

**MADDE TAKIMLARI İÇEREN TESTLERDE FARKLI
MODELLERDEN ELDE EDİLEN MADDE VE YETENEK
PARAMETRELERİNİN KARŞILAŞTIRILMASI**

**COMPARISON OF ITEM AND ABILITY PARAMETERS
OBTAINED FROM DIFFERENT MODELS ON TESTS
COMPOSED OF TESTLETS**

Esin YILMAZ KOĞAR

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı İçin
Öngördüğü

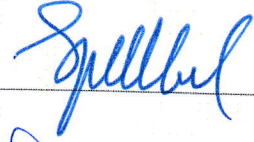
Doktora Tezi olarak hazırlanmıştır.

2016

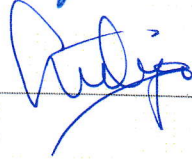
Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼'ne,

Esin YILMAZ KOđAR'ın hazırladıđı "Madde Takımları İeren Testlerde Farklı Modellerden Elde Edilen Madde ve Yetenek Parametrelerinin Karşılaştırılması" bařlıklı bu alıřma j¼rimiz tarafından **Eđitim Bilimleri Anabilim Dalı, Eđitimde Ölme ve Deđerlendirme Bilim Dalı'nda Doktora Tezi** olarak kabul edilmiřtir.

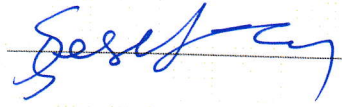
Bařkan Prof. Dr., Selahattin GELBAL



¼ye (Danıřman) Prof. Dr., H¼lya KELECİOđLU



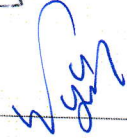
¼ye Do. Dr., řeref TAN



¼ye Do. Dr., Burcu ATAR



¼ye Do. Dr., Neře G¼LER



ONAY

Bu tez Hacettepe ¼niversitesi Lisans¼st¼ Eđitim-¼đretim ve Sınav Y¼netmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri ¼yeleri tarafından / / tarihinde uygun g¼r¼lm¼ř ve Enstit¼ Y¼netim Kurulunca / / tarihinde kabul edilmiřtir.

Prof. Dr. Berrin AKMAN
Eđitim Bilimleri Enstit¼s¼ M¼d¼r¼

MADDE TAKIMI İÇEREN TESTLERDE FARKLI MODELLERDEN ELDE EDİLEN MADDE VE YETENEK PARAMETRELERİNİN KARŞILAŞTIRILMASI

Esin YILMAZ KOĞAR

ÖZ

Bu araştırmanın amacı, madde takımları içeren testlerde madde takımı sayısı, bağımsız madde sayısı ve örneklem büyüklüğü değiştiğinde, tek boyutlu madde tepki kuramı, iki faktör ve madde takımı tepki kuramı modellerinden elde edilen madde ve yetenek parametreleri ile bu parametrelere ait standart hata değerlerini belirlemek ve elde edilen sonuçları karşılaştırmaktır. Bu amaç doğrultusunda, PISA 2012 uygulamasındaki matematik alanına ait 10. kitapçık veri toplama aracı olarak seçilmiş ve bu kitapçıkta yer alan 36 maddeden elde edilen veriler kullanılmıştır. Araştırmada bu 36 madde kullanılarak farklı sayıda madde takımı ve bağımsız madde içeren altı farklı veri seti oluşturulmuştur. Daha sonra bu oluşturulan veri setlerinden 250, 500 ve 1000 kişilik üç farklı örneklem büyüklüğü rastgele olarak seçilmiştir. Altı farklı veri seti ve üç farklı örneklem büyüklüğünden oluşan toplam 18 koşul tek boyutlu madde tepki kuramı, iki faktör ve madde takımı tepki kuramı modelleri kullanılarak analiz edilmiştir.

Verilerin analizinde IRTPRO 2.1 programı kullanılmıştır. Veri setinin yerel bağımlılık açısından test edilmesinde ise bu program aracılığıyla hesaplanan yerel bağımlılık X^2 (LD X^2) istatistiği kullanılmıştır. Daha sonra madde ve yetenek parametreleri ile bu parametrelere ait hata değerleri kestirilmiş, bütün kestirimlerin ortalaması alınarak her bir veri seti ve örneklem büyüklüğü için rapor edilmiştir. Model-veri uyumunu incelemek için ise -2LL değerleri hesaplanmış ve bu değerler her koşul için incelenmiştir. Ayrıca her bir modelden elde edilen madde ve yetenek parametrelerine ait değerlerin korelasyon katsayıları hesaplanmıştır.

Araştırmanın bulguları incelendiğinde farklı örneklem büyüklüklerinde, farklı modellerden ve farklı veri setlerinden elde edilen madde parametreleri için genel olarak en düşük ortalama hata değerlerinin tek boyutlu madde tepki kuramından (TBMTK) elde edildiği; madde takımı tepki kuramının (MTTK), iki faktör modelinden (2FM) daha düşük ortalama hata kestirimi yaptığı belirlenmiştir. Veri setindeki bağımsız madde sayısının artmasıyla parametrelere ait ortalama hata değerlerinde

daha çok azalma eğilimi olduğu görülmüştür. Veri setine madde takımı eklenmesiyle genel olarak parametrelere ait hata ortalamaları artarken, en az değişkenlik gösteren parametrenin kesişim parametresi olduğu belirlenmiştir. Ancak veri setindeki madde takımı sayısının artmasının, madde parametrelerine ait hata ortalamalarına etkisi üzerine çok net bir sonuca erişilmemiştir. Örneklem büyüklüğü arttıkça farklı modellerden elde edilen yetenek parametresinin ve hata ortalamasının çok değişmediği, bireyin yetenek kestirimlerinin örneklem büyüklüğüne pek bağlı olmadığı sonucuna varılmıştır. Bütün koşullarda yerel bağımlılığı göz önüne alan modellerin TBMTK'den daha iyi model-veri uyumu sağladığı, genel olarak 2FM ile MTTK arasında anlamlı farklılık olmadığı; her iki modelin de bu veri setleri için kullanılabileceği belirlenmiştir. Bu iki model arasında anlamlı farklılık olduğu durumda ise 2FM'nin daha iyi sonuç verdiği söylenebilir. Ayrıca her bir örneklem büyüklüğünde ve veri setinde madde ve yetenek parametreleri ile bu parametrelerinin hatalarının korelasyonlarının genel olarak yüksek olduğu belirlenmiştir.

Anahtar sözcükler: Madde tepki kuramı, madde takımı, madde takımı tepki kuramı, iki faktör modeli, yerel bağımlılık

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

COMPARISON OF ITEM AND ABILITY PARAMETERS OBTAINED FROM DIFFERENT MODELS ON TESTS COMPOSED OF TESTLETS

Esin YILMAZ KOĞAR

ABSTRACT

The purpose of this research is to estimate the item and ability parameters and the standard error values related to those parameters, obtained from unidimensional item response theory (UIRT), bifactor (BIF) and testlet response theory models (TRT) in the tests including testlets, when the number of testlets, number of independent items, and sample size change, and to compare the obtained results. In accordance with this purpose, the 10th booklet applied in the field of mathematics in PISA 2012, has been selected as a data collection tool, and data obtained from 36 items in that booklet have been used. In the research, by using those 36 items, six different data set containing different numbers of testlets and independent items have been constituted. After that, from out of those constituted data sets, three different sample sizes of 250, 500 and 1000 persons have been selected randomly. Totally 18 different conditions which are composed of six different data sets and three different sample sizes, have been analyzed using unidimensional item response theory, bifactor and testlet response theory models.

IRTPRO 2.1 program has been used in analyses of the data. In testing of the data set with regard to local dependency, the local dependency statistics of X^2 (LD X^2), which is calculated by means of this program, has been used. Then the item and ability parameters and standard error values belonging to those parameters have been estimated, all the estimations have been averaged, and reported for each data set and sample size. In order to examine the model - data compatibility, -2LL values have been calculated, and those values have been examined for each condition. In addition, correlation coefficients of the values belonging to the item and ability parameters obtained from each model have been calculated.

When the findings of the research have been investigated, it was determined that, for the item parameters estimated under different samples sizes from different models and different data sets, generally the lowest mean error values were obtained from UIRT; and TRT has made a mean of error estimation lower than BIF. With the increase in the number of independent items in the data set, it has been

seen that there was a downward tendency in the mean of error values belonging to the parameters. With the addition of the testlets into the data set, while the mean of error values belonging to the parameters have increased generally, it was determined that the least differing parameter was the intersection parameter. But no clear tendency could have been reached in respect of the effect of the increase in the number of testlets in the data set on the error means. It has been concluded that, the sample size increases, the ability parameter and error mean obtained from different models does not change too much, and ability estimations of the individual do not depend too much on the sample size. It has been determined that, under all conditions, models which take into consideration the local dependency have provided better model - data compatibility than UIRT, generally there is no meaningful difference between BIF and TRT; and both models can be used for those data sets. It can be said that, when there is a meaningful difference between those two models, generally BIF gives a better result. In addition, it has been determined that, in each sample size and data set, item and ability parameters and correlations of errors of the parameters are generally high.

Keywords: Item response theory, testlet, testlet response theory, bi-factor model, local dependency

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

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.



Esin YILMAZ KOĞAR

TEŞEKKÜR

İşin en zor kısmıydı belki de tez yazma süreci. Bu nedenle bu süreçte benimle olanlara ne yazsam ne söylesem yetersiz ama ben yine de kendilerine olan minnetlerimi kısaca dile getirmek istiyorum. Doktora ders aşamasında ve tezimin oluşmasında bana rehberlik eden ve her zaman destekçim olan kendisinden çok şey öğrendiğim sayın danışmanım Prof. Dr. Hülya Kelecioğlu'na öncelikli olarak teşekkür ederim.

Her daim güler yüzüyle, yardımcı kişiliğiyle ve birçok konudaki rehberliğiyle yanımda olan değerli hocam Prof. Dr. Selahattin Gelbal'a; aradaki uzaklığa karşın her zaman samimiyetini hissettiğim, kendime akademik ve yaşam tarzı olarak örnek aldığım, değerli görüş ve önerileri ile tezimi güçlendirmemi sağlayan sayın hocam Doç. Dr. Burcu Atar'a; tez izleme ve tez savunma komitelerimde bulunan ve görüşlerini esirgemeyip tez sürecinde bana yardımcı olan sayın hocam Doç. Dr. Şeref Tan'a ve tez savunma jürimde bulunup, tezime katkı sağlayan sayın hocam Doç. Dr. Neşe Güler'e çok teşekkür ederim. Çalışmamda kullandığım programda takıldığım çoğu noktada kendisine soru sorduğum ve her soruma cevap aldığım sayın Dr. Stephen du Toit'e çok teşekkür ederim. Ayrıca doktora öğrenimim süresince bana yurt içi doktora bursu ile destek olan TÜBİTAK'a teşekkür ederim.

Yıllarca süren eğitim-öğretim hayatımda beni her zaman destekleyen, sürekli yanımda olan, benden çok akademik hayattaki değişimleri ve yenilikleri takip edip beni de haberdar kılan canım babacığım Abdullah Yılmaz'a, canım anneciğim Fatma Yılmaz'a ve biricik canım ablacığım Esen Yılmaz'a; kardeşten öte akıl danışmanım ve arkadaşım olan Gökhan Yılmaz'a, kısacası benim güzel çekirdek aileme bütün kalbimle teşekkür ederim.

Ayrı yollardan gelip, ölçme ve değerlendirme camiasında yollarımızın kesiştiği ve sonra ise birleştiği hayat arkadaşım, eşim, canım Hakan Koğar'a benim bütün çileden çıkmalarıma katlanıp bütün sorularıma ve sorunlarıma çözüm aradığı, hep benimle olduğu ve olacağı için sonsuz teşekkür ederim. Daha birçok aşamaları birlikte ve daha güçlü bir şekilde aşmak dileğiyle...

İÇİNDEKİLER

ÖZ.....	iii
ETİK BEYANNAMESİ	vii
TEŞEKKÜR.....	viii
İÇİNDEKİLER.....	ix
TABLolar DİZİNİ	xi
ŞEKİLLER DİZİNİ.....	xiii
SİMGELER VE KISALTMALAR DİZİNİ	xiv
1. GİRİŞ.....	1
1.1. Problem Durumu	1
1.2. Araştırmanın Amacı ve Önemi:	3
1.3. Problem Cümlesi:.....	5
1.3.1. Alt Problemler:	5
1.4. Sayıtlar:	6
1.5. Sınırlılıklar:	6
1.6. Tanımlar:.....	6
1.7. Araştırmanın Kuramsal Temelleri:.....	6
A. Madde Tepki Kuramı.....	7
B. Madde Takımı Kavramı.....	13
C. Yerel Madde Bağımlılığı	14
D. Madde Takımlarına Dayalı Testlerde Kullanılan Modeller	21
2. İLGİLİ ARAŞTIRMALAR.....	33
2.1. Yerel Madde Bağımlılığı ile İlgili Çalışmalar	33
2.2. Madde Takımları İçeren Testlerin Farklı Madde Tepki Kuramı Modelleri ile Ele Alındığı Çalışmalar.....	36
2.3. Madde Takımı Tepki Kuramı ile İlgili Çalışmalar	38
2.4. İki Faktör Modeli ile İlgili Çalışmalar	47
2.5. İlgili Araştırmalar Özet.....	51
3. YÖNTEM.....	52
3.1. Araştırmanın Modeli	52
3.2. Evren ve Örneklem	52
3.3. Veri Toplama Aracı	53
3.4. Verilerin Analizi	55
4. BULGULAR VE TARTIŞMA	60
4.1. Madde Parametrelerinin Değerlendirilmesine İlişkin Bulgular	60
4.2. Yetenek Parametrelerinin Değerlendirilmesine İlişkin Bulgular	70
4.3. Farklı Modellerden Elde Edilen Uyum İndekslerine İlişkin Bulgular.....	73
4.4. Farklı Modellerden Elde Edilen Madde Parametrelerinin ve Madde Parametrelerine ait Standart Hataların Korelasyonlarına İlişkin Bulgular	75
4.5. Farklı Modellerden Elde Edilen Yetenek Parametrelerinin ve Yetenek Parametrelerine ait Standart Hataların Korelasyonlarına İlişkin Bulgular	78
4.6. Tartışma.....	79
4.6.1. Madde Parametrelerine İlişkin Tartışma	79

4.6.2. Yetenek Parametrelerine İlişkin Tartışma	81
5. SONUÇ VE ÖNERİLER	84
5.1. Sonuçlar	84
5.2. Öneriler	86
5.2.1. Araştırma Sonucuna Dönük Öneriler	87
5.2.2. Gelecek Araştırmalara Dönük Öneriler	87
KAYNAKÇA.....	89
EKLER DİZİNİ.....	97
EK 1. ETİK KURUL İZİN MUAFİYET FORMU	98
EK 2. İki Kategorili Puanlanan Maddeler için Farklı Modellerden Elde Edilen Madde Parametreleri ve Madde Parametrelerine ait Standart Hata Değerleri ...	99
EK 3. Çok Kategorili Puanlanan Maddeler için Farklı Modellerden Elde Edilen Madde Parametreleri ve Madde Parametrelerine ait Standart Hata Değerleri .	106
EK 4. ORJİNALLİK RAPORU	108
ÖZGEÇMİŞ.....	109

TABLolar DİZİNİ

Tablo 3.1: Araştırma için Oluşturulan Veri Setlerine İlişkin Özellikler	54
Tablo 3.2: Veri Setlerine göre Madde Takımlarında Yer Alan Maddelerin LD X2 Değerleri.....	56
Tablo 4.1: Veri Seti 1’de, 2’de ve 3’te Yer Alan İki Kategorili Puanlanan Maddelerin Madde Parametre Ortalamaları ile Bu Parametrelere ait Hata Ortalamaları	60
Tablo 4.2: Veri Seti 4’te, 5’te ve 6’da Yer Alan İki Kategorili Puanlanan Maddelerin Madde Parametre Ortalamaları ile Bu Parametrelere ait Hata Ortalamaları	63
Tablo 4.3: Farklı Veri Setlerinde Yer Alan Çok Kategorili Puanlanan Maddelerin Madde Parametrelerinin ve Bu Parametrelere ait Standart Hatalarının Ortalamaları	67
Tablo 4.4: Veri Seti 1, 2 ve 3 için Yetenek Parametresinin ve Bu Parametreye ait Hata Değerlerinin Ortalaması	70
Tablo 4.5: Veri Seti 4, 5 ve 6 için Yetenek Parametresinin ve Bu Parametreye ait Hata Değerlerinin Ortalaması	72
Tablo 4.6: Uyum İyiliği İndeks Değerleri	74
Tablo 4.7: Farklı Modellerden Elde Edilen Madde Parametrelerine ve Madde Parametrelerinin Standart Hata Değerlerine ait Korelasyon Değerleri.....	76
Tablo 4.8: Farklı Modellerden Elde Edilen Yetenek Parametrelerine ve Yetenek Parametrelerinin Hata Değerlerine ait Korelasyon Değerleri.....	78
Tablo 4.9: Veri Seti 1’de Yer Alan İki Kategorili Puanlanan Maddeler için Farklı Modellerden Elde Edilen Madde Parametreleri ve Madde Parametrelerine ait Standart Hata Değerleri	99
Tablo 4.10: Veri Seti 2’de Yer Alan İki Kategorili Puanlanan Maddeler için Farklı Modellerden Elde Edilen Madde Parametreleri ve Madde Parametrelerine ait Standart Hata Değerleri	100
Tablo 4.11: Veri Seti 3’te Yer Alan İki Kategorili Puanlanan Maddeler için Farklı Modellerden Elde Edilen Madde Parametreleri ve Madde Parametrelerine ait Standart Hata Değerleri	101
Tablo 4.12: Veri Seti 4’te Yer Alan İki Kategorili Puanlanan Maddeler için Farklı Modellerden Elde Edilen Madde Parametreleri ve Madde Parametrelerine ait Standart Hata Değerleri	102
Tablo 4.13: Veri Seti 5’te Yer Alan İki Kategorili Puanlanan Maddeler için Farklı Modellerden Elde Edilen Madde Parametreleri ve Madde Parametrelerine ait Standart Hata Değerleri	103
Tablo 4.14: Veri Seti 6’da Yer Alan İki Kategorili Puanlanan Maddeler için Farklı Modellerden Elde Edilen Madde Parametreleri ve Madde Parametrelerine ait Standart Hata Değerleri	104

Tablo 4.14: Devamı	105
Tablo 4.15: Çok Kategorili Puanlanan Maddeler için Farklı Modellerden Elde Edilen Madde Parametre Değerleri	106
Tablo 4.16: Çok Kategorili Puanlanan Maddeler için Farklı Modellerden Elde Edilen Madde Parametrine ait Standart Hata Değerleri.....	107



ŞEKİLLER DİZİNİ

Şekil 1.1. İki Faktör Örüntüsü	24
Şekil 1.2. İki Faktör Modeli	25



SİMGELER VE KISALTMALAR DİZİNİ

KTK: Klasik Test Kuramı

MTK: Madde Tepki Kuramı

TBMTK: Tek Boytulu Madde Tepki Kuramı

ÇBMTK: Çok Boytulu Madde Tepki Kuramı

MT: Madde Takımı

BM: Bağımsız Madde

MKE: Madde Karakteristik Eğrisi

1PLM: 1 Parametrelı Lojistik Model

2PLM: 2 Parametrelı Lojistik Model

3PLM: 3 Parametrelı Lojistik Model

MTTK: Madde Takımı Tepki Kuramı

2FM: İki Faktör Modeli

YMB: Yerel Madde Bağımlılığı

ATM: Aşamalı Tepki Modeli

M-ATM: Değiştirilmiş Aşamalı Tepki Modeli

KPM: Kısmi Puan Modeli

G-KPM: Genelleştirilmiş Kısmi Puan Modeli

DÖM: Dereceleme Ölçeği Modeli

STM: Sınıflamalı Tepki Modeli

1. GİRİŞ

Bu bölümde araştırmanın problem durumuna, amaç ve önemine, problem cümlesine, sayıltılarına, sınırlılıklarına, tanımlarına ve kuramsal temellerine yer verilmiştir.

1.1. Problem Durumu

Madde Tepki Kuramı (MTK) geleneksel gerçek puan modeline alternatif olarak sunulan etkili bir kuramdır (Thissen & Orlando, 2000) ve güçlü matematiksel alt yapısı nedeniyle geniş ölçekli testlerin geliştirilmesinde ve değerlendirilmesinde sıkça kullanılmaktadır. Ayrıca MTK, maddeler ve bireyler hakkında önemli bilgilere ulaşılmasını sağlaması açısından da eğitim çalışmalarında etkili bir kuram olarak uygulanmaktadır. MTK'nin yaygın olarak kullanılması, uygun olmayan koşullara da MTK'nin uygulanmasına ve böylece yanlış ya da eksik sonuçlara ulaşılmasına neden olabilir. Çünkü MTK'nin karşılanması zor, ancak güçlü varsayımları bulunmaktadır ve uygun MTK modelinin belirlenmesi açısından öncelikle eldeki veri setinin MTK varsayımlarını karşılayıp karşılamadığı değerlendirilmelidir.

MTK'nin birçok varsayımı bulunmasına karşın bu çalışmada yerel bağımsızlık varsayımına ve onun karakteristiği üzerine odaklanılacaktır. Yerel bağımsızlık, belli bir yetenek düzeyinde, herhangi bir ölçme aracında yer alan maddelerin doğru cevaplanma olasılıklarının yalnızca bireyin yeteneğine bağlı olmasıdır (Embretson & Reise, 2000; Hambleton & Swaminathan, 1985). Bu varsayım madde çiftleri arasındaki koşullu kovaryansın yaklaşık olarak sıfır olduğunu kabul etmektedir (Dresher, 2002). Ancak aynı uyarıcıya bağlı maddelerden oluşan maddeler kümesinde bu durum bozulabilmektedir. Aralarında yerel bağımlılık bulunan maddelere alan yazında, "süper maddeler (superitems)" (Cureton, 1965), "madde takımları (testlets)" (Wainer & Kiely, 1987), "madde desteleri (item bundles)" (Rosenbaum, 1988), "kısa testler (small tests)" (Wainer & Lewis, 1990), "içeriği-bağımlı maddeler kümesi (context-dependent item sets)" (Haladyna, 1992; Keller, Swaminathan & Sireci, 2003) gibi isimler verilmiştir. Maddeler arasında yerel bağımlılık olduğuna işaret eden bu kavramlar arasında madde takımı, analiz ve puanlama birimi olarak da ele alındığı için dikkat çeken bir kavram olmuştur. Bu terim, Wainer ve Kiely (1987) tarafından tanımlanmıştır ve bir test içindeki tek bir

uyarıcıya dayanan maddelerin oluşturduğu grup olarak ifade edilmiştir. Bir başka ifadeyle madde takımı, genel bir metni, senaryoyu ya da başka bir içeriği paylaşan maddeler kümesidir (DeMars, 2006). Örneğin, genel bir okuma parçasına dayalı bir seri maddeden oluşan bir madde takımı düşünüldüğünde bu maddelere verilen tepkiler birbirinden bağımsız değildir. Çünkü bireyin okuma parçasını anlama düzeyi bireyin parçada sunulan içerikle ilgili bilgisinden ve/veya okuma parçasında verilen içeriğin ilgi alanı olmasından etkilenebilir (Yen, 1993). Bu durumda maddeler ortak bir uyarıcıya sahip oldukları için madde takımında yer alan maddelere, birbirlerinden bağımsız olarak tepki verildiği söylenemez (Fukuhara & Kamata, 2011).

Yapılan araştırmalar, testlerde madde takımı olarak adlandırılan madde gruplarının kullanıldığı durumda, MTK'nin temel varsayımlarından biri olan yerel bağımsızlık varsayımının ihlal edildiğini göstermektedir (Lee, Dunbar & Frisbie, 2001; Lee & Frisbie, 1999; Rosenbaum, 1988; Sireci, Thissen & Wainer, 1991; Wainer & Lewis, 1990; Yen, 1993). Yerel bağımsızlık varsayımının ihlal edilmesi güvenilirliğin ya da bilginin yüksek; yetenek kestiriminin standart hatasının düşük tahmin edilmesine yol açmakta ve böylece madde parametre kestirimlerinde hatalara neden olmaktadır (Ackerman, 1987; Bradlow, Wainer & Wang, 1999; Ip, 2000; Marais & Andrich, 2008; Sireci vd., 1991; Wainer, 1995; Wainer & Wang, 2000; Yen, 1993). Aynı zamanda madde etkisini göz ardı etmek, eşitleme/ölçekleme çalışmalarında da hatalara (Bishop & Omar, 2002; Lee, Kolen, Frisbie & Ankenmann, 2001; Y. Li, Bolt & Fu, 2005) ve madde uyumsuzluğuna (Marais & Andrich, 2008) neden olabilmektedir. Bunların yanı sıra DeMars (2006), madde takımlarındaki maddelerin bağımlılıkları göz ardı edildiğinde yetenekler için elde edilen ortalama hataların karekökünün (RMSE) büyük hesaplandığını, madde takımı etkisini ele alan modeller kullanıldığında ise bu değerlerin daha küçük hesaplandığını belirtmiştir. Bu nedenlerden dolayı madde takımları arasındaki bağımlılığı göz önüne alan modellerin kullanılması güvenilirlik, yetenek ve madde parametreleri kestirimlerinde ve hata hesaplamalarında daha doğru sonuçlara ulaşılmasını sağlamaktadır.

Bu araştırmada da madde takımları arasındaki yerel bağımlılığı göz önüne almak için geliştirilen MTK'ye dayalı modellerden bazıları incelenmiş, farklı modellerden elde edilen madde ve yetenek parametreleri karşılaştırılmıştır. Ayrıca çalışmada, madde takımlarında kullanılan kuramlar hakkında bilgilere yer verilmesi de amaçlanmıştır.

1.2. Araştırmanın Amacı ve Önemi:

Bu araştırmanın amacı, madde takımı içeren testlerde; madde takımı sayısı, bağımsız madde sayısı ve örneklem büyüklüğü değiştiğinde, tek boyutlu madde tepki kuramından, iki faktör kuramından ve madde takımı tepki kuramından elde edilen madde parametreleri ve bu parametrelere ait standart hata değerleri ile yetenek parametreleri ve bu parametrelere ait standart hata değerlerini kestirmek ve elde edilen sonuçları karşılaştırmaktır.

Bu araştırma alan için çeşitli yönlerden önemli görülmektedir. Çalışmanın önemli görülmesinin ilk nedeni, son yıllarda madde takımlarının sınavlarda kullanımının artmasından kaynaklanmaktadır. Madde takımları, birbiriyle ilişkili sorular sorarak sınav süresini daha etkili hale getirmek için standartlaştırılmış testlerde sıkça kullanılmaktadır (Wainer, Bradlow & Du, 2000). Birbirine bağımlı içeriğe sahip maddelerin genellikle gerçek hayata daha uygun olduğu ve bu tür maddelerin üst düzey becerileri daha iyi ölçebildiği kabul edilmektedir (Bao, 2007; DeMars, 2006). Bu nedenle madde takımlarının geniş ölçekli testlerde kullanılması kaçınılmazdır. Ancak testlerde madde takımlarının kullanılmasının avantajlarına karşın MTK'nin en önemli varsayımlarından biri olan yerel bağımsızlığın bozulmasına neden olması da olumsuz bir durum yaratmaktadır. Bu nedenle araştırmayı önemli kılan ikinci durum, standart MTK modellerinin madde takımları arasındaki yerel bağımlılığı dikkate almadığına ve madde takımları için geliştirilmiş başka modellerin kullanımının yaygınlaştırılmasına dikkat çekmesidir.

Standartlaştırılmış eğitim testlerinin birçoğunda madde takımlarının kullanılmasındaki artış nedeniyle, madde takımlarından oluşan testlerin nasıl puanlanacağı ve analiz edileceği konusu son on yılda önemli bir araştırma başlığı olmuştur (Chien, 2008). Madde takımlarının puanlanması ve değerlendirilmesi konusunda farklı görüşler öne sürülmüştür. Genel olarak madde takımlarına da testte yer alan diğer maddelere olduğu gibi bağımsız bir madde şeklinde davranılarak iki kategorili ya da çok kategorili standart MTK modelleriyle puanlama yoluna gidilmiş ve bu durumda sonuçlarda çeşitli sorunlarla karşılaşmıştır. Kısaca belirtmek gerekirse madde takımlarına standart modellerle yaklaşmak; (a) madde güçlük kestiriminde yanlılığa, (b) madde ayırıcılıklarının gerçek değerinden yüksek değerde kestirilmesine, (c) bireylerin puanlarının olduğundan yüksek değerde kestirilmesine, (d) test bilgisinin ve güvenirliliğinin yüksek kestirilmesine, (e) test

eşitleme sonuçlarında hatalara neden olmaktadır (Ackerman, 1987; Bradlow vd., 1999; Chen & Thissen, 1997; DeMars, 2006; DeMars, 2012; Lee vd., 2001; Y. Li, Bolt & Fu, 2006; Sireci vd., 1991; Tuerlinckx & De Boeck, 2001; Wainer, 1995; Wainer vd., 2000; Wang & Wilson, 2005a; Yen, 1984, 1993). Özellikle bireylerin puanlarının olduğundan yüksek kestirilmesi durumu, bireylerin yeteneklerine göre kategorize ederken yanlış sınıflandırılma olasılığını arttırarak, hatalı çıkarımlara yol açabilmektedir (Sireci vd., 1991; Yen, 1993). Tuerlinckx ve De Boeck (2001) araştırmalarında, yerel bağımsızlık varsayımının bozulmasının madde ve yetenek parametrelerinin kestirimlerinde yanlılığa neden olduğunu belirlemişlerdir. Buna ek olarak yapılan araştırmalarda, yerel madde bağımlılığının görmezden gelinmesi durumunda, test bilgisinin ve güvenilirliğin yüksek kestirildiği görülmüştür (Thissen, Steinberg & Mooney, 1989; Sireci vd., 1991). Bu nedenle doğru modelleme, test sonuçlarına dayanarak verilen kararların doğruluğu açısından büyük önem taşımaktadır.

Bu gerekçelerle madde takımlarından oluşan testler için daha karmaşık modeller geliştirilmiştir. Böylece madde takımları arasındaki yerel bağımlılık hesaba katılabilmiş ve madde takımlarının MTK'nin en temel varsayımlarından birini bozmasına karşın bu tür veri setlerinde uygun olmayan standart MTK modellerinin kullanımının önüne geçilmeye çalışılmıştır. Min ve He (2014) tarafından yapılan araştırmada, madde takımı etkisini ele alan modellerden elde edilen model-veri uyumunun ve parametre istatistiklerinin tek boyutlu MTK (TBMTK) modelinden elde edilenlere göre daha iyi olduğunu göstermektedir. Madde takımları içeren testlerin analizlerinde iki faktör modeli (2FM) gibi daha karmaşık modeller de kullanılmaktadır (DeMars, 2006; Min & He, 2014; Rijmen, 2010). Bu araştırmada da madde takımı modelleri karşılaştırılmaktadır. TBMTK'ye göre daha karmaşık olan bu modellerin kullanılmasına dikkat çeken bu araştırma, bu açıdan da önemli görülmektedir.

Yapılan alan yazın taramasında Türkiye'de Madde Takımı Tepki Kuramı (MTTK) ile ilgili modellerin karşılaştırılmasını kapsayan herhangi bir çalışmaya rastlanmamıştır. Araştırmanın bu yönüyle Türkiye'deki alanyazına katkı getireceği düşünülmektedir. Türkiye'de uygulanan geniş ölçekli sınavlarda da madde takımları kullanılmaktadır. Bu araştırmadan elde edilecek sonuçlar madde takımlarının yer aldığı sınavların analizleri için de bilgi verici ve yönlendirici olabilecektir.

Bu arařtırmada madde takımı ieren testlerden elde edilen madde ve yetenek parametrelerindeki hatalı kestirimlerin en aza indirilmesini saėlayacak řekilde, madde takımları bulunan testlere uygulanan modellerden elde edilen parametreler karřılařtırılacaktır. Bu amala ok sayıda madde takımının yer aldıėı PISA 2012 verileri kullanılmıřtır. Arařtırmada, madde takımları ieren testlerde kullanılan tek boyutlu madde tepki kuramı modeli (iki kategorili puanlanan maddeler iin iki parametrelili lojistik model, ok kategorili puanlanan maddeler iin ise genelleřtirilmiř kısmi puan modeli), iki faktör modeli ve madde takımı tepki kuramı modeli olmak üzere üç model karřılařtırılmıř ve parametrelerin daha az hatalı kestirilmesini saėlayan model belirlenmeye alıřılmıřtır. MTK tabanlı alıřmalarda, örneklem büyüklüėünün kestirimler üzerinde etkili olduėuna dair sonuçlar elde edilmiřtir (akıcı-Eser ve Gelbal, 2015; Koėar, 2015; Sünböl ve Erkuř, 2013). Bu nedenle alıřmada örneklem büyüklüėü baėımsız deėiřken olarak ele alınmıřtır. Alan yazında madde takım sayısı ile ilgili farklı kořullar ieren alıřmalar da bulunmaktadır (Bradlow vd., 1999; Chen, 2010; DeMars, 2006). Bu nedenle madde takım sayısındaki deėiřim de alıřmada incelenmiřtir. Ancak baėımsız madde sayısı ile ilgili alan yazında alıřmaya rastlanmamıřtır ve baėımsız madde sayısının da kestirimler üzerinde etkili olabileceėi düşünölerek örneklem büyüklüėü ve madde takım sayısının yanı sıra baėımsız madde sayısına da baėımsız deėiřken olarak alıřmada yer verilmiřtir.

1.3. Problem Cümlesi:

Madde takımı ieren testlerde madde takımı sayısı, baėımsız madde sayısı ve örneklem büyüklüėü deėiřtiėinde parametreleri tahmin etmek iin kullanılan

- a) tek boyutlu madde tepki kuramı modelinden,
- b) iki faktör modelinden,
- c) madde takımı tepki kuramı modelinden

elde edilen madde parametreleri ve bu parametrelere ait standart hata deėerleri ile yetenek parametreleri ve bu parametrelere ait standart hata deėerleri nasıldır?

1.3.1. Alt Problemler:

Madde takımı ieren testlerde madde takımı sayısı (4/8), baėımsız madde_ sayısı (0/7/14) ve örneklem büyüklüėü (250/500/1000) deėiřtiėinde; tek boyutlu madde tepki kuramı, iki faktör ve madde takımı tepki kuramı modellerinden elde edilen

1. madde parametreleri ve madde parametrelerine ait standart hata deęerleri nasıldır?
2. yetenek parametreleri ve yetenek parametrelerine ait standart hata deęerleri nasıldır?
3. model-veri uyumları nasıldır?
4. madde parametreleri ve bu parametrelere ait hata deęerleri arasındaki korelasyonlar nasıldır?
5. yetenek parametreleri ve bu parametrelere ait hata deęerleri arasındaki korelasyonlar nasıldır?

1.4. Sayılılar:

Bu arařtırmada, řans parametresini içermeyen MTK ve MTTK modelleri kullanıldıęından, řans parametresinin ihmal edilebilir olduęu varsayılmıřtır.

1.5. Sınırlılıklar:

Arařtırmada, PISA 2012 uygulamasındaki matematik alanına ait biliřsel testlerden 10. kitapçıktan elde edilen veriler kullanılmıřtır. Bu bakımdan çalıřma bu kitapçıkta yer alan 14 bağımsız madde ve sekiz madde takımı ile sınırlıdır.

1.6. Tanımlar:

Madde Takımı: Aynı içerięe baęlı olan birkaç maddeler oluřan kümedir.

Bağımsız Madde: Ortak bir içerięe sahip olmayan, tek başına bir madde olarak oluřturulan maddedir.

İki-faktör Modeli: Birincil temel bir faktörün yanında bir ya da daha fazla ikincil faktörlerin ya da özel bir içerię alanının olmasına izin veren modeldir.

Madde Takımı Tepki Modeli: Madde tepki kuramı modellerine, madde takımı etkisini göz önüne alan ikincil bir boyutun eklenmesiyle oluřturulan modeldir. İki faktör modelinin sınırlandırılmıř halidir.

1.7. Arařtırmanın Kuramsal Temelleri:

Alan yazın taramasıyla oluřturulan bu kısımda ilk olarak madde tepki kuramının karakteristięinden, kuramın varsayımlarından ve iki kategorili ile çok kategorili MTK modellerinden bahsedilecektir. Daha sonra madde takımı kavramı açıklanacak ve

madde takımları içeren testlere uygulanan bazı modellere ayrıntılı olarak yer verilecektir.

A. Madde Tepki Kuramı

Madde Tepki Kuramı, kişinin ölçülen özellik düzeyi ile verdiği yanıtlar arasında bir ilişki olduğunu kabul eder ve bu ilişkiyi matematiksel bir fonksiyon ile açıklayan olasılıklı bir model önerir (Embretson & Reise, 2000). Bu kuramın kalbinde, farklı yetenek düzeyinde olan bireylerin bir maddeye nasıl tepki vereceklerinin matematiksel olarak modellenmesi vardır (Crocker & Algina, 1986). MTK bireyin yeteneği ile maddeye doğru yanıt verme olasılığı arasındaki ilişkiyi madde karakteristik eğrisi (MKE) adı verilen bir eğri yardımıyla gösterir. Gözlenen değişkenler (maddeye verilen yanıtlar) ile örtük değişken (özellik) arasındaki gerçek ilişkiyi tanımlayan bu eğriye göre olasılık arttıkça, maddeyi doğru yanıtlandırmak için gereken özellik düzeyi de artar (Embretson & Reise, 2000). MKE grafiğinin şekli, madde tepki kuramının çoğu uygulamasında S biçiminde olsa da, grafik her zaman bu şekilde olmayabilir (Crocker & Algina, 1986).

MTK'ye göre bireyin yeteneği (θ), ona sunulan maddelerden bağımsızdır ve $-\infty$ ile $+\infty$ arasında değer alır. Ancak uygulamalarda kolaylık sağlaması açısından bu aralık, eşit aralıklı bir ölçek üzerinde -3 ile +3 değerleri arasında tanımlanmıştır (van der Linden & Hambleton, 1997).

MTK'nin en önemli avantajları, yapılan kestirimlerin maddeye yanıt veren grubun özelliklerinden ve gruba sunulan maddelerin özelliklerinden bağımsız olmasıdır. Ancak MTK'nin bu avantajlarının yanı sıra, karşılanması zor varsayımları bulunmaktadır. Bu varsayımlar, kullanılan MTK modeline göre değişmektedir. Bireyin madde üzerindeki performansını etkileyen tek bir yetenek türü olduğu durumlarda TBMTK kullanılır. MTK çalışmalarında en çok TBMTK modellerinin uygulanmasına karşın bu kuramın karşılanması zor varsayımları bulunmaktadır.

Varsayımları

TBMTK'nin en temel varsayımları tek boyutluluk, yerel bağımsızlık ve testin hız testi olmamasıdır. Tek boyutluluk, test puanına katkı yapan bütün test maddelerinin aynı örtük özelliği ölçmesi ve yanıtlayıcının testteki performansının yalnızca o örtük özellikten kaynaklanması şeklinde ifade edilebilir (Pomplun, 1988). Hambleton ve Swaminathan (1985) tek boyutluluğu, bireyin testteki performansını belirleyen

faktörün testin ölçmeye çalıştığı faktör olması ile ifade ederken; Crocker ve Algina (1986) ise bu kavramı maddeler arasındaki istatistiksel bağımlılığın tek bir örtük özellikle açıklanması şeklinde betimlemişlerdir. Ancak pratikte tek boyutluluğun tamamıyla sağlanması olası değildir ve bu nedenle bireyin testteki performansı için gerekli olan tek bir başat yetenek olduğunda test tek boyutlu kabul edilir (Embretson & Reise, 2000).

MTK varsayımlarından bir diğeri olan yerel bağımsızlık ise bir bireyin bir maddedeki performansının, diğer maddelerdeki performansını olumlu ya da olumsuz yönde etkilememesidir (Hambleton & Swaminathan, 1985). Yerel bağımsızlık, tek boyutlulukla ilişkili bir özelliktir. Eğer bir ölçek tek boyutluluk özelliğini gösteriyorsa bu ölçekte yer alan maddelerin yerel bağımsızlık özelliğine sahip olduğu söylenebilir.

MTK'nin diğeri bir varsayımı, testin hız testi olmamasıdır. Bu varsayımın kontrolü için ise (1) testi tamamlayan bireylerin yüzdesi, (2) testin % 75'ini tamamlayan bireylerin sayısı, (3) bireylerin % 80'inin tamamladığı madde sayısı incelenebilir (Hambleton & Swaminathan, 1985).

Bu çalışmada tek boyutlu MTK modellerinden iki kategorili ve çok kategorili MTK modelleri kullanılmıştır.

- İki Kategorili Madde Tepki Kuramı Modelleri

Testlerde çoktan seçmeli maddelerin sıklıkla kullanılması nedeniyle, uygulamalarda MTK modellerinden en çok iki kategorili modeller kullanılmaktadır. Çünkü çoktan seçmeli maddelerin geliştirilmesi ve puanlanması diğer madde formatlarına kıyasla görece olarak daha kolaydır (Haladaya, 1997). İki kategorili MTK modelleri, modelin içerdiği parametre sayısına göre isimlendirilmektedir. İki kategorili MTK modelleri; üç parametrelili lojistik model (3PLM; Birnbaum, 1968), iki parametrelili lojistik model (2PLM; Birnbaum, 1968) ve bir parametrelili lojistik modeldir (1PLM; Rasch, 1960).

3PLM, iki kategorili MTK modellerinin en genel formudur. Bu modelde güçlük (b), ayırıcılık (a) ve şans (c) parametresi olmak üzere üç parametre bulunmaktadır. Güçlük parametresi, MKE'nin büküm noktasıdır ve $\frac{1+c_i}{2}$ olasılığının bulunduğu noktayı temsil eder. Ayırıcılık parametresi, MKE'de yetenek ölçeğinde b_i 'ye karşılık gelen noktanın eğimidir. Şans parametresi ise MKE'nin yatay asimptotudur ve en düşük yetenek düzeyindeki bireyin maddeyi doğru yanıtlama olasılığını gösterir

(Hambleton, Swaminathan & Rogers, 1991; Embretson & Reise, 2000). Bu model üç parametre içerdiğinden üç parametrelili lojistik model olarak adlandırılmıştır (Birnbaum, 1968). Bir bireyin i maddesine doğru yanıt verme olasılığı 3PLM'ye göre aşağıda verilen eşitlikle bulunmaktadır.

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \quad (1.1)$$

Eşitlik 1.1'de $P_i(\theta)$: θ yetenek düzeyindeki kişilerden rastgele olarak seçilen bir kişinin i maddesini doğru yanıtlama olasılığını, b_i : i maddesinin güçlük indeksini, a_i : i maddesinin ayıricılık indeksini, c_i : i maddesinin şansla doğru cevaplanma indeksini ve D ise 1,7 değerinde olan ölçek sabitini göstermektedir.

2PLM'de, şans başarısı göz ardı edilmekte ve şans parametresinin sıfır olduğu varsayılmaktadır. Bu nedenle bu model, 3PLM'nin matematiksel olarak basitleştirilmiş hali olarak kabul edilir. Modelde ayıricılık parametresi ve güçlük parametresi olarak iki parametre bulunduğu için modele, iki parametrelili lojistik model ismi verilmiştir (Birnbaum, 1968). Bir bireyin i maddesine doğru yanıt verme olasılığı 2PLM'ye göre aşağıda verilen eşitlikle bulunmaktadır.

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} \quad (1.2)$$

Eşitlik 1.2'de $P_i(\theta)$: θ yetenek düzeyindeki kişilerden rastgele olarak seçilen bir kişinin i maddesini doğru yanıtlama olasılığını; b_i : i maddesinin güçlük indeksini; a_i : i maddesinin ayıricılık indeksini ve D ise 1,7 değerinde olan ölçek sabitini göstermektedir.

1PLM de ayıricılık parametresini bir sabite eşitlediği için iki parametrelili lojistik modelin sınırlandırılmış bir halidir. Yalnızca güçlük parametresi, parametre olarak ele alındığından bu modele, bir parametrelili lojistik model denilmektedir. Bu modelde, bütün maddelere ait ayıricılık parametresinin eşit olduğu ve şans parametresinin ihmal edilebilir olduğu varsayımları bulunmaktadır (Hambleton & Swaminathan, 1985). 1PLM'de her bir maddeye ait ayıricılık parametrelerinin ortalaması alınarak ayırt edicilik parametresi sabitlenirken, ayırt edicilik parametresi bütün maddeler için 1'e sabitlendiğinde model "Rasch modeli" olarak isimlendirilir (Embretson & Reise, 2000). Bu modele göre bireyin bir "i" maddesine doğru yanıt verme olasılığı Eşitlik 1.3 yardımıyla hesaplanır.

$$P_i(\theta) = \frac{e^{Da(\theta-b_i)}}{1 + e^{Da(\theta-b_i)}} \quad (1.3)$$

Eşitlik 1.3'te sunulan matematiksel ifadede $P_i(\theta)$: θ yetenek düzeyindeki kişilerden rastgele olarak seçilen bir kişinin i maddesini doğru yanıtlama olasılığını; b_i : i maddesinin güçlük indeksini; a tüm maddelerde aynı olan ayırıcılık indeksi için sabit bir sayıyı ve D ise 1,7 değerinde olan ölçek sabitini göstermektedir.

İki kategorili MTK modellerinin her birinin kendine özgü varsayımları bulunmaktadır ve 1, 2 ve 3 parametrelili modellerden hangisinin seçiminin doğru olacağı model sayıtlarıyla ilgilidir (Crocker & Algina, 1986). Bu nedenle araştırmacı model seçiminde, kullanılan modelin sayıtlarını göz önünde bulundurmalıdır. Maddelerin eşit ağırlıklı olarak puanlanıp puanlanmadığı, ölçme için beklenen ölçek özellikleri, verilerin modele uygunluğu ve parametre kestirim amacı gibi ölçütler de, uygun model seçiminde dikkate alınmalıdır (Emberson & Reise, 2000). Buradan anlaşılacağı üzere model seçimi, tek bir ölçüte dayandırılmamalıdır.

- Çok Kategorili Madde Tepki Kuramı Modelleri

Çok kategorili puanlanan maddeler, iki kategorili puanlanan maddelere göre daha bilgilendiricidir ve bu modeller, iki kategorili MTK modellerinin geliştirilmesiyle oluşturulmuş modellerdir. Çok kategorili MTK modelleri, maddeye verilecek tepkinin çok kategorili olduğu durumlarda, bireyin yetenek düzeyi ile madde seçeneklerindeki belirli bir kategoriyi yanıtlaması arasındaki doğrusal olmayan ilişkiyi açıklama amacıyla kullanılmaktadır (Emberson & Reise, 2000). Bu amaçla çok kategorili puanlanan maddeler için bireyin belli bir kategori puanına ulaşma olasılığı, çok kategorili MTK modellerinden biri kullanılarak tanımlanabilmektedir (Tang, 1996). Çok kategorili MTK modelleri, bireyin yanıtının belirli bir kategorideki şartlı olasılığını belirlemede gerektirdiği işlemler bakımından “doğrudan/direct” ve “dolaylı/indirect” olmak üzere ikiye ayrılmaktadır (Emberson ve Reise, 2000). Aşamalı Tepki Modeli - ATM (Graded Response Model) ve bu modelin modifiye edilmiş hali olan Değiştirilmiş Aşamalı Tepki Modeli – M-ATM (The Modified Graded Response Model) gibi modeller, bireyin örtük özelliği ile belirli bir kategoriye cevap verme olasılığı arasındaki ilişkiyi betimlemede iki aşamalı bir süreç gerektirmesi sebebiyle dolaylı modellerde yer alırken; Kısmi Puan Modeli – KPM (Partial Credit Model), Genelleştirilmiş Kısmi Puan Modeli G-KPM (Generalized Partial Credit

Model), Dereceleme Ölçeği Modeli - DÖM (Rating Scale Model), Sınıflamalı Tepki Modeli - STM (Nominal Response Model) gibi modeller tek bir eşitlikle hesaplama yapmaları nedeniyle doğrudan modeller içerisinde yer almaktadır. Madde karakteristik eğrilerinin matematiksel tanımlamalarındaki farklılıklar nedeniyle modeller arasında da farklılıklar oluşmaktadır.

Dolaylı Modeller

Aşamalı Tepki Modeli (ATM), çok kategorili puanlamalar için Samejima (1969) tarafından geliştirilen ve 2PLM'nin uzantısı olan bir modeldir. 2PLM'den farkı ise 2PLM'de bir tane eşik parametresi bulunurken, ATM'de iki ya da daha fazla eşik parametresi bulunabilir. Bireyin cevaplarının kısmi doğruluğuna dayanan puanlamalarda (ör. yanlış, kısmen doğru, doğru) ya da davranış durumlarına katılma derecelerinin farklılık gösterdiği Likert ölçeğindeki gibi derecelenmiş kategorik cevaplar içeren ölçme araçlarının değerlendirilmesinde kullanılmaktadır (Park, 1997). D-ATM'nin ATM'den farkı, kategorilere ilişkin eşik parametresi ile ilgilidir. Modelde her madde için yer parametresi ve testteki tüm kategoriler için eşik parametresi kestirilir (Emberson & Reise, 2000). Bu modellere aynı zamanda fark modelleri (difference models) de denilmektedir.

Doğrudan Modeller

Dereceleme Ölçeği Modeli'ne (DÖM), Sınıflamalı Tepki Modeli'ne (STM), Kısmi Puan Modeli'ne (KPM) ve Genelleştirilmiş Kısmi Puan Modeli'ne (GKPM) aynı zamanda toplama bölünmeli modeller (divided by totals) da denilmektedir. Andrich (1978a, 1978b) iki kategorili puanlanan maddeler için Rasch modelini geliştirerek, davranış düzeyinin farklı derecelerini gösteren sıralı kategorileri kullanarak puanlanan dereceli ölçek maddeleri için DÖM'ü oluşturmuştur. Bu model, maddelerin aynı cevap formatında olmasını gerektirmez. Bock (1972) tarafından geliştirilen STM'de ise maddelerin seçeneklerinin sıralı olması ya da madde cevaplarının sayısal olarak sınıflandırılması zorunluluğu yoktur. Bu model de 2PLM'nin sınıflamalı tepkiler için genişletilmiş bir model olup testte yer alan bir madde içerisindeki olası kategorik tepkiler için geliştirilmiştir.

Kısmi Puan Modeli ve Genelleştirilmiş Kısmi Puan Modeli

Bu araştırmada Masters (1982) tarafından geliştirilmiş KPM, kullanılacağı için bu model diğer modellerden daha ayrıntılı anlatılmıştır. KPM çözümü çok adım

gerektiren ya da kısmi doğru yanıtı bulunan maddelerin analizlerinde kullanılmaktadır. Bu nedenle kısmi cevap içeren başarı testleri için kullanılabilir. Ayrıca bu modelin çok kategorili davranış ya da kişilik testlerine de uygun olduğu söylenebilir. Model, çok kategorili puanlanan maddeler için Rasch modelinin genişletilmiş halidir ve testteki maddelerin ayırt edicilik düzeylerine göre farklılaşmadığı ve şans faktörünün olmadığı varsayılır (Muraki, 1982). Embretson ve Reise (2000) modelin önemli iki avantajı olarak, parametre kestirimini tek bir adımla yapabilmesini ve yeteneğin kestirilmesi için Rasch modelinde olduğu gibi ham puanların yeterli bir istatistik olmasını belirtmişlerdir. Modelde tanımlanan madde parametrelerinin ve cevap kategorilerinin sıraları arasında bir ilişki olması ve işlem basamaklarının güçlüklerinin aynı olması gerekmez (Masters, 1982). Aynı zamanda modelde, işlem basamaklarını güçlüklerine göre sıralamak gibi bir zorunluluk da bulunmamaktadır. Bu avantajlarının yanı sıra, bu modelde aynı ham puana sahip bireylerin aynı yetenek düzeyinde yer alması bir dezavantaj oluşturmaktadır. Çünkü tüm maddelerin örtük özelliklerle eşit derecede ilişkili olma gerekliliğine uygulamada ulaşmak oldukça zordur (Embretson & Reise, 2000).

Masters (1982) çalışmasında aşağıdaki gibi bir örnek vermiştir:

$$(7.5/0.3 - 16)^2 = ?$$

Adım 1: 7.5/0.3 bölme işleminin sonucunu değerlendirmek

Adım 2: Adım 1'de bulduğu sonucu, 16'dan çıkartmak

Adım 3: Adım 2'deki sonucun karesini almak

Örnekte görüldüğü gibi birinci adımı tamamlamadan ikinci adımı doğru yapmak olanaksızdır. Ancak adımların zorluk dereceleri arasında bir ilişki yoktur. Bu modelde, birey ne kadar çok adımı tamamlarsa, o kadar yüksek tepki kategori puanı almaktadır. Yüksek kategori tepki puanı ise düşük tepki kategori puanına göre daha yüksek yeteneği ifade etmektedir.

KPM'de madde güçlük parametresi yerine, madde adım güçlüğü parametresi bulunmaktadır ve bu parametre örtük özellik ölçeğinde ardışık iki tepki kategori eğrisinin kesiştiği noktayı göstermektedir (Embretson & Reise, 2000). Bu parametre ne kadar yüksek olursa, bir sonraki adıma geçmek de o kadar zorlaşır. Ayrıca modelde ve tüm maddelerin eşit eğim parametresine sahip olduğu varsayılmaktadır. Modelde $K_i = m_i + 1$ tepki kategorili madde için $x = 0, \dots, m_i$ olarak puanlanan i

maddesi ele alınır, KPM için $x = j$ kategori tepki eğrisi Eşitlik 1.4'teki gibi yazılabilir (Embretson & Reise, 2000):

$$P_{ix}(\theta) = \frac{\exp[\sum_{j=0}^x (\theta - \delta_{ij})]}{\sum_{r=0}^{mi} \exp[\sum_{j=0}^r (\theta - \delta_{ij})]} \quad (1.4)$$

j : 1, 2, ..., m_i

m_i : 1'den aşağı olan puan kategorilerinin sayısı

δ_{ij} : j ile puanlanan kategoriye ait adım güçlüğü parametresi

Muraki (1992) tarafından geliştirilen bir diğer çok kategorili TBMTK modeli de G-KPM'dir. G-KPM'nin KPM'den farkı, ölçek içinde maddelerin eğim parametrelerinin farklılaşmasına olanak vermesidir. Bu model eğitimsel değerlendirmelerde oldukça sık kullanılmaktadır ve veri analizlerinde aşamalı modele etkili, popüler bir alternatiftir (Cai, Yang & Hansen, 2011). Belirli bir yetenek düzeyinde belirli bir kategori puanına (x) ait olasılık G-KPM'de Eşitlik 1.5'teki gibi ifade edilir.

$$P_{ix}(\theta) = \frac{\exp[\sum_{j=0}^x a_i (\theta - \delta_{ij})]}{\sum_{r=0}^{mi} \exp[\sum_{j=0}^r a_i (\theta - \delta_{ij})]} \quad (1.5)$$

Eşitlik 1.5'te, Eşitlik 1.4'teki ifadeler ek olarak i maddesi için ayırt edicilik parametresi olan a_i parametresi eklenmiştir.

B. Madde Takımı Kavramı

Wainer ve Kiely (1987) tarafından, bir birim olarak geliştirilen tek bir içerikle ilişkili maddeler grubu olarak ifade edilen madde takımı kavramı, bilgisayar ortamında bireyselleştirilmiş test (computer adaptive testing – CAT) uygulamalarında yaşanan bağlam etkisi (content effects) ya da madde sırası (item ordering) gibi konulardaki problemlerin üstesinden gelmek için ortaya atılmış bir kavramdır. Daha sonra Wainer ve Lewis (1990) manipüle edilecek kadar küçük ancak kendi içeriğini taşıyacak kadar büyük testler olarak bu kavram ile ilgili daha genel bir tanım yapmışlardır. Tek bir maddeden uzun ancak bütün bir testten daha kısa olarak tanımlanan bu kavram ile ilgili çalışmalar, teknik ve ekonomik donanımın iyileşmesi sonucunda 1980'lerde ortaya çıkmıştır (Wainer, Bradlow & Wang, 2007).

Sınavlarda bireylere uyarıcıların sunulmasının önemli miktarda zaman aldığı ve bir içerikle ilgili tek bir soru sorulmasının zamanı boşa harcamaya neden olduğu

düşüncesi ile birlikte test geliştirmede, bir soru kümesince paylaşılan geniş uyarıcıya dayalı madde takımları etkili bir yol olarak sıkça kullanılmaktadır (Wainer vd., 2007). k madde takımından oluşan bir test düşünülüğünde madde takımlarının her biri testle ölçülmek istenen genel bir faktörü ve bir madde takımı özellik faktörünü içermektedir. Bu nedenle bu gibi testlerin faktör yapısı bir genel faktör ile her madde takımının bir faktörü olmak üzere $k + 1$ tane ortogonal faktöre sahiptir.

C. Yerel Madde Bağımlılığı

Genel bir uyarıcıyı paylaşan maddeler grubunun bir madde kümesi içerisine yuvalanmasıyla oluşturulan testlerde, yerel bağımsızlık varsayımı bozulabilmektedir (Bao, 2007). Embretseon ve Reise'in (2000) belirttiği gibi "Uygulamada, madde tepkileri bağlantılı olduğunda yerel bağımsızlık bozulmaktadır" (s.188). Bu durumun nedeni, bireyin madde takımı içerisindeki bir maddeye verdiği tepkinin, aynı madde takımı içerisinde yer alan diğer maddelere verdiği tepkiyi etkilemesidir (Wainer ve Kiely, 1987). Bu nedenle bir uyarıcıya dayanan maddeler arasında yerel bağımlılık oluşmaktadır. Tuerlinckx ve De Boeck (2001) ise çok boyutluluğun ve maddeler arası etkileşimin yerel madde bağımlılığına (YBM) neden olabileceğini belirtmişlerdir.

YMB pozitif ya da negatif olabilmektedir (Habing & Roussos, 2003; Yen, 1993). İki madde arasındaki pozitif YMB bireyin her iki maddede de beklenenden iyi ya da düşük performans göstermesi anlamına gelirken; negatif YMB ise bireyin maddelerden birinde beklenenden yüksek, diğerinde ise beklenenden düşük performans göstermesi anlamına gelmektedir (F. Li, 2009; Lu, 2010; Nofer, 2007). Ancak Liu ve Thissen (2012) negatif YMB'nin daha az rastlanılan ve kolay açıklanamayan bir durum olduğunu belirtmişlerdir.

Yerel madde bağımlılığının (Local item dependency – LID) alanyazında birçok farklı nedenleri ve türleri olduğu görülmektedir. Hoskens ve De Boeck (1997) yerel madde bağımlılığını sıra bağımlılığı (order dependency) ve kombinasyon bağımlılığı (combination dependency) olarak iki türde sınıflandırmışlardır. Sıra bağımlılığı, önceki maddeye verilen yanıtın sonraki maddelerde verilen yanıtı etkilemesi durumunda meydana gelirken, kombinasyon bağımlılığı ise bir paragraf ya da grafik gibi ortak bir uyarıcıya dayanan maddeler kümesinin olduğu senaryolarda ortaya

çıkılmaktadır. Yen (1993) ise yerel madde bağımlılığı için on farklı neden göstermiştir. Bu nedenler kısaca şu şekilde ifade edilebilir:

Dışsal yardım ya da müdahale (External assistance or interference). Öğretmenlerden ya da diğer öğrencilerden, bazı maddeler için ilgili konu hakkında yardım alan öğrenciler, bu maddelerde daha iyi performans göstereceklerdir. Bu durum YMB'ye neden olmaktadır. Öğretmenlerden elde edilen yanlış bilgi, sınıftaki gürültü gibi dışsal müdahaleler de bazı öğrencilerin bazı maddelerde daha kötü performans göstermesine neden olur. Yine bu durum da YMB'ye yol açmaktadır.

Zaman faktörü (Speededness). Bir grup öğrencinin, bir testi tamamlamak için yeterli zamanı yok ise testin sonlarına doğru cevaplanan maddeler büyük olasılıkla YMB'ye neden olur. Eğer zaman yönetimi, testi alanlar için önemli bir faktör ise YMB oluşur.

Yorgunluk (Fatigue). Metin parçaları ya da görevler, testin sonunda yer aldığı anda testin başında yer almasına göre daha zor olma eğilimi gösterebilirler. Maddelerin sonda yer alması nedeniyle öğrencide meydana gelen yorgunluk ya da düşük motivasyon pozitif YMB'ye neden olabilir.

Alıştırma (Practice). Maddeler ile ilgili alıştırma yapma durumuna göre madde performansında artış gözleniyor ise, YMB ortaya çıkar. Eğer maddeler görel olarak homojen bir içerikteyse yerel bağımlılık yalnızca maddelerin testteki görel konumlarıyla ilgili olacaktır.

Madde ya da cevap formatı (Item or response format). Aynı içeriği ölçmek için geliştirilen maddeler farklı formatlarda dizayn edilebilir. Maddelere verilen cevaplar da farklılık gösterebilir. Yapılandırılmış cevaplar, öğrencinin bir model oluşturması, resim çizmesi ya da öykü yazması gibi türlere ya da uzunluğa göre çeşitlenebilir. Bu değişimler YMB'ye sebep olur.

Metin parçasına bağımlılık (Passage dependence). Eğer aynı metin parçasına bağlı birden fazla madde varsa YMB ortaya çıkar. Bu YMB, öğrencinin ilgilerine ve geçmiş bilgilerine göre farklılık gösterebilir.

Madde zincirleme (Item chaining). Eğer maddeler belirli adımlarla organize edilirse bir maddeyi doğru cevaplayan öğrencinin diğer maddeyi de doğru cevaplama olasılığı artacaktır.

Önceki cevabın açıklaması (Explanation of previous answer). Matematik performansının ölçülmesinde, madde zincirlemenin özel bir türü kullanılmaktadır. Bir maddede, bir öğrenciden bir cevabı hesaplaması, sonucu çizmesi ya da karar vermesi istenebilir. Daha sonraki maddede öğrenciden, bir önceki cevabı kullanarak işlemlerin nedenlerini ve süreci açıklaması istenebilir. Bu açıklama, öğrenci performansı hakkında ek bir bilgi istemekle birlikte, büyük olasılıkla YMB'ye neden olmaktadır.

Rubrik ya da değerlendirici ile puanlama (Scoring rubrics or raters). Performans değerlendirmedeki maddeler rubrikler ya da farklı kurallar ile puanlanabilmektedir. Aynı rubrik ile puanlanan maddeler, YMB'ye neden olabilir. Bunun nedeni değerlendiricilerin farklı özelliklere sahip olmasına karşın, öğrencilere ait ortak becerilerin ölçülmesidir.

İçerik, bilgi ve yetenekler (Content, knowledge, and abilities). Başarı testleri, geniş bir puan aralığına ait içeriğe sahip maddelerden oluşmaktadır. Eğer maddeler, sadece tek bir içeriği ölçmeye yönelik ise YMB oluşabilir. Örneğin, üçüncü sınıf matematik testinde saat okuma ile ilgili sorular, kesirleri ölçen maddeler ile karşılaştırıldığında daha fazla YMB içerebilir.

Geniş ölçekli testlerin içeriklerinden dolayı, yerel madde bağımlılığına neden olan bu nedenlerden en çok metin parçasına bağlılık çalışılmıştır ve bu araştırmaların hepsinde yerel bağımlılığın oluşmasına maddelerin ortak bir metne dayanmalarının neden olduğu belirlenmiştir (Keller vd., 2003; Lee, 2004; Sireci vd., 1991; Thissen vd., 1989; Yen, 1993; Zenisky, Hambleton & Sireci, 2002). Ortak bir metne dayalı maddeler kümesi, bireyin metni anlama düzeyine bağlı olarak birbiriyle ilişkili maddelerden oluşur. Wainer ve Lewis (1990) ortak bir metne dayanan maddeler arasındaki koşullu bağımlılıkla baş etmek için üç yol üzerinde tartışmışlardır. Bu yollardan ilki, formatın değiştirilerek her okuma metni için yalnızca bir madde sorulması yönündedir. Ancak bu format ile yerel bağımlılık sorunu çözülsede bireyler hakkında daha çok bilgi toplanabilmesi için daha fazla zaman gerektirmesi açısından çok etkili değildir. İkinci olarak ise araştırmacılar bağımlılığın görmezden gelinmesini ve madde takımlarındaki maddelerin iki kategorili olarak puanlanması yöntemini tartışmışlardır. Ancak bu yöntemde de her maddeden elde edilen bilgi miktarı olduğundan fazla kestirildiği için yetenek kestiriminde doğru olmayan sonuçlara ulaşılabilir. Son olarak ise araştırmada ortak metne dayanan maddeleri

tek bir madde yani madde takımı olarak ele alıp, çok kategorili MTK modelleri ile puanlama yöntemi incelenmiştir. Ancak bu model de bilgi kaybına neden olduğu için yeterli bulunmamıştır.

Zuru (1998) yerel bağımlılık ile baş etmek için geliştirilen bütün stratejilerin iki başlık altında toplanabileceği belirtmiştir: (1) test verisi toplanmadan önce izlenen stratejiler ve (2) test verisi toplandıktan sonra izlenen stratejiler. Testler zaman faktörü, yorgunluk, dışsal yardım ya da müdahale gibi yerel bağımlılık oluşmasına neden olan durumları ortadan kaldıracak uygun koşullarda bireylere uygulandığı zaman, ölçülmek istenen özellik dışında bireyin performansını başka bir faktör ya da boyut etkileyebilir. Böylece test verisi toplanmadan önce yerel bağımlılık oluşturacak durumların önüne geçilmiş olunabilir. Diğer istatistiksel yöntemler ise test verisi toplandıktan sonra izlenen stratejilerdir. Test verisi toplandıktan sonra yerel madde bağımlılığını inceleme stratejilerinin avantajı, sürecin testin yapısını ve tasarımını çok az etkilemesidir (Zuru, 1998).

Yerel Madde Bağımlılığının Belirlenmesi

Maddeler arasında yerel madde bağımlılığı (local item dependence) belirlemek için çeşitli istatistikler geliştirilmiştir. Bu istatistiklerden en çok kullanılanı Yen (1984) tarafından geliştirilen Q_3 istatistiğidir. Bu istatistik iki maddenin artıkları arasındaki Pearson korelasyonuna dayanmaktadır (Monseur, Baye, Lafontaine & Quittre, 2011; Tao, 2008; Yen, 1993). Bu değer, örtük özelliğin etkisi ortadan kaldırıldığında madde çiftleri arasındaki maddeler arası korelasyonu gösterir (O. Zhang, 2010). Artıklarının korelasyonu önemli bir düzey gösteren maddeler, yerel bağımlı olarak düşünülebilir (Yen, 1993). Herhangi iki kategorili madde çifti arasındaki Q_3 istatistiğini hesaplamak için ilk önce seçilen MTK modelini kullanarak testteki bütün maddelere verilen tepkilerden her bireyin yetenek kestirimi ($\hat{\theta}_j$) ve madde parametre kestirimleri hesaplanır (Monseur vd., 2011). Daha sonra yetenek ve madde parametre kestirimlerine dayanarak seçilen MTK modeli ile her bireyin her madde üzerindeki beklenen puanı (E_{ij}) kestirilir. Bireyin gözlenen (X_{ij}) ve beklenen puanı arasındaki farka dayanan d_{ij} değeri Eşitlik 1.5'teki gibi hesaplanır.

$$d_{ij} = X_{ij} - E_{ij} \quad (1.5)$$

X_{ij} bireyin gözlenen puanı ve E_{ij} seçilen bir MTK modelinden elde edilen beklenen puandır. Q_3 indeksi ise madde i ve i' nin artık puanları arasındaki korelasyon ile hesaplanır.

$$Q_{3ii'} = r(d_i, d_{i'}) \quad (1.6)$$

Bu istatistiğin temel varsayımı seçilen MTK modelinin iyi tanımlanmış olmasıdır. Çünkü bu varsayım sağlanmazsa Q_3 istatistiğinin yerel bağımlılık derecesine mi işaret ettiğini yoksa seçilen modelin uyumunun zayıflığından mı kaynaklandığını belirlemek zor olur (Cheng, 1996). Seçilen MTK modeli iyi tanımlanmış olduğunda, Q_3 istatistiği Fisher'ın z dönüşümü ile ortalaması sıfır ve varyansı $1/(N - 3)$ olan normal dağılıma dönüştürülecektir (N, korelasyonun hesaplandığı birey sayısı) (Yen, 1984). Ayrıca bu istatistik madde çiftleri olarak hesaplandığı için, $n(n - 1)/2$ tane Q_3 değeri olacaktır (Tao, 2008).

Q_3 istatistiğinin değeri negatif olma eğilimindedir (Chen & Thissen, 1997; Yen, 1984, 1993). Bu duruma, gözlenen puanların hem beklenen puanların hesaplanmasında hem de artık puanlarının hesaplanmasında tekrar kullanılması neden olmaktadır (Dresher, 2002; Monseur vd., 2011; Yen, 1984). Yerel bağımsızlık olduğunda Q_3 istatistiğinin beklenen değeri n toplam madde sayısını göstermek üzere, yaklaşık olarak $-1/(n-1)$ 'dir (Yen, 1993). Maddeler arasında yerel bağımsızlık olduğunu söyleyebilmek için madde takımları içerisindeki maddelerden elde edilen Q_3 istatistiğinin ortalaması, Q_3 'ün beklenen değeri ile aynı olmalıdır (Keller vd., 2003; F. Li, 2009). Yen & Fitzpatrick (2006) ortalama Q_3 değeri 0.2'den büyük olan madde takımlarına dikkat edilmesi gerektiğini belirtmişlerdir.

Chen ve Thissen (1997) de madde çiftleri arasındaki yerel bağımlılığı belirlemek için dört indeks tanımlamışlardır: (a) X^2 istatistiği, (b) G^2 olabilirlik oran istatistiği, (c) standardize edilmiş ϕ katsayı farkı ve (d) standardize edilmiş log-olabilirlik oran farkı. Bu dört istatistik genellikle beklenen ve gözlenen olasılık tablolarının iki yönlü kovaryansını incelemek için kullanılır. İki kategorili puanlanan i ve j madde çiftinin gözlenen frekansları için iki yönlü frekans tablosu şu şekilde oluşturulabilir (Chen & Thissen, 1997).

		Madde j	
		0	1
Madde i	0	O_{11}	O_{12}
	1	O_{21}	O_{22}

Bu tabloda yer alan O_{pq} , 1-dođru, 0-yanlıř cevabı gstermek zere gzlenen frekanstır. Aynı yapı beklenen frekansa da uygulanabilir (Chen & Thissen, 1997):

		Madde j	
		0	1
Madde i	0	E_{11}	E_{12}
	1	E_{21}	E_{22}

Bu tablodaki E_{pq} deđeri, MTK modelleri tarafından tahmin edilen beklenen deđerdir ve Eřitlik 1.7'deki gibi hesaplanır:

$$E_{pq} = N \int P_i(\theta)^p P_j(\theta)^q [1 - P_i(\theta)]^{(1-p)} [1 - P_j(\theta)]^{(1-q)} f(\theta) d\theta \quad (1.7)$$

Eřitlik 1.7'deki $P_i(\theta)$: i maddesi iin trace line, $f(\theta)$: poplasyon dađılımı ve N : birey sayısıdır.

Chen ve Thissen (1997)'nin gzlenen ve beklenen deđerler arasındaki farklılıklara dayanarak hesapladıđı ve yerel bađımlılık iin nerdiđi olabilirlik oran G^2 istatistiđi ve Pearson χ^2 istatistiđidir. G^2 istatistiđi Eřitlik 1.8 yardımıyla hesaplanır:

$$G^2 = -2 \sum_{i=1}^2 \sum_{j=1}^2 O_{ij} \ln \frac{E_{ij}}{O_{ij}} \quad (1.8)$$

O_{ij} : i ve j maddesi iin hesaplanan gzlenen frekans

E_{ij} : beklenen frekans

Chen ve Thissen'in (1997) nerdiđi bir diđer indeks Pearson χ^2 'dir. Bu χ^2 deđeri Eřitlik 1.9'daki gibi hesaplanır (Bishop, Fienberg & Holland, 1975, s.57; akt. Chen ve Thissen, 1997):

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} \quad (1.9)$$

Chen ve Thissen (1997) her iki istatistik iin de serbestlik derecesi bir olan ki kare dađılımının kullanılması nermektedir.

MTTK'de yer alan madde takımı parametresinin (γ) varyansı da yerel bađımlılık iin bir l oluřturmaktadır. Eđer σ_γ^2 sıfırda, yerel bađımlılık yoktur. Ancak bu deđer sıfırdan uzaklařtıka, yerel bađımlılık durumuna iřaret etmektedir. σ_γ^2 'nin byklk

normu σ_{γ}^2 ile Θ 'nın varyansının 1 olarak normalleştirilmesi arasındaki orandır (Wainer vd., 2007). γ parametresi madde takımları arasındaki maddelerin bağımlılıklarının bir ölçüsü olabilir. Bradlow vd. (1999) madde takımına ait varyansın 0.3 üzerinde olmasının oldukça büyük madde takımı etkisine işaret ettiğini belirtmiştir. Wang ve Wilson (2005a) ise tesadüfi madde takımı değişkeninin varyansını küçük etkiden büyük etkiye olacak şekilde 0.25, 0.50, 0.75 ve 1.00 olarak göstermişlerdir. Yapılan diğer araştırmalar ise tesadüfi etkinin (γ) .2, .9 ve 1.5 değerlerini küçük, orta ve büyük madde takımı etkisi olarak göstermiştir (Wang, Bradlow & Wainer, 2002; Wang & Wilson, 2005a). Y. Li vd. (2006) ise madde takımı varyansını üç düzeyde – 0 (madde takımı etkisi yok), 0.5 (orta düzeyde madde takımı etkisi) ve 1 (yüksek düzeyde madde takımı etkisi) – olacak şekilde ayarlamışlardır. Glas, Wainer ve Bradlow (2000) madde takımı varyansını 0.25 ve 1.00 olarak yürüttükleri çalışmada 0.25 değerinin madde parametre kestirimlerinde göz ardı edilebilir bir yanlılık oluşturduğunu belirlemişlerdir. Bu durum eğer düşük düzeyde madde takımı etkisi varsa, yerel bağımsızlık varsayımının bozulmasının kabul edilebilir olduğuna ve 2PL ya da 3PL gibi modellerin parametre kestirimlerinin kalitesini çok düşürmeyeceğine işaret etmektedir (Paap & Veldkamp, 2012).

Yerel madde bağımsızlığı ile boyutluluk birbiriyle yakından ilişkili iki varsayım olduğu için (Hambleton vd., 1991; Lord, 1980; Yen, 1993), bu varsayımlardan birinin sağlanması diğerinin de sağlanacağına işaret edebilmektedir. Bir başka deyişle, tek boyutluluk varsayımı karşılandığında ya da çok boyutlu MTK modelleri ile boyut sayısı doğru olarak belirlendiğinde, yerel madde bağımlılığının da olmadığı varsayılır (Tao, 2008). Bu nedenle tek boyutluluğun test edilmesi de yerel madde bağımlılığı için bir gösterge oluşturabilir.

Madde takımlarına ait veri setlerinde yerel bağımsızlık varsayımının bozulması ile baş etmek için birçok farklı madde takımı tepki modelleri geliştirilmiştir ve maddeler arasında yerel bağımlılık söz konusu olduğunda, bu yerel bağımlılığı hesaba katan modellerin kullanılması gerekmektedir. Aksi takdirde yerel madde bağımlılığı göz ardı edilerek standart MTK modelleriyle elde edilen parametrelerde, test güvenilirliğinde ve test eşitleme sonuçlarında problematik durumlarla karşılaşılabilir.

D. Madde Takımlarına Dayalı Testlerde Kullanılan Modeller

Madde takımları, MTK'nin temel varsayımlarından birini bozması nedeniyle büyük ilgi gören bir konu olmuştur (Brennan, 2001). Madde takımı ile ilgili çalışmalarda, MTK'nin yerel bağımsızlık varsayımını iyileştirmenin yolları aranmış ve bu varsayımın ihlal edilmesi durumuyla baş etmek için madde takımı içeren testlere uygun ölçme modellerini araştırmışlardır. Madde takımı dayanaklı testler için geliştirilen MTK modellerinin bazıları tek boyutlu MTK ve bazıları da çok boyutlu MTK modellerinden geliştirilmiştir (Fukuhara & Kamata, 2011).

- *İki Kategorili MTK Modelleri*

Madde takımları etkisinin göz ardı edildiği ve madde takımlarında yer alan maddelerin birbirinden bağımsız maddeler gibi düşünüldüğü durumlar için Wainer ve Lewis (1990) uygun iki kategorili tepki modellerinden birinin kullanılmasını önermişlerdir. Bu modelde madde takımında yer alan maddelere, testteki herhangi bir madde gibi davranılır (Arora, 2007; Wainer vd., 2000) ve bağımsız birimler olarak puanlanır (Wainer vd.,2007). 1PLM'den, 2PLM'den ya da 3PLM'den uygun olan bir modelle veriler analiz edilir.

- *Çok Kategorili MTK Modelleri*

Madde takımlarında yer alan maddeler arasındaki yerel bağımlılığı iyileştirmek için kullanılan diğer modelde, madde takımında bulunan maddeler çok kategorili olarak puanlanan tek bir madde olarak kabul edilir (Cook, Dodd & Fitzpatrick, 1999; Sireci vd., 1991; Thissen vd., 1989; Wainer, 1995; Wainer vd., 2000; Yen, 1993). Bu duruma örnek olarak aynı okuma parçasına dayalı beş farklı madde düşünülebilir. Bu yaklaşımda bu beş madde ayrı ayrı puanlayarak iki kategorili MTK modelleriyle kalibre edilmez. Bunun yerine bu beş maddelik küme "tek bir madde" olarak ele alınır ve çok kategorili MTK modelleriyle incelenir. Böylece madde takımlarının çok kategorili puanlanması sonucunda madde takımından elde edilecek puan, maddelere verilen doğru yanıt sayısının toplamı kadardır (Aora, 2007; Wainer vd., 2000) ve madde takımı sıfırdan ortak uyarıcı ile ilişkili toplam madde sayısına kadar puan alabilmektedir (Boyd, 2003). Madde takımı puanı, doğru sayısıyla gösterilir (Wainer vd., 2000). Çünkü maddeler, madde takımları arasında bağımsızken madde takımları içerisinde bağımsız değildir (Wainer & Lewis, 1990) ve bu

puanlama yoluyla madde takımları içerisindeki bağımlılık ortadan kaldırılabılır (Dresher, 2002).

Madde takımlarına çok kategorili madde gibi davranılmasındaki temel düşünceyi O. Zhang (2010) şu şekilde ifade etmiştir: Madde takımlarındaki maddeler arasında aşırı bir yerel bağımlılık olsa da tüm madde takımı tek bir birim olarak kabul edildiğinde ve çok kategorili olarak puanlandığında, bu yerel bağımlılık ortadan kalkabilir. Puanlama sonucunda test kısmi bilgi modeli (partial credit model), aşamalı tepki modeli (graded response model), sınıflamalı tepki modeli (nominal response model) gibi çok kategorili maddeler için uygun olan MTK modellerinden biri ile kalibre edilir (DeMars, 2006).

Maddeler arasında yerel bağımlılık olduğu durumlarda bu yaklaşımla, test bilgisinin ve güvenilirliğin olduğundan fazla kestirilmesinin önüne geçilebilir ve böylece çok kategorili MTK modellerinden elde edilen istatistikler, standart MTK modellerinden elde edilen istatistiklerden daha iyi sonuçlar verir (O. Zhang, 2010). Çok kategorili MTK, madde takımı içeren veri setlerine uygulandığı zaman puanlama örüntüleri arasındaki farklılıklar görmezden gelinir. Çünkü bu yaklaşım her bir madde takımındaki toplam puana dayalıdır. Ancak madde takımındaki maddelerin toplanması, cevap örüntüleri hakkında bilgi vermediği için bir miktar bilgi kaybına neden olabilir ve bu bilgiye ihtiyaç duyulan durumlarda daha karmaşık modellerin kullanılması gerekir (Wainer vd., 2000; Yen, 1993). Ancak Thissen, Nelson, Rosa ve McLeod (2001) bu bilgi kaybının genellikle küçük miktarda olduğunu belirtmişlerdir. DeMars (2006) madde takımındaki madde sayısının az olduğu durumlarda farklı madde takımı puanları olarak kodlanan her bir tepki örüntüsünde sınıflamalı model kullanılabileceğini ve Thissen vd. (1989) ise böylece madde puanlarının toplanmasının sonucunda oluşan bilgi kaybının bir miktar giderilebileceğini belirtmişlerdir. Cook vd. (1999) ise aşamalı tepki modelini, kısmi puan modelini ve genelleştirilmiş kısmi puan modelini kullanarak yaptıkları çalışmada hiçbir modelin net bir avantajını görememişler ve model seçimini araştırmacıların teorik, deneysel ve pratik nedenlere dayandırmalarını önermişlerdir. Alan yazın incelendiğinde ise yapılan araştırmalarda, madde takımları için çok kategorili MTK modellerinden genellikle Bock'un (1972) geliştirdiği sınıflamalı tepki modelinin kullanıldığı görülmektedir (Sireci vd., 1991; Thissen vd., 1989; Wainer ve Lewis, 1990; Wainer & Thissen, 1996). Çünkü "madde takımı

puanları sınıflamalı (ya da daha çok yarı sıralı) tepkilerdir ve şans nedeniyle 1 puanı her zaman 0 puanından daha yüksek bir yeterlik göstermez” (Thissen vd., 1989, s. 259). Aşamalı tepki modeli de madde takımları içeren testlerin eşitlenmesinde uygun olabilir (Lee vd., 2001). Ancak yapılan çalışmalardan görüldüğü üzere, madde takımlarına çok kategorili MTK modellerinin uygulanması tepki örüntüsü ve madde-düzey ayırıcılığı (item-level discrimination) hakkında bilgi kaybına neden olduğu için kısıtlı bilgi sağlamaktadır (Thissen vd., 1989; Wainer & Wang, 2000; Wang & Wilson, 2005a, 2005b; Yen, 1993). Bu nedenle madde takımları için farklı modeller geliştirilmiştir.

- *İki Faktör Modeli*

Madde takımlarına verilen tepkiler, çok boyutlu model ile modellenen ve bu bağlam için iki faktör modeli uygun bir modeldir (DeMars, 2006). ÇBMTK modelinin özel bir hali olan bu modelde, her madde iki faktörde yük verecek şekilde sınırlandırılır. Faktörlerden biri, tüm maddelerin yük değerine sahip olduğu genel faktör ve diğeri ise grup-özellik faktörü (group-specific factors) olarak isimlendirilen özel faktördür (Houts & Cai, 2013). İki faktör modeli Gibbons ve Hedeker (1992) tarafından geliştirilmiştir ve öncelikle ikili puanlanan madde tepkilerine uygun bir çerçeveye sahipken sonraları model, Gibbons vd. (2007) tarafından çok kategorili maddelerin de analiz edilmesine olanak sağlayacak şekilde aşamalı tepki modeli için genişletilmiştir. Modelde parametre kestirimleri marjinal maksimum olabilirlik yöntemi kullanılarak yapılmaktadır (Gibbons & Hedeker, 1992; Zheng, 2013).

Gibbons ve Hedeker (1992) okuduğunu anlama testlerinde ilk faktörün hedef beceriyi belirttiğini ve ek faktörünse paragraf içindeki içeriğe ait alan bilgisini belirttiğini ifade ederek bunu iki faktör modeli şeklinde tanımlamışlardır. Bu durum maddelerin paragraflar arasında koşullu bağımsız, ancak paragraflar içinde koşullu bağımlı olduklarını gösterir (Gibbons & Hedeker, 1992). Burada genel faktör “okuduğunu anlama becerisi” iken, her paragraf da kendi özel faktörünü oluşturmaktadır (Houts & Cai, 2013). İki faktör modellerinde birincil bir faktörün yanı sıra bir ya da daha fazla ikincil bir faktör de bulunmaktadır. Bir temel faktörden ve $s - 1$ grup faktöründen oluşan s -faktörlü bir matris verildiğinde, iki faktör modeli her j maddesinin birincil boyutta (α_{j1}) ve $s - 1$ grup faktöründen daha fazla olmayan ikincil bir boyutta (α_{jh} , $h = 2, 3, \dots, s$) sıfırdan farklı bir yüke sahiptir (Gibbons & Hedeker, 1992). Kısaca ifade etmek gerekirse iki faktör modelinde her madde birincil faktörde

ve ikincil faktörlerin birinde sıfır olmayan bir faktör yüküne sahipken, diğer ikincil faktörlerde faktör yükü sıfırdır. Dört maddenin iki faktör örüntüsü Şekil 1.1'deki gibidir.

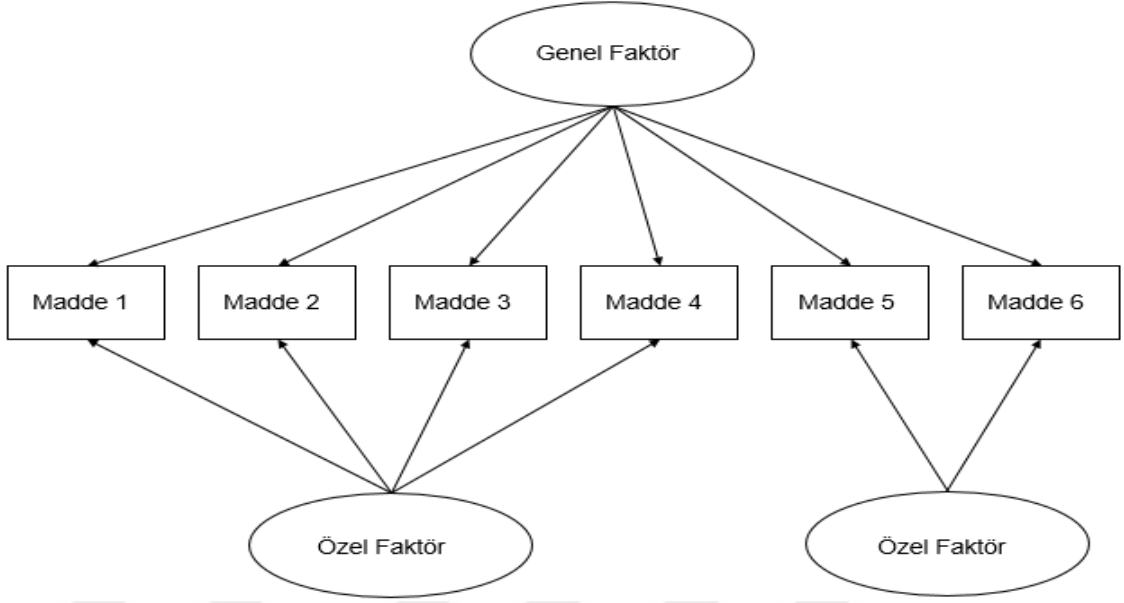
$$\alpha = \begin{vmatrix} \alpha_{11} & \alpha_{12} & 0 \\ \alpha_{21} & \alpha_{22} & 0 \\ \alpha_{31} & 0 & \alpha_{33} \\ \alpha_{41} & 0 & \alpha_{43} \end{vmatrix}$$

Şekil 1.1. İki Faktör Örüntüsü (Gibbons & Hedeker, 1992)

Bu yapı Holzinger ve Swineford (1937) tarafından “iki faktör” çözümü, Tucker (1958) tarafından bataryalar arası faktör analizi (inter-battery factor analysis) olarak adlandırılmıştır ve aynı zamanda Joreskog (1969) tarafından tartışılan doğrulayıcı faktör analizi modellerinden biridir (akt:Gibbons, Hedeker & Bock, 1990). Matriste gösterildiği gibi dört maddenin her biri birincil faktörde (α_{j1}) ve ikincil faktörlerden birinde ($\alpha_{j2}, \alpha_{j3}, \alpha_{j4}$) yük vermektedir. Örneğin, ilk iki madde yalnızca birincil boyutu değil aynı zamanda ilk alt boyutu, ikinci iki madde ise hem birincil boyutu hem de ikinci alt boyutu ölçmektedir. Faktör yükleri, madde eğim ya da ayırt edicilik parametresini göstermektedir (Cai vd., 2011). Bu modelde her madde yalnızca özellik alanından birinde ayırt edicilik parametresi sıfırdan farklı ($a_{j \in (s-1)} \neq 0$) iken diğer maddelerin ayırt edicilik parametreleri sıfıra ($a_{j \in (s-1)} = 0$) eşitlenir (Desa, 2012). Çokboyutlu MTK modellerinde matematiksel olarak faktör yükleri, madde ayırıcılık parametresine ve kesişimler ise madde güçlük parametresine dönüştürülür (McLeod, Swygert & Thissen, 2001).

İki faktör modelindeki ikincil faktörler, birinci faktörlere ve birbirlerine diktir (Gibbons & Hedeker, 1992). Bu model tanımlamasını yapmak için bütün boyutların ölçeğinin ve yerinin sabitlenmesi, ayrıca bütün boyutların ortalaması “0”, varyansı “1” olacak şekilde ayarlanması gerekmektedir (Rijmen, 2009). Bu ikincil faktörlerin birbirine dik olmasına, ikincil faktörler arasındaki kovaryansların (ya da korelasyonların) sıfır olduğu anlamına gelmektedir (Desa, 2012). Bu modelin faktörleri arası diklik varsayımı, geniş örneklem büyüklüğü gerektirmesi, en az üç alt alan faktörü gerektirmesi ve modeli tanımlanabilir yapmak için her alt alanda madde sayısının dengelenmiş olması gibi olumsuz yanları bulunmaktadır (Zheng, 2013).

İki faktör modelinin basit bir şekilsel gösterimi aşağıdaki gibidir.



Şekil 1.2. İki Faktör Modeli (Ip, 2010)

Şekil 1.2 incelendiğinde her maddenin birincil bir özelliğin yanı sıra ikincil bir özelliği de ölçtüğü görülmektedir. Ayrıca şekilden ikincil özelliklerin birincil özelliklere ve birbirlerine ortogonal olduğu da görülmektedir. Kısaca belirtmek gerekirse model her maddeyi (a) genel faktörde sıfır olmayan bir yük değerine, (b) özel faktörlerden ise yalnızca birinde ikincil yük değerine sahip olacak şekilde sınırlandırır. Ayrıca her özel faktör, genel faktöre ve diğer özel faktörlere ortogonaldir.

İki faktör ÇBMTK, madde takımları içeren testlere uyarlandığında her madde tepkisi birincil bir özelliğin ve olası bir madde takımı özelliğinin bir fonksiyonu olarak ifade edilmekte, aynı zamanda herhangi bir madde takımından bağımsız olan maddelere verilen tepkiler yalnızca birincil özelliğinin bir fonksiyonu olarak modellenenmektedir (DeMars, 2012). Burada birincil faktör/boyut temel ilgi alanına giren gizil değişken iken, ikincil faktör/boyut aynı kümeye ait maddeler arasındaki ek bağımlılıkları dikkate almak için dahil edilmiştir (Rijmen, 2010). Ayrıca bu model, madde takımı etkisinden kaynaklanan yerel madde bağımlılığını dikkate alabilmektedir (DeMars, 2006; Y. Li vd., 2006).

Madde takımları için bu model kullanıldığı zaman modeldeki ikincil özellikler, madde takımı etkisidir (testlet effect). LaFond (2014) iki faktör modelindeki faktör sayısının, özel faktörlere karşılık gelen madde grubu sayısının ve genel faktörün olduğunu belirtmiş; bu duruma örnek olarak, iki tane madde takımından oluşan bir testte bir

genel faktör ve iki özel faktör olmak üzere üç faktör olacağını söylemiştir. Bu modelde de aynı MTTK modelinde olduğu gibi maddeler madde takımları arasında koşullu bağımsız, ancak madde takımları içerisinde koşullu bağımlı olarak kabul edilir. Ancak iki faktör modeli ile MTTK modelleri arasındaki en temel fark, iki faktör modeli birincil ve ikincil (madde takımı) boyut için ayrı ayırıcılık parametresine izin verir ve bu ayırıcılık parametreleri birbirinden bağımsız olabilir. İki kategorili puanlanan maddeler için klasik 3PLM'nin iki faktör modeline genişletilmiş hali Eşitlik 1.10'daki gibi ifade edilebilir (Cai vd., 2011):

$$P(y = 1|\theta_0, \theta_s) = c + \frac{1 - c}{1 + \exp\{-(d + a_0 \theta_0 + a_s \theta_s)\}} \quad (1.10)$$

Eşitlik 1.10'da θ_0 genel faktör, θ_s özel faktör, c şans başarısı, d madde kesişim parametresi, a_0 genel faktöre ait madde eğim parametresi ve a_s özel faktöre ait madde eğim parametresidir. Eğer madde özel faktörde bir yüke sahip değilse o zaman a_s sıfır olur ve koşullu olasılık θ_s 'ye bağlı olmayacağı için model klasik 3PLM halini alır (Cai vd., 2011). 'd' parametresi, b madde güçlük ve a madde ayırt edicilik parametresi kullanılarak ve m boyut sayısını göstermek üzere Eşitlik 1.11'deki gibi hesaplanır (Reckase, 2009):

$$d_i = -b \sqrt{\sum_{v=0}^m a_v^2} \quad (1.11)$$

Daha sonraları iki faktör modeli çok kategorili puanlanan maddeler için de formüle edilmiştir. Cai vd. (2011), Thissen, Cai ve Bock'un (2010) G-KPM için yazdıkları eşitliği iki faktör modeli için G-KPM'ye uyarlayarak Eşitlik 1.12'deki gibi ifade etmişlerdir.

$$P(y = k|\theta_0, \theta_s) = \frac{\exp\{T_k[a_0\theta_0 + a_s\theta_s] + d_k\}}{\sum_{l=0}^{K-1} \exp\{T_l[a_0\theta_0 + a_s\theta_s] + d_l\}} \quad (1.12)$$

$k = 0, \dots, K - 1$ kategorisi için koşullu tepki olasılığını gösteren eşitlikte T_k : k kategorisi için sözde puanlama fonksiyonu, d_k : kategori kesişimi, a_0 ve a_s sırasıyla genel ve özel boyut için madde eğimidir. Ayrıca puanlama fonksiyonu $T_k = 0, \dots, K - 1$ olarak alınabilir.

DeMars (2006) yaptığı çalışmada, aralarında yerel bağımlılık olan madde takımları içeren testlerden elde edilen veri setlerinin, madde takımları için iki faktör ÇBMTK

modeline, MTK modelinden daha iyi uyum sağladığını bulmuşlardır. Ayrıca Chen ve arkadaşları (2012) yaptıkları araştırmada karmaşık psikolojik yapıların gösteriminde iki faktör modelinin daha gerçekçi olduğunu ifade etmişlerdir.

- *Madde Takımı Tepki Kuramı Modelleri*

Madde takımları içeren testlerin analizinde kullanılmak için geliştirilen bir diğer modelse madde takımı etki modelidir (testlet effect model). Bu model tesadüfi madde takımı etkisini içerir ve iki faktör modelinin sınırlandırılmış halidir (Y. Li, vd. 2006; Rijmen, 2009). Madde takımı için geliştirilen MTK modellerinde her madde takımı boyunca, her birey için yerel bağımlılık miktarını ifade etmek amacıyla geleneksel MTK parametrelerine ek olarak bir madde takımı parametresi bulunmaktadır (Chien, 2008; Y. Li vd., 2005). Bu bağımlılığın bir madde takımına özgü olduğu varsayılır ve test ile ölçülmek istenen asıl özellikten farklı olan ikinci bir boyut olarak düşünülür (Bao, 2007). Geliştirilen bu kuram Wainer vd. (2000) tarafından kısaca Madde Takımı Tepki Kuramı - MTTK (Testlet Response Theory-TRT) olarak adlandırılır. Orijinal MTK modellerine kişiye özgü tesadüfi madde takımı etkisinin (the person-specific random testlet effect) eklenmesi yöntemine MTTK denilmektedir (Wainer vd., 2000; Bradlow vd., 1999; Chien, 2008; Fukuhara & Kamata, 2011; Y. Li vd., 2006; Wainer & Wang, 2000; Wang vd., 2002).

Madde Takımı Tepki Kuramı, var olan madde tepki kuramı modellerinin genelleştirilmiş halidir ve standart MTK modellerinden üç yönüyle farklılık göstermektedir.

MTTK'de standart MTK modellerine, bir madde takımı içerisindeki maddeler arasında paylaşılan varyansı hesaba katan tesadüfi etki parametresi eklenmiştir (Wainer vd., 2007) ve maddeler arasındaki yerel bağımlılık bu parametre yardımıyla ele alınabilmektedir. Ayrıca bu model yalnızca madde takımlarındaki yerel bağımlılığı hesaba katmakla kalmayıp, madde takımında yer alan maddelerin madde takımı faktöründen nasıl etkilendiğine dair daha fazla bilgi verdiği için yerel bağımlılığı göz önüne alan ve madde takımı etkisinin çalışıldığı araştırmalar için umut verici bir model olarak görülmektedir (Y. Li, S. Li & Wang, 2010).

MTTK modelleri, daha çok iki kategorili MTK modellerine madde takımlarında yer alan maddeler arasındaki paylaşılan varyansın hesaba katılmasını sağlayan tesadüfi etki parametresinin eklenmesiyle düzenlenmiş modellerdir (Boyd, 2003).

Yani bu modelde test ile ölçülen gizil yeteneğin yanı sıra bir tesadüfi etki olarak, madde takımı etkisi (γ) de parametre şeklinde bulunmaktadır (Fukuhara & Kamata, 2011). MTTK modellerinin en önemli özelliği madde-birey etkileşiminin de modelde bir parametre olarak ele almasıdır. Birey-madde etkileşimi parametresi madde takımı içerisindeki bütün maddelerde ortak iken, madde takımları arasında farklıdır (Dresher, 2002). Ayrıca bu tesadüfi etkinin varyansındaki büyüklük, hangi madde takımında yer alan maddelerin yerel bağımlı olduğunun belirlenmesini sağlar (Wainer & Wang, 2000).

MTTK'de yer alan b , a , ve c parametreleri iki kategorili MTK modellerindeki parametrelerdir ve aynı şekilde yorumlanır. Bu parametrelere ek olarak madde takımı etkisini gösteren γ parametresini içerir. MTTK'de kullanılan madde takımı etki parametresi, ortalaması sıfır ve standart sapması testteki madde takımı etkisinin varyansının kareköküne eşit olan normal dağılımdan seçilen tesadüfi bir değişkendir (Boyd, 2003). Bu parametre sıfır olarak alındığında madde takımı modelinden elde edilen sonuçlar, standart MTK modelleriyle aynı olmaktadır (Kim, De Ayala, Ferdous & Nering, 2007). MTTK kullanılarak yapılan analizlerde model tanımlamaları için ikincil madde takımı boyutunun ölçeği sabitlenmezken, iki faktör modelinin diğer kısıtlamaları (bütün boyutların yeri ve genel boyutun ölçeği sabitlenir) madde takımı modeli için de devam etmektedir (Rijmen, 2009). Modelin en genel halinde güçlük (b), ayırıcılık (a) ve şans (c) parametresi olmak üzere üç madde parametresi ve madde takımı etkisi (γ) ve yetenek parametresi (θ) olmak üzere iki tane de bireye özgü parametre bulunmaktadır.

Madde takımı etki parametresi varyansının madde takımları arasında farklılaşmasına izin verilir ve her madde takımındaki yerel bağımlılık miktarını gösterir. Bu nedenle madde takımı etkisinin varyansı, yerel madde bağımlılığının bir ölçüsü olarak da kullanılabilir. Madde takımı etki parametresinin varyans kestirimi sıfırsa madde takımları içindeki maddelerin bağımsız olduğu, bu parametrenin varyans değeri arttıkça madde takımları arasındaki maddelerin bağımlı maddeler olduğu düşünülebilir ve yerel bağımlılık miktarı, bu parametrenin varyansındaki artış olarak büyür (Bao, 2007; Boyd, 2003).

MTTK kişiye özgü tesadüfi madde takımı etkisini doğrudan ele alan farklı madde takımı modelleri içermektedir (Bradlow vd., 1999; Wainer vd., 2000; Wang vd., 2002; Y. Li vd., 2006). Bütün MTTK modelleri geleneksel MTK parametrelerinin yanı sıra

her birey için YMB miktarını ifade eden bir madde takımı parametresi içerir. Bu modeller tek boyutlu ve çok boyutlu MTK modellerinin düzenlenmiş halleridir.

2PL-MTTK. Madde takımına dayanan testlerin incelenmesinde geleneksel MTK modelleri uygun olmadığı için iki parametrelili lojistik modelinin yerel madde bağımlılığını da hesaba katmasıyla genişletilmiş bir hali olan iki parametrelili madde takımı tepki modeli önerilmiştir (Bradlow vd., 1999). Aynı zamanda 2PL-MTTK modeli, iki faktör modelinin de sınırlandırılmış halidir (DeMars, 2006; Y. Li vd., 2006). Bu modelde, bir madde içerisinde yuvalanmış maddeler için tesadüfi etki parametresi bulunmaktadır. 2PL-MTTK modelinde $d_{(i)}$ d madde takımına yuvalanmış i maddesinde, θ_j yetenek düzeyli, j kişinin başarılı olma olasılığı ($x = 1$) Eşitlik 1.13'teki gibi gösterilmektedir.

$$P_{ij}(x_i = 1|\theta_j) = \frac{\exp(a_i(\theta_j - b_i - \gamma_{jd(i)}))}{1 + \exp(a_i(\theta_j - b_i - \gamma_{jd(i)}))} \quad (1.13)$$

$\gamma_{jd(i)}$ j kişinin d madde takımına yuvalanmış i maddesindeki madde takımı etkisi parametresini, b_i i maddesinin güçlük parametresini ve a_i i maddesinin ayıricılık parametresini gösterir.

Ancak Bradlow vd. (1999) 2P-MTTK modelinin bazı sınırlılıkları olduğunu belirtmişlerdir. Bu sınırlılıklardan ilki modelin şans parametresi içermemesidir. Oysa geniş ölçekli testlerde, şans parametresinin büyük ölçüde etkili olduğu çoktan seçmeli maddelerin kullanılması daha yaygındır. Ayrıca açık uçlu maddelerin yer aldığı sınavlarda bu maddeler çok kategorili olarak puanlanacağı için çok kategorili modeller ya da karışık formatlı testler için modeller kullanılmalıdır. 2PL-MTTK modeli ise yalnızca ikili puanlanan maddelere uygulanabilmektedir. Bu modelin bütün madde takımları boyunca sabit bir varyans olduğu varsayımı da modelin bir diğer sınırlılığıdır. Çünkü farklı büyüklükte yerel bağımlılık içeren farklı madde takımlarında bu varyans değişebilir.

Y. Li vd. (2006) MTTK'nin bu yönden sınırlı olduğunu düşünerek bir çalışma yapmışlardır. Hem birincil özelliğe ve hem de madde takımı faktörüne (ikincil özellik) aynı ayıricılığın verilmesini uygun görmedikleri için yeni bir model önermişlerdir. Bu model 2PL-MTTK modeli ile iki parametrelili normal ogive modelini birleştirerek, bireylerin yeteneği ve madde takımı etkisi için ayıricılık parametresinin farklılaşmasına izin veren bir modeldir. Araştırmacılar çalışmalarında bu modelin

hem simülasyon verisine hem de gerçek veriye iyi uyum gösterdiğini belirlemişlerdir. Y. Li vd. (2006) ve Rijmen (2009) yaptıkları çalışmalarda 2PL-MTTK'nin iki faktör modelinin sınırlandırılmış hali olduklarını belirlemişlerdir. Bu modelin Eşitlik 1.14'teki gibi formüle edilebilir (Y. Li vd. 2006):

$$P(y_{ij} = 1) = \Phi(a_{i1}\theta_j - t_i + a_{i2}\gamma_{d(i)j}) \quad (1.14)$$

$P(y_{ij} = 1)$ i maddesini j bireyinin doğru cevaplama olasılığı, Φ bir standart normal dağılımın yığılmalı dağılım fonksiyonu, θ_j j bireyinin yeteneği, $\gamma_{d(i)j}$ i maddesini içeren d madde takımı ile ilgili olan ikincil bir boyutu; t_i madde güçlüğüyle ilişkili olan yer parametresini, a_{i1} ve a_{i2} sırasıyla θ ve γ_d 'ya ait ayırt edicilik gücünü göstermektedir. θ_j ve $\gamma_{d(i)j}$ parametrelerinin ortalaması ve varyansı, bu parametrelerin ilişkisiz olduğunu göstermek amacıyla $N(0, 1)$ olarak sabitlenmelidir. Daha sonraları bu eşitlik çok kategorili puanlanan maddeleri de göz önüne almak amacıyla G-KPM'ye (Muraki, 1992) uyarlanarak madde takımı modelleri için genişletilmiştir. Bu genişletilmiş madde takımı modeli Eşitlik 1.15'teki gibi ifade edilebilir (Y. Li vd., 2010):

$$P_{ijk} = \frac{\exp[\sum_{v=0}^k (a_{i1}\theta_j - t_{iv} + a_{i2}\gamma_{d(i)j})]}{\sum_{c=0}^{m_i} \exp[\sum_{v=0}^c (a_{i1}\theta_j - t_{iv} + a_{i2}\gamma_{d(i)j})]} \quad (1.15)$$

Eşitlik 1.15'te Eşitlik 1.14'ten farklı olan parametreler P_{ijk} j kişinin i maddesinin $m_i + 1$ puan kategorisinin k kategorisindeki puan olasılığı, t_{iv} i maddesinin v puan kategorisi için güçlük parametresi anlamına gelmektedir. Yine θ_j ve $\gamma_{d(i)j}$ parametrelerinin ortalaması ve varyansı, bu parametrelerin ilişkisiz olduğunu göstermek amacıyla $N(0, 1)$ olarak sabitlenmelidir.

3PL-MTTK. Birnbaum (1968) tarafından geliştirilen 3PL-MTK modelin geliştirilmiş hali olan ve en genel madde takımı modeli olan üç parametrelili lojistik MTTK modeli, Wainer vd. (2000) tarafından tanıtılmıştır. 3PL-MTTK modelinde $d(i)$ madde takımına yuvalanmış i maddesinde, Θ_j yetenek düzeyli, j kişinin i maddesinde başarılı olma olasılığı ($x = 1$) Eşitlik 1.16'daki gibi gösterilmektedir.

$$P_{ij}(x_i = 1 | \theta_j) = c_i + (1 - c_i) \frac{\exp(a_i(\theta_j - b_i - \gamma_{jd(i)}))}{1 + \exp(a_i(\theta_j - b_i - \gamma_{jd(i)}))} \quad (1.16)$$

j kişinin $d(i)$ madde takımına yuvalanmış i maddesindeki $\gamma_{jd(i)}$ madde takımı etkisi parametresini, b_i i maddesinin güçlük parametresini, a_i i maddesinin ayırıcılık parametresini ve c_i şans parametresini gösterir.

3PL-MTTK'yi 2PL-MTTK'den ayıran iki önemli özellik bulunmaktadır (Chen, 2010): Birincisi 3PL-MTTK modeli daha fazla parametre içerdiğinden, bu modelle yapılacak kestirimler için daha etkili hesaplama algoritmaları geliştirilmesi gerekmektedir. İkincisi ise 2PL-MTTK'nin aksine madde takımı etkisinin varyansı madde takımları arasında değişmektedir. Ancak $c_i = 0$ ve her madde takımı için $\sigma_{\gamma d(i)}^2 = \sigma_{\gamma}^2$ olduğunda 2PL-MTTK modeli 3PL-MTTK modelinin basitleştirilmiş bir versiyonudur.

Daha sonra Wang ve arkadaşları (2002) tarafından iki kategorili veriler için geliştirilen 3PLM (Birnbaum, 1968) ve Samejima (1969) tarafından geliştirilen çok kategorili madde tepki kuramı modeli kullanılarak, 3PL-MTTK modeli genişletilerek hem iki kategorili hem de çok kategorili tepkilerin yer aldığı karışık veri setlerine de uygun hale getirilmiştir.

1PL-MTTK. Eğer $c_i = 0$ ve $a_i = 1$ ise 3PL-MTTK modeli madde takımı modellerinin Rasch versiyonu olarak basitleştirilebilir (1PL-MTTK, Wang ve Wilson, 2005a). Bu modelde de 3PL-MTTK'deki gibi madde takımı etkisinin varyanslarının, madde takımları boyunca farklılaşmasına izin verilmektedir. Modelde $d(i)$ madde takımı üzerinde, Θ yetenek düzeyli, j kişinin i maddesinde başarılı olma olasılığı ($x = 1$) Eşitlik 1.17'deki gibi gösterilmektedir.

$$P_{ij}(x_i = 1|\theta_j) = \frac{\exp(\theta_j - b_i - \gamma_{jd(i)})}{1 + \exp(\theta_j - b_i - \gamma_{jd(i)})} \quad (1.17)$$

j kişinin $d(i)$ madde takımına yuvalanmış i maddesindeki $\gamma_{jd(i)}$ madde takımı etkisi parametresini ve b_i i maddesinin güçlük parametresini gösterir.

Madde Takımlarında Kullanılan Modellerin Özeti

Özet olarak, kullanılan yöntemler üç başlık altında toplanabilir:

Madde takımı içerisinde yer alan maddeler, bağımsız maddelermiş gibi ele alınabilir. Bu yöntemde, ortak bir uyarıcıya dayanan madde takımları arasındaki yerel bağımsızlık görmezden gelinerek, bu maddelere de diğer maddeler gibi davranılır. Veri setine uygun olan standart MTK modellerinden biri kullanılarak analiz edilir.

Ancak bu yaklaşım, madde ve yetenek parametrelerinde, test bilgisinde ve testin güvenilirliğinde hatalı sonuçlara yol açması nedeniyle eleştirilmektedir.

Madde takımlarına çok kategorili bir madde gibi davranılabilir. Bu yöntemde, madde takımı çoklu puanlanan çok kategorili tek bir madde gibi kabul edilir. Madde takımında yer alan maddelere ayrı ayrı puan verilir ve doğru sayısı, madde takımından elde edilen puan olarak kabul edilir. Daha sonra çok kategorili MTK yöntemlerinden biri kullanılarak veri analiz edilir. Ancak çok kategorili MTK'nin madde takımları için kullanılması, bilgi kaybına neden olması açısından yetersiz görülmektedir.

Madde takımı etkisini açıkça ele alan yöntemler de şu şekilde ifade edilebilir: Yöntemlerden biri yerel madde bağımlılığının tesadüfi bir değişken olarak ele alınması ve bu değişkenin TBMTK modellerine eklenmesi; bir diğeri ise yerel madde bağımlılığına sebep olan madde takımı etkisinin ikinci bir boyut olarak ele alınmasıdır. Ancak bu yöntemler de diğere yöntemlere kıyasla daha karmaşık ve zor olduğu için araştırmacılar tarafından çok sık kullanılmamaktadır. Ayrıca MTTK modelleri yüksek düzeylerin sınıflandırılmasından kaynaklanan bağımlılığı hesaba katmamaktadırlar (Chen, 2010).

2. İLGİLİ ARAŞTIRMALAR

Bu bölümde alan yazındaki benzer çalışmalara yer verilmiştir.

2.1. Yerel Madde Bağımlılığı ile İlgili Çalışmalar

Yen (1984) yerel madde bağımlılığının, uyum ve eşitleme performansına etkisini incelemek için hem gerçek veri hem de simülasyon verisi kullanarak madde ve birey parametrelerini kestirmiştir. Bu amaçla 3PL modeli kullanarak tek boyutlu simülasyon verisi, Reckase ve McKinley (1983) tarafından tanımlanan iki boyutlu 3PL modeli kullanılarak da çok boyutlu veri üretilmiştir. Gerçek veri seti ise üçüncü ve altıncı sınıf öğrencilerinin matematik başarılarını ölçen üç testten elde edilmiştir. Araştırmacı, simülasyon verisi için aynı faktörden ya da gerçek veri seti için benzer içerikten etkilendiğinde, yerel bağımlılık ölçüsünün (Q_2 ve Q_3) maddelerin oluşturduğu alt kümeleri belirlemede etkili olduğunu ifade etmiştir. Araştırmanın sonucunda yüksek Q_2 ve Q_3 değerleri bulunan madde çiftlerinin madde parametrelerinin benzer olma eğiliminde olduğu, ancak benzer madde parametrelerine sahip maddelerin çoğunun yüksek Q_2 ve Q_3 değerlerine sahip olmadığı belirlenmiştir. Çalışmada testin geri kalan maddelerinden bağımsız olan yerel bağımlı maddelerin, güçlüklerinin ya da ayırt ediciliklerinin beklenmedik düzeyde yüksek ya da düşük olmadığı görülmüştür. Ayrıca farklı boyutların toplanmasını gerektiren testlerin eşitlenmesinde sistematik olmayan eşitleme hataları olduğu ifade edilmiştir.

Ackerman (1987) Spray ve Ackerman (1987) tarafından tepki bağımlılıklarında farklı dereceli veri üreten bir model kullanarak simülasyon verisi oluşturmuştur Bu veri seti için 2PLM ile madde ve yetenek parametrelerini kestirmiştir. Araştırmanın bulguları yerel madde bağımlılığının göz ardı edilmesinin, madde ayırt edicilik parametrelerinde madde güçlük parametrelerine oranla daha şişirilmiş sonuçlar vermesine neden olduğunu göstermiştir. Ayrıca düşük yetenekli bireyler için yetenek kestiriminin olduğundan yüksek ve yüksek yetenekli bireyler içinse yetenek kestiriminin olduğundan düşük kestirildiği, yani yetenek kestirimlerinde yanlış sonuçlarla ulaşıldığı belirlenmiştir.

Yen (1993) yerel madde bağımlılığı etkisinin görmezden gelinmesini iki test kullanarak incelemiştir. Hangi maddelerde yerel madde bağımlılığı olup olmadığını

belirlemek için Yen'in Q_3 istatistiğini kullanmıştır. Testlerin ikisi de üç model kullanılarak analiz edilmiştir. İlk olarak yerel madde bağımlılığı görmezden gelinmiş ve bütün maddelere bağımsız maddelermiş gibi davranılmıştır (Items condition). İkinci olarak yerel madde bağımlılığı uygun olarak modellenmiş (LID Testlets condition) ve son olarak ise bağımsız maddeler madde takımı olarak gruplandırılmış ve yerel madde bağımlılığı modellenmiştir (Non-LID Testlets condition). Araştırmanın sonucunda yerel madde bağımlılığı hesaba alınmadığında ölçmenin standart hatasının olduğundan daha düşük ve test bilgisinin olduğundan daha yüksek kestirildiği belirlenmiştir. Üçüncü model bağımsız maddelerden bir madde takımı oluşturmanın test bilgisi üzerinde hiçbir etkisi olmadığını göstermiştir. Ayrıca yerel madde bağımlılığının görmezden gelinmesinin madde ayırıcılık parametresinin ortalamadan %60 fazla kestirilmesine neden olduğu belirlenmiştir.

Reese (1995) Law School Admission Test (LSAT) verilerini kullanarak MTK'nin yerel bağımsızlık varsayımının farklı derecelerde bozulması durumunun, kalibrasyon ve puan dağılım kestirimleri üzerindeki etkisini araştırmıştır. Araştırmada LSAT'ın ve başka iki testin (Pre-American College Test Plus, P-ACT+ ve Graduate Management Admission Test, GMAT) verileri, yerel madde bağımlılığının yaklaşık değerini belirlemek için kullanılmıştır. Bu amaçla araştırmacı Yen'in Q_3 istatistiğini hesaplamış, yerel madde bağımlılığı için dört düzey belirlemiş (sıfır, düşük, orta ve yüksek) ve LSAT madde parametrelerini kullanarak her düzey için 101 maddelik bir veri seti üretmiştir. Madde parametreleri ile ilgili sonuçlar ışığında ayırt edicilik parametresinin yüksek düzeyde yerel bağımlılık olduğunda fazla kestirildiği belirlenmiştir. Özellikle gerçek parametreler ile a parametre kestirimi arasındaki korelasyonun yüksek düzeyde yerel bağımlılıkta düşük değer alırken diğer düzeylerde oldukça yüksek değerler aldığı görülmüştür. Güçlük parametresi kestirimlerinin ise YMB düzeyi arttıkça düştüğü, ancak gerçek ve kestirilen güçlük parametreleri arasında bütün düzeylerde yüksek korelasyon bulunduğu saptanmıştır. Şans başarısı parametresi incelendiğinde de b parametresine benzer şekilde YMB düzeyi arttıkça kestirimlerin düştüğü belirlenmiştir. Ancak gerçek ve kestirilen c parametreleri arasındaki korelasyon sıfır, düşük, orta düzeyde YMB varken oldukça yüksek iken yüksek düzeyde YMB olduğunda korelasyon değerinde epey bir düşüş olduğu görülmüştür. Çalışmanın bulguları ışığında yüksek düzeyde YMB'nin a parametresinin olduğundan yüksek, c parametresinin olduğundan düşük

kestirilmesine neden olduğu ve b parametresine ise oldukça az etkisi olduğu sonucuna varılmıştır. Ayrıca Reese simülasyon verisini ilgili LSAT verileriyle, YBM etkisi ve MTK kalibrasyon ve puan dağılım kestirimleri açısından karşılaştırılmıştır. Sonuçlar, maddeler arasında yerel bağımlılık olduğunda düşük puanların olduğundan düşük, yüksek puanlarınsa olduğundan yüksek kestirildiğini göstermiştir. Yalnızca yüksek düzeyde YMB olduğunda, YMB'nin madde ve yetenek parametreleri ile test karakteristik eğrisine problem yaratacak şekilde etki ettiği sonucuna varılmıştır. Reese özellikle yüksek düzeyde YMB olması durumunun madde ayırıcılık gücüne ve testin tamamının güvenilirliğine etkisinin büyük olduğunu ve bu değerlerin olduğundan fazla kestirildiğini belirlemiştir.

Tuerlinckx ve De Boeck (2001) yaptıkları çalışmada 2PL modelini kullanarak ayırt edicilik parametresi kestirimlerine, maddeler arası etkileşimin görmezden gelinmesinin etkisini incelemişlerdir ve yerel madde bağımlılığının madde ayırt edicilik parametresi üzerinde etkisi olduğunu matematiksel ve deneysel olarak göstermişlerdir. Araştırmacılar kendi simülasyon çalışmalarının sonuçlarına dayanarak üç önemli iddiada bulunmuşlardır. Bu iddialardan ilki bir maddenin ayırt edicilik kestiriminin yalnızca etkileşim içinde bulunduğu diğer maddelerin zorluğundan etkilendiği, maddenin kendi zorluğundan etkilenmediğidir. İkinci olarak testteki madde sayısını gereğinden fazla arttırmak, ayırıcılık parametresine ait ortak etkinin azalmasına neden olduğunu iddia etmişlerdir. Üçüncü iddia ise pozitif etkileşim ayırt edicilik parametresinin olduğundan fazla, negatif etkileşim ise bu parametrenin olduğundan düşük kestirilmesine neden olduğudur. Araştırmacılar eğer negatif yerel madde bağımlılığının modellenmediğinde, bağımsız maddelerin ayırt edicilik parametresinin olduğundan fazla kestirildiğini belirlemişlerdir. Aynı zamanda ayırt edicilik parametresi maddelerin güçlüğüne bağlı olduğundan etkileşime girdikleri, ancak maddenin kendi zorluğundan etkilenmediği belirlenmiştir.

Chen ve Wang (2007) maddeler arası etkileşimin görmezden gelinmesinin madde parametreleri üzerine etkisini ve yerel bağımlılık için Q_3 istatistiğinin kullanımının etkililiğini araştırmışlardır. Araştırmada iki kategorili maddeler için 3PLM ve çok kategorili maddeler için genelleştirilmiş kısmi puan modeli kullanılmıştır. Çalışmanın amacı doğrultusunda 3PLM için 1000 kişilik 50 madde, G-KPM için ise 1000 kişilik 30 madde simule edilmiştir. Simülasyon sonuçlarında pozitif madde etkileşiminin ayırt edicilik parametresini olduğundan daha fazla, güçlük parametresinin

olduğundan daha düşük ve Q_3 istatistiğinin sıfırdan daha küçük kestirilmesine yol açtığı belirlenmiştir. Ayrıca ayırt edicilik parametresindeki fazla kestirimin, şans parametresi sıfıra yaklaştığında ciddi sorunlara neden olduğu görülmüştür. Bu durumların tersine negatif madde etkileşiminin görmezden gelinmesi ayırt ediciliğin olduğundan düşük, güçlüğün olduğundan fazla ve Q_3 istatistiğinin sıfırdan daha büyük kestirilmesine yol açtığı belirlenmiştir. Ayırt edicilik parametresindeki düşük kestirim de şans parametresinin sıfıra yaklaşırken daha az sorun yarattığı bulunmuştur. Şans parametresi ile ilgili bulgular ışığında, düşük bir c-parametresinin yüksek bir c-parametresine göre a parametresini daha çok fazla kestirdiği sonucuna varılmıştır. Aynı zamanda araştırmacılar, ayırt edicilik parametresindeki yanlış kestirimlerin madde niteliğini hakkında yanlış değerlendirmelere yol açabileceğini ifade etmişlerdir.

Monseur ve arkadaşları (2011) yaptıkları çalışmada PISA 2000'de (okuma) ve PISA 2003'te (matematik) yer alan bilişsel verileri kullanarak, maddeler arasındaki yerel bağımlılığı, Yen'in Q_3 istatistiği aracılığı ile incelemişlerdir. Bazı okuma ve matematik becerilerini ölçen madde çiftleri arasında önemli madde bağımlılıkları belirlemişlerdir. Ancak matematik becerilerini ölçen maddelerde, okuma becerilerini ölçen maddelere göre daha yüksek yerel madde bağımlılık bulunmuştur. Daha sonra yerel madde bağımsızlığı varsayımının bozulmasının öğrenci performansları üzerindeki etkisini incelemişlerdir. YBM öğrencilerin yeterliklerinin değişkenliği üzerinde etkisi olduğu ve kestirimlerdeki değişkenliğin yanlılığının, ortalama ülke performansı ile güçlü bir şekilde ilişkili olduğu belirlenmiştir. Bu nedenle PISA verilerindeki yerel bağımsızlık varsayımının bozulması sonucunda, düşük performanslı ülkelerin görece değişkenliğinin olduğundan yüksek, yüksek performanslı ülkelerinkinin ise olduğundan düşük kestirildiği ifade edilmiştir. Ayrıca araştırmanın sonuçları "madde takımı etkisinin" olduğunu göstermiştir ve bu nedenle araştırmacılar YMB olduğu durumlarda tesadüfi etki yaklaşımının kullanılması gerektiğini belirtmişlerdir.

2.2. Madde Takımları İçeren Testlerin Farklı Madde Tepki Kuramı Modelleri ile Ele Alındığı Çalışmalar

Thissen vd. (1989) yaptıkları çalışmada, 3866 bireyin 22 maddelik bir okuduğunu anlama testine verdikleri yanıtları kullanmışlardır. Bu 22 maddenin boyutluluğunu belirlemek için faktör analizi (Bock, Gibbons, & Muraki, 1988) yapılmış ve dört faktör

belirlenmiştir. Araştırmacılar çok boyutlu bu yapının, yerel madde bağımsızlığı varsayımını bozacağını ve TBMTK modellerinin bu veri setinin analizinde kullanılmasının uygun olmayacağını düşünmüşlerdir. Bu nedenle dört farklı okuduğunu anlama parçasının madde takımı olarak ele alınması durumu için sınıflamalı model (Bock, 1972) kullanarak Bayes modeli ölçek puanlarını hesaplamışlardır. Test bilgisi için 3PL ve sınıflamalı tepki modeli kullanılmıştır. Araştırmada madde takımı yapısı olduğunda geleneksel modellerin çok kategorili MTK modellerine göre test bilgisini daha fazla kestirdiğini belirlemiştir. Araştırmacılar yaptıkları çalışma sonucunda, madde takımı düzeyinde analizi kolaylaştırmak için çok kategorili MTK modelleriyle madde takımlarının puanlanmasını önermişlerdir.

Sireci vd. (1991) madde takımı dayanaklı testlerin güvenilirlik hesaplamalarının nasıl olacağını göstermek amacıyla madde takımlarından oluşan iki tane okuduğunu anlama testinden elde edilen puanlar üzerinden bir çalışma yapmışlardır. Araştırmada maddeleri bağımsız maddeler gibi puanlayarak, geleneksel güvenilirlik katsayısı olan Cronbach alfa güvenilirlik katsayısı ve 3PLM ile Bock'un (1972) sınıflamalı tepki modelini kullanarak MTK dayanaklı marjinal güvenilirlik katsayısı hesaplanmıştır. Araştırmadan elde edilen bulgular, maddelerin birbirine bağımlılıklarının göz önünde bulundurulmadığı durumda elde edilen Cronbach alfa güvenilirlik katsayılarının, marjinal güvenilirlik katsayılarından daha yüksek kestirildiğini ortaya koymaktadır. Bu bulgu, güvenilirlikte görülen bu büyüklüğün test uzunluğunun iki katına çıkarılması durumunda gerçekleşebileceğini göstermiştir. Araştırmacılar madde takımı içeren testlerdeki maddelere bağımsız maddeler gibi davranılmasının yüksek oranda yanıltıcı sonuçlar verebileceğini; güvenilirlikteki bu fazla kestirimin görmezden gelmek için çok büyük olduğunu ve madde takımı içeren testlerde madde analizini geleneksel test teorileriyle yapmak mümkün olsa da madde takımının madde analizlerinin MTK'ye dayanan yöntemlerle yapılmasının daha yararlı olacağını belirlemiştir.

Thissen vd. (2001) 1991 yılında yapılan North Carolina End-of-Course Geometry testini alan 2739 bireyin verilerini kullanarak bir araştırma yapmışlardır. Araştırmacılar, 60 maddelik bu testi kalibre etmişler ve her bir madde çifti için yerel bağımlılık indeksini (LD-index) hesaplamışlardır. Hesaplanan indekslere göre iki madde arasında yerel bağımlılık belirlenmiştir. Bu maddeler tek bir madde takımı

olarak kabul edilmiş ve dört yanıt kategorisi ($\{00\}$, $\{01\}$, $\{10\}$, $\{11\}$) olacak şekilde yeniden kalibre edilmiştir. Daha sonra bu 58 madde 3PL model ile madde takımı ise sınıflamalı model ile ele alınmıştır. Araştırmanın sonuçlarına göre araştırmacılar şu bilgilere ulaşmışlardır: Eğer ölçek puanı 3PL kullanılarak hesaplanırsa iki maddeden birine doğru yanıt veren bireyler, hiçbir maddeye doğru yanıt veremeyen bireylerden daha yüksek puan almaktadır. Ancak madde takımı ölçek puanlamasında, sınıflamalı modelin kullanılması ise farklı sonuçlara neden olmaktadır: $\{11\}$ yanıt örüntüsü için bireyin ölçek puanı artarken, diğer örüntüler için ölçek puanı düşme eğilimindedir ve bu nedenle bu madde çiftleri için “bu madde çifti tek bir maddedir, eğer her iki soruyu da doğru yanıtlarsan bir puan alabilirsin ve her ikisini de yanlış yanıtlarsan sıfır puan alabilirsin” gibi bir kuralın belirlenmesi gerekmektedir. Bu MTK analizleri sonucunda araştırmacılar, geometri yeterliğini belirlemede bu puanlama kuralının kullanılmasının maddelere ayrı maddeler gibi davranılmasına göre daha iyi bir gösterge olduğunu ifade etmişlerdir.

Keller vd. (2003) yaptıkları araştırmada, maddeler arasındaki bağımlılığı göz ardı ederek madde takımlarına bağımsız maddeler gibi davranmış ve madde takımlarını iki kategorili olarak puanlamışlardır. Bunun yanı sıra madde takımlarını çok kategorili olarak puanlama yöntemini de kullanmışlardır. İki kategorili puanlamanın sonuçlarını 3 PLM ile, çok kategorili puanlamanın sonuçlarını ise aşamalı tepki modeli ile incelemişlerdir. Daha sonra bu iki modelden elde edilen test güvenilirliği, madde bilgisi, test bilgisi ve bireylerin geçti/kaldı sınıflanması sonuçları üzerinden bir karşılaştırma yapmışlardır. Karşılaştırmalar sonucunda iki kategorili puanlama yönteminden elde edilen güvenirliliğin ve test bilgisinin yerel bağımlı maddeler nedeniyle fazla tahmin edildiği; çok kategorili puanlama yönteminden elde edilen test bilgisinin ise düşük tahmin edildiği; bireylerin geçti/kaldı sınıflandırmasında ise iki yöntemin de birbirine benzer sonuçlar verdiği belirlenmiştir.

2.3. Madde Takımı Tepki Kuramı ile İlgili Çalışmalar

Bradlow vd. (1999) iki parametrelili lojistik standart MTK modeline, aynı madde takımı içerisine yuvalanmış maddelerin tesadüfi etkisini ekleyerek modifiye edilmiş MTK modeli oluşturmuşlar ve bu iki modelle ilgili bir araştırma yürütmüşlerdir. Araştırmada 30 madde takımından ve 30 bağımsız maddeden oluşan toplam 60 maddelik ve 1000 kişilik bir simülasyon verisi üretilmiştir. Simülasyon koşullarında her madde takımında yer alan madde sayısı ve madde takımı varyansı olmak üzere

iki faktör değiştirilerek toplam yedi ayrı veri seti oluşturulmuştur. Her veri seti için iki ayrı model kullanılarak a , b ve θ parametreleri hesaplanmıştır. Bütün koşullarda, bütün parametreler için modifiye edilmiş MTK modelinden elde edilen değerlerin standart MTK modeline göre daha küçük olduğu belirlenmiştir. Ayrıca kestirilen ve gerçek parametreler arasındaki korelasyonlara bakıldığında, bütün simülasyon koşullarında ve bütün parametrelerde modifiye edilmiş MTK modeli ile elde edilen değerlerin daha iyi olduğu görülmüştür. Daha sonra Scholastic Aptitude Test'ten (SAT I) elde edilen 1000 kişilik ve 60 maddelik (40 madde takımı – 20 bağımsız madde) gerçek veri setine standart ve modifiye edilmiş MTK modelleri uygulanarak, modeller karşılaştırılmıştır. Sonuçlar bağımsız model ve madde takımı modeli altında elde edilen yetenek ve ayırıcılık parametresi kestirimlerinin oldukça benzer olduğunu; ancak iki parametrelili madde takımı kuramının standart modele göre parametre kestirimlerini iyileştirdiğini göstermiştir. Standart modeldeki yanlış kestirimin büyüklüğünün madde takımı etkisinin bir fonksiyonu olduğu belirlenmiştir. Tesadüfi etki varyansının küçük olması, araştırmada kullanılan veri seti için madde takımları arasındaki bağımlılığın makul bir etkiye sahip olduğunu göstermiştir. Araştırmacılar bu bulgu ışığında, koşullu bağımlılık küçük miktarda görünse bile ilgilenilen parametreler üzerinde bağımlılık etkisinin önemsiz ya da önemli olabileceğini ve bunu değerlendirmenin zor olabileceğini belirtmişlerdir.

Wainer (1995) çalışmasında Law School Admissions Testi'nin iki bileşeni olan (LSAT) - Reading Comprehension and Analytic Reasoning – kısımlarındaki maddelerin bağımsız maddeler olarak inşa edilmediğini belirtmiş ve maddeler arasındaki yerel bağımlılığı incelemiştir. Araştırmada LAST'ın her iki kısmı için de maddelerin yerel bağımsız olarak varsayıldığı madde dayanaklı güvenilirliği ve maddeler arasındaki bağımlılığın göz önüne alındığı madde takımına dayanan güvenilirliği hesaplanmıştır. Araştırmacı madde takımı yapısındaki bağımlılığın, testin güvenilirliği gibi istatistiksel karakteristik özellikleri üzerinde önemli etkisi olduğunu ve bu etkinin bağımsızlık varsayımıyla yanlış olarak modellendiğinde göz önüne alınmadığını belirlemiştir. Bağımlılık modellendiği zaman, her iki bölümünden elde edilen güvenilirlik katsayısının daha düşük olduğu belirlenmiştir.

Wainer vd. (2000) model uyumu için maksimum marjinal olabirlik yöntemini kullanarak standart 3PL modelini (MML), model uyumu için Markov Chain Monte Carlo simülasyonunun kullanarak standart 3PL modelini (madde takımı etkisi yok)

(Gibbs), bütün madde takımları için madde takımı etkisinin sabit kabul edildiği 3 PL madde takımı modelini ($Gibbs = \gamma$), ve bütün madde takımları için madde takımı etkisinin farklı değer aldığı 3PL madde takımı modelini ($Gibbs \neq \gamma$), kullanarak 3 PL madde takımı modelinin performansını çalışmak üzere bir simülasyon çalışması tasarlamışlardır. Bu dört model, Scholastic Assessment Test'in (SAT) marjinal dağılımını kullanılarak simüle edilen üç farklı test verisi üzerinde incelenmiştir. Birinci veri setinde madde takımı etkisi ele alınmazken, ikinci veri setinde eşit madde takımı etkisi ve üçüncü veri setinde eşit olmayan madde takımı etkisi bulunmaktadır. Her veri seti 1000 kişiden oluşan ve iki kategorili puanlanan, 70 çoktan seçmeli madde (30 bağımsız madde ve 10 maddelik 4 madde takımı) içermektedir. Dört modelden ve üç farklı veri setinden elde edilen, kestirilen ve gerçek θ , a , b , c parametreleri ve ortalamalar arasındaki korelasyonlar incelendiğinde madde takımı etkisinin olmadığı veri setinde tüm modellerden elde edilen korelasyonların en yüksek olduğu belirlenmiştir. Ancak madde takımı etkisi eşit olmadığına, kestirilen ve gerçek yetenek parametresi arasındaki korelasyonun ve mutlak hatanın 3PL-MTTK modelinde daha iyi olduğu belirlenmiştir. Tüm bu bulgulara dayanarak araştırmacılar madde takımı etkisi olduğunda 3PL-MTTK modelinin veri setine daha iyi uyum sağladığı sonucuna varmışlardır. Daha sonra gerçek verilerle de analiz yapılarak, analiz sonuçları incelenmiştir. Bu amaçla SAT (Scholastic Assessment Test) ve GRE (Graduate Record Examination) sözel bölümlerine ait sırasıyla 4000 ve 5000 kişilik bireylerin yanıtları kullanılmıştır. Aynı veri setlerine iki farklı model uygulanmış ve aralarındaki korelasyonlar incelenmiştir. En yüksek korelasyon SAT-Sözel sınavından elde edilen yetenek parametreleri üzerinde (0,998), en düşük korelasyon ise GRE-Sözel sınavından elde edilen şans parametreleri üzerinde (0,75) belirlenmiştir. Diğer aşamada ise aynı modellerin farklı verilere uygulanması ile elde edilen parametreler arasındaki korelasyonlar hesaplanmıştır. Bu amaçla GRE sınavına katılan 5000 kişilik iki bağımsız örneklem ele alınmıştır. Sonuçlar b parametresinin yerel bağımlılık durumundan ciddi bir şekilde etkilenmediğini, a parametresinin madde takımı etkisi göz önüne alındığında daha iyi sonuç verdiğini, c parametre kestiriminin zor olsa da yerel bağımlılığın bu parametre üzerinde çok büyük bir etkiye sahip olmadığını göstermiştir. Tüm sonuçlar ışığında madde takımları modellerinin madde ayırıcılık kestirimlerinde daha iyi sonuçlar verdiği, modelin kestirdiği standart hatanın gerçeğe daha yakın olduğu ifade edilmiştir.

Wainer ve Wang (2000) TOEFL sonuçlarını madde takımlarının etkisini göz önüne alarak incelemişlerdir. Standart MTK modellerinin, tek bir ortak uyarana bağlı madde kümelerindeki tepkilerin birbirine bağlı olduğu gerçeğini görmezden geldiğini ve bu modellerin, maddeler arasındaki koşullu bağımlılık nedeniyle bireyin yeterliğini daha yüksek kestirdiğini belirtmişlerdir. Standart 3PL modeline tesadüfi madde takımı etkisi olarak ifade edilen γ parametresinin eklenmesiyle geliştirilen madde takımı modelinin parametre kestirimlerinde daha iyi sonuç verdiğini belirlemişlerdir. Madde takımları arasındaki bağımlılık görmezden gelindiğinde madde güçlüğü parametresi için her şekilde iyi kestirimler yapılırken, şans parametresinin ve ayırt edicilik parametresinin yanlış kestirildiği belirlenmiştir. Şans parametresi olduğundan düşük kestirildiğini ortaya koymuşlardır. Madde ayırıcılığı ise bir testte olduğundan düşük kestirilirken diğer bir testte olduğundan yüksek kestirilmiştir. Ancak test bilgisi, maddeler arasındaki bağımlılık görmezden gelindiğinde fazla tahmin edilmektedir.

Lee vd. (2001) madde takımından oluşan testlerin MTK uygulamalarıyla eşitlenmesi amacıyla bir çalışma yapmışlardır. Bu çalışmada test eşitlemede, geleneksel yöntemlerden ortalama eşitleme, doğrusal eşitleme ve eşit yüzdelikli eşitleme ile MTK modellerinden 3 parametrelili lojistik model ve aşamalı tepki modeli ile sınıflamalı model olmak üzere iki tane çok kategorili MTK modeli kullanılarak kullanılmıştır. Geleneksel eşitleme yöntemlerinin toplam test puanını kullandığı ve bu eşitleme yöntemlerinin yerel bağımsızlık varsayımından etkilenmediği göz önünde bulundurularak iki ve çok kategorili MTK modellerinden elde edilen eşitleme sonuçları, geleneksel eşitleme yöntemleriyle karşılaştırılmıştır. Araştırmacılar çok kategorili model kullanılarak, madde takımları arasındaki bağımlılığın etkisinin giderilebileceği; çok kategorili MTK modellerinin geleneksel yöntemlerle daha yakın, kabul edilebilir sonuçlar verdiği belirlemişlerdir. Bu nedenle madde takımları içeren testlerin eşitlenmesinde çok kategorili MTK modellerinin kullanılmasının, iki kategorili MTK modellerinden daha uygun olduğu sonucuna varmışlardır.

Y. Li (2004) araştırmasında madde performansı üzerinde, yerel bağımlılığı dikkate alan ve madde takımı faktörünün etkisiyle ilgili farklı varsayımları olan alternatif modelleri incelemiştir. Araştırmacı ayrıca orijinalinin Bradlow ve diğerleri tarafından önerildiği birkaç alternatif madde takımı modelinin Bayesian model karşılaştırma ölçütlerine uygunluğunu da değerlendirmiştir. Bu amaçlarla iki parametrelili normal ogive madde takımı modeli ile üç alternatif modeli karşılaştırmak için bu modellere

birkaç Bayes model seçme ölçütü uygulamış ve bunları madde takımı içeren gerçek veri setlerine uygulamıştır. Araştırmada Model 1: Genel Model (The General Model) olarak ele alınan model Gibbons ve Hedeker (1992) tarafından önerilen iki faktör modelidir. Model 2: Eğitim Parametrelerinin Sınırlandırıldığı Genel Modeldir (The General Model with Constraints on Slope Parameters). Bu modelde eğer bir madde ölçmeyi amaçlanan yetenek üzerinde yüksek ayırıcılık gücüne sahipse, bu maddenin ikinci boyuttaki ayırıcılık gücünün düşük olacağı düşünülür ve maddeler boyunca çok boyutlu ayırıcılık parametresi sabit olarak ele alınır. Son model ise Glas, Wainer ve Bradlow (2000) tarafından önerilen modelle aynıdır. Model 3: γ_d Parametresi için Ayırıcılık Parametrelerinin Sınırlandırıldığı Genel Model (The General Model with Consistent Discrimination Parameters for γ_d). Bu modelde θ -yetenek parametresi ve ikincil boyut olan γ_d parametresi aynı ayırıcılık parametresine sahiptir.

Araştırmanın simülasyon verisi, 2000 kişilik ve her biri beş maddelik dört madde takım içeren bir veri setidir. Simülasyon verisi ile yapılan analiz sonucunda sözde-Bayes faktörü (The Pseudo-Bayes factor) karşılaştırılan modeller arasında en etkili olarak belirlenmiştir. Gerçek veri ise 1992 yılında uygulanan Law School Admission Test'inin (LSAT) iki ayrı kısmından alınmıştır. Her testte dört madde takımı ve her madde takımında 5 ila 8 madde yer almaktadır. Araştırmanın sonucunda genel model, orijinal madde takımı modelinin ardından en iyi model olduğu; eğitim parametrelerinin sınırlandırıldığı modelin ise diğer iki modele göre daha az memnun edici sonuçlar verdiği belirlenmiştir. Araştırmacı elde ettiği bulgular ışığında, sonuçları daha karmaşık olmasına karşın, ayırıcılık parametresini genel yetenek ve madde takımı için ayıran modelin veriye daha uyumlu olduğunu ifade etmiştir.

Y. Li vd. (2006) madde takımlarında kullanılan dört alternatif modelin, Bayes karşılaştırma ölçütleri ile uyumunu değerlendirmek amacıyla bir simülasyon çalışması yapmışlardır. İlk model yetenek ve madde takımı parametresi için aynı madde ayırıcılığına sahip 2 PL-MTTK modelidir. İkinci model ise yetenek ve madde takımı parametresi için farklı madde ayırıcılığına izin veren bir modeldir. Üçüncü model, yetenek ve madde takımı etkisi için madde ayırıcılığı arasında ters ilişki oluşturan bir kısıtlamaya sahiptir. Son model ise madde ayırıcılığını madde takımı etki parametresi bakımından ele alan bir sabit içeren modeldir. Araştırmada Bayes

karşılaştırma ölçütlerinin performansını karşılaştırmak ve uyum ölçütlerini belirlemek için 2000 bireyden ve her biri beş maddelik dört madde takımından oluşan dört tane veri seti üretmişlerdir. Bu Bayes modeli seçme ölçütlerini modellerle karşılaştırmak içinse, madde takımına sahip gerçek bir veri seti olan 1992 yılının LSAT sonuçları ile çalışmışlardır. Bu çalışmanın sonucunda yetenek ve madde takımı etkisi için ayrı ayırıcılık parametresine izin veren ikinci modelin daha iyi sonuçlar verdiği bulunmuştur.

Arora (2007) yaptığı çalışmada madde takımlarını ölçekleme için iki ayrı yaklaşım kullanarak, bu yaklaşımlardan elde ettiği sonuçları karşılaştırmıştır. Araştırmada TIMSS 2003 uygulamasında yer alan matematik ve fen değerlendirmelerini birleştirerek sekizinci sınıflar için yeni bir problem çözme ölçeği oluşturulmuştur. Araştırma için 46 ülkeden 220000 sekizinci sınıf öğrencisinin verileri kullanılmıştır. İlk yaklaşım olarak bu ölçekte yer alan bütün maddelere, bağımsız maddeler gibi bir ölçekleme yapma yaklaşımı kullanılmıştır. Bir diğer yaklaşım olarak ise her bir madde takımındaki maddeler, çok kategorili puanlama yoluyla ölçeklendirilmiştir. Her iki ölçekleme yaklaşımı ile elde edilen yetenek kestirimleri arasındaki korelasyon yüksek bulunmuştur. İki yaklaşımdan elde edilen varyanslara bakıldığında ise çok kategorili ölçekleme yaklaşımının daha iyi olsa da, iki yaklaşımdan elde edilen açıklanan varyans yüzdeleri arasında çok küçük bir fark olduğu belirlenmiştir. Daha sonra bütün ülkeler için problem çözme başarısı iki farklı ölçekleme yaklaşımından elde edilip, bu sonuçlar karşılaştırılmıştır. Bütün ülkelerin benzer sonuçlar verdiği; yalnızca iki ülkede çok kategorili ölçekleme yaklaşımıyla kestirilen ortalama başarının bağımsız madde ölçekleme yaklaşımında daha düşük olduğu belirlenmiştir. Son olarak iki ayrı yaklaşımla elde edilen TIMSS 2003 problem çözme ölçeğine ait güvenilirlik kestirimi incelenmiş ve sonuçlar arasında çok küçük bir farklılığın olduğu görülmüştür. Madde takımlarında yer alan maddeler, bağımsız maddeler olarak ölçeklendiğinde ülkelerin çoğunda (35) daha büyük güvenilirlik kestirimi elde edilmiştir. Araştırmacı, araştırma sonuçlarını genel olarak ele aldığı anda çok kategorili ölçekleme yaklaşımının yetenek kestirimi, açıklanan varyans miktarı, ölçek güvenilirliği ya da ülkelerin ortalama başarısı üzerinde bir farklılık yaratmadığını ifade etmiştir. Bu nedenle araştırmada, çok kategorili madde yaklaşımının bağımsız madde yaklaşımına göre bir avantajının açığa çıkmadığı belirlenmiştir.

Chang ve Wang (2010) yaptıkları arařtırmada Tayvan'ın PIRLS 2006 verilerini kullanarak madde takımı etkisini belirlemiř ve standart MTK modeli ile MTTK modelinden elde edilen madde ve yetenek parametrelerinin standart hatalarını karřılařtırmıřlardır. Bu amaçla madde takımı parametresini modele alan ve Bayesian madde takımı modeli (Wang, Bradlow & Wainer, 2002) ile madde takımı parametresini sıfıra eřitleyerek yerel madde bağımlılığını modele almayan iki durumu incelemiřlerdir. Daha sonra bu iki modelden elde edilen madde ve yetenek parametrelerinin standart hatalarını ve korelasyonlarını hesaplayıp karřılařtırmıřlardır. Bulgular yerel madde bağımlılığının, madde güçlük kestirimi üzerinde önemsiz bir etkiye sahip olduğunu göstermiřtir. Ancak arařtırmada madde ayırıcılık ve yetenek parametrelerine ait kestirimlerin, maddeler arasında yerel bağımlılık olduğunda yanlı olduğu ve bu parametrelerin olduğundan fazla tahmin edildiği belirlenmiřtir. Ayrıca arařtırmacılar, maddeler arasında yerel bağımsızlık varsayımı saėlanmadığında yetenek parametresine ait standart hatanın düşük kestirildiğini ifade etmiřlerdir.

Madde takımlarındaki maddeler arasındaki yerel madde bağımlılığının multistage tests (MST) performansı üzerindeki etkisini incelemek için Lu (2010) bir arařtırma yürütmüřtür. Bu arařtırmanın ilk amacı yerel bağımlılığın geçti/kaldı kararlılığına nasıl bir etki yarattığı belirlenmesidir. Arařtırmanın ikinci amacı ise 3PL-MTK ve 3PL-MTTK modellerinin performansını karřılařtırmaktır. MST dizaynında madde takımındaki madde oranı (0, .33, .67 ve 1) manipüle edilen faktördür. Diėer faktörler ise madde takımı maddelerinin yeri (sekiz farklı yer), YMB büyüklüğü (düşük-orta-yüksek) ve test uzunluğudur (36 maddelik uzun test ve 24 maddelik kısa test). Çalışmada her simülasyon koşulu 30 kere tekrar edilmiřtir ve her durum için 2500 kiřilik veri seti oluşturulmuřtur. Kořullara uygun olarak simüle edilen veri setleri yetenek kestiriminin kesinliėi ve sınıflandırma kararlarının doėruluėu açısından analiz edilmiřtir. Simülasyon sonuçları madde takımındaki maddelerin yerinin kararların doėruluėu üzerindeki etkisinin çok az olduğunu göstermiřtir. Ancak madde takımındaki madde oranının, YMB büyüklüğünün ve test uzunluğunun kararlar üzerinde oldukça önemli etkileri olduğu belirlenmiřtir. YMB etkisinin görmezden gelinmesi ve tekboyutlu 3PL modelinin kullanılmasıyla elde edilen sonuçlar, yetenek tahmininin doėruluėu ve kesinliėi üzerinde bilgi kaybı olduğunu göstermiřtir. YMB'deki büyüklüğün ve madde takımlarındaki madde oranının

artmasının kararların doğruluk oranını azalttığı belirlenmiştir. Ayrıca araştırmının sonuçları 3PL-MTTK'nin bütün yetenek kestirimlerinde ve sınıflandırma doğruluğunda gelişmediğini göstermiştir. Araştırmacı bu sonucun yerel madde bağımlılık düzeyinin yüksek olduğu ve madde takımındaki madde oranının yüksek olduğu durumlarda 3PL modelinin madde havuzunu kalibre etmekte kullanılmasının yanlış olduğunu doğruladığı belirtmiştir. Son olarak ise YMB'nin ihmal edilmesi ve 3PL modelinin kullanılmasının kararların kesinliğinde ve doğruluğunda şişirilmiş sonuçlara neden olduğu belirtilmiştir.

B. Zhang (2010) farklı ölçme modelleri ve istatistiksel süreçleri uygulayarak, dil testlerinde yeterlik sınıflamasının doğruluğunu incelemek amaçlı bir çalışma yapmıştır. 5000 kişi üzerinden yürütülen bu çalışmada klasik test kuramı modeli (KTK), İki kategorili MTK modellerinden 3PLM, 3 PL Bayesian tesadüfi-etki madde takımı modeli ve çok kategorili MTK modellerinden ATM kullanarak, bu modellerden elde edilen sonuçlar üzerinden bir karşılaştırma yapılmıştır. Araştırmanın bulguları bütün madde takımı içeren testlerde yerel madde bağımlılığı olduğunu göstermiş ve bu gibi durumlarda MTTK modellerinin kullanılması gerektiğini doğrulamıştır. Bu çalışmada KTK'nin bütün durumlarda MTK ve MTTK modellerine göre daha düşük doğru sınıflandırma yaptığı görülmüştür. Yeterlik dağılımının MTK ve MTTK modellerinde benzer dağılım gösterdiği, ancak MTK modellerinin yeterlik sınıflaması doğruluğunu olduğundan fazla kestirdiği belirlenmiştir. Bu sonuçlar yerel bağımlı maddelerde KTK ve MTK modellerinin kullanılmasının sınıflama doğruluğunu düşürdüğünü göstermektedir.

O. Zhang (2010) madde takımı büyüklüğünün ve yerel bağımlılığın yetenek parametresinin etkililiğine, ölçmenin standart hatasına ve test güvenilirliğine etkisini belirlemek amacıyla standart Rasch modelini, Kısmi Bilgi Modelini ve Rasch Madde Takımı modelini kullanarak bir simülasyon çalışması yapmıştır. Araştırmanın bir diğer amacı ise en iyi performans gösteren modeli belirlemektir. Güvenirlik kestirimi standart Rasch modelinde Kısmi Bilgi Modeline ve Rasch Madde Takımı modeline göre daha yüksek tahmin edilmiştir. Ölçmenin standart hatası ise standart Rasch modelinde Kısmi Bilgi Modeline ve Rasch Madde Takımı modeline göre daha düşük hesaplanmıştır. Ancak bu fark 0,02'den küçük olarak belirlenmiştir ve standart hatanın düşük kestirilmesinin nedeni olarak madde takımları arasındaki yerel bağımlılığın görmezden gelinmesi gösterilmiştir. Yetenek tahmini için ise düşük

düzeyde yanlılık ve RMSE olduğunda, her üç modelin de iyi performans gösterdiği belirlenmiştir. Araştırmanın sonucunda madde takımı büyüklüğünün orta ve küçük olduğu koşullarda Kısmi Bilgi Modelinin ve Rasch Madde Takımı Modelinin standart Rasch Modeline göre daha iyi performans gösterdiği; örneklem büyüklüğünün her üç model için de analiz sonuçları üzerinde etkili olduğu ve örneklem büyüklüğü arttıkça gerçek veri ve model kestirimleri arasındaki uyumsuzluğun da arttığı; örneklem büyüklüğü arttıkça standart Rasch modelinin güvenilirlikte yüksek kestirimin derecesinin arttığı; testin güvenilirliğini kestirmede Kısmi Bilgi Modelinin ve Rasch Madde Takımı Modelinin belirgin bir fark yaratmadığı belirlenmiştir. Çalışmada son olarak çok kategorili MTK modellerinin ve Rasch Madde Takımı Modelinin her ikisinin de standart Rasch Modeline göre daha avantajlı olduğunu ifade edilmiştir.

Eckes (2014) yaptığı araştırmada madde takımına dayalı testlerde madde takımı etkisini inceleyerek, madde takımı varyansının bir yerel bağımlılık derecesi olarak incelemiş ve MTTK modelleri ile standart MTK modellerinden elde edilen madde ve yetenek parametre kestirimleri ile güvenilirlik kestirimlerini karşılaştırmıştır. Bu amaçla Test of German as a Foreign Language'ın (TestDaF) dinleme kısmının iki ayrı dönemde (Nisan ve Kasım) uygulandığı, iki bağımsız örneklemden elde edilen veriler analiz edilmiştir. İlk örneklem 2859 ve ikinci örneklem ise 2214 kişiden oluşmaktadır. TestDaF'ta yer alan üç dinleme parçası kullanılarak 8 maddelik 1. Madde Takımı, 10 maddelik 2. Madde Takımı ve 7 maddelik 3. Madde Takımı oluşturulmuştur. Veriler (a) 2PL-MTTK, (b) 2PLM ve (c) ATM olmak üzere üç farklı 2PL ölçme modeline dayanan modellerle analiz edilmiştir. Madde Takımı-2 için orta, diğer iki madde takımı içinse düşük düzeyde madde takımı etkisi olduğu belirlenmiştir. Madde takımı etkisinin görmezden gelindiği standart MTK modelleri ile yapılan analizler sonucunda test güvenirliliğinin olduğundan yüksek ve yetenek kestirimine ait standart hatanın olduğundan düşük kestirildiği sonucuna varılmıştır. 2PL-MTTK ve 2PLM kullanılarak elde edilen her iki döneme ait ayırt edicilik ve güçlük parametresi arasındaki korelasyonlar oldukça yüksek hesaplanmıştır. Bu nedenle madde güçlüğü ve madde ayırt ediciliği kestirimlerinin kullanılan modelden büyük ölçüde etkilenmediği belirlenmiştir. Yerel bağımlılığın madde güçlük parametresi ve madde ayırıcılık parametresi kestirimleri üzerinde etkisi yok iken,

test güvenilirliđin olduđundan yüksek ve yetenek kestirimin standart hatasının olduđundan düşük kestirilmesine neden olduđu sonucuna varılmıřtır.

2.4. İki Faktör Modeli ile İlgili Çalışmalar

DeMars (2006) yaptıđı çalışmada dört ayrı MTK modelini kullanarak madde takımı içeren veri setleri üzerinde bir karşılaştırma gerçekleřtirmiřtir. Çalışmanın amacı iki faktör modelini, madde takımı etki modelini, çok kategorili madde tepki modelini ve madde takımı etkisini görmezden gelen bağımsız madde modelini kullanarak elde edilen yetenek, güvenilirlik, madde ayırıcılığı ve madde güçlüđü kestirimlerini karşılařtırmaktır. Arařtırmada hem simülasyon verisi hem de gerçek veri bir arada kullanılmıřtır. Simülasyon verisi için iki faktör modeli, madde takımı etki modeli, çok kategorili MTK modeli olmak üzere üç model ile iki test uzunluđunda (25 ya da 50 maddelik) altı test oluşturulmuřtur ve her beř maddelik küme bir madde takımı olarak ele alınmıřtır. Bu kořullarda 2000 kiřilik bir simülasyon verisi oluşturulmuřtur. Arařtırmada gerçek veri olarak ise PISA 2000 uygulamasındaki matematik ve okuma testleri kullanılmıř ve 5000 kiřilik bir örneklem ile çalışılmıřtır. Simülasyon verisi kullanıldığında güvenilirlik, bağımsız madde modelinde olduđundan daha yüksek kestirilmiřtir. Gerçek veri setinden elde edilen bulgularda da güvenilirliđin bağımsız madde modeli kullanıldığında daha yüksek olduđu belirlenmiřtir. Ancak verilen simülasyon sonuçları bu güvenilirlik tahminlerinin yapay olarak yüksek olduđunu göstermiřtir. Simülasyon verisi için ortalama hataların karekökü ve yanlılık çok kategorili MTK modelinin neden olduđu bilgi kaybından dolayı diđer modellere göre daha yüksek kestirilmiřtir. Gerçek veri için farklı modellerden elde edilen yetenek kestirimleri arasındaki iliřkilerin ise oldukça yüksek olduđu görölmüřtür. İki faktör modeli ya da madde takımı tepki modeli kullanıldığında madde güçlüđünün iyileřtiđi belirlenmiř; madde ayırıcılıđının ise yanlı olduđu görölmüřtür. Arařtırmacı modele ekstra parametre eklenmesinin modeli karmařıklařtırdığını; daha az karmařık modelin daha karmařık modele uygun olduđu zaman ortalama hataların karekökünü biraz daha yükselttiđini ancak yanlılık için aynı durumun söz konusu olmadığını belirtmiřtir.

Y. Li vd. (2010) iki kategorili ve çok kategorili puanlanan maddelerden oluřan bir simülasyon çalışması yapmıřlardır. Her birinde 14 madde bulunan üç okuma parçasında yer alan bütün son maddeler çok kategorili (0, 1, 2) ve diđer maddeler iki kategorili (0, 1) olacak řekilde puanlanarak 2000 kiřilik bir veri seti üretilmiřtir.

Her durum düşük madde takımı etkisi (Koşul 1) ve yüksek madde takımı etkisi (Koşul 2) olacak şekilde tekrarlanmıştır. Araştırmada genel çok kategorili madde takımı modeli ve 2PLM ile standart genelleştirilmiş kısmi bilgi modeli kombinasyonu (2PL/G-KPM) kullanılmıştır. Sonuçlar genel çok kategorili madde takımı modelinin model veri uyumu sağlandığında, 2PL/G-KPM'ye göre madde ayırıcılık ve madde güçlük parametresini daha iyi kestirdiğini göstermiştir. 2PL/G-KPM Koşul 1'de simülasyon verisine daha iyi uyum sağladığında korelasyonlar ve güçlük ve ayırt edicilik parametresine ait RMSD sonuçlarının diğer modele benzer ya da daha iyi değerler ürettiği belirlenmiştir. Koşul 2'de ise eğitim parametresi için benzer sonuçlar elde edilirken, güçlük parametresinin RMSD değerlerinin genelleştirilmiş çok kategorili madde takımı modelinde daha düşük hesaplandığı görülmüştür. Araştırmada daha sonra AIC ve BIC olabilirlik oran testleri hesaplanmış ve bütün simülasyon koşulları için genelleştirilmiş çok kategorili madde takımı modelinin 2PL/G-KPM'ye göre daha iyi uyum gösterdiği sonucuna varılmıştır. Test güvenilirliği açısından iki modelin sonuçları karşılaştırıldığında ise 2PL/G-KPM'den elde edilen teste ait güvenilirlik kestirimlerinin daha yüksek olduğu ve bu yüksek kestirimin Koşul 2'de daha fazla olduğu belirlenmiştir.

Araştırmacılar daha sonra iki kategorili ile çok kategorili puanlanan maddelerden oluşan ve madde takımına dayalı bir test olan İngilizce okuma testinden elde edilen gerçek veriler üzerinden yerel bağımlılığı değerlendirmek ve yerel bağımlılığın istatistiksel sonuçlar üzerindeki etkisini incelemek amacıyla bir çalışma yapmışlardır. Gerçek veri ile yapılan analiz sonucunda elde edilen bulgular, simülasyon çalışmasından elde edilen bulgulara paralellik göstermiştir. Araştırmanın genel sonuçları yerel bağımlılığın madde parametre kestirimlerinde küçük bir etkisinin olduğunu, ancak test bilgisi ve güvenilirlik üzerinde ise büyük etkisi olduğunu göstermiştir.

Rijmen (2010) yaptığı çalışmada madde takımlarına dayalı testler için kullanılan üç MTK modelini uluslararası İngilizce değerlendirme testinden elde edilen verilere uygulamıştır. Bu testte 20 tane okuduğunu anlama becerisini ölçen madde, beş maddelik dört madde takımı olacak şekilde organize edilmiştir. Testi ilk kez alan ve testin sonuna kadar ulaşabilen 13508 birey, çalışmanın örnekleme olarak alınmıştır. Tek boyutlu 2PL modelini, ikinci-düzyen modelini (second-order model) ve iki faktör modelini kullanarak yürütülen çalışmada model parametre sayıları, saplamaları,

Akaike bilgi kriterleri ve Bayesian bilgi kriterleri hesaplanmıştır. Yapılan çalışmada madde takımı modeli ile ikinci-düzyen model eşit çıkmıştır. Her iki model de iki faktör modelinin sınırlandırılmış halidir ve çalışmada bu veri seti için iki faktör modelinin daha tercih edilebilir olduğu sonucuna ulaşılmıştır. Ayrıca Rijmen bu çalışması ile araştırmacıları psikometrik modellere daha geniş çerçeveden bakmaları için cesaretlendirmeyi ve yazılıma ulaşılabilirlikten çok deneysel bir yapıya dayanarak uygun modellerin seçilmesine dikkat çekmeyi amaçlamıştır.

So (2010) Certificate in Advanced English (CAE) isimle madde takımına dayalı bir testi alan 4858 birey üzerinden bir araştırma yürütmüştür. Araştırmacı bu çalışmasında dört ayrı doğrulayıcı faktör analizi (tek-faktör, ikinci-sıra faktörü, iki faktör ve correlated-uniqueness factor models) kullanarak bir seri DFA sonucunda elde edile istatistiksel sonuçları karşılaştırmıştır. İki-faktör modeli DFA sonuçları arasından en iyisi olarak belirlenmiştir. Bu uyum, maddeler arasında bir ilişki olduğunu ve temel özellik faktörü ile pasaj faktörü olmak üzere iki faktör olduğunu göstermiştir. Bu sonuca dayanarak faktör yapısını hesaba alan ve almayan durumlarla, madde ve yetenek parametreleri kestirilmiştir. Bunun için tek boyutlu ve iki faktör MTK modelleri kullanılarak kestirilen parametreler karşılaştırılmıştır. Sonuçlar iki faktör modelinin veriye daha iyi uyum sağladığını; ikinci bir boyut olarak pasaj boyutunun varlığının madde güçlük parametresine önemli bir etkisi olmazken, madde ayırıcılık parametrelerinin kestirimine önemli bir etkisi olduğunu (tek boyutlu modele göre daha büyük ayırt edicilik parametresi hesaplanmış) göstermiştir. Ayrıca tek boyutlu modelin düşük yetenekli bireylerin yeteneğini olduğundan daha düşük, yüksek yetenekli bireylerin yeteneğini olduğundan yüksek kestirdiği belirlenmiştir.

Fukuhara ve Kamata (2011) yaptıkları çalışmada madde takımı dayanaklı testler için bir değişen madde fonksiyonu (DMF) belirleme modeli önermişler ve bu modeli incelemişlerdir. Araştırmacıların önerdikleri model, madde takımları için iki faktör çok boyutlu MTK'nin genişletilmiş bir halidir ve geleneksel MTK modellerinin aksine madde takımı etkisini de ele almaktadır. Geleneksel MTK modelleri ve iki faktör modeli ile madde takımlarındaki DMF'nin belirlendiği bu çalışmada, iki faktör ÇBMTK DMF modeli ile yapılan DMF büyüklüklerinin daha iyi kestirildiği ve bütün simülasyon koşulların geleneksel MTK DMF modellerindekinden daha yüksek oranda DMF belirlendiği bulunmuştur. Simülasyon çalışmalarının yanı sıra gerçek veriyle de önerilen modelin etkililiği araştırılmış ve modelin DMF belirlemede etkili

olduğu görülmüştür. Sonuç olarak madde takımlarından oluşan testlerde DMF'nin önerilen modelle daha doğru kestirildiği belirlenmiştir.

LaFond (2014) tek boyutlu MTK, MTTK ve iki faktör modellerini kullanarak kararlara ait tutarlılık ve doğruluk indekslerini araştırmak amacıyla hem simülasyon hem de gerçek veri setleri üzerinde bir çalışma yapmıştır. Simülasyon verisi test uzunluğu (30 madde), madde takımı büyüklüğü (6 madde ve 10 madde), örneklem büyüklüğü (5000), yerel bağımlılık derecesi (düşük, orta, büyük) bakımından farklılaştırılarak incelenmiştir. Gerçek veriler ise ikili puanlanan 60 madde içeren Test A'nın (15 maddelik 4 madde takımı) ve ikili puanlanan 30 madde içeren Test B'nin (10 maddelik 3 madde takımı) üç paralel formundan (Form I, II ve III) elde edilmiştir. Her form için kayıp veri olmayan 3000 birey rastgele olarak seçilmiştir. Araştırmacı iki faktör modelinin beklentilerini çok az karşıladığını, MTTK'nin ise beklentilerini karşılamadığını belirtmiştir ve MTTK'nin genel olarak model performansının düşük olduğunu ifade etmiştir. Ayrıca araştırmacı madde takımı içeren testler için önce yerel bağımlılık analizlerinin yapılmasını, eğer verilerin önemli derecede çok boyutluluk gösterdiği belirleniyorsa çok boyutlu indekslerin kullanımını önermiştir.

Min ve He (2014) beraber yürüttükleri çalışmada madde takımına dayalı bir test olan okuduğunu anlama testi verilerine, çok boyutlu iki faktör modelini, çok boyutlu madde takımı tepki kuramı modelini ve TBMTK modellerini uygulayarak, bu modellerden elde edilen parametreleri incelemiştir. Araştırmacılar ayrıca madde takımları arasındaki yerel bağımlılığa uygun olmayan modellerin seçilmesinin ya da bu yerel bağımlılığı ihmal eden TBMTK modellerinin kullanılmasının sonuçlarını da araştırmışlardır. Bu amaçla Çin'de 2011 yılında uygulanan Graduate School Entrance English Exam'in (GSEE) okuduğunu anlama bölümünü yanıtlayan 14089 bireyden veri toplanmıştır. Çalışmada kullanılan veri toplama aracında iki kategorili ve çok kategorili puanlanan maddeler bir arada bulunmaktadır. Veri setleri 1PLM/ATM, 2PLM/ATM, 3PLM/ATM olmak üzere üç TBMTK modeli ve MTTK ve iki faktör modeli olmak üzere iki tane de ÇBMTK modeli kullanılarak analiz edilmiştir. Daha sonra, kullanılan üç ayrı TBMTK modeli arasından 2PLM/ATM seçilerek bu modelden elde edilen sonuçlar ile MTTK ve iki faktör modeli sonuçları karşılaştırılmıştır. Sonuçlar iki faktör modelinin MTTK ve 2PL/ATM'ye göre en iyi uyum sağlayan model olduğunu göstermiştir. Bu üç model madde parametreleri açısından karşılaştırıldığında, model seçimi ile kesişim parametresi kestirimi

arasında bir fark olmadığı belirlenmiştir. Ancak 2PL/ATM kullanıldığı zaman eğitim parametresi kestirimlerinde büyük bir yanlışlık olduğu görülmüştür. Tüm bulgular gözden geçirildiğinde madde takımına dayanan testlere en iyi uyum sağlayan modelin iki faktör modeli olmasına karşın, bu model madde parametreleri, yetenek kestirimi ve kestirimlerin standart hatası bakımından MTTK ile aynı sonuçları üretmiştir. Aynı zamanda tek boyutlu 2PL/ATM'nin madde parametreleri, yetenek kestirimi ve kestirimlerin standart hatası üzerine büyük bir etkisinin olduğu belirlenmiştir.

2.5. İlgili Araştırmalar Özet

Yukarıda özetlenen önceki araştırmalar incelendiğinde yerel madde bağımlılığı etkisinin göz ardı edilmesinin madde güçlük ve ayırıcılık kestirimlerinde şişirilmiş sonuçlara ulaşılmasına; yetenek parametresi kestiriminde, düşük ve yüksek yetenekli bireylerde yanlış sonuçlarla karşılaşılmasına, eşitleme sonuçlarında problemler yaşanmasına ve güvenilirlik ile test bilgisinin olduğundan fazla tahmin edilmesine neden olduğu sonucuna varılmaktadır.

Tüm bu araştırmalar ışığında madde takımı içeren testlerde, değişen madde fonksiyonu belirlemede, test eşitlemede, testin güvenilirliğini, yetenek ve madde parametrelerini hesaplamada madde takımları arasındaki yerel bağımlılığı göz önüne alan modellerin daha doğru sonuçlara ulaşmayı sağladığı söylenebilir.

3. YÖNTEM

Bu bölümde araştırmanın modeline, evren ve örnekleme, veri toplama aracına ve verilerin analizine ilişkin bilgilere yer verilmiştir.

3.1. Araştırmanın Modeli

Bu araştırma ile madde takımlarını içeren testlerde kullanılan üç farklı modelden elde edilen madde ve yetenek parametrelerinin karşılaştırılması amaçlanmıştır. Araştırma, model karşılaştırmayı amaçladığı için temel araştırma niteliğindedir. Bu çalışma, geliştirilen bir kuramı test etmeye dönüktür.

3.2. Evren ve Örneklem

Ekonomik İşbirliği ve Kalkınma Örgütü'nün (OECD - Organisation for Economic Cooperation and Development) bir projesi olan PISA - The Programme for International Student Assessment (Uluslararası Öğrenci Değerlendirme Programı), en az 7 yıl öğrenim görmüş olması koşuluyla 15 yaş öğrencilerine, üç yılda bir yapılan bir uygulamadır. PISA tarafından 15 yaş, öğrenciliğin ardından yetişkin bir vatandaşlığın başlayabileceği en küçük yaş olarak kabul edilmektedir (MEB, 2013) ve okul türüne bakılmaksızın bu özelliklere sahip öğrenciler, PISA öğrenci evrenini oluşturmaktadır. Çalışmanın örnekleme belirlenirken katılımcı ülkelerde öğrenim gören öğrencilerin kayıtlı oldukları okulların türleri (mesleki ortaöğretim, genel ortaöğretim, ilköğretim), okulların bulunduğu yerlerin konumu (köy, kasaba, şehir), katılımcı ülkenin çalışmada esas teşkil edecek olan veri toplama bölgeleri (coğrafi ya da istatistikî bölge), okulların idari biçimi (kamu ya da özel) gibi değişken tanımları kullanılmaktadır. PISA uygulamasına katılan okullar, 12 istatistikî bölgeye (NUTS 1), okul türüne ve öğrencilerin temsil edilme oranlarına göre uluslararası istatistik firması-Westat tarafından 15 yaş grubu öğrencisi olan ve olabilecek tüm okullar arasından tabakalı örnekleme yoluyla seçilmekte (Fleischman vd., 2010) ve uygulama kapsamına alınan her okuldaki rastgele seçilen 30 öğrenci, örnekleme alınmaktadır.

2012 yılında programa, 34'ü OECD üyesi olmak üzere toplam 65 ülke katılmıştır. Bu ülkelerde 15 yaşında olan yaklaşık 28 milyon öğrenci bulunmaktadır ve PISA 2012'de bu öğrencileri temsilen 510 bine yakın öğrenciden veri toplanmıştır (OECD, 2013). PISA 2012'de, onuncu kitapçığı cevaplayan 35545 öğrenci bulunmaktadır.

Bu araştırmanın örneklemini veri temizlemesi sonucunda geriye kalan 28741 öğrenciden rastgele seçilen 250, 500 ve 1000 kişilik üç farklı örneklem oluşturmaktadır.

MTTK tabanlı çalışmaların çoğunda örneklem büyüklüğü 1000 olarak alınmıştır (Bradlow vd., 1999; Chen, 2010; Lee, 2004; Jiao, Wang & Kamata, 2005; Jiao, Wang & He, 2008; Sedivy, 2009; Wainer, Kaplan & Lewis, 1992; Wang vd., 2002). Bu nedenle bu çalışmada da 1000 kişilik bir örneklem büyüklüğü üzerinden çalışma yürütülmüştür. Ayrıca alan yazında bu konuyla ilgili 500'den (Wang vd., 2002; Wang & Wilson, 2005a) ve 1000'den küçük (Sedivy, 2009) örneklem büyüklüklerinde yapılan çalışmaların olması durumu göz önünde bulundurularak, küçük ve orta büyüklükteki örneklerde farklı modellerden elde edilen parametrelerin de karşılaştırılması amaçlanmıştır. Bu amaçla 250 ve 500 kişilik örneklem grubu da çalışmanın kapsamına alınmıştır.

3.3. Veri Toplama Aracı

PISA uygulamalarında kullanılan bilişsel testler 15 yaş grubu öğrencilerinin okuma, matematik ve fen okuryazarlığı alanlarındaki performanslarını ölçmeye yöneliktir. PISA'da, çoktan seçmeli maddelerle birlikte kısa yanıt gerektiren "kapalı uçlu maddeler (closed constructed response)" ve "uzun yanıt gerektiren açık uçlu maddeler (open constructed (extended) response)" gibi farklı soru tipleri de bulunmaktadır (MEB, 2011). Bu bilişsel testlerin yanında öğrencilerin sosyo-kültürel özelliklerini ölçmeye yönelik öğrenci anketleri, aile ile ilgili özellikleri ölçmeye yönelik veli anketleri, okul ile ilgili özelliklerini ölçmeye yönelik okul anketleriyle de veri toplanmaktadır.

Üç yıllık periyotlarla yapılan PISA'da, her uygulama için okuma, matematik ve fen okuryazarlığı alanlarından biri temel olarak belirlenir ve geriye kalan iki alandaki beceriler daha az ayrıntılı ve görelî olarak incelenir (OECD, 2009). 2012 yılında gerçekleştirilen PISA uygulamasında ağırlıklı alan matematik okuryazarlığıdır (OECD, 2013). Bu nedenle bu çalışmada, PISA 2012'de yer alan matematik okuryazarlığı becerisiyle ilgili maddeler ele alınmıştır.

PISA 2012'de matematik başarı testinde çoktan seçmeli sorular ve öğrencilerin kendi yanıtlarını oluşturmasını gerektiren açık uçlu sorular birlikte kullanılmıştır. Bu uygulamada matematik okuryazarlığını ölçmek için daha önceki PISA

uygulamalarında kullanılan 36 madde ve ilk kez bu uygulamada kullanılan 74 madde olmak üzere toplam 110 madde kullanılmıştır (ACER, 2013). Bu 110 madde döndürülmüş madde deseni (rotated item design) kullanılarak, toplam 13 kitapçık oluşacak şekilde kitapçıklara yerleştirilmiştir (OECD, 2013). Kitapçıklarda en az 12, en fazla 37 madde yer almaktadır (ACER, 2013) ve her öğrencinin yanıtlayacağı kitapçık türü bilgisayar ortamında seçkisiz yöntemle belirlenmektedir.

PISA uygulamalarında tüm öğrenciler, tüm maddeleri cevaplandırmamaktadır. Bu nedenle, araştırmada aynı maddeleri yanıtlayan öğrencilerle çalışmak için 13 kitapçıktan biri seçilmiştir. Araştırmanın amacına uygun olarak, çok sayıda madde ve madde takımı içerdiği belirlenen 10 numaralı kitapçık bu araştırmanın veri toplama aracı olarak ele alınmıştır. Bu kitapçık, 14'ü bağımsız ve 8'i madde takımından oluşmak üzere toplam 36 madde içermektedir. Kitapçıkta 4 tane 2 maddelik, 2 tane 3 maddelik ve 2 tane 4 maddelik olmak üzere toplam 8 madde takımı yer almaktadır.

Araştırma için 10 numaralı kitapçıkta bulunan madde takımı ve bağımsız madde sayıları ile 6 adet veri seti oluşturulmuştur. Veri setlerine ait bağımsız madde ve madde takımı sayıları Tablo 3.1'de özetlenmiştir.

Tablo 3.1: Araştırma için Oluşturulan Veri Setlerine İlişkin Özellikler

<i>Maddeler</i>	<i>Veri Seti 1</i>	<i>Veri Seti 2</i>	<i>Veri Seti 3</i>	<i>Veri Seti 4</i>	<i>Veri Seti 5</i>	<i>Veri Seti 6</i>
Bağımsız Madde Sayısı	0	7	14	0	7	14
Madde Takımı Sayısı	4	4	4	8	8	8
Toplam Madde Sayısı	11	18	25	22	29	36

Tablo 3.1 incelendiğinde, 0/7/14 bağımsız maddeden ve 4/8 madde takımından oluşan farklı veri setlerine ait özellikler görülmektedir. Bu araştırmada 4 madde takımı içeren veri setleri, 2 tane 2 maddelik, 1 tane 3 maddelik ve 1 tane 4 maddelik, madde takımları ile oluşturulmuştur.

Kitapçık 10'da yer alan 36 maddeden 32'si ikili (0 ya da 1) olarak puanlanmaktadır. Diğer 4 madde ise açık uçlu olarak geliştirilmiştir ve kısmi olarak puanlanmaktadır (kategoriler 0, 1, 2). Bu dört maddenin üçü madde takımı (ikisi aynı, biri farklı madde takımında yer alan maddeler), diğeri ise bağımsız maddedir. Puanlanmış veri setindeki açık uçlu maddeler yanlış cevaplar için sıfır, tüm kısmi puanlar için bir, tam puanlar için iki olacak şekilde kodlanmıştır.

3.4. Verilerin Analizi

Verilerin analizinde IRTPRO 2.1 (Item Response Theory for Patient-Reported Outcomes) programı kullanılmıştır. Bu bilgisayar programı Li Cai, David Thissen ve Stephen du Toit tarafından 2011 yılında geliştirilmiştir. MTK kullanarak test puanlama ve madde kalibrasyonu yapan bu istatistiksel yazılım BILOG_MG, MULTILOG, PARSCALE ve TESTFACT programlarının yapabildiği fonksiyonları tek bir programla yapabildiği için avantajlıdır. Ayrıca iki faktör modeli ve madde takımı analizlerinde hem iki kategorili hem de çok kategorili puanlanan maddelerle birlikte kestirim yapabilmesi, programın bir diğer avantajlı yönüdür. Ancak IRTPRO D = 1.7 olan ölçekleme faktörünü içermemektedir.

Yerel Bağımsızlık ve Tek Boyutluluk Varsayımı

Veri setlerinde yer alan madde takımları içerisindeki maddelerin yerel bağımlı olup olmadığını test etmek için yerel bağımlılık X^2 (LD X^2) istatistiği kullanılmıştır. Çünkü LD X^2 hesaplanması kolay ve iyi performans gösteren bir istatistiktir (Liu & Thissen, 2012). Bu değer IRTPRO 2.1 kullanılarak tek boyutlu MTK modellemelerinden elde edilmiştir. Bu istatistik Chen ve Thissen (1997) tarafından önerilen istatistiğe dayansa da çok kategorili tepkiler için geliştirilmiştir (Min & He, 2014). X^2 değeri, her madde tepkisi ile diğer tüm maddeler arasındaki iki yönlü çapraz tablolarda beklenen ve gözlenen frekanslardan hesaplanır. Farklı sayıda tepki kategorisi olan maddeler arasında karşılaştırma yapmak için bir standartlaştırma sağladığından bu istatistik, standartlaştırılmış χ^2 değerleridir (yaklaşık olarak z-puanlarıdır) (Cai vd., 2011). Eğer 10'dan büyük bir LD X^2 değeri hesaplandıysa maddeler arasında büyük bir bağımlılık söz konusudur; eğer madde çifti arasında hesaplanan LD X^2 değeri 5 – 10 arasında ise maddeler arasında orta düzeyde yerel bağımlılık bulunmaktadır ve bu durum madde çiftleri arasında yerel bağımlılık olduğuna işaret edebilmektedir (Cai vd., 2011). IRTPRO'da LD X^2 değerlerinin bulunduğu çıktı tablosunda kırmızının ve mavinin açık ve koyu tonları ile siyah gibi farklı renkler bulunmaktadır. Çıktılarda üç değeri keyfi bir kesme noktasıdır ve renk geçişlerini ayarlamaktadır (daha kırmızı, daha mavi gibi). Eğer bir madde çifti tepkileri arasındaki gözlenen kovaryans model tarafından tahmin edileni aşarsa çıktılar kırmızı, daha küçük ise mavi renkle ifade edilmekte ve kırmızı çıktılar pozitif, mavi çıktılar ise negatif YMB'ye işaret etmektedir. Ayrıca çıktılarda yer alan kırmızı değerler modellenmemiş bir boyutu ölçebilmektedir (Cai vd., 2011).

Bu çalışmada 250/500/1000 örneklem büyüklükleri için oluşturulan altı farklı veri seti için standartlaştırılmış LD X^2 değerleri ayrı ayrı hesaplanmıştır. Ancak çalışmada, yalnızca aynı madde takımı içerisinde yer alan maddeler arasındaki LD X^2 değerleri üzerinde durulmuştur. Çünkü maddelerin madde takımları boyunca bağımlılıkları başka faktörlerden kaynaklanabilir (Min & He, 2014) ve bu durum da bu araştırmanın konusu dışında yer almaktadır.

Veri seti 1’de, veri seti 2’de ve veri seti 3’te bir tane 4 maddelik ($1 \times C_2^4 = 6$), 2 tane 2 maddelik ($2 \times C_2^2 = 2$) ve bir tane 3 maddelik ($1 \times C_2^3 = 3$) olmak üzere toplam 4 madde takımı ve incelenecek 11 madde çiftine ait LD X^2 değeri bulunmaktadır. Bu üç veri seti için bütün örneklem büyüklüklerinden elde edilen bu 11 madde çiftine ait LD X^2 değerleri incelenmiştir. Çalışmada kullanılan veri seti 4’te, veri seti 5’te ve veri seti 6’da 4 tane 2 maddelik, 2 tane 3 maddelik ve 2 tane 4 maddelik olmak üzere toplam 8 madde takımı yer almaktadır. Bu nedenle her örneklem büyüklüğünde, 4 tane 2 maddelik madde takımı için 4 madde çifti ($4 \times C_2^2$), 2 tane 3 maddelik madde takımı için 6 madde çifti ($2 \times C_2^3$), 2 tane 4 maddelik madde takımı için 12 madde çifti ($2 \times C_2^4$) olmak üzere toplam 22 madde çiftinin standartlaştırılmış LD X^2 değerleri bütün örneklem büyüklükleri için ele alınmıştır. Veri setlerinde yer alan madde takımlarındaki maddelerin LD X^2 değerleri örneklem büyüklüklerine göre Tablo 3.2’de yer almaktadır.

Tablo 3.2: Veri Setlerine göre Madde Takımlarında Yer Alan Maddelerin LD X^2 Değerleri

<i>Veri Seti</i>	<i>Örneklem</i>	<i>LD X^2 Ranji</i>
Veri Seti 1	250	(-0,6) – (3,9)
	500	(-0,6) – (11,4)
	1000	(-0,1) – (11,8)
Veri Seti 2	250	(-0,6) – (3,2)
	500	(-0,7) – (9,6)
	1000	(-0,4) – (10,8)
Veri Seti 3	250	(-0,6) – (4,1)
	500	(-0,6) – (9,2)
	1000	(-0,3) – (10,4)
Veri Seti 4	250	(-0,9) – (3,1)
	500	(-0,7) – (9,9)
	1000	(-0,7) – (10,6)
Veri Seti 5	250	(-0,7) – (3,7)
	500	(-0,7) – (6,4)
	1000	(-0,5) – (9,3)
Veri Seti 6	250	(-0,7) – (7,6)
	500	(-0,7) – (8,9)
	1000	(-0,6) – (10,1)

Tablo 3.2'deki veri setlerine ait LD X^2 deęerleri incelendięinde, veri setlerinin çoęunda yerel baęımlı madde çiftlerinin bulunduęu görölmektedir. Bu nedenle bu veri setleri üzerinden, MTTK uygulamalarının yapılabileceęi belirlenmiřtir.

Aynı zamanda 3'ten büyük deęerlerin modellenmemiř bir boyuta iřaret edebileceęi durumu da göz önüne alınırsa veri setlerinin tek boyutlu olmadığı da söylenebilir. Bu durumda incelenen veri setlerinde, iç içe geęmiř iki varsayım olan tek boyutluluk ve yerel baęımsızlık varsayımlarının saęlanmadığı sonucuna varılabilir.

Kullanılan MTK Modelleri

Çalıřmada tek boyutlu ve çok boyutlu MTK modelleri kullanılmıřtır. Arařtırmada kestirim yapmak için TBMTK modellerinden 2PL/G-KPM olmak üzere bir, ÇBMTK modellerinden ise iki faktör ve MTTK modeli olmak üzere iki model kullanılmıřtır. Tüm MTK modellerinin parametreleri IRTPRO 2.1 ile elde edilmiřtir.

IRTPRO 2.1 iki kategorili ve çok kategorili MTK modelleri ile aynı anda kestirim yapmaya olanak saęladıęından, tüm modellerde iki kategorili puanlanan maddeler için 2PLM ve çok kategorili puanlanan maddeler içinse G-KPM bir arada kullanılmıřtır. Program, 2PLM ve G-KPM kullanılarak analiz edilen maddeler için ayrı tablolar oluřturmaktadır. Tüm veri setlerinde 2PL modeli ile analiz edilen maddeler için ayrı ayrı a ve c parametreleri ile bu parametrelere ait standart hatalar elde edilmiřtir. IRTPRO çıktılarındaki a parametresi eęim, c parametresi ise "b" eřik parametresi ile iliřkili olan bir kesiřim parametresidir ($b = -c / a$). Ayrıca eęim-eřik formu çok boyutlu modellerde doęru genellemeler yapamadığı için bütün modellerde eęim-kesiřim formu kullanılmıřtır ($a(\theta - b) = a\theta + c$) (Cai vd., 2011). Çünkü ÇBMTK'de b parametresinin TBMTK'dekine benzer olarak yorumlanıp yorumlanmadığı açık deęildir (Min & He, 2014). Bu çalıřmada ÇBMTK modellerinde daha iyi genelleme yapması nedeni ile eřik parametresi (b) yerine kesiřim (c) parametresi seçilmiřtir. G-KPM ile analiz edilen maddeler için ise a ve γ parametre deęerleri ve bu parametrelere ait standart hatalar elde edilmiřtir.

MTTK'de model tanımlamak için birincil boyutun yeri ve ölçeęi sabitlenirken; ikincil (madde takımı) boyutun yeri sabitlenip, ölçeęi ise serbest bırakılmalıdır (Rijmen, 2010). Bu nedenle bu çalıřmada madde takımı modelini tanımlamak için birincil boyutun ortalaması sıfır, varyansı bir olacak řekilde sabitlenirken; ikincil boyutta ise ortalama bir ve varyans serbest olacak řekilde düzenlemeler yapılmıřtır. Ayrıca

MTTK modeli için her maddenin genel boyuttaki ayırt edicilik değeri serbest bırakılırken, her maddenin ait olduğu madde takımı boyutundaki ayırıcılığı ise genel boyuttaki ayırt ediciliğe eşitlenmiş ve diğer madde takımlarındaki boyutlar için ayırt edicilik sıfıra eşitlenmiştir. İki faktör modelini tanımlamak için ise (Cai vd., 2011) bütün boyutların varyansı bir, ortalaması sıfır olarak ele alınmıştır. Bu modelde hem genel boyut hem de madde takımı boyutu için ayrı ayırt edicilik değerleri hesaplanmıştır.

IRTPRO 2.1 yazılımında madde parametrelerinin kestirimi için Bock-Aitkin Expectation-Maximization (BA-EM), Adaptive Quadrature ve Metropolis-Hastings Robbins-Monro (MH-RM) olmak üzere üç kestirim yöntemi bulunmaktadır. Bu çalışmada parametre kestirimleri, Bock ve Aitkin (1981) tarafından geliştirilen Bock-Aitkin marjinal maksimum olabilirlik kestirimi ile yapılmıştır. Çünkü Cai ve diğerleri (2011) tek boyutlu modeller ve iki faktör MTK modelleri için Bock-Aitkin EM algoritmasının seçilmesi ile etkili kestirimler yapılabileceğini ifade etmişlerdir. Bireylerin yetenek puanlarını hesaplamak için ise expected a posteriori (EAP) yöntemi kullanılmıştır. Çünkü bu kestirim yönteminde ortalama yansız olarak kestirilmektedir ve küçük standart sapma değerleri elde edilmektedir (Zimowski, Muraki, Mislevy & Bock, 2003). Ayrıca bu yöntemle, sıfır puan ve tam puan alan bireyler için bile kestirim yapılabilir (Embretson & Reise, 2000). Kısacası herhangi bir cevap örüntüsü için kestirim yapabilmesi, diğer kestirim yöntemlerine göre daha küçük hata değerleri vermesi, iterasyon gerektirmemesi nedeniyle daha hızlı ve kolay hesaplanması EAP'ın tercih edilme nedenlerindedir. Bütün kalibrasyon sürecinde 2FM için genel ve özel boyut, TBMTK ile MTTK'de ise genel boyut standart normal dağılıma $N(0,1)$ sahip olacak şekilde sınırlandırılmıştır. Böylece farklı MTK modellerinden elde edilen parametre kestirimlerinin aynı ölçek düzeyinde olması sağlanmıştır (Y. Li vd., 2010).

Farklı modelleri karşılaştırmak ve hangi modelin veriyle uyum gösterdiğini belirlemek için log-olabilirlik (-2log-likelihood, -2LL) değerleri hesaplanmıştır. Kestirecek parametre sayısı yardımıyla hesaplanan serbestlik derecesi 0,05 anlamlılık düzeyindeki tablo X^2 değeri ile karşılaştırılarak modeller arasında anlamlı farklılık olup olmadığı ve hangi modelin veri ile daha iyi uyum gösterdiği incelenmiştir. Son olarak ise farklı modellerden elde edilen madde ve yetenek

parametrelerine ve bu parametrelere ait standart hata deęerlerine ait korelasyon deęerleri Pearson korelasyon katsayısı yardımıyla hesaplanmıřtır.



BULGULAR VE TARTIŞMA

Bu bölümde alt problem sırasına göre verilerin analiz edilmesi ile elde edilen bulgular tablolar halinde sunulmuş ve yorumlanmıştır. Elde edilen bulgular, ilgili alan yazın göz önünde bulundurularak tartışılmıştır.

4.1. Madde Parametrelerinin Değerlendirilmesine İlişkin Bulgular

Araştırmanın bu kısmında birinci alt problem olan “Madde takımı içeren testlerde madde takımı sayısı (4/8), bağımsız madde sayısı (0/7/14) ve örneklem büyüklüğü (250/500/1000) değiştiğinde; farklı modellerden elde edilen madde parametreleri ve bu parametrelere ait standart hatalar nasıldır?” sorusuna ait bulgulara yer verilmiştir.

a_1 : genel boyuta ait eğim parametresi, a_2 : ikincil boyuta ait ayırt edicilik parametresi, c : kesişim parametresi ve SH: standart hata olmak üzere; veri seti 1’de, 2’de ve 3’te yer alan iki kategorili puanlanan maddelerden elde edilen madde parametrelerine ve hata ortalamalarına ilişkin bulgulara Tablo 4.1’de yer verilmiştir.

Tablo 4.1: Veri Seti 1’de, 2’de ve 3’te Yer Alan İki Kategorili Puanlanan Maddelerin Madde Parametre Ortalamaları ile Bu Parametrelere ait Hata Ortalamaları

Veri Seti	Örneklem Büyüklüğü	Modeller	a_1	SH	a_2	SH	c	SH
Veri Seti 1 (4 madde takımı)	250	TBMTK	1,61	0,32	-	-	-0,70	0,28
		2FM	1,79	0,56	0,88	0,87	-0,76	0,41
		MTTK	1,85	0,45	1,85	0,45	-0,79	0,32
	500	TBMTK	1,64	0,23	-	-	-0,47	0,18
		2FM	1,82	0,33	0,87	0,59	-0,56	0,30
		MTTK	1,76	0,25	1,76	0,25	-0,52	0,17
	1000	TBMTK	1,53	0,15	-	-	-0,58	0,12
		2FM	1,69	0,20	0,88	0,36	-0,67	0,22
		MTTK	1,61	0,18	1,61	0,18	-0,61	0,13
Veri Seti 2 (4 madde takımı – 7 bağımsız madde)	250	TBMTK	1,60	0,33	-	-	-0,61	0,32
		2FM	1,66	0,40	0,61	1,13	-0,67	0,33
		MTTK	1,65	0,34	1,92	0,39	-0,61	0,25
	500	TBMTK	1,51	0,21	-	-	-0,55	0,20
		2FM	1,64	0,27	0,93	0,54	-0,61	0,24
		MTTK	1,63	0,22	1,82	0,25	-0,59	0,17
	1000	TBMTK	1,44	0,14	-	-	-0,61	0,12
		2FM	1,54	0,18	0,88	0,36	-0,66	0,17
		MTTK	1,49	0,16	1,62	0,18	-0,62	0,12
Veri Seti 3 (4 madde takımı – 14 bağımsız madde)	250	TBMTK	1,54	0,25	-	-	-0,40	0,33
		2FM	1,60	0,38	0,69	3,10	-0,41	0,32
		MTTK	1,61	0,28	1,89	0,33	-0,38	0,25
	500	TBMTK	1,45	0,19	-	-	-0,36	0,19
		2FM	1,55	0,25	0,97	0,48	-0,38	0,22
		MTTK	1,55	0,21	1,78	0,26	-0,37	0,16
	1000	TBMTK	1,41	0,13	-	-	-0,43	0,12
		2FM	1,50	0,17	0,93	0,34	-0,46	0,16
		MTTK	1,46	0,15	1,60	0,18	-0,42	0,11

Tablo 4.1’de, madde takım sayısı 4 iken veri setine 0/7/14 tane bağımsız madde eklenmesi ile oluşturulan veri seti 1’den, 2’den ve 3’ten örneklem büyüklüklerine ve kullanılan modellere göre elde edilen a ve c parametreleri ile bu parametrelere ait hata değerleri için şu bulgular yer almaktadır.

250 örneklem büyüklüğünde veri setine bağımsız madde eklenmesiyle birlikte TBMTK’den elde edilen ortalama a_1 parametre değerlerinin azaldığı belirlenmiştir. Bu parametreye ait ortalama hata değeri ise bağımsız maddenin bulunmadığı veri setinde, diğer koşullara göre daha düşüktür. 2FM’den ve MTTK’den elde edilen a_1 parametresinin ve bu parametreye ait hata değerlerinin ortalaması ise, veri setine bağımsız madde eklenmesiyle birlikte azaldığı; a_2 parametresi ortalaması için genel bir örüntü olmadığı, ancak bu parametreye ait hata ortalamalarının azaldığı belirlenmiştir. Ancak 2FM’den elde edilen a_2 parametresinin ortalama hata değerini, bağımsız madde sayısı arttıkça ciddi bir biçimde artış göstermiştir. Bu bulguya dayanarak düşük örneklem büyüklüklerinde, genel boyutun yanında özel boyut da içeren 2FM’den elde edilen özel boyuta ait eğim parametresi (a_2) kestirimlerinin standart hatasının daha yüksek kestirildiği söylenebilir. MTTK ile hesaplanan a_2 parametresine ilişkin kestirimlerin 2FM’ye göre daha düşük hatalı olduğu söylenebilir. Bu örneklem büyüklüğünde bütün modeller için en az hatalı a_1 parametre ortalamaları veri seti 3’ten elde edilmiştir. Ancak en az hatalı a_2 parametre ortalamaları ise MTTK için yine veri seti 3 iken 2FM için veri seti 1’dir. Bu nedenle düşük örneklemelerde 2FM ile yapılacak analizlerin, daha az sayıda madde içermesi gerektiği ifade edilebilir. TBMTK ile hesaplanan c parametresininse veri setine bağımsız madde eklenmesiyle birlikte, ortalama değerinde ve hata kestirimi ortalamasında artış gözlenmiştir. 2FM’den ve MTTK’den elde edilen c parametresine ait bulgular incelendiğinde ise bağımsız madde sayısı arttıkça c parametresinin ortalamasının arttığı, ortalama hata değerlerinin ise azaldığı görülmektedir.

500 örneklem büyüklüğünde veri setine bağımsız madde eklenmesi durumunda, TBMTK ile hesaplanan a_1 parametre değerinin ortalamasının ve a_1 parametresine ait ortalama hata kestiriminin azaldığı gözlenmiştir. Veri setine bağımsız madde eklenmesiyle 2FM’den ve MTTK’den elde edilen a_1 parametresinin ve bu parametreye ait hata değerlerinin ortalamasının düştüğü; a_2 parametresinin ise veri seti 1’e göre ortalaması artarken, hata değerinin ortalamasının çok değişmediği

belirlenmiştir. Ayrıca 250 örneklem büyüklüğüne göre a_2 parametresinin hata ortalamalarının yüksek miktarda azaldığı ve özel boyut içeren modellerde örneklem büyüklüğünün kestirimler üzerine etkili olduğu söylenebilir. 2FM'den elde edilen eğim parametrelerine ait hata değerleri, MTTK'den elde edilen ortalama hata değerlerinden daha yüksektir. a_1 parametresi için en düşük hata ortalaması TBMTK'den, a_2 parametresi için en düşük hata ortalaması MTTK'den hesaplanmıştır. Ayrıca eğim parametresi için en küçük hata değerlerine veri seti 3'ten erişilmiştir. TBMTK'den elde edilen c parametresi ve bu parametrenin hata kestirimlerinin ortalaması için genel bir örüntü belirlenmemiş olup, hata kestirimlerinin bağımsız madde eklenmesiyle az miktarda değiştiği söylenebilir. 2FM ve MTTK yardımıyla hesaplanan c parametresi için genel yorum yapılamazken ve bu parametrenin hata değerlerinin ortalamasının veri setine bağımsız madde eklenmesiyle azaldığı ifade edilebilir.

1000 örneklem büyüklüğünde TBMTK'den elde edilen ortalama a_1 parametresinin ve bu parametrenin ortalama hata değerinin veri setine bağımsız madde eklenmesiyle birlikte azaldığı belirlenmiştir. Veri setine bağımsız madde eklenmesiyle 2FM'den ve MTTK'den elde edilen ortalama a_1 parametresinin ve bu parametreye ait hata değerlerinin ortalamasının azaldığı söylenebilir. Bu örneklem büyüklüğünde 2FM ve MTTK ile hesaplanan a_2 parametre ortalama değeri ve ortalama hata değerleri, bağımsız madde sayısının artmasıyla çok değişmemiş ve birbirine yakın sonuçlar vermiştir. 2FM'den ve MTTK'den elde edilen a_2 parametresine ait ortalama hata değeri incelendiğinde, bu değerlerin MTTK'de daha düşük olduğu gözlenmiştir. 1000 örneklem büyüklüğünde eğim parametreleri için en az hata ortalamalarına genel olarak veri seti 3'te ulaşılabilirdiği söylenebilir. TBMTK'den elde edilen c parametresine ait ortalama hata değerinin veri setine bağımsız madde eklenmesiyle birlikte sabit kaldığı gözlenmiştir. Veri setine bağımsız madde eklenmesiyle 2FM'den elde edilen ortalama c parametresi içinse bağımsız madde sayısı arttıkça her iki modelden elde edilen ortalama parametre değerinin arttığı, ortalama hata değerinin ise azaldığı sonucuna varılmıştır. MTTK için de benzer şekilde c parametresi ortalama hata değerinin azaldığı gözlemlenmiştir.

Örneklem büyüklüğü arttıkça, her bir veri seti için bütün modellerden elde edilen a ve c parametrelerine ait ortalama hata değerleri azalmıştır. Ayrıca örneklem

büyükliğündeki artış ile birlikte farklı modellerden elde edilen a ve c parametrelerine ait hata ortalamalarının birbirine yaklaştığı görülmektedir. Her üç model için de elde edilen hata ortalamaları örneklem büyüklüğünün en fazla olduğu 1000 örneklem için diğer örneklem büyüklüklerine göre azaldığı belirlenmiştir. a_2 parametresi için 250 örneklem büyüklüğünde 2FM ve MTTK yüksek hatalar vermiştir. Bu nedenle genel boyutun yanı sıra özel boyut da içeren veri setlerinde doğru kestirimler yapabilmek için büyük örneklemle çalışılması gerektiği söylenebilir. Her üç modelden elde edilen madde parametreleri incelendiğinde, a ve c parametrelerine ait en az hatalı kestirimlerin 1000 örneklem büyüklüğünde madde sayısının en fazla olduğu veri seti 3 için gerçekleştirildiği görülmektedir.

a_1 : genel boyuta ait eğim parametresi, a_2 : ikincil boyuta ait ayırt edicilik parametresi, c: kesişim parametresi ve SH: standart hata olmak üzere; veri seti 4'te, 5'te ve 6'da yer alan iki kategorili puanlanan maddelerden elde edilen madde parametrelerine ve hata ortalamalarına ilişkin bulgular Tablo 4.2'de sunulmuştur.

Tablo 4.2: Veri Seti 4'te, 5'te ve 6'da Yer Alan İki Kategorili Puanlanan Maddelerin Madde Parametre Ortalamaları ile Bu Parametrelere ait Hata Ortalamaları

Veri Seti	Örneklem Büyüklüğü	Modeller	a_1	SH	a_2	SH	c	SH
Veri Seti 4 (8 madde takımı)	250	TBMTK	1,59	0,29	-	-	-0,42	0,28
		2FM	1,76	0,62	0,71	0,97	-0,53	0,64
		MTTK	1,83	0,42	1,83	0,42	-0,50	0,45
	500	TBMTK	1,51	0,29	-	-	-0,35	0,33
		2FM	1,75	0,39	0,85	1,02	-0,51	0,42
		MTTK	1,70	0,39	1,70	0,39	-0,46	0,31
	1000	TBMTK	1,44	0,15	-	-	-0,43	0,22
		2FM	1,62	0,22	0,76	0,65	-0,54	0,30
		MTTK	1,58	0,19	1,58	0,19	-0,49	0,28
Veri Seti 5 (8 madde takımı – 7 bağımsız madde)	250	TBMTK	1,51	0,32	-	-	-0,46	0,35
		2FM	1,64	0,54	0,69	1,03	-0,53	0,46
		MTTK	1,62	0,41	1,72	0,47	-0,45	0,39
	500	TBMTK	1,46	0,20	-	-	-0,41	0,20
		2FM	1,63	0,38	0,80	0,81	-0,51	0,24
		MTTK	1,63	0,24	1,72	0,26	-0,49	0,22
	1000	TBMTK	1,40	0,16	-	-	-0,49	0,19
		2FM	1,54	0,21	0,74	0,58	-0,55	0,23
		MTTK	1,51	0,16	1,58	0,17	-0,54	0,14
Veri Seti 6 (8 madde takımı – 14 bağımsız madde)	250	TBMTK	1,51	0,33	-	-	-0,36	0,25
		2FM	1,61	0,47	0,67	0,99	-0,38	0,45
		MTTK	1,55	0,40	1,64	0,48	-0,43	0,35
	500	TBMTK	1,45	0,19	-	-	-0,32	0,19
		2FM	1,57	0,32	0,77	0,82	-0,37	0,33
		MTTK	1,56	0,23	1,69	0,27	-0,35	0,22
	1000	TBMTK	1,40	0,13	-	-	-0,05	0,14
		2FM	1,50	0,19	0,72	0,49	-0,42	0,20
		MTTK	1,50	0,19	1,61	0,24	-0,41	0,20

Tablo 4.2'de madde takım sayısı 8 iken veri setine 0/7/14 tane bağımsız madde eklenmesi ile oluşturulan veri seti 4'ten, 5'ten ve 6'dan örneklem büyüklüklerine ve kullanılan modellere göre elde edilen a ve c parametreleri ile bu parametrelere ait hata değerleri için şu bulgular yer almaktadır.

250 örneklem büyüklüğünde veri setine bağımsız madde eklenmesiyle oluşturulan farklı veri setlerinin TBMTK ile analizi sonucu elde edilen a_1 parametre ortalamaları genel olarak düşerken, bu parametrelere ait ortalama hata değerleri ise artış göstermiştir. Veri setine bağımsız madde eklenmesiyle 2FM'den ve MTTK'den elde edilen ortalama a_1 parametresinin ve bu parametreye ait ortalama hata değerlerinin düştüğü söylenebilir. 2FM'den elde edilen a_2 parametresinin bağımsız madde sayısı ile birlikte ortalamasının azaldığı, hata ortalamalarının ise tamamı madde takımı olan veri seti 4'e göre diğer veri setlerinde arttığı gözlenmiştir. MTTK sonuçlarına bakıldığında ise a_2 parametresine ait ortalama değerlerin azaldığı, hata ortalamalarının ise bağımsız madde sayısının artmasıyla birlikte arttığı belirlenmiştir. a_1 parametresi için bütün veri setlerinde en düşük hata ortalamasına sahip model TBMTK ile; a_2 parametresi için ise MTTK'dir. Her koşulda MTTK, 2FM'den daha düşük hatalı eğim parametreleri kestirmiştir. TBMTK'den elde edilen c parametre ortalamaları ile bu parametrelere ait ortalama hata değerleri için de genel bir yorum yapılamadığı görülmüştür. Veri setine bağımsız madde eklenmesiyle 2FM'den ve MTTK'den elde edilen c parametresine ait bulgular incelendiğinde ise veri setinde bağımsız maddenin bulunmadığı duruma göre ortalamasının arttığı, hata değerleri ortalamasının ise azaldığı görülmektedir.

500 örneklem büyüklüğünde her üç modelden elde edilen a_1 parametre ortalaması ile bu parametrelere ait hata değerleri ortalamasının, veri setindeki bağımsız madde sayısındaki artışla birlikte azaldığı belirlenmiştir. 2FM kestirimleri için a_2 parametresi ve bu parametrenin hata değeri ortalamasının ise bağımsız madde sayısı arttıkça azalırken; MTTK'de yalnızca veri seti 4'e göre diğer iki veri setindeki ortalama hata değerleri için bir azalıştan söz edilebilir. a_1 parametrelerine ait hata ortalamaları karşılaştırıldığında üç model arasından en düşük ortalamanın TBMTK'den elde edildiği görülmüştür. a_2 parametresi incelendiğinde MTTK'nin, 2FM'den daha düşük ortalama hata verdiği belirlenmiştir. Benzer şekilde her üç modelden elde edilen c parametresinin ortalama hata değerleri bağımsız madde sayısı artışla birlikte azalma göstermiştir. Ancak bütün modellerde c parametresinin ortalaması hakkında

genel bir artış ya da azalış gözlenmemiştir. Hatalarında ise bağımsız madde artışıyla birlikte bir azalma olduğu görülmektedir. Veri setinde bulunan bağımsız madde sayısı arttıkça farklı modellerden elde edilen c parametresine ait ortalama değerlerin benzer değerler ürettiği görülmektedir.

1000 örneklem büyüklüğünde veri setine bağımsız madde eklenmesiyle oluşturulan farklı veri setlerinin TBMTK ile analizi sonucu elde edilen a_1 parametre ortalamaları ve bu parametrelere ait ortalama hata değerleri için genel bir yorum yapılamadığı görülmüştür. Veri setine bağımsız madde eklenmesiyle 2FM'den ve MTTK'den elde edilen a_1 ve a_2 parametrelerinin ve bu parametrelere ait hata değerlerinin ortalamasının genel olarak azaldığı belirlenmiştir. a_1 parametresi için bütün veri setlerinde en düşük hata ortalamasına sahip model TBMTK ile; a_2 parametresi için ise MTTK modelidir. Her koşulda MTTK, 2FM'den daha düşük hatalı eğim parametreleri kestirmiştir. TBMTK ile hesaplanan c parametresi ise bağımsız madde sayısının en fazla olduğu durumda en yüksek ortalama değeri ve en az ortalama hata değerini vermiştir. 2FM'den ve MTTK'den elde edilen c parametresinin ortalama değerinin bağımsız madde sayısındaki artışla birlikte değişim gösterdiği, ancak bu değişimin genel bir yorum yapmaya olanak sağlamadığı; bu parametreye ait ortalama hata değerlerinin ise veri setinde bulunan bağımsız madde sayısındaki artışla birlikte genel olarak azaldığı gözlenmiştir.

Tablo 4.2'de örneklem büyüklüğü arttıkça bütün modellerden elde edilen a ve c parametrelerine ait ortalama hata değerlerinin genel olarak azaldığı, ayrıca örneklem büyüklüğündeki artış ile birlikte bütün veri setleri için farklı modellerden elde edilen a ve c parametreleri hata ortalamalarının birbirine yaklaştığı görülmektedir. Özellikle 2FM ve MTTK modellerinden elde edilen hata ortalamaları örneklem büyüklüğünün en fazla olduğu 1000 örneklem için oldukça azalmıştır. Bütün modeller içerisinde a ve c parametrelerine ait en az hatalı kestirimler, 1000 örneklem büyüklüğünde madde sayısının en fazla olduğu veri seti 6 için gerçekleşmiştir.

Tablo 4.1 ve Tablo 4.2'de yer alan iki kategorili maddelere ait madde parametreleri ve hata kestirimleri ortalamaları genel olarak yorumlandığında kısaca şu bulgular elde edilmiştir. Farklı örneklem büyüklüklerinde, farklı modellerden ve farklı veri setlerinden elde edilen eğim parametresi için genel olarak en düşük ortalama hata değerlerinin TBMTK'den elde edildiği; MTTK'nin, 2FM'den daha düşük ortalama

hata kestirimi yaptığı; c parametresinin ortalama hatasının ise her zaman genel bir eğilim göstermediği belirlenmiştir. Özellikle 250 örneklem büyüklüğünde farklı modellerden hesaplanan sonuçların farklılaştığı, ancak örneklem büyüklüğü arttıkça farklı modellerden elde edilen parametre hata değerleri ortalamasının birbirine yaklaştığı; küçük örneklem büyüklüklerinde 2FM'de ve MTTK'de bulguların belirli bir örüntü göstermediği belirlenmiştir. Ayrıca veri setindeki bağımsız madde sayısının artmasıyla parametrelere ait ortalama hata değerlerinde daha çok azalma eğilimi olduğu görülmüştür. Bu durum madde sayısının artmasının kestirimler üzerinde etkili olmasından kaynaklanıyor olabilir. Ancak madde sayısının ve örneklem büyüklüğünün eğim parametrelerini daha çok etkilediği; kesişim parametresindeki etkisinin daha az olduğu söylenebilir. Veri setine madde takımı eklenmesiyle ise genel olarak parametrelere ait hata ortalamaları artarken, en az değişkenlik gösteren parametrenin c parametresi olduğu belirlenmiştir. Bu bulgular ışığında özellikle genel boyutun yanında özel boyutun da söz konusu olduğu modellerde, küçük örneklem büyüklüğünün hata kestirimleri üzerinde önemli bir etkisi olduğu ifade edilebilir. Ayrıca madde sayısının artmasının da genel boyuta ait a parametresinin hata ortalamalarının azalmasına neden olduğu araştırmanın bir diğer bulgusudur.

a_1 : genel boyuta ait eğim parametresi, a_2 : ikincil boyuta ait eğim parametresi, γ : kesişim parametresi, SH: standart hata olmak üzere, bütün veri setlerinde yer alan çok kategorili puanlanan maddeler için elde edilen madde parametrelerine ait ortalamalara ve hata ortalamalarına ilişkin bulgular Tablo 4.3'te sunulmuştur.

Tablo 4.3: Farklı Veri Setlerinde Yer Alan Çok Kategorili Puanlanan Maddelerin Madde Parametrelerinin ve Bu Parametrelere ait Standart Hatalarının Ortalamaları

Veri	Örneklem	TBMTK			2FM				MTTK			
		a (SH)	Y ₁ (SH)	Y ₂ (SH)	a ₁ (SH)	a ₂ (SH)	Y ₁ (SH)	Y ₂ (SH)	a ₁ (SH)	a ₂ (SH)	Y ₁ (SH)	Y ₂ (SH)
Veri Seti 1	250	1,57 (0,35)	-1,37 (0,34)	-0,62 (0,27)	1,53 (0,14)	0,63 (0,20)	-1,35 (0,10)	-0,54 (0,12)	1,35 (0,14)	1,35 (0,14)	-0,99 (0,11)	-0,45 (0,12)
	500	1,20 (0,16)	-0,83 (0,14)	-0,46 (0,16)	1,20 (0,19)	0,47 (0,26)	-0,81 (0,13)	-0,41 (0,18)	1,25 (0,17)	1,25 (0,17)	-0,82 (0,13)	-0,39 (0,15)
	1000	1,31 (0,13)	-0,97 (0,11)	-0,49 (0,12)	1,27 (0,14)	0,39 (0,20)	-0,95 (0,10)	-0,47 (0,12)	1,34 (0,14)	1,34 (0,14)	-0,71 (0,11)	-0,42 (0,12)
Veri Seti 2	250	1,07 (0,19)	-0,74 (0,20)	-0,50 (0,21)	1,12 (0,66)	0,24 (1,25)	-0,74 (0,16)	-0,45 (0,53)	1,15 (0,34)	1,15 (0,34)	-0,73 (0,15)	-0,45 (0,40)
	500	1,14 (0,14)	-0,78 (0,13)	-0,48 (0,16)	1,18 (0,17)	0,45 (0,23)	-0,79 (0,12)	-0,43 (0,17)	1,17 (0,16)	1,17 (0,16)	-0,79 (0,10)	-0,44 (0,16)
	1000	1,23 (0,11)	-0,94 (0,10)	-0,52 (0,11)	1,29 (0,14)	0,48 (0,20)	-0,96 (0,11)	-0,46 (0,13)	1,29 (0,12)	1,29 (0,12)	-0,95 (0,09)	-0,45 (0,12)
Veri Seti 3	250	1,32 (0,47)	-1,38 (0,77)	-1,59 (0,62)	1,24 (0,47)	0,47 (0,82)	-1,20 (0,23)	-0,47 (0,42)	1,28 (0,69)	1,28 (0,90)	-1,21 (0,24)	-0,43 (0,56)
	500	1,13 (0,15)	-1,18 (0,17)	-0,68 (0,17)	1,19 (0,17)	0,49 (0,22)	-1,18 (0,14)	-0,62 (0,20)	1,20 (0,16)	1,20 (0,15)	-1,18 (0,15)	-0,61 (0,18)
	1000	1,23 (0,11)	-1,34 (0,12)	-0,64 (0,12)	1,26 (0,12)	0,44 (0,17)	-1,35 (0,11)	-0,59 (0,13)	1,36 (0,15)	1,43 (0,17)	-1,35 (0,12)	-0,53 (0,15)
Veri Seti 4	250	0,84 (0,19)	-0,68 (0,18)	-0,57 (0,21)	1,43 (0,80)	0,40 (1,10)	-0,73 (0,24)	-0,40 (0,59)	1,44 (0,27)	1,44 (0,27)	-0,78 (0,19)	-0,39 (0,24)
	500	1,18 (0,14)	-0,64 (0,21)	-0,58 (0,15)	1,28 (0,18)	0,40 (0,24)	-0,65 (0,12)	-0,48 (0,20)	1,28 (0,20)	1,28 (0,20)	-0,65 (0,10)	-0,49 (0,18)
	1000	1,27 (0,10)	-0,79 (0,10)	-0,62 (0,11)	1,36 (0,13)	0,38 (0,17)	-0,82 (0,09)	-0,55 (0,12)	1,35 (0,11)	1,35 (0,11)	-0,80 (0,08)	-0,56 (0,12)
Veri Seti 5	250	0,82 (0,19)	-0,65 (0,24)	-0,58 (0,21)	1,34 (0,99)	0,44 (1,34)	-0,66 (0,19)	-0,45 (0,72)	1,29 (0,76)	1,29 (0,76)	-0,58 (0,16)	-0,33 (0,58)
	500	1,16 (0,13)	-0,63 (0,14)	-0,59 (0,15)	1,25 (0,18)	0,42 (0,24)	-0,63 (0,12)	-0,51 (0,17)	1,29 (0,16)	1,29 (0,16)	-0,62 (0,09)	-0,50 (0,16)
	1000	1,23 (0,11)	-0,78 (0,13)	-0,64 (0,11)	1,34 (0,13)	0,44 (0,18)	-0,80 (0,09)	-0,56 (0,12)	1,34 (0,14)	1,34 (0,14)	-0,81 (0,12)	-0,55 (0,14)
Veri Seti 6	250	1,18 (0,19)	-1,04 (0,20)	-0,59 (0,23)	1,35 (0,54)	0,54 (0,85)	-1,03 (0,24)	-0,47 (0,47)	1,28 (0,43)	1,25 (0,49)	-1,02 (0,18)	-0,53 (0,24)
	500	1,13 (0,13)	-0,96 (0,14)	-0,72 (0,17)	1,21 (0,17)	0,45 (0,23)	-0,95 (0,14)	-0,65 (0,20)	1,22 (0,15)	1,23 (0,15)	-0,95 (0,12)	-0,65 (0,18)
	1000	1,22 (0,10)	-1,12 (0,11)	-0,70 (0,12)	1,30 (0,13)	0,45 (0,17)	-1,13 (0,11)	-0,63 (0,13)	1,31 (0,13)	1,33 (0,13)	-1,13 (0,11)	-0,63 (0,13)

Tablo 4.3'te madde takım sayısı 4 iken veri setine 0/7/14 tane bağımsız madde eklenmesi ile oluşturulan veri seti 1, 2 ve 3 için elde edilen a ve Y parametreleri ile bu parametrelere ait hata değerleri yorumları aşağıdaki gibidir.

250 örneklem büyüklüğünde TBMTK'den elde edilen a₁ parametresinin ortalama değerinin ve standart hata ortalamalarının veri setine yedi tane bağımsız madde eklenmesiyle birlikte düştüğü, ancak veri setine tekrar bağımsız madde eklenmesiyle birlikte hata ortalamalarının arttığı görülmektedir. 2FM'de ve MTTK'de a₁ ve a₂ parametresi için en yüksek ortalama değer, yalnızca madde takımından oluşan veri seti için hesaplanırken, bu parametrelerin hata ortalamaları için de en düşük değer yine yalnızca madde takımından oluşan veri seti için elde edilmiştir.

TBMTK'den elde edilen Y parametresi ortalamasının veri setine yedi tane bağımsız madde eklenmesiyle birlikte arttığı, hata ortalamalarının ise azaldığı; ancak veri

setine tekrar bağımsız madde eklenmesiyle birlikte γ parametresi ortalamasının azaldığı, hata ortalamalarının ise arttığı belirlenmiştir. γ değerinin ortalaması için ise 2FM ve MTTK için genel bir yoruma erişilememiştir. Ancak madde takım sayısı sabit tutularak oluşturulan veri setlerine bağımsız madde eklenmesiyle birlikte bu parametrelerin hatalarında genel olarak bir artış olduğu görülmüştür.

500 örneklem büyüklüğünde her üç modelin a parametrelerine ait ortalama hata değerlerinin veri setine bağımsız madde eklenmesiyle genel bir azalış gösterdiği belirlenmiştir. Örneklem büyüklüğünde 500 olduğunda, γ parametrelerine ait hata ortalamalarının ise veri setine bağımsız madde eklenmesiyle birlikte ilk olarak biraz azaldığı, daha sonra bağımsız madde sayısı arttıkça hata ortalamalarının da arttığı gözlenmiştir. Her üç veri setinde 2FM'den ve MTTK'den elde edilen parametre ortalamalarının ve hata ortalamalarının birbirine benzer olduğu söylenebilir.

1000 örneklem büyüklüğünde bağımsız madde sayısının artmasıyla TBMTK'den ve 2FM'den elde edilen a parametrelerine ait hata değerlerinde azalma görülürken, MTTK'de genel bir eğilim gözlenmemiştir. γ parametrelerine ait hata ortalamaları ise her üç modelde de çok az miktarda etkilemiştir ve kısmen artış göstermiştir. Ancak bütün modellerde ve her üç veri setinde γ parametrelerine ait hata ortalamalarının birbirine yakın bulgular ürettiği belirlenmiştir.

Örneklem büyüklüğü arttıkça her üç veri setinden elde edilen madde parametreleri hata ortalamalarının genel olarak azaldığı görülmektedir. Ayrıca her üç veri seti genel olarak incelendiğinde modeller arasında standart hatanın en yüksek olduğu bulguların 250 örneklem büyüklüğünde elde edildiği; a parametresine ait hata ortalamaları için en düşük değeri TBMTK'nin verdiği; 500 ve 1000 örneklem büyüklüklerinde bağımsız madde sayısının artmasının γ parametrelerine ait hata ortalamalarını az miktarda etkilediği belirlenmiştir.

Tablo 4.3'te madde takım sayısı 8 iken veri setine 0/7/14 tane bağımsız madde eklenmesi ile oluşturulan veri seti 4, 5 ve 6 için elde edilen a ve γ parametreleri ile bu parametrelere ait hata değerleri yorumları aşağıdaki gibidir.

250 örneklem büyüklüğü için en fazla maddenin yer aldığı veri seti 6'da TBMTK ile a parametresine ait en yüksek ortalama değer elde edilirken, bu parametreye ait hataların ortalaması diğer üç veri setinde eşit olacak şekilde hesaplanmıştır. 2FM ve MTTK ile hesaplanan a parametrelerine ait ortalama değer genel olarak

düşerken, a_2 parametrelerine ait ortalama değer ise artmıştır. a parametrelerine ve γ parametrelerine ait ortalama hata değerleri için ise genel bir yorum yapılamamıştır. Ancak a parametreleri için 250 örneklem büyüklüğünde tüm veri setlerinde 2FM ile yüksek hatalar hesaplanmıştır. TBMTK'den elde edilen γ parametrelerine ait hata ortalamaları tüm veri setlerinde diğer modellerden elde edilenlere göre birbirine daha yakın sonuçlar vermiştir.

500 örneklem büyüklüğü için TBMTK'den elde edilen a parametrelerinin ortalama değeri, veri setine bağımsız madde eklenmesiyle birlikte azalış göstermiştir. 2FM ve MTTK için elde edilen a_1 ve a_2 parametresine ait ortalama hata değerlerinin azaldığı belirlenmiştir. Bütün modellerden elde edilen γ parametrelerinin ortalama değeri için ise genel olarak benzer olduğu; veri setine bağımsız madde eklenmesiyle çok fazla değişim göstermediği söylenebilir.

1000 örneklem büyüklüğünde a ve γ parametresi için ortalama değerlerin farklı veri setlerinde ciddi değişimler göstermediği ve sonuçların birbirine yakın olduğu belirlenmiştir. Ayrıca bu örneklem büyüklüğünde diğer örneklem büyüklüklerine göre her üç modelden elde edilen bulgular birbirine oldukça yakındır.

Tablo 4.3'te sunulan çok kategorili olarak puanlanan veri setlerine ait bulgular genel olarak incelendiğinde, a parametresinin tüm örneklem büyüklüklerinde, TBMTK'den elde edilen hata parametrelerinin 2FM'den ve MTTK'den daha düşük olduğu görülmektedir. a ve γ parametrelerine ait hata değerleri, her üç kuramda da, örneklem büyüklüğündeki artışla birlikte azalış göstermektedir. 2FM ve MTTK hata değerlerindeki azalış, TBMTK'den daha yüksek miktarda olmuştur. Düşük örneklem büyüklüklerinde, TBMTK lehine bulgular elde edilmiştir. Ancak, örneklem büyüklüğündeki artış ile birlikte, hata değerlerinin birbirine yaklaştığı belirlenmiştir. En optimum ve kararlı bulgular, 1000 örneklem büyüklüğünde elde edilmiştir. Düşük örneklem büyüklüklerinde ise, TBMTK'nin kullanılması daha doğru görülmektedir. veri setindeki bağımsız madde sayısının ya da madde takımı sayısının artmasının, madde parametrelerine ait hata ortalamalarına etkisi üzerine çok net bir bulguya erişilmemiştir. Farklı örneklem büyüklüklerinden ve farklı veri setlerinden elde edilen 2FM ve MTTK sonuçlarının genel olarak birbirine yakın olduğu belirlenmiştir. Farklı modellerin kullanılması durumunun özellikle γ parametresi kestirimleri üzerine az miktarda bir etkisi olduğu belirlenmiştir.

4.2. Yetenek Parametrelerinin Değerlendirilmesine İlişkin Bulgular

İkinci alt problem olan “Madde takımı içeren testlerde madde takımı sayısı (4/8), bağımsız madde sayısı (0/7/14) ve örneklem büyüklüğü (250/500/1000) değiştiğinde; farklı modellerden elde edilen yetenek parametreleri ve bu parametrelere ait standart hata değerleri nasıldır?” sorusuna ait bulgular Tablo 4.4’te ve Tablo 4.5’te yer almaktadır.

Tablo 4.4: Veri Seti 1, 2 ve 3 için Yetenek Parametresinin ve Bu Parametreye ait Hata Değerlerinin Ortalaması

Veri	Yetenek	250			500			1000		
		TBMTK	2FM	MTTK	TBMTK	2FM	MTTK	TBMTK	2FM	MTTK
		θ (SH)	θ (SH)	θ (SH)	θ (SH)	θ (SH)	θ (SH)	θ (SH)	θ (SH)	θ (SH)
Veri Seti 1 (4 madde takımı)	Teta-1	0,00 (0,44)	0,00 (0,48)	0,00 (0,47)	0,00 (0,44)	0,00 (0,46)	0,00 (0,47)	0,00 (0,45)	-0,01 (0,48)	0,00 (0,47)
	Teta-2	-	0,00 (0,21)	0,00 (0,21)	-	0,00 (0,37)	0,00 (0,39)	-	0,00 (0,39)	0,00 (0,35)
	Teta-3	-	0,00 (0,48)	0,00 (0,48)	-	0,00 (0,39)	0,00 (0,32)	-	0,00 (0,32)	0,00 (0,25)
	Teta-4	-	0,83 (0,91)	0,74 (0,88)	-	0,88 (1,00)	0,76 (0,92)	-	0,92 (0,98)	0,76 (0,90)
	Teta-5	-	0,97 (0,83)	0,88 (0,81)	-	0,89 (0,87)	0,84 (0,83)	-	0,88 (0,87)	0,85 (0,83)
Veri Seti 2 (4 madde takımı – 7 bağımsız madde)	Teta-1	0,00 (0,37)	0,00 (0,38)	0,00 (0,38)	0,00 (0,38)	0,00 (0,40)	0,00 (0,39)	0,00 (0,39)	0,00 (0,41)	0,00 (0,41)
	Teta-2	-	0,00 (0,21)	0,00 (0,19)	-	0,00 (0,35)	0,00 (0,36)	-	0,00 (0,38)	0,00 (0,38)
	Teta-3	-	0,00 (0,25)	0,00 (0,20)	-	0,00 (0,40)	0,00 (0,40)	-	0,00 (0,37)	0,00 (0,31)
	Teta-4	-	0,95 (1,00)	0,76 (0,92)	-	0,87 (0,99)	0,75 (0,91)	-	0,89 (0,98)	0,75 (0,91)
	Teta-5	-	0,97 (0,84)	0,89 (0,81)	-	0,90 (0,83)	0,85 (0,81)	-	0,89 (0,85)	0,84 (0,82)
Veri Seti 3 (4 madde takımı – 14 bağımsız madde)	Teta-1	0,00 (0,33)	-0,01 (0,43)	0,00 (0,40)	0,00 (0,34)	-0,01 (0,44)	0,00 (0,39)	0,00 (0,34)	-0,01 (0,41)	0,00 (0,35)
	Teta-2	-	0,00 (0,26)	0,00 (0,23)	-	0,00 (0,38)	0,00 (0,38)	-	0,00 (0,39)	0,00 (0,40)
	Teta-3	-	0,00 (0,34)	0,00 (0,34)	-	0,00 (0,35)	0,00 (0,36)	-	0,00 (0,36)	0,00 (0,36)
	Teta-4	-	0,87 (1,00)	0,85 (1,00)	-	0,84 (0,98)	0,73 (0,91)	-	0,87 (0,97)	0,74 (0,91)
	Teta-5	-	0,96 (0,84)	0,95 (0,84)	-	0,89 (0,83)	0,84 (0,81)	-	0,88 (0,84)	0,84 (0,81)

Tablo 4.4’te madde takım sayısı 4 iken veri setine 0/7/14 tane bağımsız madde eklenmesi ile oluşturulan veri seti 1, 2 ve 3 için her üç modelden elde edilen beş yetenek parametresi (dört özel yetenek ve bir genel yetenek) ve bu yetenek parametrelerine ait hata ortalamaları için şu bulgular elde edilmiştir.

250, 500 ve 1000 örneklem büyüklüğünde her üç modelden elde edilen θ_1 parametreleri ortalamalarının hemen hemen aynı olduğu, ancak bu parametrelere ait hata ortalamaları için en düşük değeri her koşulda TBMTK’nin verdiği

belirlenmiştir. TBMTK'den elde edilen yetenek parametrelerine ait hata ortalamalarının ise bağımsız madde sayısının artmasıyla birlikte düştüğü gözlenmiştir. 2FM ve MTTK modellerinden kestirilen yetenek ortalamaları ve yetenek ortalamalarına ait hata ortalamaları benzer olmakla birlikte, MTTK'den elde edilen hata ortalamalarının daha düşük olduğu belirlenmiştir. Ayrıca bu modellerden elde edilen son iki yeteneğin çok yüksek miktarda hatalı kestirildiği sonucuna varılmıştır. Her iki modelde de bütün örneklem büyüklükleri için en hatalı kestirime Θ_4 'te ulaşılmıştır. Bağımsız madde sayısının artmasının 250 örneklem büyüklüğünde 2FM'den ve MTTK'den elde edilen yetenek kestirimleri üzerine etkisi için genel bir yorum yapılamamışken; 500 ve 1000 örneklem büyüklüğünde veri setindeki bağımsız madde sayısı arttıkça yetenek kestirimlerine ait hata ortalamalarının genel olarak azaldığı belirlenmiştir. Veri seti 1 ve 2 için her üç modelden elde edilen yetenek parametresine ait hata ortalamalarının benzer olduğu; ancak bağımsız madde sayısının en fazla olduğu veri seti 3'te TBMTK ile kestirilen hata ortalaması ile diğer iki modelden kestirilen hata ortalamaları arasında genel olarak farklılıklar bulunduğu görülmüştür. Her koşulda Θ_1 için TBMTK'nin diğer iki modele göre daha düşük hata ortalaması verdiği; MTTK'nin ise 2FM'den daha düşük ortalama hata verdiği belirlenmiştir.

Örneklem büyüklüğü arttıkça TBMTK'den elde edilen yetenek parametrelerine ait ortalamaların değişmediği, hata ortalamalarının ise çok az değişerek arttığı belirlenmiştir. Örneklem büyüklüğünün artması, veri seti 1 ve 2 için 2FM'den ve MTTK'den elde edilen yetenek parametresi ve bu parametrelerin hata ortalaması üzerinde kısmen bir artış etkisi yaratmıştır. Ancak veri seti 3'te ise örneklem büyüklüğü arttıkça Θ_1 için hata ortalamaları azalmıştır. Bu bulgulara dayanarak örneklem büyüklüğü arttıkça modellerden elde edilen yetenek parametresinin ve hata ortalamasının çok değişmediği, bireyin yetenek kestirimlerinin örneklem büyüklüğüne pek bağlı olmadığı ifade edilebilir.

Tablo 4.5: Veri Seti 4, 5 ve 6 için Yetenek Parametresinin ve Bu Parametreye ait Hata Değerlerinin Ortalaması

Yetenek	250			500			1000			
	TBMTK	2FM	MTTK	TBMTK	2FM	MTTK	TBMTK	2FM	MTTK	
	θ (SH)	θ (SH)	θ (SH)	θ (SH)	θ (SH)	θ (SH)	θ (SH)	θ (SH)	θ (SH)	
Veri Seti 4 (8 madde takımı)	Teta-1	0,00 (0,34)	-0,01 (0,38)	-0,01 (0,38)	0,00 (0,35)	0,02 (0,41)	-0,01 (0,40)	0,00 (0,36)	0,00 (0,36)	0,00 (0,36)
	Teta-2	-	0,00 (0,26)	0,00 (0,32)	-	0,00 (0,44)	0,00 (0,42)	-	0,00 (0,41)	0,00 (0,40)
	Teta-3	-	0,00 (0,33)	0,00 (0,33)	-	0,00 (0,22)	0,00 (0,23)	-	0,00 (0,24)	0,00 (0,22)
	Teta-4	-	0,00 (0,32)	0,00 (0,30)	-	0,00 (0,31)	0,00 (0,31)	-	0,00 (0,36)	0,00 (0,33)
	Teta-5	-	0,00 (0,35)	0,00 (0,35)	-	0,02 (0,37)	0,00 (0,37)	-	0,00 (0,37)	0,00 (0,37)
	Teta-6	-	0,88 (0,98)	0,88 (0,98)	-	0,87 (0,99)	0,87 (0,98)	-	0,90 (0,97)	0,75 (0,90)
	Teta-7	-	0,95 (0,86)	0,92 (0,86)	-	0,85 (0,86)	0,86 (0,86)	-	0,87 (0,86)	0,84 (0,82)
	Teta-8	-	0,92 (0,95)	0,92 (0,95)	-	0,96 (1,00)	0,96 (1,00)	-	0,95 (1,00)	0,86 (0,92)
	Teta-9	-	0,92 (0,82)	0,93 (0,82)	-	0,90 (0,76)	0,92 (0,80)	-	0,90 (0,81)	0,83 (0,78)
Veri Seti 5 (8 madde takımı – 7 bağımsız madde)	Teta-1	0,00 (0,31)	-0,02 (0,49)	-0,02 (0,34)	0,00 (0,32)	0,00 (0,41)	-0,01 (0,41)	0,00 (0,32)	0,00 (0,38)	0,00 (0,32)
	Teta-2	-	0,00 (0,27)	0,00 (0,30)	-	0,00 (0,43)	0,00 (0,41)	-	0,00 (0,41)	0,00 (0,40)
	Teta-3	-	0,00 (0,28)	0,00 (0,29)	-	0,00 (0,21)	0,00 (0,22)	-	0,00 (0,25)	0,00 (0,22)
	Teta-4	-	0,00 (0,47)	0,00 (0,30)	-	0,00 (0,31)	0,00 (0,30)	-	0,00 (0,36)	0,00 (0,33)
	Teta-5	-	0,00 (0,32)	0,00 (0,32)	-	0,00 (0,33)	0,00 (0,33)	-	0,00 (0,34)	0,00 (0,34)
	Teta-6	-	0,81 (1,00)	0,91 (0,99)	-	0,86 (0,98)	0,86 (0,98)	-	0,88 (0,97)	0,74 (0,90)
	Teta-7	-	0,95 (0,85)	0,93 (0,85)	-	0,86 (0,84)	0,87 (0,84)	-	0,86 (0,85)	0,84 (0,82)
	Teta-8	-	0,93 (0,95)	0,93 (0,95)	-	0,97 (1,00)	0,96 (1,00)	-	0,95 (1,00)	0,86 (0,92)
	Teta-9	-	0,93 (0,83)	0,93 (0,83)	-	0,92 (0,78)	0,92 (0,82)	-	0,90 (0,82)	0,83 (0,78)
Veri Seti 6 (48madde takımı – 14 bağımsız madde)	Teta-1	0,00 (0,28)	0,00 (0,52)	-0,03 (0,41)	0,00 (0,29)	0,00 (0,44)	-0,14 (0,45)	0,00 (0,30)	0,00 (0,41)	0,00 (0,35)
	Teta-2	-	0,00 (0,29)	0,00 (0,32)	-	0,00 (0,43)	0,00 (0,36)	-	0,00 (0,42)	0,00 (0,40)
	Teta-3	-	0,00 (0,28)	0,00 (0,29)	-	0,00 (0,24)	-0,01 (0,33)	-	0,00 (0,26)	0,00 (0,24)
	Teta-4	-	0,00 (0,44)	0,00 (0,31)	-	0,00 (0,30)	0,00 (0,25)	-	0,00 (0,36)	0,00 (0,33)
	Teta-5	-	0,00 (0,29)	0,00 (0,29)	-	0,00 (0,30)	-0,01 (0,31)	-	0,00 (0,31)	0,00 (0,31)
	Teta-6	-	0,78 (1,00)	0,87 (0,99)	-	0,84 (0,98)	0,83 (0,94)	-	0,87 (0,97)	0,74 (0,90)
	Teta-7	-	0,93 (0,84)	0,92 (0,84)	-	0,86 (0,83)	0,91 (0,89)	-	0,86 (0,84)	0,83 (0,81)
	Teta-8	-	0,94 (0,97)	0,93 (0,97)	-	0,96 (1,00)	0,91 (0,96)	-	0,95 (1,00)	0,86 (0,92)
	Teta-9	-	0,86 (0,84)	0,93 (0,85)	-	0,92 (0,82)	0,95 (0,86)	-	0,90 (0,84)	0,83 (0,79)

Tablo 4.5'te madde takım sayısı 8 iken veri setine 0/7/14 tane bağımsız madde eklenmesi ile oluşturulan veri seti 4, 5 ve 6 için her üç modelden elde edilen dokuz

yetenek parametresi (sekiz özel yetenek ve bir genel yetenek) ve bu yetenek parametrelerine ait hata ortalamaları için şu bulgular elde edilmiştir.

250, 500 ve 1000 örneklem büyüklüklerinde her üç modelden elde edilen Θ_1 incelendiğinde TBMTK'nin genel olarak diğer iki modelden daha az hatalı yetenek kestirimi yaptığı; MTTK hata ortalamalarının ise 2FM sonuçlarına göre daha düşük olduğu belirlenmiştir. Her üç modelde de genel yeteneğe ait hata ortalamalarının veri setindeki bağımsız madde sayısının artmasıyla birlikte düştüğü gözlenmiştir. 2FM ve MTTK modellerinden kestirilen diğer özel yeteneklerin hata ortalamalarının ise genel olarak benzer sonuçlar verdiği söylenebilir. Her iki modelde de bütün örneklem büyüklükleri için en hatalı kestirimlere Θ_8 'de ulaşılmıştır.

Örneklem büyüklüğü arttıkça TBMTK'den elde edilen yetenek parametrelerine ait ortalamaların değişmediği, hata ortalamalarının ise çok az değişerek arttığı belirlenmiştir. 2FM'den ve MTTK'den elde edilen genel yetenek için en az hatalı sonuçlara 1000 örneklem büyüklüğünde ulaşılmıştır ve 1000 örneklem büyüklüğünde her üç modelin hata kestirim ortalamaları daha yakın sonuçlar vermiştir. Diğer yeteneklerin hata ortalamalarına ait örneklem büyüklüğü ile ilişkili bir örüntü bulunamamıştır. Bu bulgulara dayanarak örneklem büyüklüğü arttıkça modellerden elde edilen genel yeteneğe ait ve hata ortalamasının düştüğü, özel yetenek kestirimlerinin ise örneklem büyüklüğüne bağlı olmadığı ifade edilebilir.

Tablo 4.4'te ve Tablo 4.5'te yer alan bulgular ışığında yetenek kestirimi için TBMTK'nin 2FM'ye ve MTTK'ye göre daha düşük hata ortalaması verdiği; MTTK'nin ise 2FM'den daha düşük ortalama hata verdiği belirlenmiştir. Her üç modelde de genel yeteneğe ait hata ortalamalarının veri setindeki bağımsız madde sayısının artmasıyla birlikte düştüğü gözlenmiştir. Ayrıca örneklem büyüklüğünün madde takım sayısının dört olduğu koşullarda pek etkili olmadığı, madde takım sayısının sekiz olduğu koşullarda ise genel yetenek için hata ortalamasının düştüğü ve özel yetenek için belli bir örüntünün bulunmadığı sonucuna varılmıştır.

4.3. Farklı Modellerden Elde Edilen Uyum İndekslerine İlişkin Bulgular

Üçüncü alt problem olan “Madde takımı içeren testlerde madde takımı sayısı (4/8), bağımsız madde sayısı (0/7/14) ve örneklem büyüklüğü (250/500/1000) değiştiğinde; farklı modellerden elde edilen uyum iyiliği indeksleri nasıldır?” sorusuna ait bulgular Tablo 4.6'da yer almaktadır.

Tablo 4.6: Uyum İyiliği İndeks Değerleri

	<i>Veri Setleri</i>	<i>Model</i>	<i>Parametre Sayısı</i>	<i>-2 LL*</i>	<i>χ^2 Fark Testi</i>
250	Veri Seti 1	TBMTK (1)	24	3116,23	$\chi^2(9) = 28,66^{**}$ (1-2)
		2FM (2)	33	3087,57	$\chi^2(4) = 21,70^{**}$ (1-3)
		MTTK (3)	28	3094,53	$\chi^2(5) = 6,96$ (2-3)
	Veri Seti 2	TBMTK (1)	38	4946,29	$\chi^2(9) = 13,83$ (1-2)
		2FM (2)	47	4932,46	$\chi^2(4) = 10,66^{**}$ (1-3)
		MTTK (3)	42	4935,63	$\chi^2(5) = 3,17$ (2-3)
	Veri Seti 3	TBMTK (1)	53	6479,01	$\chi^2(9) = 18,61^{**}$ (1-2)
		2FM (2)	62	6460,40	$\chi^2(4) = 15,61^{**}$ (1-3)
		MTTK (3)	57	6463,40	$\chi^2(5) = 3,00$ (2-3)
	Veri Seti 4	TBMTK (1)	47	5895,90	$\chi^2(18) = 29,49^{**}$ (1-2)
		2FM (2)	65	5866,41	$\chi^2(8) = 26,88^{**}$ (1-3)
		MTTK (3)	55	5869,02	$\chi^2(10) = 2,61$ (2-3)
	Veri Seti 5	TBMTK (1)	61	7701,80	$\chi^2(18) = 28,60$ (1-2)
		2FM (2)	79	7673,20	$\chi^2(8) = 19,04^{**}$ (1-3)
		MTTK (3)	69	7682,76	$\chi^2(10) = 9,56$ (2-3)
	Veri Seti 6	TBMTK (1)	76	9195,38	$\chi^2(18) = 29,56^{**}$ (1-2)
		2FM (2)	94	9165,82	$\chi^2(8) = 19,04^{**}$ (1-3)
		MTTK (3)	84	9176,34	$\chi^2(10) = 10,52$ (2-3)
500	Veri Seti 1	TBMTK (1)	24	6295,68	$\chi^2(9) = 34,97^{**}$ (1-2)
		2FM (2)	33	6260,71	$\chi^2(4) = 29,81^{**}$ (1-3)
		MTTK (3)	28	6265,87	$\chi^2(5) = 5,16$ (2-3)
	Veri Seti 2	TBMTK (1)	38	9954,34	$\chi^2(9) = 43,87^{**}$ (1-2)
		2FM (2)	47	9910,47	$\chi^2(4) = 43,62^{**}$ (1-3)
		MTTK (3)	42	9910,72	$\chi^2(5) = 0,25$ (2-3)
	Veri Seti 3	TBMTK (1)	53	13110,84	$\chi^2(9) = 57,12^{**}$ (1-2)
		2FM (2)	62	13053,72	$\chi^2(4) = 49,26^{**}$ (1-3)
		MTTK (3)	57	13061,58	$\chi^2(5) = 7,86$ (2-3)
	Veri Seti 4	TBMTK (1)	47	11916,49	$\chi^2(18) = 72,22^{**}$ (1-2)
		2FM (2)	65	11844,27	$\chi^2(8) = 54,07^{**}$ (1-3)
		MTTK (3)	55	11862,82	$\chi^2(10) = 18,15$ (2-3)
	Veri Seti 5	TBMTK (1)	61	15496,51	$\chi^2(18) = 76,60^{**}$ (1-2)
		2FM (2)	79	15419,91	$\chi^2(8) = 61,62^{**}$ (1-3)
		MTTK (3)	69	15434,89	$\chi^2(10) = 14,98$ (2-3)
	Veri Seti 6	TBMTK (1)	76	18600,81	$\chi^2(18) = 59,78^{**}$ (1-2)
		2FM (2)	94	18541,03	$\chi^2(8) = 60,14^{**}$ (1-3)
		MTTK (3)	84	18540,67	$\chi^2(10) = 0,36$ (2-3)
1000	Veri Seti 1	TBMTK (1)	24	12581,88	$\chi^2(9) = 60,96^{**}$ (1-2)
		2FM (2)	33	12520,92	$\chi^2(4) = 39,59^{**}$ (1-3)
		MTTK (3)	28	12542,29	$\chi^2(5) = 21,37^{**}$ (2-3)
	Veri Seti 2	TBMTK (1)	38	19914,82	$\chi^2(9) = 64,01^{**}$ (1-2)
		2FM (2)	47	19850,81	$\chi^2(4) = 56,32^{**}$ (1-3)
		MTTK (3)	42	19858,50	$\chi^2(5) = 7,69$ (2-3)
	Veri Seti 3	TBMTK (1)	53	26230,22	$\chi^2(9) = 82,93^{**}$ (1-2)
		2FM (2)	62	26147,29	$\chi^2(4) = 83,02^{**}$ (1-3)
		MTTK (3)	57	26147,20	$\chi^2(5) = 0,09$ (2-3)
	Veri Seti 4	TBMTK (1)	47	23767,04	$\chi^2(18) = 112,95^{**}$ (1-2)
		2FM (2)	65	23654,09	$\chi^2(8) = 113,31^{**}$ (1-3)
		MTTK (3)	55	23653,73	$\chi^2(10) = 0,36$ (2-3)
	Veri Seti 5	TBMTK (1)	61	30952,66	$\chi^2(18) = 118,40^{**}$ (1-2)
		2FM (2)	79	30834,26	$\chi^2(8) = 118,80^{**}$ (1-3)
		MTTK (3)	69	30833,86	$\chi^2(10) = 0,40$ (2-3)
	Veri Seti 6	TBMTK (1)	76	37180,56	$\chi^2(18) = 120,64^{**}$ (1-2)
		2FM (2)	94	37059,92	$\chi^2(8) = 112,94^{**}$ (1-3)
		MTTK (3)	84	37067,62	$\chi^2(10) = 7,70$ (2-3)

*-2LL: -2 Log-olabilirlik, **p<.05

Tablo 4.6'da farklı örneklem büyüklüklerindeki farklı veri setlerine ait model-veri uyumu indeksleri özetlenmiştir. Her üç MTK modelinden elde edilen -2LL değerleri arasındaki fark, kestirilen parametre sayılarından hesaplanan serbestlik derecesine ve 0,05 önem derecesine göre belirlenen χ^2 tablo değeri ile karşılaştırılmıştır. Bu bulgulara göre 250 örneklem büyüklüğündeki bütün veri setleri için 2FM ve MTTK

modelinin, TBMTK'ye göre daha uyumlu modeller olduğu belirlenmiştir. Ancak 2FM ile MTTK arasında anlamlı farklılık olmadığı ve her iki modelin de veri setiyle uyumlu olduğu söylenebilir.

500 örneklem büyüklüğünde tüm veri setleri için 2FM'nin ve MTTK'nin TBMTK'den daha tercih edilebilir olduğu belirlenmiştir. Bütün veri setlerinde 2FM ve MTTK arasında anlamlı fark bulunamamış ve her iki modelin de veriye uyum sağlayabileceği sonucuna varılmıştır.

1000 örneklem büyüklüğünden elde edilen model-veri uyumları incelendiğinde, tüm veri setleri için 2FM'nin ve MTTK'nin TBMTK'den daha iyi model-veri uyumu sağladığı görülmektedir. Ancak veri seti 1 için 2FM ile MTTK modeli arasından, 2FM'nin daha iyi model-veri uyumunu sağladığı belirlenmiştir. Bu durum aynı madde takımlarında yer alan madde çiftlerine ait LD X^2 değerleri arasında en büyük yerel bağımlılığın 1000 örneklem büyüklüğündeki veri seti 1'de olmasından kaynaklanıyor olabilir. Diğer beş veri seti için ise 2FM ve MTTK arasında anlamlı fark bulunamamış ve her iki modelin de veriye uyum sağlayabileceği görülmüştür.

Bütün koşullarda yerel bağımlılığı göz önüne alan modellerin TBMTK'den daha iyi model veri uyumu sağladığı, genel olarak 2FM ile MTTK arasında anlamlı farklılık olmadığı; her iki modelin de bu veri setleri için kullanılabileceği belirlenmiştir. Bu iki model arasında anlamlı farklılık olduğu durumda ise genel olarak 2FM'nin daha iyi sonuç verdiği söylenebilir. Düşük örneklemde bile maddeler arasında yerel bağımlılık söz konusu olduğunda 2FM'nin ve MTTK'nin daha iyi uyum göstermesi dikkat çekici bir bulgudur. Bağımsız madde sayısının ve madde takımı sayısının model-veri uyumuna etkisi ile ilgili genel bir yorum yapılamamıştır.

4.4. Farklı Modellerden Elde Edilen Madde Parametrelerinin ve Madde Parametrelerine ait Standart Hataların Korelasyonlarına İlişkin Bulgular

Dördüncü alt problem olan “Madde takımı içeren testlerde madde takımı sayısı (4/8), bağımsız madde sayısı (0/7/14) ve örneklem büyüklüğü (250/500/1000) değiştiğinde; farklı modellerden elde edilen madde parametreleri ile bu parametrelere ait hata değerleri arasındaki korelasyonlar nasıldır?” sorusuna ait bulgular Tablo 4.7'de yer almaktadır.

Tablo 4.7: Farklı Modellerden Elde Edilen Madde Parametrelerine ve Madde Parametrelerinin Standart Hata Değerlerine ait Korelasyon Değerleri

Örneklem Büyüküğü	Veri Setleri	TBMTK-2FM		TBMTK-MTTK		2FM-MTTK	
		Eğim (SH)	Kesişim (SH)	Eğim (SH)	Kesişim (SH)	Eğim (SH)	Kesişim (SH)
250	Veri Seti 1	0,96 (0,47)	1,00 (0,84)	0,89 (0,65)	0,99 (0,89)	0,96 (0,88)	0,99 (0,97)
	Veri Seti 2	0,99 (0,59)	1,00 (0,87)	0,99 (0,79)	1,00 (0,74)	0,99 (0,86)	1,00 (0,92)
	Veri Seti 3	0,96 (0,54)	0,97 (0,97)	0,96 (0,53)	0,97 (0,94)	1,00 (0,97)	1,00 (0,97)
	Veri Seti 4	0,93 (0,52)	1,00 (0,85)	0,92 (0,59)	1,00 (0,87)	0,88 (0,87)	1,00 (0,96)
	Veri Seti 5	0,93 (0,55)	0,99 (0,87)	0,94 (0,53)	1,00 (0,80)	0,99 (0,86)	1,00 (0,92)
	Veri Seti 6	0,98 (0,62)	1,00 (0,78)	0,74 (0,95)	0,99 (0,66)	0,70 (0,92)	1,00 (0,97)
500	Veri Seti 1	0,95 (0,83)	1,00 (0,96)	0,98 (0,96)	1,00 (0,98)	0,98 (0,80)	1,00 (0,97)
	Veri Seti 2	0,97 (0,87)	0,99 (0,96)	0,98 (0,83)	1,00 (0,96)	0,99 (0,88)	1,00 (0,96)
	Veri Seti 3	0,98 (0,85)	1,00 (0,96)	0,97 (0,77)	1,00 (0,88)	0,99 (0,92)	1,00 (0,86)
	Veri Seti 4	0,93 (0,74)	0,98 (0,85)	0,95 (0,78)	0,99 (0,82)	0,93 (0,95)	1,00 (0,91)
	Veri Seti 5	0,96 (0,64)	0,99 (0,81)	0,98 (0,93)	0,99 (0,74)	0,99 (0,72)	1,00 (0,89)
	Veri Seti 6	0,97 (0,74)	0,99 (0,77)	0,98 (0,87)	0,99 (0,87)	0,99 (0,80)	1,00 (0,80)
1000	Veri Seti 1	0,93 (0,91)	1,00 (0,95)	0,99 (0,92)	1,00 (0,98)	0,96 (0,93)	1,00 (0,96)
	Veri Seti 2	0,96 (0,70)	1,00 (0,97)	0,99 (0,57)	1,00 (0,92)	0,97 (0,92)	1,00 (0,92)
	Veri Seti 3	0,97 (0,81)	0,99 (0,94)	0,98 (0,61)	1,00 (0,98)	0,97 (0,90)	1,00 (0,93)
	Veri Seti 4	0,96 (0,86)	0,99 (0,91)	0,98 (0,89)	1,00 (0,95)	0,98 (0,91)	1,00 (0,99)
	Veri Seti 5	0,97 (0,67)	0,99 (0,71)	0,98 (0,83)	1,00 (0,91)	0,99 (0,78)	1,00 (0,88)
	Veri Seti 6	0,92 (0,73)	0,97 (0,86)	0,97 (0,73)	0,99 (0,84)	1,00 (1,00)	1,00 (0,94)

Tablo 4.7'deki bulgular incelendiğinde 250 örneklem büyüklüğünde her üç modele ait a parametresi kestirimlerinin korelasyonlarının 0,74 ile 1,00 arasında değiştiği görülmektedir. Bu nedenle her üç modelden elde edilen kestirim sonuçlarının altı farklı veri seti için oldukça benzer olduğu söylenebilir. Ancak hatalara ait korelasyon değerleri incelendiğinde TBMTK-MTTK'nin ikinci ve altıncı veri seti için ve 2FM-MTTK'nin ise tüm veri setleri için yüksek korelasyon verdiği belirlenmiştir. TBMTK-2FM arasında diğer modellere göre daha düşük korelasyon aralığı hesaplanmıştır (0,52 - 0,62). c parametresine ait kestirimlerin korelasyon değerlerinin 0,66 ile 1,00; hata kestirimlerinin korelasyon aralığının ise 0,66 ile 0,97 arasında değiştiği görülmektedir. a parametresine göre c parametresinin hataları arasındaki korelasyon her üç modelde de daha yüksektir. Bu bulgulara dayanarak genel olarak bütün modeller arasında a ve c parametrelerine ait hesaplanan değerlerin benzer ve korelasyon değerlerinin yüksek olduğu; hata değerlerinin ise kullanılan veri setine ve modele göre değişkenlik gösterdiği söylenebilir.

500 örneklem büyüklüğünde her veri seti için üç modelden elde edilen a parametresinin ikili kombinasyonlarıyla elde edilen korelasyon değerleri incelendiğinde, ranjin (0,93) – (1,00) olduğu ve bütün korelasyon değerlerin oldukça yüksek olduğu görülmektedir. a parametresine ait hata değerleri korelasyonlarına bakıldığında ise korelasyon aralığının ise 0,64 ile 0,96 olduğu; TBMTK-2FM arasında 1., 2., 3., 4. ve 6. veri seti için; TBMTK-MTTK arasında ve 2FM-MTTK arasında ise tüm veri setleri için korelasyon değeri yüksek olarak hesaplanmıştır. c

parametresine ait kestirimlerin korelasyon değerlerinin 0,99 ile 1,00; hata kestirimlerinin korelasyon aralığının ise 0,74 ile 0,97 olduğu görülmektedir. a parametresine göre c parametresinin hataları arasındaki korelasyon her üç modelde de daha yüksektir. Bu bulgulara dayanarak genel olarak bütün modeller arasında a ve c parametrelerinin ve bu parametrelere ait hata değerlerinin korelasyon değerlerinin yüksek olduğu söylenebilir.

1000 örneklem büyüklüğünde bütün veri setlerine ait a parametresinin her üç modelle hesaplanan değerleri arasındaki korelasyon incelendiğinde, ranjin 0,90 – 1,00 olduğu ve bütün korelasyon değerlerin oldukça yüksek olduğu görülmektedir. a parametresine ait hata değerleri korelasyonlarına bakıldığında ise korelasyon aralığının ise 0,57 ile 1,00 olduğu; TBMTK-2FM arasında 1., 2., 3. 4. ve 6. veri seti için; TBMTK-MTTK arasında 1., 4., 5. ve 6. veri seti için; 2FM-MTTK arasında ise tüm veri setleri için korelasyon değeri yüksek olarak hesaplanmıştır. c parametresine ait kestirimlerin korelasyon değerlerinin 0,97 ile 1,00; hata kestirimlerinin korelasyon aralığının ise 0,71 ile 1,00 olduğu görülmektedir. a parametresine göre c parametresinin hataları arasındaki korelasyon her üç modelde de daha yüksektir. Bu bulgulara dayanarak genel olarak bütün modeller arasında a ve c parametrelerine ait hesaplanan korelasyon değerlerinin yüksek olduğu söylenebilir. Bu parametrelere ait hata korelasyonları ise daha düşük olmakla birlikte, farklı modellerden elde edilen hata korelasyonları için de yüksek bulgulara ulaşılmıştır.

Her bir örneklem büyüklüğünde, 2FM-MTTK arasındaki korelasyon katsayılarının yüksek düzeyde olduğu belirlenmiştir. Bu durum, bu iki modelin madde parametrelerine ait hata değerleri için benzer kestirim yaptığını, bu nedenle de madde parametrelerine ait tahminlemelerde birbirleri yerine kullanılabileceğini göstermektedir. Farklı veri setlerinden ve farklı örneklem büyüklüklerinden hesaplanan eğim ve kesişim parametresi için eğim parametresinin modele ve örneklem büyüklüğüne daha duyarlı olduğu, kesişim parametresinin ise her durumda daha benzer sonuçlar verdiği ifade edilebilir. Ancak hesaplanan korelasyon değerlerinin veri setinde yer alan madde takım sayısı ya da bağımsız madde sayısı ile ilişkisi hakkında herhangi bir yorum yapılamamıştır.

4.5. Farklı Modellerden Elde Edilen Yetenek Parametrelerinin ve Yetenek Parametrelerine ait Standart Hataların Korelasyonlarına İlişkin Bulgular

Beşinci alt problem olan “Madde takımı içeren testlerde madde takımı sayısı (4/8), bağımsız madde sayısı (0/7/14) ve örneklem büyüklüğü (250/500/1000) değiştiğinde; farklı modellerden elde edilen yetenek parametreleri ile bu parametrelere ait hata değerleri arasındaki korelasyonlar nasıldır?” sorusuna ait bulgular Tablo 4.8’de yer almaktadır.

Tablo 4.8: Farklı Modellerden Elde Edilen Yetenek Parametrelerine ve Yetenek Parametrelerinin Hata Değerlerine ait Korelasyon Değerleri

Örneklem Büüklüğü	Veri Setleri	TBMTK-2FM		TBMTK-MTTK		2FM-MTTK	
		Yetenek	Standart Hata	Yetenek	Standart Hata	Yetenek	Standart Hata
250	Veri Seti 1	0,994	0,984	0,995	0,997	0,990	0,990
	Veri Seti 2	0,999	0,981	0,999	0,999	1,000	0,984
	Veri Seti 3	0,999	0,997	0,999	0,996	1,000	0,998
	Veri Seti 4	0,999	0,991	0,999	0,992	1,000	0,999
	Veri Seti 5	0,999	0,996	0,999	0,994	1,000	0,996
	Veri Seti 6	0,999	0,998	0,999	0,987	1,000	0,987
500	Veri Seti 1	0,995	0,971	0,998	0,993	0,998	0,974
	Veri Seti 2	0,998	0,984	0,998	0,995	1,000	0,991
	Veri Seti 3	0,998	0,995	0,998	0,994	1,000	0,999
	Veri Seti 4	0,998	0,989	0,998	0,988	1,000	0,993
	Veri Seti 5	0,999	0,994	0,999	0,994	1,000	0,999
	Veri Seti 6	0,999	0,998	0,999	0,992	1,000	0,993
1000	Veri Seti 1	0,995	0,933	0,999	0,997	0,994	0,973
	Veri Seti 2	0,998	0,990	0,998	0,998	1,000	0,993
	Veri Seti 3	0,999	0,998	0,999	0,997	1,000	0,999
	Veri Seti 4	0,999	0,994	0,999	0,991	1,000	0,998
	Veri Seti 5	0,999	0,997	0,999	0,996	1,000	0,999
	Veri Seti 6	0,999	0,999	0,999	0,993	1,000	0,994

Tablo 4.8’e göre, tüm örneklem büyüklüklerinde, tüm modellerden elde edilen yetenek parametreleri arasındaki korelasyon katsayılarının (0,990) – (1,000) arası değiştiği görülmektedir. Bu durum, hangi modelle yetenek tahmini yapılırsa yapılsın neredeyse aynı bulguların elde edileceğini göstermektedir. Yetenek parametrelerine ait hata değerleri arasındaki korelasyon katsayıları incelendiğinde ise, TBMTK-2FM arasındaki korelasyon katsayılarının (0,971) – (0,999); TBMTK-MTTK arasındaki korelasyon katsayılarının ise (0,987) – (0,999) arasında değiştiği belirlenmiştir. 2FM-MTTK arasındaki korelasyon katsayıları incelendiğinde ise bu katsayıların (0,973) – (0,999) arasında değiştiği gözlenmiştir. Yetenek parametrelerinin ve bu parametrelere ait hataların korelasyon katsayılarının yüksek olması, bütün modellerin yetenek parametresi tahmini için birbirlerine benzer sonuçlar verdiği işaret etmektedir. Bu nedenle yetenek parametresine ait kestirimlerin kullanılan

modelden, örneklem büyüklüğünden, bağımsız madde sayısından ve madde takımı sayısından çok fazla etkilenmediği söylenebilir.

4.6. Tartışma

Bu çalışmada farklı örneklem büyüklükleri kullanılarak farklı sayıda madde takımı ve bağımsız madde içeren veri setleriyle çalışılmış, ayrıca üç farklı model ile analizler gerçekleştirilmiştir. Elde edilen bulgular madde ve yetenek parametrelerine ilişkin olmak üzere iki başlık altında tartışılmıştır.

4.6.1. Madde Parametrelerine İlişkin Tartışma

Alan yazında, MTK'nin temel varsayımlarından biri olan yerel bağımsızlık varsayımının ihlal edildiği durumlarda standart MTK modellerinin kullanılmasının, madde parametreleri kestirimlerinde ve madde parametrelerine ait hata hesaplamalarında sorunlar yarattığını gösteren birçok çalışma bulunmaktadır (Ackerman, 1987; Bradlow vd., 1999; Ip, 2000; Marais & Andrich, 2008; Sireci vd., 1991; Wainer, 1995; Wainer & Wang, 2000). Bu çalışmada da bütün koşullarda TBMTK'den elde edilen ayırt edicilik parametresine ait ortalama hata değerinin, diğer iki modele göre daha düşük hesaplandığı belirlenmiştir. Ackerman (1987), Yen (1993), Chang ve Wang (2010) da çalışmalarında yerel bağımsızlık varsayımlarının bozulması durumunda ayırt edicilik parametresinin standart hatasının olduğundan düşük kestirildiğini belirlemişlerdir.

Wainer ve arkadaşları (2000) madde takımı modellerinin madde ayırt edicilik kestirimlerinde daha iyi sonuçlar verdiğini, modelin kestirdiği standart hatanın gerçeğe daha yakın olduğunu ifade etmişlerdir. So (2009) ile Min ve He (2014) de benzer şekilde madde takımı içeren testlerin, tek boyutlu modellerle analiz edildiğinde eğim parametresi kestirimlerinde büyük bir yanlışlık olduğunu belirtmişlerdir. Bu çalışmada da madde bağımlılığını göz önüne alan modellerin, tek boyutlu modele göre ayırt edicilik parametresinin hata kestirimini daha büyük hesapladığı görülmüştür. Bu nedenle alan yazındaki çalışmalara dayanarak yerel bağımlılığın söz konusu olduğu durumlarda, 2FM ya da MTTK kullanılarak eğim parametresine ilişkin daha doğru sonuçlara ulaşılabileceği söylenebilir. Ancak yerel bağımlılığın madde parametresi kestirimleri üzerinde etkisinin olmadığını ya da çok küçük bir etkisi olduğunu belirten çalışmalar da bulunmaktadır (Eckes, 2014; Y. Li

vd., 2010). Bu çalışmada ise kullanılan modelin eğim parametresine ilişkin bulguları etkilediği belirlenmiştir.

Alan yazında maddeler arasındaki yerel bağımlılığı göz önüne alan modellerden ve standart MTK modellerinden elde edilen kesişim parametresi ile ilgili bulgulara rastlanılmamıştır. Min ve He (2014) yaptıkları çalışmada TBMTK, 2FM ve MTTK modellerini madde parametreleri açısından karşılaştırıldığında, model seçimi ile kesişim parametresi kestirimi arasında bir fark olmadığını; kesişim parametresinin madde takımları arasındaki yerel bağımlılıktan etkilenmediğini belirtmişlerdir. Bu çalışmada ise genel olarak TBMTK ile elde edilen kesişim parametresine ait hata ortalamalarının diğer modellerden daha düşük kestirildiği belirlenmiştir. Min ve He'nin (2014) çalışmasına benzer olarak kesişim parametresi kestirimlerinin, eğim parametresi kestirimlerine göre kullanılan modelden daha az etkilendiği bulunmuştur.

Bu çalışmada örneklem büyüklüğü arttıkça madde parametrelerine ait hata değerlerinin de birbirine yaklaşarak, benzer sonuçlar ürettiği gözlenmiştir. O. Zhang (2010) da yaptığı çalışmada iki kategorili ve çok kategorili MTK modelleri ile madde takımı modeli kullanmış; örneklem büyüklüğünün her üç model için de analiz sonuçları üzerinde etkili olduğu, büyük örneklemelerden daha iyi sonuçlar elde edildiği sonucuna varmıştır. Yine Ra (2011), 1000 ve 2000 kişilik simülasyon koşullarında madde takımları ile ilgili yaptığı çalışmada madde parametrelerine ait RMSE değerlerinin 2000 örneklem büyüklüğünde azaldığını belirtmiştir. Bu araştırmada da madde parametrelerine ait kestirimlerin en büyük örneklem olan 1000 örneklem büyüklüğünde daha iyi sonuçlar verdiği belirlenmiştir.

DeMars (2006) madde takımlarıyla ilgili yaptığı çalışmada madde sayısının artmasının eğim parametrelerinin RMSE değerlerini çok değiştirmediğini belirlemiştir. Alan yazında madde sayısının kesişim parametresi üzerinde nasıl bir değişim gösterdiğine dair bir çalışmaya rastlanılmamıştır.

Farklı modellerden elde edilen madde parametreleri ve madde parametrelerinin hataları arasındaki korelasyonların çok düşük olmadığının belirlendiği bu çalışmanın bulguları Bradlow, Wainer ve Wang (1999) tarafından gerçek veri setiyle yapılan çalışmanın bulguları ile benzerlik göstermektedir. Araştırmacılar bağımsız model ve madde takımı modeli altında elde edilen ayırt edicilik kestirimlerinin yakın olduğunu

belirlemiştir. Min ve He (2014) tarafından yürütülen başka bir araştırma da 2PL ile 2FM ve 2FM ile MTTK modellerinden kestirilen kesişim parametreleri arasındaki korelasyon her ikisinde de 0,98 olarak hesap edilmiştir. Yine aynı çalışmada bu modeller arasındaki eğim parametrelerine ait korelasyon değerleri daha düşük bulunmuştur. Bu çalışmada da benzer şekilde kesişim parametreleri korelasyonları tüm modeller arasında yüksek korelasyon vermiş; eğim parametresine ait korelasyon değerleri ise veri setlerine ve kullanılan modele göre daha düşük olarak hesaplanmıştır. Ancak eğim parametresi için de elde edilen korelasyon değerlerinin genel olarak yüksek olduğu belirlenmiştir. Eckes (2014) de çalışmasında madde parametrelerini farklı MTK tabanlı modellerle kestirmiş ve kestirimler arası korelasyonları oldukça yüksek olarak belirlemiştir. Chang ve Wang (2010) da standart MTK modeli ile madde takımı modelini kullanarak kestirdiği ayırt edicilik parametrelerinin korelasyonlarını 0,98 ile 0,99 arasında ve çok yüksek olarak hesaplamıştır.

4.6.2. Yetenek Parametrelerine İlişkin Tartışma

Maddeler arasında yerel bağımlılık söz konusu olduğunda standart MTK modellerinin kullanılmasının, yetenek parametreleri ve bu parametrelere ait hata kestirimlerinde yanlış sonuçlar verdiğini gösteren birçok çalışma bulunmaktadır (Ackerman, 1987; Chang & Wang, 2010; Eckes, 2014; Monseur vd., 2011; Reese, 1995; So, 2010; Tuerlinckx & De Boeck, 2001). Bu çalışmada da altı farklı veri seti ile üç farklı örneklem büyüklüğü ve üç farklı model kullanılmış, yetenek parametreleri ve bu parametrelere ait hata ortalamaları incelenmiştir. Bütün örneklem büyüklükleri ve veri setleri için genel yeteneğe ait hata ortalamalarının 2FM'ye ve MTTK'ye göre TBMTK'de daha düşük değerler verdiği gözlenmiştir. Chang ve Wang (2010) da yaptıkları çalışmada gerçek veri seti üzerinden standart MTK modeli ile MTTK modelinden elde edilen yetenek parametrelerinin standart hatalarını karşılaştırmış; bu çalışmanın sonuçlarına benzer olarak maddeler arasında yerel bağımsızlık varsayımı sağlanmadığında yetenek parametresine ait standart hatanın düşük kestirildiğini ifade etmişlerdir. Yine Eckes (2014) çalışmasında madde takımı etkisinin görmezden gelindiği standart MTK modelleri ile yapılan analizler sonucunda yetenek kestirimine ait standart hatanın olduğundan düşük kestirildiğini ifade etmiştir. Bu araştırmacıların çalışmalarındaki sonuçlara dayanarak, bu çalışmada TBMTK'den elde edilen yetenek parametresi hata ortalamalarının diğer

iki modelden daha düşük deęer vermesi, maddeler arasında yerel baęımlılık olmasına baęlanabilir. Bu nedenle alıřmanın yetenek parametresinin hatalarına iliřkin bulgularının alan yazınla paralellik gsterdięi sylenebilir. Ancak Chang ve Wang (2010) yerel baęımlılıęın grmezden gelinerek standart MTK modellerinin kullanılmasıyla elde edilen yetenek parametresinin fazla kestirildięi sonucuna varsa da, bu alıřmada yetenek parametresine iliřkin bulgular her  modelde de olduka yakın sonular vermiřtir.

DeMars (2006) yaptıęı simlasyon alıřmasında 25 ve 50 maddelik iki veri seti iin; iki faktr modeli, madde takımı modeli, baęımsız madde modeli ve ok kategorili madde modelini kullanarak birincil yetenek puanlarının RMSE deęerleri ortalamasını hesaplamıř ve btn modeller iin test uzunluęu artınca RMSE ortalamalarının dřtęn belirlemiřtir. Benzer řekilde arařtırmacı, test uzunluęu arttıķa yeteneęe ait gerek puan ile kestirilen puan arasındaki korelasyonların da test uzunluęu arttıķa arttıęını gzlemiřtir. Bu alıřmada ise yalnızca TBMTK iin madde sayısı arttıķa hataların ortalaması azalmıř; dięer modeller iin ise belli bir eęilim belirlenmemiřtir. rneklem byklę aısından yetenek parametresine ait ortalama hatalar inceledięi zaman da belli bir bulguya varılmamıřtır. Oysa Ra (2011) byk rneklem kořulunda (2000) yetenek parametresine ait hataların daha dřk olduęunu belirlemiřtir. alıřmanın madde sayısı ve rneklem byklę ile ilgili bu bulguları alan yazından farklılık gstermiřtir.

Genel yeteneęe ait farklı modellerden elde edilen kestirimlere ait korelasyonların yksek ıktıęı bu alıřmanın bulguları alan yazındaki alıřmalarla rtřmektedir. DeMars (2006) PISA sonularıyla yaptıęı alıřmada iki faktr modeli, madde takımı modeli, baęımsız madde modeli ve ok kategorili madde modeli ile yetenek kestirimleri yapmıř ve farklı modellerden elde edilen yetenek kestirimleri arasındaki korelasyonları 1,00'e yakın olarak hesaplamıřtır. Eckes (2014), Min ve He (2014) ile Bradlow ve arkadaşları (1999) yaptıkları alıřmalarda yetenek parametrelerinin benzer olduęunu ifade etmiřlerdir. Ancak bu arařtırmada yetenek parametrelerine ait hataların korelasyonu iin benzer bir durum bulunamamıřtır. Farklı modellerden elde edilen hataların farklılık gsterdięi Eckes (2014) ile Min ve He (2014) tarafından da ifade edilmiřtir. Ancak Eckes (2014) 2PL ile MTTK arasındaki hata kestirimlerini, Min ve He (2014) ise 2FM ile MTTK arasındaki hata kestirimlerini daha benzer olarak bulmuřlardır. DeMars (2006) ise alıřmasında farklı modellerden elde edilen

yetenek parametrelerine ait korelasyonların oldukça yüksek olduğunu belirlemiştir. Bu çalışmanın yetenek parametresine ait bulguları, DeMars'ın (2006) çalışmasına benzerlik göstermektedir.



5. SONUÇ VE ÖNERİLER

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

5.1. Sonuçlar

MTK güçlü bir matematiksel model olduğu için günümüzde eğitimsel testlerde oldukça sık kullanılmaktadır. Bu kuramın kullanılabilmesi için veri setinin karşılaması gereken varsayımları bulunmaktadır ve en önemli varsayımlarından biri de yerel bağımsızlıktır. Ancak ortak bir uyarıcıya dayanan madde takımlarının testlerde kullanılması sonucunda, yerel bağımsızlık varsayımı bozulmaktadır. İlgili araştırmalar bölümünde de belirtildiği üzere alan yazında yapılan birçok çalışma yerel bağımsızlık varsayımının bozulmasının standart MTK modelleriyle elde edilen madde ve yetenek parametrelerinde yanlış kestirime neden olduğunu göstermiştir. Bu araştırmada, bu nedenle madde takımları üzerinde durulmuş ve PISA 2012'deki matematik okuryazarlığının ölçüldüğü sorulardan oluşan onuncu kitapçıkta yer alan maddelerden farklı veri setleri oluşturularak, farklı MTK tabanlı modeller ile bu veri setleri analiz edilmiştir. Araştırmanın amacını kısaca belirtmek gerekirse altı adet farklı veri setinin TBMTK, 2FM ve MTTK modelleriyle analiz edilmesi sonucunda elde edilen madde ve yetenek parametrelerini ve bu parametrelere ait hataları karşılaştırmaktır. Ayrıca her bir veri seti için model-veri uyumu değerleri incelenmiş; daha sonra bu üç modelden elde edilen madde ve yetenek parametreleri ile bu parametrelere ait hatalar arasındaki korelasyon değerleri hesaplanmıştır. Çalışmanın bulguları ışığında, bu çalışmadan elde edilen sonuçlar şöyledir.

- Farklı örneklem büyüklüklerinde, farklı modellerden ve farklı veri setlerinden elde edilen madde parametreleri için genel olarak en düşük ortalama hata değerlerinin TBMTK'den elde edildiği; MTTK'nin, 2FM'den daha düşük ortalama hata kestirimi yaptığı belirlenmiştir.
- Özellikle 250 örneklem büyüklüğünde farklı modellerden hesaplanan madde parametrelerine ait bulguların farklılaştığı, ancak örneklem büyüklüğü arttıkça farklı modellerden elde edilen madde parametre hata değerleri ortalamasının birbirine yaklaştığı sonucuna varılmıştır.

- Veri setindeki bağımsız madde sayısının artmasıyla parametrelere ait ortalama hata değerlerinde daha çok azalma eğilimi olduğu görülmüştür. Veri setine madde takımı eklenmesiyle genel olarak parametrelere ait hata ortalamaları artarken, en az değişkenlik gösteren parametrenin kesişim (c) parametresi olduğu belirlenmiştir. Ancak veri setindeki madde takımı sayısının artmasının, madde parametrelerine ait hata ortalamalarına etkisi üzerine çok net bir sonuca erişilmemiştir.
- Farklı model kullanımıyla elde edilen kesişim parametresine ait bulguların, eğim parametresine ait bulgulara göre daha az değişim gösterdiği; dolayısıyla kesişim parametresinin kullanılan modelden daha az etkilendiği belirlenmiştir.
- 250, 500 ve 1000 örneklem büyüklüğünde her üç modelden elde edilen birincil yetenek parametresine ait ortalamalarının hemen hemen aynı olduğu ve bu parametrelere ait hata ortalamaları için en düşük değeri her koşulda TBMTK'nın verdiği belirlenmiştir.
- TBMTK'den elde edilen yetenek parametrelerine ait hata ortalamaları veri setindeki bağımsız madde sayısının artmasıyla birlikte genel olarak azalırken; 2FM ve MTTK için genel bir sonuç elde edilememiştir.
- 2FM ve MTTK modellerinden kestirilen yetenek ortalamaları ve yetenek ortalamalarına ait hata ortalamaları benzer olmakla birlikte, MTTK'den elde edilen hata ortalamalarının daha düşük olduğu belirlenmiştir.
- Örneklem büyüklüğü arttıkça farklı modellerden elde edilen yetenek parametresinin ve hata ortalamasının çok değişmediği, bireyin yetenek kestirimlerinin örneklem büyüklüğüne pek bağlı olmadığı sonucuna varılmıştır.
- Bütün koşullarda yerel bağımlılığı göz önüne alan modellerin TBMTK'den daha iyi model-veri uyumu sağladığı, genel olarak 2FM ile MTTK arasında anlamlı farklılık olmadığı; her iki modelin de bu veri setleri için kullanılabileceği belirlenmiştir. Bu iki model arasında anlamlı farklılık olduğu durumda ise genel olarak 2FM'nin daha iyi sonuç verdiği söylenebilir.

- Her bir örneklem büyüklüğünde eğim ve kesişim parametreleri hatalarının korelasyonlarına ait ortalama değerler incelendiğinde, 2FM-MTTK arasındaki korelasyon katsayılarının tüm örneklem büyüklüklerinde yüksek düzeyde olduğu belirlenmiştir. Bu durum, bu iki modelin madde parametrelerine ait hata değerleri için yüksek düzeyde benzer kestirim yaptığını, bu nedenle de madde parametrelerine ait tahminlemelerde birbirleri yerine kullanılabileceğini göstermektedir.
- TBMTK-2FM'den ve TBMTK-MTTK'den elde edilen a parametrelerine ait hataların korelasyon ortalamaları incelendiğinde, genel olarak örneklem büyüklüğündeki artışla birlikte korelasyon katsayısının da yüksek düzeyde olduğu söylenebilir. Kesişim parametresi için ise bütün örneklem büyüklüklerinde, bütün modeller için yüksek ilişkiler hesaplanmıştır.
- Farklı veri setlerinden ve farklı örneklem büyüklüklerinden hesaplanan eğim ve kesişim parametresi için eğim parametresinin modele ve örneklem büyüklüğüne daha duyarlı olduğu, kesişim parametresinin ise her durumda daha benzer sonuçlar verdiği ifade edilebilir.
- Farklı modellerden hesaplanan korelasyon değerlerinin veri setinde yer alan madde takım sayısı ya da bağımsız madde sayısı ile ilişkisi hakkında herhangi bir yorum yapılamamıştır.
- Araştırmanın bütün koşullarında, yetenek parametresi korelasyonları ve yetenek parametresine ait standart hata korelasyonları incelendiğinde her üç modelinde birbirleriyle yüksek korelasyon gösterdiği belirlenmiştir. Bu nedenle yetenek parametresine ait değerler için her üç modelin de birbiriyle ilişkili sonuçlar verdiği ifade edilebilir.
- Farklı modellerden elde edilen madde ve yetenek parametresine ilişkin sonuçların benzer olması, veri setlerinde aynı madde takımında yer alan madde çiftleri arasındaki yerel bağımlılıkların çok büyük olmamasından kaynaklanıyor olabilir.

5.2. Öneriler

Araştırma sonucuna ve gelecek araştırmalara dönük önerilere iki başlık halinde aşağıda yer verilmiştir.

5.2.1. Arařtırma Sonucuna Dönük Öneriler

1. Tüm testler için, çeřitli yerel bağımlılık indeksleri kullanılarak maddelerin yerel bağımsızlığı kontrol edilmelidir. Eğer yerel bağımlı maddeler varsa standart MTK modelleri yerine, maddeler arasındaki yerel bağımlılık durumunu göz önüne alan modellerin kullanılması önerilmektedir.
2. MTK tabanlı modeller için daha büyük örneklem büyüklükleri gerekmektedir. Bu çalışmada da genel olarak örneklem büyüklüğündeki artışla beraber daha az hata içeren sonuçlar elde edilmiştir. Bu nedenle arařtırmacılar daha az hatalı madde parametre değerleri elde edebilmek için daha büyük örneklemlemlerle çalışmalıdırlar.
3. Arařtırmacıların yaptıkları çalışmada eğer eğim parametresi kestirimleri önemli ise yerel bağımlılık durumunu göz önüne alan modeller kullanması önerilmektedir. Ancak yalnızca kesişim parametresi ile ilgilenilen çalışmalarda tek boyutlu MTK modelleri de kullanılabilir.
4. Bu çalışmada madde sayısının yüksek olduğu veri setlerinde genel olarak daha az hatalı sonuçlar elde edildiği için testlerin madde sayılarının artırılması önerilebilir.
5. Yalnızca yetenek parametresi kestirimleriyle ilgilenilen durumlarda, kullanılan model kestirimlerinin birbiriyle ilişkili sonuçlar vermesi nedeniyle bu modellerden herhangi biri kullanılabilir.
6. 2FM ile MTTK kestirimleri yüksek korelasyon verdiği için bu iki modelden herhangi biri arařtırmacılar tarafından tercih edilebilir.

5.2.2. Gelecek Arařtırmalara Dönük Öneriler

1. Bu arařtırmada 250, 500 ve 1000 örneklem büyüklükleri ile çalışılmıştır. Örneklem büyüklüğü daha azaltılarak ya da artırılarak benzer bir çalışma yapılabilir.
2. Bu çalışmada kullanılan madde takımları arasında çok aşırı düzeyde yerel bağımlılık bulunmamıştır. Bu nedenle daha yüksek yerel bağımlı maddeler kullanıldığında standart MTK modelleri ile yerel bağımlılığı göz önüne alan modeller arasındaki kestirimlerin nasıl değiştiğine dair bir çalışma gerçekleştirilebilir.

3. Madde takımları arasında yerel bağımlılık bulunması durumunda, bu yerel bağımlılığın nedenlerini araştıracak bir çalışma yapılabilir.
4. Araştırmada madde takımları arasındaki yerel bağımlılık incelemesi LD χ^2 ile yapılmıştır. Gelecek araştırmalarda başka indeksler kullanılarak yerel bağımlılık kontrol edilebilir.
5. Bu araştırma gerçek veri seti üzerinden yürütülmüş ve var olan duruma herhangi bir müdahalede bulunulamamıştır. Bu nedenle benzer bir çalışma simülasyon koşulları altında, istenilen durum yaratılarak tekrarlanabilir.
6. Bu çalışmada tüm modellerde iki kategorili puanlanan maddeler için 2PLM kullanılmıştır. İleri yapılacak çalışmalarda 3PLM ile analizler gerçekleştirilebilir.
7. Bu araştırmada parametre kestirimleri için IRTPRO 2.1 yazılımı kullanılmıştır. Başka yazılımlar kullanılarak araştırma tekrar edilebilir ya da başka yazılımlardan elde edilen sonuçlar karşılaştırılabilir.

KAYNAKÇA

- ACER (Australian Council for Educational Research) (2013). *PISA 2012: How Australia measures up*. [Çevrim-içi: <http://www.acer.edu.au/documents/PISA-2012-Report.pdf>, Erişim tarihi: 12 Nisan 2014.]
- Ackerman, T. A. (1987). *The robustness of LOGIST and BILOG IRT estimation programs to violations of local independence*. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.
- Andrich, D. (1978a). Application of a psychometric model to ordered categories which are scored with successive integers. *Applied Psychological Measurement*, 2, 581–594.
- Andrich, D. (1978b). A rating formulation for ordered response categories. *Psychometrika*, 43, 561–573.
- Arora, A. (2007). *Creating a TIMSS 2003 problem-solving scale and examining the problem-solving achievement of United States eight-grade students in TIMSS 2003*. Unpublished Doctoral Dissertation. Boston Collage.
- Bao, H. (2007). *Investigating differential item function amplification and cancellation in application of item response testlet models*. Unpublished Doctoral Dissertation. University of Maryland.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick (Eds.), *Statistical theories of mental test scores* (pp. 397–472). Reading, MA: Addison- Wesley.
- Bishop, N. S., & Omar, M. H. (2002). *Comparing vertical scales derived from dichotomous and polytomous IRT models for a test composed of testlets*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29-51.
- Bock, R. D., & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: An application of the EM algorithm. *Psychometrika*, 46(4), 443–459.
- Boyd, A. M. (2003). *Strategies for controlling testlet exposure rates in computerized adaptive testing systems*. Unpublished Doctoral Dissertation. The University of Texas.
- Bradlow, E. T., Wainer, H., & Wang, X. (1999). A Bayesian random effects model for testlets. *Psychometrika*, 64, 153–168.
- Brennan, R. I. (2001). An essay on the history and future of reliability from the perspective of replications. *Journal of Educational Measurement*, 38(4), 295-317.
- Cai, L., du Toit, S. H. C., & Thissen, D. (2011). *IRTPRO: User guide*. Lincolnwood, IL: Scientific Software International.
- Cai, L., Yang, J. S., & Hansen, M. (2011). Generalized full-information item bifactor analysis. *Psychol Methods*, 16(3), 221–248.

- Chang, Y., & Wang, J. (2010). *Examining testlet effects on the PIRLS 2006 assessment*. Paper presented at the 4th IEA International Research Conference, Gothenburg, Sweden.
- Chen, C. T., & Wang W. C. (2007). Effects of ignoring item interaction on item parameter estimation and detection of interacting items. *Applied Psychological Measurement, 31*(5), 388–411.
- Chen, F. F., Hayes, A., Carver, C. S., Laurenceau, J. P., & Zhang, Z. (2012). Modeling general and specific variance in multifaceted constructs: a comparison of the bifactor model to other approaches. *Journal of Personality, 80*(1), 219-251.
- Chen, T. A., (2010). *Random or fixed testlet effects: a comparison of two multilevel testlet models*. Unpublished Doctoral Dissertation. The University of Texas.
- Chen, W. H., & Thissen, D. (1997). Local dependence indexes for item pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*(3), 265–289.
- Cheng, S. C. (1996). *A comparison of testlet and individual item scoring*. Unpublished Doctoral Dissertation. The University of Texas.
- Chien, Y. M. (2008). *An investigation of testlet-based item response models with a random facets design in generalizability theory*. Unpublished Doctoral Dissertation. The University of Iowa.
- Cook, K. F., Dodd, B. G., & Fitzpatrick, S. J. (1999). A comparison of three polytomous item response theory models in the context of testlet scoring. *Journal of Outcome Measurement, 3*(1), 1–20.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Fort Worth: Holt, Rinehart and Winston.
- Cureton, E. E. (1965). Reliability and validity: Basic assumptions and experimental design. *Educational and Psychological Measurement, 25*, 326–346.
- Çakıcı-Eser, D., & Gelbal, S. (2015). Farklı boyutluluk özelliklerindeki basit ve karmaşık yapıları testlerin çok boyutlu madde tepki kuramına dayalı parametre kestirimlerinin incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 6*(2), 331-350.
- DeMars, C. E. (2006). Application of the bi-factor multidimensional item response theory model to testlet-based tests. *Journal of Educational Measurement, 43*, 145–168.
- DeMars, C. E. (2012). Confirming testlet effects. *Applied Psychological Measurement, 36*, 104–121.
- Desa, Z. N. D. M. (2012). *Bi-factor Multidimensional Item Response Theory Modeling for Subscores Estimation, Reliability, and Classification*. Unpublished Doctoral Dissertation. University of Kansas.
- Dresher, A. R. (2002). *The examination of local item dependency of NAEP assessments using the testlet model*. Unpublished Doctoral Dissertation. University of Pittsburgh.
- Eckes, T. (2013). Examining testlet effects in the TestDaF listening section: A testlet response. *Language Testing, 31*(1), 39–61.

- Embretson, S., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- Fleischman, H. L., Hopstock, P. J., Pelczar, M. P., & Shelley, B. E. (2010). *Highlights From PISA 2009: Performance of U.S. 15-Year-Old Students in Reading, Mathematics, and Science Literacy in an International Context* (NCES 2011-004). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.
- Fukuhara, H., & Kamata, A. (2011). Functioning analysis on testlet-based items a bifactor multidimensional item response theory model for differential items. *Applied Psychological Measurement, 35*(8), 604–622.
- Gibbons, R. D., Bock, R. D., Hedeker, D., Weiss, D. J., Segawa, E., Bhaumik, D. K., Kupfer, D. J., Frank, E., Grochocinski, V. J., & Stover, A. (2007). Full-Information item bifactor analysis of graded response data. *Applied Psychological Measurement, 31*(1), 4–19.
- Gibbons, R. D., & Hedeker, D. R. (1992). Full-information bi-factor analysis. *Psychometrika, 57*, 423–436.
- Gibbons, R. D., Hedeker, D. R., & Bock, R. D. (1990). *Full-Information Item Bi-Factor Analysis: ONR Technical report*. Technical report, Biometric Laboratory, Illinois State Psychiatric Institute, Chicago.
- Glas, C. A. W., Wainer, H., & Bradlow, E. T. (2000). MML and EAP estimation in testlet-based adaptive testing. In W. Jvan der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (p. 271-287). Boston, MA: Kluwer-Nijhoff.
- Habing, B., & Roussos, L. A. (2003). On the need for negative local item dependence. *Psychometrika, 68*, 435–451.
- Haladyna, T. M. (1992). Context-dependent item sets. *Educational Measurement: Issues and Practice, 11*(1), 21–25.
- Haladyna, T. M. (1997). *Writing test items to evaluate higher order thinking*. Boston, MA: Allyn and Bacon.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications, Inc.
- Hoskens, M., & De Boeck, P. (1997). A parametric model for local dependence among test items. *Psychological Methods, 2*(3), 261–277.
- Houts, C. R., & Cai, L. (2013). *flexMIRT® user's manual version 2: Flexible multilevel multidimensional item analysis and test scoring*. Chapel Hill, NC: Vector Psychometric Group.
- Ip, E. H. (2000). Adjusting for information inflation due to local dependency in moderately large item clusters. *Psychometrika, 65*, 73–91.
- Ip, E. H. (2010). Empirically indistinguishable multidimensional IRT and locally dependent unidimensional item response models. *British Journal of Mathematical and Statistical Psychology, 63*, 395–416.

- Jiao, H., Wang, S., & He, W. (2008). *Comparison of estimation methods of one-parameter testlet model*. Paper presented at the annual meeting of the American Educational Research Association, New York.
- Jiao, H., Wang, S., & Kamata, A. (2005). Modeling local item dependence with the hierarchical generalized linear model. *Journal of Applied Measurement*, 6, 311-321.
- Keller, L. A., Swaminathan, H., & Sireci, S. G. (2003). Evaluating scoring procedures for context-dependent item sets. *Applied Measurement in Education*, 16, 207–222.
- Kim, D., De Ayala, R. J., Ferdous, A. A., & Nering, M. L. (2007). *Assessing relative performance of local item dependence (LID) indexes*. Paper presented at the annual meeting of the National Council on Measurement in Education, Chicago, IL.
- Koğar, H. (2015). Madde tepki kuramına ait parametrelerin ve model uyumlarının karşılaştırılması: Bir Monte Carlo çalışması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6(1), 142-157.
- LaFond, L. J. (2014). *Decision consistency and accuracy indices for the bifactor and testlet response theory models*. Unpublished Doctoral Dissertation. The University of Iowa.
- Lee, Y. W. (2004). Examining passage-related local item dependence (LID) and measurement construct using Q_3 statistics in an EFL reading comprehension test. *Language Testing*, 21(1), 74–100.
- Lee, G., Dunbar, S. B., & Frisbie, D. A. (2001). The relative appropriateness of eight measurement models for analyzing scores from tests composed of testlets. *Educational and Psychological Measurement*, 61, 958–975.
- Lee, G., & Frisbie, D. A. (1999). Estimating reliability under a generalizability theory model for test scores composed of testlets. *Applied Measurement in Education*, 12(3), 237–255.
- Lee, G., Kolen, M. J., Frisbie, D. A., & Ankenmann, R. D. (2001). Comparison of dichotomous and polytomous item response models in equating scores from tests composed of testlets. *Applied Psychological Measurement*, 25, 357–372.
- Li, F. (2009). *An information correction method for testlet-based test analysis: From the perspectives of item response theory and generalizability theory*. Unpublished Doctoral Dissertation. The University of Maryland.
- Li, Y. (2004). *Applications and extensions of IRT testlet models*. Unpublished Doctoral Dissertation. University Of Wisconsin-Madison.
- Li, Y., Bolt, D. M., & Fu, J. (2005). A test characteristic curve linking method for the testlet model. *Applied Psychological Measurement*, 29(5), 340–356.
- Li, Y., Bolt, D. M., & Fu, J. (2006). A comparison of alternative models for testlets. *Applied Psychological Measurement*, 30(1), 3–21.
- Li, Y., Li, S., & Wang, L. (2010). *Application of a general polytomous testlet model to the reading section of a large-scale English language assessment (ETS RR-10–21)*. Princeton, NJ: Educational Testing Service.

- Liu, Y., & Thissen, D. (2012). Identifying local dependence with a score test statistic based on the bifactor logistic model. *Applied Psychological Measurement, 36*(8), 670–688.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, N. J.: Lawrence Erlbaum Associates.
- Lu, R. (2010). *Impacts of local item dependence of testlet items with the multistage tests for pass-fail decisions*. Unpublished Doctoral Dissertation. The University of Maryland.
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*, 149–173.
- Marais, I. D., & Andrich, D. (2008). Effects of varying magnitude and patterns of local dependence in the unidimensional Rasch model. *Journal of Applied Measurement, 9*, 105–124.
- McLeod, L. D., Swygert, K. A., & Thissen, D. (2001). Factor analysis for items scored in two categories. In D. Thissen & H. Wainer (Eds.), *Test scoring* (pp. 189-216). Mahwah, NJ: Lawrence Erlbaum Associates.
- MEB (Mili Eğitim Bakanlığı) (2011). *PISA Türkiye*. [Çevrim-içi: <http://pisa.meb.gov.tr/wp-content/uploads/2013/07/PISA-kitab%C4%B1.pdf>, Erişim tarihi: 22 Nisan 2014.]
- MEB (Mili Eğitim Bakanlığı) (2013). *PISA 2012 ulusal ön raporu*. [Çevrim-içi: http://yegitek.meb.gov.tr/meb_iys_dosyalar/2013_12/13053601_pisa2012_ulusal_n_raporu.pdf, Erişim tarihi: 21 Nisan 2014.]
- Min, S., & He, L. (2014). Applying unidimensional and multidimensional item response theory models in testlet-based reading assessment. *Language Testing, 31*(4) 453–477.
- Monseur, C., Baye, A., Lafontaine, D., & Quittre, V. (2011). PISA test format assessment and the local independence assumption. *IERI Monograph Series: Issues and Methodologies in Large-Scale Assessments, 4*, 131–155.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement, 16*, 159–176.
- Nofer, D. C. (2007). *Conditional item dependence for testlet items*. Unpublished Doctoral Dissertation. The University of Maryland.
- OECD (Organisation for Economic Cooperation and Development) (2009). *PISA 2006 technical report*. [Çevrim-içi: <http://www.oecd.org/pisa/pisaproducts/42025182.pdf>, Erişim tarihi: 17 Nisan 2014.]
- OECD (Organisation for Economic Cooperation and Development) (2013). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy* [Çevrim-içi: http://www.oecd.org/pisa/pisaproducts/PISA%202012%20framework%20e-book_final.pdf, Erişim tarihi: 17 Nisan 2014.]
- Paap, M. C. S., & Veldkamp, B. P. (2012). Minimizing the testlet effect: Identifying critical testlet features by means of tree-based regression. In T. J. H. M. Eggen & B. P.

- Veldkamp (Eds.), *Psychometrics in Practice at RCEC*. Enschede, the Netherlands: RCEC, Cito/University of Twente.
- Park, C. (1997). *Accuracy of parameter estimation in polytomous IRT models*. Unpublished Doctoral Dissertation. Fordham University.
- Pomplun, M. R. (1988). *Effects of local dependence in achievement tests on IRT ability estimation*. Unpublished Doctoral Dissertation. The Florida State University.
- Ra, J. (2011). *Sensitivity of prior specification within testlet model*. Unpublished Doctoral Dissertation. The University of Georgia.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.
- Reese, L. M. (1995). *The impact of local item dependence on some LSAT outcomes*. (Report No. LSAC-R-95-02). Newtown, PA: Law School Admission Council, Inc.
- Rijmen, F. (2009). *Three multidimensional models for testlet based tests: Formal relations and an empirical comparison* (ETS Research Rep. No. RR-09-37). Princeton, NJ: ETS.
- Rijmen, F. (2010). Formal relations and an empirical comparison among the bi-factor, the testlet, and a second-order multidimensional IRT model. *Journal of Educational Measurement*, 47(3), 361–372.
- Rosenbaum, P. R. (1988). Item bundles. *Psychometrika*, 53, 349–359.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, 17.
- Sedivy, S. K. (2009). *Using traditional methods to detect differential item functioning in testlet data*. Doctoral dissertation, University of Wisconsin-Milwaukee.
- Sireci, S. G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement*, 28, 237–247.
- So, Y. (2010). *Dimensionality of responses to a reading comprehension assessment and its implications to scoring test takers on their reading proficiency*. Unpublished Doctoral Dissertation. University of California.
- Sünbül, Ö., & Erkuş, A. (2013). Madde parametrelerinin değişmezliğinin çeşitli boyutluluk özelliği gösteren yapılarda madde tepki kuramına göre incelenmesi. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 9(2), 378-398.
- Tang, K. L. (1996). *Polytomous item response theory (IRT) models and their applications in large-scale testing problems: Review of the literature*. Princeton, New Jersey: Educational Testing Service.
- Tao, W. (2008). *Using the score-based testlet method to handle local item dependence*. Unpublished Doctoral Dissertation. Boston Collage.
- Thissen, D., Cai, L., & Bock, R. D. (2010). The nominal categories item response model. In: M. Nering & R. Ostini (Eds.), *Handbook of polytomous item response theory models: Developments and applications*. New York, NY: Taylor & Francis.

- Thissen, D., Nelson, L., Rosa, K., & McLeod, L. D. (2001). Item response theory for items scored in more than two categories. In D. Thissen & H. Wainer (Eds.), *Test scoring* (p. 141–184). Mahwah, NJ: Lawrence Erlbaum Associates.
- Thissen, D., Steinberg, L., & Mooney, J. A. (1989). Trace lines for testlets: A use of multiple-categorical-response models. *Journal of Educational Measurement, 26*(3), 247–260.
- Thissen, D., & Orlando, M. (2001). Item response theory for items scored in two categories. In D. Thissen & H. Wainer (Eds.), *Test scoring* (pp. 73 - 140). Mahwah, NJ: Lawrence Erlbaum Associates.
- Tuerlinckx, F., & De Boeck, P. (2001). The effect of ignoring item interactions on the estimated discrimination parameters in item response theory. *Psychological Methods, 6*(2), 181–195.
- van der Linden, W. J., & Hambleton, R. K. (1997). *Handbook of modern item response theory*. New York: Springer.
- Wainer, H. (1995). Precision and differential item functioning on a testlet-based test: The 1991 Law School Admissions Test as an example. *Applied Measurement in Education, 8*, 157–186.
- Wainer, H., Bradlow, E. T., & Du, Z. (2000). Testlet response theory: An analog for the 3PL model useful in testlet-based adaptive testing. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice* (pp. 245–269). Dordrecht, Netherlands: Kluwer.
- Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its applications*. Cambridge: Cambridge University Press.
- Wainer, H., Kaplan, B., & Lewis, C. (1992). A comparison of the performance of simulated hierarchical and linear testlets. *Journal of Educational Measurement, 29*, 243–251.
- Wainer, H., & Kiely, G. L. (1987). Item clusters and computerized adaptive testing: A case for testlets. *Journal of Educational Measurement, 24*, 185–201.
- Wainer, H., & Lewis, C. (1990). Toward a psychometrics for testlets. *Journal of Educational Measurement, 27*(1), 1–14.
- Wainer, H., & Thissen, D. (1996). How is reliability related to the quality of tests cores? What is the effect of local dependence on reliability? *Educational Measurement: Issues and Practice, 15*(1), 22–29.
- Wainer, H., & Wang, X. (2000). Using a new statistical model for testlets to score TOEFL. *Journal of Educational Measurement, 37*(3), 203–220.
- Wang, X., Bradlow, E. T., & Wainer, H. (2002). A general Bayesian model for testlets: Theory and applications. *Applied Psychological Measurement, 26*, 190–128.
- Wang, W. C., & Wilson, M. (2005a). The Rasch testlet model. *Applied Psychological Measurement, 29*(2), 126–149.
- Wang, W. C., & Wilson, M. (2005b). Assessment of differential item functioning in testlet-based items using the Rasch testlet model. *Educational and Psychological Measurement, 65*, 549–576.

- Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, 8(2), 125–145.
- Yen, W. M. (1993). Scaling performance assessments Strategies for managing local item dependence. *Journal of Educational Measurement*, 30, 187–213.
- Yen, W. M., & Fitzpatrick, A. R. (2006). Item response theory. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 111-153). Westport, CT: American Council on Education/Praeger.
- Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (2002). Identification and evaluation of local item dependencies in the Medical College Admissions Test. *Journal of Educational Measurement*, 39(4), 291–309.
- Zhang, B. (2010). Assessing the accuracy and consistency of language proficiency classification under competing measurement models. *Language Testing*, 27(1), 119–140.
- Zhang, O. (2010). *Polytomous IRT or testlet model: An evaluation of scoring models in small testlet size situations*. Unpublished Master Dissertation. University of Florida.
- Zheng, C. (2013). *Examination of the parameter estimate bias when violating the orthogonality assumption of the bifactor model*. Unpublished Doctoral Dissertation. Columbia University.
- Zimowski, M. F., Muraki, E., Mislevy, R. J., & Bock, R. D. (2003). *BILOG-MG 3* [Computer Program]. Chicago: Scientific Software Corporation.
- Zuru, D. (1998). *Modeling conditional item dependencies with a three-parameter logistic testlet model*. Unpublished Doctoral Dissertation. University of Kansas.



EKLER DİZİNİ

EK 1. ETİK KURUL İZİN MUAFİYET FORMU

Form: 40

Tez Çalışması Etik Kurul İzin Muafiyeti Formu

27 / 01 / 2016

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

Tez Başlığı / Konusu:	Madde Takımları İçeren Testlerde Farklı Modellerden Elde Edilen Madde ve Yetenek Parametrelerinin 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 Kurullar ve Komisyonlarının Yönergelerini inceledim ve bunlara göre tez çalışmamın yürütülebilmesi için herhangi bir Etik 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.


Esin YILMAZ KOĞAR
(Öğrencinin Adı Soyadı, İmzası)

Öğrenci Bilgileri

Adı Soyadı	Esin YILMAZ KOĞAR
Öğrenci No	N11143700
Anabilim Dalı	Eğitim Bilimleri
Programı	Eğitimde Ölçme ve Değerlendirme
Statüsü	<input type="checkbox"/> Yüksek Lisans <input checked="" type="checkbox"/> Doktora <input type="checkbox"/> Bütünleşik Dr.

Danışman Görüşü ve Onayı

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



EK 2. İki Kategorili Puanlanan Maddeler için Farklı Modellerden Elde Edilen Madde Parametreleri ve Madde Parametrelerine ait Standart Hata Değerleri

Tablo 4.9: Veri Seti 1'de Yer Alan İki Kategorili Puanlanan Maddeler için Farklı Modellerden Elde Edilen Madde Parametreleri ve Madde Parametrelerine ait Standart Hata Değerleri

N	Madde*	TBMTK				2FM				MTTK							
		a	SH	c	SH	a ₁	SH	a ₂	SH	c	SH	a ₁	SH	a ₂	SH	c	SH
250	MT11	1,94	0,33	0,45	0,22	2,13	0,77	1,77	1,23	0,57	0,32	2,16	0,45	2,16	0,45	0,55	0,22
	MT14	1,37	0,24	-0,41	0,19	1,23	0,25	0,53	0,33	-0,41	0,18	1,30	0,22	1,30	0,22	-0,42	0,16
	MT21	1,71	0,29	-0,64	0,22	1,98	0,39	0,93	0,36	-0,74	0,25	2,05	0,34	2,05	0,34	-0,76	0,20
	MT22	0,93	0,19	-0,60	0,16	0,97	0,23	0,93	0,36	-0,68	0,20	1,00	0,23	1,00	0,23	-0,70	0,17
	MT31	1,70	0,30	0,77	0,21	2,02	0,67	0,67	1,28	0,89	0,34	1,96	0,43	1,96	0,43	0,83	0,20
	MT32	2,91	0,81	-4,19	0,94	3,07	0,96	0,67	1,28	-4,55	1,23	3,41	1,02	3,41	1,02	-4,78	1,08
	MT41	1,72	0,32	-1,24	0,26	2,12	0,80	0,43	1,16	-1,60	0,60	1,99	0,43	1,99	0,43	-1,52	0,34
	MT42	1,26	0,23	0,44	0,17	1,58	0,78	1,54	1,52	0,58	0,36	1,78	0,75	1,78	0,75	0,64	0,35
	MT43	0,98	0,21	-0,91	0,18	0,99	0,23	0,44	0,29	-0,94	0,19	1,01	0,21	1,01	0,21	-0,96	0,17
500	MT11	2,65	0,33	0,83	0,19	2,64	0,33	0,92	0,40	0,87	0,20	2,68	0,39	2,68	0,39	0,87	0,16
	MT14	1,31	0,16	-0,02	0,12	1,31	0,20	0,92	0,42	-0,02	0,13	1,35	0,19	1,35	0,19	-0,02	0,12
	MT21	1,89	0,23	-0,34	0,15	2,09	0,28	0,18	0,87	-0,37	0,16	1,97	0,23	1,97	0,23	-0,34	0,13
	MT22	1,13	0,15	-0,11	0,11	1,16	0,16	0,18	0,87	-0,12	0,12	1,16	0,16	1,16	0,16	-0,11	0,11
	MT31	1,31	0,17	0,95	0,13	1,71	0,40	1,35	0,72	1,24	0,31	1,48	0,19	1,48	0,19	1,08	0,14
	MT32	2,20	0,43	-3,84	0,46	2,66	0,77	1,35	0,72	-4,75	1,15	2,42	0,44	2,42	0,44	-4,26	0,49
	MT41	1,51	0,20	-1,13	0,16	1,60	0,23	0,73	0,31	-1,23	0,18	1,58	0,21	1,58	0,21	-1,21	0,16
	MT42	1,36	0,17	0,52	0,12	1,63	0,32	1,17	0,58	0,64	0,19	1,62	0,25	1,62	0,25	0,64	0,15
	MT43	1,40	0,19	-1,12	0,15	1,59	0,29	1,02	0,46	-1,30	0,23	1,59	0,19	1,59	0,19	-1,31	0,10
1000	MT11	2,27	0,20	0,62	0,12	2,27	0,21	0,75	0,33	0,66	0,12	2,24	0,22	2,24	0,22	0,64	0,11
	MT14	1,22	0,11	-0,06	0,08	1,20	0,13	0,68	0,30	-0,06	0,09	1,26	0,13	1,26	0,13	-0,07	0,06
	MT21	1,94	0,17	-0,51	0,11	2,11	0,20	0,43	0,26	-0,53	0,12	2,06	0,17	2,06	0,17	-0,53	0,10
	MT22	1,04	0,10	-0,33	0,08	1,09	0,11	0,43	0,26	-0,34	0,08	1,09	0,11	1,09	0,11	-0,35	0,08
	MT31	1,40	0,13	0,83	0,09	1,91	0,16	1,46	0,61	1,13	0,24	1,51	0,14	1,51	0,14	0,88	0,09
	MT32	2,01	0,29	-4,08	0,34	2,44	0,53	1,46	0,61	-5,13	0,97	2,16	0,34	2,16	0,34	-4,32	0,41
	MT41	1,44	0,13	-1,09	0,10	1,51	0,15	0,70	0,22	-1,16	0,12	1,51	0,15	1,51	0,15	-1,16	0,12
	MT42	1,24	0,11	0,46	0,08	1,40	0,18	1,08	0,37	0,55	0,12	1,43	0,20	1,43	0,20	0,56	0,11
	MT43	1,18	0,12	-1,02	0,09	1,27	0,16	0,93	0,30	-1,14	0,14	1,27	0,13	1,27	0,13	-1,13	0,11

* MTij, i: madde takımı numarası, j: madde takımındaki soru sırası

Tablo 4.14: Devami

N	Madde	TBMTK				2FM				MTTK							
		a	SH	c	SH	a ₁	SH	a ₂	SH	c	SH	a ₂	SH	c	SH		
1000	B1	1,54	0,15	1,56	0,11	1,55	0,16	-	-	-2,39	0,16	1,55	0,16	-	-	-2,38	0,16
	B2	1,12	0,10	-1,15	0,12	1,13	0,11	-	-	1,30	0,10	1,13	0,11	-	-	1,30	0,10
	B3	1,40	0,11	0,64	0,07	1,42	0,13	-	-	-0,88	0,10	1,42	0,13	-	-	-0,87	0,10
	MT11	1,94	0,14	-0,30	0,07	2,13	0,22	0,98	0,30	0,68	0,13	2,12	0,22	2,12	0,22	0,69	0,13
	MT14	1,12	0,09	0,05	0,08	1,18	0,12	0,71	0,21	-0,05	0,09	1,18	0,12	1,18	0,12	-0,05	0,09
	B4	1,01	0,09	0,21	0,08	1,02	0,10	-	-	-0,20	0,08	1,02	0,10	-	-	-0,19	0,08
	B5	1,26	0,10	0,55	0,08	1,27	0,11	-	-	-0,69	0,09	1,27	0,11	-	-	-0,68	0,09
	MT21	1,87	0,14	0,26	0,06	1,95	0,17	0,48	0,22	-0,49	0,12	1,95	0,17	1,95	0,17	-0,48	0,12
	MT22	1,05	0,09	0,32	0,08	1,09	0,11	0,48	0,22	-0,33	0,09	1,09	0,11	1,09	0,11	-0,33	0,09
	B6	0,90	0,09	0,19	0,09	0,90	0,09	-	-	-0,16	0,08	0,90	0,09	-	-	-0,16	0,08
	B7	2,04	0,16	0,65	0,06	2,06	0,18	-	-	-1,31	0,14	2,07	0,18	-	-	-1,29	0,14
	MT31	1,44	0,12	-0,59	0,08	1,84	0,37	1,34	0,62	1,11	0,24	1,87	0,37	1,87	0,37	1,13	0,24
	MT32	2,29	0,30	1,94	0,12	2,80	0,56	1,34	0,62	-5,44	0,98	2,83	0,57	2,83	0,57	-5,47	0,98
	B8	1,50	0,12	-0,59	0,08	1,52	0,13	-	-	0,90	0,11	1,52	0,13	-	-	0,91	0,11
	B10	2,02	0,17	0,96	0,06	2,04	0,18	-	-	-1,93	0,16	2,05	0,18	-	-	-1,92	0,16
	B11	1,15	0,10	-0,64	0,09	1,15	0,11	-	-	0,75	0,09	1,16	0,11	-	-	0,75	0,09
	B12	1,05	0,10	-0,65	0,10	1,06	0,10	-	-	0,70	0,09	1,07	0,10	-	-	0,71	0,09
	B13	0,40	0,11	-5,62	1,52	0,40	0,11	-	-	2,25	0,12	0,40	0,11	-	-	2,25	0,12
	B14	1,87	0,16	1,04	0,07	1,88	0,17	-	-	-1,92	0,15	1,89	0,17	-	-	-1,91	0,15
	MT41	1,46	0,12	0,75	0,07	1,58	0,15	0,70	0,21	-1,18	0,13	1,58	0,16	1,58	0,16	-1,17	0,13
	MT42	1,10	0,10	-0,41	0,09	1,33	0,20	1,22	0,42	0,57	0,13	1,33	0,20	1,33	0,20	0,58	0,13
	MT43	1,09	0,10	0,91	0,09	1,24	0,16	0,96	0,29	-1,13	0,14	1,24	0,16	1,24	0,16	-1,12	0,14
	MT51	1,10	0,10	-0,32	0,08	1,14	0,11	0,47	0,15	0,38	0,09	1,14	0,11	1,14	0,11	0,38	0,09
	MT61	1,05	0,10	0,55	0,08	1,06	0,10	0,08	1,32	-0,57	0,09	1,07	0,10	1,07	0,10	-0,56	0,09
	MT62	1,52	0,12	-0,63	0,08	1,53	0,14	0,08	1,32	0,97	0,11	1,53	0,14	1,53	0,14	0,99	0,11
	MT71	0,71	0,11	-3,03	0,44	0,78	0,15	0,96	0,49	2,48	0,33	0,78	0,15	0,78	0,15	2,48	0,33
	MT72	0,85	0,09	0,85	0,11	0,94	0,12	0,68	0,31	-0,78	0,10	0,95	0,13	0,95	0,13	-0,77	0,11
	MT73	0,93	0,09	-0,70	0,11	0,93	0,09	0,10	0,17	0,66	0,08	0,94	0,09	0,94	0,09	0,67	0,08
	MT74	1,44	0,11	0,12	0,07	1,59	0,17	0,65	0,30	-0,17	0,10	1,59	0,17	1,59	0,17	-0,16	0,10
	MT81	1,27	0,12	-1,13	0,11	1,58	0,30	1,24	0,64	1,82	0,35	1,58	0,30	1,58	0,30	1,83	0,35
MT82	1,75	0,16	1,21	0,07	2,30	0,69	1,68	1,05	-2,83	0,85	2,29	0,67	2,29	0,67	-2,80	0,82	
MT83	3,42	0,46	1,50	0,06	3,59	0,58	1,03	0,49	-5,47	0,78	3,60	0,58	3,60	0,58	-5,46	0,78	

EK 4. ORJİNALLİK RAPORU

Turnitin x

https://www.turnitin.com/t_inbox.asp?r=72.0698004090675&svr=10&lang=tr&aid=47676367

Esin Yılmaz Koğar | Kullanıcı Bilgisi | Mesajlar | Öğretmen ▼ | Türkçe ▼ | Roadmap | Yardım | Çıkış

turnitin

Ödevler Öğrenciler Not Defteri Kütüphaneler Takvim Tartışma Tercihler

GÖRÜNTÜLENİYOR: ANASAYFA > DENEME > DENEME

Bu sayfa hakkında

Bu sizin ödev gelen kutunuzdur. Bir ödevi görüntülemek için, ödev başlığına tıklayın. Orijinallik Raporu'nu görmek için, benzerlik kolonundaki orijinallik raporu ikonuna tıklayın. Bu ikon tıklanabilir durumda değilse, orijinallik raporu henüz oluşturulmamış demektir.

deneme

GELEN KUTUSU | GÖRÜNTÜLENİYOR: YENİ ÖDEVLER ▼

Dosyayı Gönder

GradeMark Raporu | Ödev ayarlarını düzenle | E-posta bildirmeyenler

<input type="checkbox"/>	YAZAR	BAŞLIK	BENZERLİK	PUANLA	CEVAP	DOSYA	ÖDEV NUMARASI	TARİH
<input type="checkbox"/>	Esın Yılmaz Koğar	MADDE TAKIMLARI İÇEREN TESTLERDE FARKLI ...	%3 ■		•		640727006	04-Mar-2016

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı	Esin Yılmaz Koğar
Doğum Yeri	Samsun
Doğum Tarihi	27/04/1987

Eğitim Durumu

Lise	Ünye Mehmet Refik Güven Anadolu Öğretmen Lisesi - Ordu	2004
Lisans	Başkent Üniversitesi – İlköğretim Matematik Öğretmenliği - Ankara	2008
Yüksek Lisans	Hacettepe Üniversitesi İlköğretim Bölümü - Ankara	2011
Yabancı Dil	İngilizce: Okuma (Çok iyi), Yazma (İyi), Konuşma (Orta)	

İş Deneyimi

Stajlar	Ünye Özel Pi Eğitim Dersanesi – Matematik Öğretmenliği	2008 - 2009
Çalıştığı Kurumlar	Niğde Üniversitesi – Araştırma Görevlisi	2012 - 2013
	Hacettepe Üniversitesi – Araştırma Görevlisi	2013 - 2016

Akademik Çalışmalar

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

Aygün, B., & Yılmaz, E. (2015). *Temel eğitimden orta öğretime geçiş sınavı (TEOG) matematik temel alanın kapsam geçerliği açısından incelenmesi*. 24. Ulusal Eğitim Bilimleri Kongresi, 16-19 Nisan 2015, Niğde, Türkiye.

İnal, H., Yılmaz, E., Demirdüzen, E., & Gelbal, S. (2014). *Cronbach alfa katsayısı: Bir meta analizi çalışması*. IV. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi (Uluslararası Katılımlı) 9-13 Haziran 2014, Ankara, Türkiye.

Koğar, H., & Yılmaz, E. (2015). *Comparison of different estimation techniques for categorical and ordered data in confirmatory factor analysis*, International Congress on Education for the Future: Issues and Challenges (ICEFIC 2015) 13-15 May 2015, Ankara, Türkiye.

Koğar, H., & Yılmaz-Koğar, E. (2015). Comparison of different estimation methods for categorical and ordinal data in confirmatory factor analysis. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6(2), 351-364.

Yılmaz, E., & Koğar, H. (2014). *Uç değerle baş etmede kullanılan farklı tekniklerin bazı istatistiksel analiz sonuçları üzerindeki etkisi*. IV. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi (Uluslararası Katılımlı) 9-13 Haziran 2014, Ankara, Türkiye.

Yılmaz, E., & Koğar, H. (2015). Uç değerle baş etmede kullanılan farklı tekniklerin bazı istatistiksel analiz sonuçları üzerindeki etkisi. *Başkent University Journal of Education*, 2(1), 61-67.

Yılmaz-Koğar, E., & Aygun, B. (2015). Temel eğitimden orta öğretime geçiş sınavı (TEOG)'nın matematik temel alanına ait testlerin kapsam geçerliğinin incelenmesi. *Pegem Eğitim ve Öğretim Dergisi*, 5(5), 667-680.

Yılmaz-Koğar, E., & Çakıcı-Eser, D. (2015). Tek ve çok boyutlu madde tepki kuramına dayalı bir veri analizi yazılımı: IRTPRO 2.1. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6(2), 207-217.

Yılmaz-Koğar, E., & Demircioğlu, E. (2016). 6.-7.-8. sınıf matematik dersi öğrenme alanlarına ilişkin ölçekleme çalışması. *Başkent University Journal of Education*, 3(1), 44-52.

Yılmaz-Koğar, E., & Koğar, H. (2015). *Examination of the variable that influence the engagement with and at school through a path analysis*. Multidisciplinary Academic Conference on Education, Teaching and E-learning 4-5 December 2015, Czech Republic (MAC-ETeL 2015), Prague.

İletişim

e-Posta Adresi	esinyilmaz@hacettepe.edu.tr
Jüri Tarihi	27/01/2016