

**TOPLAM TEST PUANI VE ALT TEST PUANLARININ
KESTİRİMİNİN HİYERARŞİK MADDE TEPKİ KURAMI
MODELLERİ İLE KARŞILAŞTIRILMASI**

**COMPARISON OF ESTIMATION OF TOTAL SCORE AND
SUBSCORES WITH HIERARCHICAL ITEM RESPONSE
THEORY MODELS**

Sümevra SOYSAL

Hacettepe Üniversitesi

Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı

Doktora Tezi

olarak hazırlanmıştır.

2017

KABUL ve ONAY

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼'ne,

S¼meyra SOYSAL'ın hazırladıđı "Toplam Test Puanı ve Alt Test Puanlarının Kestiriminin Hiyerarşik Madde Tepki Kuramı Modelleri ile Karşılaştırılması" başlıklı bu çalışma j¼rimiz tarafından **Eđitim Bilimleri Anabilim Dalı, Eđitimde Ölçme ve Deđerlendirme Bilim Dalı'nda Doktora Tezi** olarak kabul edilmiştir.

Başkan Prof. Dr. Selahattin GELBAL



¼ye (Danışman) Prof. Dr. H¼lya KELEÇİOđLU



¼ye Doç. Dr. Nuri DOđAN



¼ye Doç. Dr. Kaan Z¼lfikar DENİZ



¼ye Yrd. Doç. Dr. Deniz G¼LLEROđLU



ONAY

Bu tez Hacettepe ¼niversitesi Lisans¼st¼ Eđitim-¼đretim ve Sınav Y¼netmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri ¼yeleri tarafından 11/ 12 / 2017 tarihinde uygun g¼r¼lm¼ş ve Enstit¼ Y¼netim Kurulunca / / tarihinde kabul edilmiştir.

Prof. Dr. Ali Ekber ŐAHİN
Eđitim Bilimleri Enstit¼s¼ M¼d¼r¼

YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI

Enstitü tarafından onaylanan lisansüstü tezimin/raporumun tamamını veya herhangi bir kısmını, basılı (kağıt) ve elektronik formatta arşivleme ve aşağıda verilen koşullarla kullanıma açma iznini Hacettepe Üniversitesine verdiğimi bildiririm. Bu izinle Üniversiteye verilen kullanım hakları dışındaki tüm fikri mülkiyet haklarım bende kalacak, tezimin tamamının ya da bir bölümünün gelecekteki çalışmalarda (makale, kitap, lisans ve patent vb.) kullanım hakları bana ait olacaktır.

Tezin kendi orijinal çalışmam olduğunu, başkalarının haklarını ihlal etmediğimi ve tezimin tek yetkili sahibi olduğumu beyan ve taahhüt ederim. Tezimde yer alan telif hakkı bulunan ve sahiplerinden yazılı izin alınarak kullanılması zorunlu metinlerin yazılı izin alınarak kullandığımı ve istenildiğinde suretlerini Üniversiteye teslim etmeyi taahhüt ederim.

Tezimin/Raporumun tamamı dünya çapında erişime açılabilir ve bir kısmı veya tamamının fotokopisi alınabilir.

(Bu seçenekle teziniz arama motorlarında indekslenebilecek, daha sonra tezinizin erişim statüsünün değiştirilmesini talep etseniz ve kütüphane bu talebinizi yerine getirse bile, teziniz arama motorlarının önbelleklerinde kalmaya devam edebilecektir)

Tezimin/Raporumun 01/01/2019 tarihine kadar erişime açılmasını ve fotokopi alınmasını (İç Kapak, Özet, İçindekiler ve Kaynakça hariç) istemiyorum.

(Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir, kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisi alınabilir).

Tezimin/Raporumun tarihine kadar erişime açılmasını istemiyorum ancak kaynak gösterilmek şartıyla bir kısmı veya tamamının fotokopisinin alınmasını onaylıyorum.

Serbest Seçenek/Yazarın Seçimi:

.....

10 /02 /2018



Sümeyra SOYSAL

ETİK BEYANNAMESİ

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, tez yazım kurallarına uygun olarak hazırladığım bu tez çalışmada,

- tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- görsel, işitsel ve yazılı tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.

İmza
Sümeyra SOYSAL

TEŐEKKÜR

Doktora öğrenimim boyunca aldığım dersler aracılığıyla bilgisinden ve deneyiminden faydalandığım, en zor zamanlarımda dahi her zaman desteđi ile yanımda olan tez danışmanın canım hocam Prof. Dr. Hülya KELECİOđLU'na;

Ölçme ve Deđerlendirme bilim dalında eğitime başladığım ilk günden itibaren bilgisi ve deneyiminden faydalandığım ve özellikle yaşadığım en zor dönemlerimde idari amirim olarak gösterdiği anlayış ve destek ile iş kaygımı azaltan sevgili hocam Selahattin GELBAL'a;

Tez jürimde yer alarak deđerli görüş ve önerileri ile tezimin son halini almasında katkıları bulunan Doç. Dr. Nuri DOđAN'a;

Bu günlere gelmemde büyük emeđi ve payı olan, benliğimin her zerresinde izlerini taşıdığım, her geçen gün eksikliđini daha fazla hissettiğim ve özlediğim canımın içi annem Miyase SOYSAL ve babaannem Döne SOYSAL'a;

Yine bu günlere gelmemde emeđi ve desteđi olan babam Mevlüt SOYSAL ve kardeşlerim Hadis SOYSAL, Ahmet SOYSAL ve Osman SOYSAL'a;

Doktora öğrenimim boyunca yurt içi doktora burs olanađı sağlayan TÜBİTAK'a;

Destek ve katkıları ile hep yanımda olan dostlarıma ve meslektaşlarıma;

Hayatım boyunca beni destekleyip bilgisini benimle paylaşan adını saymadığım tüm hocalarım ve arkadaşlarıma

teşekkürlerimi sunarım.

TOPLAM TEST PUANI VE ALT TEST PUANLARININ KESTİRİMİNİN HİYERARŞİK MADDE TEPKİ MODELLERİ İLE KARŞILAŞTIRILMASI

Sümevra SOYSAL

ÖZ

Bu araştırmada güvenilir alt test ve toplam test puanı kestirimleri konusuna katkı sağlamak amacıyla alt test ve toplam test arasındaki ilişki hiyerarşik madde tepki kuramı modelleri ile araştırılmak istenmiştir. Çalışmada Üst Düzey Sıralı (Higher Order), İki Faktör (Bi-factor) ve hiyerarşik çok boyutlu madde tepki kuramı (ÇBMTK) modelleri ile kestirilen toplam test puanının ve alt test puanlarının RMSE ve güvenilirlik değerleri alt test sayısı, alt test uzunluğu ve alt testler arasındaki korelasyonların büyüklüğü koşulları altında karşılaştırılmıştır. Ayrıca TEOG 2015 verileri üzerinde araştırmada kullanılan üç kestirim modelinin performansı incelenmiştir.

TEOG 2015 verisinin madde parametrelerine dayalı olarak bu çalışmada kullanılan verilerin madde ayırt edicilik parametresi ranjı [0.8-3] arasında olacak şekilde ortalaması 1.5 ve varyansı 0.5 olan bir normal dağılımdan; güçlük parametresi ranjı [-2, 2] arasında olacak şekilde ortalaması 0.0 ve varyansı 1.0 olan bir normal dağılımdan ve en düşük asimtot (şans) parametresi ise (6,16) olan bir beta dağılımdan üretilmiştir. Yetenek parametreleri çok değişkenli normal dağılıma dayalı olarak ortalaması sıfır (0), varyansı ise araştırma koşullarında belirlenmiş olan varyans-kovaryans matrisine göre üretilmiştir. Üretilen madde ve yetenek parametreleri kullanılarak alt test sayısı 2 ve 3, alt test uzunluğu 20, 30 ve 40; alt testler arası korelasyonlar 0.0, 0.3, 0.5 ve 0.8 olan iki kategorili 3000 kişilik veri setleri 50 tekrara dayalı olarak SimuMIRT programı kullanılarak üretilmiştir. Araştırmada üretilen veriler ve 29 Nisan 2015'te yapılan TEOG sınavına ait veriler BMIRT programı kullanılarak analiz edilmiştir. Parametre kestirimlerinde 3PL model ile MCMC kestirim yöntemi kullanılmıştır.

Araştırmanın sonucunda iki ve üç boyutlu verilerde hemen hemen tüm koşullarda alt test uzunluğu ve alt testler arasındaki korelasyonun arttıkça üç kestirim modelinden elde edilen toplam test puanı için yetenek parametreleri kestirim hatasının azaldığı, kestirim güvenilirliğinin ise arttığı bulunmuştur. Toplam test

puanları için Hiyerarşik ÇBMTK model ile tüm koşullarda en düşük RMSE değeri ve en yüksek güvenilirlik değeri elde edilmiştir. Ayrıca korelasyonun 0.8 düzeyinde toplam test puanı için tüm modeller birbirine yakın RMSE ve güvenilirlik değerleri ile kestirim yapmıştır. İki ve üç boyutlu verilerde alt test puanı için kestirilen yetenek parametrelerinin RMSE değerleri, Hiyerarşik ÇBMTK modelde alt test uzunluğu arttıkça azalırken alt testler arasındaki korelasyon düzeyinden etkilenmediği; Üst Düzey Sıralı modelde alt test uzunluğu ve alt testler arasındaki korelasyon arttıkça azaldığı; İki Faktör modelde ise alt test uzunluğu arttıkça azalırken alt testler arasındaki korelasyon arttıkça önemli düzeyde arttığı bulunmuştur. Tüm koşullar altında Hiyerarşik ÇBMTK model alt test yetenek parametresini kabul edilebilir güvenilir düzeyde kestirmiştir. Ayrıca koşulların çoğunda Hiyerarşik ÇBMTK ve Üst Düzey Sıralı Modelin alt test puanlarını benzer düzeyde hatalar ile kestirmiş fakat korelasyonun yüksek düzeylerinde Üst Düzey Sıralı model daha iyi performans göstermiştir.

Araştırmadan elde edilen bulgulara dayanarak; geniş ölçekli testlerin raporlanmasında Hiyerarşik ÇBMTK modelin kullanımı önerilmektedir. Ayrıca alt testler arasında orta ve düşük düzeyde ilişkilerin olduğu bilinen sınavların raporlanmasında hiyerarşik ÇBMTK Model'e alternatif olarak bu modelle çok yakın analizler yapabilen Üst Düzey Sıralı Model'in kullanımı da tercih edilebilir.

Anahtar sözcükler: Alt test puan kestirimi, toplam test puan kestirimi, hiyerarşik madde tepki kuramı modelleri, üst düzey sıralı model, İki Faktör model

Danışman: Prof. Dr. Hülya KELECİOĞLU, Hacettepe Üniversitesi, Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı

COMPARISON OF ESTIMATION OF TOTAL SCORE AND SUBSCORES WITH HIERARCHICAL ITEM RESPONSE THEORY MODELS

Sümeyra SOYSAL

ABSTRACT

In this study, the relationship between subtest and total test was investigated by using hierarchical item response theory models in order to contribute to reliable subtest and total test score estimates. The RMSE and reliability of the total test score and subtest scores estimated by the Higher Order, Bi-factor and hierarchical MIRT models in the study were compared under the conditions of the size of the correlations between the subtest number, subtest length and number of subtests. In addition, the performance of three models used in the research was examined on TEOG 2015 data.

To generate data sets based on the item parameters of the TEOG 2015 data, item discrimination parameters were drawn from normal distribution with a mean of 1.5 and a variance of 0.5; item difficulty parameters were drawn from normal distribution with a mean of 0.0 and a variance of 1.0, and guessing (lower asymptote) parameters were drawn from beta distribution with (6,16). The true subtest abilities were drawn from a multivariate normal distribution with variance-covariance matrix based on the correlations between the dimensions explained under simulation conditions. Finally, given subtest abilities and item parameters, binary responses were simulated for number of subtest (2,3), subtest length (20,30,40) and correlation between subtest (0.0, 0.3, 0.5, 0.8) by SimuMIRT software. The simulated data and TEOG 2015 data was analyzed by BMIRT software. For the parameter estimates, 3PL model and MCMC estimation method are used.

As a result of the study, in almost all conditions, the correlation between the subtest length and the subtests increased, the RMSE of the ability parameters decreased and the reliability increased for the total test score obtained from the three estimation models. For all test scores, the Hierarchical MIRT model yielded the lowest RMSE value and highest reliability value under all conditions. In addition, all models estimated RMSE and reliability values close to each other at 0.8 level of correlation. The RMSE values of the ability parameters for the subtest scores in two and three

dimensional data were found to be not affected by the correlation level between the subtests while the subtest length decreased in the Hierarchical MIRT model; were found to decrease as the correlation between subtest length and subtest in the Higher Order model and were found to decrease as the subtest length increased, but significantly increased as the correlation between the subtests increased in the Bi-factor model. Under all conditions, the Hierarchical MIRT model reliably estimated the subtest ability parameter at an acceptable level. In addition, in the majority of the conditions, the subtest scores of the Hierarchical MIRT and the Higher Order model were estimated with similar errors, but the Higher Order model showed better performance at higher levels of correlation.

Based on findings from the study; the use of the Hierarchical CBMTK model is recommended for the reporting of large scale tests. In reporting exams known to have moderate and low correlations among the sub-tests, it may also be preferable to use the Higher Order model, which is able to perform close analyzes with the Hierarchical MIRT Model, as an alternative to the Hierarchical MIRT Model.

Keywords: Subtest scoring, overall test scoring, hierarchical item response theory models, higher order model, Bi-factor model

Advisor: Prof. Dr. Hülya KELECİOĞLU, Hacettepe University, Department of Educational Science, Division of Measurement and Evaluation

İÇİNDEKİLER

KABUL ve ONAY.....	ii
YAYIMLAMA VE FİKRİ MÜLKİYET HAKLARI BEYANI	iii
ETİK BEYANNAMESİ	iv
TEŞEKKÜR.....	v
ÖZ.....	vi
ABSTRACT.....	viii
İÇİNDEKİLER.....	x
TABLolar DİZİNİ	xiii
ŞEKİLLER DİZİNİ.....	xv
SİMGELER VE KISALTMALAR DİZİNİ	xvi
1. GİRİŞ 1	
1.1.Problem Durumu.....	1
1.2. Araştırmanın Amacı ve Önemi.....	4
1.3. Problem	6
1.3.1. Alt Problemler.....	6
1.4 Sınırlılıklar	7
1.5. Araştırmanın Kuramsal Temelleri	7
1.5.1. Alt Test Puanı Kestirim Yöntemleri.....	7
1.5.2. Araştırmada Kullanılan Hiyerarşik MTK Modelleri	9
1.5.2.1. Hiyerarşik Çok Boyutlu Madde Tepki Kuramı (Hiyerarşik ÇBMTK)10	
1.5.2.2. Üst Düzey Sıralı Madde Tepki Kuramı Modeli (Higher Order Item Response Theory Model)	11
1.5.2.3. İki Faktörlü Madde Tepki Kuramı Modeli (Bi-factor Item Response Theory Model)	12
2. İLGİLİ ARAŞTIRMALAR.....	14
2.5. İlgili Araştırmalar Özet	23
3. YÖNTEM	25
3.1. Araştırmanın Türü	25
3.2. Simülasyon Koşulları	25
3.2.1. Çalışma Grubu Büyüklüğü	25
3.2.2. Alt Test Sayısı	26
3.2.3. Alt Test Uzunluğu	26
3.2.4. Alt Testler Arasındaki Korelasyon	27
3.2.5. İterasyon Sayısı.....	27
3.3. Verilerin Üretilmesi.....	28
3.4. Üretilen Verilerin Geçerliğinin Test Edilmesi	30
3.5. Değerlendirme Kriteri	30
3.6. Verilerin Analizi	31
4. BULGULAR VE TARTIŞMA	32
4.1. Birinci Alt Probleme İlişkin Bulgular ve Tartışma.....	32

4.1.1. İki Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri.....	32
4.1.2. Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri.....	35
4.1.3. İki Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri	39
4.1.4. Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri	42
4.1.5. Birinci Alt Probleme İlişkin Tartışma	46
4.2. İkinci Alt Probleme İlişkin Bulgular ve Tartışma	47
4.2.1. İki Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri.....	48
4.2.2. Üç Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri.....	51
4.2.3. İki Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri	55
4.2.4. Üç Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri	59
4.2.5. İkinci Alt Probleme İlişkin Tartışma.....	63
4.3. Üçüncü Alt Probleme İlişkin Bulgular ve Tartışma	64
4.3.1. Toplam Test Puanlarına ait RMSE Değerleri İçin Varyans Analizi Sonuçları	65
4.3.2. Toplam Test Puanlarına ait Güvenirlik Değerleri İçin Varyans Analizi Sonuçları	66
4.3.3. Alt Test Puanlarına ait RMSE Değerleri İçin Varyans Analizi Sonuçları	68
4.3.4. Alt Test Puanlarına ait Güvenirlik Değerleri İçin Varyans Analizi Sonuçları	70
4.3.5. Üçüncü Alt Probleme İlişkin Tartışma.....	72
4.4. Dördüncü Alt Probleme İlişkin Bulgular ve Tartışma.....	73
5. SONUÇ ve ÖNERİLER.....	75
5.1. Sonuçlar.....	75
5.1.1. Birinci Alt Probleme İlişkin Sonuçlar	75
5.1.2. İkinci Alt Probleme İlişkin Sonuçlar.....	76
5.1.3. Üçüncü Alt Probleme İlişkin Sonuçlar.....	78
5.1.4. Dördüncü Alt Probleme İlişkin Sonuçlar	80
5.2. Öneriler.....	80
5.2.1. Uygulamaya Dönük Öneriler	80
5.2.2. Araştırmaya Dönük Öneriler.....	81
KAYNAKÇA.....	82
EKLER DİZİNİ	89
EK 1. ETİK KOMİSYONU ONAY BİLDİRİMİ	89
EK-2. TEOG 2015 Verisinin Faktör Analizi Sonuçları	91
EK-3. SimuMIRT Veri Üretme Kodları	94
EK-4. Örnek Simülasyon Verisinin Faktör Analizi Sonuçları	95
EK-5. Madde Tepki Kuramı Modellerine Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerlerine İlişkin Karşılaştırılmalı Grafikler.....	97

EK-6. Madde Tepki Kuramı Modellerine Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerlerine İlişkin Karşılaştırılmalı Grafikler	99
EK-7. İkinci Alt Probleme Ait RMSE Değerleri	101
EK-8. Madde Tepki Kuramı Modellerine Göre İki ve Üç Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerlerine İlişkin Karşılaştırılmalı Grafikler	102
EK-9. İkinci Alt Probleme Ait Güvenirlik Değerleri.....	104
EK-10. Madde Tepki Kuramı Modellerine Göre İki ve Üç Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerlerine İlişkin Karşılaştırılmalı Grafikler	105
EK-11. Gerçek Veri Setinde Modellerin Kestirdiği Alt Testler Arası Korelasyonların Standart Sapma Matrisi	107
EK-12. ORJİNALLİK RAPORU	108
EK-13. ÖZGEÇMİŞ	110

TABLolar DİZİNİ

Tablo 1.1: Alt Test Puan Kestirim Yöntemlerinin Bir Özeti	8
Tablo 3.1: Simülasyon Koşulları ve Düzeyleri	27
Tablo 3.2: TEOG 2015 Sınavının Alt Testlerine Ait Madde Parametreleri.....	29
Tablo 4.1.: İki Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri.....	32
Tablo 4.2.: Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri.....	36
Tablo 4.3.: İki Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri	39
Tablo 4.4.: Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri	43
Tablo 4.5.: İki Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri.....	48
Tablo 4.6.: Üç Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri.....	51
Tablo 4.7.: İki Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri	55
Tablo 4.8.: Üç Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri	59
Tablo 4.9: Toplam Test Puanlarına Ait RMSE Değerleri İçin Varyans Analizi Sonuçları	65
Tablo 4.10: Toplam Test Puanlarına Ait Güvenirlik Değerleri İçin Varyans Analizi Sonuçları.....	67
Tablo 4.11: Alt Test Puanlarına Ait RMSE Değerleri İçin Varyans Analizi Sonuçları	68
Tablo 4.12: Alt Test Puanlarına Ait Güvenirlik Değerleri İçin Varyans Analizi Sonuçları	70
Tablo 4.13: Sonsal Dağılımın Ortalama ve Standart Sapması	73
Tablo 4.14: Modellerin Kestirdiği ve TEOG Verisinin Alt Testler Arası Korelasyon Matrisi.....	74
Tablo EK-7.1.: İki Boyutlu Veri Setlerinden Her Bir Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri.....	101
Tablo EK-7.2.: Üç Boyutlu Veri Setlerinden Her Bir Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri.....	101
Tablo EK-9.1.: İki Boyutlu Veri Setlerinden Her Bir Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri	104
Tablo EK-9.2.: Üç Boyutlu Veri Setlerinden Her Bir Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri	104

Tablo EK-14.1: Modellerin Kestirdiđi Alt Testler Arası Korelasyonların Standart Sapma Matrisi	107
---	-----

ŞEKİLLER DİZİNİ

Şekil 1.1. Araştırmada Kullanılan Hiyerarşik Çok Boyutlu MTK Modelleri	9
Şekil 4.1. İki Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri	34
Şekil 4.2. Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri	38
Şekil 4.3. İki Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri	42
Şekil 4.4. Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri	45
Şekil 4.5. İki Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri	50
Şekil 4.6. Üç Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri	53
Şekil 4.7. İki Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri	58
Şekil EK-5.1. İki Faktör Modele Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri	97
Şekil EK-5.2. Hiyerarşik ÇBMTK Modele Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri	97
Şekil EK-5.3. Üst Düzey Sıralı Modele Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri	98
Şekil EK-6.1. İki Faktör Modele Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri	99
Şekil EK-6.2. Hiyerarşik ÇBMTK Modele Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri	100
Şekil EK-6.3. Üst Düzey Sıralı Modele Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri	100
Şekil EK-8.1. İki Faktör Modele Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri	102
Şekil EK-8.2. Hiyerarşik ÇBMTK Modele Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri	103
Şekil EK-8.3. Üst Düzey Sıralı Modele Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri	103
Şekil EK-10.1. İki Faktör Modele Göre İki ve Üç Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri	105
Şekil EK-10.2. Hiyerarşik ÇBMTK Modele Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri	106
Şekil EK-10.3. Üst Düzey Sıralı Modele Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri	106

SİMGELER VE KISALTMALAR DİZİNİ

KTK: Klasik Test Kuramı

MTK: Madde Tepki Kuramı

ÇBMTK: Çok Boyutlu Madde Tepki Kuramı

TEOG: Temel Düzeyden Orta Düzeye Geçiş

MCMC: Markov Zinciri Monte Carlo

BS: Beklenen Sonsal (Expected a Priori)

MS: Maksimum Sonsal (Maximum a Posteriori)

EÇO: En Çok Olabilirlik (Maximum Likelihood)

1. GİRİŞ

Bu bölümde araştırmaya ait problem durumu, amaç ve önem, problem cümlesi, alt problemler, sayılılar, tanımlar ve sınırlılıklar üzerinde durulmuştur.

1.1.Problem Durumu

Pek çok gelişmiş ülkede, geniş ölçekli standart testler, eğitimde ve psikolojide kullanılan en yaygın ölçme araçlarıdır. Ülkemizde bu araçlar, bir eğitim programına giriş, sertifika alımı ya da personel seçimi gibi önemli kararların verildiği durumlarda sıklıkla kullanılmaktadır. Birçok ülke geniş ölçekli testleri kendi eğitim sistemlerine yönelik bilgiler toplama, eğitime yönelik karar verme ve planlama aşamalarında sıkça kullanmaktadır. SAT ve ACT gibi Amerika'da yapılan geniş ölçekli sınavların burs verme ve eyaletlerin eğitim politikalarının değerlendirilmesi gibi ikincil amaçları vardır. Öğrenci gelişimlerinin izlenmesinde, okulların başarı durumlarının/performanslarının yıllara göre incelenmesinde, öğretim programlarının araştırılması, değerlendirilmesi ve geliştirilmesinde geniş ölçekli testlerden yararlanılmaktadır. Bu testlerin sonuçlarından öğrenciler, öğretmenler, veliler, yöneticiler ve diğer paydaşlar farklı şekillerde faydalanmaktadır.

Geniş ölçekli sınavlar, genellikle hem farklı yapıları hem de bir yapının farklı alanlarını ölçen alt bölümlerden oluşur. Bu alt bölümlere genellikle alt test, alt bölümlerden elde edilen puanlara da alt test puanı denir. Örneğin, KPSS sınavındaki eğitim bilimleri alt testi, kendi içinde sekiz konu alanına ayrılır (öğrenme psikolojisi, gelişim psikolojisi, ölçme ve değerlendirme, rehberlik ve özel eğitim, öğretim ilke ve yöntemleri, program geliştirme, sınıf yönetimi, öğretim teknolojileri ve materyal tasarımı). Genellikle bir alt testteki maddeler bir yeteneği, bir konu alanını ya da bir örtük yapıyı ölçmek için düzenlenirler. Ülkemizde bu testlerin sonuçları alt testlerin ağırlıklı ortalamalarından elde edilen toplam puanlar ile ifade edilir. Böyle birleşik puanlar genel olarak birey başarısını değerlendirmek için yeterli bilgiyi sağlayabilirler. Çünkü geniş ölçekli testlerin en başta oluşturulma amacı bireyleri sıralamaktır.

ABD'de 2001 yılında kabul edilen Hiçbir Çocuk Geride Kalmasın (No Child Left Behind) Yasası gereğince her eyalet okul ilerleme durumunu ölçmek amacıyla toplam test puanı yanında öğrencilerin matematik, okuma, yazma, fen bilimleri gibi

temel konu alanlarındaki puanlarını da raporlaması gerekmektedir. Brennan (2012) test puanlarını kullananların çoğunlukla toplam test puanıyla birlikte alt testlerin tanısallama amaçları için raporlanmasını talep ettiğini belirtmiştir. Ayrıca Haladyna ve Kramer (2005), testin başlangıçtaki temel amacı ne olursa olsun, öğrenciler, öğretmenler, veliler, yöneticiler ve diğer paydaşlar tarafından farklı alt alanlar ya da alt bölümlere ait puanların büyük talep gördüğünü raporlamışlardır. (akt. Ling, 2012).

Son yıllarda tanı amaçlı değerlendirmelere artan ilgi düşünüldüğünde, sıralama amaçlı yapılan bir testin spesifik alanları ya da alt bölümleri de tanı amacıyla kullanılabilirler. Geniş ölçekli testlerde olduğu gibi sıralama amaçlı bir test testin herhangi bir alt kümesine odaklanmadan geçerlik ve adalet açısından geniş bir materyal yelpazesi kapsarken tanısallama amaçlı bir test spesifik bir alana yoğunlaşmaktadır. Uygun istatistiksel süreçler altında bir testin sıralama ve tanı açısından birbiriyle çelişen bu iki amaca hizmet etmesi sağlanabilmektedir.

Alt testler, formatif (biçimlendirici) ve summatif (özetleyici) değerlendirmelere, eğitim programlarının değerlendirilmesine ve öğretmen değerlendirmelerine bilgi sağlayabilecek bir potansiyele sahiptir. Aynı şekilde, alt testler, toplam puanla karşılaştırıldığında, bireylerin yeteneklerinin farklı alanlarda nasıl değiştiğini/çeşitlendiğini belirlemek için daha bilgilendirici olabilmektedir. Bilişsel bir yapıyı, bir yeteneği ya da psikolojik bir yapıyı temsil eden ve bunlara yönelik tanılayıcı bilgiler sağlayan alt testler, sınıf içi ve dışı etkinliklerin düzenlenmesinde de yararlı olabilmektedir. Testlerde başarısız olan bireyler, testin kapsamında yer alan konu alanları, yeterli alanları ya da bilişsel yapılar içinde başarılı ve başarısız oldukları noktaları bilmek istemektedir. Böylece bireyler, çalışma planlarını eksik ya da zayıf oldukları konuları tamamlayabilmek için daha etkili şekilde düzenleyebilme imkanı bulmaktadır (Haladyna & Kramer, 2004). Ayrıca alt testler öğrencilerin güçlü ve zayıf oldukları noktalar hakkında bilgi sağlayarak öğretmenlerin ders programlarını düzenlemesine katkı sağlayabilmektedir. Yine alt testlerin sağladığı bilgilerle veliler çocuklarının durumları ile ilgili bilgilendirirken, onların eksik veya başarısız oldukları konular için destekleyici tedbirler alma ya da onların potansiyellerine göre yönlendirici imkanlar sağlama konusunda daha etkili çözüm üretebilmektedir.

Alt testlerden elde edilebilecek bilgilerden yararlanabilmek için öncelikle alt test puanlarının test geliştirme süreçleri açısından bazı önemli özellikleri sağlaması gerekmektedir. (ETS, 2014; Ferrara & DeMauro, 2007). İlk olarak, alt test puanları

güvenirlik, geçerlik, ayırt edicilik açısından yeterli psikometrik niteliklere sahip olmalıdır. Psikolojide ve Eğitimde Test Geliştirme Standartları 5.12'ye (AERA, APA, NCME,1999) göre test puanlarının geçerliği, güvenilirliği ve karşılaştırılabilirliği sağlanmadıkça raporlanmaması gerekir. Yine aynı standartların 1.12'ye göre, bir test birden fazla puan sağlıyorsa farklı puanların ayıricılıklarının gösterilmesi gerekir. Benzer şekilde, Ferrara ve DeMauro (2007) orta düzeyde ilişkili ve yüksek güvenilirliğe sahip alt testlerin raporlanmasını, düşük güvenilirlikli alt testlerin ise raporlanmaması gerektiğini belirtmişlerdir. Alt testlerin güvenilirliğine ek olarak testin yapısının da incelenmesi gerekir. Messick'e (1989, s.43) göre maddeler arası ilişkiler yapının alt testlerini ya da alt alanlarını yansıtmalı ve bu da test puanları ve onların yorumlanması düzeyinde ele alınmalıdır. İkinci olarak, Haberman (2008) ve Haberman, Sinharay ve Puhan (2009) alt test puanlarının toplam puan üzerinde bir değeri olup olmadığının belirlenmesi gerektiğini belirtmişlerdir.

Alt test puan ya da birleşik toplam puan kestirimlerinde geleneksel Klasik Test Kuramı'na (KTK) dayalı toplam puanı ya da düzeltilmiş toplam puanı kullanan yöntemler bulunmaktadır (Kelley, 1927, 1947; Wainer, Vevea, Camacho, Reeve, Rosa, Nelson, Swygert & Thissen, 2001). KTK'nın madde ve birey örnekleme bağlı olması nedeniyle daha güvenilir alt test puan kestirimlerinde tek boyutlu Madde Tepki Kuramı'na (MTK) dayalı yöntemler geliştirilmiştir (Wainer & ark 2001;Yen, 1987). Her ne kadar alt test puanı kestirimlerinde tek boyutlu MTK'ya dayalı yöntemlerin KTK'ya dayalı yöntemlere göre daha güvenilir sonuçlar verdiği gösterilse de tek boyutlu yaklaşımlar alt testler arası ilişkileri göz ardı etmektedir. Geniş ölçekli testlerde toplam puanların alt testlerin tek boyutlu olduğu varsayımı altında ve birbirleri arasındaki ilişkilerden bağımsız olarak analiz edilmesi durumunda eğer test maddeleri yerel bağımlı ise güvenirliliğin ve test parametrelerinin yanlış kestirilmesi beklenen bir sonuçtur (Brandt & Duckor, 2013; Wang & Wilson, 2005;Yen, 1980). Bu sorunun çözümünde alt testlerden oluşan testler için toplam test yetenek kestirimlerinde MTK 'ya dayalı çeşitli modellerin kullanıldığı görülmektedir. Alt testlerden kaynaklanan yerel bağımlılığın, test yapısına (madde demetleri kullanımı gibi) veya ölçülen psikolojik yapıya (alt yeterlik alanları gibi) bağlı olup olmadığına göre bu modeller sırasıyla testlet modeller (Bradlow, Wainer & Wang, 1999; Wang & Wilson, 2005) ya da hiyerarşik modeller

(de la Torre & Song, 2009; Gibbons & Hedeker, 1992; Sheng & Wikle, 2008) olarak gösterilmektedir.

Alt testleri çok boyutlu MTK çerçevesinde analiz eden birkaç çalışma bulunmaktadır (de la Torre, Song & Hong, 2011; Sheng & Wikle, 2007; Wang, Chen & Cheng, 2004; Yao, 2010; Yao & Boughton, 2007). Çok boyutlu MTK'nın alt test puanı kestirimlerinde kullanılması konusunda daha fazla araştırmaya ihtiyaç olduğu düşünülmektedir. Alt testlerden oluşan ölçme araçlarından elde edilen puanların güvenilirliği; ölçme aracının ölçtüğü teorik yapıların açıklanması ya da bileşke yapıya ilişkin kurulan modelin test edilmesi gibi yapı geçerliği; sınıflama geçerliği; çapraz geçerliği gibi ölçme araçlarının psikometrik özellikleri üzerinde MTK'ya dayalı yöntemlerin performansının daha fazla incelenmesi gerektiği düşünülmektedir. Belirlenen bu ihtiyaca bağlı olarak, ölçme araçlarının psikometrik özelliklerinden güvenilirlik üzerinde çalışılmaya karar verilmiştir. Güvenilir alt test ve toplam test puanı kestirimleri konusuna katkı sağlamak amacıyla alt test ve toplam test arasındaki ilişki hiyerarşik MTK modelleri ile araştırılmak istenmiştir.

1.2. Araştırmanın Amacı ve Önemi

Ülkemizde yapılan geniş ölçekli sınavlarda olduğu gibi PISA, TIMSS, PIRLS gibi uluslararası geniş ölçekli sınavlarda da çoğu zaman politikacıların, eğitimcilerin veya velilerin daha çok ülke performansı/sıralaması üzerinde durduğu ve testlerin içeriğindeki alt ölçeklerin, alt yeterlik alanlarının ya da alt konuların üzerinde çok fazla durulmadığı dikkat çekmektedir. Oysaki eğitim araştırmacıları ve uygulayıcıları için böyle uluslararası geniş ölçekli testlerin anlamlı daha küçük alt birimlerinden yola çıkarak yapılacak çok boyutlu analiz sonuçlarının potansiyel olarak öğrencilerin öğrenme şeklinin ve dolayısıyla öğrenme çıktılarının belirlenmesi açısından daha anlamlı ve otantik olacağı söylenebilir. Bu bağlamda araştırmanın temel amacı alt testlerden elde edilebilecek tanısal bilgilere dikkat çekmek ve güvenilir alt test puanı kestirimlerine katkı sağlamaktır.

Bir testin alt birimlerinden (alt yeterlik alanları, alt testler, alt ölçekler vb) elde edilen puanların genel olarak zayıf psikometrik özelliklere sahip olduğu belirtilmiştir (Monaghan, 2006; Skorupski & Carvajal, 2010). Alt birimler testin toplamına göre daha az sayıda madde içermesi nedeniyle daha düşük güvenilirliğe sahip olabilmektedir. Sinharay (2010) yetersiz test uzunluğuna sahip olması nedeniyle alt

testlerden alınan puanların güvenilir olmasa dahi potansiyel tanı değerleri nedeniyle raporlanmasının yararlı olabileceğini belirtmesine rağmen güvenilir alt test puanlarının raporlanması gerekli ve önemlidir (Haberman, 2008). Bu durumda, “Az sayıda madde içermesi nedeniyle bir testin daha küçük alt birimlerinden güvenilir puanlar elde edilebilir mi veya alt testler arası korelasyon, alt test uzunluğu gibi test özelliklerinin alt test puanı kestirimleri üzerindeki etkisi nedir?” gibi sorular ortaya çıkmaktadır.

Genellikle bu amaç için tasarlanmamış testlerden güvenilir alt puanlar elde etme zorunluğu göz önüne alındığında, alt puan kestirimlerinin kalitesini arttırmak için daha güvenilir ve geçerli bir puanlama mekanizması sağlayabilen puanlama yöntemleri kullanmak gerekli ve önemlidir. Çok boyutlu MTK'nın tek boyutlu MTK ve KTK'ya dayalı yöntemlerle karşılaştırıldığında alt test puanı kestirimlerinde daha küçük yanlılık ve standart hata ile daha yüksek güvenilirliğe sahip yetenek tahminleri sağladığı bulunmuştur (Boughton, 2007; de la Torre & Patz, 2005, de la Torre, 2009; Sheng & Wikle, 2007,2008; Thissen & Edwards, 2005; Yao & Sheng, 2005; Yao, 2010). Bu nedenle bu araştırmada çok boyutlu MTK'nın sağladığı avantajlardan yola çıkarak alt test ve toplam test arasındaki ilişki hiyerarşik MTK modelleri çerçevesinde ele alınmış ve incelenmiştir.

Ülkemizde uygulanan geniş ölçekli sınavlara ait toplam puanlar, alt testlerin ağırlıklı ortalamaları olarak ifade edilmektedir. Bu sınavların test yapısını çok boyutlu ve hiyerarşik modeller çerçevesinde ele alınmasının ve analiz etmenin alt test ve toplam test puan kestirimlerinin doğruluğuna ve güvenilirliğine etkisinin nasıl olacağı araştırılmak istenmiştir. Bu amaçla TEOG veri setinin özelliklerine göre güvenilirliğe etki edebilecek faktörlerden alt test uzunluğu, alt test sayısı ve alt testler arasındaki korelasyonların büyüklüğü koşulları altında üretilen veriler üzerinde hiyerarşik çok boyutlu MTK Model, İki faktör Model ve Üst Düzey Sıralı Model'in alt test ve toplam test puanı kestirimleri üzerindeki performansı incelenmiştir.

Alan yazında madde parametresi doğrulama çalışmalarına göre alt test puanı ve toplam test puanı kestirimlerinde çok boyutlu yaklaşımları ve özellikle hiyerarşik modelleri kullanan sınırlı sayıda çalışma vardır (Chen, West & Souza, 2006; de la Torre, Song & Hong, 2011; Sheng & Wikle, 2009; Wang, Chen , & Cheng, 2004; Yao & Boughton, 2007; Yao, 2010). Ülkemizde yapılan bilimsel araştırmalarda ise çok boyutlu yetenek parametresi kestirimleri veya alt testlere yönelik çalışmaların

oldukça az olduğu dikkat çekmektedir (Çakıcı Eser, 2015; Köse, 2010; Özkan, 2012). Özellikle, son yıllarda tanılayıcı ya da sınıflandırıcı modellere olan ilginin arttığı düşünüldüğünde, alt testlerin tanısallık potansiyellerine dikkat çeken çalışmalara ihtiyaç olduğu düşünülmüştür. Bu durum bu çalışmayı önemli kılan noktalardan biridir. Alt testlerden oluşan geniş ölçekli testlere ilişkin hem genel hem de alt test puanı kestirimlerine ülkemizde kullanılan klasik yöntemlerden farklı bir bakış açısı getirmesi bu çalışmanın en önemli özelliğidir.

1.3. Problem

Farklı test koşullarında türetilen veriler ile gerçek verilerin toplam test puanı ve alt test puanları kestirimlerinin doğruluğu ve güvenilirlikleri çok boyutlu hiyerarşik MTK modellerine göre nasıl değişmektedir?

1.3.1. Alt Problemler

1. Üst Düzey Sıralı (Higher Order), İki Faktör (Bi-factor) ve hiyerarşik çok boyutlu madde tepki kuramı modellerine göre kestirilen toplam test puanının RMSE ve güvenilirlik değerleri

- a. alt test sayısına (2,3,)
- b. alt test uzunluğuna (20,30,40)
- c. alt testler arasındaki korelasyonların büyüklüğüne (0.0,0.3, 0.5, 0.8)

göre nasıl değişmektedir?

2. Üst Düzey Sıralı (Higher Order), İki Faktör (Bi-factor) ve hiyerarşik çok boyutlu madde tepki kuramı modellerine göre kestirilen alt test puanlarının RMSE ve güvenilirlik değerleri

- a. alt test sayısına (2,3)
- b. alt test uzunluğuna (20,30,40)
- c. alt testler arasındaki korelasyonların büyüklüğüne (0.0,0.3, 0.5, 0.8)

göre nasıl değişmektedir?

3. Üst Düzey Sıralı (Higher Order), İki Faktör (Bi-factor) ve hiyerarşik çok boyutlu madde tepki kuramı modellerine göre kestirilen alt test ve toplam test puanlarının RMSE değerleri ve güvenilirlikleri üzerinde alt test sayısı, alt test uzunluğu ve alt testler arası korelasyon koşullarının etkisi nasıldır?

4. TEOG (2015) verilerinin Üst Düzey Sıralı (Higher Order), İki Faktör (Bi-factor) ve hiyerarşik çok boyutlu madde tepki kuramı modellerine göre alt test puan kestirimleri nasıl değişmektedir?

1.4 Sınırlılıklar

Bu araştırmanın simülatif verileri 29 Nisan 2015'te yapılan TEOG sınavının parametrelerine dayalı olarak üretilmiştir. TEOG sınavının alt testleri ayrı oturumlarda adaylara sunulmaktadır. Araştırmada aynı oturumda cevaplanan alt testlere sahip gerçek bir veri seti kullanılmak istenmiş fakat ÖSYM sınavları verilerine erişim sağlanamadığı için ülkemizde sınırlı olarak erişime açık olan TEOG sınavı parametreleri kullanılmıştır. Bu durum araştırmanın sınırlılığını oluşturmaktadır.

1.5. Araştırmanın Kuramsal Temelleri

1.5.1. Alt Test Puanı Kestirim Yöntemleri

Crocker ve Algina (1986,s.87-100) birleşik toplam puanı oluşturan alt testlerin "mini test" olarak kabul edilerek analiz edilebileceğini belirtmiştir. Bu yaklaşım geleneksel KTK'ya dayalı toplam doğru sayısı, doğru yüzdesi veya yüzdelik sıra bilgilerini kullanarak alt test puan kestirimlerini içermektedir. Hambleton ve Swaminathan (1985, s.53-72) alt yetenek kestirimlerini bağlı oldukları maddeler kullanılarak tek boyutlu madde tepki kuramı modellerine göre hesaplanabileceğini belirtmiştir.

Alt testlerin toplam teste göre daha az madde içermesi nedeniyle oldukça düşük güvenilirlik düzeyinde puanlar ürettiği belirtilmiştir (Green, Yen & Burket, 1989; Skoropski & Carvajal, 2010; Yen, 1980). Bu probleme çeşitli yaklaşımlarla çözüm aranmıştır. Yen (1987) tarafından bir bireyin gerçek alt test puan kestirimlerini elde etmek için birey toplam test puanı ile birey alt test/ölçek cevaplarından gelen bilgiyi birleştiren MTK'ya dayalı bir yöntem önerilmiştir. Konu Performans İndeksi (Objective Performans Index) olarak adlandırılan bu yöntemde her bir alt test puanı kestiriminde, bireyin testin geri kalanına verdiği cevaplar bir birim sayılır. Bayes MTK'ya dayalı alt yetenek kestirimlerinde olasılık fonksiyonu kullanılırken aynı zamanda sonsal dağılıma ulaşmak için önsel bir dağılım kullanılmaktadır. Bu yöntem için kullanılan önsel dağılım alt testten geriye kalan maddelere verilen

yanıtlardan elde edilmektedir. Ayrıca bu yaklaşımda önsel dağılımın binom dağılım, sonsal dağılımın ise beta dağılım gösterdiği varsayılmaktadır.

Tablo 1.1: Alt Test Puan Kestirim Yöntemlerinin Bir Özeti

<i>Yaklaşım</i>	<i>Yöntem</i>	<i>İşlem</i>
Ağırlıklandırılmamış	Geleneksel KTK	Her bir alt test mini bir test gibi kabul edilir. Alt test puanı kestirimlerinde toplam doğru sayısı, düzeltilmiş toplam doğru sayısı, doğru sayısı oranı veya yüzdelik sıra bilgisi kullanılır.
	Tek Boyutlu MTK	Her bir alt test puan kestirimi yalnızca alt testlerin bağlı olduğu maddelere dayalı olarak tek boyutlu kalibrasyonlar ile hesaplanır.
Yardımcı bilgiler kullanılarak ağırlıklandırılmış (Augmented)	Önsel bilgileri kullanan tek boyutlu Bayes MTK	Her bir alt test puan kestirimleri testin geri kalanındaki maddelerden kestirilen yetenek puanları kullanılarak elde edilen önsel bir dağılım ile birleştirilerek kestirilir.
	Regresyona dayalı tek boyutlu KTK veya MTK	Gözlenen alt test puanlarının diğer puanlar üzerinde lineer bir fonksiyonu vardır. Gözlenen puanlar KTK'ya dayalı toplam puanlar olabileceği gibi MTK'ya dayalı yetenek kestirimleri de olabilir.
	Çok Boyutlu MTK	Veri çok boyutlu olarak ele alınır. Alt testler arası ilişkiler kullanılarak ilgilenilen her bir alt ölçek için bir yetenek kestirim vektörü hesaplanır. Alt testler arası ilişkiler eğik ya da dik olabilir.

Kaynak:Skorupski & Carvajal (2010).

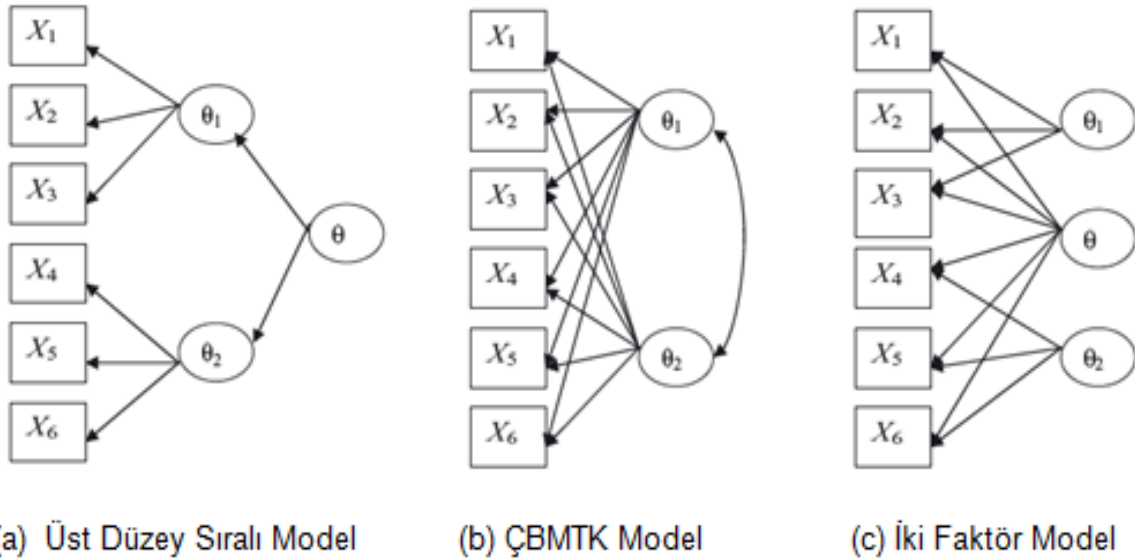
Wainer ve ark. (2000) ve Wainer ve ark. (2001) alt testlerin toplam test göre daha az bilgi sağlasa da bir bireyin alt test performansı ile toplam test performansı arasında yüksek düzeyde ilişki olduğunu belirtmişler ve alt test puan kestirimlerinde regresyona dayalı bir yöntem önermişlerdir. Kelley (1927, 1947) tarafından gerçek puan kestirimlerinde kullanılan regresyon denklemi Wainer ve ark. (2001) tarafından alt test puan kestirimlerinde kullanılmak üzere genişletilmiş ve Ağırlıklandırılmış Puan Yöntemi (Subscore Augmentation) olarak adlandırılmıştır. Bu yaklaşımda bir alt test puanının diğer alt testler üzerindeki regresyonuna dayalı olarak alt test puanı kestirimleri yapılmaktadır. Ayrıca bu yaklaşımda hem MTK hem de geleneksel ham puanlara dayalı kestirimler yapılabilmektedir. Alt test puanı kestirimlerinde

yardımcı/tamamlayıcı (collateral) bilgilerden yararlanması açısından Yen'in (1987) yaklaşımı ile Wainer ve ark.'nın (2001) yaklaşımı benzese de iki yöntem birkaç noktada farklılık göstermektedir. Yen (1987) alt test puanı kestirimlerinde toplam puanı yardımcı değişken olarak kullanırken Wainer ve ark. (2001) alt test puanlarını kullanmaktadır.

Alt test puan kestirimlerinde yardımcı bilgileri kullanan ve son yıllarda KTK ve tek boyutlu MTK'ya dayalı üstünlüğü gösterilen bir diğer yaklaşım çok boyutlu MTK modelleridir. Test dışından ya da alt testler arası korelasyon gibi test içi bilgileri kullanılarak alt test ve toplam test puanları daha yüksek güvenilirlik ve daha düşük hata ile kestirebilmektedir (de la Torre Song 2008; DeMars, 2006; Sheng & Wikle, 2007; Tao, 2009; Tate, 2004; Yao & Boughton, 2007). Alt test puan kestirim yöntemlerinin genel bir özeti Tablo1.1'de sunulmuştur.

1.5.2. Araştırmada Kullanılan Hiyerarşik MTK Modelleri

Araştırmada kullanılan çok boyutlu madde tepki kuramı modelleri alt testler ile toplam test arasındaki ilişkiyi hiyerarşik düzeyde ele alması nedeniyle "hiyerarşik MTK modelleri" olarak adlandırılmıştır. Modellerin yapısal gösterimleri Şekil1.1'de sunulmuştur.



Şekil 1.1. Araştırmada Kullanılan Hiyerarşik Çok Boyutlu MTK Modelleri

1.5.2.1. Hiyerarşik Çok Boyutlu Madde Tepki Kuramı (Hiyerarşik ÇBMTK)

Araştırmada Hiyerarşik ÇBMTK Model için alt testlerin madde ve yetenek parametre kestirimlerinde matematiksel formülü Reckase (1997) tarafından ifade edilen çok boyutlu üç parametrelili lojistik model kullanılmıştır. Bir bireyin j maddesine cevap verme olasılığı Reckase (1987) tarafından aşağıdaki gibi ifade edilmiştir:

$$P_{ij1} = P\{X_{ij} = 1 | \vec{\theta}_i, \vec{\beta}_{2j}, \beta_{1j}, \beta_{3j}\} = \beta_{3j} + (1 - \beta_{3j}) \frac{\exp(\sum_{d=1}^D \beta_{2jd} \theta_{id} + \beta_{1j})}{1 + \exp(\sum_{d=1}^D \beta_{2jd} \theta_{id} + \beta_{1j})} \quad (1)$$

Burada $\vec{\theta}_i = (\theta_{i1}, \theta_{i2}, \dots, \theta_{iD})$, bireyin yetenek vektörünü; $\vec{\beta}_{2j} = (\beta_{2j1}, \beta_{2j2}, \dots, \beta_{2jD})$, çoklu ayırt edicilik parametrelerinin vektörünü, β_{1j} , güçlük parametresini, β_{3j} , düşük asimptot ya da diğer adıyla şans parametresini ve D boyut indisini göstermektedir.

Araştırma kapsamında üretilen verilerin basit yapılı olması nedeniyle yalnızca bir boyutta ayırt edicilik parametresine sahip olduğu için bu formül aşağıdaki gibi sadeleştirilerek ifade edilebilmektedir:

$$P_{ij1} = P\{X_{ij} = 1 | \theta_{id}, \beta_{2jd}, \beta_{1j}, \beta_{3j}\} = \beta_{3j} + (1 - \beta_{3j}) \frac{\exp(\beta_{2jd} \theta_{id} + \beta_{1j})}{1 + \exp(\beta_{2jd} \theta_{id} + \beta_{1j})} \quad (2)$$

Hiyerarşik ÇBMTK Model'de toplam test puanı için yetenek kestirimleri Yao ve Schwarz (2006) tarafından tanımlanan *Maksimum Bilgi Yöntemi* ile elde edilmiştir. Yao ve Schwarz (2006) 3PL modele göre j maddesi için $\vec{\theta}$ düzeyinde bilgi fonksiyonunu aşağıdaki eşitlikle tanımlamışlardır:

$$I_j(\vec{\theta}) = \frac{(P_{j1} - \beta_{3j})^2 (1 - P_{j1})}{P_{j1} (1 - \beta_{3j})^2} \vec{\beta}_{2j} \otimes \vec{\beta}_{2j} \quad (3)$$

Burada P_{j1} Eşitlik 1'in i indisi haricinde $\vec{\theta}$ düzeyindeki olasılık ifadesini; $\vec{\beta}_{2j} \otimes \vec{\beta}_{2j}$ ise DxD elemanlı matrisin dış elemanlarını göstermektedir. Yukarıdaki eşitlikten yola çıkarak test bilgi fonksiyonu Eşitlik 4 ile hesaplanmaktadır.

$$I(\vec{\theta}) = \sum_{j=1}^J I_j(\vec{\theta}) \quad (4)$$

$$\theta_\alpha = \sum_{d=1}^D \theta_d \omega_d \quad (5)$$

$$V(\theta_\alpha) = \vec{\omega} V(\theta) \vec{\omega}^{-1} \quad (6)$$

$$\vec{\omega} = (\cos^2 \alpha_1, \cos^2 \alpha_2, \dots, \cos^2 \alpha_D) \quad (7)$$

Maksimum test bilgi yöntemi ile mümkün olan tüm açı değerleri için Eşitlik 6'da verilen varyans değerini en küçük yapacak açı değeri hesaplanmaktadır. Böylece

maksimum bilgiye sahip en güvenilir θ_α -birleşik (composite) puan- elde edilmektedir. Sonuç olarak bu yöntem ile Hiyerarşik ÇBMTK modelde toplam/genel test puanı ile alt test puanları arasında lineer bir ilişki kurulmaz. Aksine toplam puan ve alt test puanları arasındaki ilişkilerin farklı yetenek düzeylerinde veya farklı puan düzeylerinde farklılaşabileceği gerçeğini dikkate alınarak toplam test puanları elde edilmektedir (Yao, 2010). Modelin yapısal gösterimi Şekil1.1b'de verilmiştir.

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} \quad \Lambda_y = \begin{bmatrix} \lambda_{1,1} & 0 \\ \lambda_{2,1} & 0 \\ \lambda_{3,1} & 0 \\ 0 & \lambda_{4,2} \\ 0 & \lambda_{5,2} \\ 0 & \lambda_{6,2} \end{bmatrix} \quad \eta = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix} \quad (8)$$

$$Y = \Lambda_y \eta + \epsilon \quad (9)$$

Şekilde 1.1b'deki modelin faktör analitik yapısı Eşitlik 9 ile gösterilmektedir. Eşitlik 8'de verilen Y vektörü gözlenen değişkenleri; Λ_y matrisi, gözlenen değişkenlerin alt faktörler üzerindeki faktör yüklerini; η vektörü, alt faktörleri ve ϵ vektörü ise artık varyansı ifade etmektedir.

1.5.2.2. Üst Düzey Sıralı Madde Tepki Kuramı Modeli (Higher Order Item Response Theory Model)

Model hiyerarşik bir Bayesiyen yaklaşımı kullanmaktadır. De la Torre (2011) tarafından modelin matematiksel formülü aşağıdaki gibi ifade edilmektedir:

$$\theta_i \sim N(0,1) \quad (10)$$

$$\lambda_d \sim U(-1,1) \quad (11)$$

$$\theta_{i(d)} | (\theta_i, \lambda_d) \sim N(\lambda_d \theta_i, 1 - \lambda_d^2) \quad (12)$$

$$P(\theta, \lambda, \Theta | X) = L(X | \Theta) P(\Theta | \theta, \lambda) P(\theta) P(\lambda) \quad (13)$$

$$\theta_{i(d)} = \lambda_d \theta_i + \eta_{i(d)} \quad (14)$$

Üst Düzey Sıralı MTK modelinde Eşitlik 10'da görüldüğü gibi bireyin tüm teste ait yetenek parametresi olan θ_i 'nin normal dağılımdan geldiği ve $\theta_{i(d)}$ 'nin ise Eşitlik 12'de verilen dağılıma sahip olduğu varsayılmaktadır. λ_d , (-1,1) aralığında değer alabilen d.alt testin ($\theta_{i(d)}$) tüm test üzerindeki (θ_i) regresyon katsayısını gösterirken $\eta_{i(d)}$ ise $N(0, 1 - \lambda_d^2)$ 'e yaklaşan hata terimini göstermektedir. Parametrelerin sonsal

dağılımı ise Eşitlik 13'te verilen koşullu olasılık fonksiyonu ile ifade edilmektedir. Modelde Eşitlik 14'te gösterildiği gibi alt test puanları ile toplam test puanı arasında lineer bir ilişki kurulmaktadır. Bu ilişki toplam test puanı ile alt test puanları arasındaki korelasyonlara dayanmaktadır. Bu yaklaşıma göre alt testler kendi içlerinde tek boyutludur ama bütün alt testler dolaylı olarak genel bir boyutla ilişkilidir. Alan yazında böyle yapılara çoklu-tek boyutlu (multi-unidimensional) test yapıları da denilmektedir (Sheng & Wickle, 2007). Modelin yapısal gösterimi Şekil1.1a'da gösterilmiştir.

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \end{bmatrix} \quad \Lambda_y = \begin{bmatrix} \lambda_{1,1} & 0 \\ \lambda_{2,1} & 0 \\ \lambda_{3,1} & 0 \\ 0 & \lambda_{4,2} \\ 0 & \lambda_{5,2} \\ 0 & \lambda_{6,2} \end{bmatrix} \quad \eta = \begin{bmatrix} \eta_1 \\ \eta_2 \end{bmatrix} \quad \Gamma = \begin{bmatrix} \Gamma_{1,1} \\ \Gamma_{2,1} \end{bmatrix} = \begin{bmatrix} \zeta_1 \\ \zeta_2 \end{bmatrix} \quad \xi = [\xi] \quad \zeta \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix} \quad (15)$$

$$\eta = \Gamma \xi + \zeta \quad (16)$$

$$Y = \Lambda_y \eta + \epsilon \quad (17)$$

Şekilde 1.1a'daki modelin faktör analitik yapısı Eşitlik 16 ve 17 ile gösterilmektedir. Eşitlik 15'de verilen Y vektörü gözlenen değişkenleri; Λ_y matrisi, gözlenen değişkenlerin birinci düzeydeki faktörler üzerindeki faktör yüklerini; η vektörü, birinci düzey faktörleri; Γ vektörü, birinci düzey faktörlerin üst düzeydeki faktör üzerindeki faktör yüklerini; ξ vektörü, üst düzey faktörü; ζ vektörü, birinci düzeydeki faktörlerin tekil varyansını ve ϵ vektörü ise artık varyansı ifade etmektedir.

1.5.2.3. İki Faktörlü Madde Tepki Kuramı Modeli (Bi-factor Item Response Theory Model)

Bu model ilk defa Holzinger ve Swineford (1937) tarafından bir faktör analizi yöntemi olarak tanımlanmıştır. Modelde bir maddenin hem spesifik bir alt boyutla hem de genel bir boyutla ilişkili olduğu varsayılır. Bu modelin parametre kestirimleri Eşitlik 2'deki matematiksel ifadeye dayalı olarak Cai, Yang ve Hansen (2011) tarafından genişletildiği şekliyle kullanılmıştır:

$$P_{ij1} = P\{X_{ij} = 1 | \theta_{ig}, \theta_{id}, \beta_{2jd}, \beta_{1j}, \beta_{3j}\} = \beta_{3j} + (1 - \beta_{3j}) \frac{\exp(\beta_{2jg}\theta_{ig} + \beta_{2jd}\theta_{id} + \beta_{1j})}{1 + \exp(\beta_{2jg}\theta_{ig} + \beta_{2jd}\theta_{id} + \beta_{1j})} \quad (18)$$

Burada Eşitlik 2'ye ek olarak β_{2jg} , j maddesinin genel boyutta ait madde ayırt edicilik parametresini gösterirken θ_{ig} , i bireyinin genel boyuta ait yetenek parametresini göstermektedir.

Bu modelin yapısal gösterimi ise Şekil1.1c'de gösterilmiştir.

$$Y = \begin{bmatrix} y1 \\ y2 \\ y3 \\ y4 \\ y5 \\ y6 \end{bmatrix} \quad \Lambda_y = \begin{bmatrix} \lambda_{1,g} & \lambda_{1,1} & 0 \\ \lambda_{2,g} & \lambda_{2,1} & 0 \\ \lambda_{3,g} & \lambda_{3,1} & 0 \\ \lambda_{4,g} & 0 & \lambda_{4,2} \\ \lambda_{5,g} & 0 & \lambda_{5,2} \\ \lambda_{6,g} & 0 & \lambda_{6,2} \end{bmatrix} \quad \eta = \begin{bmatrix} \eta_g \\ \eta_1 \\ \eta_2 \end{bmatrix} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \epsilon_5 \\ \epsilon_6 \end{bmatrix} \quad (19)$$

$$Y = \Lambda_y \eta + \epsilon \quad (20)$$

Şekilde 1.1c'deki modelin faktör analitik yapısı Eşitlik 20 ile gösterilmektedir. Eşitlik 19'de verilen Y vektörü gözlenen değişkenleri; Λ_y matrisi, gözlenen değişkenlerin alt faktörler üzerindeki faktör yüklerini; η vektörü, alt faktörleri ve ϵ vektörü ise artık varyansı ifade etmektedir.

Ölçme modeli açısından İki faktör Model hem Üst Düzey Sıralı Model hem de ÇBMTK model ile birbirine benzerlik gösterse de modeller arasında önemli farklılıklar vardır. ÇBMTK modelde alt testler arasındaki korelasyon model kestirimlerinde serbest bırakılırken İki faktör modelde genel boyut ile alt testlerin birbirine dik olduğu varsayılmaktadır. Üst Düzey Sıralı ve ÇBMTK modeldeki maddeler genel boyut ile doğrudan ilişkili olmazken İki Faktör modelde maddeler genel boyut ile doğrudan ilişkilidir. Ayrıca İki Faktör modelin temel amacı genel boyuta ilişkin yeteneği kestirmektir. Alt testler ikincil çıktılar olarak kabul edilmektedir ve genel yeteneğin artıklarından açıklanmaktadır. Schmid ve Leiman (1957) tarafından İki Faktör model ile Üst Düzey Sıralı modelin belirli matematiksel sınırlamalar altında eşit olduğu belirtilmiştir.

2. İLGİLİ ARAŞTIRMALAR

Bu bölümde, alt test ve toplam test puanı kestirimleri ile ilişkili alan yazındaki arařtırmalar incelenmiřtir.

Skorupski ve Carvajal (2010), alt test puanlarının gvenirliklerinin geliřtirilmesini konu aldıkları alıřmalarında gerek veri zerinde MTK, Weiner'in Yntemi'ne dayalı MTK ve Yen'in Yntemine dayalı MTK olmak zere  modelin performanslarını 3PL model kullanarak karřılařtırmıřlardır. Sırasıyla 15, 12, 14 ve 11 test uzunluęuna sahip drt alt testin Weiner ve Yen Yntemlerini kullanmadan nce kestirilen gvenirlik deęerlerinin 0.61-0.74 arasında deęiřtięini bu yntemler kullanılarak kestirilen gvenirliklerin ise 0.89-0.91 arasında deęiřtięini belirtmiřlerdir. Ayrıca bu alıřmada, Yen'in Ynteminin ise daha dřk hatalı alt test puan kestirimleri yaptığı bulunmuřtur.

Edwards ve Vevea (2006), alt test puan kestirimlerinde geleneksel toplam puan yntemi ile aęırlıklandırılmıř MTK yntemini karřılařtırmıřlardır. alıřmanın simlasyon kořulları alt testler arası korelasyonlar (0.3, 0.6 ve 0.9), alt test sayısı (2,4) ve alt test uzunluęu (iki boyutlu testler iin 5-40 arasında deęiřen,  boyutlu testler iin 5-30 arasında deęiřen) olarak belirlenmiřtir. Arařtırmanın sonucunda aęırlıklandırma ynteminin geleneksel ynteme gre daha yansız sonular verdięi ve aradaki farkın alt test uzunluęu ve alt testler arası korelasyonlara baęlı olarak deęiřkenlik gsterdięi belirtilmiřtir.

Luecht (2003) toplam 74 maddeden ve yeterlik alanı olarak tanımlanan 4 alt alandan oluřan bir sertifika programı sınavı verilerini kullanarak 3PL modele dayalı drt yntem ile test puanlamaya (scoring) ynelik bir arařtırma yapmıřtır. alıřmasında geleneksel ynteme dayalı elde edilen puanlar (standartlařtırılmıř ham puanlar), toplam test puanının tek boyutlu BS kestirimi, her bir yeterlik alanının tek boyutlu MS kestirimi ile her bir yeterlik alanının ok boyutlu MS kestirimini ortalama standart hatalar ve birey alt test profilleri aısından karřılařtırmıřtır. alıřmanın sonucunda alt test puanlarının tek boyutlu MS kestirimlerinin daha iyi sonu verdięi bulunmuřtur.

Wainer ve ark. (2001) alt test puan kestirimleri iin Kelley'nin (1927, 1947) gerek puan teorisine dayalı regresyon ve Bayesiyen yaklařımlar ile geliřtirdikleri modeli tanıtarak gerek veri setleri zerinden rnek uygulamalar gstermiřlerdir. Alt test

puanlarının ağırlıklandırılması olarak belirttikleri bu yöntemin KTK'ya dayalı elde edilen toplam puanlar ve standartlaştırılmış puanlar ile MTK'ya dayalı ölçek puanları ile uygulanabilirliğini ve etkililiğini tartışmışlardır. İlk olarak, Amerika'da yükseköğretim ve üniversitelere girişte uygulanan geniş ölçekli bir yetenek testi olan SAT üzerinde Kelley'in gerçek puan teorisi ile kendi modellerini geleneksel toplam puanlara dayalı olarak karşılaştırmışlardır. SAT sınavının matematik ve sözel iki alt testinden matematik alt testinin gerçek puan kestirimlerinde sözel teste ait puanları yardımcı bilgi olarak kullanmışlar ve bu yöntemin sonsal dağılımın varyansını genişlettiğinden Kelley'in modelinden daha avantajlı olduğunu belirtmişlerdir. İkinci olarak, Amerikan Üretim ve Stok Kontrol Derneği'nin (APICS) 100 madde ve 6 alt ölçekten oluşan sertifika sınavının 2410 bireye ait verileri kullanmışlardır. Testin alt ölçeklere sahip olmasına karşın tek boyutlu yapıyı desteklediğini kanıtlamışlar ve standart toplam puanlara dayalı olarak geleneksel yaklaşım ile regresyona dayalı ağırlıklandırma yaklaşımının alt test puan kestirimlerini karşılaştırmışlardır. Sonuç olarak, geleneksel standart puan yaklaşımının birey içi (within person) kestirimlerinin yüksek değişkenlik gösterdiğini fakat regresyona dayalı alt test kestirimlerinin ise birbirine yakın değerler alarak tek boyutluluğu desteklediğini göstermişlerdir. Üçüncü olarak, 26 madde ve dört alt testten oluşan Kuzey Carolina Bilgisayar Becerileri Testi'nin kitapçıklarından birine ait 267 kişinin verileri üzerinde MTK'ya dayalı olarak MS ve BS kestirim yöntemlerini kullanarak ağırlıklandırılmış alt test puan kestirimleri gösterilmiştir.

Brandt (2012) ve Brandt ve Duckor (2013) gerçek veri setleri üzerinde tek boyutlu MTK ve Genelleştirilmiş Alt Boyut Model'ini BS kestirim yöntemi kullanarak karşılaştırmışlardır. Modellerden kestirilen yetenek parametrelerinin standart hatalarına göre Genelleştirilmiş Alt Boyut Model'in daha az hatalı kestirimler yaptığı bulunmuştur.

Yao (2010) alt test ve toplam test puanlarının raporlanmasında tek boyutlu MTK, çok boyutlu MTK, Üst Düzey Sıralı Model ve İki faktör Model'in performanslarını hem simülasyon verileri hem de gerçek veri üzerinde karşılaştırmıştır. ÇBMTK ve Üst Düzey Sıralı Model'e göre örneklem büyüklüğü (500, 1000, 2000), alt test sayısı (4), alt test uzunluğu (5, 8, 12) ve alt testler arası korelasyon (0.0, 0.3, 0.5, 0.7, 0.9) koşulları altında basit yapıları üretilmiş ve sonuçlar RMSE, yanlılık, güvenilirlik ve sınıflama oranı istatistikleri ile raporlanmıştır. Çalışmada ÇBMTK ve Üst Düzey

Sıralı Model'in birçok koşulda benzer ve daha yansız sonuçlar verdiğini bulurken İki faktör Model'in alt test puan kestirimlerinde yalnızca alt testler arası korelasyonun düşük olduğu durumlarda, toplam puan kestirimlerin de ise alt testler arası korelasyonun yüksek olduğu durumlarda daha iyi performans gösterdiği bulunmuştur. Gerçek veri uygulamasını, Okuma (46 madde), Dil (34 madde), Matematik (57 madde), Fen (40 madde) ve Sosyal Bilimler (40 madde) alt testlerden oluşan 3953 bireye ait bir test üzerinde gerçekleştirmiştir. Tek boyutlu model kestirimleri gerçek model olarak kabul edilmiş ve ÇBMTK, Üst Düzey Sıralı Model ve İki Faktör Model kestirimleri ile karşılaştırılmıştır. Gerçek veri uygulamasında da ÇBMTK ve Üst Düzey Sıralı Model'in İki Faktör Model'e göre daha iyi performans gösterdiği belirtilmiştir.

De la Torre ve Song (2009) örneklem büyüklüğü (1000, 2000, 4000), alt test sayısı (2, 5), alt test uzunluğu (10, 20, 30) ve alt testler arası korelasyon (0.0, 0.4, 0.7, 0.9) koşulları altında Üst Düzey Sıralı Model'e dayalı olarak veriler üretilmiş ve tek boyutlu 3PL Model ve Üst Düzey Sıralı Model'i alt test ve toplam test puanları kestirimleri açısından karşılaştırmışlardır. Toplam test puanı kestirimlerinde özellikle alt testler arası korelasyon arttıkça iki modelin benzer sonuçlar verdiği bulunmuştur. Alt testler arası korelasyon 0.0 olduğunda tek boyutlu 3PL Model ve Üst Düzey Sıralı Model'in aynı sonuçlar verdiği belirtilmiş ve genel olarak Üst Düzey Sıralı Model'in alt test kestirimlerinde daha etkili olduğu raporlanmıştır. Ayrıca 4 alt testten oluşan 2255 dokuzuncu sınıf öğrencisinin verileri üzerinde iki yöntemin performansı karşılaştırılmış ve Üst Düzey Sıralı Model'in alt test puan kestirimlerinin daha doğru olduğu belirtilmiştir.

Kahraman ve Thompson (2011) alt testlerin geçerlik ve güvenilirliklerinin geliştirilmesi amacıyla tek boyutlu MTK çerçevesinde analitik ve ampirik yaklaşıma dayalı iki yöntem ile gerçek veri setleri üzerinde uygulama yapmışlardır. Sonuç olarak yardımcı bilgiler kullanan MTK yaklaşımının güvenilirlik ve geçerliği arttırdığı bulunmuştur.

Feinberg (2012) örneklem büyüklüğü, alt testler arası korelasyon, alt test uzunluğu, örneklem büyüklüğü ve birey yetenek düzeyi koşulları altında gerçek verilere dayalı olarak ürettiği veriler üzerinde Haberman'ın (2008) yöntemine dayalı olarak alt test puanları kestirmiştir. Araştırma sonucunda yüksek güvenilirliğe ve toplam puanı ile orta düzeyde korelasyona sahip alt testlerin katma değere sahip olma olasılığının

daha yüksek olduğu bulunmuştur. Ayrıca çalışmada dikkate alınan alt ölçeklerin sadece küçük bir yüzdesi ($45/865 = \% 5$) değer kazandığından alt ölçeğin nadiren raporlanmaya değer olduğu belirtilmiştir.

Shin, Ansley, Tsai, ve Mao (2005) çalışmalarında Yen'in Konu Performans İndeksi Yöntemini, Wainer'in Ağırlıklandırma Yöntemini ve Bock, Thissen ve Zimowski'nin (1997) MTK yaklaşımını alt test puan kestirimlerinde mutlak yanlılık, standart hata, RMSE ve %95 güven aralığı istatistikleri açısından karşılaştırmışlardır. Yöntemlerin performansını alt test uzunluğu ve alt test sayısının orta düzeyde, alt testler arası korelasyonların ise yüksek düzeyde etkilediğini bulmuşlardır. Ham puanlara dayalı yaklaşıma göre diğer yöntemlerin daha düşük RMSE'ye sahip olduğu belirtilmiştir. Ayrıca her bir yöntemin MCMC kestirimleri ile orijinal kestirimleri karşılaştırılmış ve MCMC kestirimlerinin ya aynı ya da daha iyi performans gösterdiği bulunmuştur.

Shin ve Chien (2007), alt test puan kestirimlerinde önsel etkilerin değerlendirilmesinde ardışık Bayesian analizlerin kullanılması üzerine bir çalışma yapmışlardır. Çalışmada örneklem sayısı (250,1000), alt test uzunluğu (6,18) ve açık uçlu soruların çoktan seçmeli sorulara oranı (%0, %50) koşulları altında birinci, ikinci, üçüncü ve dördüncü yıllara ait veriler kendinden önceki yıllar önsel bilgi olarak kullanılacak şekilde veri setleri üretilmiş ve Shin (2007)'in önerdiği Bayesian yaklaşım ile analiz edilmiştir. Araştırma sonucunda, ikinci ve sonraki yıllar için elde edilen alt test puan kestirimlerinde standart hata, RMSE ve güven aralığı açısından iyileşme gözlemlendiği bulunmuştur.

De la Torre (2009), demografik ya da eğitimsel bilgiler gibi yardımcı (ancillary) değişkenlerin çok boyutlu değerlendirmelerde yetenek kestirimlerinin geliştirilmesi üzerindeki etkisini hiyerarşik bir çerçevede araştırmıştır. Örneklem büyüklüğü (2000), alt test sayısı (2, 5), alt test uzunluğu (10, 20) , alt testler arası korelasyon (0.5, 0.9) ve alt testlerle kovaryant değişkenler arasındaki ilişki düzeyi (0.25 ve 0.50) koşulları altında üretilen veriler üzerinde dört yöntemin kestirimleri incelenmiştir. Yöntemler sırasıyla, alt testler arası korelasyonları ve alt testler ile kovaryant değişkenler arasındaki ilişkileri görmezden gelen model (A), yalnızca alt testler arası korelasyonları kullanan model (B), yalnızca alt testler ile kovaryant değişkenler arası ilişkiyi kullanan model (C) ve hem alt testler arası korelasyonları hem de alt testler ile kovaryant değişkenler arası ilişkiyi kullanan model (D) olarak tanımlanmıştır. Modeller hata kareleri ortalaması, gerçek ve kestirilen yetenek arasındaki

korelasyon ve kestirimlerin ortalama sonsal (posterior) standart sapması istatistiklerine göre karşılaştırılmıştır. Tüm modellerde alt test sayısı artışının modellerin performansını minimal düzeyde etkilediği fakat madde artışının kestirim hatalarını azalttığı bulunmuştur. Ayrıca alt testler arası korelasyonları ve alt testler ile kovaryant değişkenler arası ilişkileri ayrı ayrı ya da birlikte kullanmanın yetenek parametresi kestirimlerinde daha iyi sonuçlar sağladığı belirtilmiştir. Aynı çalışmada 1500 sekizinci sınıf öğrencisinin 41 matematik ve 36 fen bilimleri testi verileri üzerinde yapılan gerçek veri uygulamasında da simülasyon çalışması ile benzer sonuçlar elde edilmiştir.

Lin (2012), İşletmede Temel Alan Testi'nin (Major Field Test in Business) alt testlerinin neden raporlanmadığını sorgulamış ve toplamda 120 madde ve yedi alt alandan oluşan bu testi alt test puanları kestirimleri açısından incelemiştir. Alt test puan kestirimlerinde Haberman'ın (2005) yöntemini kullanırken testin yapı geçerliği için dört farklı modeli LISREL ile analiz etmiştir. Araştırma sonucunda gözlenen toplam test puanının gözlenen alt test puanına göre alt testlerin gerçek puanına daha fazla yaklaştığını bulmuşlar ve bu durumu alt testlerin raporlanma nedeni olarak göstermişlerdir.

Md Desa (2012) telafisel ve kısmi telafisel İki Faktör modellere göre alt test uzunluğu (10,20) ile genel ve alt testler arasındaki madde ayırt edicilik parametresi (0.50, 0.75, 1.0, 1.25, 1.50) koşulları altında üretilen verileri analiz etmiştir. Araştırma sonunda telafisel İki Faktör modelin daha iyi sonuç verdiği bulunmuştur.

Shin (2007), çalışmasında Yen'in Konu Performans İndeksi Yöntemini, Wainer'in Ağırlıklandırma Yöntemini, Bock, Thissen ve Zimowski'nin (1997) MTK yaklaşımını, Shin'in (2005) MCMC Regresyon yaklaşımını ve doğru cevap oranı yaklaşımını karışık madde formatlı testlerde alt test puan kestirimlerinde güvenilirlik, %95 güven aralığı, RMSE, yanlılık ve standart hata açısından karşılaştırmıştır. Bu amaçla örneklem büyüklüğü (250, 500, 1000), alt test uzunluğu (6,12,18), alt testler arası korelasyon (0.5, 0.8, 1.0) ve açık uçlu soruların çoktan seçmeli sorulara oranı (%0, %20, %50) koşulları altında veri setleri üretilmiştir. Örneklem büyüklüğünün Bock'ın yöntemi dışında kestirimleri etkilemediği, en yüksek alt test güvenirliliğinin Shin'in ve Wainer'in yaklaşımlarından elde edildiği, alt test uzunluğu artışının Bock'ın Yöntemi dışında güvenirliliği arttırdığı, alt testler arası korelasyonun doğru cevap oranı

yaklaşımı dışında güvenilirliği arttırdığı ve açık uçlu soruların çoktan seçmeli sorulara oranı arttıkça yöntemlerin kestirim güvenilirliğinin arttığı bulunmuştur.

Sha ve McCoy'un (2014), çalışmaları iki bölümden oluşmaktadır. Birinci bölümde Wainer'in Ağırlıklandırma yöntemi ile Haberman'ın Ağırlandırılmış Ortalama (weighted average) yöntemini normal ve çarpık veri setlerinde alt test güvenilirlik kestirimleri açısından karşılaştırmışlardır. Bu amaçla aralarında 0.65 korelasyon olan sırasıyla 55 ve 18 maddeye sahip iki alt test ve iki alt test ile toplam testin güvenilirliği sırasıyla 0.85, 0.59, 0.87 olacak şekilde normal ve çarpık dağılıma göre 1000 kişilik veri setleri üretilmiştir. İki yöntemin de normal ve çarpık dağılımlarda aynı güvenilirlik kestirimi yaptığını ve kestirimlerin dağılımdan etkilenmediği bulunmuştur. Çalışmanın ikinci kısmında dağılımın alt test puanlarının katma değerlerini (added value) nasıl etkilediğini hata kareleri ortalamasının azalma oranı (PRMSE) açısından incelemişler ve dağılımın etkili olmadığını bulmuşlardır.

Skorupski (2008), çalışmasında alt test yetenek kestirimlerinde testin geri kalanını bilgilendirici önsel dağılım (informative prior distribution) olarak kullanan Bayesiyen MTK yaklaşımı, ham KTK puanlarını kullanan regresyona dayalı yaklaşım ve ham MTK puanlarını kullanan regresyona dayalı yaklaşımı hem birbirleriyle hem de ağırlıklandırılmamış (ham) KTK ve MTK eşleriyle karşılaştırmıştır. Araştırmada test özellikleri sırasıyla $N=10000$, 6952, 4895 ve 17226, alt test sayısı=4, 4, 6, 4 ve toplam test uzunluğu=36, 54, 64, 52 olan dört geniş ölçekli testin verileri kullanılmıştır. Sonuç olarak ham KTK ve regresyona dayalı KTK kestirimleri karşılaştırıldığında her iki yöntemde göre alt testlerin yetenek ortalamaları aynı kestirilirken regresyona dayalı yöntemde standart sapmalar daha düşük kestirilmiştir. Ham MTK, regresyona dayalı MTK ve Bayesiyen MTK sonuçları karşılaştırıldığında Bayesiyen MTK yaklaşımında alt test yetenek ortalamaları daha yüksek kestirilirken standart sapmanın en düşük kestirildiği yaklaşım regresyona dayalı MTK olarak bulunmuştur.

De la Torre ve Patz (2005), çalışmalarında çok boyutlu değerlendirmeler için pratik bir uygulama örneği göstermişlerdir. Çalışmalarının simülasyon kısmı için örneklem büyüklüğü=1000, yetenek sayısı=2 ve 5, test uzunluğu=10,30 ve 50, yetenekler arası korelasyon=0.0, 0.4, 0.7 ve 0.9 olan veri setleri üretmişlerdir. 3PL model için gerçek toplam test yeteneği ve kestirilen toplam test yeteneği arasındaki korelasyonlara göre MCMC kestirimlerinin alt test uzunluğu , yetenek sayısı ve

korelasyon artışı ile daha iyi kestirimler yaptığı bulunmuştur. Genel yeteneğin tek boyutlu ve çok boyutlu BS kestirimleri karşılaştırıldığında çok boyutlu kestirimlerin etkinliğinin yetenekler arası korelasyon ve yetenek sayısı artışı ile daha belirgin fakat test uzunluğuna duyarlılığının az olduğu belirtilmiştir. Alt test kestirimlerinde Wainer ve arkadaşlarının (2001) Ağırlıklandırma Yöntemini kullanmışlar ve bu yöntemi çok boyutlu kestirim sonuçları karşılaştırmışlardır. İki yöntemin kestirim sonuçları arasındaki korelasyonların çok yüksek olduğu yöntemlerin yalnızca uç değerlerde biraz farklılaştığı bulunmuştur. Ayrıca bir lisenin 2255 dokuzuncu sınıf öğrencilerinin matematik (25 madde), matematiksel hesaplama (20 madde), Heceme (20 madde) ve Sosyal Bilimler (25 madde) testlerine ait verileri tek boyutlu ve çok boyutlu BS kestirim yöntemlerine göre analiz etmişlerdir. Sonuç olarak gerçek veri üzerinde de çok boyutlu yöntemin görece etkinliğinin daha iyi olduğu bulunmuştur.

Shin ve Li (2007) bireysel ve okul düzeyinde büyümenin alt ölçek puanları üzerindeki istatistiksel modellemesi üzerine bir araştırma yapmışlardır. Büyüme modelleri üzerinde okul büyüklüğü, alt testler arası korelasyon, okul özellikleri gibi çeşitli koşullar altında üretilen verilerden kestirilen yetenek puanları doğru cevap oranı, Bock'ın (1997) Yöntemi, Yen'in (1987) Yöntemi ve Wainer'in(2000) Yöntemi'ne göre karşılaştırmışlardır. Araştırma sonucunda he okul hem de birey düzeyinde en güvenilir kestirimlerin Yen'in Yöntemi'nden elde edildiği bulunmuştur.

Stone, Ye, Zhu ve Lane (2009), 10545 sekizinci sınıf öğrencisine uygulanan 17,17,15 ve 10 maddeli dört alt testten oluşan geniş ölçekli bir matematik değerlendirme sınavının verileri üzerinde Yen'in Konu Performans İndeksi Yöntemi, Wainer'in Ağırlıklandırma Yöntemi ve ÇBMTK Yöntemi alt test puan güvenilirlikleri açısından karşılaştırılmıştır. Alt testler arası korelasyonların 0.74-0.79 arasında değiştiği ve testin boyutluluk analizlerine göre baskın bir tek boyuta sahip olduğu bulunmuştur. Yöntemler alt test puan ortalamaları ve standart sapmaları ile alt testler arası korelasyon kestirimleri açısından karşılaştırılmıştır. Yen'in Konu Performans İndeksi Yönteminin diğer yöntemlere göre daha iyi performans gösterdiği belirtilmiştir.

De la Torre, Song ve Hong (2011), alt test puanlarının kestiriminde ÇBMTK Modeli, Üst Düzey Sıralı Model, Ağırlıklandırma Yöntemi ve Konu Performans İndeksi Yöntemini karşılaştırmışlardır. Çalışmanın simülasyon verileri örneklem büyüklüğü

(1000), alt test sayısı (2, 5), alt test uzunluğu (10, 20, 30) ve alt testler arası korelasyon (0.0, 0.4, 0.7, 0.9) koşulları altında Üst Düzey Sıralı Model'e dayalı olarak üretilmiştir. Kısa alt test uzunluğu ve yüksek korelasyon koşullarında dahi ÇBMTK Modeli, Üst Düzey Sıralı Model ve Ağırlıklandırma Yöntemlerinin en iyi sonuçları verdiği fakat Konu Performas İndeksi yönteminin diğer üç yöntemle göre zayıf kaldığı bulunmuştur. Ayrıca 4 alt testten oluşan 2255 dokuzuncu sınıf öğrencisinin verileri üzerinde dört yöntemin performansı karşılaştırılmış ve gerçek veri analizinin dört yöntem arasındaki benzerlikleri ve farklılıkları daha da vurguladığı belirtilmiştir.

Yao (2017), çok boyutlu MTK çerçevesinde 3PL modele dayalı olarak en çok olabilirlik (EÇÖ) ağırlıklandırılmış en çok olabilirlik (AEÇÖ) ve maksimum sonsal (MS) yöntemlerinin alt test ve toplam test yetenek kestirim performansını basit ve karmaşık yapıları karışık formatlı testlerde hem iki hem de çok kategorili veriler üzerinde karşılaştırmıştır. Bu amaçla bir, iki, üç ve dört boyutlu testler çeşitli yetenek dağılımlarına göre üretilmiştir. Tüm koşullarda WLME yönteminin daha küçük yanlılık ve daha yüksek güvenilirlik verdiği ancak daha büyük RMSE ve standart hata değerleri verdiği bulunmuştur. Ayrıca alt testler arası korelasyon varlığında boyut sayısı fazla olan yapılar için MS kestirim yöntemi önerilmiştir.

Seybert (2013), sıralama ölçeğindeki çok boyutlu veriler için madde ve yetenek parametreleri kestirimlerinde Hiperbolik Kosin Model (Hyperbolic Cosine Model) adında bir madde tepki kuramı modeli önermiştir. Çalışmasında örneklem büyüklüğü (250,500), alt test sayısı (4,8) ve alt test uzunluğu (20) koşulları altında oluşturulan verileri parametre kestirimlerinde doğrudan ve iki aşamalı olarak iki farklı kestirim stratejisi kullanarak önerilen modele dayalı analizler yapılmıştır.

Haberman (2005) alt test puanlarının toplam puanlar üzerinde katma değerlerinin (added value) olup olmadığını belirlemek için KTK'ya dayalı bir yöntem önermiştir. Bu yöntemde bir alt ölçeğin gözlenen puanı ve testin geri kalan kısmına ait gözlenen puanlardan o alt ölçeğin gerçek puanının kestirilmesine odaklanılır. Yönteme göre bir alt ölçeğin gerçek puanı, gözlenen toplam puanlar yerine o alt ölçeğin gözlenen puanından daha iyi kestirilmesi durumunda alt ölçek katma değer kazanır. Sinharay (2013) Haberman (2005) tarafından önerilen bu yöntemin matematiksel yapısını paralel formlardan elde edilen puanlara göre yeniden düzenlemiş ve gerçek puan kavramına göre paralel form kavramının açıklanmasının daha kolay olduğunu

belirtilerek kendi yöntemi ile Haberman'ın yönteminin anlaşılabilirliğini arttırdığını belirtmiştir. Sinharay (2013) araştırmasında gerçek veri setlerinde uygulama yapmış ve bazı geniş ölçekli testlerdeki alt testlerin katma değerinin olmadığını bulmuştur.

Brennan (2012) alt testlerin raporlanıp raporlanmaması konusunda kullanılabilecek bir indeks geliştirmiştir. Alan yazında kullanılan Haberman (2008) ve Wainer ve ark. (2001)'in regresyona dayalı yaklaşımlarına eleştiri getirmiş ve tamamen klasik test teorisine ve klasik güvenilirlik kavramlarına dayanan bir yaklaşım önermiştir. Çalışmasında Haberman (2005)'in kullandığı gerçek veriler üzerinde analiz yapmış ve diğer çalışma ile aynı karara vardığını belirtmiştir.

Bulut (2013) alt test güvenilirliğini bireyler arası ve birey içi alt test güvenilirliği olarak ikiye ayırdığı yeni bir yaklaşımla ele almıştır. Çalışmasında alt test sayısı (3,5,7), alt test uzunluğu (10,20,40) ve alt testler arası korelasyon (0.3, 0.5, 0.8) olacak şekilde veri setleri üretmiş ve 3PL model ve MS kestirim yöntemine göre tek boyutlu MTK ve çok boyutlu MTK açısından analiz etmiştir. Araştırma sonucunda tüm koşullarda çok boyutlu modelin tek boyutlu modele göre daha iyi performans gösterdiği bulunmuştur.

Gignac, Palmer ve Stough (2007) ve duyguları belirleme zorluğu, duyguları açıklanma zorluğu ve dışa yönelik düşünme adlı üç alt ölçekten oluşan Toronto Aleksimi Ölçeği'nin 355 kişiden elde edilen verileri üzerinde beş hiyerarşik modeli model veri uyumu açısından incelemişlerdir. Araştırma sonucunda modifiye edilmiş İki faktör Model'in veriye en uyumlu model olduğu bulunmuştur. Ayrıca çalışmanın alt ölçeklerinin güvenilirlikleri Cronbach alfa ve omega katsayılarına göre incelenmiştir. Aynı ölçeğin Reise, Bonifay ve Haviland (2013) tarafından 1612 bireyden elde edilen veriler üzerinde çok boyutlu MTK ve İki Faktör Modellere göre model veri uyumu incelenmiştir.

Goodman ve Hambleton (2004), üç ABD'li test firmasında, 11 ABD ve iki Kanada eyaletinde yapılmış standart testleri inceleyerek alt testlerin sağladığı bilgilere ve raporlanması konusuna dikkat çekmişlerdir.

Huang, Wang ve Chen (2013) 1PL, 2PL ve 3PL modellere göre Üst Düzey Sıralı ve Üçüncü Düzey Sıralı hiyerarşik MTK modelleri üzerinde madde ve yetenek parametre doğrulama çalışması yapmışlardır. Araştırmada genel ve alt testler arasındaki faktör yükü ile üst düzey faktör sayısı koşullarının farklı düzeyleri için

örneklem büyüklüğü=1000 alt test uzunluğu =20 olacak şekilde üç farklı simülasyon çalışması için veriler üretilmiştir.

Chang (2015) İki faktör Modelde alt testlerin genel faktörün artıklarından elde edilmesinden kaynaklanan problemlere çözüm getirebilmek amacıyla sınırlandırılmış bir İki Faktör Modeli önermiştir. Çalışmasında test uzunluğu (15,30), boyut sayısı (3,4), alt testler arası korelasyon (0.3, 0.6) koşulları altında üretilen verilerin Rasch ve 2PL modeli BS ve MS kestirim yöntemlerinin performansını toplam test yetenek ve alt test yetenek kestirimleri açısından karşılaştırmıştır. Araştırma sonucunda BS ve MS kestirimlerinin birbirine benzer sonuç vermekle birlikte MS'in biraz daha düşük hata varyansına sahip olduğu belirtilmiştir.

Meijer, Boeve, Tendeiro, Bosker ve Albers (2017) çoktan seçmeli ve karma yapıları iki üniversite sınavının alt testlerinin toplam puan üzerinde bir değeri olup olmadığını araştırmışlardır. Araştırma sonucunda alt testlerin toplam puanın üzerinde bir değeri olmadığını belirterek raporlanmamasını önermişlerdir. Ayrıca araştırma sonucunda açık uçlu soruların olduğu alt testlerden elde edilen hataların daha düşük olduğu bulunmuştur.

2.5. İlgili Araştırmalar Özeti

Alt test puanlarının raporlanıp raporlanmaması konusu üzerinde tartışmaların 1980'lerden günümüze kadar devam ettiği görülmektedir. Alt testlerin tanıtılabilirliği sağlama potansiyellerine rağmen yetersiz materyal içermeleri nedeniyle alt testlerin raporlanmadan önce ampirik olarak performanslarının incelenmesi gerektiği araştırmacılar tarafından sıklıkla belirtilen bir konudur. Bu bağlamda araştırmalarda alt testlerden elde edilebilecek bilgilerden yararlanılabilmesi için alt test puanlarının sağlanması gereken bazı özellikler olduğu belirtilmiştir. Alt test ve toplam test puanı kestirimlerine ilişkin alan yazın çalışmalarının alt testlerin toplam test puanı üzerinde bir değeri olup olmadığının belirlenmesi ve yeterli psikometrik özelliklere sahip alt test puan kestirimlerinin sağlanması olarak iki başlıkta toplandığı görülmektedir.

Alt testlerin birbiri ile yüksek derece korelasyon içermesi gibi durumlarda alt testlerden ayrı ayrı elde edilecek bilgilerin toplam testten elde edilecek bilgilerden bir üstünlüğü olmadığı durumlarda sadece toplam test puanlarının raporlanmasının yeterli olduğu belirtilmiştir. Bu amaçla alt test puanlarının toplam test puanı üzerinde katma değeri olup olmadığının belirlenebilmesi için çeşitli istatistiklerin önerildiği ve

bu istatistiklerin gerçek ve simülatif veriler üzerinde test edildiği arařtırmalar bulunmaktadır.

Güvenirlik ve geçerlik analizleri üzerinden alt test puanlarının yeterli psikometrik özellik gösterip göstermediğini inceleyen arařtırmalar mevcuttur. Güvenilir olmayan puanların raporlanmaması gerektiğini belirten arařtırmacıların alt test puanları için daha doğru ve güvenilir kestirimler sađlayan bir puanlama yöntemi sunmak amacıyla KTK, tek boyutlu MTK ve çok boyutlu MTK'ya dayalı çeřitli yöntemler önerdiği görölmektedir. Bir alt test puanı kestiriminde testin geri kalanından sađlanan bilgileri kullanan ađırlıklandırma yöntemlerinin geleneksel toplam doğru sayısı, düzeltilmiş toplam doğru sayısı veya doğru sayısı oranı gibi puanlama yöntemlerine göre üstünlüğünü raporlayan arařtırmalar bulunmaktadır. Fakat alt test ve toplam test puanı kestirimlerinde alt testler arasındaki ilişkileri ve test dışı bilgileri kullanmaya olanak sađlayan çok boyutlu analizlere ilginin arttığı görölmektedir. Ayrıca son yıllarda alt test ve toplam test arasındaki ilişkiyi hiyerarşik olarak ele alan çalışmaların dikkat çekici olduđu düşünölmektedir. Hiyerarşik çok boyutlu MTK'ya dayalı çalışmaların hem daha doğru ve güvenilir alt test puan kestirimlerinde hem de model-veri uyumunun incelenmesi amacıyla yapısal modellerin test edilmesinde kullanıldığı görölmektedir.

Arařtırmalarda alt test puanı kestirimlerinde geleneksel toplam doğru sayısına dayalı yöntemlerin kullanılmasının yanı sıra MTK'ya dayalı BS, MS, ML ve MCMC gibi çeřitli yöntemlerin kullanıldığı görölmektedir. Birçok çalışmada MCMC yöntemlerinin analiz süresi uzunluğunun dezavantajından bahsedilse de kestirim doğruluđu açısından yöntemin kullanılmasının sıklıkla önerildiği görölmektedir. Ayrıca MTK'ya dayalı alt test puan kestirimlerinde 1PL, 2PL, 3PL, Rasch, Kısmi Puanlama gibi çeřitli modellerin kullanıldığı gözlenmektedir. Arařtırmaların simülatif veya gerçek veri setleri üzerinde yapıldığı görölürken simülatif verilerin üretilmesinde alt test sayısı, alt testler arası korelasyon, alt test uzunluđu, örneklem büyüklüđu, genel boyut ile alt testler arası korelasyonlar, genel boyut ile alt testin ayırt edicilik parametresi, yetenek parametresi dağılımı, madde ve test yapıları gibi çeřitli koşulların kullanıldığı görölmektedir.

3. YÖNTEM

Bu bölümde, araştırma yöntemi, araştırma verilerinin üretilmesi ve simülasyon koşulları ile kestirim modelleri ve verilerin analizine yer verilmektedir

3.1. Araştırmanın Türü

Deneysel desenler değişkenler arasındaki neden sonuç ilişkilerini test etmeyi amaçlayan araştırma desenleridir. Bu amacı gerçekleştirebilmek için Fraenkel, Wallen ve Hyun'a (2011, s.265-266) göre deneysel desenler, bağımsız değişken/lerin bağımlı değişken/ler üzerindeki etkisini incelemek için en az iki koşulun karşılaştırılmasını ve bağımsız değişkenin araştırmacı tarafından doğrudan değişimlenmesini (manipüle edilmesini) gerektirir. Ayrıca, araştırmacılar iç geçerliği korumak için dışsal değişkeni (ilgilenilmeyen ya da istenmeyen değişken) kontrol altına alarak bağımlı değişken üzerinde ölçüm yapmalıdır (Kerlinger, 1973, s.300-313; Gall, Gall & Borg, 2003, s.367-368). Simülasyon çalışmaları doğası gereği araştırmacılara bağımsız değişkenleri değişimleme ve dışsal değişkenleri kontrol altına alma imkânı sağlar. Bir bölümünde simülasyon verisi kullanılan bu araştırmada farklı test koşullarında üretilmiş verilerin toplam test puanı ve alt test puanları kestirimlerinin doğruluğu ve güvenilirliği farklı modeller ve farklı test koşulları açısından karşılaştırıldığından çalışma simülatif verilerle yürütülen deneysel araştırma özelliği taşımaktadır. Ayrıca araştırmanın gerçek veri uygulaması içeren diğer bölümü, TEOG sınavı ile ilgili mevcut duruma ait bilgiler vereceğinden bu çalışmanın betimsel araştırma özelliği de bulunmaktadır.

3.2. Simülasyon Koşulları

3.2.1. Çalışma Grubu Büyüklüğü

Alan yazın incelendiğinde, çok boyutlu ya da alt testlerden oluşan ölçme yapıları için madde ve yetenek parametresi doğrulama çalışmalarında çeşitli büyüklüklerde örneklemelerin kullanıldığı görülmektedir. Örneğin, Yao ve Boughton (2009) çalışmalarında N=1000, 3000 ve 6000 örneklem büyüklüğü koşulları altında ve Yao (2010) çalışmasında N=500, 1000 ve 2000 örneklem büyüklüğü koşulları altında madde ve yetenek parametresi doğrulanması üzerinde örneklem büyüklüğü etkisini incelemiş ve örneklem büyüklüğü artışının madde parametresi kestirim hatalarını azalttığını fakat yetenek parametresi kestirim hataları üzerinde etkisi olmadığını belirtmişlerdir. Benzer şekilde, de la Torre ve Song (2009) çalışmalarında N=1000,

2000 ve 4000 örneklem büyüklüğü koşulları altında yetenek parametre kestirimleri üzerinde örneklem büyüklüğünün etkisi olmadığını göstermişlerdir. Bazı yetenek parametresi doğrulama çalışmalarında ise araştırmacılar N=1000 örneklem büyüklüğünün yeterli olduğunu belirterek araştırmalarını tek bir örneklem üzerinden simüle etmişlerdir (de la Torre & Patz, 2005; de la Torre, Song & Hong, 2011; Koziol, 2016). Gerçek veri uygulamalarına dayalı yetenek parametre kestirimlerinde N=1500 (de la Torre, 2009), N=17226 (Skorupski & Carvajal, 2010), N=2255 (de la Torre, Song & Hong, 2011) veya N=4031 (Hartig & Höhler, 2008) gibi çeşitli büyüklüklerde veri setlerinin analiz edildiği görülmüştür. Alan yazın çalışmalarından yetenek parametresi kestirimlerinde 1000 ve üzeri örneklem arasında fark gözlenmediği bulgusuna dayalı olarak bu araştırmada örneklem büyüklüğü bağımsız değişken olarak seçilmemiş ve araştırmanın verileri N=3000 olacak şekilde üretilmiştir. Ayrıca geniş ölçekli testleri temsil edebilmesi için 3000 örneklem büyüklüğünün yeterli olduğuna karar verilmiştir.

3.2.2. Alt Test Sayısı

Alan yazındaki çok boyutlu ya da alt testlerden oluşan ölçme yapıları için parametre doğrulama çalışmaları incelendiğinde hem simülasyon hem de gerçek veri setleri üzerinde yapılan araştırmalarda kullanılan ölçme araçlarının iki ila altı boyut arasında değişen boyutlarda olduğu görülmüştür. Örneğin; Yao (2017) çalışmasında 2 ve 4 boyutlu veri setleri kullanırken, Lee (2012) 3 ve 6 boyutlu veri setleri, Edwards ve Vevea (2006) ise 2 ve 4 boyutlu veri setleri kullanmışlardır. Ülkemizde yapılan çeşitli geniş ölçekli sınavlar incelendiğinde TEOG sınavının aynı oturumda yapılan sınavlarının üç toplamda altı, KPSS'de aynı oturumda yapılan çeşitli sınavların 2 (Genel Kültür ve Genel yetenek gibi) ya da beş (Çalışma ekonomisi, Ekonometri, İstatistik, Kamu Yönetimi ve Uluslararası ilişkiler gibi), ALES sınavının ise 2 (sayısal ve sözel) alt testten oluştuğu görülmüştür. Alan yazında yapılan araştırmalar ve ülkemizde yapılan geniş ölçekli sınavlar ve test uzunlukları göz önüne alındığında gerçek durumları temsil etmesi açısından bu araştırmada alt test sayısı iki ve üç olarak belirlenmiştir.

3.2.3. Alt Test Uzunluğu

Alan yazındaki çok boyutlu ya da alt testlerden oluşan ölçme yapıları için parametre doğrulama çalışmaları incelendiğinde hem simülasyon hem de gerçek veri setleri üzerinde yapılan araştırmalarda kullanılan ölçme araçlarının 5-60 madde arası alt

test uzunluđuna sahip olduđu grlmektedir. lkemizde yapılan geniř lekli sınavlar incelendiđinde TEOG sınavlarının toplam altı alt testten ve 20'řer maddeden oluřtuđu, ALES sınavının sayısal blmnn 40'ar sorudan oluřan sayısal1 ve sayısal2 olarak iki alt testten, szel blmnn ise 40'ar sorudan oluřan szel1 ve szel2 olarak iki alt testten oluřtuđu belirlenmiřtir. KPSS sınavları incelendiđinde aynı oturumda yapılan sınavlardan genel kltr ve genel yetenek alt testlerinin 60'ar sorudan, A grubu sınavlarından aynı oturumda yapılan sınavlardan Hukuk, İktisat, İřletme, Maliye ve Muhasebe alt testlerinin 30'ar sorudan ve Din Hizmetleri Alan Bilgisi Testi'nin (DHAB) ise DHAB1 ve DHAB2 alt testlerinin 20'řer sorudan oluřtuđu grlmřtr. Alan yazın ve lkemizde yapılan sınavlara dayalı olarak gerek durumları temsil etmesi aısından bu arařtırmada alt test uzunluđu 20, 30 ve 40 madde olarak belirlenmiřtir.

3.2.4. Alt Testler Arasındaki Korelasyon

Alan yazındaki ok boyutlu ya da alt testlerden oluřan lme yapıları iin parametre dođrulama alıřmaları incelendiđinde alt testler arasındaki korelasyonların parametre kestirimi zerinde etkisinin olduđu belirtilmiřtir (Yao, 2010; Shin, Ansley, Tsai, & Mao, 2005; de la Torre & Patz, 2005; Shin, 2007). Yine alan yazındaki arařtırmalarda alt testler ya da alt testler arası korelasyon kořulu iin 0.0-1.0 arasında deđiřen dzeylerde byklkler seildiđi grlmřtr. lkemizde 29 Nisan 2015'te yapılan TEOG sınavının altı alt testi arasındaki korelasyonların 0.0 civarında olması da gz nnde bulundurularak (Ek 2.1) bu arařtırma iin alt testler arası korelasyon dzeyleri 0.0, 0.3, 0.5 ve 0.8 olarak belirlenmiřtir.

Tablo 3.1: Simlasyon Kořulları ve Dzeyleri

<i>Alt Test Sayısı</i>	<i>Alt Test Uzunluđu</i>	<i>Alt Testler Arası Korelasyon</i>
2	20	0.0
3	30	0.3
	40	0.5
		0.8

Tekrar (replikasyon) sayısı:50

3.2.5. İterasyon Sayısı

Arařtırmada hem verilerin retilmesi hem de retilen verilerin ve gerek verinin analizinde iteratif bir yntem olan MCMC kullanılmıřtır. Alan yazında MCMC kestirimlerinin dođruluđu ve sresine etki eden faktrlerden birinin iterasyon sayısı

olduđu belirtilmektedir (SAS/STAT, 2009). Arařtırmalarda kullanılan iterasyon sayısının kestirim modelinin karmařıklıđına gre deđiřmekle birlikte birkaç bin olması gerektiđi belirtilmiřtir (Calderhead, 2014; Raftery & Lewis, 1991; SAS/STAT, 2009). MCMC yntemini kullanan arařtırmalar incelendiđinde kullanılması gereken burn in (bařlangıç noktasının sonsal kestirimler üzerindeki etkisini en aza indirmek iin Markov zincirinin dıřında tutulan kısımdır) ve iterasyon sayısı ile ilgili belirli bir sınır olmadıđı grlmektedir. Browne, Rasbash ve Charlton (2009), eřitli veri yapıları ve modeller üzerinde 500' burn in olmak zere 5000 iterasyon; de la Torre (2009) ve de la Torre ve Patz (2005) 3PL model kestirimlerinde 2000'i burn in olmak zere 10000 iterasyon; Holmes ve Held (2006) regresyon kestirimlerinde, Kieftenbeld ve Natesan (2010) ise ařamalı tepki modeli kestirimlerinde 1000'i burn in olmak zere 10000 iterasyon; Hadfield (2010) genelleřtirilmiř lineer karıřık modellerde 10000'i burn in olmak zere 60000 iterasyon; Patz ve Junker (1999a) 2PL model kestirimlerinde 400' burn in olmak zere 7400 iterasyon kullanırken Patz ve Junker (1999b) 3PL model kestirimlerinde 2000'i burn in olmak zere 50000 iterasyon kullanmıřlardır. Alan yazın arařtırmaları da gz nnde bulundurularak bu arařtırma kapsamında verilerin retilmesi ve analizinde burn in sayısı 2000 olmak zere 10000 iterasyon kullanılmıřtır.

3.3. Verilerin retilmesi

Arařtırmada kullanılan veri setleri 29 Nisan 2015'te yapılan TEOG sınavının psikometrik zelliklerine dayalı olarak retilmiřtir. Bu sınav verisinin faktr yapısını belirlemek iin Factor 10.3 programında Ađırlıklandırılmamıř En Kk Kareler (ULS) yntemi ve varimax dndrme tekniđine gre veriler analiz edilmiřtir. Analiz sonucunda elde edilen maddelerin faktr yk matrisi EK-2.1'de, alt testler arası korelasyon matrisi ise EK-2.2'de sunulmuřtur. EK-2.1 ve EK-2.2'de grldđ gibi TEOG verisi basit yapılı rtk yetenek konfigrasyonuna sahip olduđu iin bu alıřmada veriler basit yapılı olacak řekilde retilmiřtir. TEOG 2015 sınavının madde parametreleri BILOG-MG (Zimowski, 1998) programı ile 3PL modele gre analiz edilmiř ve Tablo 3.2'de sunulmuřtur.

Tablo 3.2: TEOG 2015 Sınavının Alt Testlerine Ait Madde Parametreleri

Alt Test	Parametre					
	a		b		c	
	En Küçük	En Büyük	En Küçük	En Büyük	En Küçük	En Büyük
Din Kültürü	0.525	3.395	-1.855	0.161	0.000	0.330
Fen Bilgisi	0.928	3.905	-0.692	0.890	0.002	0.316
İngilizce	0.163	6.237	-0.534	2.108	0.000	0.336
Matematik	0.764	4.829	-1.284	1.654	0.000	0.313
Tarih	1.351	5.960	-0.615	0.691	0.001	0.299
Türkçe	0.810	3.909	-1.342	0.802	0.000	0.357

Tablo 3.2’de yer alan madde parametreleri ve alan yazında yer alan simülasyon çalışmaları göz önünde bulundurularak bu çalışmada kullanılan verilerin madde ayırt edicilik parametresi ranjı [0.8-3] arasında olacak şekilde ortalaması 1.5 ve varyansı 0.5 olan bir normal dağılımdan; güçlük parametresi ranjı [-2-2] arasında olacak şekilde ortalaması 0.0 ve varyansı 1.0 olan bir normal dağılımdan ve en düşük asimtot (şans) parametresi ise (6,16) olan bir beta dağılımdan üretilmiştir. Yetenek parametreleri çok değişkenli normal dağılıma $\theta_i \sim MVN(0, \Sigma)$ dayalı olarak ortalaması 0, varyansı ise araştırma koşullarında belirlenmiş olan varyans-kovaryans matrisine göre üretilmiştir. Üretilen madde ve yetenek parametreleri kullanılarak Tablo 3.1’de özetlendiği gibi alt test sayısı 2 ve 3, alt test uzunluğu 20, 30 ve 40 ve alt testler arası korelasyonlar 0.0, 0.3, 0.5 ve 0.8 olan iki kategorili 3000 kişilik veri setleri 50 tekrara dayalı olarak SimuMIRT (Yao, 2003) programı kullanılarak üretilmiştir. Harwell, Stone, Hsu ve Kirisci (1996) monte carlo simülasyon çalışmaları için optimal koşulları belirleme, mevcut programları inceleme ve simülasyon çalışmalarının kavramsallaştırılmasının önemini açıklama konusundaki çalışmalarında, simülasyon çalışmalarında en az 25 replikasyon kullanılması gerektiğini belirtmişlerdir. Bu konuda alan yazın incelendiğinde, Yao’nun (2010) ve Huang, Wang ve Chen (2013) 20 tekrar, Çakıcı Eser’in (2015) 25 tekrar ve de la Torre’nin (2009) 100 tekrar ile çalışmalarını yürüttükleri görülmüştür. Bu araştırmada ise tekrar sayısı= 50 olarak belirlenmiştir. Araştırmada SimuMIRT programında kullanılan kontrol dosyasına ilişkin bir örnek EK-3’te verilmiştir.

3.4. Üretilen Verilerin Geçerliğinin Test Edilmesi

Üretilen verilerin arařtırmada incelenen kořullara uygun olarak üretilip üretilmediđi Factor 10.3 programı kullanılarak incelenmiřtir. Üretilen veriler uyum indeksleri, basit yapılı olup olmama ve alt testler arası korelasyonlar ađısından incelenmiř ve verilerin istenilen düzeyde üretildiđi dođrulanmıřtır. Örnek olması ađısından alt test sayısı=3, alt testler arası korelasyon=0.8 ve alt test uzunluđu=20 olan kořula ait veri setinin faktör yükleri matrisi ve alt testler arası korelasyon matrisi sırasıyla EK-4.1 ve EK-4.2'de verilmiřtir. Ayrıca bu örnek verinin uyum istatistikleri RMSEA=0.01 ve CFI=0.99 olarak hesaplanmıřtır.

3.5. Deđerlendirme Kriteri

Arařtırmada alt test puanlarının ve toplam test puan kestirimlerinin dođruluklarını deđerlendirmek için RMSE (Ortalama hata kareler kökü) ve güvenirlilik istatistikleri kullanılmıřtır. Bu istatistiklere ait matematiksel ifadeler ařađıdaki gibidir:

$$RMSE(\tau_j) = \sqrt{\frac{1}{n * N} \sum_{d=1}^n (\tau_j^* - \tau_j)^2}$$
$$Güvenirlilik = \frac{1}{n} \sum_{d=1}^n cor(\tau_j^*, \tau_j)^2$$

τ_j : j parametresinin gerçek deđer

τ_j^* : j parametresinin kestirilen deđer

n: tekrar (replikasyon) sayısı

N: örneklem büyüklüđu

RMSE deđer, gerçek parametre ile kestirilen parametreler arasındaki farkların ortalamasının karekökünü ifade etmektedir. Güvenirlilik deđer ise gerçek parametre ile kestirilen parametreler arasındaki korelasyonun kareler ortalamasını ifade etmektedir.

3.6. Verilerin Analizi

Araştırmada üretilen veriler ve 29 Nisan 2015'te yapılan TEOG sınavına ait veriler BMIRT (Yao, 2003) programı kullanılarak ÇBMTK, İki faktör ve Üst Düzey Sıralı modellere göre analiz edilmiştir. Parametre kestirimlerinde 3PL model ile MCMC kestirim yöntemi kullanılmıştır. Gerçek parametreler ile kestirilen parametrelerin karşılaştırılabilmesi için kestirilen parametreler ile gerçek değerlerin aynı ölçekte olması gerekir. Bunu sağlayabilmek için parametre kestirimlerinde popülasyon parametrelerinin onların gerçek değerine sabitlenmesi gerekir. Normalde gerçek değerleri bilinemez ama üretilen verilerin ortalama ve varyans-kovaryans matrisi, madde parametrelerinin dağılımları gerçek değer yerine kullanılabilir. Yao (2010) yetenek parametreleri kestirimlerinde madde parametrelerinin sabitlenmesi ile sabitlenmemesi yaklaşımı arasında bir fark olmadığını belirtse de bu araştırmada üretilen verilerin özellikleri önsel bilgi (prior) olarak kullanılmıştır. Ayrıca kestirim modelleri ve simülasyon koşullarına göre elde edilen RMSE değerleri ortalaması arasında anlamlı fark olup olmadığı varyans analizi ile test edilmiştir. Varyans analizi sonucunda en az iki grup arasında anlamlı farklılığın gözlemlendiği durumlar için farklılığın hangi gruplar arasında olduğunu belirlemek amacıyla çoklu karşılaştırma testi yapılmıştır. Karşılaştırma testi olarak Fisher'in LSD Testi kullanılmıştır. SPSS programı ara yüzü seçeneklerine göre yalnızca ana etkiler için karşılaştırma testi yapabilmesi nedeniyle etkileşimler için karşılaştırma testleri syntax yazılarak yapılmıştır.

Araştırmada verilerin analizi için Intel Core i7-6700K CPU 4.00GHz işlemci, 16GB (RAM) ve 64 bit işletim sistemi özelliklerine sahip bir bilgisayar kullanılmıştır. Bir hiyerarşik model için 1200 çapraz koşuldaki yalnızca birinin analizi ortalama 25-30 dakika sürmüştür.

4. BULGULAR VE TARTIŞMA

Bu bölümde, sırasıyla her bir alt probleme ait araştırma bulguları ve bu bulgularla ilgili değerlendirmeler yer almaktadır.

4.1. Birinci Alt Probleme İlişkin Bulgular ve Tartışma

Araştırmanın birinci alt problemi kapsamında “Üst Düzey Sıralı (Higher Order), İki Faktör (Bi-factor) ve hiyerarşik çok boyutlu madde tepki kuramı (ÇBMTK) modellerine göre **toplam test puanı** için kestirilen yetenek parametrelerinin **alt test sayısı (2,3), alt test uzunluğu (20, 30, 40) ve alt testler arasındaki korelasyonların büyüklüğü (0.0, 0.3, 0.5, 0.8)** koşullarından nasıl etkilendiği” güvenilirlik ve RMSE değerleri ile incelenmiştir. Yukarıda verilen test koşullarına ait RMSE ve güvenilirlik değerleri alt test sayısı koşuluna göre ayrı ayrı tablo ve grafikler ile verilmiştir.

4.1.1. İki Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri

Tablo 4.1’de araştırmada ele alınan koşullara dayalı olarak iki boyutlu veri setlerinden toplam test puanı için üç hiyerarşik madde tepki kuramı modeli kullanılarak kestirilen yetenek parametrelerine ilişkin RMSE değerleri verilmiştir.

Tablo 4.1.: İki Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri

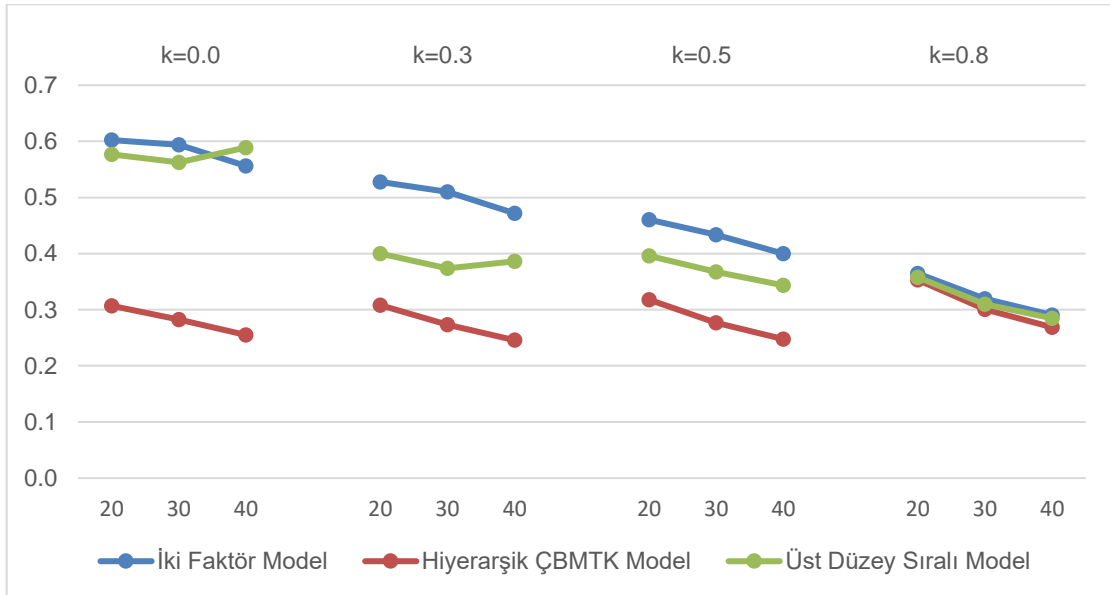
<i>Alt test sayısı</i>	<i>Korelasyon</i>	<i>Alt test uzunluğu</i>	<i>İki-faktör model</i>	<i>Hiyerarşik ÇBMTK model</i>	<i>Üst Düzey Sıralı model</i>
2	0.0	20	0.602	0.307	0.577
		30	0.594	0.282	0.562
		40	0.556	0.255	0.589
	0.3	20	0.528	0.308	0.400
		30	0.510	0.273	0.374
		40	0.472	0.246	0.386
	0.5	20	0.460	0.318	0.396
		30	0.433	0.276	0.367
		40	0.400	0.248	0.344
	0.8	20	0.364	0.353	0.358
		30	0.319	0.301	0.310
		40	0.291	0.268	0.285

Tablo 4.1’de verilen alt test uzunluğu ve alt testler arası korelasyon koşullarının tüm düzeylerine ait RMSE değerleri bir arada değerlendirildiğinde Hiyerarşik ÇBMTK modele ait hataların 0.246-0.353 arasında değiştiği görülmektedir. Alt testler arası korelasyonun ilk üç düzeyi (0.0, 0.3, 0.5) için alt test uzunluğu sabit tutulduğunda kestirim hataları arasındaki farkın minimal düzeyde olduğu gözlenirken en yüksek hataların korelasyon düzeyinin 0.8 olduğu durumda elde edildiği gözlenmektedir. Bunun yanında, alt test uzunluğu arttıkça hataların azaldığı görülmektedir. Örneğin, alt test uzunluğu 20 iken alt testler arası korelasyonun her bir düzeyi (0.0, 0.3, 0.5, 0.8) için elde edilen hataların sırasıyla 0.307, 0.308, 0.318 ve 0.353 olduğu görülmektedir. Bu sonuca göre, alt testler arası korelasyonun 0.8 düzeyi için elde edilen hata, 0.0 düzeyi için elde edilen hatadan yaklaşık olarak %15 daha fazladır. Buna karşılık alt test uzunluğundaki artış ile kestirim hatalarındaki azalma oranı alt testler arası korelasyon arttıkça daha fazla artmaktadır. Örneğin; alt test uzunluğu 20’den 30’a çıkarıldığında yetenek parametresinin kestirim hatasının alt testler arası korelasyon koşulunun her bir düzeyi için sırasıyla yaklaşık olarak %8, %11, %13 ve %15 oranında azaldığı gözlenmektedir. Fakat alt testler arası korelasyon koşulunun tüm düzeyleri göz önüne alındığında alt test uzunluğunu 30’dan 40’a çıkarmanın kestirim hatalarının azalmasına etkisinin yaklaşık %10 oranında olduğu görülmektedir.

İki-faktör modeline dayalı olarak elde edilen RMSE değerleri incelendiğinde, hataların 0,291-0.602 arasında değiştiği görülmektedir. Alt testler arası korelasyon arttığında alt test uzunluğunun tüm düzeyleri için bu yöntemden elde edilen hataların önemli düzeyde azaldığı dikkat çekmektedir. Örneğin, alt test uzunluğu 20 ve alt testler arası korelasyon 0.0 iken yetenek parametresinin kestirim hatası 0.602 olarak elde edilirken aynı alt test uzunluğu düzeyinde alt testler arası korelasyon 0.8 iken elde edilen hatanın 0.291 olduğu gözlenmektedir. Bu sonuca göre iki korelasyon düzeyi açısından hatanın yaklaşık olarak %52 oranında azaldığı söylenebilir. Alt test uzunluğu açısından hatalar incelendiğinde, alt test uzunluğu artışı ile hataların azaldığı ve alt testler arası korelasyon arttıkça hataların daha fazla oranda azaldığı görülmektedir. Örneğin, alt testler arası korelasyon 0.0 iken alt test uzunluğu 20 den 30’a çıkarıldığında İki Faktör modeline dayalı elde edilen yetenek parametresi kestirim hatasının yaklaşık %1 oranında azaldığı gözlenirken aynı koşullar için alt testler arası korelasyon 0.3 iken kestirim hatasının yaklaşık %3, 0.5 iken kestirim

hatasının yaklaşık %6 ve 0.8 iken kestirim hatasını yaklaşık %12 oranında azaldığı gözlenmektedir. Fakat alt testler arası korelasyon koşulunun tüm düzeyleri göz önüne alındığında alt test uzunluğunu 30'dan 40'a çıkarmanın kestirim hatalarının azalmasına etkisi alt testler arası korelasyon düzeylerine göre sırasıyla yaklaşık %6, %7, %8 ve %9 oranları arasındadır.

Üst Düzey Sıralı modele dayalı elde edilen RMSE değerleri incelendiğinde hataların 0.285-0.589 arasında değiştiği görülmektedir. Alt test uzunluğu sabit iken alt testler arası korelasyon düzeyi arttıkça yetenek parametresi kestirim hatalarının korelasyonun 0.3 ve 0.5 düzeyi arasında minimal olmakla birlikte azaldığı gözlenirken en yüksek hataların 0.0 düzeyinden, en düşük hataların ise 0.8 düzeyinden elde edildiği gözlenmektedir. Örneğin, alt test uzunluğu 20 iken alt testler arası korelasyonun her bir düzeyi için elde edilen hataların sırasıyla 0.577, 0.400, 0.396 ve 0.358 olduğu görülürken kestirim hatalarının korelasyonun 0.0 düzeyi ile diğer düzeyleri arasında sırasıyla %31 ve %38 oranında azaldığı görülmektedir. Alt test uzunluğu artışı ile modelden elde edilen kestirim hatalarının çoğunlukla azaldığı görülmektedir. Örneğin, alt testler arası korelasyon düzeyi 0.8 iken alt test uzunluğu 20'den 30'a çıkarıldığında hataların %13 oranında azaldığı gözlenirken alt test uzunluğunun 30'dan 40'a çıkarılması ile hataların %8 oranında azaldığı gözlenmektedir.



Şekil 4.1. İki Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri

Şekil 4.1’de iki boyutlu veri setleri için alt testler arası korelasyon ve alt test uzunluğu koşulları altında İki-faktör Model, Üst Düzey Sıralı Model ve Hiyerarşik ÇBMTK Model’in yetenek parametresi kestirim hatalarına ait grafiklere yer verilmiştir. Şekil 4.1’e göre üç kestirim modelinin tüm koşullar altındaki performansları karşılaştırıldığında, en düşük hata düzeyine sahip kestirimlerin İki Faktör modelinden elde edildiği görülürken en düşük hata düzeyine sahip kestirimlerin Hiyerarşik ÇBMTK modelden edildiği görülmektedir. Alt testler arası korelasyon düzeyi attıkça Hiyerarşik ÇBMTK dışında modellerin kestirim hatalarının genel olarak azaldığı ve üç modele ait sonuçların birbirine yaklaştığı gözlenmektedir. Özellikle alt testler arası korelasyon düzeyinin 0.8 olduğu durumda üç yöntemin benzer hatalar ile parametre kestirimi yaptığı söylenebilir. Alt test uzunluğundaki artışın üç yöntem için de genel olarak parametre kestirim hatasını azalttığı görülmektedir. İki-faktör Model’e ait kestirim hataları ile alt test uzunluğu ve alt testler arası korelasyon arasında azalan doğrusal yönde bir ilişki olduğu gözlenirken Üst Düzey Sıralı Model için bazı koşullarda değişken düzeyde ama çoğu koşul için benzer bir ilişki gözlenmektedir. Hiyerarşik ÇBMTK Model’de ise kestirim hataları ile alt test uzunluğu arasında azalan fakat alt testler arası korelasyon arasında artan doğrusal yönde bir ilişki olduğu görülmektedir.

4.1.2. Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri

Tablo 4.2’de araştırmada ele alınan koşullara dayalı olarak üç boyutlu veri setlerinden toplam test puanı için üç hiyerarşik madde tepki kuramı modeli kullanılarak kestirilen yetenek parametrelerine ilişkin RMSE değerleri verilmiştir.

Tablo 4.2’de verilen alt test uzunluğu ve alt testler arası korelasyon koşullarının tüm düzeylerine ait RMSE değerleri bir arada değerlendirildiğinde Hiyerarşik ÇBMTK modele ait hataların 0.192-0.310 arasında değiştiği görülmektedir. Alt testler arası korelasyon koşulunun ilk iki düzeyinden (0.0, 0.3) elde edilen kestirim hatalarının birbirine yakın değerler aldığı gözlenirken en düşük hata düzeyine sahip kestirimlerin korelasyonun 0.8 düzeyinde elde edildiği gözlenmektedir. Örneğin, alt test uzunluğu 20 iken alt testler arası korelasyonun 0.0, 0.3, 0.5 ve 0.8 düzeyi için elde edilen hataların sırasıyla 0.246, 0.249, 0.261 ve 0.310 olduğu görülmektedir. Bu sonuca göre, alt testler arası korelasyonun 0.8 düzeyi için elde edilen hata, 0.0

düzeyi için elde edilen hatadan yaklaşık olarak %26 daha fazladır. Buna karşılık, alt test uzunluğu arttıkça kestirim hatalarının azaldığı hatta alt testler arası korelasyon düzeyinin ilk iki düzeyinde çok değişmemekle birlikte korelasyonun 0.8 olduğu durumda hataların azalma oranının daha fazla olduğu görülmektedir. Örneğin; alt testler arası korelasyon düzeyi 0.0 ve 0.3 iken alt test uzunluğu 20'den 30'a çıkarıldığında yetenek parametresinin kestirim hatasının yaklaşık olarak %11 oranında azaldığı gözlenirken korelasyon düzeyi 0.5 iken aynı durumda kestirim hatasının yaklaşık olarak %15 oranında, korelasyon düzeyi 0.8 iken yaklaşık %16 oranında azaldığı gözlenmektedir. Alt testler arası korelasyon düzeyinin tüm düzeyleri için alt test uzunluğunu 30'dan 40'a çıkarmanın kestirim hatalarına etkisi ise %10-%12 arasında değişmektedir.

Tablo 4.2.: Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri

<i>Alt test sayısı</i>	<i>Korelasyon</i>	<i>Alt test uzunluğu</i>	<i>İki-faktör model</i>	<i>Hiyerarşik ÇBMTK model</i>	<i>Üst Düzey Sıralı model</i>
3	0.0	20	0.586	0.246	0.410
		30	0.576	0.219	0.445
		40	0.538	0.196	0.417
	0.3	20	0.462	0.249	0.344
		30	0.452	0.216	0.337
		40	0.420	0.192	0.319
	0.5	20	0.373	0.261	0.344
		30	0.359	0.223	0.330
		40	0.337	0.197	0.312
	0.8	20	0.309	0.310	0.308
		30	0.275	0.260	0.273
		40	0.249	0.228	0.246

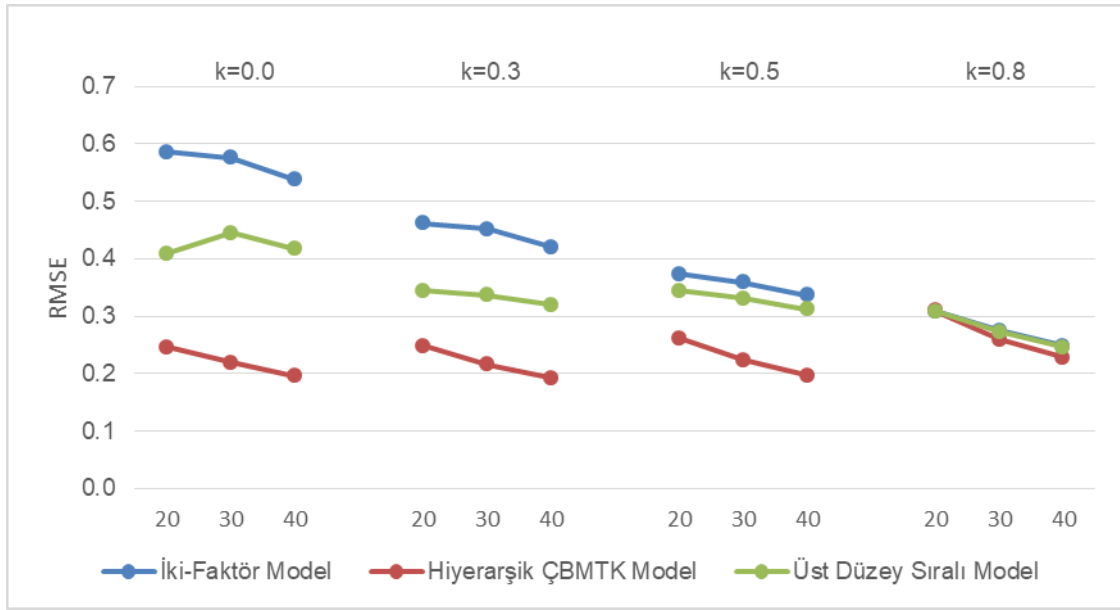
İki-faktör modeline dayalı olarak elde edilen RMSE değerleri incelendiğinde, hataların 0,249-0.586 arasında değiştiği görülmektedir. Alt testler arası korelasyon arttığında alt test uzunluğunun tüm düzeyleri için bu yöntemden elde edilen hataların önemli düzeyde azaldığı dikkat çekmektedir. Örneğin, alt test uzunluğu 20 ve alt testler arası korelasyon 0.0 iken yetenek parametresinin kestirim hatası 0.586 olarak elde edilirken aynı alt test uzunluğu düzeyinde alt testler arası korelasyon 0.8 iken elde edilen hatanın 0.309 olduğu gözlenmektedir. Bu sonuca göre iki korelasyon düzeyi açısından hatanın yaklaşık olarak %47 oranında azaldığı söylenebilir. Alt test

uzunluđu açısından hatalar incelendiđinde, alt test uzunluđu artışı ile hataların azaldığı ve alt testler arası korelasyon arttıkça hataların daha fazla oranda azaldığı görölmektedir. Örneđin, alt testler arası korelasyon 0.0 ve 0.3 iken alt test uzunluđu 20 den 30'a çıkarıldığında İki Faktör modeline dayalı elde edilen yetenek parametresi kestirim hatasının yaklaşık %2 oranında azaldığı gözlenirken alt testler arası korelasyon 0.5 ve 0.8 iken aynı koşullar için kestirim hatasının sırasıyla yaklaşık %4 ve %11 oranında azaldığı gözlenmektedir. Fakat alt testler arası korelasyon koşulunun tüm düzeyleri göz önüne alındığında alt test uzunluđunu 30'dan 40'a çıkarmanın kestirim hatalarının azalmasına etkisi yaklaşık %6-%9 oranları arasındadır.

Üst Düzey Sıralı modele dayalı elde edilen RMSE deđerleri incelendiđinde hataların ise 0.246-0.445 arasında deđiştiiği görölmektedir. Alt test uzunluđu sabit iken alt testler arası korelasyon düzeyi arttıkça yetenek parametresi kestirim hatalarının korelasyonun 0.3 ve 0.5 düzeyi arasında minimal olmakla birlikte azaldığı gözlenirken en yüksek hataların 0.0 düzeyinden, en düşük hataların ise 0.8 düzeyinden elde edildiđi gözlenmektedir. Örneđin, alt test uzunluđu 20 iken alt testler arası korelasyonun her bir düzeyi için elde edilen hataların sırasıyla 0.410, 0.344, 0.344 ve 0.308 olduđu görülürken kestirim hatalarının korelasyonun 0.0 düzeyi ile diđer düzeyleri arasında sırasıyla %16 ve %25 oranında azaldığı görölmektedir. Alt testler arası korelasyon koşulunun her bir düzeyinde alt test uzunluđu artışı ile modelden elde edilen kestirim hatalarının çođunlukla azaldığı görölmektedir. Örneđin, alt testler arası korelasyon düzeyi 0.3, 0.5 ve 0.8 iken alt test uzunluđu 20'den 30'a çıkarıldığında hataların sırasıyla yaklaşık olarak %2, %4 ve %12 oranında azaldığı gözlenirken alt test uzunluđunun 30'dan 40'a çıkarılması ile hataların sırasıyla yaklaşık olarak %5, %6 ve %10 oranında azaldığı gözlenmektedir.

Şekil 4.2'de üç boyutlu veri setleri için alt testler arası korelasyon ve alt test uzunluđu koşulları altında İki-faktör Model, Üst Düzey Sıralı Model ve Hiyerarşik ÇBMTK Model'in yetenek parametresi kestirim hatalarına ait grafiklere yer verilmiştir. Şekil 4.2'ye göre üç kestirim modelinin tüm koşullar altındaki performansları karşılaştırıldığında, en düşük hata düzeyine sahip kestirimlerin İki Faktör modelinden elde edildiđi görülürken en düşük hata düzeyine sahip kestirimlerin Hiyerarşik ÇBMTK modelden edildiđi görölmektedir. Alt testler arası korelasyon

düzeyi attıkça modellerin kestirim hatalarının genel olarak azaldığı ve üç modele ait sonuçların birbirine yaklaştığı gözlenmektedir. Özellikle alt testler arası korelasyon düzeyinin 0.8 olduğu durumda İki-faktör Model ve Üst Düzey Sıralı Model'in aynı olmakla birlikte üç yöntemin benzer hatalar ile parametre kestirimi yaptığı söylenebilir. Alt test uzunluğunun üç yöntem için de genel olarak parametre kestirim hatasını azalttığı; alt testler arası korelasyon düzeyindeki artışın İki Faktör ve Üst Düzey Sıralı Modelde kestirim hatasını azalttığı hiyerarşik ÇBMTK Modelde ise arttırdığı görülmektedir. İki-faktör Model'e ait kestirim hataları ile alt test uzunluğu ve alt testler arası korelasyon arasında azalan doğrusal yönde bir ilişki gözlenirken Üst Düzey Sıralı Model için de 0.0 korelasyon düzeyindeki koşullar haricinde benzer bir ilişki gözlenmektedir. Hiyerarşik ÇBMTK Model'e ait kestirim hataları ile alt test uzunluğu arasında azalan fakat alt testler arası korelasyon ile artan doğrusal yönde bir ilişki olduğu gözlenmektedir.



Şekil 4.2. Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri

Tablo 4.1, Tablo 4.2, Şekil 4.1 ve Şekil 4.2'de verilen iki ve üç boyutlu veri setlerine ait değerler birlikte değerlendirildiğinde alt test sayısındaki artışın üç modelin toplam test puanı için yetenek parametresi kestirim hatalarını azalttığı gözlenmektedir. İki faktör modeli için alt test sayısı ikiden üçe çıkarıldığında yetenek parametresi kestirim hatası alt testler arası korelasyon koşulunun her bir düzeyi için sırasıyla ortalama %3, %12, %17 ve %15 oranında azalmaya neden olurken bu durum

Hiyerarşik ÇBMTK model için sırasıyla ortalama %22, %21, %19 ve %14 oranında, Üst Düzey Sıralı Model için sırasıyla %26, %14, %11 ve %13 oranında azalmaya neden olduğu görülmektedir. Her bir madde tepki kuramı modeli için iki ve üç boyutlu verilerden elde edilen alt test yetenek parametresi kestirimlerine ait RMSE değerlerine ilişkin karşılaştırılmalı grafikler EK-5'te verilmiştir.

4.1.3. İki Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri

Tablo 4.3'te araştırmada ele alınan koşullara dayalı olarak iki boyutlu veri setlerinden toplam test puanı için üç hiyerarşik madde tepki kuramı modeli kullanılarak kestirilen yetenek parametrelerine ilişkin güvenirlik değerleri verilmiştir.

Tablo 4.3.: İki Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri

<i>Alt test sayısı</i>	<i>Korelasyon</i>	<i>Alt test uzunluğu</i>	<i>İki-faktör Model</i>	<i>Hiyerarşik ÇBMTK Model</i>	<i>Üst Düzey Sıralı Model</i>
2	0.0	20	0.305	0.721	0.294
		30	0.339	0.778	0.315
		40	0.407	0.818	0.326
	0.3	20	0.531	0.789	0.690
		30	0.572	0.840	0.731
		40	0.634	0.871	0.732
	0.5	20	0.677	0.818	0.755
		30	0.720	0.865	0.798
		40	0.766	0.892	0.824
	0.8	20	0.834	0.847	0.841
		30	0.873	0.888	0.882
		40	0.896	0.910	0.902

Tablo 4.3'te verilen güvenirlik değerleri incelendiğinde, öncelikle iki boyutlu Hiyerarşik ÇB3PK'ya göre üretilen verilerin tüm koşullar altında en güvenilir parametre kestirimlerinin yine aynı modele dayalı olarak elde edildiği gözlenmektedir. Alt test uzunluğu ve alt testler arası korelasyon koşullarının tüm düzeyleri bir arada değerlendirildiğinde Hiyerarşik ÇBMTK modele ait güvenirliklerin 0.721-0.910 arasında değiştiği görülmektedir. Alt testler arası korelasyon koşulunun en düşük düzeyinde dahi toplam test puanı kestirimlerinin en az 0.7 ile iyi düzeyde güvenirlikle elde edildiği gözlenmekle birlikte korelasyon düzeyi arttıkça modelden elde edilen kestirimlerin güvenirliklerinin arttığı ve en güvenilir

kestirimlerin korelasyonunun 0.8 düzeyinde elde edildiği gözlenmektedir. Örneğin, alt test uzunluğu 20 iken alt testler arası korelasyonun 0.0, 0.3, 0.5 ve 0.8 düzeyi için elde edilen güvenilirliklerin sırasıyla 0.721, 0.789, 0.818 ve 0.847 olduğu görülmektedir. Bu sonuca göre, alt testler arası korelasyonun 0.8 düzeyi için elde edilen güvenilirlik, 0.0 düzeyi için elde edilen güvenilirlikten yaklaşık olarak %17 daha fazladır. Benzer şekilde, korelasyon koşulunun tüm düzeylerinde alt test uzunluğu arttıkça modelden elde edilen kestirimlerin daha güvenilir olduğu görülmektedir. Örneğin; alt testler arası korelasyon koşulunun tüm düzeylerinde alt test uzunluğu 20'den 30'a çıkarıldığında yetenek parametresinin korelasyon koşulunun tüm düzeylerinde sırasıyla yaklaşık olarak %8, %6, %6 ve %5 oranında daha fazla güvenilirlikle kestirildiği gözlenirken alt test uzunluğu 30'dan 40'a çıkarıldığında korelasyon koşulunun tüm düzeylerinde sırasıyla yaklaşık olarak %5, %4, %3 ve %2 oranında daha fazla güvenilirlikle kestirildiği gözlenmektedir.

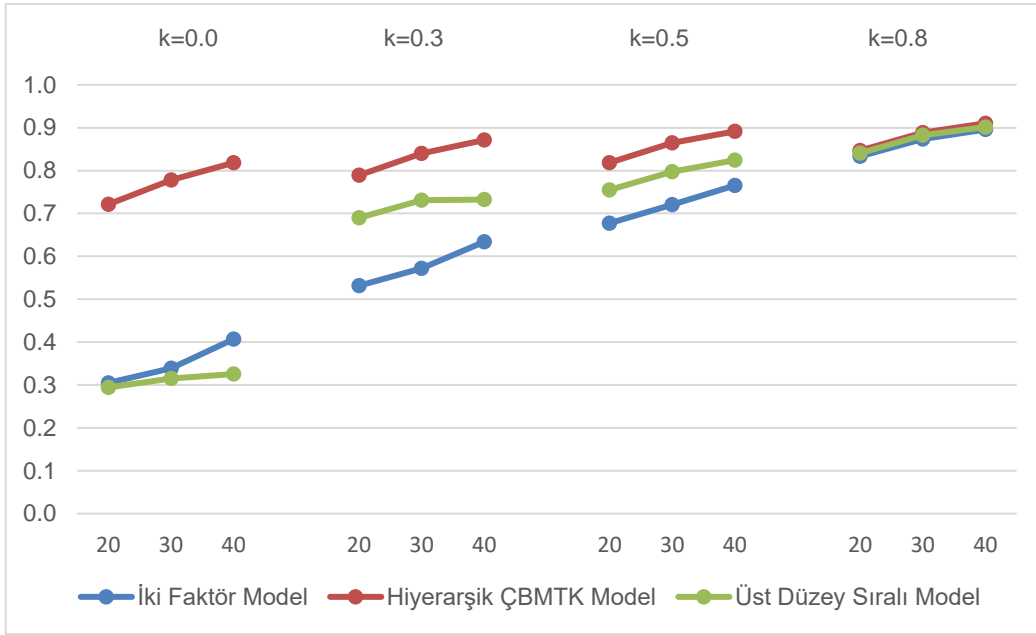
İki-faktör modeline dayalı olarak elde edilen güvenilirlik değerleri incelendiğinde, alt test uzunluğu ve alt testler arası korelasyon koşullarının tüm düzeyleri bir arada değerlendirildiğinde modele ait güvenilirliklerin 0.305-0.896 arasında değiştiği görülmektedir. Alt testler arası korelasyon arttığında alt test uzunluğunun tüm düzeyleri için bu yöntemden elde edilen kestirimlerin güvenilirliğinin önemli düzeyde arttığı dikkat çekmektedir. Fakat kabul edilebilir düzeyde güvenilir kestirimlerin en az 0.5 korelasyon düzeyinde 30 madde için elde edildiği görülmektedir. Örneğin, yetenek parametresinin alt test uzunluğu 20 iken alt testler arası korelasyon düzeyleri için sırasıyla 0.305, 0.531, 0.677 ve 0.834 düzeyinde güvenilirlik ile kestirildiği gözlenirken alt test uzunluğu 40 iken sırasıyla 0.407, 0.634, 0.766 ve 0.896 düzeyinde güvenilirlik ile kestirildiği gözlenmektedir. Bu sonuca göre alt test uzunluğu koşulunun tüm düzeyleri için alt testler arası korelasyonun 0.8 düzeyinde elde edilen yetenek parametresi kestirimlerinin korelasyonun 0.0 düzeyine göre %100'den daha fazla oranda, 0.5 düzeyine göre ise yaklaşık olarak %23 oranında daha güvenilir olduğu söylenebilir. Alt test uzunluğu açısından güvenilirlikler incelendiğinde, alt test uzunluğu artışı ile güvenilirliğin arttığı fakat alt testler arası korelasyon arttıkça güvenilirliklerin daha az oranda arttığı görülmektedir. Örneğin, alt test uzunluğu 20 den 30'a çıkarıldığında İki Faktör modeline dayalı elde edilen yetenek parametresi kestiriminin alt testler arası korelasyon düzeylerine göre sırasıyla yaklaşık olarak %11, %8, %6 ve %5 oranında daha fazla güvenilirlikle

kestirildiđi; alt test uzunluđu 30'dan 40'a ıkarıldıđında ise sırasıyla yaklaşık olarak %20, %11, %6 ve %3 oranında daha fazla güvenilirlikle kestirildiđi gözlenmektedir.

Üst Düzey Sıralı modele dayalı olarak elde edilen güvenilirlik deđerleri incelendiđinde, alt test uzunluđu ve alt testler arası korelasyon koşullarının tüm düzeyleri bir arada deđerlendirildiđinde modele ait güvenilirliklerin 0.294-0.902 arasında deđiřtiđi görülmektedir. Alt testler arası korelasyon arttıđında alt test uzunluđunun tüm düzeyleri için bu yöntemden elde edilen yetenek parametresi kestirimlerin güvenilirliđinin arttıđı ve en güvenilir kestirimlerin korelasyonun 0.8 düzeyinde elde edildiđi gözlenmektedir. Ayrıca kabul edilebilir düzeyde güvenilir kestirimlerin en az 0.3 korelasyon düzeyinde elde edildiđi dikkat çekmektedir. Örneđin, alt test uzunluđu 20 iken alt testler arası korelasyonun 0.0, 0.3, 0.5 ve 0.8 düzeyi için elde edilen güvenilirliklerin sırasıyla 0.294, 0.690, 0.755 ve 0.841 olduđu görülmektedir. Bu sonuca göre, alt testler arası korelasyonun 0.8 düzeyi için elde edilen yetenek parametresi kestirimlerinin, 0.0 düzeyine göre %100'den daha fazla oranda, 0.3 düzeyine göre ise yaklaşık %22 oranında daha güvenilir elde edildiđi söylenebilir. Benzer řekilde, korelasyon koşulunun tüm düzeylerinde alt test uzunluđu arttıđıca modelden elde edilen kestirimlerin daha güvenilir olduđu görülmektedir. Örneđin; alt testler arası korelasyon koşulunun tüm düzeylerinde alt test uzunluđu 20'den 30'a ıkarıldıđında yetenek parametresinin yaklaşık olarak %5-7 arası oranlarda daha fazla güvenilirlikle kestirildiđi gözlenirken alt test uzunluđu 30'dan 40'a ıkarıldıđında korelasyon koşulunun tüm düzeylerinde sırasıyla yaklaşık olarak %0-%3 arası oranlarda daha fazla güvenilirlikle kestirildiđi gözlenmektedir.

řekil 4.3'te iki boyutlu veri setleri için alt testler arası korelasyon ve alt test uzunluđu koşulları altında İki-faktör Model, Üst Düzey Sıralı Model ve Hiyerarřik BMTK Model'in yetenek parametresi kestirimine iliřkin güvenilirlik deđerlerine ait grafiklere yer verilmiřtir. řekil 4.3'e göre üç kestirim modelinin tüm koşullar altındaki performansları karřılařtırıldıđında kabul edilebilir en güvenilir kestirimlerin Hiyerarřik BMTK modelden edildiđi görülmektedir. Alt testler arası korelasyon düzeyi attıkça modellerin yetenek parametresi kestirim güvenilirliđinin arttıđı, Üst Düzey Sıralı Model kestirim güvenilirliđinin korelasyonun 0.3 ve üzeri düzeylerde kabul edilebilir düzeylerde olduđu ve üç modelin kestirim güvenilirliđinin korelasyonun 0.5 ve üzeri düzeylerde birbirine yaklařtıđı gözlenmektedir. Özellikle alt testler arası korelasyon düzeyinin 0.8 olduđu durumda üç yöntemin benzer

güvenirlik ile parametre kestirimi yaptığı söylenebilir. Alt test uzunluğu ve alt testler arası korelasyon düzeyindeki artışın her üç yöntem için de genel olarak parametre kestirim güvenilirliğini arttırdığı görülmektedir. İki-faktör Model ve Üst Düzey Sıralı Model'in alt testler arası korelasyonun 0.0 düzeyinde yetenek parametre kestirimlerinin kabul edilemez düzeyde güvenilirlik ile kestirildiği dikkat çekmektedir. Her üç modele ait kestirim güvenilirliği ile alt test uzunluğu ve alt testler arası korelasyon arasında artan doğrusal yönde bir ilişki olduğu gözlenmektedir.



Şekil 4.3. İki Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri

4.1.4. Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri

Tablo 4.4'te araştırmada ele alınan koşullara dayalı olarak üç boyutlu veri setlerinden toplam test puanı için üç hiyerarşik madde tepki kuramı modeli kullanılarak kestirilen yetenek parametrelerine ilişkin güvenilirlik değerleri verilmiştir.

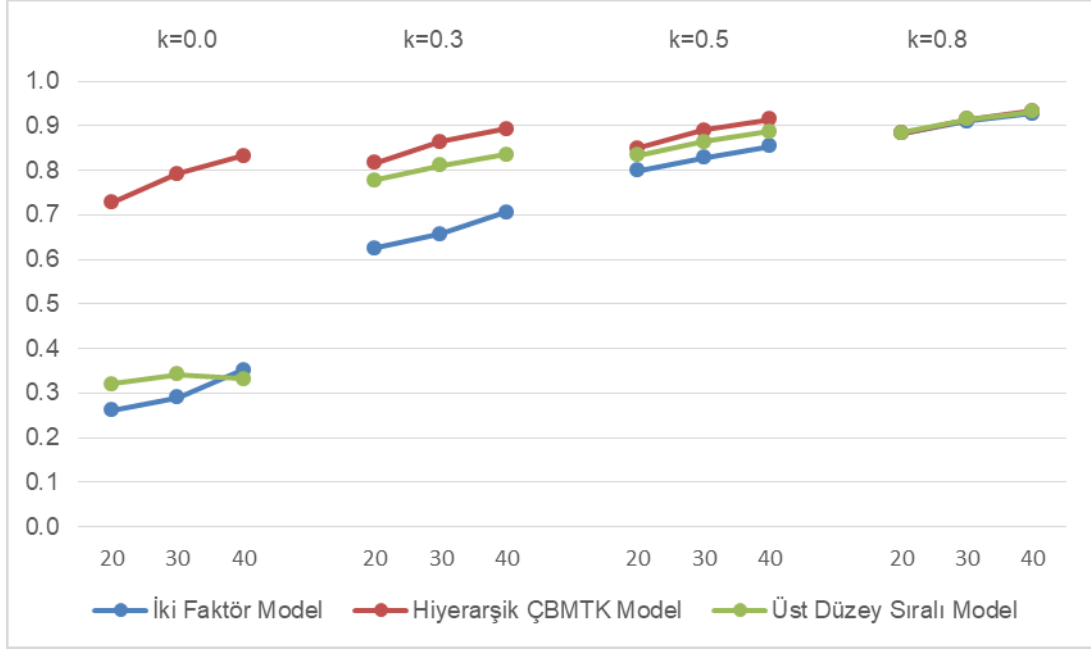
Tablo 4.4.: Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri

<i>Alt test sayısı</i>	<i>Korelasyon</i>	<i>Alt test uzunluğu</i>	<i>İki-faktör Model</i>	<i>Hiyerarşik ÇBMTK Model</i>	<i>Üst Düzey Sıralı Model</i>
3	0.0	20	0.262	0.729	0.321
		30	0.290	0.793	0.342
		40	0.353	0.833	0.333
	0.3	20	0.626	0.816	0.778
		30	0.657	0.864	0.811
		40	0.707	0.893	0.836
	0.5	20	0.800	0.851	0.835
		30	0.829	0.891	0.865
		40	0.854	0.915	0.887
	0.8	20	0.883	0.883	0.885
		30	0.911	0.915	0.915
		40	0.929	0.934	0.933

İki-faktör modeline dayalı olarak elde edilen güvenilirlik değerleri incelendiğinde, alt test uzunluğu ve alt testler arası korelasyon koşullarının tüm düzeyleri bir arada değerlendirildiğinde modele ait güvenilirliklerin 0.262-0.929 arasında değiştiği görülmektedir. Alt testler arası korelasyon arttığında alt test uzunluğunun tüm düzeyleri için bu yöntemden elde edilen kestirimlerin güvenilirliğinin önemli düzeyde arttığı ancak kabul edilebilir düzeyde güvenilir kestirimlerin en az 0.3 korelasyon düzeyinde 40 madde için elde edildiği görülmektedir. Örneğin, yetenek parametresinin alt test uzunluğu 20 iken alt testler arası korelasyon düzeyleri için sırasıyla 0.262, 0.626, 0.800 ve 0.883 düzeyinde güvenilirlik ile kestirildiği gözlenirken alt test uzunluğu 40 iken sırasıyla 0.353, 0.707, 0.854 ve 0.929 düzeyinde güvenilirlik ile kestirildiği gözlenmektedir. Bu sonuca göre alt test uzunluğu koşulunun tüm düzeyleri için alt testler arası korelasyonun 0.8 düzeyinde elde edilen yetenek parametresi kestirimlerinin korelasyonun 0.0 düzeyine göre %100'den daha fazla, 0.5 düzeyi için ise yaklaşık olarak %10 oranında daha güvenilir olduğu söylenebilir. Alt test uzunluğu açısından güvenilirlikler incelendiğinde, alt test uzunluğu artışı ile güvenilirliğin arttığı fakat alt testler arası korelasyon arttıkça güvenilirliklerin daha az oranda arttığı görülmektedir. Örneğin, alt test uzunluğu 20 den 30'a çıkarıldığında İki Faktör modeline dayalı elde edilen yetenek parametresi kestiriminin alt testler arası korelasyon düzeylerine göre sırasıyla yaklaşık olarak %11, %5, %4 ve %3 oranında daha fazla güvenilirlikle kestirildiği; alt test uzunluğu

30'dan 40'a çıkarıldığında ise sırasıyla yaklaşık olarak %21, %8, %3 ve %2 oranında daha fazla güvenilirlikle kestirildiği gözlenmektedir.

Üst Düzey Sıralı modele dayalı olarak elde edilen güvenilirlik değerleri incelendiğinde, alt test uzunluğu ve alt testler arası korelasyon koşullarının tüm düzeyleri bir arada değerlendirildiğinde modele ait güvenilirliklerin 0.321-0.933 arasında değiştiği görülmektedir. Alt testler arası korelasyon arttığında alt test uzunluğunun tüm düzeyleri için bu yöntemden elde edilen yetenek parametresi kestirimlerin güvenilirliğinin arttığı ve en güvenilir kestirimlerin korelasyonun 0.8 düzeyinde elde edildiği gözlenmektedir. Ayrıca kabul edilebilir düzeyde güvenilir kestirimlerin en az 0.3 korelasyon düzeyinde elde edildiği dikkat çekmektedir. Örneğin, alt test uzunluğu 20 iken alt testler arası korelasyonun 0.0, 0.3, 0.5 ve 0.8 düzeyi için elde edilen güvenilirliklerin sırasıyla 0.321, 0.778, 0.835 ve 0.885 olduğu görülmektedir. Bu sonuca göre, alt testler arası korelasyonun 0.8 düzeyi için elde edilen yetenek parametresi kestirimlerinin, 0.0 düzeyine göre %100 oranından fazla, 0.3 düzeyine göre ise yaklaşık %14 oranında daha güvenilir elde edildiği söylenebilir. Benzer şekilde, korelasyon koşulunun tüm düzeylerinde alt test uzunluğu arttıkça modelden elde edilen kestirimlerin daha güvenilir olduğu görülmektedir. Örneğin; alt test uzunluğu 20'den 30'a çıkarıldığında yetenek parametresinin alt testler arası korelasyon koşulunun 0.0 düzeyinde değişken sonuçlar elde edilmekle birlikte diğer düzeylerde sırasıyla yaklaşık olarak %4, %4 ve %3 oranında daha fazla güvenilirlikle kestirildiği gözlenirken alt test uzunluğu 30'dan 40'a çıkarıldığında korelasyon koşulunun 0.3, 0.5 ve 0.8 düzeylerinde sırasıyla yaklaşık olarak %3 oranında daha fazla güvenilirlikle kestirildiği gözlenmektedir.



Şekil 4.4. Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri

Şekil 4.4'te üç boyutlu veri setleri için alt testler arası korelasyon ve alt test uzunluğu koşulları altında İki-faktör Model, Üst Düzey Sıralı Model ve Hiyerarşik ÇBMTK Model'in yetenek parametresi kestirimine ilişkin güvenirlik değerlerine ait grafiklere yer verilmiştir. Şekil 4.3'e göre üç kestirim modelinin tüm koşullar altındaki performansları karşılaştırıldığında kabul edilebilir en güvenilir kestirimlerin Hiyerarşik ÇBMTK modelden edildiği görülmektedir. Alt testler arası korelasyon düzeyi attıkça modellerin yetenek parametresi kestirim güvenirliliğinin arttığı, Üst Düzey Sıralı Model kestirim güvenirliliğinin korelasyonun 0.3 ve üzeri düzeylerde kabul edilebilir düzeylerde olduğu ve üç modelin kestirim güvenirliliğinin korelasyonun 0.5 ve üzeri düzeylerde birbirine yaklaştığı gözlenmektedir. Özellikle alt testler arası korelasyon düzeyinin 0.8 olduğu durumda üç yöntemin aynı güvenirlik ile parametre kestirimi yaptığı söylenebilir. Alt test uzunluğu ve alt testler arası korelasyon düzeyindeki artışın her üç yöntem için de genel olarak parametre kestirim güvenirliliğini arttırdığı görülmektedir. İki-faktör Model ve Hiyerarşik ÇBMTK Model'e ait kestirim güvenirliliği ile alt test uzunluğu ve alt testler arası korelasyon arasında artan doğrusal yönde bir ilişki olduğu gözlenmektedir. Üst Düzey Sıralı Model içinse korelasyonun 0.0 düzeyinde değişken düzeyde bir ilişki olduğu fakat diğer düzeylerde benzer şekilde artan doğrusal bir ilişki olduğu görülmektedir.

Tablo 4.3, Tablo 4.4, Şekil 4.3 ve Şekil 4.4'te verilen iki ve üç boyutlu veri setlerine ait değerler birlikte değerlendirildiğinde boyut sayısındaki artışın üç modelin toplam test puanı için yetenek parametresi kestirim güvenilirliğini koşulların çoğunda arttırdığı gözlenmektedir. İki faktör modelde alt test sayısındaki artışın kabul edilebilir düzeyde parametre kestirimi koşullarında iyileşme sağladığı görülmektedir. İki Faktör Model için alt test sayısını ikiden üçe çıkarmanın yetenek parametresi kestirim güvenilirliğinin alt testler arası korelasyon düzeyinin 0.0 olduğu durum dışındaki diğer düzeyleri için sırasıyla ortalama %15, %15 ve %5 oranında artışa neden olduğu görülmektedir. Bu durum Hiyerarşik ÇBMTK model için alt testler arası korelasyon koşulunun tüm düzeylerinde ortalama %3 oranında artışa neden olurken Üst Düzey Sıralı Model için boyut sayısı artışı ile alt testler arası korelasyon koşulunun tüm düzeyleri için yetenek parametresi kestirim güvenilirliğinde sırasıyla ortalama %7, %13, %9 ve %4 oranında artış olduğu gözlenmektedir. Her bir madde tepki kuramı modeli için iki ve üç boyutlu verilerden elde edilen toplam test yetenek parametresi kestirimlerine ait güvenilirlik değerlerine ilişkin karşılaştırılmalı grafikler EK-6'da verilmiştir.

4.1.5. Birinci Alt Probleme İlişkin Tartışma

İki ve üç boyutlu Hiyerarşik ÇB3PK'ya göre üretilen verilerin aynı modele dayalı elde edilen toplam test yetenek parametresi kestirimlerinin belirli bir hata düzeyinde elde edildiği gözlenmiştir. Simülasyon çalışmalarında verilerin belirli bir hata düzeyinde üretilmesi beklenen bir durumdur. RMSE istatistiği için belirli bir sınır olmadığından yalnızca daha düşük değerlerin daha iyi olduğu belirtilmektedir. Parametre doğrulama çalışmalarında yetenek parametresine ait hataların madde parametresine ait hatalardan daha yüksek elde edildiği görülmektedir (Çakıcı Eser, 2014; Jiang, Wang & Weiss,2016; Lee, 2012). Ayrıca de la Torre ve Patz (2005) ile Yao'nun (2010) çok boyutlu MTK'ya dayalı ürettikleri verileri ve de la Torre, Song ve Hong'un (2011) Üst Düzey Sıralı Modele dayalı ürettikleri verileri aynı model ile analiz etmeleri sonucu yetenek parametre kestirimlerinde bu araştırma ile benzer düzeyde hataların elde edildiği görülmüştür. Bu durumun veri üretilirken yetenek parametresinin geniş bir normal dağılımdan gelmesi ve az sayıda madde örneklemini birey yeteneğinin kestirilmesinden kaynaklandığı düşünülmektedir.

Hem iki hem de üç boyutlu veri setlerinde alt testler arasındaki korelasyon düzeyinin artmasıyla Hiyerarşik ÇBMTK Model'den elde edilen hataların artmasının nedeni

olarak testin boyutluluk derecesindeki azalma gösterilebilir. Bir başka ifadeyle alt testler arasındaki yüksek düzeyde ilişkiler testin tek boyutluluğa yaklaşmasına neden olmaktadır. Aynı nedenle, İki Faktör Model’de toplam test puanı kestirimlerinde alt testler arasındaki korelasyon düzeyi arttıkça hataların azaldığı görülmektedir. Çünkü İki Faktör Model’de toplam test puanı testteki tüm maddelerin kalibrasyonundan elde edilir. Korelasyon düzeyindeki artışı ile Üst Düzey Sıralı Model’de toplam test puanı kestirimlerinde hataların azalmasının nedeni ise modelde kullanılan regresyon katsayılarının alt testler arası ilişkilerden türetilmesidir. Yukarıdaki sayılan benzer nedenler ile hem iki hem üç boyutlu verilerde İki Faktör ve Üst Düzey Sıralı Model için korelasyon arttıkça toplam test güvenilirliği artmaktadır. Ayrıca alt testler arası korelasyonlar yüksek olsa da yapı modelinin maddeleri tek bir boyutla ilişkilendirmesi ve verilerin yine aynı modelle üretilmesi nedeniyle tüm koşullarda Hiyerarşik ÇBMTK Model en hatasız ve en güvenilir sonuçlar vermektedir. Daha uzun alt testlerde üç yöntem için de kestirim hatalarının azalması ve güvenilirliğin artması beklenen bir durumdur.

Hiyerarşik ÇBMTK Model’in toplam test puanı kestirimlerinde maksimum bilgi yöntemini kullanması nedeniyle alt test sayısındaki artışın yetenek parametre kestirimleri üzerinde en fazla katkısı olan model bu modeldir. Alt test sayısındaki artışın toplam madde sayısını arttırdığı için İki Faktör Model’de toplam test yetenek kestirimlerine katkısı alt testler arası korelasyon arttıkça daha fazla artmaktadır. Üst Düzey sıralı modelde ise genel yetenek kestirimde kullanılan birinci düzey değişken sayısının artması nedeniyle alt test sayısındaki artış kestirim hatalarını azaltmakta ve güvenilirliği arttırmaktadır.

4.2. İkinci Alt Probleme İlişkin Bulgular ve Tartışma

Araştırmanın ikinci alt problemi kapsamında “Üst Düzey Sıralı (Higher Order), İki Faktör (Bi-factor) ve Hiyerarşik çok boyutlu madde tepki kuramı (ÇBMTK) modellerine göre **alt test puanı** için kestirilen yetenek parametrelerinin **alt test sayısı (2,3), alt test uzunluğu (20, 30, 40) ve alt testler arasındaki korelasyonların büyüklüğü (0.0, 0.3, 0.5, 0.8)** koşullarından nasıl etkilendiği” güvenilirlik ve RMSE değerleri ile incelenmiştir. Yukarıda verilen test koşullarına ait RMSE ve güvenilirlik değerleri alt test sayısı koşuluna göre ayrı ayrı tablo ve grafikler ile verilmiştir.

4.2.1. İki Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri

Tablo 4.5'te araştırmada ele alınan koşullara dayalı olarak iki boyutlu veri setlerinden alt test puanı için üç hiyerarşik madde tepki kuramı modeli kullanılarak kestirilen yetenek parametrelerine ilişkin RMSE değerleri verilmiştir. Tablo 4.5'de her bir madde tepki kuramı modeline ilişkin verilen değerler iki alt testten kestirilen yetenek parametresine ait RMSE değerlerinin ortalamasını göstermektedir. Her bir alt testten elde edilen RMSE değerleri EK-7.1'de verilmiştir.

Tablo 4.5.: İki Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri

<i>Alt test sayısı</i>	<i>Korelasyon</i>	<i>Alt test uzunluğu</i>	<i>İki-faktör model</i>	<i>Hiyerarşik ÇBMTK model</i>	<i>Üst Düzey Sıralı model</i>
2	0.0	20	0.553	0.477	0.480
		30	0.502	0.411	0.413
		40	0.473	0.369	0.371
	0.3	20	0.613	0.476	0.475
		30	0.571	0.410	0.409
		40	0.551	0.369	0.368
	0.5	20	0.684	0.476	0.466
		30	0.653	0.411	0.404
		40	0.639	0.369	0.363
	0.8	20	0.830	0.476	0.435
		30	0.823	0.410	0.378
		40	0.820	0.369	0.343

Tablo 4.5'te verilen RMSE değerleri incelendiğinde, öncelikle Hiyerarşik ÇB3PK'ya göre üretilen verilerin aynı modele dayalı elde edilen parametre kestirimlerinin belirli bir hata düzeyinde elde edildiği gözlenmektedir. Alt test uzunluğu ve alt testler arası korelasyon koşullarının tüm düzeyleri bir arada değerlendirildiğinde Hiyerarşik ÇBMTK modele ait hataların 0.369-0.477 arasında değiştiği görülmektedir. Alt test uzunluğu sabit tutulduğunda alt testler arası korelasyonun tüm düzeyleri için alt test yetenek kestirim hataları arasında fark olmadığı dikkat çekmektedir. Bunun yanında, alt test uzunluğu arttıkça hataların azaldığı gözlenmektedir. Örneğin, alt test uzunluğu 20 iken alt testler arası korelasyonun her bir düzeyi (0.0, 0.3, 0.5, 0.8) için elde edilen hataların sırasıyla 0.477, 0.476, 0.476 ve 0.476 olduğu görülürken, alt testler arası korelasyonun herhangi bir düzeyi için 20, 30 ve 40 maddelik alt testlerden elde edilen hataların sırasıyla 0.477, 0.411, 0.369 olduğu görülmektedir. Bu sonuca göre, alt testler arası korelasyon koşulundan bağımsız olarak alt test uzunluğu 20'den 30'a çıkarıldığında alt test yetenek parametresinin kestirim

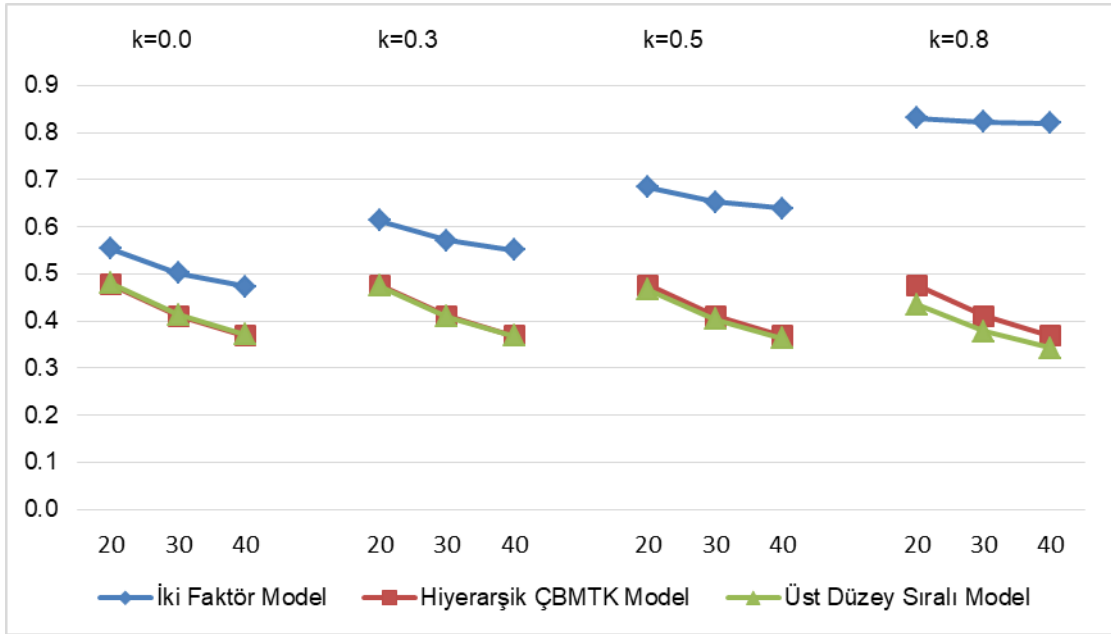
hatasının %14 oranında ve alt test uzunluğu 30'dan 40'a çıkarıldığında ise %10 oranında azaldığı gözlenmektedir.

İki-faktör modeline dayalı olarak elde edilen RMSE değerleri incelendiğinde, hataların 0,473-0.830 arasında değiştiği görülmektedir. Alt testler arası korelasyon arttığında alt test uzunluğunun tüm düzeyleri için bu yöntemden elde edilen hataların önemli düzeyde arttığı dikkat çekmektedir. Örneğin, alt test uzunluğu 20 ve alt testler arası korelasyon 0.0 iken alt test yetenek parametresinin kestirim hatası 0.553 olarak elde edilirken aynı alt test uzunluğu düzeyinde alt testler arası korelasyon 0.8 iken elde edilen hatanın 0.830 olduğu gözlenmektedir. Bu sonuca göre iki korelasyon düzeyi açısından hatanın yaklaşık olarak %50 oranında arttığı söylenebilir. Alt test uzunluğu açısından hatalar incelendiğinde, alt test uzunluğu artışı ile hataların azaldığı ve alt testler arası korelasyon arttıkça hataların daha az oranda azaldığı görülmektedir. Örneğin, alt testler arası korelasyon 0.0 iken alt test uzunluğu 20 den 30'a çıkarıldığında İki Faktör modeline dayalı elde edilen alt test yetenek parametresi kestirim hatasının yaklaşık %9 oranında azaldığı gözlenirken aynı koşullar için alt testler arası korelasyon 0.3 iken kestirim hatasının yaklaşık %7 oranında, 0.5 iken kestirim hatasının yaklaşık %5 oranında ve 0.8 iken yaklaşık %1 oranında azaldığı gözlenmektedir. Fakat alt testler arası korelasyon koşulunun tüm düzeyleri göz önüne alındığında alt test uzunluğunu 30'dan 40'a çıkarmanın alt test yetenek parametresi kestirim hatalarının azalmasına etkisi alt testler arası korelasyon düzeylerine göre sırasıyla yaklaşık %6, %4, %2 ve %0 oranları arasındadır.

Üst Düzey Sıralı modele dayalı elde edilen RMSE değerleri incelendiğinde hataların ise 0.343-0.480 arasında değiştiği görülmektedir. Alt test uzunluğu sabit iken alt testler arası korelasyon düzeyi arttıkça korelasyonun ilk üç düzeyi arasında minimal olmakla birlikte alt test yetenek parametresi kestirim hatalarının azaldığı gözlenirken en yüksek hataların 0.0 düzeyinden, en düşük hataların ise 0.8 düzeyinden elde edildiği gözlenmektedir. Örneğin, alt test uzunluğu 20 iken alt testler arası korelasyonun her bir düzeyi için elde edilen hataların sırasıyla 0.480, 0.475, 0.466 ve 0.435 olduğu görülürken kestirim hatalarının korelasyonun 0.0 düzeyi ile diğer düzeyleri arasında sırasıyla %1, %3 ve %9 oranında azaldığı görülmektedir. Alt test uzunluğu 20 iken bu iki korelasyon düzeyinden elde edilen parametre kestirim hatası sırasıyla 0.480 ve 0.435'tir ve bu iki hata arasındaki fark yaklaşık olarak %9

oranındadır. Alt test uzunluğu artışı ile modelden elde edilen kestirim hatalarının azaldığı; fakat bu durumun alt testler arası korelasyon düzeyine göre değişkenlik göstermediği görülmektedir. Örneğin, alt testler arası korelasyon koşulunun herhangi bir düzeyinde alt test uzunluğu 20'den 30'a çıkarıldığında alt test yetenek parametresi kestirim hatasının yaklaşık olarak %13-%14 arası oranda azaldığı gözlenirken alt test uzunluğunun 30'dan 40'a çıkarılması ile hataların yaklaşık olarak %9-%10 arası oranda azaldığı gözlenmektedir.

Şekil 4.5'te iki boyutlu veri setleri için alt testler arası korelasyon ve alt test uzunluğu koşulları altında İki-faktör Model, Üst Düzey Sıralı Model ve Hiyerarşik ÇBMTK Model'in alt test yetenek parametresi kestirim hatalarına ait grafiklere yer verilmiştir.



Şekil 4.5. İki Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri

Şekil 4.5'e göre üç kestirim modelinin tüm koşullar altındaki performansları karşılaştırıldığında, en düşük hata düzeyine sahip kestirimlerin İki Faktör modelinden elde edildiği görülürken en düşük hata düzeyine sahip kestirimlerin Üst Düzey Sıralı modelden edildiği görülmektedir. Alt testler arası korelasyon düzeyi attıkça Üst Düzey Sıralı Model ve Hiyerarşik ÇBMTK Model için alt test yetenek kestirim hatalarının azaldığı gözlenirken İki-faktör Model için hataların arttığı gözlenmektedir. Alt testler arası korelasyon koşulunun ilk iki düzeyinde Üst Düzey Sıralı Model ve Hiyerarşik ÇBMTK Model'in alt test yetenek kestirim hatalarının benzer olduğu görülürken korelasyon düzeyi arttıkça Üst Düzey Sıralı Model'in daha

düşük hatalı kestirimler yaptığı görülmektedir. İki-faktör Model'e ait kestirim hataları ile alt test uzunluğu arasında azalan fakat alt testler arası korelasyon arasında artan doğrusal yönde bir ilişki olduğu gözlenirken Üst Düzey Sıralı Model'e ait kestirim hataları ile alt test uzunluğu ve alt testler arası korelasyon koşulları arasında azalan doğrusal yönde bir ilişki olduğu gözlenmektedir. Hiyerarşik ÇBMTK Model'e ait kestirim hataları ile korelasyon koşulu arasında bir ilişki olmadığı ama alt test uzunluğu ile kestirim hataları arasında azalan doğrusal yönde bir ilişki olduğu görülmektedir.

4.2.2. Üç Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri

Tablo 4.6'da araştırmada ele alınan koşullara dayalı olarak üç boyutlu veri setlerinden alt test puanı için üç hiyerarşik madde tepki kuramı modeli kullanılarak kestirilen yetenek parametrelerine ilişkin RMSE değerleri verilmiştir. Tablo 4.6'da her bir madde tepki kuramı modeline ilişkin verilen değerler üç alt testten kestirilen yetenek parametresine ait RMSE değerlerinin ortalamasını göstermektedir. Her bir alt testten elde edilen RMSE değerleri EK-7.2'de verilmiştir.

Tablo 4.6.: Üç Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri

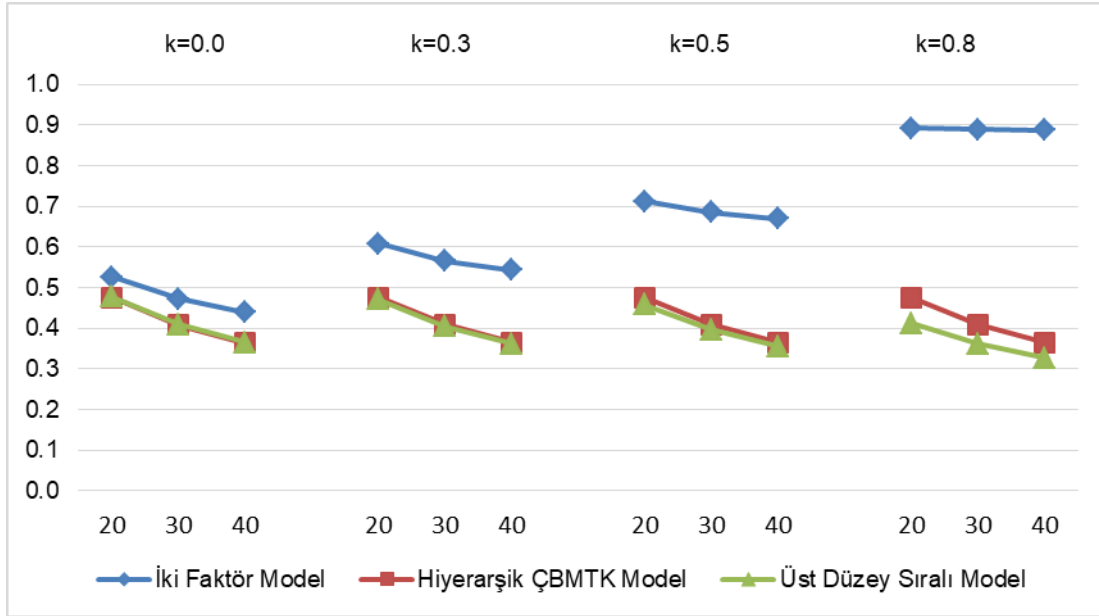
<i>Alt test sayısı</i>	<i>Korelasyon</i>	<i>Alt test uzunluğu</i>	<i>İki-faktör model</i>	<i>Hiyerarşik ÇBMTK model</i>	<i>Üst Düzey Sıralı model</i>	
3	0.0	20	0.527	0.475	0.478	
		30	0.472	0.408	0.410	
		40	0.439	0.364	0.365	
	0.3	20	0.608	0.475	0.475	0.471
		30	0.566	0.409	0.409	0.405
		40	0.544	0.365	0.365	0.362
	0.5	20	0.713	0.476	0.476	0.458
		30	0.686	0.409	0.409	0.396
		40	0.670	0.365	0.365	0.355
	0.8	20	0.892	0.476	0.476	0.413
		30	0.890	0.409	0.409	0.362
		40	0.888	0.365	0.365	0.327

Tablo 4.6'da verilen RMSE değerleri incelendiğinde, öncelikle Hiyerarşik ÇB3PK'ya göre üretilen verilerin aynı modele dayalı elde edilen parametre kestirimlerinin belirli bir hata düzeyinde elde edildiği gözlenmektedir. Alt test uzunluğu ve alt testler arası

korelasyon koşullarının tüm düzeyleri bir arada değerlendirildiğinde Hiyerarşik ÇBMTK modele ait hataların 0.365-0.475 arasında değiştiği görülmektedir. Alt test uzunluğu sabit tutulduğunda alt testler arası korelasyonun tüm düzeyleri için alt test yetenek kestirim hataları arasında fark olmadığı dikkat çekmektedir. Bunun yanında, alt test uzunluğu arttıkça hataların azaldığı gözlenmektedir. Örneğin, alt test uzunluğu 20 iken alt testler arası korelasyonun her bir düzeyi için elde edilen hataların sırasıyla 0.475, 0.475, 0.476 ve 0.476 olduğu görülürken, alt testler arası korelasyonun herhangi bir düzeyi için 20, 30 ve 40 maddelik alt testlerden elde edilen kestirim hatalarının sırasıyla 0.476, 0.409, 0.365 olduğu görülmektedir. Bu sonuca göre, alt testler arası korelasyon koşulundan bağımsız olarak alt test uzunluğu 20'den 30'a çıkarıldığında alt test yetenek parametresinin kestirim hatasının %14 oranında ve alt test uzunluğu 30'dan 40'a çıkarıldığında ise %11 oranında azaldığı gözlenmektedir.

İki-faktör modeline dayalı olarak elde edilen RMSE değerleri incelendiğinde, hataların 0,439-0.888 arasında değiştiği görülmektedir. Alt testler arası korelasyon arttığında alt test uzunluğunun tüm düzeyleri için bu yöntemden elde edilen hataların önemli düzeyde arttığı dikkat çekmektedir. Örneğin, alt test uzunluğu 20 ve alt testler arası korelasyon 0.0 iken alt test yetenek parametresinin kestirim hatası 0.527 olarak elde edilirken aynı alt test uzunluğu düzeyinde alt testler arası korelasyon 0.8 iken elde edilen hatanın 0.880 olduğu gözlenmektedir. Bu sonuca göre iki korelasyon düzeyi açısından hatanın yaklaşık olarak %67 oranında arttığı söylenebilir. Alt test uzunluğu açısından hatalar incelendiğinde, alt test uzunluğu artışı ile hataların azaldığı ve alt testler arası korelasyon arttıkça hataların daha az oranda azaldığı görülmektedir. Örneğin, alt testler arası korelasyon 0.0 iken alt test uzunluğu 20 den 30'a çıkarıldığında İki Faktör modeline dayalı elde edilen alt test yetenek parametresi kestirim hatasının yaklaşık %10 oranında azaldığı gözlenirken aynı koşullar için alt testler arası korelasyon 0.3 iken kestirim hatasının yaklaşık %7 oranında, 0.5 iken kestirim hatasının yaklaşık %4 oranında ve 0.8 iken yaklaşık %0 oranında azaldığı gözlenmektedir. Fakat alt testler arası korelasyon koşulunun tüm düzeyleri göz önüne alındığında alt test uzunluğunu 30'dan 40'a çıkarmanın alt test yetenek parametresi kestirim hatalarının azalmasına etkisi alt testler arası korelasyon düzeylerine göre sırasıyla yaklaşık %7, %4, %2 ve %0 oranları arasındadır.

Üst Düzey Sıralı modele dayalı elde edilen RMSE değerleri incelendiğinde hataların ise 0.327-0.478 arasında değiştiği görülmektedir. Alt test uzunluğu sabit iken alt testler arası korelasyon düzeyi arttıkça korelasyonun ilk üç düzeyinde minimal olmakla birlikte alt test yetenek parametresi kestirim hatalarının azaldığı gözlenirken en yüksek hataların 0.0 düzeyinden, en düşük hataların ise 0.8 düzeyinden elde edildiği gözlenmektedir. Örneğin, alt test uzunluğu 20 iken alt testler arası korelasyonun her bir düzeyi için elde edilen hataların sırasıyla 0.478, 0.471, 0.458 ve 0.413 olduğu görülürken kestirim hatalarının korelasyonun 0.0 düzeyi ile diğer düzeyleri arasında sırasıyla %1, %4 ve %14 oranında azaldığı görülmektedir. Alt test uzunluğu artışı ile modelden elde edilen kestirim hatalarının azaldığı; fakat bu durumun alt testler arası korelasyon düzeyine göre değişkenlik göstermediği görülmektedir. Örneğin, alt testler arası korelasyon koşulunun herhangi bir düzeyinde alt test uzunluğu 20'den 30'a çıkarıldığında alt test yetenek parametresi kestirim hatasının yaklaşık olarak %13-14 arası oranda azaldığı gözlenirken alt test uzunluğunun 30'dan 40'a çıkarılması ile hataların yaklaşık olarak %10-11 arası oranda azaldığı gözlenmektedir.



Şekil 4.6. Üç Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri

Şekil 4.6'da üç boyutlu veri setleri için alt testler arası korelasyon ve alt test uzunluğu koşulları altında İki-faktör Model, Üst Düzey Sıralı Model ve Hiyerarşik ÇBMTK Model'in alt test yetenek parametresi kestirim hatalarına ait grafiklere yer verilmiştir.

Şekil 4.6'ya göre üç kestirim modelinin tüm koşullar altındaki performansları karşılaştırıldığında, en düşük hata düzeyine sahip kestirimlerin İki Faktör modelinden elde edildiği görülürken en düşük hata düzeyine sahip kestirimlerin Üst Düzey Sıralı modelden edildiği görülmektedir. Alt testler arası korelasyon düzeyi attıkça Üst Düzey Sıralı Model ve Hiyerarşik ÇBMTK Model için alt test yetenek kestirim hatalarının azaldığı gözlenirken İki-faktör Model için hataların arttığı gözlenmektedir. Alt testler arası korelasyon koşulunun ilk iki düzeyinde Üst Düzey Sıralı Model ve Hiyerarşik ÇBMTK Model'in alt test yetenek kestirim hatalarının benzer olduğu görülürken korelasyon düzeyi arttıkça Üst Düzey Sıralı Model'in daha iyi performans gösterdiği görülmektedir. İki-faktör Model'e ait kestirim hataları ile alt test uzunluğu arasında azalan fakat alt testler arası korelasyon arasında artan doğrusal yönde bir ilişki olduğu gözlenirken Üst Düzey Sıralı Model'e ait kestirim hataları ile alt test uzunluğu ve alt testler arası korelasyon koşulları arasında azalan doğrusal yönde bir ilişki olduğu gözlenmektedir. Hiyerarşik ÇBMTK Model'e ait kestirim hataları ile korelasyon koşulu arasında bir ilişki olmadığı ama alt test uzunluğu ile kestirim hataları arasında azalan doğrusal yönde bir ilişki olduğu görülmektedir.

Tablo 4.5, Tablo 4.6, Şekil 4.5 ve Şekil 4.6'da verilen iki ve üç boyutlu veri setlerine ait değerler birlikte değerlendirildiğinde boyut sayısındaki artışın üç modelin alt test puanı için yetenek parametresi kestirim hataları üzerindeki etkisinin değişkenlik gösterdiği görülmektedir. İki-faktör modeli için boyut sayısı ikiden üçe çıkarıldığında yetenek parametresi kestirim hatası alt testler arası korelasyon koşulunun düşük düzeylerinde azalmaya neden olurken korelasyonun yüksek düzeylerinde artmaya neden olduğu görülmektedir. Bu model için boyut sayısı arttırıldığında alt testler arası korelasyonun 0.0 düzeyinde 20, 30 ve 40 maddelik alt testlerden elde edilen yetenek parametre kestirim hatalarının sırasıyla yaklaşık olarak %5, %6 ve %7 oranlarında azalmasına neden olurken korelasyonun 0.3 düzeyinde yaklaşık olarak %1 oranda azalmasına neden olmaktadır. Yine İki faktör model için boyut sayısı arttırıldığında alt testler arası korelasyonun 0.5 düzeyinde 20, 30 ve 40 maddelik alt testlerden elde edilen yetenek parametre kestirim hatalarının sırasıyla yaklaşık olarak %4, %5 ve %5 oranlarında artmasına neden olurken korelasyonun 0.8 düzeyinde sırasıyla yaklaşık olarak %1 oranda artmasına neden olmaktadır. Hiyerarşik ÇBMTK model için boyut sayısının ikiden üçe çıkarılmasının alt test

yetenek kestirim hatalarında bir etkisi olmadığı görülmektedir. Üst Düzey Sıralı Model için boyut sayısı ikiden üçe çıkarıldığında alt testler arası korelasyon koşulunun düşük düzeylerinde minimal düzeyde olmakla birlikte alt test yetenek parametresi kestirim hatasının korelasyonun artışı ile azalmaya neden olduğu görülmektedir. Bu model için boyut sayısı artırıldığında alt testler arası korelasyonun 0.0 ve 0.3 düzeylerinde 20, 30 ve 40 maddelik alt testlerden elde edilen yetenek parametre kestirim hatalarının yaklaşık olarak %0-%2 arası oranlarında azalmasına neden olurken korelasyonun 0.3 düzeyinde yaklaşık olarak %1-%2 arası oranlarda ve 0.8 düzeyinde yaklaşık olarak %4-%5 arası oranda azalmasına neden olmaktadır. Her bir madde tepki kuramı modeli için iki ve üç boyutlu verilerden elde edilen alt test yetenek parametresi kestirimlerine ait RMSE değerlerine ilişkin karşılaştırılmalı grafikler EK- 8’de verilmiştir.

4.2.3. İki Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri

Tablo 4.7’de araştırmada ele alınan koşullara dayalı olarak iki boyutlu veri setlerinden alt test puanı için üç hiyerarşik madde tepki kuramı modeli kullanılarak kestirilen yetenek parametrelerine ilişkin güvenirlik değerleri verilmiştir. Tablo 4.7’de her bir madde tepki kuramı modeline ilişkin verilen değerler iki alt testten kestirilen yetenek parametresine ait güvenirlik değerlerinin ortalamasını göstermektedir. Her bir alt testten elde edilen güvenirlik değerleri EK-9.1’de verilmiştir.

Tablo 4.7.: İki Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri

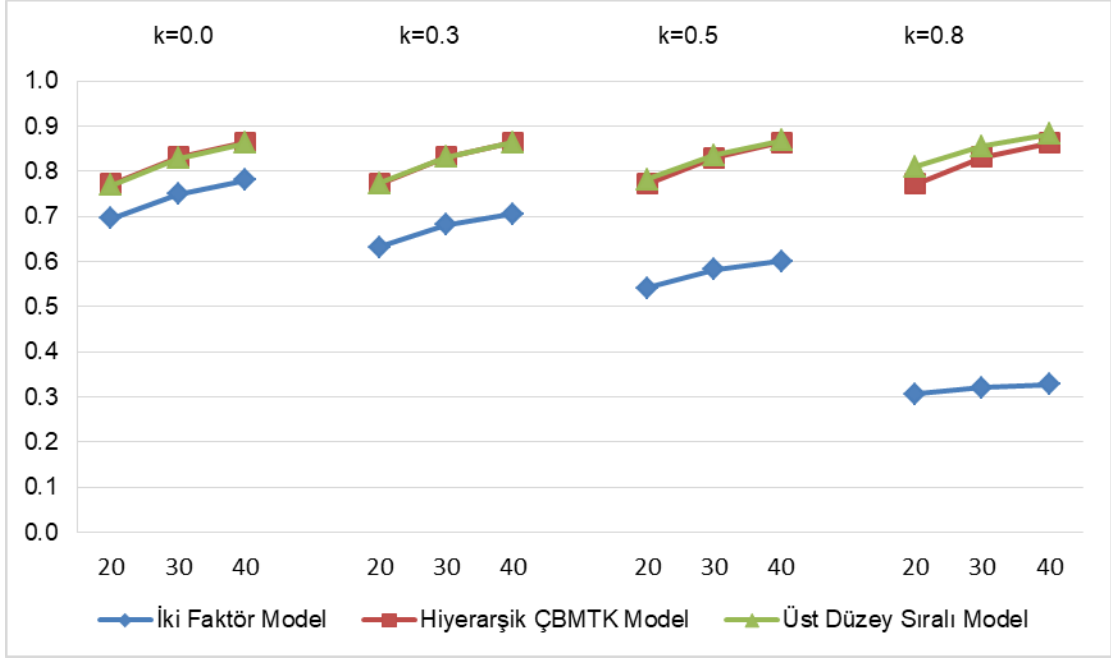
<i>Alt test sayısı</i>	<i>Korelasyon</i>	<i>Alt test uzunluğu</i>	<i>İki-faktör Model</i>	<i>ÇBMTK Model</i>	<i>Üst Düzey Sıralı Model</i>
2	0.0	20	0.696	0.772	0.769
		30	0.750	0.831	0.829
		40	0.780	0.864	0.863
	0.3	20	0.632	0.772	0.773
		30	0.682	0.831	0.832
		40	0.705	0.864	0.864
	0.5	20	0.541	0.772	0.782
		30	0.583	0.830	0.836
		40	0.601	0.864	0.868
	0.8	20	0.307	0.772	0.810
		30	0.320	0.831	0.856
		40	0.328	0.863	0.882

Tablo 4.7’de verilen güvenilirlik deęerleri incelendięinde, öncelikle iki boyutlu Hiyerarşik ÇB3PMTK’ya göre üretilen verilerin tüm koşullar altında en güvenilir parametre kestirimlerinin Üst Düzey Sıralı modele dayalı olarak elde edildięi gözlenmektedir. Alt test uzunluęu ve alt testler arası korelasyon koşullarının tüm düzeyleri bir arada deęerlendirildięinde modele ait güvenilirliklerin 0.769-0.882 arasında deęiştii görölmektedir. Ayrıca tüm koşullar altında alt test puan kestirimlerinin en az 0.7 ile kabul edilebilir sınırlarda güvenilirlikle elde edildięi dikkat çekmektedir. Alt testler arası korelasyon arttırıldıęında korelasyonun ilk üç düzeyi arasında minimal olmakla birlikte alt test yetenek parametresi kestirimlerin güvenilirlięinin arttıęı ve en güvenilir kestirimlerin korelasyonun 0.8 düzeyinde elde edildięi gözlenmektedir. Örneęin, alt test uzunluęu 20 iken alt testler arası korelasyonun 0.0, 0.3, 0.5 ve 0.8 düzeyi için elde edilen güvenilirliklerin sırasıyla 0.769, 0.773, 0.782 ve 0.810 olduęu görölmektedir. Bu sonuca göre, alt testler arası korelasyonun 0.8 düzeyi için elde edilen yetenek parametresi kestirimlerinin, 0.0 ve 0.3 düzeylerine göre yaklaşık olarak %5 oranında, 0.5 düzeyine göre ise yaklaşık olarak %4 oranında daha güvenilir elde edildięi söylenebilir. Benzer şekilde, korelasyon koşulunun tüm düzeylerinde alt test uzunluęu arttıęı modelden elde edilen kestirimlerin daha güvenilir olduęu görölmektedir. Örneęin; alt testler arası korelasyon koşulunun tüm düzeylerinde alt test uzunluęu 20’den 30’a çıkarıldıęında yetenek parametresinin yaklaşık olarak %6-8 arası oranlarda daha fazla güvenilirlikle kestirildięi gözlenirken alt test uzunluęu 30’dan 40’a çıkarıldıęında korelasyon koşulunun tüm düzeylerinde sırasıyla yaklaşık olarak yaklaşık olarak %3-%7 arası oranlarda daha fazla güvenilirlikle kestirildięi gözlenmektedir.

Alt test uzunluęu ve alt testler arası korelasyon koşullarının tüm düzeyleri bir arada deęerlendirildięinde Hiyerarşik ÇBMTK modele ait güvenilirliklerin 0.722-0.863 arasında deęiştii görölmektedir. Alt testler arası korelasyon koşulunun en düşük düzeyinde dahi alt test yetenek parametresi kestirimlerinin en az 0.7 ile yeterli düzeyde güvenilirlikle elde edildięi gözlenmektedir. Alt test uzunluęu sabit tutulduęunda alt testler arası korelasyonun tüm düzeyleri için alt test yetenek kestirim güvenilirlikleri arasında fark olmadıęı fakat korelasyon düzeyi sabit tutulduęunda alt test uzunluęu arttıęı kestirim güvenilirlięinin arttıęı gözlenmektedir. Örneęin, alt test uzunluęu 20 iken alt testler arası korelasyonun her bir düzeyi için elde edilen güvenilirliklerin sırasıyla 0.772, olduęu görülürken, alt testler arası

korelasyonun herhangi bir düzeyi için 20, 30 ve 40 maddelik alt testlerden elde edilen kestirim güvenilirliğinin sırasıyla 0.772, 0.831, 0.864 olduğu görülmektedir. Bu sonuca göre, alt testler arası korelasyon koşulundan bağımsız olarak alt test uzunluğu 20'den 30'a çıkarıldığında alt test yetenek parametresinin kestirim güvenilirliğinin yaklaşık olarak %8 oranında ve alt test uzunluğu 30'dan 40'a çıkarıldığında ise yaklaşık %4 oranında arttığı gözlenmektedir.

İki-faktör modeline dayalı olarak elde edilen güvenilirlik değerleri incelendiğinde, alt test uzunluğu ve alt testler arası korelasyon koşullarının tüm düzeyleri bir arada değerlendirildiğinde modele ait güvenilirliklerin 0.307-0.780 arasında değiştiği görülmektedir. Alt testler arası korelasyon arttığında alt test uzunluğunun tüm düzeyleri için bu yöntemden elde edilen alt test yetenek parametresi kestirim güvenilirliğinin önemli düzeyde azaldığı dikkat çekmektedir. Ayrıca bu yöntem için kabul edilebilir düzeyde güvenilir kestirimlerin korelasyonun 0.0 düzeyinde alt test uzunluğu koşulunun tüm düzeylerinde ve korelasyonun 0.3 düzeyinde alt test uzunluğu koşulunun 30 ve 40 madde düzeyinde elde edildiği görülmektedir. Örneğin, alt test yetenek parametresinin alt test uzunluğu 20 iken alt testler arası korelasyon düzeyleri için sırasıyla 0.696, 0.632, 0.541 ve 0.601 düzeyinde güvenilirlik ile kestirildiği gözlenirken alt test uzunluğu 40 iken sırasıyla 0.780, 0.705, 0.601 ve 0.328 düzeyinde güvenilirlik ile kestirildiği gözlenmektedir. Bu sonuca göre alt test uzunluğu koşulunun tüm düzeyleri için alt testler arası korelasyon arttıkça alt test yetenek parametresi kestirim güvenilirliğinin korelasyonun 0.0 düzeyinde elde edilen korelasyonun diğer düzeylerine (0.3, 0.5 ve 0.8) göre sırasıyla yaklaşık olarak %9, %22 ve %52 oranında azaldığı söylenebilir. Alt test uzunluğu açısından güvenilirlikler incelendiğinde, alt test uzunluğu artışı ile güvenilirliğin arttığı fakat alt testler arası korelasyon düzeyi artışına göre değişmediği görülmektedir. Örneğin, alt test uzunluğu 20 den 30'a çıkarıldığında İki Faktör modeline dayalı elde edilen alt test yetenek parametresi kestiriminin alt testler arası korelasyonun ilk üç düzeyinde yaklaşık olarak %8 oranında, korelasyonun 0.8 düzeyinde yaklaşık olarak %4 oranında fazla güvenilirlikle kestirildiği; alt test uzunluğu 30'dan 40'a çıkarıldığında ise korelasyonun tüm düzeylerinde yaklaşık olarak %3 oranında daha fazla güvenilirlikle kestirildiği gözlenmektedir.



Şekil 4.7. İki Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri

Şekil 4.7’de iki boyutlu veri setleri için alt testler arası korelasyon ve alt test uzunluğu koşulları altında İki-faktör Model, Üst Düzey Sıralı Model ve Hiyerarşik ÇBMTK Model’in alt test yetenek parametresi kestirimine ilişkin güvenilirlik değerlerine ait grafiklere yer verilmiştir. Şekil 4.7’e göre üç kestirim modelinin tüm koşullar altındaki performansları karşılaştırıldığında kabul edilebilir en güvenilir kestirimlerin Üst Düzey Sıralı Modelden elde edildiği görülmektedir. Alt testler arası korelasyon düzeyi attıkça Üst Düzey Sıralı Model ve Hiyerarşik ÇBMTK Model’in yetenek parametresi kestirim güvenilirliğinin arttığı ve kestirim güvenilirliğinin korelasyonun tüm düzeylerinde kabul edilebilir düzeylerde olduğu gözlenmektedir. Ayrıca bu iki modelin alt test yetenek kestirim güvenilirliğinin korelasyonun ilk üç düzeyinde aynı/benzer olduğu ve korelasyonun 0.8 düzeyinde Üst Düzey Sıralı Modelin daha güvenilir sonuçlar verdiği söylenebilir. İki faktör Modelin alt test yetenek kestirim güvenilirliğinin korelasyon düzeyi arttıkça önemli düzeyde azaldığı ve bu yöntemin kabul edilebilir düzeyde güvenilir kestirimlerinin düşük korelasyon düzeyinde elde edildiği gözlenmektedir. Fakat düşük korelasyon düzeyinde dahi İki faktör Model parametre kestirim güvenilirliğinin Üst Düzey Sıralı Model ve Hiyerarşik ÇBMTK Model kestirimlerine göre daha düşük olduğu dikkat çekmektedir. Alt test uzunluğu ve alt testler arası korelasyon düzeyindeki artışın her üç yöntem için de parametre

kestirim güvenilirliğini arttırdığı görülmektedir. Alt test yetenek parametresi kestirim güvenilirliği ile alt test uzunluğu ve alt testler korelasyon arasında Üst Düzey Sıralı Model ve Hiyerarşik ÇBMTK Model için artan doğrusal yönde bir ilişki olduğu gözlenirken İki faktör Model için alt test uzunluğu ile artan fakat korelasyon ile azalan doğrusal yönde bir ilişki olduğu gözlenmektedir.

4.2.4. Üç Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri

Tablo 4.8’de araştırmada ele alınan koşullara dayalı olarak üç boyutlu veri setlerinden alt test puanı için üç hiyerarşik madde tepki kuramı modeli kullanılarak kestirilen yetenek parametrelerine ilişkin güvenilirlik değerleri verilmiştir. Tablo 4.8’de her bir madde tepki kuramı modeline ilişkin verilen değerler üç alt testten kestirilen yetenek parametresine ait güvenilirlik değerlerinin ortalamasını göstermektedir. Her bir alt testten elde edilen güvenilirlik değerleri E-9.2’de verilmiştir.

Tablo 4.8.: Üç Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri

<i>Alt test sayısı</i>	<i>Korelasyon</i>	<i>Alt test uzunluğu</i>	<i>İki-faktör Model</i>	<i>Hiyerarşik ÇBMTK Model</i>	<i>Üst Düzey Sıralı Model</i>
3	0.0	20	0.721	0.774	0.771
		30	0.777	0.833	0.832
		40	0.807	0.867	0.866
	0.3	20	0.631	0.774	0.778
		30	0.680	0.833	0.836
		40	0.704	0.867	0.869
	0.5	20	0.492	0.774	0.791
		30	0.530	0.833	0.843
		40	0.552	0.867	0.874
	0.8	20	0.215	0.773	0.830
		30	0.229	0.833	0.870
		40	0.239	0.867	0.893

Tablo 4.8’de verilen güvenilirlik değerleri incelendiğinde, öncelikle iki boyutlu Hiyerarşik ÇB3PK’ya göre üretilen verilerin hemen hemen tüm koşullar altında en güvenilir parametre kestirimlerinin Üst Düzey Sıralı modele dayalı olarak elde edildiği gözlenmektedir. Alt test uzunluğu ve alt testler arası korelasyon koşullarının tüm düzeyleri bir arada değerlendirildiğinde modele ait güvenilirliklerin 0.771-0.893 arasında değiştiği görülmektedir. Ayrıca tüm koşullar altında alt test puan kestirimlerinin en 0.7 ile kabul edilebilir sınırlarda güvenilirlikle elde edildiği dikkat

çekmektedir. Alt testler arası korelasyon arttırıldığında korelasyonun ilk üç düzeyi arasında minimal olmakla birlikte alt test yetenek parametresi kestirimlerin güvenilirliğinin arttığı ve en güvenilir kestirimlerin korelasyonun 0.8 düzeyinde elde edildiği gözlenmektedir. Örneğin, alt test uzunluğu 20 iken alt testler arası korelasyonun 0.0, 0.3, 0.5 ve 0.8 düzeyi için elde edilen güvenilirliklerin sırasıyla 0.771, 0.778, 0.791 ve 0.830 olduğu görülmektedir. Bu sonuca göre, alt testler arası korelasyonun 0.8 düzeyi için elde edilen yetenek parametresi kestirimlerinin, 0.0 ve 0.3 düzeylerine göre yaklaşık %6 oranında, 0.5 düzeyine göre ise yaklaşık %5 oranında daha güvenilir elde edildiği söylenebilir. Benzer şekilde, korelasyon koşulunun tüm düzeylerinde alt test uzunluğu arttıkça modelden elde edilen kestirimlerin daha güvenilir olduğu görülmektedir. Örneğin; alt testler arası korelasyon koşulunun tüm düzeylerinde alt test uzunluğu 20'den 30'a çıkarıldığında yetenek parametresinin yaklaşık olarak %5-8 arası oranlarda daha fazla güvenilirlikle kestirildiği gözlenirken alt test uzunluğu 30'dan 40'a çıkarıldığında korelasyon koşulunun tüm düzeylerinde sırasıyla yaklaşık olarak %3-%7 arası oranlarda daha fazla güvenilirlikle kestirildiği gözlenmektedir.

Alt test uzunluğu ve alt testler arası korelasyon koşullarının tüm düzeyleri bir arada değerlendirildiğinde Hiyerarşik ÇBMTK modele ait güvenilirliklerin 0.774-0.867 arasında değiştiği görülmektedir. Alt testler arası korelasyon koşulunun en düşük düzeyinde dahi alt test yetenek parametresi kestirimlerinin en az 0.7 ile iyi düzeyde güvenilirlikle elde edildiği gözlenmektedir. Alt test uzunluğu sabit tutulduğunda alt testler arası korelasyonun tüm düzeyleri için alt test yetenek kestirim güvenilirlikleri arasında fark olmadığı fakat korelasyon düzeyi sabit tutulduğunda alt test uzunluğu arttıkça kestirim güvenilirliğinin arttığı gözlenmektedir. Örneğin, alt test uzunluğu 20 iken alt testler arası korelasyonun her bir düzeyi için elde edilen güvenilirlikler sırasıyla 0.774, 0.774, 0.774 ve 0.773 olduğu görülürken, alt testler arası korelasyonun herhangi bir düzeyi için 20, 30 ve 40 maddelik alt testlerden elde edilen kestirim hatalarının sırasıyla 0.774, 0.833, 0.867 olduğu görülmektedir. Bu sonuca göre, alt testler arası korelasyon koşulundan bağımsız olarak alt test uzunluğu 20'den 30'a çıkarıldığında alt test yetenek parametresi kestirim güvenilirliğinin yaklaşık olarak %8 oranında ve alt test uzunluğu 30'dan 40'a çıkarıldığında ise %4 oranında azaldığı gözlenmektedir.

İki-faktör modeline dayalı olarak elde edilen güvenilirlik değerleri incelendiğinde, alt test uzunluğu ve alt testler arası korelasyon koşullarının tüm düzeyleri bir arada değerlendirildiğinde modele ait güvenilirliklerin 0.215-0.807 arasında değiştiği görülmektedir. Alt testler arası korelasyon arttığında alt test uzunluğunun tüm düzeyleri için bu yöntemden elde edilen alt test yetenek parametresi kestirim güvenilirliğinin önemli düzeyde azaldığı dikkat çekmektedir. Ayrıca bu yöntem için kabul edilebilir düzeyde güvenilir kestirimlerin korelasyonun 0.0 düzeyinde alt test uzunluğu koşulunun tüm düzeylerinde ve korelasyonun 0.3 düzeyinde alt test uzunluğu koşulunun 30 ve 40 madde düzeyinde elde edildiği görülmektedir. Örneğin, alt test yetenek parametresinin alt test uzunluğu 20 iken alt testler arası korelasyon düzeyleri için sırasıyla 0.721, 0.631, 0.492 ve 0.215 düzeyinde güvenilirlik ile kestirildiği gözlenirken alt test uzunluğu 40 iken sırasıyla 0.807, 0.704, 0.552 ve 0.239 düzeyinde güvenilirlik ile kestirildiği gözlenmektedir. Bu sonuca göre alt test uzunluğu koşulunun tüm düzeyleri için alt testler arası korelasyon arttıkça alt test yetenek parametresi kestirim güvenilirliğinin korelasyonun 0.0 düzeyinde elde edilen korelasyonun diğer düzeylerine (0.3, 0.5 ve 0.8) göre sırasıyla yaklaşık olarak %12, %32 ve %70 oranında azaldığı söylenebilir. Alt test uzunluğu açısından güvenilirlikler incelendiğinde, alt test uzunluğu artışı ile güvenilirliğin arttığı fakat alt testler arası korelasyon düzeyi artışına göre değişmediği görülmektedir. Örneğin, alt test uzunluğu 20 den 30'a çıkarıldığında İki Faktör modeline dayalı elde edilen alt test yetenek parametresi kestiriminin alt testler arası korelasyonun tüm düzeyinde yaklaşık %8 oranında daha fazla güvenilirlikle kestirildiği; alt test uzunluğu 30'dan 40'a çıkarıldığında ise korelasyonun tüm düzeylerinde yaklaşık %4 oranında daha fazla güvenilirlikle kestirildiği gözlenmektedir.



Şekil 4.8. Üç Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri

Şekil 4.8’de iki boyutlu veri setleri için alt testler arası korelasyon ve alt test uzunluğu koşulları altında İki-faktör Model, Üst Düzey Sıralı Model ve Hiyerarşik ÇBMTK Model’in alt test yetenek parametresi kestirimine ilişkin güvenirlilik değerlerine ait grafiklere yer verilmiştir. Şekil 4.8’e göre üç kestirim modelinin tüm koşullar altındaki performansları karşılaştırıldığında kabul edilebilir en güvenilir kestirimlerin Üst Düzey Sıralı Modelden elde edildiği görülmektedir. Alt testler arası korelasyon düzeyi attıkça Üst Düzey Sıralı Model ve Hiyerarşik ÇBMTK Model’in yetenek parametresi kestirim güvenirliliğinin arttığı ve kestirim güvenirliliğinin korelasyonun tüm düzeylerinde kabul edilebilir düzeylerde olduğu gözlenmektedir. Ayrıca bu iki modelin alt test yetenek kestirim güvenirliliğinin korelasyonun ilk üç düzeyinde aynı/benzer olduğu ve korelasyonun 0.8 düzeyinde Üst Düzey Sıralı Modelin daha güvenilir sonuçlar verdiği söylenebilir. İki faktör Modelin alt test yetenek kestirim güvenirliliğinin korelasyon düzeyi arttıkça önemli düzeyde azaldığı ve bu yöntemin kabul edilebilir düzeyde güvenilir kestirimlerinin düşük korelasyon düzeyinde elde edildiği gözlenmektedir. Fakat düşük korelasyon düzeyinde dahi İki faktör Model parametre kestirim güvenirliliğinin Üst Düzey Sıralı Model ve Hiyerarşik ÇBMTK Model kestirimlerine göre daha düşük olduğu dikkat çekmektedir. Alt test uzunluğu ve alt testler arası korelasyon düzeyindeki artışın her üç yöntem için de parametre

kestirim güvenilirliğini arttırdığı görülmektedir. Alt test yetenek parametresi kestirim güvenilirliği ile alt test uzunluğu ve alt testler korelasyon arasında Üst Düzey Sıralı Model ve Hiyerarşik ÇBMTK Model için artan doğrusal yönde bir ilişki olduğu gözlenirken İki faktör Model için güvenilirlik ile alt test uzunluğu arasında artan fakat korelasyon ile azalan doğrusal yönde bir ilişki olduğu gözlenmektedir.

Tablo 4.7, Tablo 4.8, Şekil 4.7 ve Şekil 4.8'de verilen iki ve üç boyutlu veri setlerine ait değerler birlikte değerlendirildiğinde boyut sayısındaki artışın üç modelin alt test puanı için yetenek parametresi kestirim güvenilirliğine etkisinin değişken olduğu gözlenmektedir. İki faktör modeli için boyut sayısını ikiden üçe çıkarmanın alt test kestirim güvenilirliğini korelasyon düzeyinin 0.0 olduğu durumda yaklaşık olarak %4 oranında arttırdığı, korelasyonun 0.3 düzeyinde etkisinin olmadığı, korelasyonun 0.5 düzeyinde yaklaşık olarak %9 oranında azalttığı ve korelasyonun 0.8 düzeyinde yaklaşık olarak %28 oranında azalttığı görülmektedir. Hiyerarşik ÇBMTK model için boyut sayısının ikiden üçe çıkarılmasının alt test yetenek kestirim hatalarında bir etkisi olmadığı görülmektedir. Üst Düzey Sıralı Model için boyut sayısını ikiden üçe çıkarmanın alt test kestirim güvenilirliğine korelasyon düzeyinin 0.0 düzeyinde etkisinin olmadığı, korelasyonun 0.3 ve 0.5 düzeyinde güvenilirliği yaklaşık olarak %1 oranında arttırdığı, korelasyonun 0.8 düzeyinde güvenilirliği yaklaşık olarak %2 oranında artırdığı görülmektedir. Her bir madde tepki kuramı modeli için iki ve üç boyutlu verilere ait güvenilirlik değerlerine ilişkin karşılaştırılmalı grafikler EK-10'da verilmiştir.

4.2.5. İkinci Alt Probleme İlişkin Tartışma

Hem iki hem de üç boyutlu veri setlerinde alt testler arasındaki korelasyon düzeyinin Hiyerarşik ÇBMTK Model'den elde edilen alt test puan kestirim hataları ve güvenilirliği üzerinde bir etkisi olmadığı sonucuna ulaşılmıştır. Fakat Bulut (2013) ve Yao (2010) çalışmalarında alt testler arası korelasyon düzeyindeki artışın ÇBMTK Model kestirimlerinin güvenilirliği ve doğruluğunu arttırdığını bulmuşlardır. Alt test sayısındaki artışın bu çalışmada kestirim hataları ve güvenilirliği üzerinde bir etkisi gözlenmezken Bulut (2013) alt test sayısının güvenilirlik üzerinde minimal düzeyde etkisi olduğunu belirtmiştir. Bu çalışma ile diğer çalışmalar arasındaki farkın veri üretme koşulları arasındaki farklılıktan kaynaklanabileceği düşünülmektedir.

Hem iki hem de üç boyutlu veri setlerinde alt testler arasındaki korelasyon düzeyindeki artış ile İki Faktör Model'den elde edilen alt test puan kestirim hatalarının arttığı ve güvenilirliğin azaldığı sonucu Yao (2010) ve Chang'nın (2015) araştırma sonuçlarıyla uyumludur. Bu modelde alt testler arasındaki ilişkilerin dik olduğu varsayımı nedeniyle model yüksek ilişki gösteren alt testlerde yüksek hatalı ve düşük güvenilirlikli kestirimler yapmaktadır. Dolayısıyla kabul edilebilir düzeyde güvenilir kestirimlerin ancak 0.0 korelasyon düzeyinde ve 0.3 korelasyon düzeyinin de uzun alt test düzeylerinde elde edilmektedir. Ayrıca alt test sayısındaki artışın bu model için optimal koşul olan düşük korelasyon düzeylerinde hataların azalması ve güvenilirliğin artması benzer nedenlerden kaynaklanmaktadır. Optimal koşullardan uzaklaştıkça alt test sayısını arttırmak daha düşük güvenilirlikli ve daha yüksek hatalı kestirimlere sebep olmaktadır.

Hem iki hem de üç boyutlu veri setlerinde alt testler arasındaki korelasyon düzeyindeki artış ile Üst Düzey Sıralı Model'den elde edilen alt test puanı kestirim hatalarının azalması modelin doğası gereği genel ve alt boyutlar arasındaki ilişkileri kullanmasının doğal bir sonucudur. De la Torre, Song ve Hong (2011) Üst Düzey Sıralı Model'e dayalı ürettikleri veriler üzerinde de bu araştırmayla benzer sonuçlara ulaşmıştır. Aynı zamanda 0.8 korelasyon düzeyinde Üst Düzey Sıralı Model'in biraz daha doğru ve güvenilir kestirimler yapmasıyla birlikte bu model ile Hiyerarşik ÇBMTK model benzer performans göstermiştir. Bu durumun iki modelin de alt testler arasındaki korelasyonları kullanarak yetenek parametresi kestirmesinden kaynaklandığı düşünülmektedir. Bu araştırma sonuçlarından farklı olarak Yao (2010) bu iki modeli benzer ama minimal farkla ÇBMTK modelin daha iyi performans gösterdiğini bulurken de la Torre, Song ve Hong (2011) bu iki modelin performansını birbirine eşit bulmuştur. Bu araştırmanın alt test sayısındaki artış ile Üst Düzey Sıralı Model'in alt test puan kestirimlerinin hatası ve güvenilirliği üzerinde minimal düzeyde iyileşme sağladığı sonucu ile de la Torre, Song ve Hong'un (2011) araştırma sonuçları benzerdir. Daha uzun alt testlerde üç yöntem için de kestirim hatalarının azalması ve güvenilirliğin artması beklenen bir durumdur.

4.3. Üçüncü Alt Probleme İlişkin Bulgular ve Tartışma

Araştırmanın üçüncü alt problemi kapsamında "Üst Düzey Sıralı (Higher Order), İki Faktör (Bi-factor) ve hiyerarşik çok boyutlu madde tepki kuramı (ÇBMTK) modellerine göre **alt test** ve **toplam test puanları** için kestirilen yetenek

parametrelerinin RMSE ve güvenilirlik değerlerinin **kestirim modelleri, alt test sayısı, alt test uzunluğu ve alt testler arasındaki korelasyonların büyüklüğü** koşullarından nasıl etkilendiği” varyans analizi ile incelenmiştir. Alt test ve toplam test puanlarına ait RMSE değerleri için yapılan varyans analizi sonuçları ayrı ayrı tablolarda verilmiştir.

4.3.1. Toplam Test Puanlarına ait RMSE Değerleri İçin Varyans Analizi Sonuçları

Simülasyon çalışmasında tüm koşullar altında elde edilen toplam test puanlarına ait RMSE değerleri üzerinde model, alt test sayısı, alt test uzunluğu ve alt testler arasındaki korelasyonun etkisini incelemek için yapılan varyans analizi sonuçları Tablo 4.9’da verilmiştir.

Tablo 4.9: Toplam Test Puanlarına Ait RMSE Değerleri İçin Varyans Analizi Sonuçları

<i>Varyans kaynağı</i>	<i>Kareler toplamı</i>	<i>df</i>	<i>Kareler ortalaması</i>	<i>F</i>	<i>p</i>	<i>Kısmi η^2</i>
Alt test sayısı	2,964	1	2,964	2239,519	0,000	0,388
Alt test uzunluğu	1,482	2	0,741	559,900	0,000	0,241
Korelasyon	10,010	3	3,337	2521,279	0,000	0,682
Model	19,153	2	9,576	7236,319	0,000	0,804
Alt test sayısı * alt test uzunluğu	0,016	2	0,008	6,104	0,002	0,003
Alt test sayısı * korelasyon	0,127	3	0,042	31,884	0,000	0,026
Alt test sayısı * model	0,066	2	0,033	24,815	0,000	0,014
Alt test uzunluğu * korelasyon	0,167	6	0,028	21,091	0,000	0,035
Alt test uzunluğu * model	0,136	4	0,034	25,756	0,000	0,028
Korelasyon * model	8,567	6	1,428	1078,886	0,000	0,647
Alt test sayısı * Alt test uzunluğu * korelasyon	0,009	6	0,002	1,196	0,305	0,002
Alt test sayısı * Alt test uzunluğu * model	0,018	4	0,004	3,358	0,009	0,004
Alt test sayısı * korelasyon * model	0,687	6	0,114	86,494	0,000	0,128
Alt test uzunluğu * korelasyon * model	0,056	12	0,005	3,501	0,000	0,012
Alt test sayısı * Alt test uzunluğu * korelasyon * model	0,016	12	0,001	0,999	0,447	0,003

Tablo 4.9 incelendiğinde toplam test puanlarına ait RMSE değerleri üzerinde tüm ana ve tüm ikili ortak etkilerin anlamlı düzeyde etkisi olduğu gözlenmektedir. Ana etkiler açısından etki büyüklükleri incelendiğinde en fazla etkiye sahip değişkenlerin sırasıyla model (kısmi $\eta^2=0.804$) ve alt testler arası korelasyon (kısmi $\eta^2=0.682$) olduğu görülmektedir. İkili ortak etkiler incelendiğinde toplam testlere ait RMSE değerlerinin varyansını en fazla açıklayan etkileşimin korelasyon*model (kısmi

$\eta^2=0.647$) olduğu ve diğer ikili etkileşimlerin etki büyüklüklerinin (kısımlı $\eta^2 \leq 0.035$) çok düşük olduğu gözlenmektedir. Üçlü ortak etkiler içerisinde en fazla etkiye sahip etkileşimin alt test uzunluğu*korelasyon*model (kısımlı $\eta^2=0.128$) olduğu görülürken diğer üçlü etkileşimlerin etkisinin ya olmadığı (alt test sayısı*alt test uzunluğu*korelasyon etkileşimi, $p=0.305$) ya da çok düşük (kısımlı $\eta^2 \leq 0.012$) olduğu görülmektedir. Dörtlü ortak etkinin RMSE değerlerinin varyansına anlamlı bir katkısı olmadığı görülmektedir ($p=0.447$).

Çoklu karşılaştırma testi sonucunda ana etkiler için koşulların tüm düzeyleri arasında anlamlı farklılık olduğu bulunmuştur. Korelasyon, alt test uzunluğu ve alt test sayısı koşullarının düzeyleri arttıkça hatanın azaldığı gözlenmiştir. Modeller açısından en az hatalı kestirim yapan modeller sırasıyla Hiyerarşik ÇBMTK, Üst Düzey Sıralı ve İki Faktör modeldir. Korelasyon*model ikili etkileşimin modellere göre ikili karşılaştırma testi sonucunda yalnızca 0.8 korelasyon düzeyi için İki Faktör Model ile Üst Düzey Sıralı Model arasında anlamlı farklılık gözlenmezken diğer tüm ikili karşılaştırmalar anlamlı bulunmuştur. Yine, korelasyon*model etkileşimin korelasyona göre ikili karşılaştırma testi sonucunda ise İki Faktör ve Üst Düzey Sıralı modeller için korelasyonun tüm ikili etkileşimleri arasında anlamlı farklılık gözlenirken Hiyerarşik ÇBMTK model için yalnızca 0.8 korelasyon ile diğer korelasyon düzeyleri ve 0.5 korelasyon ile 0.3 korelasyon düzeyi arasında anlamlı farklılık gözlenmiştir.

4.3.2. Toplam Test Puanlarına ait Güvenirlik Değerleri İçin Varyans Analizi Sonuçları

Simülasyon çalışmasında tüm koşullar altında elde edilen toplam test puanlarına ait güvenirlik değerleri üzerinde model, alt test sayısı, alt test uzunluğu ve alt testler arasındaki korelasyonun etkisini incelemek için yapılan varyans analizi sonuçları Tablo 4.10'da verilmiştir.

Tablo 4.10 incelendiğinde toplam test puanlarına ait güvenirlik değerleri üzerinde tüm ana ve tüm ikili ortak etkilerin anlamlı düzeyde etkisi olduğu gözlenmektedir. Ana etkiler açısından etki büyüklükleri incelendiğinde en fazla etkiye sahip değişkenlerin sırasıyla alt testler arası korelasyon (kısımlı $\eta^2=0.894$) ve model (kısımlı $\eta^2=0.700$) olduğu görülmektedir. İkili ortak etkiler incelendiğinde toplam testlere ait güvenirlik değerlerinin varyansını en fazla açıklayan etkileşimin korelasyon*model (kısımlı $\eta^2=0.706$) olduğu ve diğer ikili etkileşimlerin etki büyüklüklerinin (kısımlı $\eta^2 \leq 0.070$)

çok düşük olduğu gözlenmektedir. Üçlü ortak etkiler içerisinde yalnızca alt test sayısı*korelasyon*model ve alt test uzunluğu*korelasyon*model ortak etkileşimlerinin anlamlı fakat çok düşük düzeyde (kısmi $\eta^2 \leq 0.044$) etkiye sahip olduğu görülmektedir. Dörtlü ortak etkinin güvenilirlik değerlerinin varyansına anlamlı bir katkısı olmadığı görülmektedir ($p=0.964$).

Tablo 4.10: Toplam Test Puanlarına Ait Güvenirlik Değerleri İçin Varyans Analizi Sonuçları

<i>Varyans kaynağı</i>	<i>Kareler toplamı</i>	<i>df</i>	<i>Kareler ortalaması</i>	<i>F</i>	<i>p</i>	<i>Kısmi η^2</i>
Alt test sayısı	1,509	1	1,509	497,963	0,000	0,124
Alt test uzunluğu	2,721	2	1,361	448,974	0,000	0,203
Korelasyon	90,152	3	30,051	9915,606	0,000	0,894
Model	24,928	2	12,464	4112,629	0,000	0,700
Alt test sayısı * alt test uzunluğu	0,020	2	0,010	3,371	0,034	0,002
Alt test sayısı * korelasyon	0,806	3	0,269	88,663	0,000	0,070
Alt test sayısı * model	0,154	2	0,077	25,379	0,000	0,014
Alt test uzunluğu * korelasyon	0,038	6	0,006	2,117	0,048	0,004
Alt test uzunluğu * model	0,162	4	0,040	13,349	0,000	0,015
Korelasyon * model	25,631	6	4,272	1409,532	0,000	0,706
Alt test sayısı * Alt test uzunluğu * korelasyon	0,009	6	0,002	0,499	0,810	0,001
Alt test sayısı * Alt test uzunluğu * model	0,008	4	0,002	0,637	0,636	0,001
Alt test sayısı * korelasyon * model	0,497	6	0,083	27,322	0,000	0,044
Alt test uzunluğu * korelasyon * model	0,140	12	0,012	3,842	0,000	0,013
Alt test sayısı * Alt test uzunluğu * korelasyon * model	0,015	12	0,001	0,400	0,964	0,001

Çoklu karşılaştırma testi sonucunda ana etkiler için koşulların tüm düzeyleri arasında anlamlı farklılık olduğu bulunmuştur. Korelasyon, alt test uzunluğu ve alt test sayısı koşullarının düzeyleri arttıkça güvenilirliğin arttığı gözlenmiştir. Modeller açısından en güvenilir kestirim yapan modeller sırasıyla Hiyerarşik ÇBMTK, Üst Düzey Sıralı ve İki Faktör modeldir. Korelasyon*model ikili etkileşimin modellere göre ikili karşılaştırma testi sonucunda korelasyonun 0.0 ve 0.8 düzeyi için İki Faktör Model ile Üst Düzey Sıralı Model arasında anlamlı farklılık gözlenmezken diğer tüm ikili karşılaştırmalar anlamlı bulunmuştur. Yine, korelasyon*model etkileşimin korelasyona göre ikili karşılaştırma testi sonucunda ise üç model için de korelasyonun tüm ikili etkileşimleri arasında anlamlı farklılık gözlenmiştir.

4.3.3. Alt Test Puanlarına ait RMSE Değerleri İçin Varyans Analizi Sonuçları

Simülasyon çalışmasında tüm koşullar altında elde edilen alt test puanlarına ait RMSE değerleri üzerinde model, alt test sayısı, alt test uzunluğu ve alt testler arasındaki korelasyonun etkisini incelemek için yapılan varyans analizi sonuçları Tablo 4.11'de verilmiştir.

Tablo 4.11: Alt Test Puanlarına Ait RMSE Değerleri İçin Varyans Analizi Sonuçları

<i>Varyans kaynağı</i>	<i>Kareler toplamı</i>	<i>df</i>	<i>Kareler ortalaması</i>	<i>F</i>	<i>p</i>	<i>Kısmi η^2</i>
Alt test sayısı	,002	1	,002	104,263	,000	,029
Alt test uzunluğu	4,617	2	2,309	114331,070	,000	,985
Korelasyon	5,680	3	1,893	93761,951	,000	,988
Model	45,951	2	22,976	1137853,566	,000	,998
Alt test sayısı * alt test uzunluğu	,000	2	,000	11,190	,000	,006
Alt test sayısı * korelasyon	,101	3	,034	1662,740	,000	,586
Alt test sayısı * model	,091	2	,045	2245,775	,000	,560
Alt test uzunluğu * korelasyon	,092	6	,015	760,861	,000	,564
Alt test uzunluğu * model	,425	4	,106	5257,361	,000	,856
Korelasyon * model	16,664	6	2,777	137549,079	,000	,996
Alt test sayısı * Alt test uzunluğu * korelasyon	,001	6	,000	9,986	,000	,017
Alt test sayısı * Alt test uzunluğu * model	,000	4	,000	5,141	,000	,006
Alt test sayısı * korelasyon * model	,308	6	,051	2538,759	,000	,812
Alt test uzunluğu * korelasyon * model	,088	12	,007	364,595	,000	,554
Alt test sayısı * Alt test uzunluğu * korelasyon * model	,001	12	,000	4,206	,000	,014

Tablo 4.11 incelendiğinde alt test puanlarına ait RMSE değerleri üzerinde tüm ana ve tüm ikili ortak etkilerin anlamlı düzeyde etkisi olduğu gözlenmektedir. Ana etkiler açısından etki büyüklükleri incelendiğinde alt test puanlarına ait RMSE değerleri üzerinde alt test sayısı değişkeni (kısmi $\eta^2=0.029$) dışındaki diğer değişkenlerin yüksek düzeyde (kısmi $\eta^2 \geq 0.985$) etkiye sahip olduğu görülmektedir. İkili ortak etkiler incelendiğinde alt testlere ait RMSE değerlerinin varyansını en fazla açıklayan etkileşimin sırasıyla alt test uzunluğu* model (kısmi $\eta^2=0.996$) ve korelasyon*model (kısmi $\eta^2=0.856$) olduğu; en düşük etkiye sahip etkileşimin ise alt test sayısı*alt test uzunluğu (kısmi $\eta^2=0.006$) olduğu gözlenmektedir. Üçlü ortak etkiler içerisinde en fazla etkiye sahip etkileşimlerin sırasıyla alt test sayısı*korelasyon*model (kısmi $\eta^2=0.812$) ve alt test uzunluğu*korelasyon*model (kısmi $\eta^2=0.554$) olduğu görülürken diğer üçlü etkileşimlerin etkisinin çok düşük

(kısmi $\eta^2 \leq 0.017$) olduğu görülmektedir. Dörtlü etkilerin RMSE değerlerinin varyansına katkısının çok düşük (kısmi $\eta^2=0.014$) olduğu görülmektedir.

Çoklu karşılaştırma testi sonucunda ana etkiler için koşulların tüm düzeyleri arasında anlamlı farklılık olduğu bulunmuştur. Korelasyon arttıkça hatanın arttığı fakat alt test uzunluğu arttıkça hatanın azaldığı gözlenmiştir. İki ve üç boyutlu alt testlerden elde edilen alt test puanı kestirim hataları arasında bir fark olmadığı görülmüştür. Her bir boyut için elde edilen RMSE ortalaması sırasıyla 0,490 ve 0,491'dir. Varyans analizinde gözlenen anlamlı etkinin örneklem büyüklüğünden kaynaklandığı düşünülmektedir. Modeller açısından en az hatalı kestirim yapan modeller sırasıyla Üst Düzey Sıralı, Hiyerarşik ÇBMTK ve İki Faktör modeldir. Alt test uzunluğu*model, alt test sayısı*model, alt test sayısı*korelasyon ve alt test uzunluğu*korelasyon ikili etkileşimlerinin çoklu karşılaştırma testi sonuçlarında tüm ikili karşılaştırmaları anlamlı bulunmuştur. Korelasyon*model ikili etkileşimin modellere göre ikili karşılaştırma testi sonucunda korelasyonun tüm koşulları için üç modelin tüm ikili karşılaştırmaları arasında anlamlı farklılık bulunurken korelasyona göre ikili karşılaştırma testi sonucunda ise İki Faktör ve Üst Düzey Sıralı modeller için korelasyonun tüm ikili etkileşimleri arasında anlamlı farklılık gözlenmiş fakat Hiyerarşik ÇBMTK model için korelasyonun tüm düzeyleri arasında anlamlı farklılık gözlenmemiştir. Alt test uzunluğu*korelasyon*model üçlü etkileşimin modellere göre ikili karşılaştırma testi sonucunda korelasyon ve alt test uzunluğunun tüm koşulları için üç modelin tüm ikili karşılaştırmaları arasında anlamlı farklılık bulunmuştur. Yine bu üçlü etkileşimin korelasyona göre ikili karşılaştırma testi sonucunda ise İki Faktör ve Üst Düzey Sıralı modeller için korelasyonun tüm ikili etkileşimleri arasında anlamlı farklılık gözlenmiş fakat Hiyerarşik ÇBMTK model için korelasyonun tüm düzeyleri arasında anlamlı farklılık gözlenmemiştir. Alt test sayısı*korelasyon*model üçlü etkileşimin modellere göre ikili karşılaştırma testi sonucunda 2 boyut ve 0.3 korelasyon düzeyi için Hiyerarşik ÇBMTK model ile Üst Düzey Sıralı model arasında anlamlı farklılık gözlenmezken diğer tüm ikili karşılaştırmalar arasında anlamlı farklılık bulunmuştur. Yine bu üçlü etkileşimin alt test sayısına göre tüm ikili etkileşimleri arasında anlamlı farklılıklar gözlenmiştir. Ayrıca korelasyona göre ikili karşılaştırma testi sonucunda İki Faktör ve Üst Düzey Sıralı modeller için korelasyonun tüm ikili etkileşimleri arasında anlamlı farklılık gözlenmiş fakat

Hiyerarşik ÇBMTK model için korelasyonun tüm düzeyleri arasında anlamlı farklılık gözlenmemiştir.

4.3.4. Alt Test Puanlarına ait Güvenirlik Değerleri İçin Varyans Analizi Sonuçları

Simülasyon çalışmasında tüm koşullar altında elde edilen alt test puanlarına ait güvenirlik değerleri üzerinde model, alt test sayısı, alt test uzunluğu ve alt testler arasındaki korelasyonun etkisini incelemek için yapılan varyans analizi sonuçları Tablo 4.12'de verilmiştir.

Tablo 4.12: Alt Test Puanlarına Ait Güvenirlik Değerleri İçin Varyans Analizi Sonuçları

<i>Varyans kaynağı</i>	<i>Kareler toplamı</i>	<i>df</i>	<i>Kareler ortalaması</i>	<i>F</i>	<i>p</i>	<i>Kısmi η^2</i>
Alt test sayısı	0,037	1	0,037	1487,220	0,000	0,297
Alt test uzunluğu	3,835	2	1,918	76934,827	0,000	0,978
Korelasyon	11,430	3	3,810	152859,783	0,000	0,992
Model	56,748	2	28,374	1138361,878	0,000	0,998
Alt test sayısı * alt test uzunluğu	0,000	2	0,000	2,827	0,059	0,002
Alt test sayısı * korelasyon	0,163	3	0,054	2175,031	0,000	0,649
Alt test sayısı * model	0,235	2	0,118	4724,061	0,000	0,728
Alt test uzunluğu * korelasyon	0,080	6	0,013	532,873	0,000	0,475
Alt test uzunluğu * model	0,115	4	0,029	1152,489	0,000	0,566
Korelasyon * model	28,721	6	4,787	192049,700	0,000	0,997
Alt test sayısı * Alt test uzunluğu * korelasyon	0,000	6	0,000	1,993	0,063	0,003
Alt test sayısı * Alt test uzunluğu * model	0,001	4	0,000	5,515	0,000	0,006
Alt test sayısı * korelasyon * model	0,450	6	0,075	3008,391	0,000	0,837
Alt test uzunluğu * korelasyon * model	0,059	12	0,005	197,838	0,000	0,402
Alt test sayısı * Alt test uzunluğu * korelasyon * model	0,001	12	0,000	2,238	0,008	0,008

Tablo 4.12 incelendiğinde alt test puanlarına ait güvenirlik değerleri üzerinde tüm ana etkilerin anlamlı düzeyde etkisi olduğu gözlenmektedir. Ana etkiler açısından etki büyüklükleri incelendiğinde alt test puanlarına ait güvenirlik değerleri üzerinde alt test sayısı değişkeni (kısmi $\eta^2=0.297$) dışındaki diğer değişkenlerin yüksek düzeyde (kısmi $\eta^2 \geq 0.978$) etkiye sahip olduğu görülmektedir. İkili ortak etkiler incelendiğinde alt testlere ait güvenirlik değerlerinin varyansına alt test sayısı*alt test uzunluğu etkileşiminin katkı sağlamadığı görülürken ($p=0.059$) en fazla katkı sağlayan etkileşimin korelasyon*model (kısmi $\eta^2=0.997$) olduğu ve diğer ikili etkileşimlerin etki büyüklüklerinin (kısmi $\eta^2 \geq 0.475$) en az orta düzeyde olduğu görülmektedir. Üçlü ortak

etkiler içerisinde yalnızca alt test sayısı*alt test uzunluğu*korelasyon etkileşiminin anlamlı etkisinin olmadığı gözlenirken en fazla etkiye sahip üçlü ortak etkileşimin alt test sayısı*korelasyon*model (kısmi $\eta^2=0.837$) olduğu, en az etkiye sahip etkileşimin ise alt test sayısı*alt test uzunluğu*model (kısmi $\eta^2=0.006$) olduğu görülmektedir. Dörtlü etkinin alt testlere ait güvenilirlik değerlerinin varyansına katkısının çok düşük düzeyde (kısmi $\eta^2=0.997$) olduğu görülmektedir.

Çoklu karşılaştırma testi sonucunda ana etkiler için koşulların tüm düzeyleri arasında anlamlı farklılık olduğu bulunmuştur. Korelasyon arttıkça güvenilirliğin azaldığı fakat alt test uzunluğu ve alt test sayısı arttıkça güvenilirliğin arttığı gözlenmiştir. Modeller açısından en güvenilir kestirim yapan modeller sırasıyla Üst Düzey Sıralı, Hiyerarşik ÇBMTK ve İki Faktör modeldir. Alt test uzunluğu*model, alt test sayısı*model, alt test sayısı*korelasyon ve alt test uzunluğu*korelasyon ikili etkileşimlerinin çoklu karşılaştırma testi sonuçlarında tüm ikili karşılaştırmaları anlamlı bulunmuştur. Korelasyon*model ikili etkileşimin modellere göre ikili karşılaştırma testi sonucunda korelasyonun tüm koşulları için üç modelin tüm ikili karşılaştırmaları arasında anlamlı farklılık bulunurken korelasyona göre ikili karşılaştırma testi sonucunda ise İki Faktör ve Üst Düzey Sıralı modeller için korelasyonun tüm ikili etkileşimleri arasında anlamlı farklılık gözlenmiş fakat Hiyerarşik ÇBMTK model için korelasyonun tüm düzeyleri arasında anlamlı farklılık gözlenmemiştir. Alt test uzunluğu*korelasyon*model üçlü etkileşimin modellere göre ikili karşılaştırma testi sonucunda 40 madde ve 0.0 korelasyon düzeyi için Hiyerarşik ÇBMTK model ile Üst Düzey Sıralı model arasında anlamlı farklılık gözlenmezken üç modelin korelasyon ve alt test uzunluğunun diğer tüm ikili karşılaştırmaları arasında anlamlı farklılık bulunmuştur. Yine bu üçlü etkileşimin korelasyona göre ikili karşılaştırma testi sonucunda ise İki Faktör ve Üst Düzey Sıralı modeller için korelasyonun tüm ikili etkileşimleri arasında anlamlı farklılık gözlenmiş fakat Hiyerarşik ÇBMTK model için alt test uzunluğu ve korelasyonun tüm düzeyleri arasında anlamlı farklılık gözlenmemiştir. Alt test sayısı*korelasyon*model üçlü etkileşimin modellere göre ikili karşılaştırma testi sonucunda 2 boyut ve 0.3 korelasyon düzeyi için Hiyerarşik ÇBMTK model ile Üst Düzey Sıralı model arasında anlamlı farklılık gözlenmezken üç modelin korelasyon ve alt test sayısının diğer tüm ikili karşılaştırmaları arasında anlamlı farklılık bulunmuştur. Yine bu üçlü etkileşimin alt test sayısına göre tüm ikili etkileşimleri arasında anlamlı farklılıklar gözlenmiştir.

Ayrıca korelasyona göre ikili karşılaştırma testi sonucunda İki Faktör ve Üst Düzey Sıralı modeller için korelasyonun tüm ikili etkileşimleri arasında anlamlı farklılık gözlenmiş fakat Hiyerarşik ÇBMTK model için korelasyonun tüm düzeyleri arasında anlamlı farklılık gözlenmemiştir.

4.3.5.Üçüncü Alt Probleme İlişkin Tartışma

Simülasyon çalışmasında tüm koşullar altında elde edilen **toplam test puanlarına** ait RMSE ve güvenilirlik değerleri üzerinde araştırma koşullarından korelasyon ve model ana etkilerinin ve bu iki değişkenin ortak etkisinin en fazla etkiye sahip olması araştırmada ele alınan hiyerarşik modellerin toplam test puanını ele alma varsayımlarındaki farklılıktan kaynaklanmaktadır. Hiyerarşik ÇBMTK Model toplam test puanını maksimum test bilgi fonksiyonuna bağlı hesaplarken Üst Düzey Sıralı Model toplam test puanının alt testlerin lineer bir fonksiyonu olduğunu kabul etmektedir. İki Faktör Model’de ise toplam test puanını tüm test maddeleri üzerinden tek boyutlu yaklaşımla hesaplanmaktadır. Karşılaştırma testlerinde Hiyerarşik ÇBMTK modele ait RMSE değerleri üzerinde yalnızca korelasyonun en yüksek düzeyi ile diğer düzeyleri arasında anlamlı farklılık gözlenmesinin nedeni alt testler arasındaki korelasyon yüksek olduğunda testin boyutluluk derecesinin düşmesidir. Dolayısıyla bu model için korelasyonun ilk üç düzeyinde kestirim hataları ve güvenilirliği anlamlı düzeyde değişmemektedir.

Simülasyon çalışmasında tüm koşullar altında elde edilen **alt test puanlarına** ait RMSE ve güvenilirlik değerleri üzerinde araştırma koşullarından test uzunluğu, korelasyon ve model ana etkileri, alt test uzunluğu*alt test sayısı dışındaki ikili ortak etkileri ile alt test sayısı*korelasyon*model ve alt test uzunluğu*korelasyon*model üçlü etkilerin anlamlı ve en az orta düzeyde etkiye sahip olması araştırmada ele alınan hiyerarşik modellerin varsayımlarından kaynaklanmaktadır. Üst Düzey Sıralı Model toplam test puanının alt testlerin bir lineer bir fonksiyonu olduğunu kabul edilirken İki Faktör Model’de alt test puanları ile toplam test puanlarının ilişkisiz olduğu varsayılır ve alt test puanları toplam test puanlarının artık varyansından hesaplanmaktadır. Bu varsayımlar nedeniyle model ve korelasyona bağlı ana ve ortak etkilere ait karşılaştırma testleri sonucunun hemen hemen tüm ikili karşılaştırmaları anlamlı düzeyde farklılık göstermektedir. Korelasyonun Hiyerarşik ÇBMTK model kestirimlerinin hatası ve güvenilirliği üzerinde etkili olması beklenirken ikili karşılaştırma testleri sonucunda korelasyon düzeyleri için bu modelde anlamlı

farklılık gözlenmemiştir. Bu durum Bulut'un (2013) varyans analizi sonucunda korelasyonun ÇBMTK modele ait güvenilirlik değerleri üzerinde orta düzeyde etkisi olduğunu gösterdiği çalışmasının sonucundan farklıdır ve araştırmanın beklenmeyen sonucudur. Bu sonucun veri üretme koşullarında kullanılan madde ve yetenek parametrelerinin dağılımı gibi değişkenlerin etkisinden kaynaklanabileceği düşünülmektedir.

4.4. Dördüncü Alt Probleme İlişkin Bulgular ve Tartışma

Araştırmanın dördüncü alt problemi kapsamında "TEOG (2015) verilerinin Üst Düzey Sıralı (Higher Order), İki Faktör (Bi-factor) ve hiyerarşik çok boyutlu madde tepki kuramı modellerine göre alt test puan kestirimlerinin nasıl değiştiği" modellerin kestirdiği sonsal dağılımın ortalama ve standart sapması ve alt testler arası korelasyon değerleri ve standart sapması ile incelenmiştir. TEOG 2015 verisinin üç kestirim modeline göre her bir alt test için kestirilen sonsal dağılımın ortalama ve standart sapma değerleri Tablo 4.13'te verilirken kestirilen alt testler arası korelasyon matrisi ise Tablo 4.14'te verilmiştir.

Tablo 4.13: Sonsal Dağılımın Ortalama ve Standart Sapması

Alt testler	<i>İki Faktör Model</i>		<i>Hiyerarşik ÇBMTK Model</i>		<i>Üst Düzey Sıralı Model</i>	
	<i>Ortalama</i>	<i>Std. Sapma</i>	<i>Ortalama</i>	<i>Std. Sapma</i>	<i>Ortalama</i>	<i>Std. Sapma</i>
Din Kültürü	0.004	0.009	0.009	0.012	0.009	0.013
Fen Bilgisi	-0.008	0.014	-0.004	0.010	-0.001	0.009
İngilizce	-0.017	0.021	-0.026	0.028	-0.024	0.026
Matematik	-0.015	0.017	-0.017	0.019	-0.014	0.017
Tarih	-0.005	0.009	-0.005	0.009	-0.002	0.008
Türkçe	0.002	0.008	0.005	0.013	0.000	0.008

Tablo 4.13 incelendiğinde İki Faktör modelin Hiyerarşik ÇBMTK modele göre Din Kültürü, İngilizce, Matematik ve Türkçe alt testlerinde Üst Düzey Sıralı modele göre ise Din Kültürü ve İngilizce alt testlerinde daha düşük sonsal standart sapmaya sahip olduğu görülmektedir. Üst Düzey Sıralı modelin Hiyerarşik ÇBMTK modele göre Fen Bilgisi, İngilizce, Matematik, Tarih ve Türkçe alt testlerinde İki Faktör modele göre ise Fen Bilgisi ve Tarih alt testlerinde daha düşük sonsal standart sapmaya sahip olduğu görülmektedir. Hiyerarşik ÇBMTK modelin İki Faktör modele göre yalnızca Fen Bilgisi alt testinde ise Üst Düzey Sıralı modele göre ise yalnızca Din Kültürü alt testinde daha düşük sonsal standart sapmaya sahip olduğu

görülmektedir. Tüm yöntemlerin bütün alt testler açısından sonsal dağılım ortalama ve standart sapmaları bir arada değerlendirildiğinde üç yönteminde genel olarak düşük sonsal standart sapmaya sahip olduğu fakat İki Faktör modelin diğer yöntemlere göre az farkla daha iyi sonuç verdiği söylenebilir.

Tablo 4.14: Modellerin Kestirdiği ve TEOG Verisinin Alt Testler Arası Korelasyon Matrisi

<i>Model</i>	<i>Alt testler</i>	<i>Din Kültürü</i>	<i>Fen Bilgisi</i>	<i>İngilizce</i>	<i>Matematik</i>	<i>Tarih</i>	<i>Türkçe</i>
İki Faktör	Fen Bilgisi	-0.032	1	0.001	-0.009	-0.012	-0.007
	İngilizce	-0.037	0.001	1	-0.005	-0.018	-0.009
	Matematik	-0.036	-0.009	-0.005	1	0.000	-0.009
	Tarih	-0.037	-0.012	-0.018	0.000	1	0.006
	Türkçe	-0.041	-0.007	-0.009	-0.009	0.006	1
ÇBMTK	Fen Bilgisi	0.009	1	0.006	-0.007	-0.006	-0.005
	İngilizce	-0.008	0.006	1	-0.004	-0.012	-0.006
	Matematik	-0.005	-0.007	-0.004	1	0.005	-0.005
	Tarih	0.012	-0.006	-0.012	0.005	1	0.010
	Türkçe	-0.007	-0.005	-0.006	-0.005	0.010	1
Üst Düzey	Fen Bilgisi	0.008	1	0.006	-0.005	-0.005	-0.003
	İngilizce	-0.008	0.006	1	-0.004	-0.015	-0.010
	Matematik	-0.004	-0.005	-0.004	1	0.007	-0.007
	Tarih	0.015	-0.005	-0.015	0.007	1	0.013
	Türkçe	-0.006	-0.003	-0.010	-0.007	0.013	1
TEOG	Fen Bilgisi	-0.002	1	0.006	-0.005	-0.005	-0.003
	İngilizce	-0.018	0.013	1	-0.004	-0.015	-0.010
	Matematik	0.013	-0.011	-0.007	1	0.007	-0.007
	Tarih	0.017	0.011	-0.009	0.000	1	0.013
	Türkçe	0.040	-0.001	0.006	-0.081	0.010	1

Tablo 4.14'te üç model tarafından kestirilen alt testler arası korelasyon matrisi incelendiğinde üç modelin de gerçek veri matrisinin korelasyon matrisine çok benzer kestirimler yaptığı görülmektedir. Tablo 4.14'teki korelasyon matrisinin standart sapma değerleri ise EK-11'de verilmiştir. Standart sapma değerleriyle birlikte korelasyon değerleri incelendiğinde Hiyerarşik ÇBMTK model ile Üst Düzey Sıralı modelin birbirine çok benzer sonuçlar verdiği ve çok az farkla İki Faktör modelin diğer modellere göre standart sapma değerlerinin yüksek olduğu söylenebilir. Gerçek veri setinden elde edilen bu sonucun simülasyon çalışmasınının 0.0 korelasyon ve 20 maddelik alt testlerden elde edilen sonuçları ile uyumludur.

5. SONUÇ ve ÖNERİLER

Bu bölümde araştırmının bulgu ve yorumlarına dayalı olarak ulaşılan sonuçların özetine ve bu sonuçlardan yola çıkarak geliştirilen önerilere yer verilmiştir.

5.1. Sonuçlar

5.1.1. Birinci Alt Probleme İlişkin Sonuçlar

- 1) İki ve üç boyutlu verilerde toplam test puanı için kestirilen yetenek parametrelerinin RMSE değerleri, İki Faktör modelde alt test uzunluğu ve alt testler arasındaki korelasyon arttıkça azalmaktadır.
- 2) İki ve üç boyutlu verilerde toplam test puanı için kestirilen yetenek parametrelerinin RMSE değerleri, Hiyerarşik ÇBMTK modelde alt test uzunluğu arttıkça azalırken alt testler arasındaki korelasyon arttıkça artmaktadır.
- 3) İki boyutlu verilerde toplam test puanı için kestirilen yetenek parametrelerinin RMSE değerleri, Üst Düzey Sıralı modelde alt testler arası korelasyonun düşük düzeyleri için (0.0 ve 0.3) alt test uzunluğu 30'dan 40'a çıkarıldığında bir miktar artış göstermekle birlikte diğer koşullarda alt test uzunluğu arttıkça azalmaktadır. Üç boyutlu verilerde yine bu model için korelasyonun 0.0 düzeyi için benzer bir durum görülmektedir. Ayrıca bu model için İki Faktör modelde olduğu gibi alt testler arasındaki korelasyon arttıkça RMSE değerleri azalmaktadır.
- 4) Hiyerarşik ÇBMTK model ile tüm koşullarda en düşük RMSE değeri elde edilmiş, ikinci sırada ise Üst Düzey Sıralı model yer almıştır. Korelasyon 0.80 olduğunda, tüm modeller birbirine yakın RMSE değerleri vermiştir.
- 5) Alt test sayısındaki artış tüm modellerin toplam test puanı için yetenek parametresi kestirim hatasını azaltmaktadır. Alt test sayısındaki artış yetenek parametrelerinin RMSE değerlerinde İki Faktör modelde alt testler arası korelasyonun koşulunun yüksek düzeylerinde; Hiyerarşik ÇBMTK ve Üst Düzey Sıralı modellerde ise korelasyonun düşük düzeylerinde daha fazla azalmaya neden olmaktadır.

- 6) İki ve üç boyutlu verilerde toplam test puanı için kestirilen yetenek parametrelerinin güvenilirlik kestirimleri tüm modeller için korelasyonlar ve alt test uzunlukları arttıkça artmaktadır.
- 7) İki ve üç boyutlu verilerde toplam test puanı için kestirilen yetenek parametreleri için en güvenilir kestirim Hiyerarşik ÇBMTK model ile elde edilmiş, ikinci sırada Üst Düzey Sıralı model yer almıştır. Hiyerarşik ÇBMTK model, düşük korelasyon ve alt test uzunluğunda bile kabul edilebilir güvenilirlik kestirimleri vermiştir.
- 8) İki ve üç boyutlu verilerde toplam test puanı için kestirilen yetenek parametreleri için korelasyon 0.80 iken tüm modeller birbirine yakın güvenilirlik değerleri vermiştir.
- 9) Hiyerarşik ÇBMTK ve Üst Düzey Sıralı modellerde tüm koşullarda üç boyutlu testten elde edilen güvenilirlik kestirimleri iki boyutlu testten daha yüksek bulunmuştur.
- 10) İki Faktör modelde korelasyonun 0.0 düzeyi dışındaki tüm koşullarda alt test sayısındaki artış toplam test puanı için yetenek parametresi kestirim güvenilirliğini arttırmaktadır. Alt test sayısını ikiden üçe çıkarmanın toplam test yetenek parametresi kestirim güvenilirliğinde en fazla ortalama artış olan ve kabul edilebilir düzeydeki güvenilir kestirimlerde en fazla iyileşme olan modelin İki Faktör Model olduğu bulunmuştur.

5.1.2. İkinci Alt Probleme İlişkin Sonuçlar

- 1) İki ve üç boyutlu verilerde alt test puanı için kestirilen yetenek parametrelerinin RMSE değerleri, Hiyerarşik ÇBMTK modelde alt test uzunluğu arttıkça azalırken alt testler arasındaki korelasyon düzeyinden etkilenmemektedir.
- 2) İki ve üç boyutlu verilerde alt test puanı için kestirilen yetenek parametrelerinin RMSE değerleri, Üst Düzey Sıralı modelde alt test uzunluğu ve alt testler arasındaki korelasyon arttıkça azalmaktadır.
- 3) İki ve üç boyutlu verilerde alt test puanı için kestirilen yetenek parametrelerinin RMSE değerleri, İki Faktör modelde alt test uzunluğu

artıkça azalırken alt testler arasındaki korelasyon artıkça önemli düzeyde artmaktadır.

- 4) İki ve üç boyutlu verilerde hemen hemen tüm koşullarda en düşük RMSE değerleri Üst Düzey Sıralı modelden elde edilmiştir. Ayrıca koşulların çoğunda Hiyerarşik ÇBMTK ve Üst Düzey Sıralı Modelin alt test puanlarını benzer hatalar ile kestirmiş fakat korelasyonun yüksek düzeylerinde Üst Düzey Sıralı model daha iyi performans göstermiştir.
- 5) Alt test sayısındaki artış İki Faktör modelde alt test puanı için yetenek parametresi kestirim hatasını artırırken Üst Düzey Sıralı modelde azalmaktadır. Hiyerarşik ÇBMTK modelde iki ve üç boyutlu verilerden elde edilen RMSE değerleri ise aynıdır.
- 6) İki ve üç boyutlu verilerde alt test puanı için kestirilen yetenek parametrelerinin güvenilirlik değerleri, Üst Düzey Sıralı modelde alt test uzunluğu ve alt testler arasındaki korelasyon artıkça artmaktadır.
- 7) İki ve üç boyutlu verilerde tüm koşullar altında Hiyerarşik ÇBMTK modelin alt test yetenek parametresini kabul edilebilir güvenilir düzeyde kestirmektedir. Ayrıca bu model için güvenilirlik değerleri alt test uzunluğu artıkça artarken korelasyon düzeyinden etkilenmemektedir.
- 8) İki ve üç boyutlu verilerde alt test puanı için kestirilen yetenek parametrelerinin güvenilirlik değerleri, İki Faktör modelde alt test uzunluğu artıkça artarken alt testler arasındaki korelasyon artıkça önemli düzeyde azalmaktadır. Ayrıca bu yöntem için kabul edilebilir düzeyde güvenilir kestirimlerin korelasyonun 0.0 düzeyinde alt test uzunluğu koşulunun tüm düzeylerinde ve korelasyonun 0.3 düzeyinde alt test uzunluğu koşulunun 30 ve 40 madde düzeyinden elde edilmiştir.
- 9) İki ve üç boyutlu verilerde hemen hemen tüm koşullarda en yüksek güvenilirlik değerleri Üst Düzey Sıralı modelden; ikinci olarak Hiyerarşik ÇBMTK modelden elde edilmiştir.
- 10) Tüm koşullarda Üst Düzey Sıralı model için üç boyutlu testten elde edilen güvenilirlik kestirimleri iki boyutlu testten daha yüksek bulunurken boyut sayısının Hiyerarşik ÇBMTK model kestirim güvenilirliği üzerinde etkisi bulunmamıştır.

- 11) Alt test sayısındaki artış ile İki Faktör modelde alt test puanı için yetenek parametresi kestirim güvenilirliği korelasyonun 0.0 düzeyinde artarken korelasyonun yüksek düzeylerinde azalmaktadır.

5.1.3. Üçüncü Alt Probleme İlişkin Sonuçlar

Toplam test puanlarına ait RMSE ve güvenilirlik değerleri üzerinde yapılan varyans analizi sonucuna göre;

- 1) Tüm ana varyans kaynakları anlamlı düzeyde etkili olmakla birlikte en fazla etkiye sahip değişkenler model ve alt testler arası korelasyondur.
- 2) Ana etkilere ait tüm ikili karşılaştırmalar anlamlıdır ve alt testler arası korelasyon, alt test uzunluğu ve madde sayısı arttıkça kestirim hataları azalırken kestirim güvenilirliği artmaktadır.
- 3) Modellere ait kestirimler birbirinden anlamlı düzeyde farklıdır. En az hatalı ve en yüksek güvenilirlikli kestirimler sırasıyla Hiyerarşik ÇBMTK model, Üst Düzey Sıralı model ve İki Faktör Modelden elde edilmiştir.
- 4) Ortak etkiler açısından her iki değer üzerinde anlamlı en yüksek etkiye sahip varyans kaynağı ise korelasyon*model etkileşimidir.
- 5) İki Faktör ve Üst Düzey Sıralı modelde korelasyona artışı ile kestirim hatası anlamlı düzeyde azalırken kestirim güvenilirliği anlamlı düzeyde artmaktadır. İki faktör Model ile Üst Düzey Sıralı modelin kestirim güvenilirliği korelasyonun en düşük ve en yüksek düzeyinde aynıdır.
- 6) Hiyerarşik ÇBMTK modelde yalnızca korelasyonun ilk üç düzeyinden korelasyonun 0.8 düzeyine geçişte kestirim hatası ve güvenilirliği anlamlı düzeyde değişmektedir.

Alt test puanlarına ait RMSE ve güvenilirlik değerleri üzerinde yapılan varyans analizi sonucuna göre;

- 1) Tüm ana varyans kaynaklarının anlamlı düzeyde etkisi olmakla birlikte her iki değer üzerinde model, alt testler arası korelasyon ve alt test uzunluğu ana değişkenleri oldukça yüksek düzeyde etkiye sahiptir.

- 2) Ana etkilere ait tüm ikili karşılaştırmalar anlamlıdır. Korelasyon arttıkça kestirim hataları artıp kestirim güvenilirliği azalırken alt test uzunluğu arttıkça kestirim hataları azalıp kestirim güvenilirliği artmaktadır.
- 3) Modellere ait kestirimler birbirinden anlamlı düzeyde farklıdır. En az hatalı ve en yüksek güvenilirlikli kestirimler sırasıyla Üst Düzey Sıralı model, Hiyerarşik ÇBMTK model ve İki Faktör modelden elde edilmiştir.
- 4) Alt test sayısının kestirim hatası üzerinde anlamlı etkisi gözlenmezken kestirim güvenilirliği üzerinde düşük düzeyde anlamlı etkisi vardır. Alt test sayısı arttıkça kestirim güvenilirliği artmaktadır.
- 5) İkili ortak etkiler açısından alt test sayısı*alt test uzunluğu dışındaki tüm etkileşimler anlamlıdır. Ortak etkileşimlere ait hemen hemen tüm ikili karşılaştırmalar anlamlıdır. Alt test sayısı artışı İki Faktör Modelde hatayı azaltıp güvenliği artırırken Üst Düzey Sıralı ve Hiyerarşik ÇBMTK modellerde hatayı azaltıp güvenliği arttırmaktadır.
- 6) Tüm korelasyon düzeylerinde yöntemler arasında farklılık gözlenmektedir. İki faktör modelde korelasyon arttıkça hatalar artıp güvenilirlik azalırken Üst Düzey Sıralı modelde hatalar azalıp güvenilirlik artmaktadır. Hiyerarşik ÇBMTK modelde kestirim güvenilirliği ve hatası korelasyon düzeylerinden etkilenmemektedir.
- 7) Üçlü ortak etkiler açısından alt test sayısı*korelasyon*model anlamlı ve yüksek düzeyde etkiye sahiptir. Yalnızca 40 madde ve 0.0 korelasyon düzeyinde Hiyerarşik ÇBMTK model ile Üst Düzey model aynı güvenilirlik düzeyinde kestirim yapmakla birlikte alt test sayısı*korelasyon*model etkileşimine ait diğer tüm ikili karşılaştırmalar arasında anlamlı farklılıklar vardır.
- 8) Üçlü ortak etkiler açısından alt test uzunluğu*korelasyon*model ise anlamlı ve orta düzeyde etkiye sahiptir. Yalnızca 2 boyut ve 0.3 korelasyon düzeyinde Hiyerarşik ÇBMTK model ile Üst Düzey model aynı güvenilirlik ve hata düzeyinde kestirim yapmakla birlikte alt test uzunluğu*korelasyon*model etkileşimine ait diğer tüm ikili karşılaştırmalar arasında anlamlı farklılıklar mevcuttur.

5.1.4. Dördüncü Alt Probleme İlişkin Sonuçlar

- 1) TEOG 2015 gerçek verisinin faktör analizi sonucunda elde edilen alt testler arası korelasyon matrisi ile Hiyerarşik ÇBMTK Model, Üst Düzey Sıralı Model ve İki Faktör Model tarafından kestirilen matris karşılaştırıldığında modellerin gerçek duruma çok benzer kestirimler yaptığı görülmüştür.
- 2) Din Kültürü, Fen Bilgisi, İngilizce, Matematik, Tarih ve Türkçe alt testlerinin her bir yöntemden elde edilen sonsal dağılımın ortalama standart sapma değerleri karşılaştırıldığında üç yöntemde genel olarak düşük sonsal standart sapmaya sahip olduğu bulunmuştur. Fakat daha fazla alt düşük sonsal dağılım ortalama ve standart sapma değerleri verdiği için İki Faktör modelin diğer yöntemlere göre az farkla daha iyi sonuç verdiği sonucuna ulaşılmıştır.

5.2. Öneriler

5.2.1. Uygulamaya Dönük Öneriler

Araştırmadan elde edilen bulgulara dayanarak;

1. Hiyerarşik modellerin varsayımının hem toplam test hem de alt test puanlarının kestiriminde farklı performans göstermesi nedeniyle daha doğru ve güvenilir alt test ve toplam test puanı kestirimleri için öncelikle mevcut testin yapısal modeli tespit edilmelidir.
2. Toplam test puan kestirimlerinde araştırmada ele alınan tüm koşullar altında ve alt test puanların kestirimde ise hemen hemen tüm koşullarda en düşük hatalı ve en güvenilir kestirimlerin hiyerarşik ÇBMTK modelden elde edilmesi nedeniyle geniş ölçekli testlerin raporlanmasında bu modelin kullanımı önerilebilir.
3. Alt testler arasında orta ve düşük düzeyde ilişkilerin olduğu bilinen sınavların raporlanmasında hiyerarşik ÇBMTK Model'e alternatif olarak bu modelle çok yakın analizler yapabilen Üst Düzey Sıralı Model'in kullanımı da tercih edilebilir.
4. Alan yazında alt testler arasında yüksek düzeyde ilişki olduğu bilinen sınavlarda toplam test puanı kestirimleri için İki Faktör Model'in kullanımı önerilirken bu araştırmada ele alınan koşullara sahip sınavlar için alt test puan kestirimlerinde bu yöntemin kullanımı önerilmez.

5. Toplam test süresi göz önünde bulundurmak koşuluyla hem toplam hem de alt test puanları kestirimin doğruluğunu ve güvenilirliğini arttıracığı için alt test uzunluklarının arttırılması önerilebilir.

5.2.2. Araştırmaya Dönük Öneriler

1. Araştırmada verilerin madde ve yetenek parametrelerinin üretilmesinde kullanılan dağılımlarından farklı dağılım türleri için belirlenen koşullar altında benzer bir araştırma yapılabilir.
2. Araştırmada basit yapılı iki kategorili veri setleri kullanılmıştır. Karmaşık veri yapıları ya da farklı test yapıları için araştırmalar yapılabilir.
3. Açık uçlu maddelerin de kullanıldığı testler için alt test ve toplam test puanı kestirimleri üzerine araştırma yapılabilir.
4. Araştırmada bir testi oluşturan konu alanlarına göre alt testler analiz edilmiştir. Spesifik bir alt testin kendi içerisindeki konu alanı ya da yeterlik alanı gibi daha küçük birimleri üzerinde güvenilir alt puan kestirimine ilişkin araştırma yapılabilir.
5. Araştırmada kullanılan hiyerarşik ÇBMTK Model'de toplam test puanı kestirimleri maksimum test bilgi fonksiyona dayalı olarak hesaplanmıştır. Bu model için benzer bir çalışma alt testlerin toplam testin bir fonksiyonu olarak kabul edildiği bir ÇBMTK model ile tekrarlanabilir.
6. Araştırmada ikinci dereceden hiyerarşik modeller kullanılmıştır. Daha üst dereceli modeller için toplam ve alt test puan kestirimlerinin performansı incelenebilir.
7. Araştırmada iteratif bir yöntem olan MCMC yöntemi kullanılmıştır. Bunun yerine BS, MS gibi farklı kestirim yöntemlerinin performansı araştırılabilir.
8. MCMC analizleri için 2000'i burn in olmak üzere toplam 10000 iterasyon kullanılmıştır. MCMC analizlerinin uzun sürmesi nedeniyle daha düşük iterasyon sayılarının alt test ve toplam test kestirim doğruluğu ve güvenilirliğine etkisi araştırılabilir.

KAYNAKÇA

- American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing (US). (1999). Standards for educational and psychological testing. *Amer Educational Research Assn.*
- Bock, R. D., Thissen, D., & Zimowski, M. F. (1997). IRT estimation of domain scores. *Journal of Educational Measurement, 34*, 197-211.
- Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika, 64*, 153–168.
- Brandt, S. (2012). *Definition and classification of a generalized subdimension model*. Conference of the National Council on Measurement in Education. Van-couver, Kanada, 14-16 Nisan 2012. [Çevrim-içi: http://www.artofreduction.com/wp-content/uploads/anleitungen/Definition%20and%20Classification%20of%20a%20Generalized%20Subdimension%20Model_v0.4b.pdf, Erişim tarihi: Ekim 2014.]
- Brandt, S., & Duckor, B. (2013). Increasing unidimensional measurement precision using a multidimensional item response model approach. *Psychological Test and Assessment Modeling, 55*(2), 148-161.
- Brennan, R. L. (2012). *Utility indexes for decisions about subscores*. Center for Advanced Studies in Measurement and Assessment (CASMA). Research Report 33.
- Browne, W. J., Rasbash, J. & Charlton, C. (2009). *MCMC estimation in MLwiN*. Centre of Multilevel Modelling, University of Bristol. United Kingdom.
- Bulut, O. (2013). *Between-person and within-person subscore reliability: Comparison of unidimensional and multidimensional IRT models*. Doctoral Dissertation. University of Minnesota.
- Calderhead, B. (2014). A general construction for parallelizing Metropolis– Hastings algorithms. *Proceedings of the National Academy of Sciences, 111*(49), 17408-17413.
- Cai, L., Yang, J. S., & Hansen, M. (2011). Generalized full-information item bifactor analysis. *Psychological methods, 16*(3), 221.
- Chang, Y. F. (2015). *A Restricted Bi-factor Model of Subdomain Relative Strengths and Weaknesses*. Doctoral Dissertation. University of Minnesota.
- Chen, F. F., West, S. G., & Sousa, K. H. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research, 41*(2), 189-225.
- Crocker, L. & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: CBS College Publishing.
- Çakıcı Eser, D. (2015). *Çok boyutlu madde tepki kuramının farklı modellerinden çeşitli koşullar altında kestirilen parametrelerin incelenmesi*. Doktora tezi. Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

- de la Torre, J., & Patz, R.J. (2005). Making the most of what we have: A practical application of multidimensional IRT in test scoring. *Journal of Educational and Behavioral Statistics*, 30, 295–311.
- de la Torre, J. (2009). Improving the quality of ability estimates through multidimensional scoring and incorporation of ancillary variables. *Applied Psychological Measurement*, 33, 465–485.
- de la Torre, J., & Song, H. (2009). Simultaneous estimation of overall and domain abilities: A higher-order IRT model approach. *Applied Psychological Measurement*, 33, 620–639.
- de la Torre, J., Song, H., & Hong, Y. (2011). A comparison of four methods of IRT subscore. *Applied Psychological Measurement*, 35(4), 296-316.
- 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.
- ETS. (2014). *ETS standards for quality and fairness*. Princeton, NJ: Educational Testing Service. [Çevrim-içi: <https://www.ets.org/s/about/pdf/standards.pdf>, Erişim tarihi:11 Ekim 2017.]
- Edwards, M. C., & Vevea, J. L. (2006). An empirical Bayes approach to subscore augmentation: How much strength can we borrow?. *Journal of Educational and Behavioral Statistics*, 31(3), 241-259.
- Feinberg, R. (2012). *A simulation study of the situations in which reporting subscores can add value to licensure examinations*. Doctoral Dissertation, University of Delaware.
- Ferrara, S., & DeMauro, G. E. (2007). Standardized assessment of individual achievement in K–12. In R. L. Brennan (Eds.). *Educational measurement*, 579–622. Westport, CT: Praeger.
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2011). *How to design and evaluate research in education*. (8th edition). Boston: McGraw – Hill.
- Gall M. D., Gall, J. P., & Borg, W., R. (2003). *Educational research: An introduction*. (7th. Edition). Pearson Education, Inc.
- Gibbons, R. D., & Hedeker, D. (1992). Full-information item Bi-factor analysis. *Psychometrika*, 57, 423–436.
- Gignac, G. E., Palmer, B. R., & Stough, C. (2007). A confirmatory factor analytic investigation of the TAS–20: Corroboration of a five-factor model and suggestions for improvement. *Journal of Personality Assessment*, 89(3), 247-257.
- Goodman, D. P., & Hambleton, R. K. (2004). Student test score reports and interpretive guides: Review of current practices and suggestions for future research. *Applied Measurement in Education*, 17, 145–220.
- Green, D. R., Yen, W. M., & Burket, G. R. (1989). Experiences in the application of item response theory in test construction. *Applied Measurement in Education*, 2(4), 297-312.

- Haberman, S. J. (2005). When can subscores have value?. ETS Research Report No. RR-05-08. [Çevrim-içi: <http://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.2005.tb01985.x/epdf>, Erişim tarihi:21 Kasım 2014.]
- Haberman, S. J. (2008). When can subscores have value? *Journal of Educational and Behavioral Statistics*, 33, 204–229.
- Haberman, S., Sinharay, S., & Puhan, G. (2009). Reporting subscores for institutions. *British Journal of Mathematical and Statistical Psychology*, 62(1), 79–95.
- Hadfield, J. D. (2010). MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *Journal of Statistical Software*, 33(2), 1-22.
- Haladyna, T. M., & Kramer, G. A. (2004). The validity of subscores for a credentialing Test. *Evaluation & The Health Professions*, 27(4), 349–368.
- Hambleton, R. K. & Swaminathan, H. (1985). *Item response theory principles and applications*. Boston, MA: Kluwer-Nijhoff Publishing.
- Hartig, J., & Höhler, J. (2008). Representation of competencies in multidimensional IRT models with within-item and between-item multidimensionality. *Zeitschrift für Psychologie/Journal of Psychology*, 216(2), 89-101.
- Harwell, M., Stone, C. A., Hsu, T. C., & Kirisci, L. (1996). Monte Carlo studies in item response theory. *Applied Psychological Measurement*, 20(2), 101-125.
- Holmes, C. C., & Held, L. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian analysis*, 1(1), 145-168.
- Holzinger, K. J., & Swineford, F. (1937). The Bi-factor method. *Psychometrika*, 2, 41–54.
- Huang, H. Y., Wang, W. C., Chen, P. H., & Su, C. M. (2013). Higher-order item response models for hierarchical latent traits. *Applied Psychological Measurement*, 37(8), 619-637.
- Jiang, S., Wang, C., & Weiss, D. J. (2016). Sample size requirements for estimation of item parameters in the multidimensional graded response model. *Frontiers in psychology*, 7(109), 1-10.
- Kahraman, N. & Thompson, T. (2011). Relating unidimensional IRT parameters to a multidimensional response space: A review of two alternative projection IRT models for scoring subscales. *Journal of Educational Measurement*, 48(2), 146–164.
- Kelley, T. L. (1927). *The interpretation of educational measurements*. New York: World Book.
- Kelly, T. L. (1947). *Fundamentals of statistics*. Cambridge: Harvard University Press
Kerlinger.
- Koziol, N. A. (2016). Parameter recovery and classification accuracy under conditions of testlet dependency: a comparison of the traditional 2pl, testlet, and bifactor models. *Applied Measurement In Education*, 29(3), 184–195

- Köse, İ.A. (2010). *Madde tepki kuramına dayalı tek boyutlu ve çok boyutlu modellerin test uzunluğu ve örneklem büyüklüğü açısından karşılaştırılması*. Doktora Tezi. Ankara Üniversitesi Eğitim Bilimleri Enstitüsü.
- Lee, J. (2012). *Multidimensional item response theory: an investigation of interaction effects between factors on item parameter recovery using Markov Chain Monte Carlo*. Doctoral Dissertation, Michigan State University, Measurement and Quantitative Methods.
- Ling, G. (2012). *Why the major field test in business does not report subscores: Reliability and construct validity evidence*. ETS Research Report No. RR-12-11. [Çevrim-içi: <http://onlinelibrary.wiley.com/doi/10.1002/j.2333-8504.2012.tb02293.x/epdf>. Erişim tarihi:21 Kasım 2014.]
- Luecht, R. M. (2003). *Applications of Multidimensional Diagnostic Scoring for Certification and Licensure Tests*. Annual Meeting of the National Council on measurement in Education, Chicago, IL, 22-24 Nisan 2003. [Çevrim-içi: http://www.aicpa.org/BecomeACPA/CPAExam/PsychometricsandScoring/TechnicalReports/DownloadableDocuments/Luecht_Applications.pdf Erişim tarihi:18 Kasım 2014.]
- Md Desa, Z. N. D. (2012). *Bi-factor multidimensional item response theory modeling for subscore estimation, reliability, and classification*. Doctoral dissertation. University of Kansas
- Meijer, R. R., Boevé, A. J., Tendeiro, J. N., Bosker, R. J., & Albers, C. J. (2017). The Use of Subscores in Higher Education: When Is This Useful?. *Frontiers in Psychology*, 8, 1-6.
- Messick, S. (1989). Validity. In R. L. Linn (Eds.). *Educational measurement*, 13-103, New York, NY: Macmillan.
- Monaghan, W. (2006). *The fact about subscores*. ETS Research Report No. RDC-04. [Çevrim-içi: https://www.ets.org/Media/Research/pdf/RD_Connections4.pdf. Erişim tarihi:21 Kasım 2014.]
- Özkan, Y. Ö. (2012). *Öğrenci başarılarının belirlenmesi sınavından (ÖBBS) klasik test kuramı, tek boyutlu ve çok boyutlu madde tepki kuramı modelleri ile kestirilen başarı puanlarının karşılaştırılması*. Doktora Tezi. Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Patz, R. J., & Junker, B. W. (1999a). A straightforward approach to Markov chain Monte Carlo methods for item response models. *Journal of educational and behavioral Statistics*, 24(2), 146-178.
- Patz, R. J., & Junker, B. W. (1999b). Applications and extensions of MCMC in IRT: Multiple item types, missing data, and rated responses. *Journal of educational and behavioral statistics*, 24(4), 342-366.
- Raftery, A. E., & Lewis, S. (1991). *How many iterations in the Gibbs sampler?*. Reported from University of Washington, Department of Statistics. [Çevrim-içi: <http://www.dtic.mil/docs/citations/ADA640705>. Erişim tarihi: Eylül 2017]
- Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement*, 21, 25–36.

- Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of personality assessment*, 95(2), 129-140.
- Schmid, J., & Leiman, J. M. (1957). The development of hierarchical factor solutions. *Psychometrika*, 22(1), 53-61.
- Sinharay, S. (2013). A note on assessing the added value of subscores. *Educational Measurement: Issues and Practice*, 32(4), 38-42.
- SAS/STAT (2009), 9.2 User's Guide.[Çevrim-içi:
<https://support.sas.com/documentation/cdl/en/statug/63033/PDF/default/statug.pdf>. Erişim tarihi: Eylül 2017]
- Seybert, J. (2013). *A new item response theory model for estimating person ability and item parameters for multidimensional rank order responses*. Doctoral Dissertation, University of South Florida.
- Sha, S., & McCoy, T. (2014). *A Comparison of two augmented subscore methods and the role of score distribution*, Annual meeting of National Council on Measurement and Education, Philadelphia, 2-3 Nisan 2014. [Çevrim-içi:
https://www.researchgate.net/publication/271203558_A_Comparison_of_Two_Augmented_Subscore_Methods_and_the_Role_of_Score_Distribution, Erişim tarihi: Eylül 2017.]
- Sheng, Y. (2005). *Bayesian analysis of hierarchical IRT models: comparing and combining the unidimensional & multi-unidimensional IRT models*. Doctoral dissertation, University of Missouri--Columbia.
- Sheng, Y., & Wikle, C. K. (2007). Comparing Multiunidimensional and unidimensional item response theory models. *Educational and Psychological Measurement*, 67(6) 899–919.
- Sheng, Y., & Wikle, C. K. (2008). Bayesian multidimensional IRT models with a hierarchical structure. *Educational and Psychological Measurement*, 68(3), 413–430.
- Sheng, Y. & Wikle, C. K. (2009). Bayesian IRT models incorporating general and specific abilities. *Behaviormetrika*, 36(1), 27–48.
- Shin, D. (2007). *A comparison of methods of estimating subscale scores for mixed-format tests*. Report for Pearson Educational Measurement. [Çevrim-içi
http://images.pearsonassessments.com/images/tmrs/tmrs_rq/EstimatingSubscaleScoresforMixedFormatItemsforPEMreportfinal.pdf?WT.mc_id=TMRS_A_Comparison_of_Methods_of_Estimating , Erişim tarihi: Eylül 2017.]
- Shin, C. D., Ansley, T., Tsai, T., & Mao X. (2005). *A comparison of methods of estimating objective scores*. Annual meeting of the National Council on Measurement in Education, Montreal, Canada, 12-14 Nisan 2005. [Çevrim-içi:
<https://elibrary.ru/item.asp?id=8884120> , Erişim tarihi: Eylül 2017.]
- Shin, D., & Chien, Y. (2007). *Using Bayesian sequential analyses in evaluating the prior effect for the estimation of subscale scores*. Annual meeting of the National Council on Measurement in Education, Chicago, 10-12 Nisan 2007. [Çevrim-içi:
http://assets.pearsonglobalschools.com/asset_mgr/legacy/200746/Using%20Bayesian%20Sequential%20Analyses_3972_1.pdf, Erişim tarihi: Eylül 2017.]

- Shin, D., & Li, D.(2007). *An Exploration of Methods for Evaluation of Individual and School Progress at the Subscale Level*. Annual Meeting of the National Council on Measurement in Education, Chicago, 10-12 Nisan 2007, [Çevrim-içi: http://assets.pearsonglobalschools.com/asset_mgr/legacy/200746/Using%20Bayesian%20Sequential%20Analyses_3972_1.pdf, Erişim tarihi: Eylül 2017.]
- Sinharay, S. (2010). How often do subscores have added value? Results from operational and simulated data. *Journal of Educational Measurement*, 47(2),150-174.
- Sinharay, S., Haberman, S. J., & Puhan, G. (2007). Subscores based on classical test theory: to report or not to report. *Educational Measurement: Issues and Practice*, 26(4), 21–28
- Sinharay, S., Puhan, G., & Haberman, S. J. (2011). An NCME instructional module on subscores. *Educational Measurement: Issues and Practice*, 30(3), 29–40.
- Skorupski, W. P. (2008). *A review and empirical comparison of approaches for improving the reliability of objective level scores*. Council of Chief State School Officers Konferansı, Ağustos 2008. [Çevrim-içi: http://www.ccsso.org/Documents/2008/Approaches_for_Improving_the_Reliability_2008.pdf, Erişim tarihi: Eylül 2017.]
- Skorupski, W. P., & Carvajal, J. (2010). A comparison of approaches for improving the reliability of objective level scores. *Educational and Psychological Measurement*, 70(3), 357-375.
- Stone, C. A., Ye, F., Zhu, X., & Lane, S. (2009). Providing subscale scores for diagnostic information: A case study when the test is essentially unidimensional. *Applied Measurement in Education*, 23(1), 63-86.
- Sun, S., Tao, J., Chang, H., & Shi, N. (2012).Weighted maximum-a-posteriori estimation in tests composed of dichotomous and polytomous items. *Applied Psychological Measurement*, 369(4),271–290.
- Tate, R. L. (2004). Implications of multidimensionality for total score and subscore performance. *Applied measurement in education*, 17(2). 89-112.
- Tao, S. (2009). *Using collateral information in the estimation of sub-scores—a fully Bayesian approach*. Doctoral Thesis, The University of Iowa.
- Thissen, D. & Edwards, M. C. (2005). *Diagnostic scores augmented using multidimensional item response theory: Preliminary investigation of MCMC strategies*. Annual meeting the National Council on Educational Measurement, Montreal, 12-14 April 2005.[Çevrim-içi: <http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=EA7D4F90C6FB267C4594DA6B579EA7BB?doi=10.1.1.551.9317&rep=rep1&type=pdf>. Erişim tarihi: Eylül 2014.]
- Wainer, H., Vevea, J. L., Camacho, F., Reeve, B. B., Rosa, K., Nelson, L.,Swygert, K. A., & Thissen, D. (2001). Test scoring. In D. Thissen and H. Wainer (Eds.). *Augmented score—“borrowing strength” to compute scores based on small numbers of items*, 343-387. Mahwah, Lawrence Erlbaum Associates, Inc
- Wang, W. C., & Wilson, M. (2005). The Rasch testlet model. *Applied Psychological Measurement*, 29(2), 126–149.

- Wang, W. C., Chen, P. H., & Cheng, Y. Y. (2004). Improving measurement precision of test batteries using multidimensional item response models. *Psychological Methods*, 9(1), 116.
- Yao, L. (2003). BMIRT: Bayesian multivariate item response theory [Software]. Monterey, CA: Defense Manpower Data Center. [Çevrim-içi: <http://www.bmirt.com>. Erişim tarihi: Aralık 2014]
- Yao, L. (2003). SimuMIRT [Software]. Monterey, CA: Defense Manpower Data Center. [Çevrim-içi: <http://www.bmirt.com>. Erişim tarihi: Aralık 2014]
- Yao, L. (2010). Reporting valid and reliable overall scores and domain scores. *Journal of Educational Measurement*, 47(3), 339-360.
- Yao, L. (2017). Comparing methods for estimating the abilities for the multidimensional models of mixed item types. *Communications in Statistics-Simulation and Computation*, 1-18.
- Yao, L., & Boughton, K. A. (2007). A multidimensional item response modeling approach for improving subscale proficiency estimation and classification. *Applied Psychological Measurement*, 31(2), 83-105.
- Yao, L., & Schwarz, R. D. (2006). A multidimensional partial credit model with associated item and test statistics: An application to mixed-format tests. *Applied Psychological Measurement*, 30, 469–492.
- Yen, W. M. (1980). The extent, causes and importance of context effects on item parameters for 2 latent trait models. *Journal of Educational Measurement*, 17(4), 297–311.
- Yen, W. M. (1987,). *A Bayesian/IRT index of objective performance*. Annual meeting of the Psychometric Society, Montreal, Quebec, Canada. Haziran 1987. [Çevrim-içi: http://www.ets.org/Media/Research/pdf/Yen_OPI_1987.pdf Erişim tarihi: Kasım 2014.]
- Zimowski, M. F. (1998). *BILOG-MG: Multiple-group IRT analysis and test maintenance for binary items*. Scientific Software International.

EKLER DİZİNİ

EK 1. ETİK KOMİSYONU ONAY BİLDİRİMİ

Form: 40

Tez Çalışması Etik Komisyon İzin Muafiyeti Formu

08 / 11 / 2017

Hacettepe Üniversitesi
Eğitim Bilimleri Enstitüsü
Eğitim Bilimleri Anabilim Dalı Başkanlığı'na

Tez Başlığı / Konusu:	Toplam Test Puanı ve Alt Test Puanlarının Kestiriminin Hiyerarşik Madde Tepki Kuramı Modelleri ile Karşılaştırılması
------------------------------	--

Yukarıda başlığı/konusu gösterilen tez çalışmam:

1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır,
2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.
3. Beden bütünlüğüne müdahale içermemektedir.
4. Gözlemsel ve betimsel araştırma (anket, ölçek/skala çalışmaları, dosya taramaları, veri kaynakları taraması, sistem-model geliştirme çalışmaları) niteliğinde değildir.

Hacettepe Üniversitesi Etik Kurulları ve Komisyonlarının Yönergelerini inceledim ve bunlara göre tez çalışmamın yürütülebilmesi için herhangi bir Etik Komisyondan/Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Sümeysa SOYSAL
(Öğrencinin Adı Soyadı, İmzası)

Öğrenci Bilgileri

Adı Soyadı	Sümeysa SOYSAL
Öğrenci No	H11162841
Anabilim Dalı	Eğitim Bilimleri
Programı	Eğitimde Ölçme ve Değerlendirme
Statüsü	<input type="checkbox"/> Yüksek Lisans <input type="checkbox"/> Doktora <input checked="" type="checkbox"/> Bütünleşik Dr.

Danışman Görüşü ve Onayı

"Toplam Test Puanı ve Alt Test Puanlarının Kestiriminin Hiyerarşik Madde Tepki Kuramı Modelleri ile Karşılaştırılması" başlıklı tez çalışmasında simülasyon ile veri üretilmiş olup herhangi bir birey ya da denek üzerinden veri toplanmamıştır.

Prof. Dr. Hülya KELEÇİOĞLU
(İmza)
(Danışmanın Ünvanı, Adı ve Soyadı)

EK-2. TEOG 2015 Verisinin Faktör Analizi Sonuçları

EK2.1 TEOG 2015 Verisinin Madde Faktör Yükü Matrisi

ROTATED LOADING MATRIX

(loadings lower than absolute 0.300 omitted)

Variable	F 1	F 2	F 3	F 4	F 5	F 6
V 1					0.598	
V 2					0.546	
V 3					0.623	
V 4					0.671	
V 5					0.829	
V 6					0.519	
V 7					0.727	
V 8					0.843	
V 9					0.740	
V 10					0.758	
V 11					0.612	
V 12					0.672	
V 13					0.722	
V 14					0.698	
V 15					0.629	
V 16					0.691	
V 17					0.323	
V 18					0.763	
V 19					0.699	
V 20					0.703	
V 21						0.631
V 22						0.739
V 23						0.663
V 24						0.616
V 25						0.654
V 26						0.607
V 27						0.456
V 28						0.705
V 29						0.522
V 30						0.526
V 31						0.607
V 32						0.657
V 33						0.741
V 34						0.731
V 35						0.611
V 36						0.590
V 37						0.669
V 38						0.631
V 39						0.642
V 40						0.679
V 41			0.687			
V 42			0.587			
V 43			0.564			
V 44			0.509			
V 45			0.716			
V 46			0.710			
V 47			0.738			
V 48			0.631			
V 49			0.540			
V 50						
V 51			0.434			
V 52			0.618			
V 53			0.712			
V 54			0.554			
V 55			0.627			
V 56			0.609			
V 57			0.374			
V 58			0.545			
V 59			0.624			
V 60						

EK-2.1 Devamı

V 60		
V 61		0.601
V 62		0.618
V 63		0.626
V 64		0.508
V 65		0.743
V 66		0.487
V 67		0.673
V 68		0.393
V 69		0.690
V 70		0.643
V 71		0.433
V 72		
V 73		0.532
V 74		0.492
V 75		0.458
V 76		0.597
V 77		0.679
V 78		0.514
V 79		0.642
V 80		0.579
V 81	0.733	
V 82	0.688	
V 83	0.709	
V 84	0.558	
V 85	0.682	
V 86	0.710	
V 87	0.550	
V 88	0.598	
V 89		
V 90	0.684	
V 91	0.634	
V 92	0.627	
V 93	0.497	
V 94	0.640	
V 95	0.638	
V 96	0.512	
V 97	0.564	
V 98	0.624	
V 99	0.714	
V 100	0.676	
V 101	0.598	
V 102	0.687	
V 103	0.503	
V 104	0.698	
V 105	0.797	
V 106	0.652	
V 107	0.793	
V 108	0.560	
V 109	0.570	
V 110	0.661	
V 111	0.540	
V 112	0.770	
V 113	0.724	
V 114	0.627	
V 115	0.580	
V 116	0.528	
V 117	0.636	
V 118	0.558	
V 119	0.457	
V 120	0.551	

EK-2.2. TEOG 2015 Verisinin Alt Testler Arası Korelasyon Matrisi

INTER-FACTORS CORRELATION MATRIX

Factor	F 1	F 2	F 3	F 4	F 5	F 6
F 1	1.000					
F 2	-0.002	1.000				
F 3	-0.018	0.013	1.000			
F 4	0.013	-0.011	-0.007	1.000		
F 5	0.017	0.011	-0.009	0.000	1.000	
F 6	0.004	-0.001	0.006	-0.081	0.001	1.000

F1: Din Kültürü, F2: Fen Bilgisi, F3:İngilizce, F4: Matematik, F5:Tarih, F6: Türkçe

EK-4. Örnek Simülasyon Verisinin Faktör Analizi Sonuçları

EK4.1 Örnek Simülasyon Verisinin Madde Faktör Yüğü Matrisi

(alt test sayısı=3, alt testler arası korelasyon=0.8, alt test uzunluğu=20)

Variable	F 1	F 2	F 3
V 1			0.468
V 2			0.427
V 3			
V 4			0.525
V 5			0.377
V 6			0.629
V 7			0.473
V 8			0.493
V 9			0.637
V 10			0.431
V 11			
V 12			0.436
V 13			0.373
V 14			0.564
V 15			0.414
V 16			0.339
V 17			0.508
V 18			0.626
V 19			0.375
V 20			0.587
V 21		0.421	
V 22			
V 23		0.582	
V 24		0.366	
V 25		0.433	
V 26		0.399	
V 27		0.377	
V 28		0.672	
V 29		0.579	
V 30		0.510	
V 31		0.393	
V 32		0.723	
V 33		0.609	
V 34		0.530	
V 35		0.576	
V 36		0.546	
V 37		0.535	
V 38		0.760	
V 39		0.490	
V 40		0.656	
V 41	0.500		
V 42	0.588		
V 43	0.490		
V 44	0.362		
V 45			
V 46	0.576		
V 47	0.634		
V 48	0.469		
V 49	0.437		
V 50	0.439		
V 51	0.692		
V 52	0.361		
V 53	0.334		
V 54	0.463		
V 55	0.420		
V 56	0.435		
V 57	0.380		
V 58	0.604		
V 59	0.496		
V 60	0.573		

EK-4.2. Örnek Simulasyon Verisinin Alt Testler Arası Korelasyon Matrisi

(alt test sayısı=3, alt testler arası korelasyon=0.8, alt test uzunluğu=20)

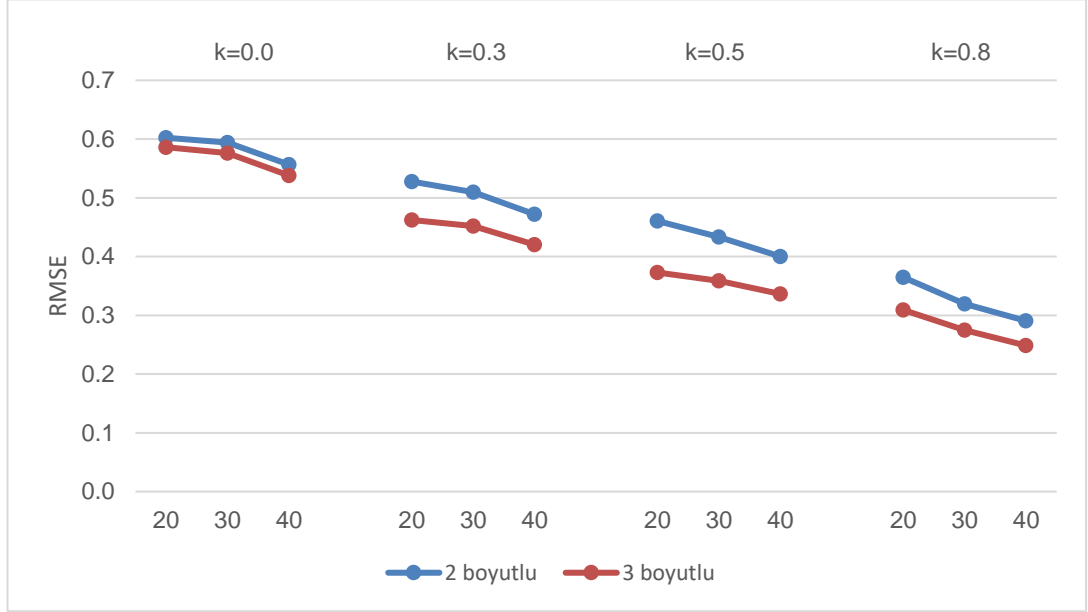
INTER-FACTORS CORRELATION MATRIX

Factor	F 1	F 2	F 3
F 1	1.000		
F 2	0.772	1.000	
F 3	0.767	0.776	1.000

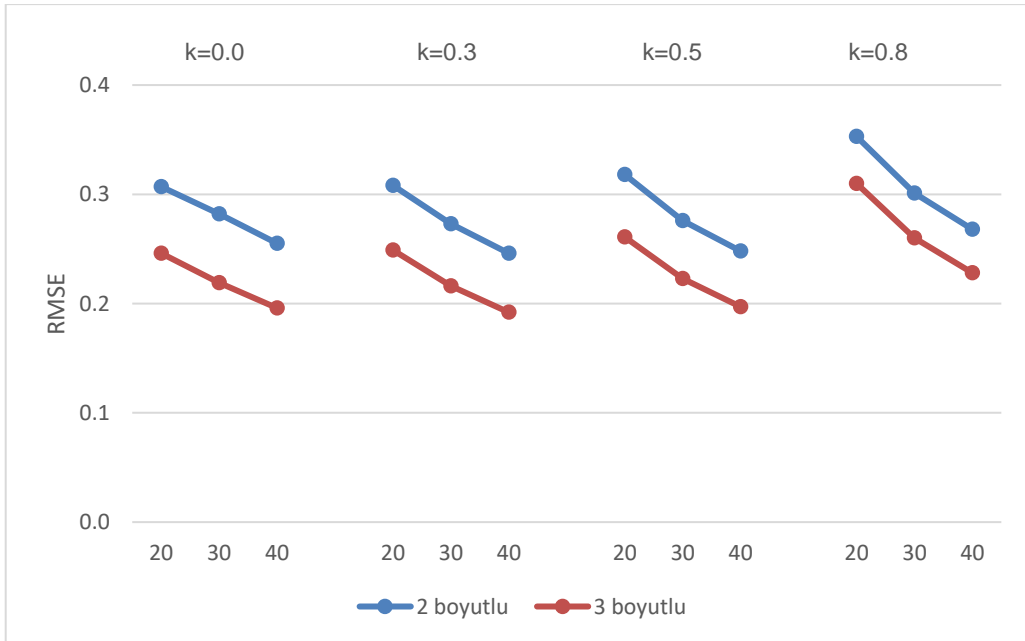
RMSEA=0.01

CFI=0.99

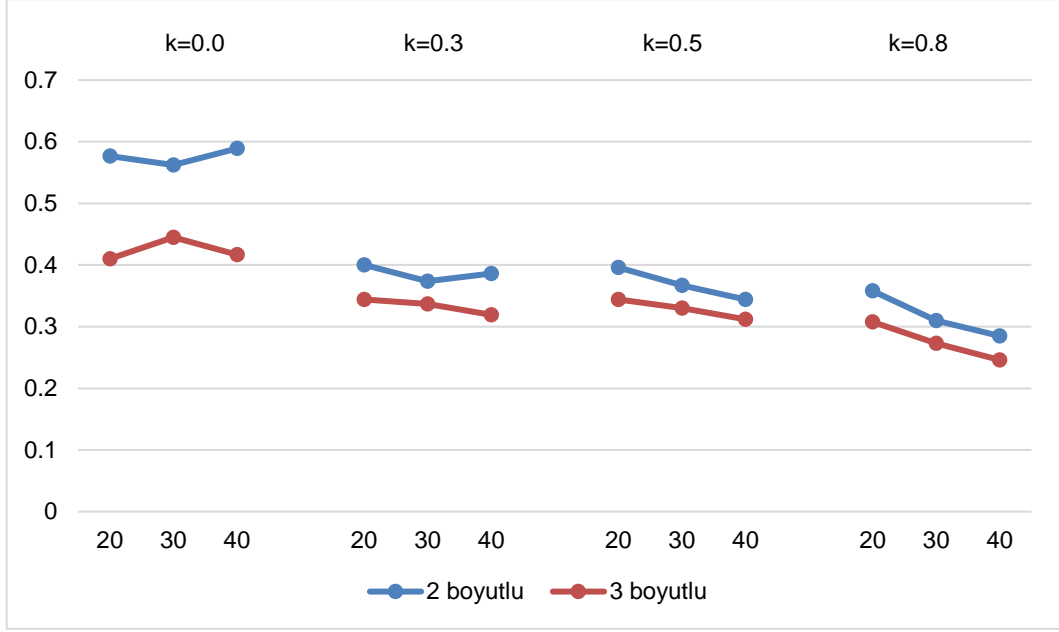
EK-5. Madde Tepki Kuramı Modellerine Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerlerine İlişkin Karşılaştırılmalı Grafikler



Şekil EK-5.1. İki Faktör Modele Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri

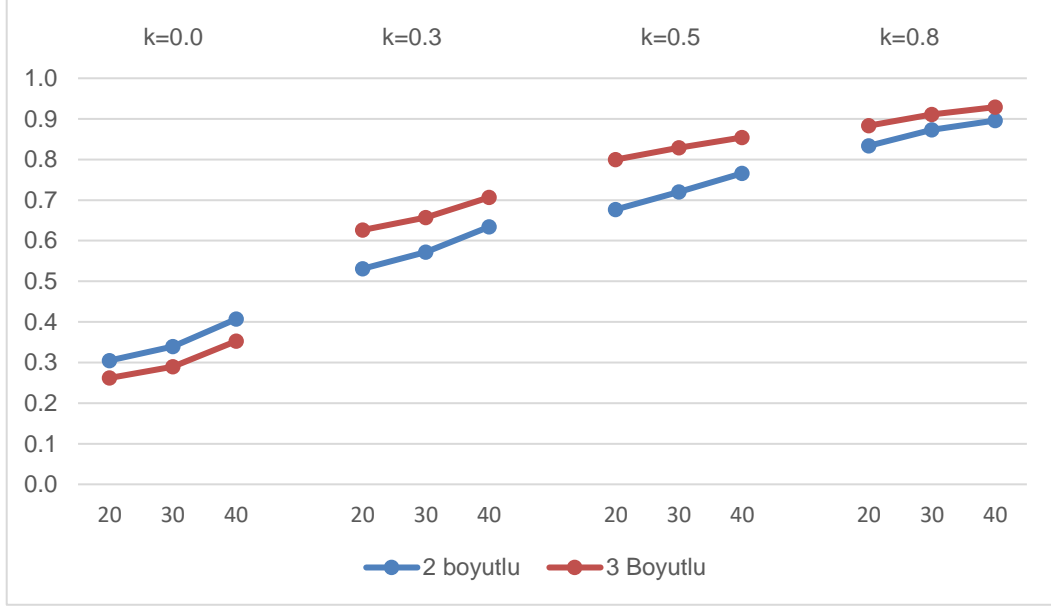


Şekil EK-5.2. Hiyerarşik ÇBMTK Modele Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri

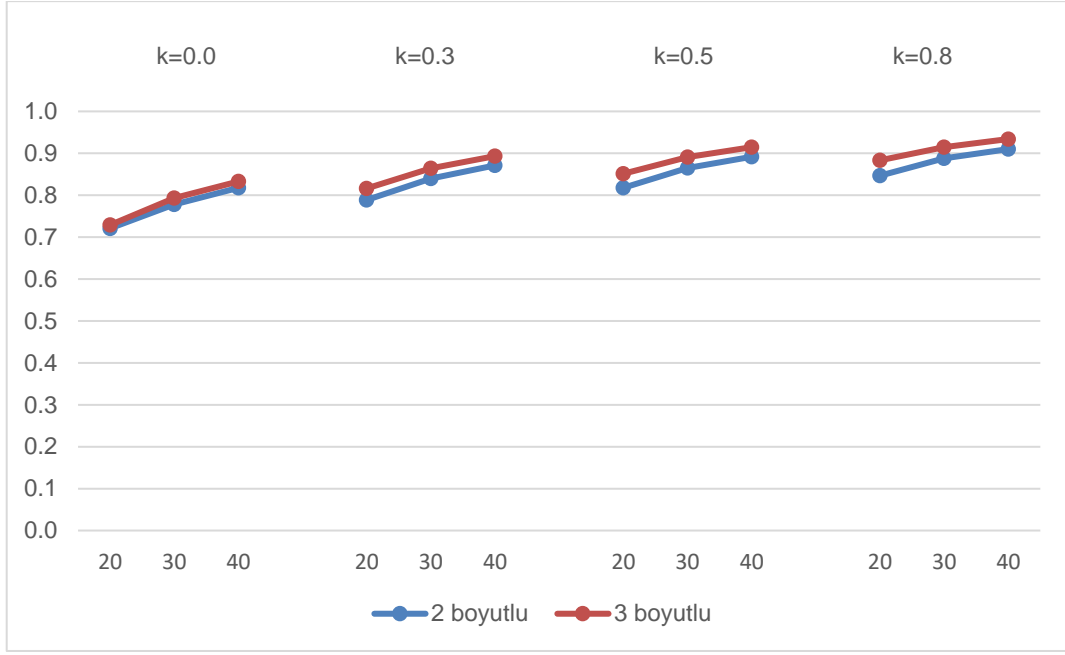


Şekil EK-5.3. Üst Düzey Sıralı Modele Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri

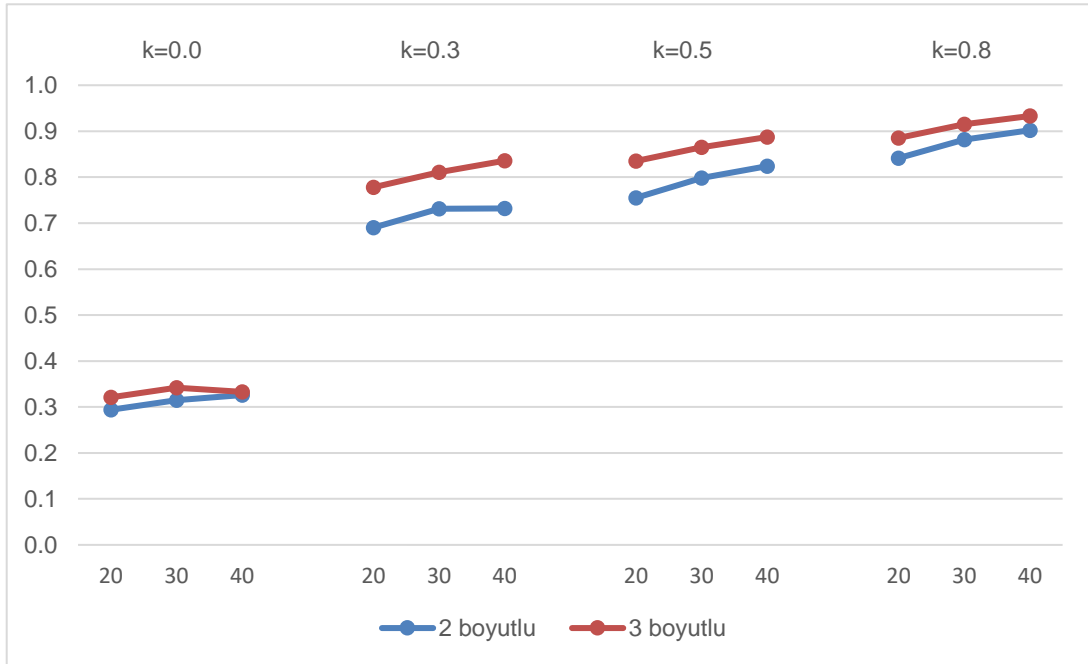
EK-6. Madde Tepki Kuramı Modellerine Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerlerine İlişkin Karşılaştırılmalı Grafikler



Şekil EK-6.1. İki Faktör Modele Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri



Şekil EK-62. Hiyerarşik ÇBMTK Modele Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri



Şekil EK-6.3. Üst Düzey Sıralı Modele Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri

EK-7. İkinci Alt Probleme Ait RMSE Değerleri

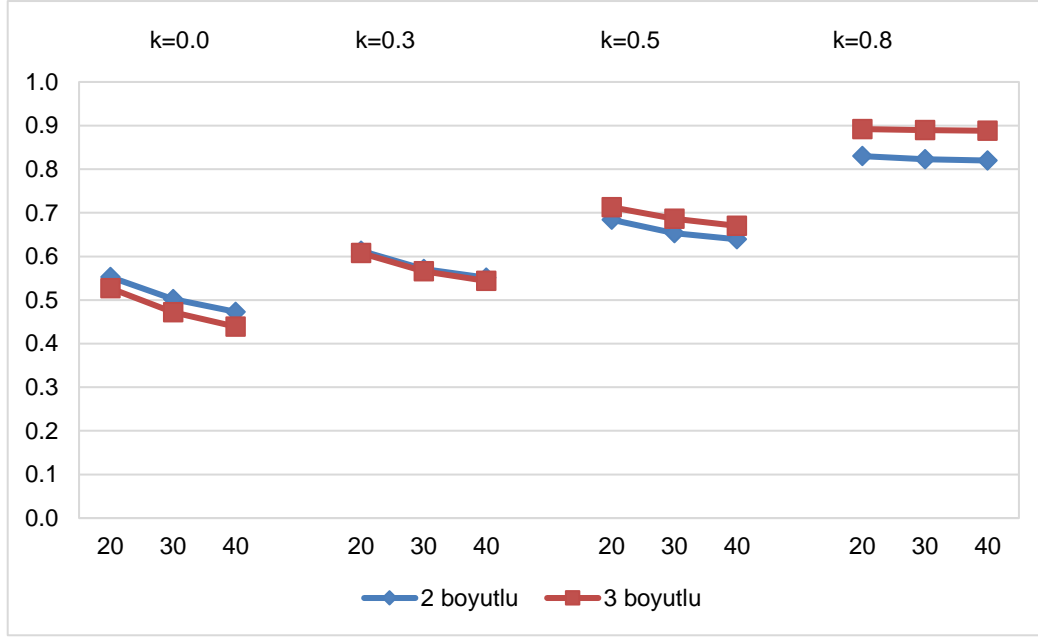
Tablo EK-7.1.: İki Boyutlu Veri Setlerinden Her Bir Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri

Alt test sayısı	Korelasyon	Alt test uzunluğu	İki-faktör Model		Hiyerarşik ÇBMTK Model		Üst Düzey Sıralı Model	
			θ_1	θ_2	θ_1	θ_2	θ_1	θ_2
2	0	20	0.635	0.471	0.498	0.456	0.501	0.459
		30	0.583	0.421	0.427	0.394	0.430	0.396
		40	0.535	0.410	0.366	0.373	0.368	0.374
	0.3	20	0.694	0.532	0.498	0.454	0.496	0.454
		30	0.647	0.495	0.427	0.393	0.426	0.392
		40	0.610	0.493	0.366	0.372	0.365	0.371
	0.5	20	0.748	0.620	0.498	0.453	0.486	0.446
		30	0.709	0.597	0.428	0.393	0.420	0.387
		40	0.681	0.598	0.366	0.372	0.360	0.366
	0.8	20	0.856	0.805	0.498	0.454	0.449	0.420
		30	0.841	0.805	0.427	0.393	0.391	0.366
		40	0.822	0.817	0.366	0.372	0.341	0.345

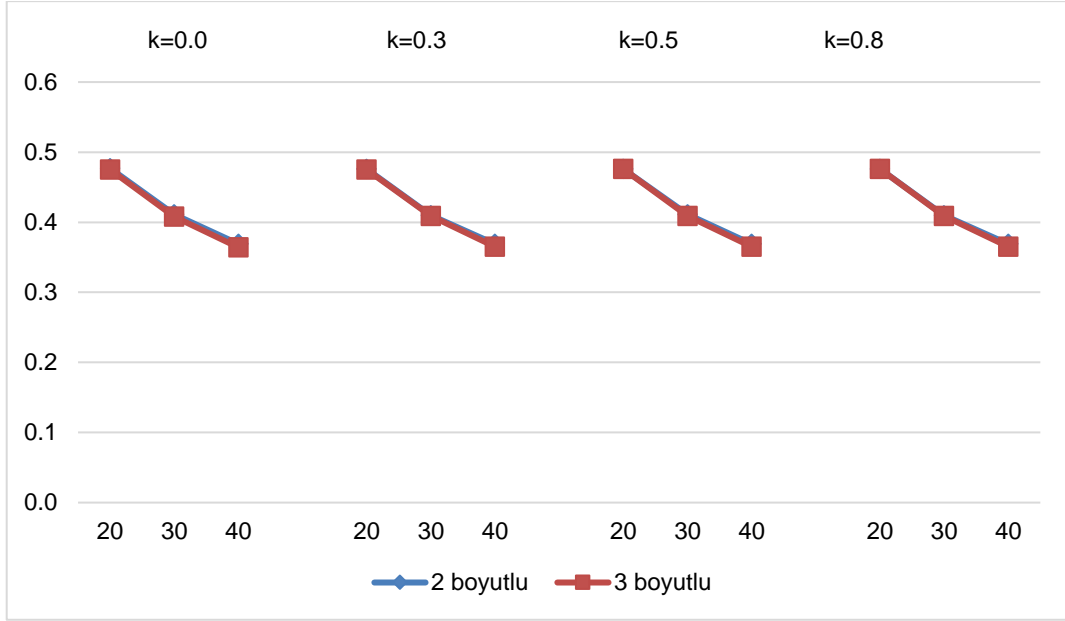
Tablo EK-7.2.: Üç Boyutlu Veri Setlerinden Her Bir Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri

Alt test sayısı	Korelasyon	Alt test uzunluğu	İki-faktör Model			Hiyerarşik ÇBMTK Model			Üst Düzey Sıralı Model		
			θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
2	0	20	0.622	0.464	0.497	0.498	0.448	0.479	0.501	0.450	0.482
		30	0.561	0.416	0.437	0.426	0.389	0.408	0.428	0.391	0.410
		40	0.510	0.406	0.400	0.364	0.370	0.359	0.365	0.371	0.360
	0.3	20	0.685	0.552	0.587	0.498	0.449	0.479	0.494	0.445	0.474
		30	0.638	0.519	0.543	0.426	0.392	0.408	0.422	0.389	0.404
		40	0.597	0.518	0.517	0.364	0.372	0.359	0.361	0.369	0.356
	0.5	20	0.754	0.676	0.708	0.498	0.450	0.479	0.478	0.434	0.461
		30	0.721	0.656	0.679	0.426	0.394	0.408	0.412	0.382	0.395
		40	0.693	0.658	0.660	0.364	0.373	0.358	0.354	0.363	0.348
	0.8	20	0.906	0.874	0.897	0.498	0.452	0.479	0.426	0.397	0.415
		30	0.901	0.876	0.893	0.426	0.393	0.408	0.373	0.351	0.360
		40	0.891	0.886	0.888	0.364	0.374	0.358	0.326	0.333	0.322

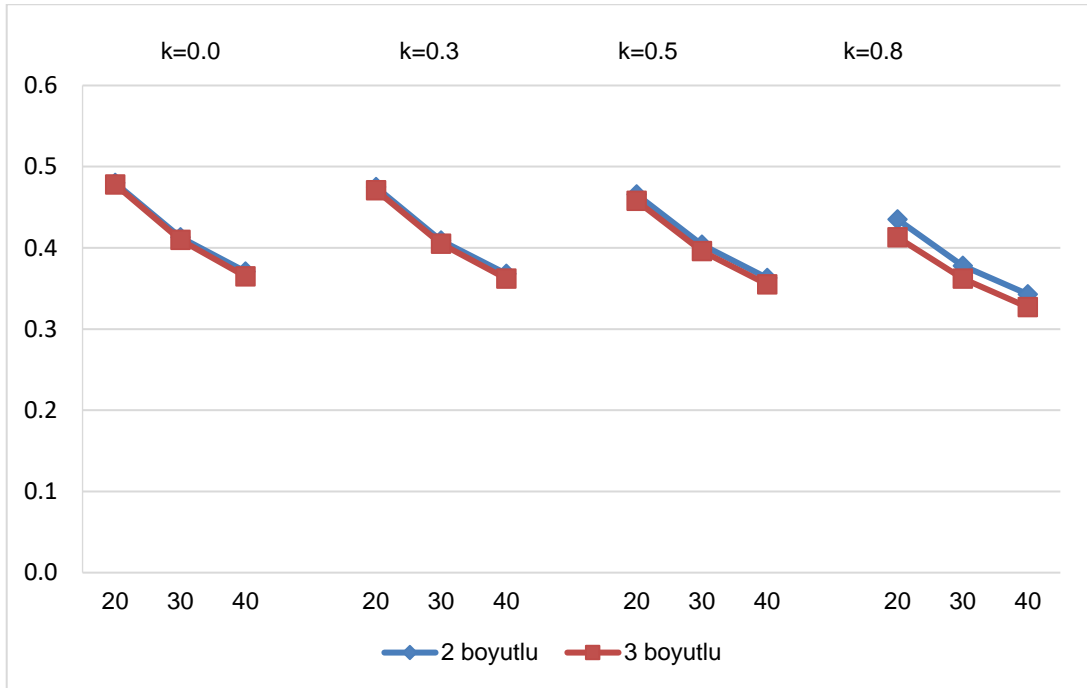
EK-8. Madde Tepki Kuramı Modellerine Göre İki ve Üç Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerlerine İlişkin Karşılaştırılmalı Grafikler



Şekil EK-8.1. İki Faktör Modeline Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri



Şekil EK-8.2. Hiyerarşik ÇBMTK Modele Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri



Şekil EK-8.3. Üst Düzey Sıralı Modele Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait RMSE Değerleri

EK-9. İkinci Alt Probleme Ait Güvenirlik Değerleri

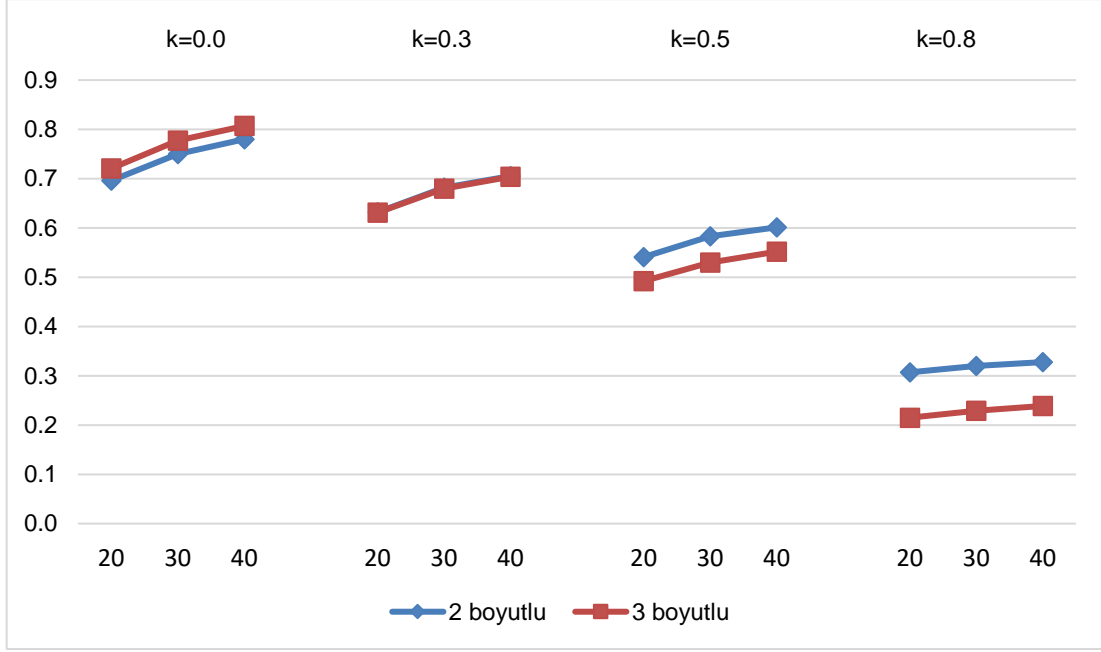
Tablo EK-9.1.: İki Boyutlu Veri Setlerinden Her Bir Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri

Alt test sayısı	Korelasyon	Alt test uzunluğu	İki-faktör Model		Hiyerarşik ÇBMTK Model		Üst Düzey Sıralı Model	
			θ_1	θ_2	θ_1	θ_2	θ_1	θ_2
2	0	20	0.613	0.779	0.751	0.793	0.747	0.791
		30	0.676	0.823	0.817	0.845	0.814	0.844
		40	0.727	0.833	0.865	0.862	0.864	0.861
	0.3	20	0.548	0.717	0.751	0.794	0.752	0.794
		30	0.608	0.756	0.817	0.846	0.818	0.847
		40	0.652	0.758	0.865	0.862	0.866	0.863
	0.5	20	0.467	0.615	0.751	0.794	0.762	0.801
		30	0.521	0.644	0.816	0.845	0.822	0.850
		40	0.559	0.643	0.865	0.862	0.870	0.866
	0.8	20	0.266	0.349	0.751	0.793	0.797	0.823
		30	0.289	0.351	0.817	0.845	0.846	0.865
		40	0.320	0.337	0.865	0.861	0.883	0.880

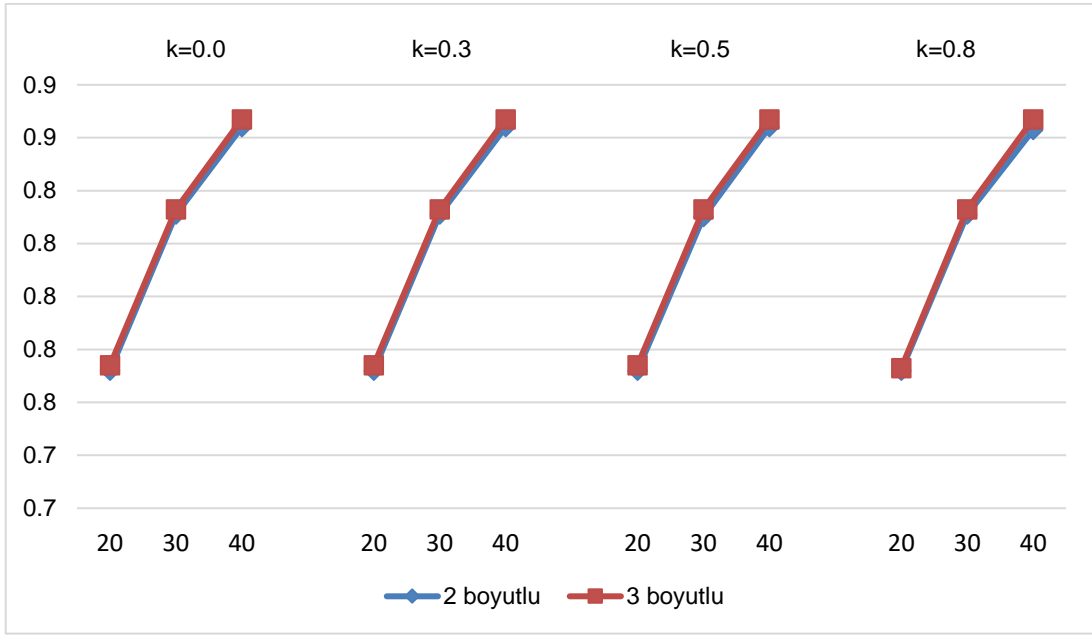
Tablo EK-9.2.: Üç Boyutlu Veri Setlerinden Her Bir Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri

Alt test sayısı	Korelasyon	Alt test uzunluğu	İki-faktör Model			Hiyerarşik ÇBMTK Model			Üst Düzey Sıralı Model		
			θ_1	θ_2	θ_3	θ_1	θ_2	θ_3	θ_1	θ_2	θ_3
2	0	20	0.624	0.780	0.757	0.752	0.795	0.775	0.748	0.793	0.772
		30	0.695	0.823	0.812	0.818	0.845	0.836	0.816	0.844	0.835
		40	0.747	0.831	0.843	0.867	0.860	0.874	0.867	0.859	0.873
	0.3	20	0.540	0.688	0.664	0.751	0.794	0.777	0.756	0.798	0.781
		30	0.601	0.725	0.713	0.818	0.843	0.838	0.821	0.845	0.841
		40	0.648	0.726	0.739	0.867	0.858	0.875	0.869	0.861	0.877
	0.5	20	0.431	0.534	0.511	0.752	0.794	0.777	0.771	0.808	0.793
		30	0.478	0.563	0.550	0.818	0.842	0.838	0.830	0.851	0.849
		40	0.519	0.562	0.576	0.867	0.858	0.875	0.874	0.866	0.882
	0.8	20	0.188	0.237	0.221	0.752	0.794	0.775	0.818	0.840	0.831
		30	0.206	0.244	0.236	0.818	0.844	0.837	0.860	0.875	0.873
		40	0.230	0.236	0.249	0.867	0.859	0.874	0.894	0.888	0.898

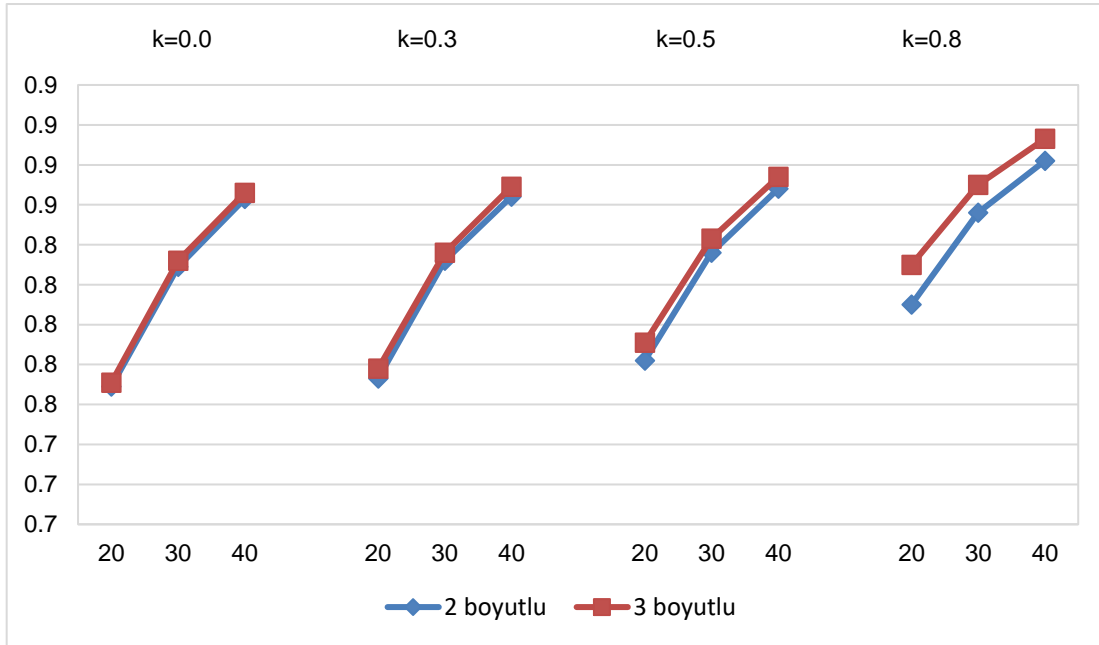
EK-10. Madde Tepki Kuramı Modellerine Göre İki ve Üç Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlilik Değerlerine İlişkin Karşılaştırılmalı Grafikler



Şekil EK-10.1. İki Faktör Modele Göre İki ve Üç Boyutlu Veri Setlerinden Alt Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlilik Değerleri



Şekil EK-10.2. Hiyerarşik ÇBMTK Modele Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri



Şekil EK-10.3. Üst Düzey Sıralı Modele Göre İki ve Üç Boyutlu Veri Setlerinden Toplam Test Puanı İçin Kestirilen Yetenek Parametrelerine Ait Güvenirlik Değerleri

EK-11. Gerçek Veri Setinde Modellerin Kestirdiği Alt Testler Arası Korelasyonların Standart Sapma Matrisi

Tablo EK-14.1: Modellerin Kestirdiği Alt Testler Arası Korelasyonların Standart Sapma Matrisi

<i>Model</i>	<i>Alt testler</i>	<i>Din Kültürü</i>	<i>Fen Bilgisi</i>	<i>İngilizce</i>	<i>Matematik</i>	<i>Tarih</i>	<i>Türkçe</i>
İki Faktör	Fen Bilgisi	0.033	1.000	0.007	0.011	0.014	0.009
	İngilizce	0.039	0.007	1	0.009	0.020	0.011
	Matematik	0.038	0.011	0.009	1	0.006	0.010
	Tarih	0.038	0.014	0.020	0.006	1	0.008
	Türkçe	0.041	0.009	0.011	0.010	0.008	1
ÇBMTK	Fen Bilgisi	0.010	1	0.009	0.009	0.008	0.007
	İngilizce	0.010	0.009	1	0.008	0.013	0.008
	Matematik	0.008	0.009	0.008	1	0.007	0.008
	Tarih	0.013	0.008	0.013	0.007	1	0.011
	Türkçe	0.009	0.007	0.008	0.008	0.011	1
Üst Düzey	Fen Bilgisi	0.010	1	0.010	0.009	0.008	0.007
	İngilizce	0.011	0.010	1	0.008	0.017	0.012
	Matematik	0.007	0.009	0.008	1	0.009	0.009
	Tarih	0.016	0.008	0.017	0.009	1	0.014
	Türkçe	0.008	0.007	0.012	0.009	0.014	1

EK-12. ORJİNALLİK RAPORU



HACETTEPE ÜNİVERSİTESİ EĞİTİM BİLİMLERİ ENSTİTÜSÜ DOKTORA TEZ ÇALIŞMASI ORJİNALLİK RAPORU

HACETTEPE ÜNİVERSİTESİ EĞİTİM BİLİMLER ENSTİTÜSÜ EĞİTİM BİLİMLERİ ANA BİLİM BAŞKANLIĞI'NA

Tarih: 31/01/2018

Tez Başlığı : Toplam Test Puanı ve Alt Test Puanlarının Kestiriminin Hiyerarşik Madde Tepki Kuramı Modelleri ile Karşılaştırılması

Yukarıda başlığı verilen tez çalışmamın tamamı (kapak sayfası, özetler, ana bölümler, kaynakça) aşağıdaki filtreler kullanılarak Turnitin adlı intihal programı aracılığı ile kontrol edilmiştir. Kontrol sonucunda aşağıdaki veriler elde edilmiştir.

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Endeksi	Gönderim Numarası
31/01/2018	82	154,201	11/12/2017	%3	909184757

Uygulanan filtreler:

- 1- Kaynakça hariç
- 2- Alıntılar dâhil
- 3- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı inceledim ve çalışmamın herhangi bir intihal içermediğini; aksinin tespit edileceği muhtemel durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.


31/01/2018

Adı Soyadı: Sümeyra Soysal
Öğrenci No: H11162841
Anabilim Dalı: Eğitim Bilimleri
Programı: Eğitimde Ölçme ve Değerlendirme
Statüsü: Y.Lisans Doktora Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.
Prof. Dr. Hülya KELECİOĞLU



HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF EDUCATIONAL SCIENCES
THESIS/DISSERTATION ORIGINALITY REPORT

HACETTEPE UNIVERSITY
GRADUATE SCHOOL OF EDUCATIONAL SCIENCES
TO THE DEPARTMENT OF EDUCATIONAL SCIENCE

Date 31/01/2018

Thesis Title : Comparison of Estimation of Total Score and Subscores with Hierarchical Item Response Theory Models

The whole thesis that includes the *title page, introduction, main chapters, conclusions and bibliography section* is checked by using **Turnitin** plagiarism detection software take into the consideration requested filtering options. According to the originality report obtained data are as below.

Time Submitted	Page Count	Character Count	Date of Thesis Defence	Similarity Index	Submission ID
31/01/2018	82	154,201	11/12 /2017	%3	909184757

Filtering options applied:

1. Bibliography excluded
2. Quotes included
3. Match size up to 5 words excluded

I declare that I have carefully read Hacettepe University Graduate School of Educational Sciences Guidelines for Obtaining and Using Thesis Originality Reports; that according to the maximum similarity index values specified in the Guidelines, my thesis does not include any form of plagiarism; that in any future detection of possible infringement of the regulations I accept all legal responsibility; and that all the information I have provided is correct to the best of my knowledge.

I respectfully submit this for approval.


31/01/2018

Name Surname: Sümeyra SOYSAL

Student No: H11162841

Department: Educational Science

Program: Educational Measurmenet and Evalutaion

Status: Masters Ph.D. Integrated Ph.D.

ADVISOR APPROVAL


APPROVED
Prof. Dr. Hülya KELECIOĞLU

EK-13. ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı	Sümevra Soysal
Doğum Yeri	Sivas
Doğum Tarihi	1985

Eğitim Durumu

Lise	Cumhuriyet Anadolu Lisesi /SİVAS	2003
Lisans	Erciyes Üniversitesi Fen Bilgisi Öğretmenliği	2009
Yüksek Lisans	-	-
Yabancı Dil	İngilizce: Okuma (iyi), Yazma (İyi), Konuşma (Orta)	

İş Deneyimi

Stajlar		Buraya tarih aralığı yazılacak
Projeler		
Çalıştığı Kurumlar	Sinop Üniversitesi	2011-2018

Akademik Çalışmalar

Yayınlar (Ulusal, uluslararası makale, bildiri, poster vb gibi.)

Gündeğer, C., Soysal, S., Yağcı, E., (2012). ÖYP Araştırma Görevlilerinin Yurt İçi ve Yurt Dışı Dil Eğitimlerine İlişkin Görüşleri, <i>Pegem Eğitim ve Öğretim Dergisi</i> , 2(4), s.55-70
Impact Of Missing Data in Rasch Model Estimation, (2016). <i>The Turkish Online Journal of Educational Technology Special Issue for INTE 2016</i> , s.1166-1177
Soysal, S., Karaman, H., Doğan, N., (2016). Örneklem Büyüklüğünün ve Kayıp Veri Oranının Genellenebilirlik Katsayısına Etkisi. Paper presented in III. International Eurasian Educational Research Congress, 31 May - 03 June 2016, Muğla Sıtkı Koçman Üniversitesi
Arıkan Akın, Ç., Soysal, S. (2016). Güvenirlik Katsayılarının Kayıp Veri Atama Yöntemlerine Göre İncelenmesi, V. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi (Uluslararası Katılımlı), 1-3 Eylül 2016, Akdeniz Üniversitesi.
Doğan, N., Soysal, S. Karaman, H. (2017). Aynı Örneklem Açıklayıcı ve Doğrulayıcı Faktör Analizi Uygulanabilir Mi? 26. Uluslararası Eğitim Bilimleri Kongresinde bildiri olarak sunulmuştur. 20-23 Nisan 2017, Antalya.
Soysal, S., Arıkan Akın, Ç., (2017). Kayıp Veri Atama Yöntemlerinin Faktörleştirme Yöntemleri Üzerindeki Etkisi, 26. Uluslararası Eğitim Bilimleri Kongresinde bildiri olarak sunulmuştur. 20-23 Nisan 2017, Antalya.

Soysal, S. Arıkan Akın, Ç., (2017). Examining The Cross-Cultural Applicability via Generalizability Theory, Paper presented at International Meeting of the Psychometric Society, 18-21 July, 2017, Zürich, Switzerland

Seminer ve Çalıştaylar

Genellenebilirlik Kuramı Çalıştayı (01.04.2013/Gazi Üniversitesi)
Eğitim Bilimlerinde Araştırma Yöntem Sorunu-HUTF TEBAD Eğitim Toplantıları 17 (22.04.2013/Gazi Üniversitesi)
Cognitive Diagnosis Modeling: A General Framework Approach (26.06.2014/Hacettepe Üniversitesi)
An Introduction to Diagnostic Classification Modeling (15.04.2015 ,Chicago, Amerika)
Fundamentals of Item Response Theory and Computerized Adaptive Testing (16.04.2015, Amerika)
Using IRT for standard setting in performance based assessments (16.04.2015, Amerika)
Hierarchical Linear Modeling with Large-Scale International Databases (19.04.2015, Amerika)

Sertifikalar

--

İletişim

e-Posta Adresi	sumeyrasoysal@hotmail.com
Jüri Tarihi	11/12/2017

