681
	S1	-0.010	0.480	-2.513	-0.305	-0.025	0.276	3.041
	S2	-0.033	0.460	-2.032	-0.316	-0.040	0.241	2.145
	S3	0.022	0.460	-2.210	-0.248	0.041	0.294	2.240
	S4	0.010	0.440	-2.753	-0.219	0.023	0.260	2.340
Model 1.3	G	-0.010	0.360	-1.758	-0.244	-0.014	0.226	1.508
	S1	0.004	0.470	-1.977	-0.288	-0.007	0.301	2.372
	S2	-0.020	0.480	-2.117	-0.323	-0.020	0.279	2.708
	S3	0.020	0.520	-2.878	-0.280	0.041	0.337	2.543
	S4	-0.010	0.480	-2.724	-0.248	0.007	0.256	2.400

Değişkenliğin diğer modellere göre son modelde daha yüksek olması nedeniyle, parametre kestirim güvenilirliğinin diğer modellere göre bu modelde daha düşük olduğu söylenebilir Yanlılık puanlarının ranjı 12 madde için -4.092 ile 3.511 arasında iken, 40 madde için -2.878 ile 3.041 arasındadır. Tüm faktörler (genel ve spesifik) arasında, dördüncü spesifik faktörde madde kestirimleri 12 maddelik test uzunluğu için en az güvenilir iken, 40 maddelik test uzunluğunda tüm faktörlerdeki parametre kestirim güvenilirliği 12 maddelik test uzunluğuna göre daha yüksektir. Ranjin geniş ve değişkenliğin yüksek olmasından dolayı spesifik modeller için birey parametrelerinin kestirim güvenilirlikleri düşüktür.

Ek-F'de birey (theta) parametrelerinin 40 madde için yanlılık puan grafikleri verilmiştir. Grafikler incelendiğinde tüm spesifik modellerde tüm faktörlerin (genel ve spesifik) ortalamaları yaklaşık olarak aynı ve sıfır civarındadır. Bununla birlikte yanlılık puanları geniş bir ranja sahiptir. Genel olarak kestirilen birey parametreleri

gerçek birey parametrelerinden oldukça farklıdır.

3.1.2.3. Test Uzunluğu 100 Madde İken Birey Parametrelerinin Kestirim Yanlılıkları

12 ve 40 maddelik test uzunluğunda olduğu gibi, madde parametre kestirimlerindeki iyileşmeyi (recovery) incelemek için, gerçek parametrelerin, kestirilen parametrelerden farkının ortalaması alınarak yanlılık hesaplanmıştır. Çizelge 8’de 100 madde için kestirim yanlılıkları verilmiştir.

Çizelge 8

Birey parametrelerinin 100 madde için kestirim yanlılıkları

		\bar{X}	σ	Min	Q1	Q2	Q3	Max
Model 1.1	G	0.021	0.240	-1.437	-0.132	0.023	0.180	2.135
	S1	-0.005	0.330	-1.800	-0.225	-0.013	0.207	1.830
	S2	0.012	0.370	-0.220	-0.434	0.016	0.253	1.970
	S3	0.017	0.340	-1.706	-0.200	0.013	0.234	1.848
	S4	-0.010	0.320	-1.772	-0.194	-0.006	0.180	2.005
Model 1.2	G	0.002	0.250	-1.436	-0.171	0.004	0.172	1.122
	S1	0.005	0.350	-1.753	-0.217	0.000	0.222	1.718
	S2	0.005	0.380	-2.496	-0.238	0.020	0.255	1.940
	S3	0.017	0.360	-2.758	-0.194	0.017	0.235	2.486
	S4	0.015	0.330	-2.243	-0.173	0.021	0.213	1.970
Model 1.3	G	-0.007	0.280	-1.086	-0.195	-0.004	0.180	1.190
	S1	-0.011	0.360	-2.224	-0.235	-0.014	0.208	2.600
	S2	-0.031	0.400	-2.307	-0.290	-0.015	0.234	1.721
	S3	0.001	0.390	-3.251	-0.230	0.001	0.223	1.860
	S4	-0.001	0.350	-2.790	-0.193	0.000	0.204	2.432

Çizelge 8’de 100 maddelik test uzunluğu için birey parametrelerinin kestirim yanlılıkları verilmiştir. Sonuçlar 12 maddelik test uzunluğu ve 40 maddelik test uzunluğu ile benzerlik göstermektedir. 100 maddelik test uzunluğu için tüm faktörlerdeki (genel ve spesifik) yanlılık ortalamaları birbirine yakın ve sıfır civarındadır. Oldukça fazla uç değer bulunmakla birlikte puanların ranjı 40 maddelik test uzunluğuna göre geniştir. Genel faktördeki parametre kestirimleri diğer faktörlere göre daha güvenilirdir çünkü yanlılık puanların ranjı diğer faktörlere göre genel faktörde daha dardır. Test uzunluğunu 100 maddeye çıkarmak yanlılık puanlarının standart sapma değerlerini bir miktar düşürmüştür. Yanlılık puanlarının %50’lik kısmı yani 1. ve 3. çeyrekler arası, 100 maddelik test uzunluğu için, 40

maddelik test uzunluğu ile nerdeyse aynı, 12 maddelik test uzunluğundan ise daha dardır ($Q1_{12} = -0.587$, $Q3_{12} = 0.528$; $Q1_{40} = -0.323$, $Q3_{40} = 0.279$; $Q1_{100} = -0.434$, $Q3_{100} = 0.253$).

Ek-G'de birey (theta) parametrelerinin 100 madde için yanlılık puanlarının grafikleri verilmiştir. Grafikler tablo bulgularını doğrular nitelikte olmakla birlikte, tüm spesifik modellerde tüm faktörlerin (genel ve spesifik) ortalamalarının yaklaşık olarak aynı ve sıfır olduğu görülmüştür.

Son olarak tüm test uzunlukları (12, 40, 100) birlikte incelendiğinde, en fazla bozulmuş parametrelerin 12 maddelik test uzunluğunda her model için dördüncü spesifik faktörde olduğu görülmüştür. Genel olarak tüm test uzunluklarında yanlılık puanlarının değişkenliği yüksek ve ranjı geniştir. En geniş ranj 12 maddelik test uzunluğundadır. Test uzunluğunu artırmak yanlılık puanlarının ortalamasını düşürmüş ve ranjını daraltmıştır. Bu bulgu Zheng(2013)'ün çalışması ile uyum göstermektedir. 12 maddelik test uzunluğuna göre 40 ve 100 maddelik test uzunluğunda kestirim kesinliğini (accuracy) daha yüksektir. Test uzunluğu ile doğru orantılı şekilde ranjin daralması test uzunluğunun parametre iyileşmesinde etkisi olabileceğine işaret etmektedir. Yine de ranj ve değişkenlik tüm test uzunluklarında yüksektir. Bu durum parametre kestirim güvenilirliklerini düşürmektedir.

3.2. Tüm spesifik faktörler arasındaki farklı diklik ihlal düzeylerinde (Model-2), madde ve birey parametrelerinin kestirimi, ideal duruma (diklik ihlalinin olmadığı) göre nasıldır?

3.2.1. Model 2 İçin Madde Parametrelerinin Kestirim Yanlılıkları

Bu bölümde tüm spesifik faktörler arasındaki diklik ihlali durumunda madde parametre yanlılıkları verilmiştir. Çizelge 9, 10 ve 11'de de verildiği üzere en düşük diklik ihlal düzeyi $r=0.020$ olarak en yüksek diklik ihlal düzeyi ise $r=0.660$ olarak belirlenmiştir. Kestirilen ve gerçek parametreler arasındaki yanlılık test uzunluğu ve model bazında tablolaştırılmıştır. Yanlılık puanlarının ortalaması, standart sapması ve çeyrekler gibi betimsel istatistikler, farklı koşullar için gerçek parametreler ile kestirilen parametreleri karşılaştırarak parametreler arasındaki iyileşme (recovery) bakımından incelenmiştir.

Çizelge 9

Model 2.1 İçin Korelasyon Değerleri

	P	S1	S2	S3	S4
P	1.000				
S1	0.000	1.000			
S2	0.000	0.020	1.000		
S3	0.000	0.040	0.080	1.000	
S4	0.000	0.060	0.090	0.100	1.000

Çizelge 10

Model 2.2 İçin Korelasyon Değerleri

	P	S1	S2	S3	S4
P	1.000				
S1	0.000	1.000			
S2	0.000	0.160	1.000		
S3	0.000	0.200	0.280	1.000	
S4	0.000	0.240	0.320	0.360	1.000

Çizelge 11

Model 2.3 İçin Korelasyon Değerleri

	P	S1	S2	S3	S4
P	1.000				
S1	0.000	1.000			
S2	0.000	0.460	1.000		
S3	0.000	0.500	0.580	1.000	
S4	0.000	0.540	0.620	0.660	1.000

3.2.1.1. Test Uzunluğu 12 Madde İken Madde Parametrelerinin Kestirim Yanlılıkları

Madde parametre kestirimlerindeki iyileşmeyi (recovery) incelemek için, gerçek parametrelerin, kestirilen parametrelerden farkının ortalaması alınarak yanlılık hesaplanmıştır. Çizelge 12’de 12 madde için kestirim yanlılıkları verilmiştir.

Çizelge12’de 12 madde için yanlılık puanlarının betimsel istatistikleri verilmektedir. Güçlük (intercept) katsayıları incelendiğinde, katsayılarının ortalama yanlılıkları ($\bar{X}_{Model21} = 0.245$, $\bar{X}_{Model22} = 0.250$, $\bar{X}_{Model23} = 0.250$), tüm modellerde yüksektir ve diklik ihlal düzeylerine göre fazla değişkenlik göstermemiştir. Benzer şekilde güçlük katsayılarının standart sapmaları

($\sigma_{Model21} = 0.330$, $\sigma_{Model22} = 0.340$, $\sigma_{Model23} = 0.340$) yani yanlılık puanları arasındaki deęişkenlik üç model için de hemen hemen aynıdır. Deęişkenliğin yüksek olması kestirim güvenirliliğini düşürmektedir.

Çizelge 12

Ayırt edicilik ve güçlük katsayılarının 12 madde için kestirim yanlılıkları

		\bar{X}	σ	Min	Q1	Q2	Q3	Max
Model 2.1	d katsayısı	0.245	0.330	-0.322	-0.009	0.209	0.523	0.727
	G	0.060	0.060	-0.030	0.016	0.038	0.098	0.187
	S1	0.020	0.010	0.009	0.014	0.018	0.024	0.030
	S2	-0.012	0.050	-0.050	-0.040	-0.030	0.005	0.040
	S3	-0.070	0.030	-0.090	-0.084	-0.080	-0.055	-0.030
	S4	-0.022	0.090	-0.125	-0.062	0.000	0.030	0.058
Model 2.2	d katsayısı	0.250	0.340	-0.354	0.016	0.233	0.515	0.760
	G	0.125	0.140	-0.090	0.053	0.122	0.225	0.377
	S1	-0.050	0.020	-0.067	-0.057	-0.050	-0.040	-0.030
	S2	-0.107	0.010	-0.123	-0.111	-0.100	-0.010	-0.010
	S3	-0.170	0.050	-0.241	-0.205	-0.170	-0.154	-0.139
	S4	-0.170	0.210	-0.407	-0.240	-0.070	-0.051	-0.035
Model 2.3	d katsayısı	0.250	0.340	-0.366	0.003	0.230	0.514	0.756
	G	0.290	0.260	-0.216	0.190	0.281	0.483	0.662
	S1	-0.220	0.040	-0.255	-0.240	-0.222	-0.195	-0.170
	S2	-0.320	0.060	-0.361	-0.350	-0.340	-0.293	-0.250
	S3	-0.390	0.130	-0.524	0.442	-0.360	-0.320	-0.280
	S4	-0.330	0.270	-0.594	-0.463	-0.332	-0.200	-0.050

Üç modelde de uçdeęerler bulunmaktadır ($Max_{Model22} = 0.760$, $Min_{Model21} = -0.322$). Şekil 10 (Ek-1) incelendiğinde, üç model için de güçlük parametresinin yanlılık puan ortalamasının sıfırın üzerinde olduđu görülmüştür. Yani güçlük parametresi her model için gerçek deęerden daha büyük kestirilmiştir. Ortalama ve standart sapmaların modeller arası deęişkenliğinin az olması nedeniyle diklik varsayımı ihlalinin güçlük parametre kestirimlerinde etkisi olmadığı söylenebilir. Yani spesifik modellerin güçlük parametresinin kestirimine etkisi yoktur denilebilir. Aynı zamanda standart sapmanın yüksek olması nedeniyle parametre kestirim güvenirlilikleri tüm modellerde düşük olduđu söylenebilir. Bu bulgular Model 1 bulguları ile benzerlik göstermektedir.

Kestirilen ayırt edicilik parametreleri ile gerçek parametreler arasındaki ortalama yanlılık, diklik ihlali ile doğru orantılı olarak artmaktadır (örn; $\bar{X}_{G\&Model21} = 0.060$, $\bar{X}_{G\&Model22} = 0.125$, $\bar{X}_{G\&Model23} = 0.290$;

$\bar{X}_{S1\&Model21} = 0.020$, $\bar{X}_{S1\&Model22} = -0.050$, $\bar{X}_{S1\&Model23} = -0.220$). Bu sonuca göre Model 2.2'deki ortalama yanlılık Model 2.1'den, Model 2.3'deki ortalama yanlılığın ise Model 2.2'den yüksek olduğu söylenebilir. Yanlılık puanlarının standart sapmaları incelendiğinde, diklik ihlali arttıkça yanlılık puanlarının değişkenliğinin arttığı görülmektedir. Bu sonuçlar diklik varsayımının ihlalinin ayırt edicilik parametre kestirim yanlılığına etkisi olduğunu göstermektedir. Bu bulgular Model 1 ile benzerlik göstermektedir.

3.2.1.2. Test Uzunluğu 40 Madde İken Madde Parametrelerinin Kestirim Yanlılıkları

Madde parametre kestirimlerindeki iyileşmeyi (recovery) incelemek için, gerçek parametrelerin, kestirilen parametrelerden farkının ortalaması alınarak yanlılık hesaplanmıştır. Çizelge 13'de 40 madde için kestirim yanlılıkları verilmiştir.

Çizelge 13

Ayırt edicilik ve güçlük katsayılarının 40 madde için kestirim yanlılıkları

		\bar{X}	σ	Min	Q1	Q2	Q3	Max
Model 2.1	d katsayısı	0.023	1.780	-3.890	-0.816	0.054	1.290	3.450
	G	-0.050	0.640	-1.190	-0.380	-0.030	0.090	1.310
	S1	-0.020	0.010	-0.030	-0.030	-0.022	-0.012	-0.003
	S2	-0.163	0.560	-1.035	-0.610	-0.007	0.180	0.545
	S3	-0.040	0.880	-1.630	-0.521	-0.274	0.700	1.101
	S4	-0.012	0.870	-1.234	-0.672	-0.040	0.510	1.720
Model 2.2	d katsayısı	0.050	1.780	-3.864	-0.776	0.084	1.321	3.422
	G	0.090	0.650	-1.100	-0.234	0.094	0.300	1.530
	S1	-0.070	0.020	-0.100	-0.080	-0.071	-0.065	-0.024
	S2	-0.230	0.530	-1.080	-0.660	-0.075	0.090	0.430
	S3	-0.130	0.870	-1.763	-0.600	-0.390	0.620	0.970
	S4	-0.140	0.850	-1.275	-0.725	-0.173	0.314	1.630
Model 2.3	d katsayısı	0.040	1.780	-3.920	-0.785	0.080	1.316	3.410
	G	0.190	0.640	-1.040	-0.206	0.215	0.430	1.720
	S1	-0.270	0.090	-0.380	-0.311	-0.250	-0.236	-0.070
	S2	-0.420	0.460	-1.170	-0.776	-0.284	-0.136	0.152
	S3	-0.390	0.820	-2.004	-0.874	-0.650	0.325	0.563
	S4	-0.400	0.840	-1.372	-0.853	-0.540	-0.070	1.450

Çizelge 13'de verilen güçlük (intercept) katsayıları incelendiğinde, 12 maddelik test uzunluğunun aksine 40 maddelik test uzunluğunda, güçlük katsayılarının ortalama yanlılıkları, tüm modellerde düşüktür ve diklik ihlal düzeylerine göre fazla değişkenlik göstermemiştir. Benzer şekilde güçlük

katsayılarının standart sapmaları yani yanlılık puanları arasındaki değişkenlik üç model için de aynı olmakla birlikte 12 maddelik test uzunluğuna göre oldukça yüksektir. Üç modelde de uçdeğerler bulunmaktadır ($Max_{Model21} = 3.450$, $Min_{Model21} = -3.920$). Ek-C'de verilen grafikler incelendiğinde, üç model için de güçlük parametresinin yanlılık puan ortalamasının sıfırın üzerinde olduğu görülmüştür. Yani güçlük parametresi her model için gerçek değerden daha büyük kestirilmiştir. Ortalama ve standart sapmaların modeller arası değişkenliğinin az olması nedeniyle diklik varsayımı ihlalinin güçlük parametre kestirimlerinde etkisi olmadığı söylenebilir. Bunun yanı sıra ranjin geniş, değişkenliğin yüksek olması nedeniyle parametre kestirim güvenilirliğinin de düştüğü görülmüştür. Bu bulgular Model 1.2 bulguları ile benzerlik göstermektedir.

Kestirilen ayırt edicilik parametreleri ile gerçek parametreler arasındaki ortalama yanlılık, 12 maddelik test uzunluğunda olduğu gibi, diklik ihlali arttıkça daha da artmaktadır (örn; $\bar{X}_{G\&Model21} = -0.050$, $\bar{X}_{G\&Model22} = 0.090$, $\bar{X}_{G\&Model23} = 0.190$; $\bar{X}_{S1\&Model21} = -0.020$, $\bar{X}_{S1\&Model22} = -0.070$, $\bar{X}_{S1\&Model23} = -0.270$). Bu sonuca göre Model 2.2'deki ortalama yanlılık Model 2.1'den, Model 2.3'deki ortalama yanlılık ise Model 2.2'den yüksek olduğu söylenebilir. Yanlılık puanlarının standart sapmaları incelendiğinde, yanlılık puanlarının değişkenliği genel olarak modeller arası neredeyse aynıdır. Bu sonuçlar diklik varsayımı ihlalinin sadece ayırt edicilik parametre kestirimlerinin ortalamasına etki ettiği söylenebilir. Değişkenliğin genel olarak tüm faktörlerde yüksek olması ise kestirimlerin güvenilirliğini tüm faktörler için düşürmektedir.

Ek-C'de 40 maddelik test uzunluğu için yanlılık puanlarının grafikleri verilmiştir. Verilen bu grafikler incelendiğinde tüm spesifik faktörlerdeki ayırt edicilik parametrelerinin ortalama yanlılıklarının sıfırın altında olduğu görülmüştür. Bu sonuca göre ayırt edicilik parametrelerinin tüm spesifik faktörler için olduğundan daha düşük kestirildiği söylenebilir.

3.2.1.3. Test Uzunluğu 100 Madde İken Madde Parametrelerinin Kestirim Yanlılıkları

Madde parametre kestirimlerindeki iyileşmeyi (recovery) incelemek için, gerçek parametrelerin, kestirilen parametrelerden farkının ortalaması alınarak

yanlılık hesaplanmıştır. Çizelge 14’de 100 madde için kestirim yanlılıkları verilmiştir.

Çizelge 14

Ayirt edicilik ve güçlük katsayılarının 100 madde için kestirim yanlılıkları

		\bar{X}	σ	Min	Q1	Q2	Q3	Max
Model 2.1	d katsayısı	-0.040	1.830	-4.259	-1.040	-0.005	1.313	4.084
	G	0.034	0.660	-1.645	-0.403	0.005	0.405	1.620
	S1	-0.064	0.610	-1.389	-0.017	-0.003	0.038	1.284
	S2	0.002	0.850	-1.990	-0.604	-0.002	0.717	1.230
	S3	-0.023	0.860	-1.715	0.036	0.036	0.590	1.790
	S4	0.000	0.560	-1.014	-0.382	-0.134	0.453	1.171
Model 2.2	d katsayısı	-0.060	1.830	-4.293	-1.066	-0.020	1.282	4.060
	G	0.090	0.660	-1.544	-0.375	0.023	0.490	1.632
	S1	-0.097	0.610	-1.389	-0.067	-0.028	0.001	1.252
	S2	-0.109	0.820	-1.990	-0.736	-0.072	0.556	1.103
	S3	-0.100	0.860	-1.715	-0.400	-0.050	0.517	1.720
	S4	-0.077	0.580	-1.132	-0.490	-0.190	0.360	1.137
Model 2.3	d katsayısı	-0.060	1.830	-4.263	-1.050	-0.032	1.300	4.053
	G	0.020	0.650	-1.313	-0.223	0.180	0.564	1.691
	S1	-0.270	0.600	-1.417	-0.297	-0.218	-0.095	1.048
	S2	-0.290	0.810	-1.990	-0.975	-0.215	0.328	1.011
	S3	-0.300	0.820	-1.767	-0.615	-0.213	0.240	1.433
	S4	-0.300	0.590	-1.400	-0.732	-0.316	0.120	0.960

Çizelge 14’de verilen güçlük (intercept) katsayıları incelendiğinde, 40 maddelik test uzunluğuna benzer şekilde güçlük katsayılarının ortalama yanlılıkları tüm modellerde düşüktür ve diklik ihlal düzeylerine göre fazla değişkenlik göstermemiştir. Benzer şekilde güçlük katsayılarının standart sapmaları yani yanlılık puanları arasındaki değişkenlik üç model için de aynı ve 40 maddelik test uzunluğuna göre yüksektir. Üç modelde de uçdeğerler bulunmaktadır ($Max_{Model21} = 4.084$, $Min_{Model21} = -4.293$). Ek-D’de verilen grafikler incelendiğinde, üç model için de güçlük parametresinin yanlılık puan ortalamasının sıfırın altında olduğu görülmüştür. Yani güçlük parametresi her model için gerçek değerden daha düşük kestirilmiştir. Ortalama ve standart sapmaların modeller arası değişkenliğinin az olması nedeniyle diklik varsayımı ihlalinin güçlük parametre kestirimlerinde etkisi olmadığı söylenebilir. Yani, spesifik modellerin güçlük parametresinin kestirimine etkisi olmadığı söylenebilir. Bu bulgular Model 1 bulguları ile örtüşmektedir.

Spesifik faktörler için kestirilen ayirt edicilik parametreleri ile gerçek

parametreler arasındaki ortalama yanlılık, 12 ve 40 maddelik test uzunluklarında olduğu gibi, diklik ihlali arttıkça artmaktadır (örn; $\bar{X}_{S1\&Model21} = -0.064$, $\bar{X}_{S1\&Model22} = -0.097$, $\bar{X}_{S1\&Model23} = -0.270$; $\bar{X}_{S2\&Model21} = 0.002$, $\bar{X}_{S2\&Model22} = -0.109$, $\bar{X}_{S2\&Model23} = -0.290$). Bu sonuca göre Model 2.2'deki ortalama yanlılığın Model 2.1'den, Model 2.3'deki ortalama yanlılık ise Model 2.2'den yüksek olduğu söylenebilir Genel faktördeki ayırt edicilik katsayılarının ortalama yanlılıkları 12 ve 40 maddelik test uzunluğuna kıyasla oldukça düşük ve modeller arası neredeyse aynıdır. Tüm faktörler için (genel ve spesifik) yanlılık puanlarının standart sapmaları incelendiğinde, yanlılık puanlarının değişkenliğinin genel olarak modeller arası neredeyse aynı olduğu görülmüştür. Bu sonuçlar diklik varsayımı ihlalinin sadece ayırt edicilik parametre kestirim ortalamalarının kesinliğine (accuracy) etkisi olduğu söylenebilir.

Ek-D'de 100 maddelik test uzunluğuna ait grafikleri verilmiştir. Verilen bu grafikler incelendiğinde 12 ve 40 maddelik test uzunluklarında olduğu gibi tüm spesifik faktörlerdeki ayırt edicilik parametrelerinin ortalama yanlılıklarının sıfırın altında olduğu görülmüştür. Bu sonuca göre ayırt edicilik parametrelerinin tüm spesifik faktörler için daha düşük kestirildiği söylenebilir. Sonuç olarak üç test uzunluğunda (12,40,100) da diklik ihlalinin artması tüm faktörlerdeki (genel ve spesifik) yanlılık ortalamalarını arttırmıştır. Değişkenlik ise Model 1'de olduğu gibi diklik ihlal düzeyinden test uzunluğu ile doğru orantılı şekilde etkilenmiştir. Yani test uzunluğu arttıkça genel olarak tüm faktörlerde değişkenlik artmıştır. Bununla birlikte değişkenlik, diklik ihlal düzeylerine göre neredeyse aynı kalmıştır. Test uzunluğuna göre ranj incelendiğinde, ranjin 12 madde için 1.256 (Min=-0.594, Max=0.662), 40 madde için 3,724 (Min=-2.004, Max=1.720) ve 100 madde için ise 3.780 (Min=-1.990, Max=1.790) olduğu görülmüştür. Test uzunluğu arttıkça ranjin genişlediği görülmüştür. Ranjin geniş, değişkenliğin yüksek olması parametre kestirim güvenilirliğini düşürmektedir. Güçlük katsayıları incelendiğinde ise test uzunluğu 12 den 40'a çıktığında yanlılık ortalamalarında düşüş gözlenmekle birlikte bu durum 40 maddeden 100 maddeye çıktığında aynı kalmamıştır. Ayırt edicilik parametrelerinde gözlenen değişkenlik durumu güçlük parametresi için de gözlenmiştir. Yani test uzunluğu arttıkça değişkenlik artmış fakat bu değişkenlik diklik ihlal düzeyleri (modeller arası) arası aynı kalmıştır. Test uzunluğu arttıkça değişkenliğin artması parametre kestirim güvenilirliğini etkilememektedir. Fakat

değişkenliğin modeller arası neredeyse aynı olması nedeniyle diklik ihlalinin güçlük parametre kestirimleri kesinliğine (accuracy) etkisi olmadığı söylenebilir.

3.2.2. Model 2 İçin Birey Parametrelerinin Kestirim Yanlılıkları

Bu bölümde tüm spesifik faktörler arasındaki diklik ihlali durumunda birey parametrelerindeki yanılıklar verilmektedir. Çizelge 9, 10, 11’de verildiği üzere en düşük diklik ihlal düzeyi $r=0.020$ olarak en yüksek diklik ihlal düzeyi ise $r=0.660$ olarak belirlenmiştir. Kestirilen ve gerçek parametreler arasındaki yanılılık, test uzunluğu ve model bazında tablolştırılmıştır. Yanlılık puanlarının ortalaması, standart sapması ve çeyrekler gibi betimsel istatistikler, farklı koşullarda gerçek parametreler ile kestirilen parametreleri karşılaştırarak, parametreler arasındaki iyileşme (recovery) bakımından incelenmiştir.

3.2.2.1. Test Uzunluğu 12 Madde İken Birey Parametrelerinin Kestirim Yanlılıkları

Çizelge 15 incelendiğinde Model 1’de (üçüncü ve dördüncü spesifik faktörlerin ilişkili olma durumu) olduğu gibi kestirilen parametreler, gerçek parametrelerden oldukça farklıdır çünkü yanılılık puanlarının ranjı ($\min=-3.792$, $\max=3.767$) oldukça geniştir.

Genel faktör ve spesifik faktör için modeller arası ortalamalar çok fazla değişkenlik göstermemekle birlikte sıfır civarındadır. Standart sapma değerleri incelendiğinde tüm modeller için genel faktörün diğer faktörlere kıyasla daha az değişkenlik gösterdiği görülmüştür. Bu sonuca dayanarak genel faktördeki parametre kestirimlerinin diğer faktörlere göre daha güvenilir olduğu söylenebilir. Model 1’de olduğu gibi, Model 2’de de dördüncü faktör, değişkenliğin en yüksek olduğu faktör olmasından dolayı, en yanılı kestirimlerin olduğu faktördür. Ranjın geniş, değişkenliğin yüksek olmasından dolayı parametre kestirim güvenirliliği düşüktür. Bu nedenlerden dolayı spesifik modellerin birey parametrelerinin kestirimine etkisi olmadığı söylenebilir. Ek-E’de birey parametrelerinin yanılılık grafikleri verilmiştir. Bu grafikler incelendiğinde yanılılık puanlarının geniş bir ranja sahip olduğu görülmektedir. Ayrıca yanılılık puanlarının dağılımı simetriktir ve ortalamalar sıfır civarındadır.

Çizelge 15

*Birey parametrelerinin 12 madde için kestirim yanlışlıkları**3.2.2.2. Test Uzunluğu 40 Madde İken Birey Parametrelerinin Kestirim Yanlışlıkları*

Çizelge 16'da 40 maddelik test uzunluğu için parametre kestirim yanlışlıkları verilmiştir. 12 maddelik test uzunluğu ile aynı şekilde yanlışlık puanlarının ranjı

		\bar{X}	σ	Min	Q1	Q2	Q3	Max
Model 2.1	G	-0.010	0.530	-2.247	-0.343	-0.007	0.332	2.257
	S1	0.010	0.760	-2.487	-0.513	0.001	0.498	2.855
	S2	0.020	0.650	-2.714	-0.399	-0.009	0.395	3.044
	S3	0.020	0.650	-3.792	-0.409	0.017	0.415	3.440
	S4	-0.020	0.800	-2.958	-0.555	-0.032	0.372	2.410
Model 2.2	G	0.000	0.570	-2.136	-0.373	0.010	0.326	2.280
	S1	0.001	0.780	-3.229	-0.532	-0.009	0.512	4.044
	S2	-0.020	0.680	-2.633	-0.454	-0.041	0.384	2.755
	S3	-0.010	0.690	-2.797	-0.457	-0.026	0.431	3.417
	S4	-0.010	0.840	-3.006	-0.552	-0.050	0.529	3.232
Model 2.3	G	-0.001	0.650	-2.300	-0.436	-0.021	0.429	2.636
	S1	0.000	0.840	-3.048	-0.577	-0.005	0.556	3.047
	S2	-0.001	0.780	-3.055	-0.492	-0.031	0.469	3.213
	S3	-0.002	0.800	-3.630	-0.523	0.007	0.515	3.451
	S4	-0.004	0.910	-3.447	-0.618	-0.010	0.610	3.767

oldukça geniştir (Min=-2.850, Max= 3.206).

Çizelge 16'da verilen yanlışlık puanlarının ortalamaları incelendiğinde 12 maddelik test uzunluğunda olduğu gibi ortalamaların sıfır civarında ve dağılım simetriktir. Standart sapma değerleri incelendiğinde tüm modeller için genel faktördeki yanlışlık puanlarının diğer faktörlere kıyasla daha az değişkenlik gösterdiği görülmüştür. Bu sonuca dayanarak genel faktördeki parametre kestirimlerinin diğer faktörlere göre daha güvenilir olduğu söylenebilir. Model 1'de olduğu gibi, Model 2'de de üçüncü faktör en yanlış kestirimlerin olduğu faktördür çünkü bu faktör değişkenliğin en yüksek olduğu faktördür.

Çizelge 16

Birey parametrelerinin 40 madde için kestirim yanlılıkları

		\bar{X}	σ	Min	Q1	Q2	Q3	Max
Model 2.1	G	0.020	0.280	-1.182	-0.157	0.013	0.191	1.615
	S1	0.040	0.460	-2.484	-0.227	0.028	0.295	2.679
	S2	-0.010	0.400	-2.054	-0.237	-0.017	0.232	1.851
	S3	-0.020	0.450	-2.040	-0.271	0.009	0.258	2.851
	S4	0.030	0.420	-2.621	-0.178	0.034	0.250	2.451
Model 2.2	G	0.013	0.270	-1.236	-0.155	0.009	0.182	1.224
	S1	0.001	0.340	-1.827	-0.185	-0.013	0.170	2.330
	S2	0.001	0.330	-2.597	-0.163	0.014	0.176	1.769
	S3	0.002	0.360	-1.952	-0.175	0.012	0.196	2.009
	S4	0.030	0.350	-2.318	-0.168	0.033	0.224	2.626
Model 2.3	G	-0.011	0.280	-1.714	-0.266	-0.019	0.250	1.517
	S1	0.014	0.360	-2.417	-0.216	-0.003	0.225	2.371
	S2	0.020	0.400	-2.226	-0.192	0.031	0.243	2.404
	S3	0.030	0.390	-3.290	-0.199	0.039	0.266	2.596
	S4	0.030	0.350	-2.318	-0.168	0.033	0.224	2.626

Ek-F’de birey parametrelerinin yanlılık grafikleri verilmiştir. Bu grafikler incelendiğinde yanlılık puanlarının geniş bir ranja sahip olduğu görülmektedir. Ayrıca yanlılık puanlarının dağılımı simetriktir ve sıfır civarındadır.

3.2.2.3. Test Uzunluğu 100 Madde İken Birey Parametrelerinin Kestirim Yanlılıkları

Çizelge 17’de 100 maddelik test uzunluğu için parametre kestirim yanlılıkları verilmiştir. 12 ve 40 maddelik test uzunluğu ile aynı şekilde yanlılık puanlarının ranjı oldukça geniştir (Min=-3.290, Max= 2.626).

Çizelge 17’de verilen yanlılık puanlarının ortalamaları incelendiğinde, yanlılık puanlarının ortalamaları, 12 ve 40 maddelik test uzunluğunda olduğu gibi sıfır civarında ve dağılım simetriktir. Bu bulgu Model 1 ile uyum göstermektedir. Standart sapma değerleri incelediğinde tüm modeller için genel faktördeki yanlılık puanlarının, diğer faktörlere kıyasla daha az değişkenlik gösterdiği görülmüştür. Bu sonuca dayanarak genel faktördeki parametre kestirimlerinin diğer faktörlere göre daha güvenilir olduğu söylenebilir. Model 1 ve Model 2 karşılaştırıldığında standart sapma değerlerinin oldukça benzer olduğu görülmüştür. Hem Model 1 hem Model 2

için test uzunluğu arttıkça değişkenliğin düştüğü görülmüştür. Test uzunluğunun parametre kestirimlerindeki iyileşmede (recovery) etkisi olduğu söylenebilir.

Çizelge 17

Birey parametrelerinin 100 madde için kestirim yanlılıkları

		\bar{X}	σ	Min	Q1	Q2	Q3	Max
Model 2.1	G	-0.010	0.200	-1.087	-0.132	-0.008	0.117	0.903
	S1	-0.010	0.290	-1.829	-0.171	-0.025	0.129	1.832
	S2	0.000	0.310	-2.232	-0.153	0.019	0.177	1.812
	S3	0.000	0.320	-2.501	-0.159	0.010	0.173	1.907
	S4	0.002	0.280	-2.551	-0.130	0.012	0.137	1.524
Model 2.2	G	0.013	0.270	-1.236	-0.155	0.009	0.182	1.224
	S1	0.001	0.340	-1.827	-0.185	-0.013	0.170	2.330
	S2	0.001	0.330	-2.597	-0.163	0.014	0.176	1.769
	S3	0.002	0.360	-1.952	-0.175	0.012	0.196	2.009
	S4	-0.004	0.310	-2.044	-0.142	0.005	0.147	2.155
Model 2.3	G	-0.011	0.280	-1.714	-0.266	-0.019	0.250	1.517
	S1	0.014	0.360	-2.417	-0.216	-0.003	0.225	2.371
	S2	0.020	0.400	-2.226	-0.192	0.031	0.243	2.404
	S3	0.030	0.390	-3.290	-0.199	0.039	0.266	2.596
	S4	0.030	0.350	-2.318	-0.168	0.033	0.224	2.626

Model 1 de olduğu gibi Model 2’de de, yanlılık puanlarının ranjı 100 madde için 40 ve 12 maddeden daha dardır ($Q1_{12} = -0.618$, $Q3_{12} = 0.610$; $Q1_{40} = -0.369$, $Q3_{40} = 0.361$; $Q1_{100} = -0.266$, $Q3_{100} = 0.250$). Test uzunluğunun birey parametre kestirim kesinliğine etkisi olduğu söylenebilir. Ek-G’de birey parametrelerinin yanlılık grafikleri verilmiştir. Bu grafikler incelendiğinde yanlılık puanlarının geniş bir ranja sahip olduğu görülmektedir. Ayrıca yanlılık puanlarının dağılımı simetriktir ve sıfır civarındadır.

3.3. Model türüne ve test uzunluğuna göre, madde ve birey parametrelerinin yanlılık puanları nasıldır?

Bu bölümde parametre kestirim yanlılıklarına model türünün ve test uzunluğunun etkisi incelenmiştir. Amaç iki faktördeki diklik ihlalinden (Model 1) elde edilen parametrelerin, tüm faktörlerdeki diklik ihlalinden (Model 2) elde edilen parametrelerden farklılık gösterip göstermediğidir. Standart hata ortalamaları ve ortalama RMSE değerleri madde ve birey parametreleri için tablolaştırılmıştır.

3.3.1. Ayırt Edicilik Parametrelerinin Model Türüne ve Test Uzunluğuna Göre

Standart Hata ve RMSE Değerleri

Çizelge 18’de Model 1 ve Model 2 için ayırt edicilik parametrelerinin standart hata değerleri ve ortalama RMSE değerleri verilmiştir. Bu değerler önce model bazında daha sonra ise test uzunluğuna göre yorumlanmıştır.

Çizelge 18

Model 1 ve Model 2 için Standart Hata ve ortalama RMSE değerleri

	Model	Test uzunluğu		
		12 madde	40 madde	100 madde
SH	Model 1.1	0.120(0.090)	0.060(0.020)	0.050(0.010)
	Model 1.2	0.140(0.120)	0.060(0.020)	0.050(0.010)
	Model 1.3	0.160(0.150)	0.060(0.020)	0.050(0.010)
	Model 2.1	0.130(0.110)	0.060(0.020)	0.050(0.010)
	Model 2.2	0.140(0.130)	0.060(0.020)	0.050(0.010)
	Model 2.3	0.180(0.160)	0.060(0.020)	0.050(0.010)
RMSE	Model 1.1	0.140(0.120)	0.480(0.440)	0.540(0.430)
	Model 1.2	0.180(0.120)	0.480(0.440)	0.540(0.430)
	Model 1.3	0.290(0.230)	0.490(0.450)	0.550(0.430)
	Model 2.1	0.150(0.110)	0.480(0.440)	0.540(0.430)
	Model 2.2	0.220(0.140)	0.490(0.430)	0.540(0.440)
	Model 2.3	0.390(0.180)	0.540(0.430)	0.570(0.440)

Model bazında standart hatalar incelendiğinde tablo değerlerinin (ortalama ve standart sapma) her iki model için de (12 maddelik test uzunluğu dışında) aynı olduğu görülmüştür ($\bar{X}_{SE} \& Model 1 = 0.083$, $\bar{X}_{SE} \& Model 2 = 0.086$). RMSE ortalamaları incelendiğinde ise Model 1 ve Model 2 için tablo değerlerinin neredeyse aynı ($\bar{X}_{RMSE} \& Model 1 = 0.410$, $\bar{X}_{RMSE} \& Model 2 = 0.435$) olduğu görülmüştür. Bu bulgulara göre; iki faktörün ilişkili olması durumu ile tüm faktörlerin ilişkili olması durumunun, ayırt edicilik parametrelerinin kestiriminde neredeyse aynı etkiye sahip olduğu söylenebilir. Sonuç olarak her iki model (Model 1 ve Model 2) için de parametre kestirim doğruluğu arasında farklılık yoktur yani model türünün parametre kestirim doğruluğuna etkisi olmadığı söylenebilir.

Modeller kendi içlerinde incelendiklerinde ise (diklik ihlal düzeylerine göre), Model 1.1., Model 1.2., Model 1.3’ün standart hata ortalamalarının çok fazla

farklılaşmadığı görülmüştür ($\bar{X}_{SE \& Model 1.1} = 0.076$, $\bar{X}_{SE \& Model 1.2} = 0.083$, $\bar{X}_{SE \& Model 1.3} = 0.090$). RMSE ortalamaları incelendiğinde, ($\bar{X}_{RMSE \& Model 1.1} = 0.386$, $\bar{X}_{RMSE \& Model 1.2} = 0.400$, $\bar{X}_{RMSE \& Model 1.3} = 0.443$), diklik ihlal düzeyi artıkça parametre kestirim doğruluğunun düştüğü görülmüştür. Model 2 alt modeller bazında incelendiğinde, Model 2.1., Model 2.2., Model 2.3'ün standart hata ortalamalarının çok fazla farklılaşmadığı görülmüştür ($\bar{X}_{SE \& Model 2.1} = 0.080$, $\bar{X}_{SE \& Model 2.2} = 0.083$, $\bar{X}_{SE \& Model 2.3} = 0.096$). RMSE ortalamaları incelendiğinde ($\bar{X}_{RMSE \& Model 2.1} = 0.390$, $\bar{X}_{RMSE \& Model 2.2} = 0.416$, $\bar{X}_{RMSE \& Model 2.3} = 0.500$), diklik ihlal düzeyi artıkça parametre kestirim doğruluğunun düştüğü görülmüştür. Diklik ihlal düzeyindeki artış, ayırt edicilik parametrelerinin kestirim doğruluğunu her iki model için de düşürmüştür. Gözlenen bu durum Zheng (2013)'ün bulguları ile örtüşmektedir.

Test uzunluğuna göre RMSE değerleri incelendiğinde ise, test uzunluğunun parametre kestirim doğruluğuna etkisi olabileceği düşünülmüştür. Test uzunluğu arttıkça ortalama RMSE değerleri artmıştır ($\bar{X}_{RMSE \& 12} = 0.228$, $\bar{X}_{RMSE \& 40} = 0.493$, $\bar{X}_{RMSE \& 100} = 0.546$). Yani madde sayısı arttıkça, ayırt edicilik parametrelerinin kestirim doğruluğu düşmektedir. Standart hatalar incelendiğinde ise test uzunluğu artıkça standart hatanın düştüğü gözlenmiştir ($\bar{X}_{SE \& 12} = 0.145$, $\bar{X}_{SE \& 40} = 0.060$, $\bar{X}_{SE \& 100} = 0.050$). Standart hata, simülasyon örneklerinin standart sapmasıdır yani bir nevi uzaklık ölçüsüdür. Bu neden dolayı standart hata aslında kesinliğin (precision) bir ölçüsüdür (Walther ve Moore, 2005). Bu durumda test uzunluğu arttıkça ayırt edicilik parametre kestirimlerinin daha güvenilir olduğu yani test uzunluğunun parametre kestirim kesinliğine etkisi olduğu söylenebilir.

3.3.2. Güçlük Parametrelerinin Model Türüne ve Test Uzunluğuna Göre Standart Hata ve RMSE Değerleri

Çizelge 19'da sırasıyla güçlük parametrelerinin Model 1 ve Model 2 için Standart hata ortalamaları ve ortalama RMSE değerleri verilmiştir. Modellerin standart hataları incelendiğinde tablo değerlerinde (ortalama ve standart sapma) çok

fazla farklılık olmadığı görülmüştür ($\bar{X}_{SE \& Model 1} = 0.046$, $\bar{X}_{SE \& Model 2} = 0.047$).

Çizelge 19’da verilen RMSE ortalamaları incelendiğinde ise, standart hata için görülen durum burada da geçerlidir. Model 1 ve Model 2 için tablo değerleri aynıdır ($\bar{X}_{RMSE \& Model 1} = 1.056$, $\bar{X}_{RMSE \& Model 2} = 1.056$). Bu bulgulara göre güçlük parametresinin kestiriminde, iki spesifik faktörün ilişkili olma durumu ile tüm spesifik faktörlerin ilişkili olma durumunun neredeyse aynı etkiye sahip olduğu söylenebilir. Yani model türünün güçlük parametre kestirim doğruluğuna bir etkisi olmadığı söylenebilir.

Çizelge 19

Model 1 ve Model 2 için standart hata ve ortalama RMSE değerleri

		Test uzunluğu		
Model		12 madde	40 madde	100 madde
SH	Model 1.1	0.050(0.030)	0.040(0.010)	0.040(0.010)
	Model 1.2	0.060(0.040)	0.040(0.010)	0.040(0.010)
	Model 1.3	0.070(0.050)	0.040(0.010)	0.040(0.010)
	Model 2.1	0.060(0.040)	0.040(0.010)	0.040(0.010)
	Model 2.2	0.060(0.050)	0.040(0.010)	0.040(0.010)
	Model 2.3	0.070(0.050)	0.040(0.010)	0.040(0.010)
		Model	12 madde	40 madde
RMSE	Model 1.1	0.330(0.260)	1.400(1.060)	1.430(1.130)
	Model 1.2	0.330(0.260)	1.420(1.060)	1.430(1.140)
	Model 1.3	0.330(0.270)	1.410(1.060)	1.430(1.140)
	Model 2.1	0.330(0.250)	1.400(1.070)	1.430(1.130)
	Model 2.2	0.340(0.260)	1.410(1.060)	1.430(1.130)
	Model 2.3	0.330(0.260)	1.410(1.060)	1.430(1.140)

Modeller kendi içlerinde incelendiklerinde ise (diklik ihlal düzeylerine göre), Model 1.1., Model 1.2., Model 1.3’ün standart hata ortalamalarının çok fazla farklılaşmadığı görülmüştür ($\bar{X}_{SE \& Model 1.1} = 0.043$, $\bar{X}_{SE \& Model 1.2} = 0.046$, $\bar{X}_{SE \& Model 1.3} = 0.050$). Aynı durum RMSE ortalamaları için de görülmüştür ($\bar{X}_{RMSE \& Model 1.1} = 1.053$, $\bar{X}_{RMSE \& Model 1.2} = 1.060$, $\bar{X}_{RMSE \& Model 1.3} = 1.056$). Model 2 alt modeller bazında incelendiğinde, Model 2.1., Model 2.2., Model 2.3’ün standart hata ortalamalarının çok fazla farklılaşmadığı görülmüştür (\bar{X}_{SE}

$\bar{X}_{SE \& Model 2.1} = 0.046$, $\bar{X}_{SE \& Model 2.2} = 0.046$, $\bar{X}_{SE \& Model 2.3} = 0.050$). Aynı durum RMSE ortalamaları için de görülmüştür ($\bar{X}_{RMSE \& Model 2.1} = 1.053$, $\bar{X}_{RMSE \& Model 2.2} = 1.060$, $\bar{X}_{RMSE \& Model 2.3} = 1.056$). Güçlük parametrelerinin kestirim doğruluğuna, diklik ihlal düzeyinin etkisi olmamıştır. Bu durum Zheng (2013)'ün bulguları ile örtüşmektedir.

Test uzunluğuna göre RMSE değerleri incelendiğinde, test uzunluğu 12'den 40 a çıktığında kestirim doğruluğunun düştüğü fakat 40 maddeden 100 maddeye çıktığında bu durumun çok fazla farklılaşmadığı gözlenmiştir ($\bar{X}_{RMSE \& 12} = 0.616$, $\bar{X}_{RMSE \& 40} = 1.408$, $\bar{X}_{RMSE \& 100} = 1.430$). Zheng (2013)'ün araştırmasının aksine bu araştırma kapsamında kestirilen parametreler arasında kestirim doğruluğu, hem test uzunluğuna hem de model türüne göre en düşük parametreler, güçlük parametreleri olmuştur. Jennrich ve Bentler (2012)'nin da araştırmalarında belirttiği üzere faktörler arası ilişkiye izin verildiğinde yorulmayan sonuçlar ortaya çıkmıştır. Test uzunluğu arttıkça standart hata değerleri de artmıştır ($\bar{X}_{SE \& 12} = 0.310$, $\bar{X}_{SE \& 40} = 0.493$, $\bar{X}_{SE \& 100} = 0.546$). Standart hatadaki bu artış, test uzunluğu arttıkça kestirim kesinliğinin düştüğünü göstermektedir.

3.3.3. Birey Parametrelerinin Model Türüne ve Test Uzunluğuna Göre Ortalama Standart Hata ve RMSE Değerleri

Çizelge 20'de sırasıyla birey parametrelerinin Model 1 ve Model 2 için standart hata ve RMSE ortalamaları verilmiştir. Modellerin standart hata değerleri incelendiğinde tablo değerlerinin (ortalama ve standart sapmalar) benzerlik gösterdiği görülmüştür ($\bar{X}_{SE \& Model 1} = 0.464$, $\bar{X}_{SE \& Model 2} = 0.400$). RMSE ortalamaları incelendiğinde ise, standart hata için görülen durum burada da ortaya çıkmıştır. Model 1 ve Model 2 için tablo değerleri neredeyse aynıdır ($\bar{X}_{RMSE \& Model 1} = 0.660$, $\bar{X}_{RMSE \& Model 2} = 0.615$). Sonuçlar birey parametrelerinin kestiriminde, iki spesifik faktörün ilişkili olma durumu ile tüm spesifik faktörlerin ilişkili olma durumunun neredeyse aynı etkiye sahip olduğu söylenebilir.

Modeller kendi içlerinde incelendiklerinde ise (diklik ihlal düzeylerine göre),

Model 1.1., Model 1.2., Model 1.3'ün standart hata ortalamalarının çok fazla farklılaşmadığı görülmüştür ($\bar{X}_{SE \& Model 1.1} = 0.477$, $\bar{X}_{SE \& Model 1.2} = 0.477$, $\bar{X}_{SE \& Model 1.3} = 0.473$). Aynı durum RMSE ortalamaları için de görülmüştür ($\bar{X}_{RMSE \& Model 1.1} = 0.650$, $\bar{X}_{RMSE \& Model 1.2} = 0.660$, $\bar{X}_{RMSE \& Model 1.3} = 0.670$). Model 2 alt modeller bazında incelendiğinde, Model 2.1., Model 2.2., Model 2.3'ün standart hata ortalamalarının çok fazla farklılaşmadığı görülmüştür ($\bar{X}_{SE \& Model 2.1} = 0.400$, $\bar{X}_{SE \& Model 2.2} = 0.400$, $\bar{X}_{SE \& Model 2.3} = 0.397$). Aynı durum RMSE ortalamaları için de görülmüştür ($\bar{X}_{RMSE \& Model 2.1} = 0.583$, $\bar{X}_{RMSE \& Model 2.2} = 0.603$, $\bar{X}_{RMSE \& Model 2.3} = 0.660$).

Çizelge 20

Model 1 ve Model 2 için standart hata ve ortalama RMSE değerleri

		Test uzunluğu		
	Model	12 madde	40 madde	100 madde
SH	Model 1.1	0.450(0.080)	0.490(0.130)	0.490(0.170)
	Model 1.2	0.450(0.080)	0.480(0.130)	0.500(0.170)
	Model 1.3	0.450(0.080)	0.480(0.130)	0.490(0.170)
	Model 2.1	0.420(0.080)	0.420(0.080)	0.360(0.070)
	Model 2.2	0.410(0.070)	0.420(0.090)	0.370(0.070)
	Model 2.3	0.380(0.070)	0.420(0.100)	0.390(0.100)
		Model	12 madde	40 madde
RMSE	Model 1.1	0.750(0.320)	0.620(0.210)	0.580(0.200)
	Model 1.2	0.760(0.340)	0.630(0.220)	0.590(0.200)
	Model 1.3	0.770(0.350)	0.640(0.230)	0.600(0.210)
	Model 2.1	0.740(0.340)	0.560(0.200)	0.450(0.140)
	Model 2.2	0.750(0.360)	0.590(0.220)	0.470(0.160)
	Model 2.3	0.790(0.410)	0.650(0.270)	0.540(0.210)

Test uzunluğuna göre RMSE değerleri incelendiğinde, test uzunluğunun parametre kestirim doğruluğuna etkisi olabileceği düşünülmüştür. Test uzunluğu arttıkça ortalama RMSE değerleri düşmüştür ($\bar{X}_{RMSE \& 12} = 0.760$, $\bar{X}_{RMSE \& 40} = 0.615$, $\bar{X}_{RMSE \& 100} = 0.540$). Madde sayısını arttırmak birey parametrelerinin kestirim doğruluğunu arttırmıştır. Madde sayısını arttırmak bir nevi örtük değişkenin (birey parametreleri/theta) açıklanan varyansını arttırdığı için kestirim doğruluğunda

artış görülmüştür. Standart hatalar incelendiğinde ise test uzunluğuna göre standart hatanın çok fazla değişkenlik göstermediği görülmüştür ($\bar{X}_{SE \&12} = 0.426$, $\bar{X}_{SE \&40} = 0.431$, $\bar{X}_{SE \&100} = 0.433$). Test uzunluğunun birey parametre kestirim kesinliğine yani güvenilirliğine etkisi olmadığı söylenebilir. Bununla birlikte bu araştırma kapsamında, kestirim güvenilirliği en düşük parametrelerin birey parametreleri olduğu görülmüştür. Bu bulgular Zheng (2013)'ün çalışması ile uyum göstermektedir.



BÖLÜM 4

SONUÇLAR VE ÖNERİLER

Bu bölümde araştırmadan elde edilen sonuç ve tartışmalara yer verilmiş ve araştırma sonuçlarına uygun olarak öneriler sunulmuştur.

4.1. Sonuçlar

Bu araştırmada, iki faktör (Bifactor) modelde farklı düzeylerdeki diklik ihlali varsayımının parametre (madde ve birey) kestirimlerine etkisinin incelenmesi amaçlanmıştır.

Bu amaç doğrultusunda elde edilen sonuçlar aşağıda özetlenmiştir;

1. Ayırt edicilik parametrelerinin kestirim yanlılığı, iki spesifik faktörün ilişkili olma durumunda (Model 1), diklik ihlali arttıkça artış göstermiştir.
2. Ayırt edicilik, güçlük ve birey parametrelerinin kestirim doğruluğunda modelin (Model1 ve Model 2) önemi yoktur. Yani iki spesifik faktörün ilişkili olma durumu ile tüm faktörlerin ilişkili olma durumu hem birey hem de madde parametre kestirim doğruluğunda aynı etkiye sahiptir.
3. Her iki modelde de (Model1 ve Model 2) test uzunluğu arttıkça, ayırt edicilik ve güçlük parametrelerinin kestirim kesinliği yani güvenilirliği düşmüştür. Bu durum ilişkili madde sayısının artması ile açıklanabilir.
4. Ayırt edicilik parametrelerinin kestiriminde, iki spesifik faktörün ilişkili olma durumunda (Model 1), test uzunluğuna bağlı olarak parametre kestiriminde iyileşme görülürken, tüm spesifik faktörlerin ilişkili olması durumunda (Model 2) bu iyileşme görülmemiştir.
5. Güçlük parametreleri, kestirim doğruluğu tüm modellerde (Model1 ve Model 2) en düşük parametrelerdir.
6. Madde sayısını arttırmak, hem Model1 hem de Model 2 için birey parametrelerinin kestirim kesinliğini yani güvenilirliğini arttırmıştır. Birey

parametrelerinde gözlenen bu durum, madde sayısı arttıkça bireyin örtük özelliğinin daha iyi açıklandığının bir sonucudur.

7. Birey parametrelerinin kestiriminde, güvenilirliği en düşük parametre kestirimleri her iki model için de (Model 1ve Model 2) en küçük test uzunluğundadır. Test uzunluğu arttıkça kestirim güvenilirliği de artmıştır. Buna rağmen tüm test uzunluklarında ve diklik ihlal düzeylerinde kestirim güvenilirliği en düşük parametreler birey parametreleridir.
8. Literatürde varolan araştırmalar iki faktör kuramının ilişkili yapılarda bile çok iyi düzeyde uyum verdiği ve robust bir model olduğu belirtmektedir. Bu araştırma kapsamında ise parametre bazında yanlılık incelendiğinde, belirtilen bu robust yapı görülememiştir.



4.2. Öneriler

Bu araştırmanın sonuçlarına dayanarak iki faktör (bifactor) madde tepki kuramının uygulamalarında uygulayıcılar veya araştırmacılar için aşağıdaki öneriler sunulabilir:

1. İki faktör (bifactor) kuramı, birey parametrelerinin kestiriminde test uzunluğu arttıkça diklik varsayımı ihlalini daha iyi tolere edebilmektedir. Bu kuramı kullanmak isteyen uygulayıcılar 40 ve üstü madde sayısı ile çalışmalıdırlar.
2. Çok boyutlu MTK modelleri için literatürde en az 20 madde olması gerektiği belirtilmektedir. Bu araştırmada en küçük test uzunluğu 12 madde olarak belirtilmiştir. Madde parametrelerinin kestiriminde daha yansız sonuçlar elde etmek için, ileriki araştırmalarda en küçük test uzunluğu 20 madde olarak belirlenerek aynı koşullar tekrar denenebilir.
3. Tüm parametreler arasında kestirim güvenilirliği en düşük parametreler (SH ortalamaları en yüksek) birey parametreleri bulunmuştur. İleriki araştırmalar, güvenilirliği arttırmak amacıyla farklı replikasyon sayıları ve farklı örneklem büyüklükleri ile denenebilir.
4. 12 ile 40 madde arasındaki yanlılık ortalamalarındaki artış/düşüş 40 madde ile 100 madde arasındakine göre daha keskindir. Bu nedenden dolayı ileriki araştırmalarda ara madde sayıları ile çalışılabilir.
5. Bu araştırma kapsamında Bayesci yaklaşım (Bayesian approach)'ın EAP yöntemi ile parametre kestirimleri yapılmıştır. İleriki araştırmalar, farklı parametre kestirim yöntemleri denenerek yapılabilir.
6. Bu bir simülasyon çalışmasıdır ve belirlenen koşullar için geçerlidir. İleriki araştırmalarda, farklı test uzunlukları ve korelasyon düzeyine odaklanıla bilir.
7. İleriki araştırmalarda, çoklu puanlanan maddeler ile diklik varsayımı aynı koşullar için denenebilir.

KAYNAKÇA

- Ackerman T.A. (1994). Using Multidimensional Item Response Theory to Understand What Items and Tests are Measuring. *Applied Measurement In Education*, 7 (4), 255–278.
- Baker, Frank, B. (1985). *The Basics of Item Response Theory*. University of Maryland, College Park, MD: Heinemann and ERIC Clearinghouse on Assessment and Evaluation, second, revised and updated (2001) edition.
- Baker, Frank, B. & Kim, S. (2004). *Item response theory: parameter estimation techniques*. New York: Marcel Dekker, Inc, second edition, revised and expanded edition.
- Baykul, Y. (2000). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: ÖSYM Yayınları.
- Boomsma, A. (2013) Reporting Monte Carlo Studies in Structural Equation Modeling, *Structural Equation Modeling. A Multidisciplinary Journal*, 20(3), 518-540.
- Brouwer, D., Meijer, R. R., Weekers, A. M., & Baneke, J. J. (2008). On the dimensionality of the Dispositional Hope Scale. *Psychological assessment*, 20(3), 310.
- Cai, L., Yang, J. S., & Hansen, M. (2011). Generalized full-information item bifactor analysis. *Psychological Methods*, 16(3), 221–248. [11]
[SEP]
- Carmines, E. G., & Zeller, R. A. (1979). *Reliability and validity assessment*. Beverly Hills, CA: Sage Publications.
- Carver, C. S. (1989). How should multifaceted personality constructs be tested? Issues illustrated by self-monitoring, attributional style, and hardiness. *Journal of personality and social psychology*, 56(4), 577.
- Chalmers, P. (2016). Mirt: Multidimensional Item Response Theory. *R package version 1.19*, URL: <https://cran.r-project.org/web/packages/mirt/index.html>
- Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J. P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: A comparison of the bifactor model to other approaches. *Journal of personality*, 80(1), 219-251.

- Chen, F.F., West, S. G., Sousa, K. H., (2006). A Comparasion of Bifactor and Second Order Models of Quality of Life. *Multivariate Behavioral Research*. 41:2. 189-225
- Cohen, J. (1988). *Statistical power analysis for the behavioural sciences*. Hillside. NJ: Lawrence Earlbaum Associates.
- Crocker, L., and Algina, J. (1986). *Introduction classical and modern test theory*. New York: Harcourt Brace Javonovich College Publishers.
- Cucina, J., & Byle, K. (2017). The Bifactor Model Fits Better Than the Higher-Order Model in More Than 90% of Comparisons for Mental Abilities Test Batteries. *Journal of Intelligence*, 5(3), 27.
- De Ayala, R.J. (1993). An Introduction to Polytomous Item Response Theory Models. *Measurement and Evaluation in Counseling and Development*, 3, 172-189.
- DeMars, C. E. (2006). Application of the Bi-Factor multidimensional item response theory model to Testlet-Based tests. *Journal of educational measurement*, 43(2), 145–168.
- DeMars, C. E. (2013). A tutorial on interpreting bifactor model scores. *International Journal of Testing*, 13, 354–378. <http://dx.doi.org/10.1080/15305058.2013.799067>.
- Desa, Z. N. D. M. (2012). *Bi-factor Multidimensional Item Response Theory Modeling for Subscores Estimation, Reliability, and Classification*. (Doctoral Dissertation). Retrieved from <http://hbanaszak.mjr.uw.edu.pl/>.
- Doğan, N. (2002). Klasik test kuramı ve örtük özellikler kuramının örneklemeler bağlamında karşılaştırılması, Yayınlanmamış doktora tezi, Hacettepe Üniversitesi, Ankara.
- Drasgow, F., & Parsons, C. (1983). Applications of unidimensional item response theory models to multidimensional data. *Applied Psychological Measurement*, 7, 189-199.
- Ebesutani, C., Smith, A., Bernstein, A., Chorpita, B. F., Higa-McMillan, C., ve Nakamura, B. (2011). A bifactor model of negative affectivity: Fear and distress components among younger and older youth. *Psychological Assessment*, 23(3), 679–691

- Edwards, M. C., & Edelen, M. O. (2009). Special topics in item response theory. In R. E. Millsap & A. Maydeu-Olivares (Eds.), *Handbook of quantitative methods in psychology* (pp. 178–198). New York, NY : Sage.
- Embretson, S.E. (1999). Issue in the measurement of cognitive abilities. Embretson, S.E., & Hershberger, S.L. (Eds.). *The new rules of measurement: What every psychologist and educator should know*. New Jersey: Lawrence Erlbaum Associates.
- Embretson, S.E. and Reise, S.P. (2000). *Item Response Theory for Psychologists*. New Jersey: Lawrence Erlbaum Associates
- Erkuş, A., Sünbül, Ö., Sünbül, S. Ö., Yormaz, S., & Aşiret, S. (2017). Psikolojide ölçme ve ölçek geliştirme-II ölçme araçlarının psikometrik nitelikleri ve ölçme kuramları. *PegemA*.
- Fox J, Weisberg, S. (2016). Car: Companion to Applied Regression. . *R package version 2.1-4*, URL: <https://cran.r-project.org/web/packages/car/index.html>
- Garn, A. C. (2017). Multidimensional measurement of situational interest in physical education: Application of bifactor exploratory structural equation modeling. *Journal of Teaching in Physical Education*, 36(3), 323-339.
- Gibbons, R. D., & Hedeker, D. (1992). Full-information item bi-factor analysis. *Psychometrika*, 57, 423–436.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Golay, P., & Lecerf, T. (2011). Orthogonal higher order structure and confirmatory factor analysis of the French Wechsler Adult Intelligence Scale (WAIS-III). *Psychological assessment*, 23(1), 143.
- Hambleton, R. K. ve Swaminathan, H. (1985). *Item response theory: principles and applications*. Boston: Academic Publishers Group.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.
- Hambleton, R. K., and Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement*, 12, 38-47.
- Harrison, D. (1986). Robustness of IRT parameter estimation to violations of the unidimensionality assumption. *Journal of Educational Statistics*, 11, 91-115.

- Hattie, J. A. (1985). Methodology review: Assessing unidimensionality of tests and items. *Applied Psychological Measurement*, 9, 139–164.
- Harwell, M., Stone, C. A., Hsu, T. and Kirisci, L. (1996). Monte carlo studies in item response theory. *Applied Psychological Measurement*, 20, 2,101- 125.
- Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, 2(1), 41-54.
- Hull, J. G., Lehn, D. A., & Tedlie, J. C. (1991). A general approach to testing multifaceted personality constructs. *Journal of Personality and Social Psychology*, 61, 932–945.
- Hyland, P., Boduszek, D., Dhingra, K., Shevlin, M., & Egan, A. (2014). A bifactor approach to modelling the Rosenberg Self Esteem Scale. *Personality and Individual Differences*, 66, 188-192.
- Immekus, J. C., & Imbrie, P. K. (2008). Dimensionality assessment using the full-information item bifactor analysis for graded response data: An illustration with the state metacognitive inventory. *Educational and Psychological Measurement*, 18, 695–709. <http://dx.doi.org/10.1177/0013164407313366>.
- Jang, H., (2014). Exploring The Estimation Of Examinee Locations Using Multidimensional Latent Trait Models Under Different Distributional Assumptions. (Doktora Tezi) (UMI: 36119231).
- Jennrich, R. I., & Bentler, P. M. (2012). Exploratory bi-factor analysis: The oblique case. *Psychometrika*. doi: 10.1007/s1136-012-9269-1.
- Karasar, N. (2010). *Bilimsel araştırma yöntemi*. Ankara. Nobel Yayın Dağıtım.
- Kavcar, B. (2004). Simülasyon Yöntemi Kullanılarak Yapılan Satış Tahminleriyle Satış Bütçesi Hazırlanması. Yayımlanmamış yüksek lisans tezi, Ankara Üniversitesi Sosyal Bilimleri Enstitüsü, Ankara.
- Khojasteh, J. (2012). Investigating the Sensitivity of Goodness-of-fit Indices to Detect Measurement Invariance in the Bifactor Model. (Doctoral Dissertation). Retrieved from <http://scholarworks.uark.edu/>
- Kirisci, L., Hsu, T., & Yu, L. (2001). Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. *Applied Psychological Measurement*, 25, 146–162.
- Kline, P. (1994). *An easy guide to factor analysis*. London: Routledge.

- Köse, İ. A. (2010). Madde Tepki Kuramına Dayalı Tek Boyutlu ve Çok Boyutlu Modellerin Test Uzunluğu ve Örneklem Büyüklüğü Açısından Karşılaştırılması. (Yayımlanmamış Doktora Tezi). Ankara Üniversitesi. Ankara
- Krykova, I. (2003). *Evaluating of path-dependent securities with low discrepancy methods* (Master's thesis). Retrieved from <http://pdfs.semanticscholar.org/>
- L'Ecuyer, P. (2014). Random Number Generation and Quasi-Monte Carlo. *Wiley StatsRef: Statistics Reference Online*.
- Li, Y. & Lissitz, R. W. (2012). Exploring the full-information bifactor model in vertical scaling with construct shift. *Applied Psychological Measurement*, 36(1), 3.
- Liu, Y. (2011). *Identifying local dependence with a score test statistic based on the bifactor 2-parameter logistic model*. (Master's thesis). Available from ProQuest Dissertations and Theses database. (UMI No. 1506674).
- Lord, F. (1980). *Applications of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum Associates Publishers.
- Lord, F. M. ve Novick M. R. (1968). *Statistical theories of mental test scores*. New York: Addison- Wesley Publishing Company.
- MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, 4, 84–99.
- Magnusson, D. (1968). *Test theory*. Massachusetts. Addison-Wesley.
- Martel, M. M., Von Eye, A., & Nigg, J. T. (2010). Revisiting the latent structure of ADHD: is there a 'g' factor?. *Journal of Child Psychology and Psychiatry*, 51(8), 905-914.
- McDonald, R. P. (1981). The dimensionality of tests and items. *British Journal of Mathematical and Statistical Psychology*, 34, 100-117.
- McDonald, R.P. (1982). Linear Versus Models in Item Response Theory Applied *Psychological Measurement*, 6, 379-396.
- McKinley, R.L. and Reckase, M.D. (1982). *Multidimensional Latent Trait Models*. Paper presented at the National Council on Measurement in Education, New York.

- Niederreiter, H. (1992). *Random number generation and quasi-Monte Carlo methods*. Society for Industrial and Applied mathematics. Philadelphia, Pennsylvania.
- Patrick, C. J., Hicks, B. M., Nichol, P. E., ve Krueger, R. F. (2007). A bifactor approach to modeling the structure of the psychopathy checklist-revised. *Journal of Personality Disorders, 21*(2), 118–141.
- Quinn, H. O. (2014). Bifactor models, explained common variance (ECV), and the usefulness of scores from unidimensional item response theory analyses. (Master's thesis)
- Reckase, M.D. (1997) The Past and Future of Multidimensional Item Response Theory, *Applied Psychological Measurement, 21*(1), 25–36.
- Reckase, M. D. (2009). *Multidimensional item response theory*. New York, NY: Springer.
- Reise, S. P., Morizot, J., & Hays, R. D. (2007). The role of the bifactor model in resolving dimensionality issues in health outcomes measures. *Quality of Life Research, 16*, 19–31.
- Reise, Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*(6), 544–559.
- Reise, S. P., Ventura, J., Keefe, R. S., Baade, L. E., Gold, J. M., Green, M. F., ... & Bilder, R. (2011). Bifactor and item response theory analyses of interviewer report scales of cognitive impairment in schizophrenia. *Psychological assessment, 23*(1), 245.
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate behavioral research, 47*(5), 667-696.
- Reise, S. P., & Revicki, D. A. (2014). *Handbook of item response theory modeling*. Taylor & Francis.
- Revelle, W. (2017). Psych: Procedures for Psychological, Psychometric, and Personality Research. *R package version 1.7.5*. URL: <https://cran.r-project.org/web/packages/psych/index.html>
- Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research, 23*, 51–67.
- Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Applying bifactor statistical indices in the evaluation of psychological measures. *Journal of Personality Assessment, 98*(3), 223-237.

- Simms, L. J., Grös, D. F., Watson, D., & O'Hara, M. W. (2007). Parsing the general and specific components of depression and anxiety with bifactor modeling. *Depression and Anxiety*, 25(7), E34–E46.
- Smith, J. (2009). Some Issues in Item Response Theory: Dimension- ability Assessment and Models for Guessing. Unpublished Doctoral Dissertation. University of South California.
- Spencer, G.S. (2004). The Strength of Multidimensional Item Response Theory in Exploring Constuct Space That is Multidimensional and Correlated. Unpublished Doctoral Dissertation. Brigham Young University.
- Stout, W. F. (1990). A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. *Psychometrika*, 55, 293-325.
- Stucky, B. D., & Edelen, M. O. (2015). Using hierarchical IRT models to create unidimensional measures from multidimensional data. In S. P. Reise, & D. A. Revicki (Eds.), *Handbook of item response theory modeling: Applications to typical performance assessment*. New York, NY: Routledge.
- 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 İncelenmesi. Yayınlanmamış doktora tezi, Mersin Üniversitesi Sosyal Bilimler Enstitüsü, Mersin.
- Şahin, A (2012). *Madde Tepki Kuramı'nda Test Uzunluğu ve Örneklem Büyüklüğünün Model Veri Uyumu, Madde Parametreleri ve Standart Hata Değerlerine Etkisinin İncelenmesi*. (Yayınlanmamış doktora tezi). Hacettepe Üniversitesi, Ankara.
- Thissen, D. & Orlando, M. (2001). Item response theory for items scored in two categories. In *Test Scoring*.
- Thissen, D., & Wainer, H. (Eds.). (2001). *Test scoring*. Routledge
- Wainer, H., & Wang, C. (2000). Using a new statistical model for testlets to score TOEFEL. *Journal of Educational Measurement*, 37, 203–220.
- Wainer, H., and Thiisen, D. (2001) True score theory: The traditional method. Thiisen, D., and Wainer, H. (Ed.). *Test scoring*. New Jersey: Lawrence Erlbaum Associates Publishers.
- Walther, B. A., & Moore, J. L. (2005). The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance. *Ecography*, 28(6), 815-829.

- Wang, Y., & Kim, E. S. (2017). Evaluating Model Fit and Structural Coefficient Bias: A Bayesian Approach to Multilevel Bifactor Model Misspecification. *Structural Equation Modeling: A Multidisciplinary Journal*, 24(5), 699-713.
- Wickham, H. (2008). ggplot2: An implementation of the grammar of graphics. *R package version 0.7*, URL: <http://CRAN.R-project.org/package=ggplot2>.
- Zhang, B. (2008). Application of Unidimensional Item Response Models to Tests With Items Sensitive to Secondary Dimensions. *The Journal of Experimental Education*, 77(2), 147-166.
- Zheng, C., (2013). Examination of the Parameter Estimate Bias When Violating the Orthogonality Assumption of the Bifactor Model. (Doctoral Dissertation) Available from ProQuest Dissertations and Theses database.(UMI: 3559167)





EKLER

EK-A: Simülasyon Kod Örneği (100 madde 0.40 korelasyon)

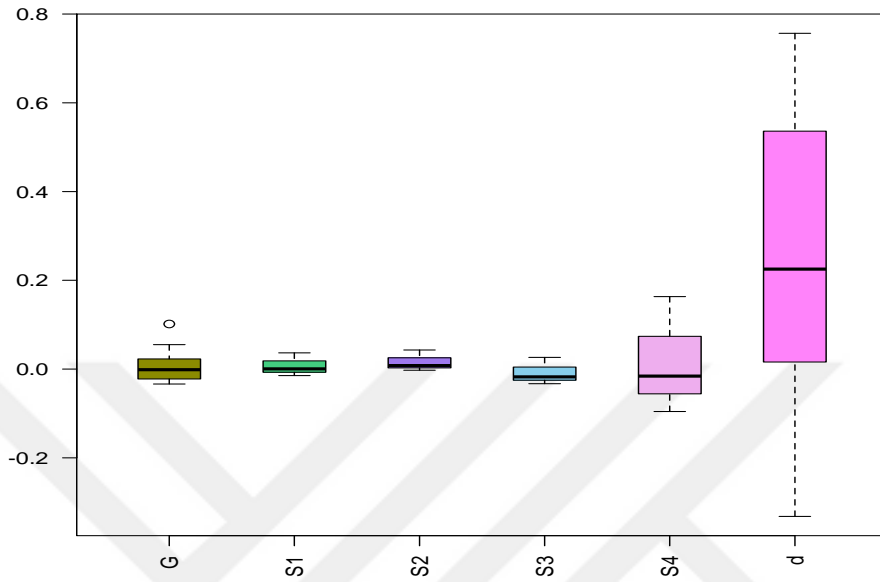
```

library(mvtnorm)
N <- 5000
D <- 5
npb <- 25
M <- (D-1)*npb
meanTheta <- rep(0, D)
sdTheta <- rep(1, D)
rhoTheta <-
matrix(c(1,rep(0,4),0,1,rep(0,3),0,0,1,rep(0,2),0,0,0,1,rep(0.4,1),rep(0,3),0.4,1),
ncol=5)
covTheta <- rhoTheta * sdTheta %o% sdTheta
theta <- round(rmvnorm(N, mean = meanTheta, sigma = covTheta, method=
"chol"),digits=3)
write.csv(theta,file ="theta100.csv", row.names=FALSE)
set.seed(255)
diffp <- round(rnorm(M, 0, 1),digits=3)
set.seed(256)
discraw <- round(matrix(runif(D*M, min=0.2, max=2.0), ncol=D),digits=3)
blotterMatrix <- matrix(0, nrow=M, ncol=D-1)
for (i in 1:(D-1)){
blotterMatrix[(1 + (i-1)*npb):(i*npb), i] <- 1
BlotterMatrix <- cbind(1, blotterMatrix)
discp <- discraw * BlotterMatrix
discssum <- sqrt(matrix(apply(discp*discp,1,sum),ncol=1))
intercept <- cbind(-diffp*discssum)
DiffAndIntercept <- cbind(discp,intercept)
eta <- discp %*% t(theta) - diffp
invlink <- 1/(1+exp(-eta))
}
write.csv(DiffAndIntercept,file ="DiffAndIntercept.csv", row.names=FALSE)
write.csv(invlink ,file ="invlink .csv", row.names=FALSE)
rs <- 200
for (i in 1:rs){
items <- t(apply(invlink,2, function(col) rbinom(n=M, pr=col, size=1)))
myname <- paste("bifactor", i, ".csv")
write.table(items, file=myname,sep = ",",quote=TRUE,row.name=TRUE,
col.names=TRUE)
}

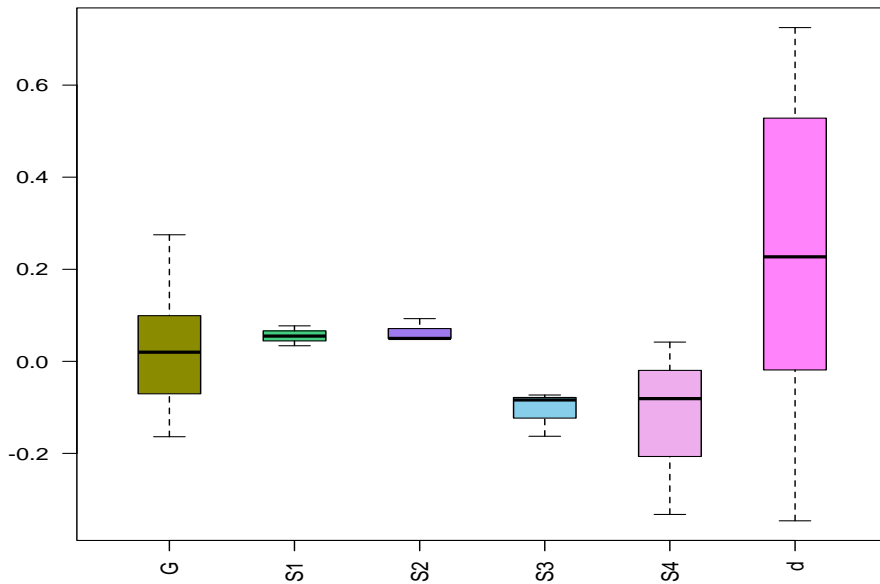
```

Ek B: 12 Maddelik Test Uzunluğu İçin Spesifik Modellerdeki Madde Parametreleri Kestirim Yanlılıkları

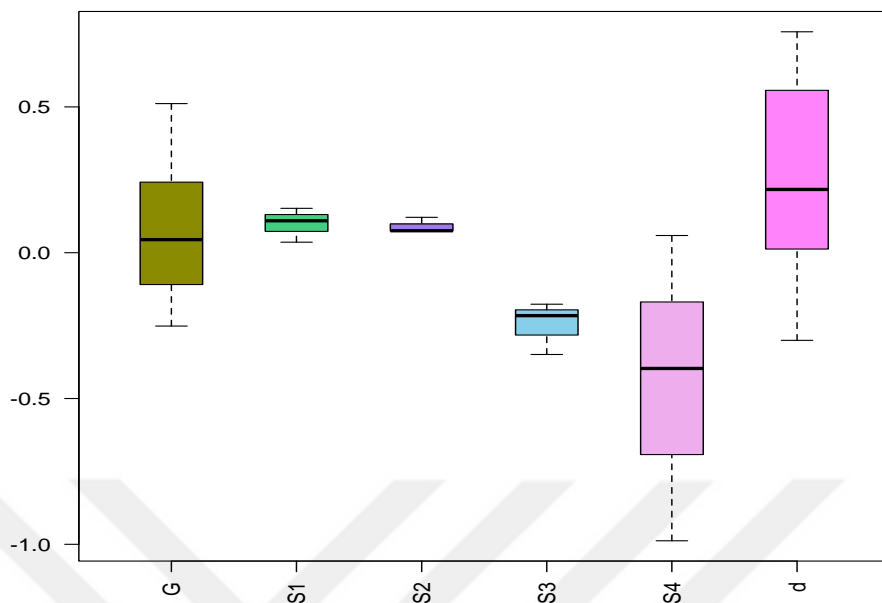
Model 11_12



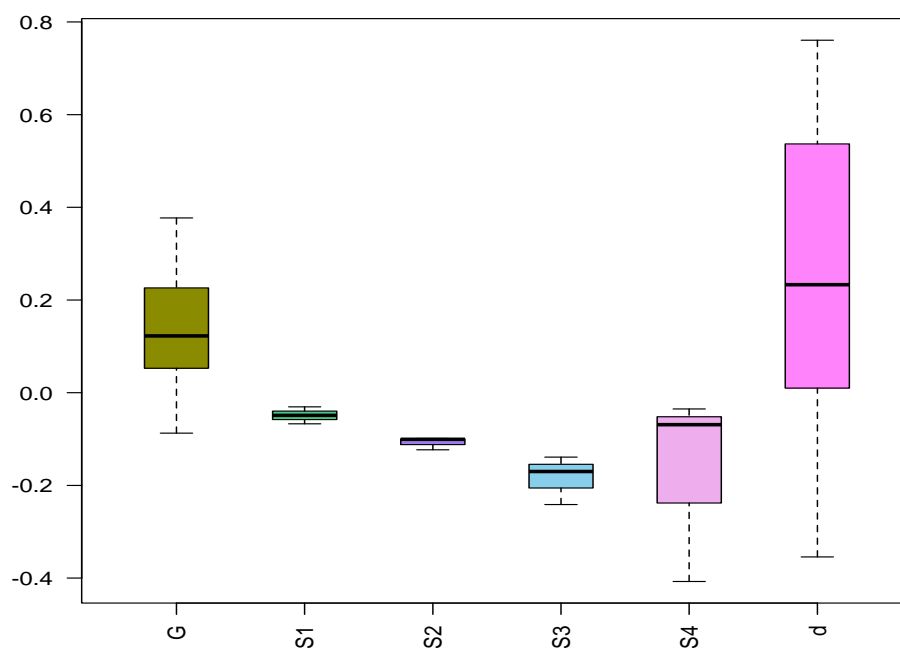
Model 12_12



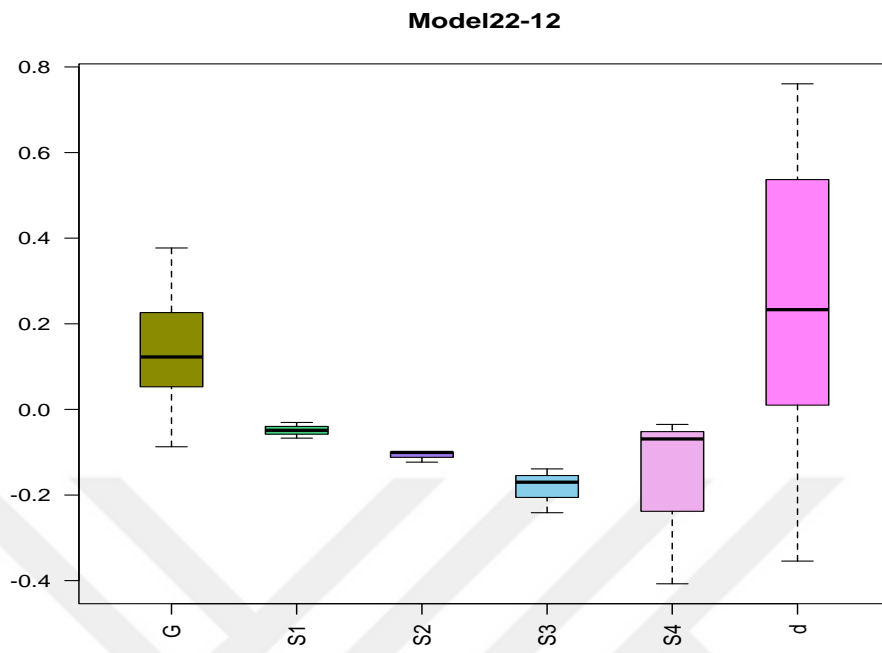
Model 13_12



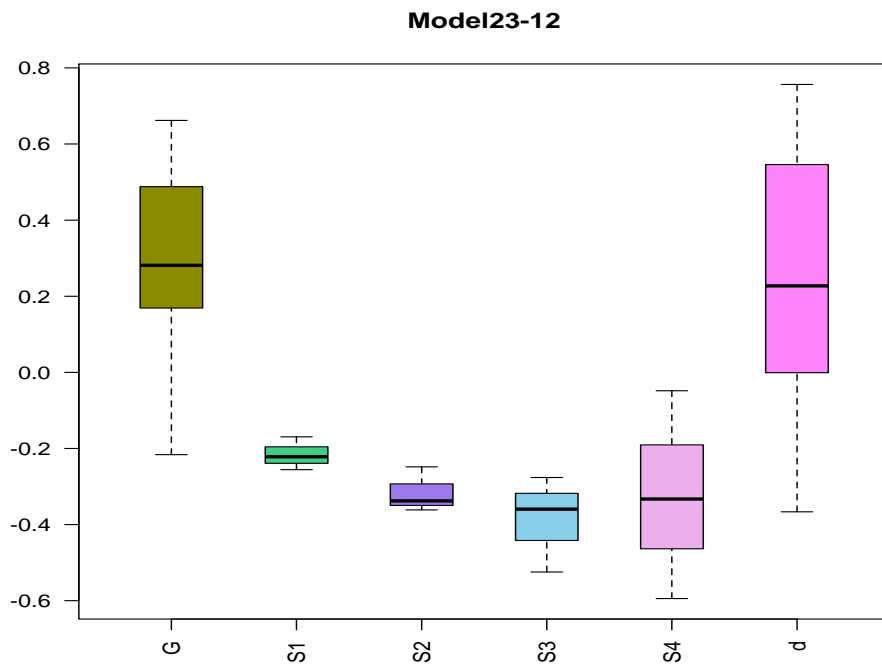
Model 21_12

Model21-12

Model 22_12

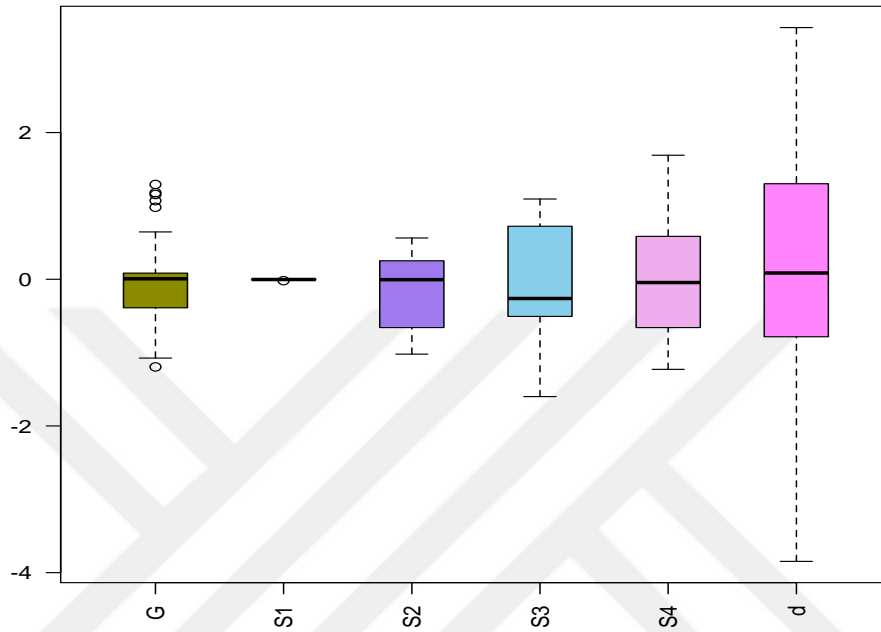


Model 23_12

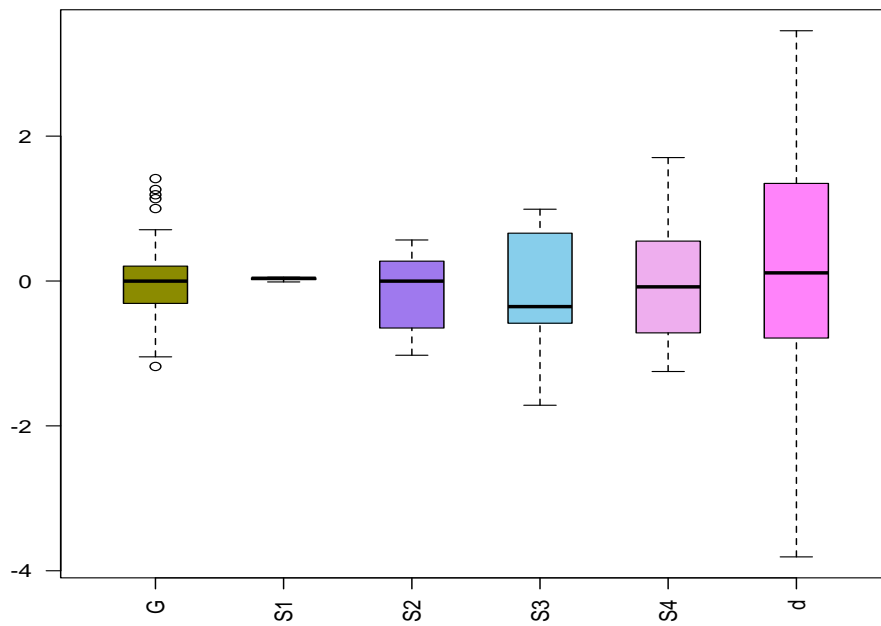


Ek C: 40 Maddelik Test Uzunluđu İin Spesifik Modellerdeki Madde Parametreleri Kestirim Yanlılıkları

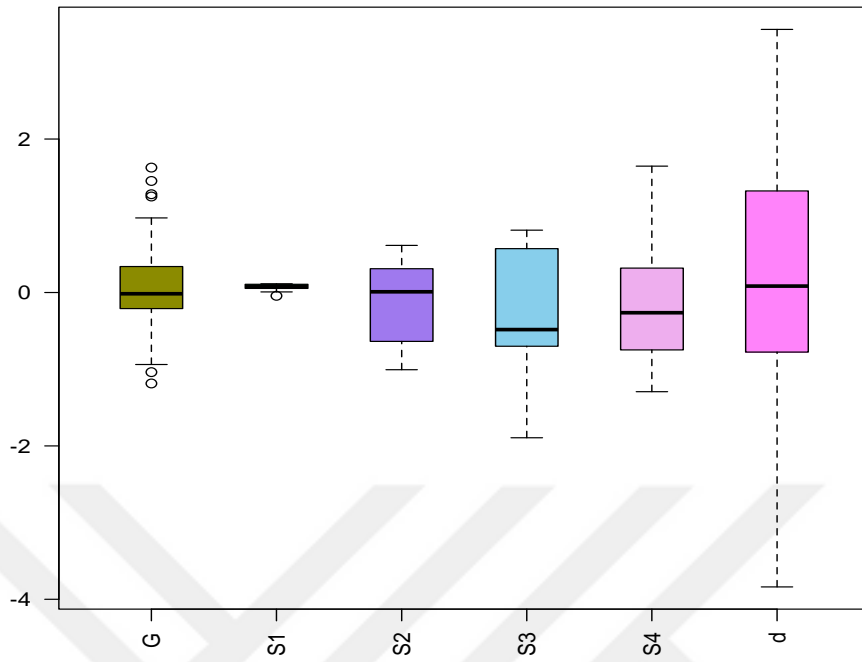
Model11_40



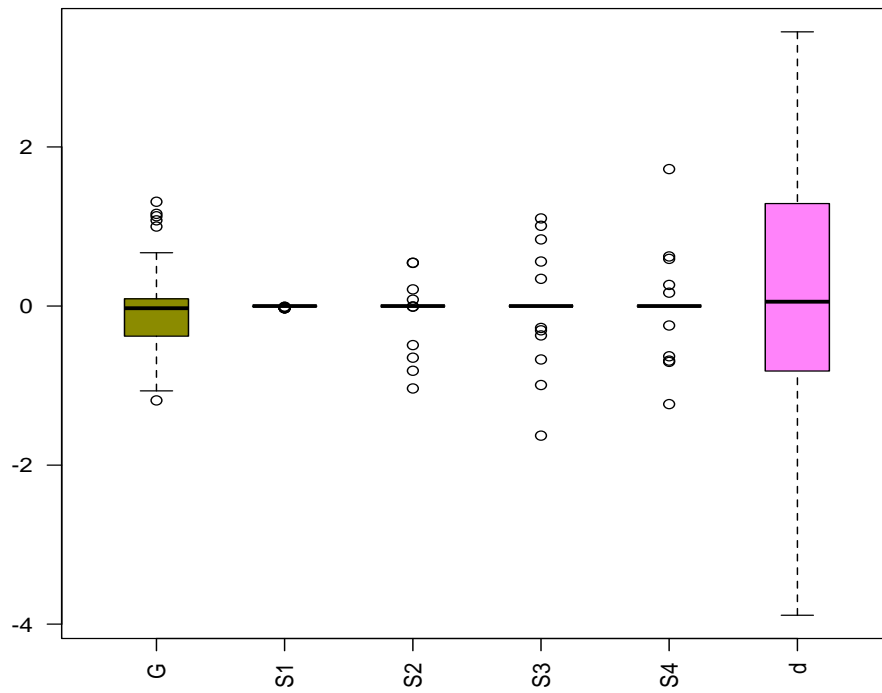
Model12_40



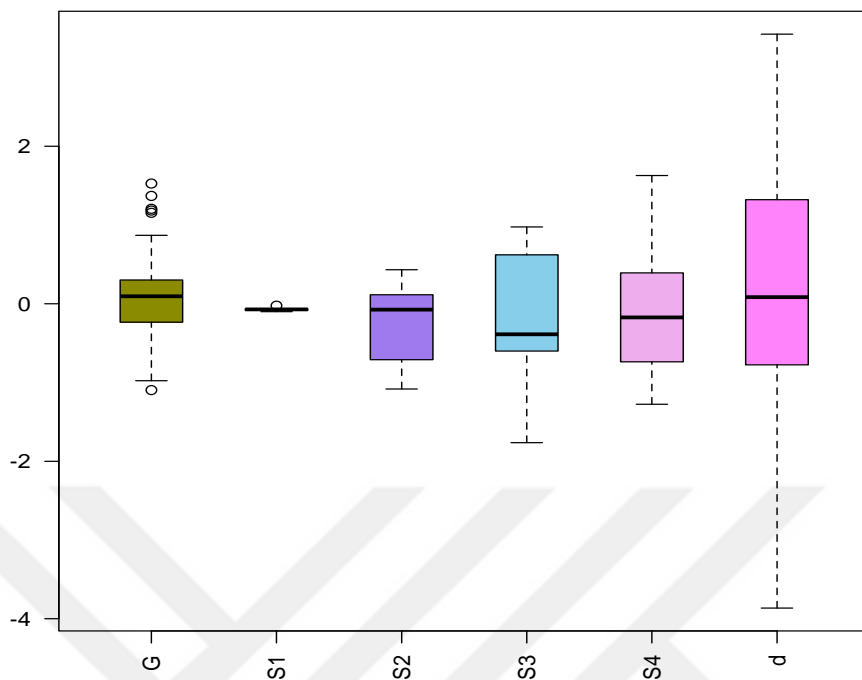
Model13_40



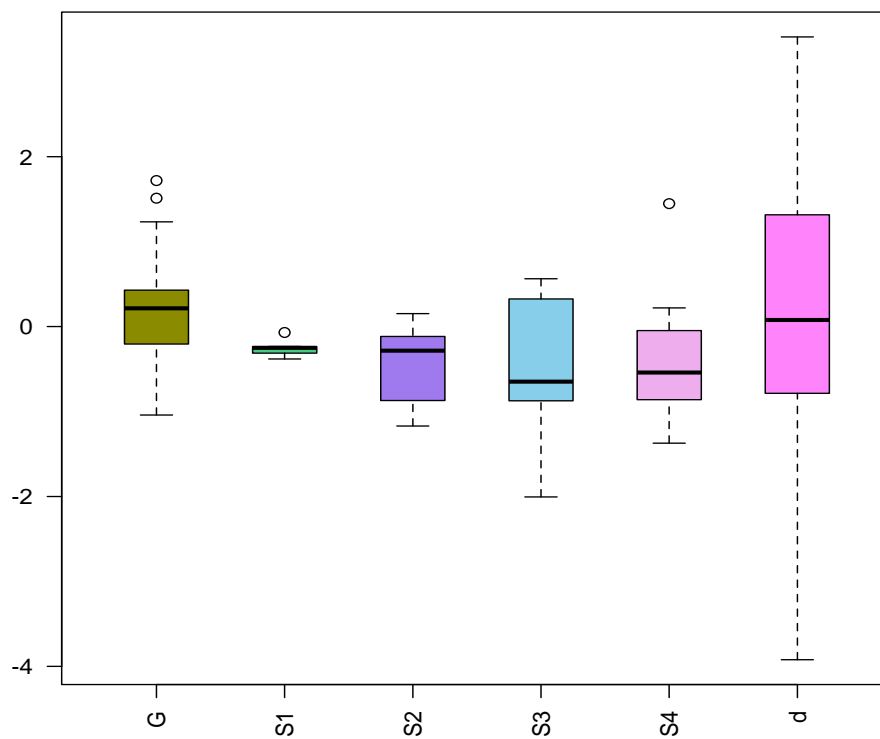
Model21_40



Model22-40

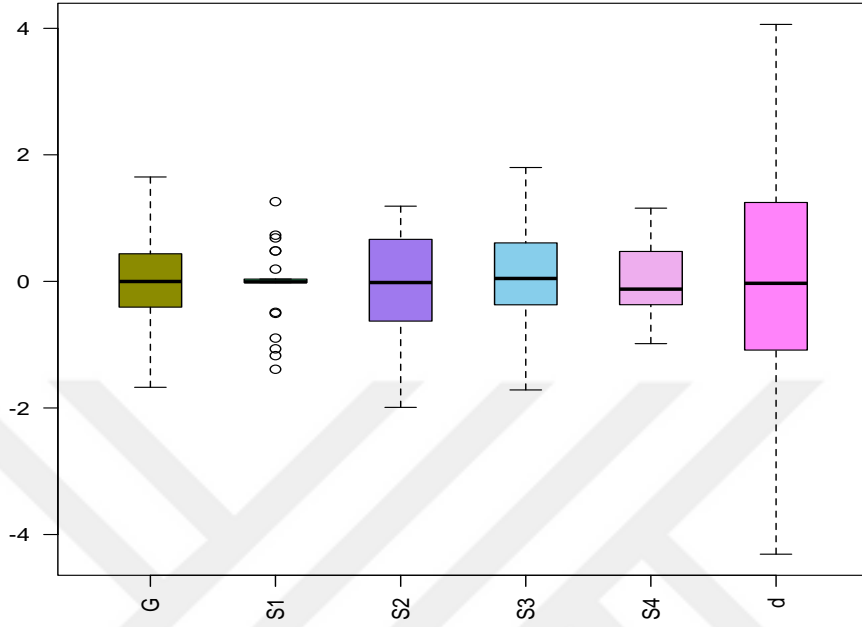


Model23-40

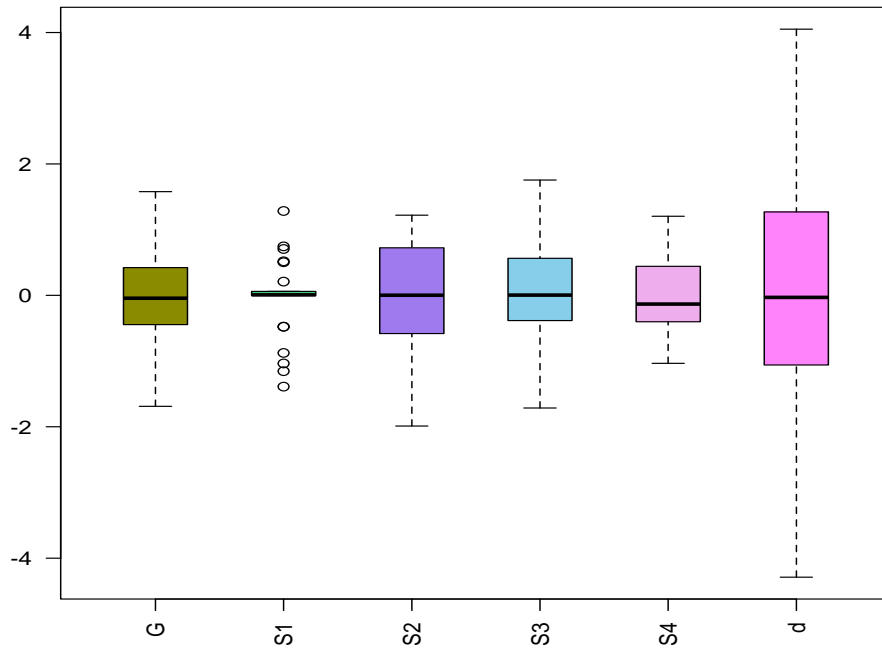


Ek D: 100 Maddelik Test Uzunluğu İçin Spesifik Modellerdeki Madde Parametreleri Kestirim Yanlılıkları

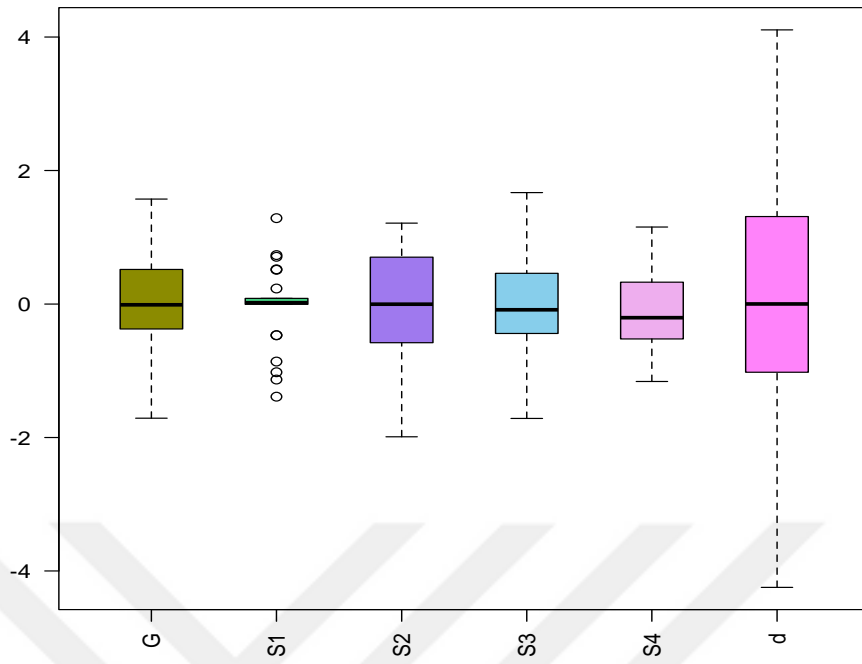
Model11_100



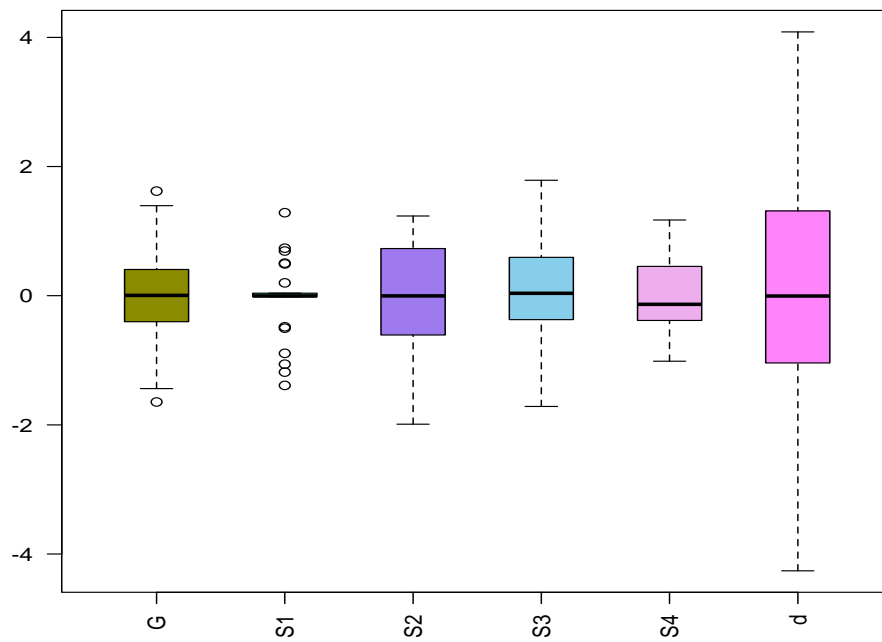
Model12_100



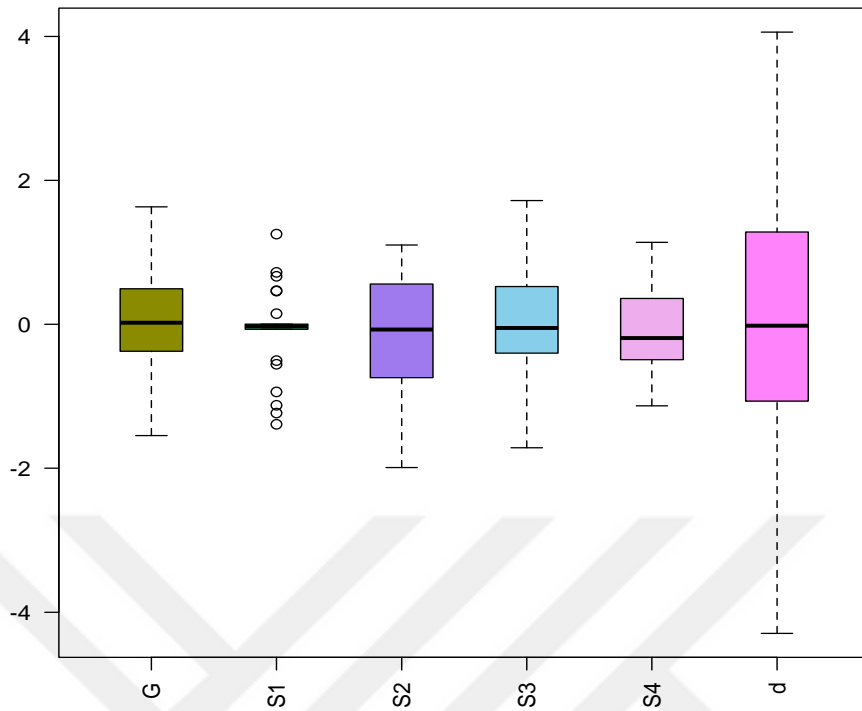
Model13_100



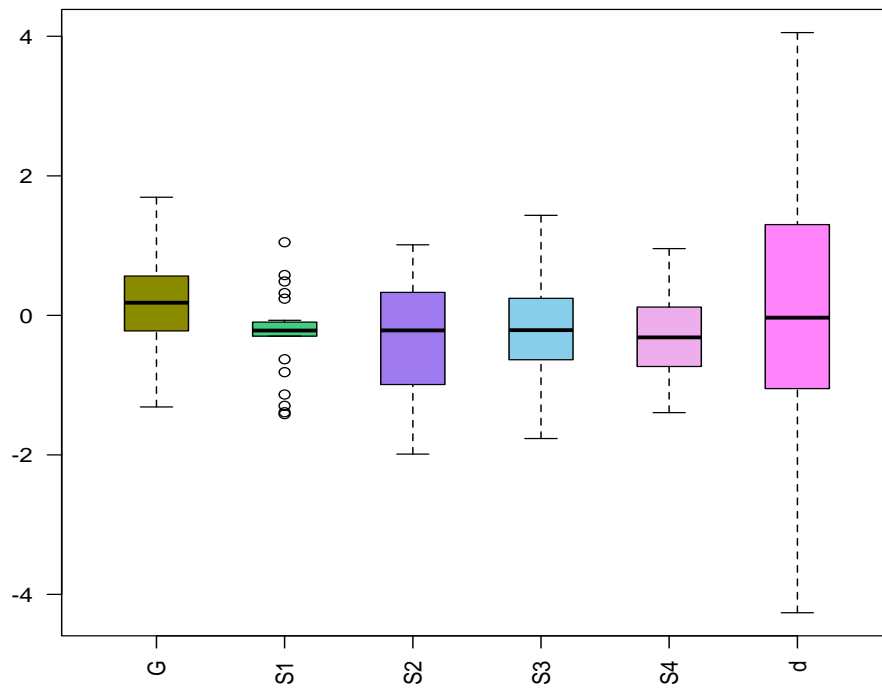
Model21-100



Model22-100

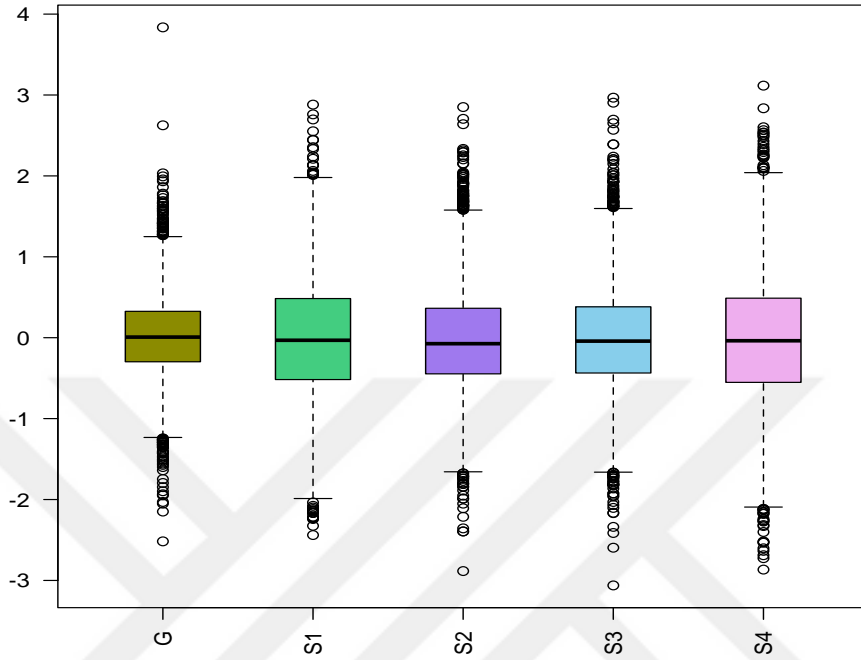


Model23-100

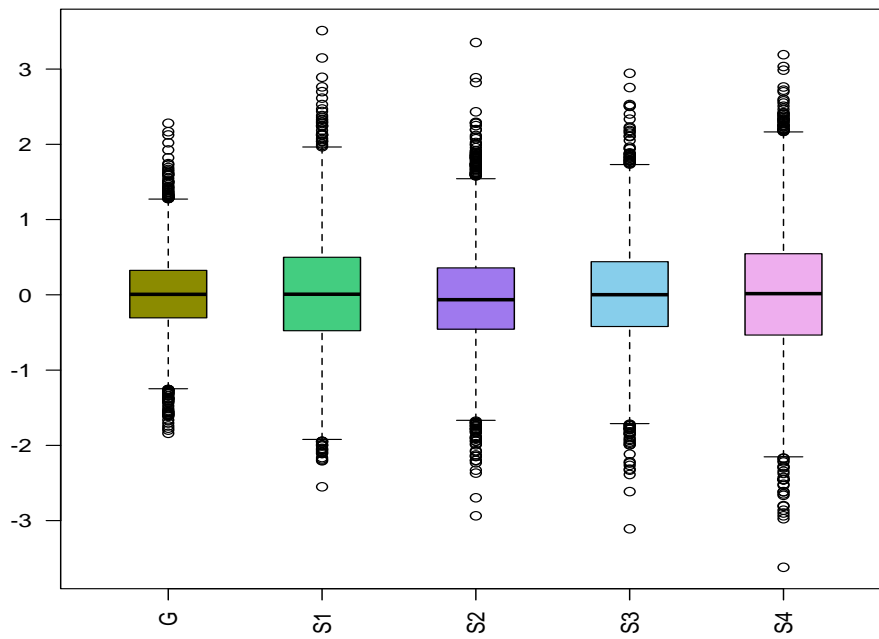


Ek E: 12 Maddelik Test Uzunluğu İçin Spesifik Modellerdeki Birey Parametrelerinin Kestirim Yanlılıkları

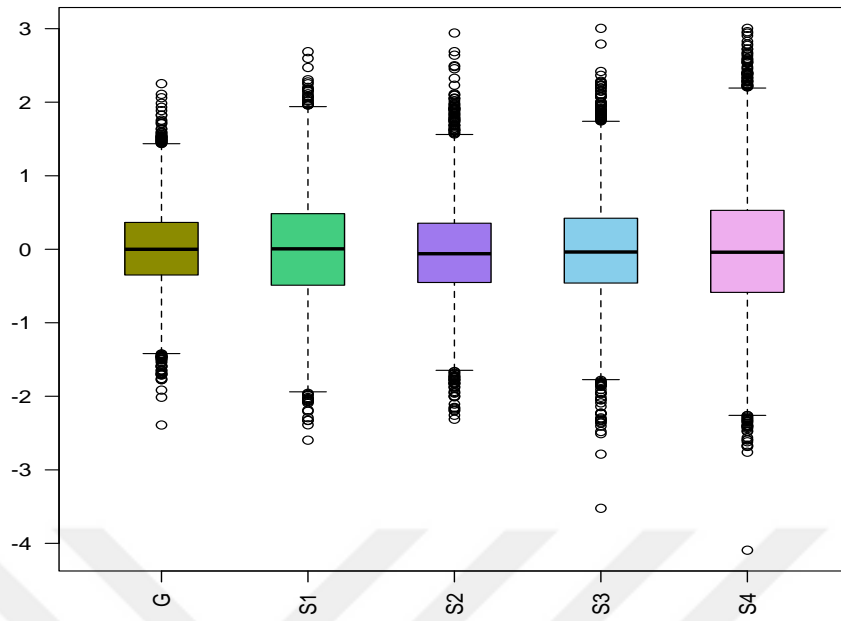
Model11_12



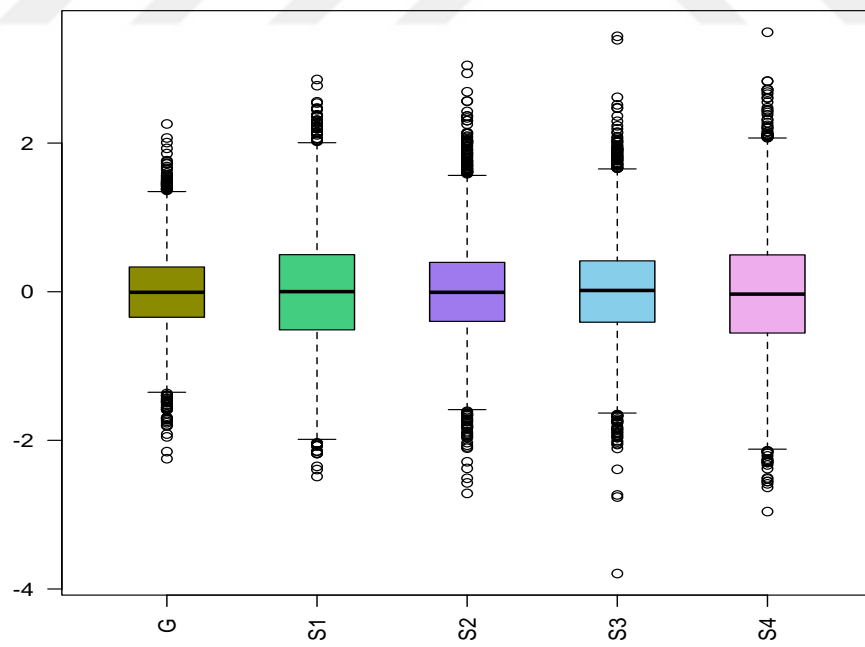
Model12_12



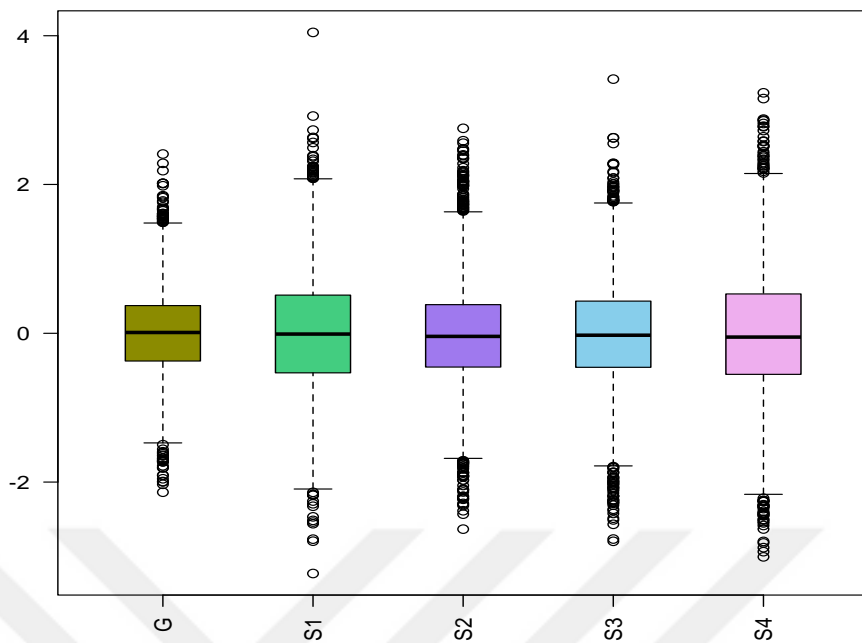
Model13_12



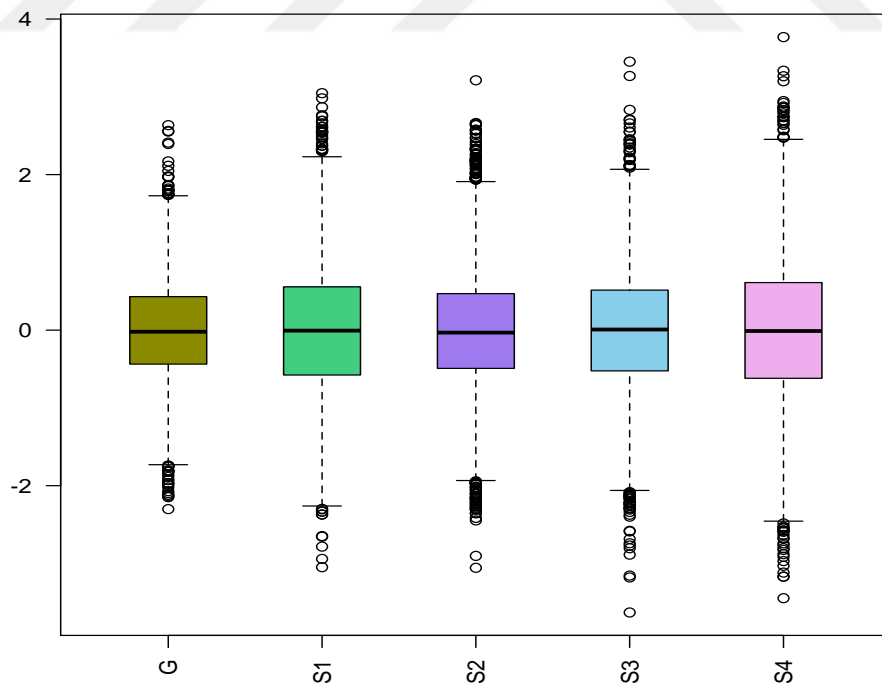
Model21_12



Model22_12

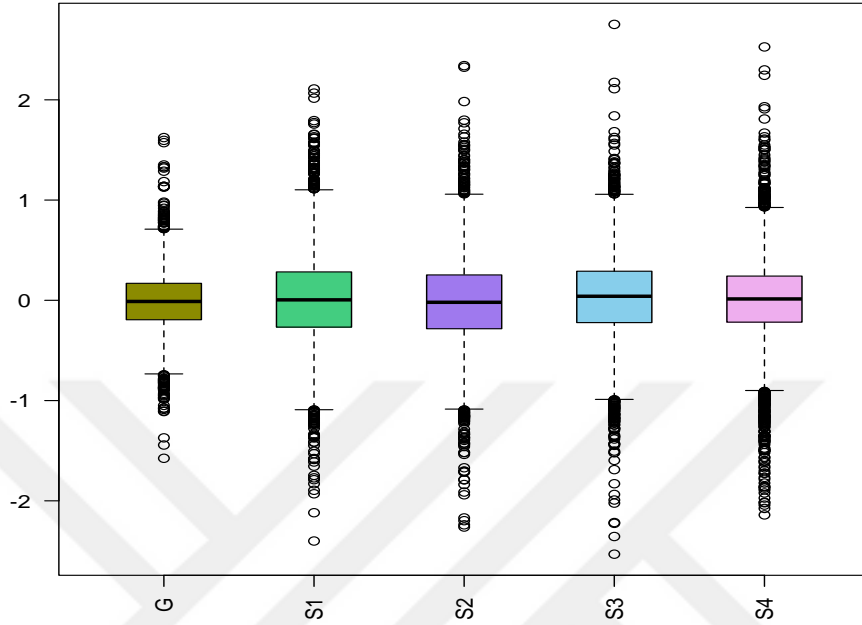


Model23_12

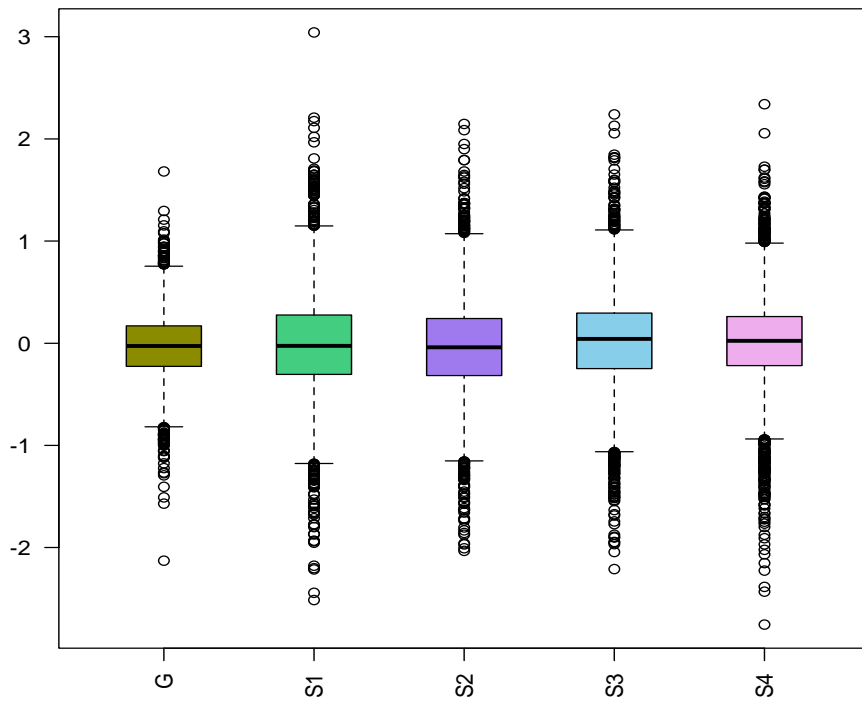


Ek F: 40 Maddelik Test Uzunluğu İçin Spesifik Modellerdeki Birey Parametrelerinin Kestirim Yanlılıkları

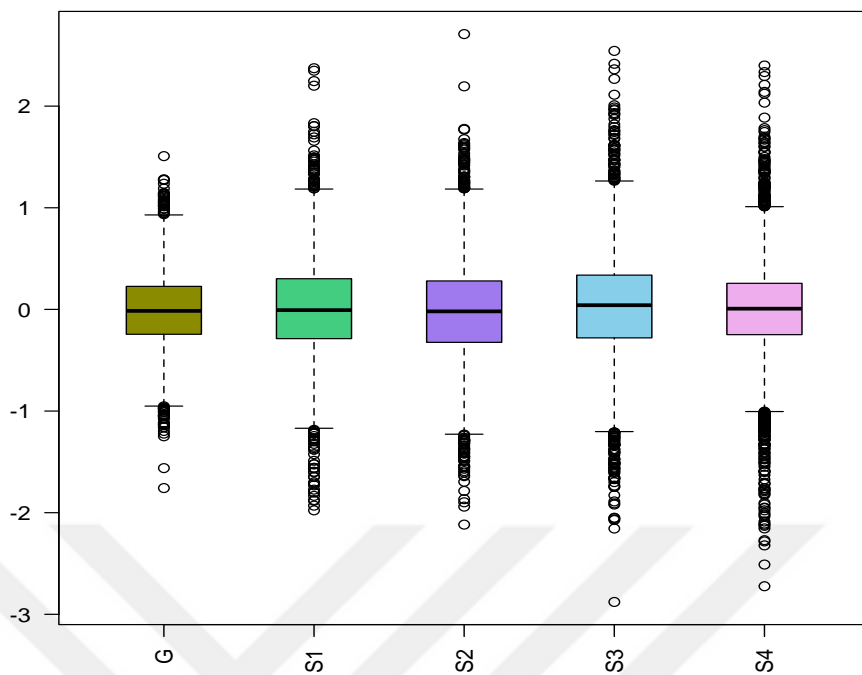
Model11_40



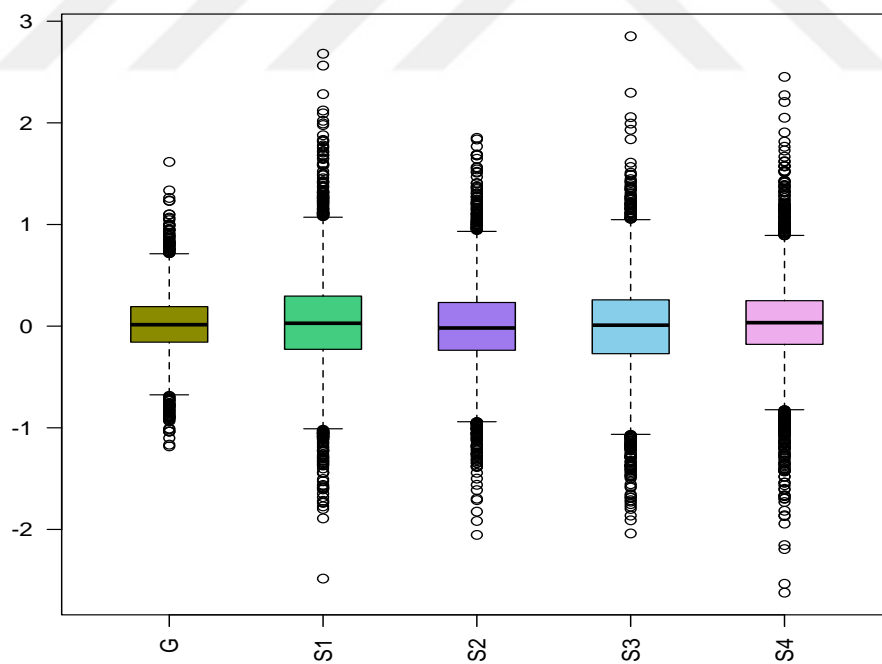
Model12_40



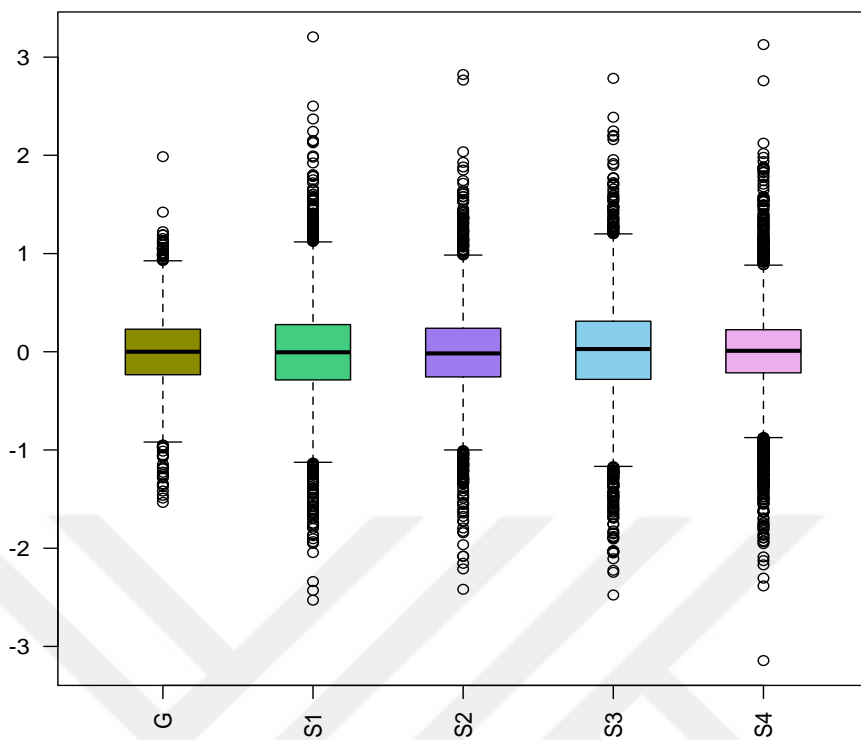
Model13_40



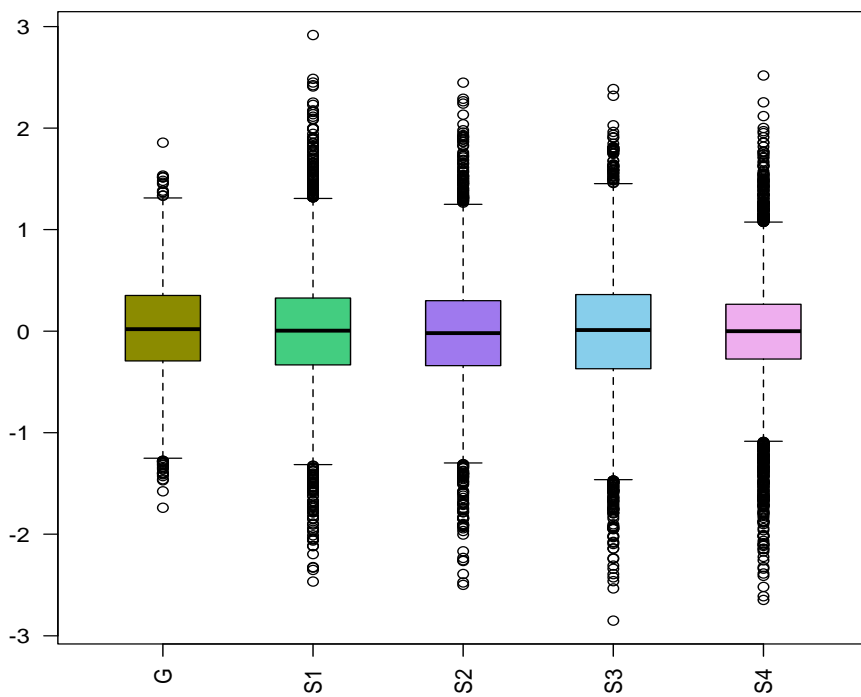
Model21_40



Model22_40

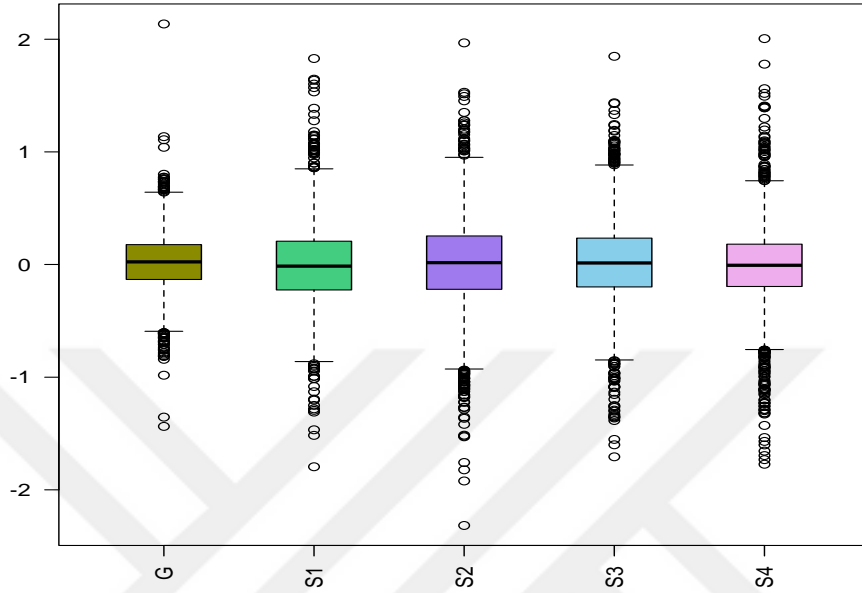


Model23_40

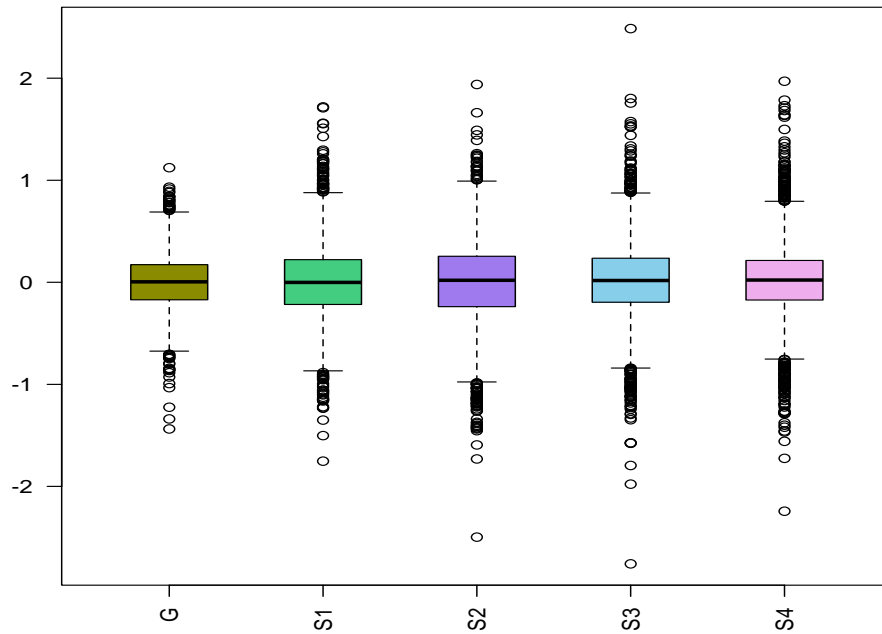


Ek G: 100 Maddelik Test Uzunluğu İçin Spesifik Modellerdeki Birey Parametrelerinin Kestirim Yanlılıkları

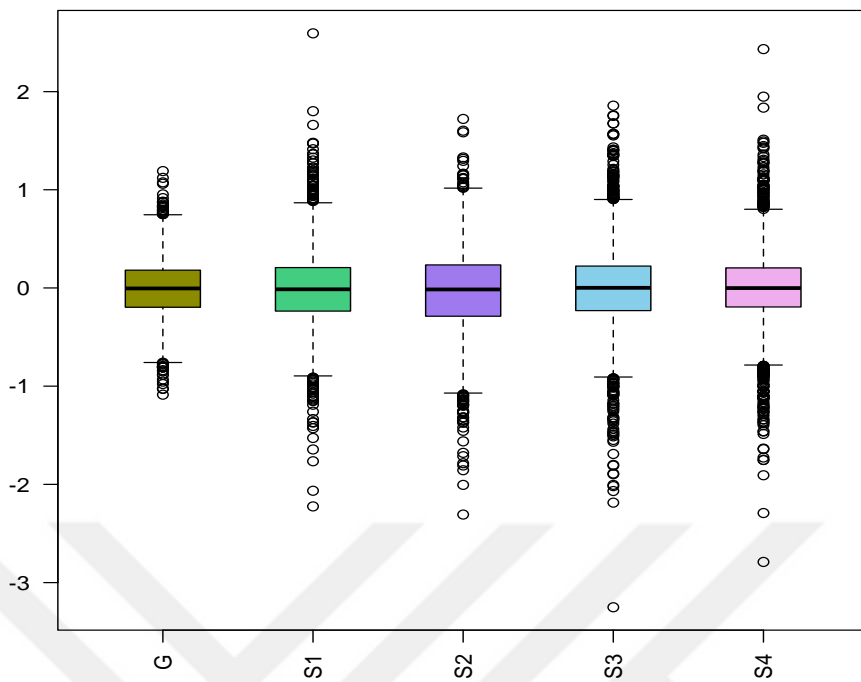
Model11_100



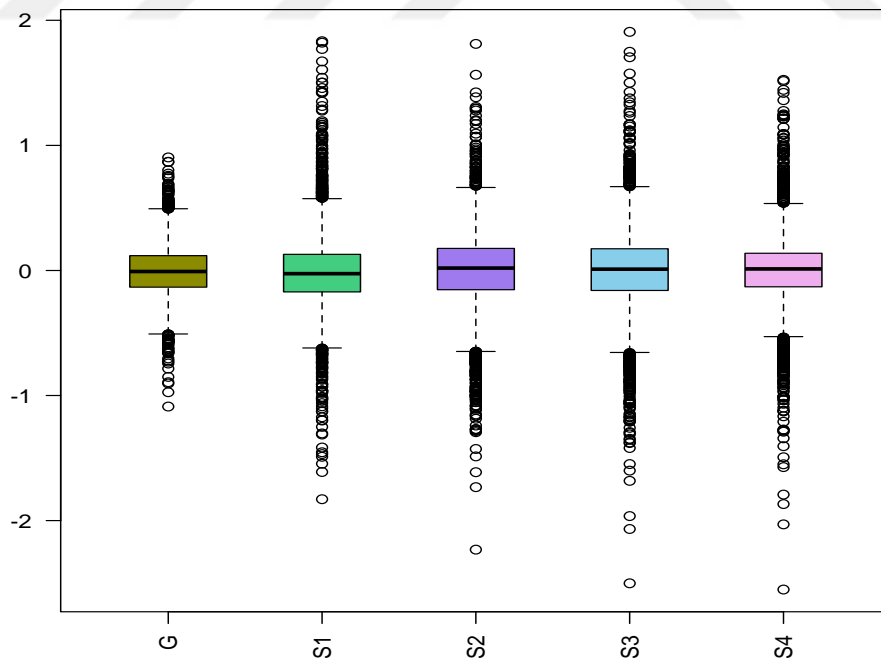
Model12_100



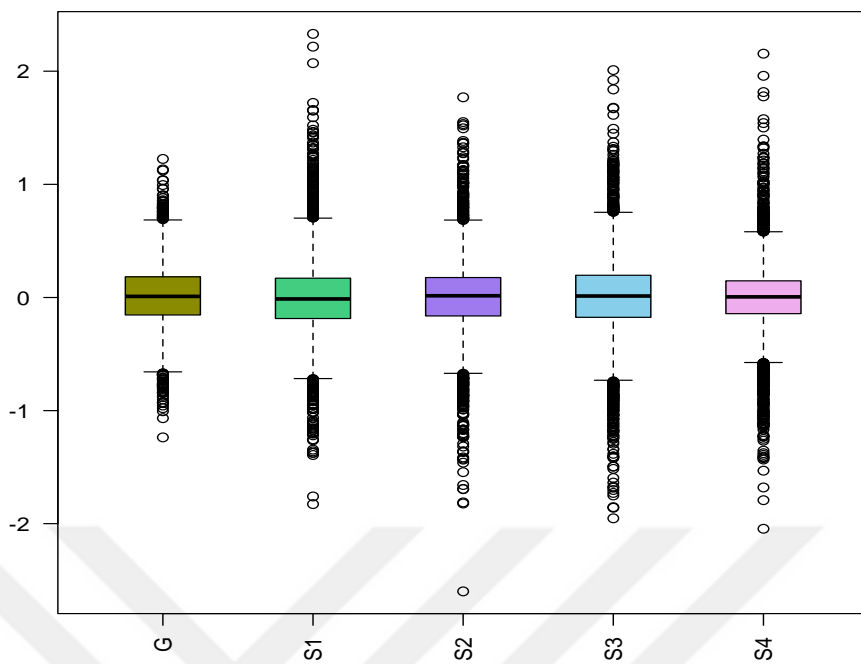
Model13_100



Model21_100



Model22_100



Model23_100

