



**TIMSS 2011 MATEMATİK ALT TESTİ MADDE PARAMETRELERİNİN
KTK ile MTK'YA GÖRE BILOG MG, MULTILOG ve R
PROGRAMLARIYLA KARŞILAŞTIRMALI OLARAK İNCELENMESİ**

Fazilet Gül İnce

YÜKSEK LİSANS TEZİ
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME ANA BİLİM DALI

GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

ŞUBAT, 2016

TELİF HAKKI ve TEZ FOTOKOPİ İZİN FORMU

Bu tezin tüm hakları saklıdır. Kaynak göstermek koşuluyla tezin teslim tarihinden itibaren yirmi dört (24) ay sonra tezden fotokopi çekilebilir.

YAZARIN

Adı: Fazilet Gül

Soyadı: İnce

Bölümü: Eğitimde Ölçme ve Değerlendirme

İmza:

Teslim tarihi:

TEZİN

Türkçe Adı: TIMSS 2011 Matematik Alt Testi Madde Parametrelerinin KTK ile MTK'ya Göre BILOG MG, MULTILOG ve R Programlarıyla Karşılaştırmalı Olarak İncelenmesi

İngilizce Adı: Comparative Investigation of the TIMSS 2011 Mathematics Subtest's Item Parameters According to Classical Test Theory And Item Response Theory With BILOG MG, MULTILOG and R Programs

ETİK İLKELERE UYGUNLUK BEYANI

Tez yazma sürecinde bilimsel ve etik ilkelere uyduğumu, yararlandığım tüm kaynakları kaynak gösterme ilkelerine uygun olarak kaynakçada belirttiğimi ve bu bölümler dışındaki tüm ifadelerin şahsıma ait olduğunu beyan ederim.

Yazar Adı Soyadı: Fazilet Gül İnce

İmza:

Jüri onay sayfası

Fazilet Gül İnce tarafından hazırlanan “TIMSS 2011 Matematik Alt Testi Madde Parametrelerinin KTK ile MTK’ya Göre BILOG MG, MULTİLOG ve R Programlarıyla Karşılaştırmalı Olarak İncelenmesi” adlı tez çalışması aşağıdaki jüri tarafından oy birliği ile Gazi Üniversitesi Eğitimde Ölçme ve Değerlendirme Anabilim Dalı’nda Yüksek Lisans tezi olarak kabul edilmiştir.

Danışman: Prof. Dr. Mehtap Çakan

(Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Gazi Üniversitesi)

Başkan: Doç. Dr. Nuri Doğan

(Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Hacettepe Üniversitesi)

.....

Üye: Doç. Dr. Şeref Tan

(Eğitimde Ölçme ve Değerlendirme Anabilim Dalı)

.....

Tez Savunma Tarihi: 17.02.16

Bu tezin Eğitimde Ölçme ve Değerlendirme Anabilim Dalı’nda Yüksek Lisans tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Servet KARABAĞ

Eğitim Bilimleri Enstitüsü Müdürü

.....



Canim aileme

TEŞEKKÜR

Tez yazma sürecimin başından sonuna kadar beni destekleyen, çalışmalarım konusunda beni kısıtlamayan danışmanım Prof. Dr. Mehtap ÇAKAN'a anlayışı ve destekleri için teşekkürlerimi sunarım.

Yüksek lisans eğitimine bu alanda başlama nedenim Prof. Dr. Şener BÜYÜKÖZTÜRK'e, sorularımı asla yanıtsız bırakmayan, değerli vaktini bana ayıran Doç. Dr. Şeref TAN'a, derin bilgisinden faydalandığım, bana Hacettepe Üniversitesi'nin kapılarını açan Doç. Dr. Nuri DOĞAN'a ve tez yazma sürecimde büyük katkısı olan Yrd. Doç. Dr. Emine ÖNEN'e desteklerinden dolayı teşekkürlerimi bir borç bilirim.

Bana okumayı sevdiren, eğitim hayatımı değiştiren, eğitimci olmamı sağlayan ilkokul öğretmenim Fahri AKÇA'ya, yüksek lisans yapma isteğimin başkahramanı olan kıymetli dostum Yağmur Uysal'a, bana Ankara'yı sevdiren, her konuda desteklerini esirgemeyen manevi kardeşim Arş. Gör. Yıldız YILDIRIM'a ve desteklerini esirgemeyen dostlarım Arş. Gör. Tuba GÜNDÜZ, Arş. Gör. Elif SEZER ve Arş. Gör. Fatma Gökçen AYVA'ya çok teşekkür ederim.

Son olarak beni hayata hazırlayan güçlü insan anneme, bir dediğimi iki etmeyen ve bana her zaman inanan babama, ablalıktan çok annelik yapan ablalarım, varlığı bile beni rahatlatan abime ve bana her zaman şanslı olduğumu hissettiren nişanlıma derinden teşekkür ederim.

**TIMSS 2011 MATEMATİK ALT TESTİ MADDE
PARAMETRELERİNİN KTK İLE MTK'YA GÖRE BILOG MG,
MULTILOG VE R PROGRAMLARIYLA KARŞILAŞTIRMALI
OLARAK İNCELENMESİ
(Yüksek Lisans Tezi)**

**Fazilet Gül İnce
GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ
Şubat, 2016**

ÖZ

Bu araştırmanın amacı, TIMSS 2011 4. Sınıf Türkiye matematik alt testinin madde parametreleri ve özelliklerini Madde Tepki Kuramı (MTK)'nın bir, iki ve üç parametrelili lojistik modellerine ve Klasik Test Kuramına (KTK) göre BILOG MG, R, MULTILOG programları ile karşılaştırmalı olarak incelemektir. Araştırmanın örneklemini TIMSS 2011 Türkiye uygulamasına katılan 4. sınıf öğrencilerinden 11. matematik alt testini alan 543 öğrenci oluşturmaktadır. Bu araştırma var olan yöntem ve tekniklerin sınanması yapıldığı için betimsel araştırma türündedir. Araştırma sonuçlarına göre ikili puanlanan maddeler için hesaplanan KTK güçlük katsayıları ve kestirilen MTK b parametreleri arasındaki korelasyon negatif yönde yüksek ve manidar bulunmuştur. KTK ayırt edicilik katsayıları ile MTK a parametreleri arasındaki korelasyonun 1PL ve 2PL için pozitif yönde, yüksek ve manidar, 3PL model için ise pozitif yönde, orta düzeyde ve manidar olduğu görülmüştür. BILOG MG, MULTILOG ve R yazılımlarına göre yapılan parametre kestirimlerinden elde edilen sonuçlar incelendiğinde ise BILOG MG ve MULTILOG programlarının birbirine yakın parametre kestirimleri verdiği tespit edilmiştir. Model veri uyumu incelendiğinde, BILOG yazılımına göre en uyumlu modelin 2PL model; MULTILOG ve R yazılımlarına göre ise 3PL model olduğu görülmüştür. Güvenirlik analizleri için KTK'ya göre Cronbach alfa iç tutarlık katsayısı hesaplanmış, MTK'ya göre ise marjinal ve görgül güvenirlilik katsayıları hesaplanmıştır. KTK ve MTK'dan elde edilen güvenirlilik katsayıları arasındaki farkın manidarlığı test edilmiş, 3PL model için her üç yazılımda; 2PL model için BILOG yazılımında kestirilen güvenirlilik katsayısı ile Cronbach alfa katsayısı arasında manidar fark bulunmuştur. Son olarak programların kestirdiği parametrelerin standart hataları karşılaştırılmış ve genel olarak R ve BILOG MG programlarının MULTILOG programına göre daha düşük standart hata verdiği tespit edilmiştir.

Bilim Kodu: 6015

Anahtar Kelimeler: Klasik Test Kuramı, Madde Tepki Kuramı, Lojistik Model, BILOG MG, R, MULTILOG, Madde Parametreleri.

Sayfa Adedi: xvi + 99

Danışman: Prof. Dr. Mehtap ÇAKAN

**COMPARATIVE INVESTIGATION OF THE TIMSS 2011
MATHEMATIC SUBTEST ACCORDING TO CLASSICAL TEST
THEORY AND ITEM RESPONSE THEORY WITH BILOG MG,
MULTILOG and R
(M.S Thesis)**

**Fazilet Gül İnce
GAZI UNIVERSITY
GRADUATE SCHOOL OF EDUCATIONAL SCIENCES**

February, 2016

ABSTRACT

The purpose of this study is to scrutinize the item parameters and characteristics of 4th grade mathematics subtest in TIMSS 2011, Turkey, according to Classical Test Theory (CTT) and one parameter, two parameter and three parameter logistic models of Item Response Theory (IRT) by using BILOG, MG, R, MULTILOG programs. The sample of the research consists of 543 students who participated in TIMSS 2011, Turkey and took a subtest 11th which from 4th grade mathematics tests. This research is a descriptive study of background study as the existing institutions were examined using real data, analyzed and interpreted. According to research results, it was found out that the correlation between difficulty coefficients of dichotomous items of CTT and b parameters of IRT was found negative, high and significant. In addition, while correlation between discrimination coefficients of CTT and a parameters of IRT for one parameter logistic and two parameter logistic models was positive, high and significant, for 3 parameter logistic model, it was positive and significant in the middle level. When the results obtained using BILOG, MG, MULTILOG and R were investigated, it was found out that BILOG MG and MULTILOG programs yielded almost similar results. When model-data fit was scrutinized, it was deduced that while the fittest model for BILOG was 2 parameter logistic model, according to R and MULTILOG 3 parameter logistic model was the fittest one. For reliability analysis, Cronbach alfa internal consistency coefficient was calculated using CTT, and marginal reliability coefficients were calculated using IRT. The difference between reliability coefficients from CTT and IRT was tested and for three software in 3 parameter model, and in 2 parameter model, difference between reliability coefficient estimated by BILOG and Cronbach alpha coefficient was significant. Finally, the standard error of the programs were calculated and it was found that there was a significant difference for b parameter of 1PL model. In general, it was detected that R, BILOG and MG programs yielded lower error than MULTILOG program.

Science Code: 6015

Key Words: Classical Test Theory, Item Response Theory, Logistic Model, BILOG MG, R, MULTILOG, Item Parameters.

Page Number: xvi + 99

Supervisor: Prof.Mehtap Çakan



İÇİNDEKİLER

TELİF HAKKI ve TEZ FOTOKOPİ İZİN FORMU.....	i
ETİK İLKELERE UYGUNLUK BEYANI.....	ii
Jüri onay sayfası.....	iii
TEŞEKKÜR	v
ÖZ.....	vi
ABSTRACT.....	vii
İÇİNDEKİLER	ix
TABLolar LİSTESİ.....	xiii
ŞEKİLLER LİSTESİ	xv
SİMGE VE KISALTMALAR LİSTESİ.....	xvi
BÖLÜM I.....	1
GİRİŞ	1
Test Kuramları	2
Klasik Test Kuramı	3
Klasik Test Kuramında Güvenirlik.....	4
Klasik Test Kuramında Geçerlik.....	4
Madde Tepki Kuramı	6
Madde Karakteristik Eğrisi (MKE)	8
Madde Tepki Kuramı Sayıtları.....	9
Madde Tepki Kuramı Avantajları.....	11
Madde Tepki Kuramının Sınırlılıkları.....	11
Madde Tepki Kuramında Güvenirlik ve Geçerlik	12
Madde Tepki Kuramı Modelleri	12
Madde Tepki Kuramı ile Klasik Test Kuramı Arasındaki Temel Farklılıklar	15
MTK Çalışmalarında Kullanılan Yaygın Programlar	16
BILOG MG Programı.....	16

MULTILOG 7.03 Programı	16
R Studio Yazılımı ve paketleri.....	17
Araştırmanın Amacı	18
Problem İfadesi.....	18
Araştırma Soruları.....	18
Araştırmanın Önemi	20
Sınırlılıklar	20
Tanımlar.....	20
BÖLÜM II	23
İLGİLİ ARAŞTIRMALAR	23
KTK'ya Göre Kestirilen Madde İstatistikleri ve MTK'ya Göre Kestirilen Madde Parametrelerinin Karşılaştırıldığı Çalışmalar	23
MTK Kestirimleri Yapan Programların Karşılaştırıldığı Araştırmalar	32
BÖLÜM III.....	35
YÖNTEM.....	35
Araştırmanın Deseni	35
Evren ve Örneklem	35
Veri Toplama Araçları.....	35
Verilerin Analizi	36
1. Araştırma Sorusuna İlişkin Veri Analizi.....	37
2. Araştırma Sorusuna İlişkin Veri Analizi.....	38
3. Araştırma Sorusuna İlişkin Veri Analizi.....	39
4. Araştırma Sorusuna İlişkin Veri Analizi.....	39
5. Araştırma Sorusuna İlişkin Veri Analizi.....	40
KTK'ya İlişkin Geçerlik Çalışması	40
Madde Tepki Kuramının Sayıtlarının Kontrolü	41
Normallik.....	41
Tek Boyutluluk	42
Yerel Bağımsızlık	43
Testin Hız Testi Olup Olmadığının Kontrol Edilmesi	44
Model Veri Uyumu	44
Model-Madde Uyumu	45
Madde Parametrelerinin Değişmezliği	46
BÖLÜM IV	49
BULGULAR ve YORUM.....	49

1. Araştırma Sorusuna İlişkin Bulgular	49
1.1. Araştırma Sorusuna İlişkin Bulgular	50
1.2. Araştırma Sorusuna İlişkin Bulgular	52
1.3. Araştırma Sorusuna İlişkin Bulgular	54
2. Araştırma Sorusuna İlişkin Bulgular	56
2.1. Araştırma Sorusuna İlişkin Bulgular	57
2.2. Araştırma Sorusuna İlişkin Bulgular	59
2.3. Araştırma Sorusuna İlişkin Bulgular	60
3. Araştırma Sorusuna İlişkin Bulgular	62
4. Araştırma Sorusuna İlişkin Bulgular	63
4.1. Araştırma Sorusuna İlişkin Bulgular	63
4.2. Araştırma Sorusuna İlişkin Bulgular	63
5. Araştırma Sorusuna İlişkin Bulgular	64
BÖLÜM V.....	71
SONUÇ ve TARTIŞMA.....	71
1. Araştırma Sorusuna İlişkin Sonuçlar.....	71
1.1. Araştırma Sorusuna İlişkin Sonuçlar	71
1.2. Araştırma Sorusuna İlişkin Sonuçlar	72
1.3. Araştırma Sorusuna İlişkin Sonuçlar	72
2. Araştırma Sorusuna İlişkin Sonuçlar.....	73
2.1. Araştırma Sorusuna İlişkin Sonuçlar	73
2.2. Araştırma Sorusuna İlişkin Sonuçlar	73
2.3. Araştırma Sorusuna İlişkin Sonuçlar	74
3. Araştırma Sorusuna İlişkin Sonuçlar.....	74
4. Araştırma Sorusuna İlişkin Sonuçlar.....	75
4.1. Araştırma Sorusuna İlişkin Sonuçlar	75
4.2. Araştırma Sorusuna İlişkin Sonuçlar	75
5. Araştırma Sorusuna İlişkin Sonuçlar.....	75
BÖLÜM VI.....	77
ÖNERİLER	77
Gelecekte Yapılacak Araştırmalar İçin Öneriler	77
KAYNAKÇA	79
EK 1: Mplus Programı Tetrakorik AFA Syntax	88
EK 2: LISREL Programı DFA Syntax	89

EK 3: BILOG Programında Kullanılan Syntax	90
EK 4: MULTİLOG Programında Kullanılan Syntax	91
EK 5: R Programı ile Elde Edilen Q3 İstatistikleri	92



TABLÖLAR LİSTESİ

Tablo 1. Veri Setine İlişkin Betimsel İstatistikler.....	36
Tablo 2. Veri Setine Ait DFA Sonuçları.....	41
Tablo 3. Veri Setine İlişkin Kolmogorov-Smirnov ve Shapiro Wilk Testi Sonuçları.....	41
Tablo 4. Veri Setine Ait Merkezi Dağılım Ölçüleri	42
Tablo 5. Tetrakorik AFA'ya Ait Özdeğerler	42
Tablo 6. Model Madde Uyumuna İlişkin Kay Kare (KK) ve p değerleri.....	46
Tablo 7. KTK'ya Göre Kestirilen p Katsayıları.....	50
Tablo 8. BILOG Programında Kestirilen b Parametreleri	51
Tablo 9. MULTILOG'a Göre Elde Edilen b Parametreleri	53
Tablo 10. R'ye göre elde edilen b parametreleri.....	55
Tablo 11. KTK'ya Göre Kestirilen r Katsayıları	57
Tablo 12. BILOG MG Programında Kestirilen a Parametreleri.....	58
Tablo 13. MULTILOG Programında Kestirilen a Parametreleri.....	59
Tablo 14. R Programında Kestirilen a Parametreleri.....	61
Tablo 15. BILOG MG, R ve MULTILOG Programlarında Lojistik Modellere Göre Hesaplanan -2 log likelihood Değerleri	62
Tablo 16. BILOG MG, R ve Programları için χ^2 Sonuçları.....	62
Tablo 17. Görgül ve Marjinal Güvenirlilik Katsayıları	63
Tablo 18. 1PL Model İçin BILOG MG, R ve MULTILOG Programlarında Kestirilen b Parametresi Standart Hataları için Friedman Sonucu	64
Tablo 19. 2PL Model İçin BILOG MG, R ve MULTILOG Programlarında Kestirilen b Parametresi Standart Hataları için Friedman Sonucu	66
Tablo 20. 3PLM için BILOG MG ve R Programlarında Kestirilen b Parametresi Standart Hataları için Friedman Testi Sonucu	67
Tablo 21. 2PL Model için BILOG MG, R ve MULTILOG Programlarında Kestirilen a Parametresi Standart Hataları için Friedman Testi Sonucu	67

Tablo 22. 3PLM için BILOG MG, R Programlarında Kestirilen a Parametresi Standart Hataları için Friedman Testi Sonucu	69
Tablo 23. 3PL Model için BILOG MG ve R Programlarında Kestirilen c Parametresi Standart Hataları için Friedman Testi Sonucu	69



ŞEKİLLER LİSTESİ

Şekil 1. Madde parametrelerinin madde karakteristik eğrisinde gösterimi	7
Şekil 2. Madde karakteristik eğrisi (madde karakteristik fonksiyonu)	8
Şekil 3. TIMSS 2011 4. sınıf matematik alt testine ait saçılma grafiği	51

SİMGE VE KISALTMALAR LİSTESİ

AFA	Açımlayıcı Faktör Analizi
DFA	Doğrulayıcı Faktör Analizi
KTK	Klasik Test Kuramı
MKE	Madde Karakteristik Eğrisi
MML	Marjinal Maksimum Likelihood
MTK	Madde Tepki Kuramı
1PLM	Bir parametrelili lojistik model
2PLM	İki parametrelili lojistik model
3PLM	Üç parametrelili lojistik model
a	Madde Ayırt Edicilik Parametresi
b	Madde Güçlük Parametresi
c	Şans Parametresi
θ	Yetenek Düzeyi
$P_i(\theta)$	Maddeye Doğru Yanıt Verme Olasılığı
$SE(\theta)$	Tahminin Standart Hatası
p	Madde güçlük indeksi
r	Madde ayırt edicilik indeksi
Sd	Serbestlik derecesi

BÖLÜM I

GİRİŞ

Problem Durumu

Türkiye’de eğitim sistemimizin öğrenci başarısı üzerindeki etkilerini açık bir şekilde görmek amacıyla çeşitli testler ve anketler uygulanmaktadır. Öğretim programlarında düzenleme yapabilmek, öğretim sistemlerindeki eksiklikleri giderebilmek ve ülkelerin uluslararası düzeyde kendi düzeylerini görebilmesi amacıyla son yıllarda TIMSS (Trends in International Mathematics and Science Study), PISA (Program for International Student Assessment) ve PIRLS (Progress in International Reading Literacy Study) gibi sınavlar gerçekleştirilmektedir. Türkiye’de, eğitimin öğrenci başarısı üzerindeki etkisini görmek amacı ile ulusal ve uluslararası alanda ve farklı sınıf düzeylerinde standart testler ve anketler aracılığı ile ölçme değerlendirme çalışmaları yapılmaktadır (Özer ve Anıl, 2011).

Kuruluşların yönettiği çalışmaların en önemlilerinden biri Uluslararası Matematik ve Fen Eğilimleri Araştırması’dır. TIMSS öğrencilerin matematik ve fen alanlarındaki bilgi ve becerilerin değerlendirilmesi amacıyla uygulanan bir tarama araştırmasıdır (Şişman, Acat, Aypay ve Karadağ, 2011). Dünyadaki en büyük ve en kapsamlı uluslararası öğrenci başarılarını değerlendirme çalışması olarak kabul görmektedir. 4. ve 8. sınıf düzeyindeki öğrencilere uygulanır ve 4 yılda bir yapılmaktadır. TIMSS öğrenci başarılarındaki eğilimleri izlemekte ve ulusal eğitim sistemleri arasındaki farklılıkları belirlemektedir (EARGED [Eğitimi Araştırma ve Geliştirme Dairesi], 2007). TIMSS sınavı ilk olarak 1995 yılında yapılmış olup 4 yıl aralıklarla tekrarlanmaktadır. IEA (International Association for the Evaluation of Educational Achievement) tarafından gerçekleştirilen bu proje uzun soluklu bir projedir; çünkü asıl hedef ülkelerin TIMSS’den elde ettikleri çıkarımları sisteme yansıtıktan sonra değişimlerin izlenmesidir. 4 yıl arayla yapıldığı için, önceki sınav ile sonraki sınav karşılaştırılmakta her ülkenin gelişim durumu incelenmektedir (Robitaille ve Beaton, 2002).

Uluslararası sınavlar amaç bakımından ulusal sınavlardan farklıdır. TIMSS araştırmalarında amaç eğitim politikasını belirlemek, öğretim programlarını hazırlayan uzmanların ve araştırmacıların kendi eğitim sistemlerinin işleyişini daha iyi anlayabilmeleri konusunda yardımcı olmaktadır. 1996 yılında TIMSS 1995 raporlarının yayınlanmasından sonra elde edilen sonuç ülkemizde yenilik hareketleri için tartışma ortamı yaratmış ve sonuçlar önemli bilgiler sağlamıştır. (EARGED, 2003).

Bilindiği üzere, TIMSS matematik sınavlarında öğrencilere 26 kitapçık ve her kitapçıkta farklı maddeler sunulmaktadır. Bu durum, beraberinde farklı maddeleri alan öğrencilerin başarılarının değerlendirilmesindeki sorunları getirmektedir. Ancak TIMSS çalışmaları seçkisiz örneklem üzerinde gerçekleştirdiğinden sonuçları genellenebilir olarak kabul edilmektedir (Berberoğlu, Çelebi, Özdemir, Uysal ve Yayan, 2003).

TIMSS projesinin temel amacı, matematik ve fen eğitim öğretimine evrensel olarak katkıda bulunmaktır. TIMSS, katılımcı ülkelere aşağıdaki soruların cevabının bulunmasında yardımcı olmaktadır (MEB [Milli Eğitim Bakanlığı], 2015).

- Öğrencilerin matematik ve fende durumu nedir?
- Ülkelerin matematik ve fende durumu pozitif yönde ilerliyor mu?
- İçinde bulunulan durum nasıl geliştirebilir?
- Diğer ülkelerle kıyaslandığında ülkenin durumu nasıl?
- Diğer ülkeler başarılarını artırmak amacıyla neler yapıyorlar?

Test Kuramları

Psikolojik yapılar soyuttur ve dolaylı yollarla ölçülebirlirler. Bu sebeple ölçülmek istenen bir yapı için ölçme aracı geliştirmek zordur. Yapıların ölçümünde kullanılabilecek tek ve evrensel bir yaklaşım yoktur. Psikoloji ve eğitim alanındaki ölçme problemlerine ve bunların çözüm yöntemlerine ilişkin çalışmalardan oluşan disiplin alanı test kuramı olarak tanımlanmaktadır. Test kuramlarının genel amacı yapıya ilişkin ölçmelerin ölçme problemlerinden ne kadar etkilendiğini saptamak ve ölçme hatalarını en aza indirecek yöntemleri bulmaktır (Crocker ve Algina, 2008).

Eğitim ve psikolojide yaygın olarak iki ölçme kuramı kullanılmaktadır. Bunlar Madde Tepki Kuramı (MTK) ve Klasik Test Kuramı (KTK) olarak adlandırılmaktadır. KTK'nın

geçmiş daha eskiye dayandığı için, alana katkısı daha çoktur. Ancak MTK, ölçme problemlerini çözmedeki potansiyelinin daha yüksek olduğu kabul edildiği için günümüzde daha çok tercih edilmektedir (Embretson ve Reise, 2000).

Test kuramı ile eğitimsel ve psikolojik değerlendirme birbirinden farklı kavramlardır. Eğitimsel ve psikolojik değerlendirmeler bir testin uygulanması ve yorumlanmasına dönük işlemlere odaklanırken, test kuramı bir aracın geliştirilme işlemlerine ışık tutacak genel bir çerçeve sağlamaya odaklanmaktadır (Crocker ve Algina, 2008). Psikolojide kullanılan modeller, davranışı yapılarla ilişkilendiren matematiksel eşitliklerdir. Psikolojik ölçmeler için model kavramının kullanılan en geçerli anlamı, bağımsız değişkenleri bağımlı değişkenlerle ilişkilendiren ve sayısal ifadeleri içeren matematiksel ifadelerdir (Embretson ve Reise, 2000).

Klasik Test Kuramı

Klasik Test Kuramı (KTK), bireyin gözlenen özelliğinin puanı içinde hem gerçek puan hem de hata puanı bulunduğunu kabul eder.

Klasik Test Kuramı'nın temel denklemi bazı sayıtları olmakla birlikte aşağıdaki gibidir:

$$X = T + E$$

X: Gözlenen Puan T: Gerçek Puan E: Hata Puanı (Crocker ve Algina, 2008).

Bu denklem bir bireyin belirlenen ölçekten elde ettiği puanın, bireyin gerçek puanı ve ölçme işlemi hatasının toplamına eşit olduğu anlamına gelmektedir. Ölçme işlemlerdeki hataların azlığı, gerçeğe yakın puanlar elde etmemizi sağlar (Baykul, 2000). Ayrıca yukarıda tanımlanan gerçek puan modeline ilişkin bazı varsayımlar vardır.

Bu varsayımlar aşağıda sıralanmıştır:

1. Evrendeki öğrencilerin hata puanlarının ortalaması sıfırdır ($\mu_e = 0$).
2. Evrendeki öğrencilerin gerçek puan ve hata puanları arasındaki korelasyon daima sıfırdır ($\rho_{Te} = 0$).
3. Öğrenciler iki farklı test ya da aynı testi iki kez alırsa her öğrencinin iki testten aldığı puanlardan seçkisiz olarak seçilen hata puanları arasındaki korelasyon sıfır ($\rho_{e1e2} = 0$) kabul edilir (Crocker ve Algina, 2008; Lord, 1980).

Klasik Test Kuramında, madde ve test istatistikleri örnekleme bağılıdır. Hata varyansı ve standart hata tüm bireyler için eşittir. KTK' da bireylerin yetenek düzeyleri maddelere

bağılıdır, bu yüzden üst ve alt gruptaki yetenek kestirimleri için uygun değildir. (Hambleton ve Swaminathan, 1985).

Klasik Test Kuramında Güvenirlik

Güvenirlik, ölçme aracının duyarlı ve tutarlı ölçme sonuçları verme gücüdür. Klasik test kuramında güvenilirlik katsayısı, gerçek puan varyansının gözlenen puan varyansına oranlanması ile elde edilir. Gözlenen puan varyansı bazı sayıtlar sağlandığında gerçek puan varyansı ile hata-puanı varyansının toplamıdır. Gerçek varyans dışındaki tüm varyansın hatadan kaynaklanabileceği düşünülür. Aynı testin iki ya da daha çok uygulanması ile elde edilen puanlar arasındaki ilişki kararlılık veya tutarlılık olarak adlandırılır. Ölçme aracının maddelerinden elde edilen puanların tutarlılığı ise iç tutarlılık güvenilirliğidir. İç tutarlılıktaki hata kaynağının, ölçme aracıda yer alan maddeler olduğu kabul edilir. Diğer bir yöntem olan paralel formlar güvenilirliğinde ise iki paralel formun uygulanması sonucu elde edilen puanların tutarlılığı incelenir ve hata kaynağı ölçme aracı olarak kabul edilmektedir. Ölçmede birden fazla puanlayıcının ölçülen bir özelliğe verdikleri puanlar arasındaki tutarlılık incelenirken ise puanlayıcılar arası güvenilirlik söz konusudur ve hata kaynağı puanlayıcılar olarak kabul edilir (Crocker ve Algina, 2008; Güler, 2008).

Cronbach alfa iç tutarlık katsayısı hesaplanırken kullanılan formül şu şekildedir:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{l=1}^k S_l^2}{S_x^2} \right)$$

Formülde k soru sayısını S_l^2 madde varyansını, S_x^2 ise testin varyansını temsil etmektedir. Cronbach alfa katsayısı 1,00'a yaklaştıkça maddelerin iç tutarlılığının yüksek olduğunu gösterir.

Klasik Test Kuramında Geçerlik

Geçerlik bir ölçme aracının ölçülmek istenen özelliği amacına uygun olarak ölçebilme gücüdür. Testin geçerliğini tahmin etmek için farklı yöntemler mevcuttur. Bunlar yapı geçerliği, kapsam geçerliği, ölçüt geçerlik ve görünüş geçerliğidir.

Doğrudan ölçmelerde geçerlik daha yüksek, dolaylı ölçmelerde ise değişkenin yeterince tanımlanamaması ve kriterlerinin yeterince duyarlı olmamasından dolayı daha düşük sonuçlar verebilmektedir. Ölçeğin geçerlik düzeyi onun geçerlik katsayısının

hesaplanmasıyla anlaşılır ve $-1,00$ ile $+1,00$ arasında değerler alır. Katsayı ne kadar yüksekse ölçek amacına o kadar hizmet ediyor demektir (Karasar, 2000; Tekin, 2014).

Klasik Test Kuramı'nın avantajı uygulamaların karmaşık yapıda olmaması ve uygulama kolaylığıdır. Ancak kuram bazı eleştiriler almaktadır ve bunlar, çok sayıda sayılıtısının olması, sayılıtların her zaman karşılanamaması, bu kuramdan elde edilen madde ve test istatistiklerinin testin uygulandığı gruba ve teste bağımlı olması gibi sınırlılıklardır (Embretson ve Reise, 2000; Lord ve Novick, 1968).

KTK'da yetenek veya cevaplayıcı özellikleri test ile belirlenirken test zor ise cevaplayıcının yetenek düzeyinin düşük olduğu; eğer test kolay ise cevaplayıcının yeteneğinin yüksek olduğu yorumu yapılır. Ayrıca madde güçlük indeksi ve madde ayırt ediciliği parametreleri de cevaplayıcı grubun yeteneğine bağlı olduğundan, madde ayırtıcılık ve güçlük indeksi, testin geçerlik ve güvenilirliği belirlenen grup açısından tanımlanabilir (Hambleton, Swaminathan ve Rogers, 1991).

KTK diğer kuramlar gibi varsayımlara dayanır ve bu varsayımları karşılamak kolay sayılır. Ancak kuramın doğasından kaynaklı bazı sınırlılıklar mevcuttur (Hambleton ve Swaminathan, 1985).

Klasik test kuramının bazı sınırlılıkları şu şekilde sıralanabilir:

1. KTK'da madde istatistikleri gruba bağımlı olarak kestirilir. Örneklem yüksek yetenekli bireylerden oluşuyorsa maddeler kolay madde olarak nitelendirilir, ancak bireylerin yeteneği düşükse madde zor olarak nitelendirilir. Ayırt edicilik, bireylerin yeteneklerinden etkilenmektedir.
2. Sadece aynı formu veya paralel formları alan bireylerin puanları karşılaştırılabilmektedir. Paralel formlar ile yapılan ölçmelerde, ölçülen özellik, madde sayıları aynı olmalıdır. Ayrıca standart sapmalar, güvenilirlik katsayıları ve puanların aritmetik ortalaması da yaklaşık olarak birbirine eşit olmalıdır ki bu eşitliklerin sağlanması oldukça güçtür.
3. Tüm maddelerin aynı cevap formatına sahip olması gereklidir.
4. Testi alan bir bireyin bir maddeye ilişkin doğru cevaplama olasılığı test uygulanmadan önce bilinemez.
5. KTK modelleri test geliştirme, test eşitleme ve madde/test yanlılığını belirlemede yetersiz kalır, tatmin edici çözümler sunmaz. Madde güçlükleri ve ayırt edicilikleri gruba bağlı olduğu için, benzer testi alan bireylerin yetenekleri karşılaştırılırken

farklı gruplar için tek bir testten elde edilen puanların karşılaştırılamaması gibi sorunlar yaşanmaktadır.

6. KTK ile geliştirilen ölçme araçları, genelde orta yetenek düzeyindeki bireylere hitap eder. Bu durum KTK' nın dezavantajı olarak sayılır (Hambleton vd., 1991).
7. KTK bireyin maddeyi nasıl yanıtladığı hakkında bilgi vermez, dolayısıyla madde merkezli değildir.
8. Bütün yetenek ranjı için tek hata kestirimi yapar.
9. KTK şans başarısını göz ardı eder (Lord ve Novick,1968).

KTK'nın sınırlılıklarının yanı sıra kolay test edilebilir olması açısından avantajı vardır. Böylece, uygulamalarda KTK'nın kullanılması çoğu durum için daha kolay olmaktadır (Hambleton ve Jones, 1993, Baykul 2000).

Crocker ve Algina (2008) maddeler arasındaki farklılıkların belirlenmesinin test hazırlayanlar için önemli olduğunu, farklı yetenek düzeyindeki bireylerin madde bazındaki performans düzeyleri hakkında KTK'nın bilgi vermediği üzerinde durmuştur. Bu durumdan yola çıkarak, test geliştirenlerin bir maddenin bütün fonksiyonlarının görüntüsünü görmelerinin, MTK ile mümkün olduğunu belirtmektedir.

Madde Tepki Kuramı

Örtük özellik, bir testin maddeleri üzerinde gösterilen performansın altında yatan gizil özelliktir. Örtük özellik ve test ile ölçülmeye çalışılan yapı aynı kavramlar değildir. Örtük özellik kavramı, genel yapı kavramı ile eş tutulmamalıdır (Crocker ve Algina, 2008).

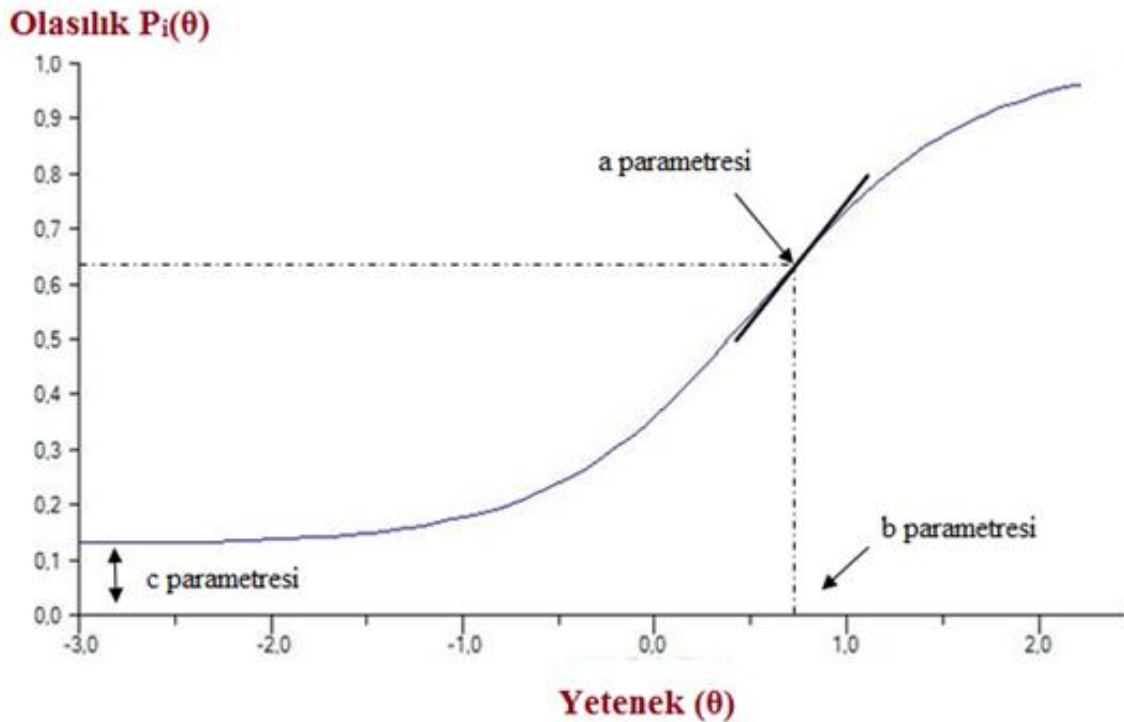
1970'lerden itibaren Madde Tepki Kuramı (MTK) ölçme uzmanları için önemli bir kuram olmuştur. Ancak MTK'nın oluşumu 1930'ların ortasında başlamıştır. MTK konularının başında Tucker'in 1946' da keşfettiği madde karakteristik eğrileri bulunur. Bu erken çalışmalara rağmen, MTK' ya olan ilgi doğru puan teorisiyle artmaya başlar. 1950 lerdeki gelişiminin yavaş olmasının sebebi matematiksel konulardaki karışıklık ve bilgisayar programlarının olmamasıdır (Hambleton ve Swaminathan, 1985).

MTK, bireylerin yetenek düzeylerinin, belirli madde grubundan bağımsız olarak kestirilebileceği sayılısı üzerine kurulmuştur (Hambleton ve Swaminathan, 1985). Bu kuram, KTK'ya göre matematiksel altyapısı daha güçlü olan bir kuramdır. Merkezinde matematik bulunan bu test kuramının temeli Binet, Simon ve Terman'ın 1916 yılında

yaptıkları psikolojik ölçme çalışmalarına dayanmaktadır ve madde temelli bir kuramdır (Ostini ve Nering, 2006).

Ölçme işlemlerinde kullanılan en önemli madde parametreleri, a (ayrıt edicilik) parametresi, b (güçlük) parametresi ve c (şans) parametresidir. MTK modellerinde b parametresi, c parametresinin dahil edilmediği durumlarda, bireyin bir maddeyi 0,5 olasılıkla doğru cevaplandığı noktaya karşılık gelen yetenek (θ) düzeyidir. $c = 0$ olduğunda, b parametresi doğru cevap olasılığı 0,5 olan yetenek düzeyidir. c parametresinin sıfırdan farklı değer aldığı durumda b ; doğru yanıt olma olasılığının c parametresi ve 1,0'ın orta noktasına denk geldiği yetenek düzeyidir (Kolen ve Brennan, 2014).

b parametresinin kuramsal olarak $-3 \leq b \leq +3$ aralığında yer aldığı kabul edilmektedir. a parametresi $\theta = b$ noktasındaki eğrinin eğimidir. a parametresinin $0 \leq a \leq +2$ aralığında yer aldığı kabul edilmektedir. c parametresi ise madde karakteristik eğrisinin asimptotudur ve $0 \leq c \leq 1$ aralığında yer aldığı kabul edilir (Lord ve Stocking, 1997; Baker, 2001; Baker ve Kim, 2004).



Şekil 1. Madde parametrelerinin madde karakteristik eğrisinde gösterimi

Madde tepki kuramı, bireyin doğrudan gözlenemeyen yetenekleri ile doğru cevaplama olasılığı arasındaki ilişkiyi tanımlamaya çalışır (Hambleton ve Swaminathan, 1985).

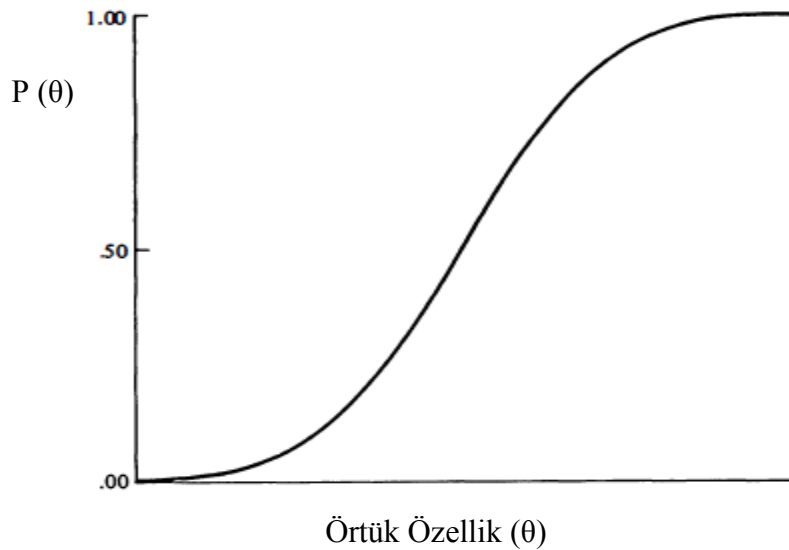
KTK’da yetenek ayırımının yapılamaması, MTK’nın gelişimine olanak sağlamıştır. MTK’nın iki temel varsayımı:

1. Bireyin test maddesindeki performansı; örtük özellik veya yetenekler vasıtasıyla yordanabilir.
2. Bireyin madde performansı ile madde performansının altında yatan özellikler arasındaki ilişki, madde karakteristik eğrisi (MKE) veya fonksiyonu adı verilen ve monoton olarak artan bir fonksiyonla ifade edilebilir (Hambleton vd., 1991).

Madde Karakteristik Eğrisi (MKE)

Hambleton ve Swaminathan (1985)’e göre Madde Tepki kuramının oluşumunda MKE'nin gelişimi büyük rol oynamıştır. MKE madde puanlarının yetenek üzerindeki fonksiyonu olarak tanımlanabilir. MKE, bireyin bir maddede başarılı olma olasılığını, ölçülen yetenekle yordayan bir fonksiyondur (Crocker ve Algina, 2008). Embretson ve Reise (2000), MKE'yi maddenin güçlük, ayırt edicilik ve şans parametresi ile o maddenin doğru cevaplanma olasılığını tüm yetenek düzeyleri için veren eğriler olarak tanımlamıştır.

MTK’da yetenek θ (teta) simgesi ile gösterilir. MTK’ya göre bireyin bir testte gösterdiği performansın ile testte ölçülmeye çalışılan yetenek ilişkilidir. MTK’da bireylerin maddelere verdiği cevaplardan o bireylerin yetenek düzeyleri kestirilmeye çalışılır. Gözlenen ve gözlenemeyen özellikler arasındaki ilişki matematiksel fonksiyonlar ile kurulur (Hambleton ve Swaminathan, 1985).



Şekil 2. Madde karakteristik eğrisi (madde karakteristik fonksiyonu)

Şekil 2'de gösterildiği gibi MKE'de, yatay eksen yetenek düzeylerini, dikey eksen ise doğru cevaplama olasılığını gösterir. Yetenek düzeyi θ ile ve doğru cevaplama olasılığı $P(\theta)$ ile gösterilir. MKE aracılığı ile bir maddenin parametreleri tanımlanabilir. MKE'nin bükülme noktası maddenin güçlüğünü, eğimi maddenin ayırt ediciliğini ve eğrinin y-eksenini kestiği nokta bize maddenin şans parametresini verir (Hambleton ve Swaminathan, 1985).

Madde tepki kuramına göre MKE'nin çoğu zaman -S şeklinde olduğu varsayılır. Eğrinin -S şeklinde olması örtük özellik puanı arttıkça doğru cevaplanma olasılığının artması anlamına gelir. MKE ile örtük özelliklere bağlı olan doğru cevaplama ihtimalinin nasıl değiştiğini görmemiz mümkün olur (Crocker ve Algina, 2008).

Madde Tepki Kuramı Sayıltıları

Madde tepki kuramının tek boyutluluk, yerel bağımsızlık, normallik ve testin hız testi olup olmaması gibi sayıltıları vardır. Aşağıda bu sayıltılar incelenecektir.

Normallik

Test kuramlarındaki önemli kavramlardan birisi normal dağılımdır. Normal dağılım evrendeki sürekli değişkenlere ilişkin teorik bir dağılımdır. Normallik varsayımı 19. yüzyılda Quetelet ve Galton gibi matematikçilerin büyük gruplarda ölçülen özelliklerin frekans dağılımlarının çoğunlukla normal dağılım denilen çan eğrisine benzediğini keşfetmeleri ile ortaya çıkmıştır (Crocker ve Algina, 2008). Bir dağılımın normal olup olmadığı farklı yöntemlerle irdelenebilir. Çarpıklık katsayısı, aritmetik ortalama ve mod gibi betimsel betimsel istatistikler kullanılarak, dağılıma ilişkin grafikler incelenerek normallik sınanabilir. Ayrıca testlerden faydalanılarak da normallik incelenebilmektedir. Çarpıklık katsayısını standart hatasına bölerek elde edilen Z istatistiğinin incelenmesi, grup büyüklüğünün 50'den küçük olması halinde Shapiro-Wilks, büyük olması halinde Kolmogorov Smirnov (K-S) testlerine başvurmak ise diğer yöntemlerdir (Büyüköztürk,2014).

Tek boyutluluk

Tek boyutluluk, testin tek bir örtük özelliği ölçülüyor olması durumudur. Bu varsayımın bazı bilişsel, kişisel ve test performansını etkileyen nedenlerle tam olarak sağlanması güç

olduğundan, testin baskın gelen tek bir faktörü ölçtüğünün gösterilmesi tek boyutluluğun sağlanması için yeterli sayılmaktadır (Hambleton vd.,1991). Kısaca bireyin madde üzerindeki performansını etkileyen sadece bir yetenek türü olması durumudur (Hambleton ve Swaminathan, 1985).

Veri setinin tek boyutluluk varsayımını karşılaması, test performansını etkileyen baskın bir bileşen ya da faktörün olması demektir. Bu baskın bileşen ya da faktör test ile ölçülen yetenek olarak tanımlanır. Ancak bu yetenek değiştirilemez değildir; öğrenme ve unutma gibi bazı faktörlerden etkilenebilir (Hambleton vd., 1991).

Yerel Bağımsızlık

Aynı yetenek düzeyindeki bireylerin farklı maddelere verdiği tepkilerin istatistiksel olarak bağımsız olmasıdır. Bu varsayım bireyin bir maddeye verdiği cevabın diğer bir maddeyi olumlu ya da olumsuz etkilememesi anlamına gelir. Ayrıca, tek boyutluluk sağlandığında, yerel bağımsızlığın da sağlanmış olduğu görülür (Hambleton vd., 1991). Bir maddeyi yanıtlarken gösterilen performansın diğer maddeleri yanıtlarken gösterilen performansı etkilememesi beklenir (Hambleton ve Swaminathan,1985).

Tek boyutluluk varsayımının sağlanması için testte yer alan maddelerin ilişkili olması şartı vardır ancak yerel bağımsızlık varsayımında maddelerin belirli yetenek düzeylerinde bağımsız olması gerekir (Hambleton ve Swaminathan, 1985).

Yerel bağımsızlığı değerlendirmenin bir yolu Q_3 istatistiğinin hesaplanmasıdır (Yen,1984). Q_3 istatistiği yerel bağımsızlığın belirlenmesinde kullanılan bir indekstir. Q_3 istatistiği yetenek parametresi tahminlerinin korelasyonlarının tespitinden sonra hesaplanan artıkların korelasyonudur. Q_3 istatistiğinin hesaplanması için bireylerin her maddedeki performanslarına dayalı olarak θ (gizil yetenek) hesaplanır. Artık değer ise bireylerin beklenen ve gözlenen performans değerleri arasındaki fark alınarak hesaplanmaktadır. Diğer bir deyişle Q_3 istatistiği her bir madde için tüm bireylerin sapma puanlarının korelasyonudur (Yen,1984).

Testin Hız Testi Olup Olmaması

MTK' nın diğer bir varsayımı kullanılan testin hız testi olmamasıdır. Bir testin hız testi olması testin sadece ilgilenilen yeteneği değil, ek olarak performans hızını da ölçmesi

durumdur. Bu durumda testin tek boyutluluk varsayımı ihlal edilmiş olur (Hambleton ve Swaminathan, 1985).

Testin hız testi olup olmadığı, her bir kişinin boş bıraktığı madde sayılarının varyansının, yanlış yaptıkları madde sayılarının varyansına oranlanarak değerlendirilmektedir. Bu oranın sifıra yakın olduğu durumlarda test hız testi olarak çalışmamaktadır (Berberoğlu, Çömlekoğlu ve Yıldırım, 2003).

MTK'ın üstünlüklerinin yanında bazı sınırlılıkları da vardır. MTK matematiksel modele dayandırıldığı için, varsayımlarının sağlanması KTK'ya göre daha zordur (Crocker ve Algina, 2008). MTK ile KTK'nın ortak yönleri şu şekilde sıralanabilir:

1. Her iki kuramda da test maddelerine verilen tepkilerin doğruluğu veya yanlışlığı üzerine odaklanılmaktadır.
2. Normallik varsayımı her iki kuram için de söz konusudur. DeMars (2010), normallik varsayımının MTK modelleri için gerekli olmadığını ancak normal dağılım varsayımının sağlanmadığı takdirde yetenek kestiriminde sorunlara yol açabileceğini belirtmiştir.

Madde Tepki Kuramı Avantajları

MTK'nın değişmezlik özelliği ve tanımlanan madde karakteristik eğrisi, matematiksel altyapısıyla beraber birçok avantaja sahiptir. MTK, birey yetenekleri uygulanan ölçme aracından bağımsızdır. Böylece, farklı gruplar ölçme sonuçları açısından karşılaştırılabilir.

MTK'da hata katsayısı her birey için ayrı ayrı kestirilebilmektedir ve her madde düzeyi için güvenirlik katsayısı kestirilebilir (Hambleton vd., 1991). MTK' da madde parametrelerinin gruptan bağımsız olması, testlerin geçerlik ve güvenirliklerinin belirlenmesinde kararlılık sağlamaktadır (Kelecioğlu, 2001).

Madde Tepki Kuramının Sınırlılıkları

1. Maksimum olabilirlik hesapları geniş veri grupları ve fazla işlem gerektirir. Ayrıca, sonuçların anlamsız ve güvenirliğin düşük çıkma olasılıkları da vardır.
2. Belirlenen MTK modelindeki parametre sayısı arttıkça, hesaplama ve kestirim işlemleri zaman almaktadır.
3. Tekboyutluluk varsayımı sağlanması zor olan bir varsayımdır.

4. MTK uygulamaları, bilgisayar kullanımı gerektirir ve programların ücretli olmasından dolayı mali değeri yüksek olabilir (Hambleton ve Swaminathan, 1985).

Madde Tepki Kuramında Güvenirlik ve Geçerlik

Madde ve test bilgi fonksiyonları KTK'daki güvenirlilik kavramının karşılığı olarak görülmektedir. MTK'da test bilgi fonksiyonunun sahip olunan farklı yetenek düzeylerinin tamamında tek biçimli dağılım vermesi halinde, elde edilen test bilgi fonksiyonlarının ortalaması alınır ve güvenirlilik katsayısı elde edilir (De Ayala, 2009).

MTK'ya göre geçerlik çalışmalarında genellikle faktör analizi ile yapı geçerliği incelenmektedir. Faktör analizinin amacı çok sayıda maddenin daha az sayıda faktörle ifade edilmesidir. Aynı faktörü ölçen maddeler bir araya gelerek gruplar oluşturulur. Her faktör grubuna, içinde bulunan maddelerin özelliğine göre bir faktör adı verilir. Bu faktörlerden her biri ölçmedeki kuramsal yapıyı ifade etmelidir (Karasar, 2013).

Madde Tepki Kuramı Modelleri

Modeller belirli psikolojik özelliği ölçmek, test maddelerine verilen cevapların dağılımını belirlemek için kullanılan matematiksel fonksiyonlardır. Psikolojik ölçmeler için model kavramı, bağımsız değişkenleri bağımlı değişkenlerle ilişkilendiren ve sayısal ifadeleri içeren matematiksel ifadelerdir (Hambleton, 1985; Embretson ve Reise, 2000). Madde Tepki Kuramı modellerinin dayandığı formül aşağıdaki gibidir:

$$P_g(\theta) = \frac{e^x}{1 + e^x}$$

MTK'da her model, cevaplayıcının maddeye cevabı ile yeteneği arasındaki ilişkiyi veren matematiksel bir eşitlikten ve tanımlanmış temel varsayımlardan oluşur (Hambleton ve Swaminathan, 1985).

MTK, literatürde boyutlarına göre tek boyutlu ve çok boyutlu, puan kategorilerine göre iki kategorili, çok kategorili ve parametrelerine göre ise parametrik ve parametrik olmayan olarak çeşitli şekillerde sınıflandırılmaktadır. MTK modelleri, fonksiyonlarındaki parametrelere göre normal ogive ve lojistik olmak üzere iki farklı fonksiyon türünde sınıflandırılabilir. Normal ogive model ve lojistik model bir parametrelili, iki parametrelili ve üç parametrelili olmak üzere üçe ayrılır.

KTK ve MTK’de kullanılan puanlama yöntemleri iki kategorili ve çok kategorili yöntemler olarak üzere iki grupta incelenebilir (Linden ve Hambleton, 1997).

İki kategorili modeller, 1-0 şeklinde puanlama yapılan doğru-yanlış, kısa cevaplı, boşluk doldurma, çoktan seçmeli ve eşleştirmeli testlere uygulanır. Cevapları 1-0 şeklinde kodlanamayan maddeler ise çok kategorili modeller ile incelenir (Embretson ve Reise, 2000; Reckase, 2009). MTK’nın bir, iki ve üç parametrelili lojistik modelleri eğitim alanındaki ölçmelerde sık kullanılan modellerdir. Çalışmada lojistik modeller inceleneceğinden, bu modellerle ilgili ayrıntılı bilgi verilecektir.

Bir Parametrelili Lojistik (1PL) Model

Bir parametrelili lojistik model çok yaygın olarak kullanılan bir MTK modelidir. Bu modelde tüm maddelerin ayırt edicilik parametresi (a) ve şans parametresi (c) aynıdır. Yalnız, testteki maddelerin güçlük parametreleri (b) maddeye göre değişir. 1PL model, kısıtlayıcı varsayımlara dayanır. Çoktan seçmeli maddelerde, düşük yetenekteki cevaplayıcıları tahmin etme olasılığı yoktur (Hambleton vd., 1991). Ancak bilinmesi gereken bir farklılık vardır. Rasch modelinde a parametre değeri 1,00 olarak alınır, bir parametrelili lojistik modelde ise kestirilen bir a değeri, yani ortalama bir değer kullanılır (Embretson ve Reise, 2000).

1PL model için madde karakteristik eğrisi eşitliği şu şekilde yazılabilir:

$$P_j(\theta) = \frac{e^{(\theta-b_j)}}{1 + e^{(\theta-b_j)}}$$

$P_j(\theta)$ = θ yetenek düzeyindeki bireyin j. maddeyi doğru yanıtlama olasılığı

b_j = madde güçlük parametresi

$e = 2,718$

Bir parametrelili lojistik modelde bir maddenin doğru cevaplanma olasılığı 0,5 kabul edildiğinde, θ yetenek düzeyine karşılık gelen değer, madde güçlük indeksi b parametresidir. Maddenin b parametresi değeri arttıkça, bireylerin o maddeye doğru cevap vermeleri için sahip olması gereken yetenek düzeyi de artmaktadır. b parametresinin değeri grubun yetenek ortalaması 0 ve standart sapması 1 olacak şekilde alındığında -3 ile +3 aralığında değer alır. Madde kolaylaştıkça b parametresi -3’e; madde zorlaştıkça +3’ye yaklaşır (Baker, 2001).

İki Parametrelili Lojistik (2PL) Model

Lord tarafından 1952' de Normal Ogive modelin temeli atıldıktan sonra 1957 yılında Birnbaum tarafından 2PL model geliştirilmiştir. Bir parametrelili modeldeki madde güçlüğü (b) parametresine ek olarak, ayırt edicilik (a) parametresi eklenmiştir. İki parametrelili lojistik modelde şans parametreleri aynı (0) kabul edilir, maddelerin güçlükleri ve ayırt edicilikleri değişiklik gösterir (Hambleton ve Swaminathan, 1985; Hambleton vd., 1991).

2PL model için madde karakteristik eğrisi eşitliği şu şekilde yazılabilir:

$$P_j(\theta) = \frac{e^{Da_j(\theta-b_j)}}{1 + e^{Da_j(\theta-b_j)}}$$

$P_j(\theta)$ = θ yetenek düzeyindeki bireyin j. maddeyi doğru yanıtlama olasılığı

b_j = madde güçlük parametresi

a_j = madde ayırt edicilik parametresi

$D = 1.7$ (ölçekleme faktörü)

MTK'da a parametresi madde karakteristik eğrisinin büküm noktasındaki eğimidir. a parametresi $(-\infty, +\infty)$ arasında değişir. Madde karakteristik eğrisinin eğimi arttıkça maddenin ayırtıcılık gücü o kadar artar. Eğri yatay eksene paralelleştikçe, maddenin ayırtıcılık gücü düşer. Düşük ve yüksek yetenekteki bireylerin maddeyi doğru cevaplama olasılıkları birbirine benzer sonuçlar verir (Baker, 2001).

Üç Parametrelili Lojistik (3PL) Model

Bir ve iki parametrelili modeller, cevaplayıcıların doğru cevabı şans eseri bulma olasılıklarını dikkate almadığından eleştiri almıştır. Bu nedenle, şans parametresinin de fonksiyona dâhil edildiği, üç parametrelili model doğmuştur. Üç parametrelili lojistik modelde testteki maddelerin ayırt edicilik, güçlük ve şans parametreleri farklılık gösterir (Baker, 2001).

Şans parametresi ile alt yetenek düzeylerindeki bireylerin maddeyi şans başarısı ile doğru cevaplama olasılığı kestirilebilir. 3PL model bu yönüyle 1PL ve 2PL modellerden ayrılmaktadır. c parametresi, madde karakteristik eğrisinin alt asimptotudur (Hambleton ve Swaminathan, 1985).

3PL model için madde karakteristik eğrisi eşitliği şu şekilde gösterilebilir:

$$P_j(\theta) = c_j + (1 - c_j) \cdot \frac{e^{Da_j(\theta - b_j)}}{1 + e^{Da_j(\theta - b_j)}}$$

$P_j(\theta)$ = θ yetenek düzeyindeki bireyin j. maddeyi doğru yanıtlama olasılığı

b_j = madde güçlük parametresi

a_j = madde ayırt edicilik parametresi

c_j = Şans parametresi

$D = 1,7$ (ölçekleme faktörü)

c parametresinin değeri yetenek düzeyinden etkilenmediğinden, en yüksek ve en düşük yetenek düzeyleri için şans eseri doğru cevap verme olasılığı eşit kabul edilir (Baker, 2001). Bu model diğer modellere göre daha karmaşık yapıdadır. Ayrıca yanıtın şans eseri doğru olma olasılığı madde ayırt edicilik gücünü düşürmektedir (Harvey ve Hammer, 1999).

Madde Tepki Kuramı ile Klasik Test Kuramı Arasındaki Temel Farklılıklar

1. MTK'ya göre geliştirilen testlerden elde edilen yetenek ölçüleri, bireylere uygulanan testlerden bağımsızdır. Bu özellik sayesinde, test puanları eşitlendiğinde, bireylerin yetenekleri gruptan bağımsız olarak karşılaştırılabilmektedir. KTK' ya göre ise, birey puanları uygulanan teste bağımlıdır. Bu sebeple, kolay ve zor testlerden aynı puanı alan farklı öğrenciler karşılaştırılamaz.
2. KTK'da elde edilen madde istatistikleri, testin uygulandığı grubun özelliklerine bağlıdır. MTK'da ise madde özellikleri matematiksel modeller esas alınarak hesaplandığından, testin uygulandığı gruba bağlı değildir.
3. KTK'nın sayıtlarını karşılamak oldukça kolaydır; ancak MTK sayıtlarını karşılamak oldukça güçtür.
4. KTK'da madde ve birey istatistiklerinde değişmezlik özelliği yoktur, MTK'da ise olduğu iddia edilmektedir.
5. KTK'da test istatistiği olarak p ve r istatistikleri; MTK'da ise modele göre a, b ve c parametreleri hesaplanır.

6. KTK modelleri doğrusal modellerdir, MTK modelleri ise doğrusal olmayan modeller sınıfına girer (Hambleton ve Swaminathan, 1985; Hambleton vd., 1991, Klein, 2005).

MTK Çalışmalarında Kullanılan Yaygın Programlar

BILOG MG Programı

BILOG-MG programı madde ve katılımcı sınırlaması olmaksızın geniş kapsamlı uygulamalarda kullanılabilen pratik bir programdır. Tek bir programlama süresinde alt testlerin ve alt ölçeklerin madde analizi ve puanlamasını yapabilmektedir. BILOG MG program çıktıları, maddeleri seçme ya da test puanlarının raporlarını hazırlama amacıyla metin dosyaları şeklinde çıktı vermektedir. Program ikili puanlamalı cevaplar için 1, 2 ve 3 parametrelili lojistik modeller ile inceleme sağlamaktadır. Ayrıca eşit olmayan grupları eşitleme ve test formlarını dikey eşitlemeye yardım eder (Du Toit, 2003).

Testlerdeki madde sayısının 20'den fazla olduğu durumlarda, katılımcılar θ 'nın tahmini değerine dayanarak aralıklara atanabilirler. Daha sonra belirlenen maddeye doğru bir şekilde cevap vermiş olan her aralıktaki cevaplayıcının madde skorları elde edilebilir.

BILOG MG programında, olabilirlik oranı olan χ^2 test istatistiği verilmekte ve uygun olan modelin seçilmesinde kullanılmak üzere sunulmaktadır. Bu yöntem doğru ve yanlış cevapların sıklığını kıyaslamak için de kullanılabilir. Du Toit (2003), BILOG MG programının arka planında kullanılan formülleri nihai biçimde belirtmiştir.

BILOG MG programı çıktıları 3 farklı output olarak vermektedir. Phase 1'de veri bilgileri ve KTK'ye göre kestirilen madde istatistikleri bulunur. Phase 2'de MTK kalibrasyon sonuçları, ve Phase 3'de birey kestirimleri bulunur. Phase 3'de gösterilen madde parametresi kestirimleri, Phase 2'den farklıdır. Bunun sebebi Phase 3'de ölçeklendirilmiş kestirim yapılmasıdır. Bu sebeple madde parametreleri Phase 2'den temin edilir (De Ayala, 2009).

MULTILOG 7.03 Programı

MULTILOG 7.03 programı çoktan seçmeli testler veya likert türü anketler gibi çoklu maddelerin olduğu analizler için uygun bir programdır. Uygun modelleri kullanan ve bu tür verilerle model uyumunu sağlayabilen geniş çaplı tek programdır. 1PL, 2PL ve 3PL

modeller ile çoklu nominal kategoriler, kademeli değerlendirme ölçeği, kısmi (parçalı) kredi modeli, çoktan seçmeli model ve sıkıştırılmış değişken modellerle çalışabilmektedir (Du Toit, 2003).

MULTILOG 7.03 programı parametre kestiriminde lojistik metrik (ölçüm) kullanmaktadır. Ayrıca MULTILOG 7.03, expectation-maximization (EM) algoritmasına dayanır (Childs ve Chen, 1999).

İkili puanlanan veriler için lojistik modeller yaygın olarak kullanılmaktadır. MULTILOG 7.03 programı Maksimum Olasılık (ML), Marjinal Maksimum Olasılık (MML) kestirim yöntemlerini kullanır. Modelin uyum derecesi χ^2 indeksleri ile incelenir (Du Toit, 2003).

R Studio Yazılımı ve paketleri

R yazılımı, istatistiksel hesaplamalar ve grafik çizimleri için tasarlanan bir sistemdir. R yazılımı ile birlikte sağlanan 25 kadar paket (standart ve önerilen paketler olarak adlandırılan) vardır ve daha birçoğu mevcuttur (Du Toit, 2003). Veri manipülasyonu ve hesaplamaların yanı sıra veri işleme ve depolamada da etkilidir (Venables, Smith ve R Core Team, 2015).

R yazılımı Amerika ve Avrupa'da istatistikçiler tarafından en çok kullanılan programlardan biridir. R, ticari amaç gütmeksizin model ve yöntem geliştirmeye açık bir programdır. Yazılımcılar yeni model ve yönteme ait kodları R programında yazabilmekte ve bu kodlar diğer araştırmacılar tarafından kullanılıp geliştirilebilmektedir (İlk, 2011).

Bu araştırmada R yazılımında parametre kestirimlerini Marginal Maximum Likelihood (MML) kestirim yöntemine göre yapan 'sirt' ve 'ltm' paketleri kullanılacaktır. 1PL modele göre yapılan kestirimler 'ltm' (Rizopoulos, 2013) paketi ile, 2PL ve 3PL modele göre yapılan kestirimler ise 'sirt' (Robitzsch, 2015) paketi ile yapılacaktır. Bu paketler dichotomous ya da polytomous verilerde MTK'ya göre parametre kestirimi yapan bir paketlerdir.

TIMSS 2011 Türkiye 4.sınıf matematik alt testinin psikometrik özelliklerinin KTK ve MTK'ya göre karşılaştırılmasının, farklı kuramlara dayalı yöntemlerle belirlenen değerlerin uyumunun belirlenmesine katkı sağlayacağı düşünülmektedir. Ayrıca alanyazında bu programların karşılaştırıldığı bir çalışma bulunmamaktadır. Bu yüzden BILOG MG, MULTILOG 7.03 ve R programlarının madde parametrelerinin karşılaştırılmasının alanyazına katkı sağlaması beklenmektedir. Belli özellikleri ve

kullandığı formüller verilen programlar aracılığı ile kestirimler yapılarak, programlardan elde edilen sonuçların ne ölçüde benzerlik ve farklılık gösterdiği incelenecektir.

Araştırmanın Amacı

Bu çalışmada TIMSS Türkiye 2011 4. sınıf matematik alt testi maddelerinin KTK'ya göre hesaplanan madde istatistikleri ile BILOG MG, MULTILOG 7.03 ve R programlarıyla MTK'nın 1PL, 2PL ve 3PL modellerine göre kestirilen madde parametrelerini karşılaştırmalı olarak incelemek, kullanılan programlara göre model veri uyumunun farklılaşıp farklılaşmadığını belirlemek, kullanılan programların kestirdiği parametrelerin standart hatalarını ve güvenilirlik katsayılarını karşılaştırmak amaçlanmıştır.

Problem İfadesi

TIMSS 2011 Türkiye 4.sınıf matematik alt testinin KTK'ya göre hesaplanan madde istatistikleriyle, MTK'ya göre BILOG MG, R yazılımı 'ltm' ve 'sirt' paketleri ve MULTILOG 7.03 programları ile kestirilen madde parametreleri arasındaki ilişkiler ve BILOG MG, MULTILOG 7.03 ve R programlarına göre, standart hatalar, güvenilirlik katsayıları ve model veri uyumu nasıldır?

Araştırma Soruları

1. TIMSS 2011 Türkiye 4.sınıf matematik testindeki maddelerin MTK'ya göre kestirilen b parametreleri ve KTK'ya göre hesaplanan madde güçlük katsayıları arasındaki ilişkinin düzeyi nedir?

1.1. R programı ile TIMSS 2011 Türkiye 4.sınıf matematik alt testindeki her bir maddeye ait MTK'nın 1PL, 2PL ve 3PL modeline göre kestirilen b parametreleri ve bu parametreler ile KTK'ya göre hesaplanan madde güçlük (p) katsayıları arasındaki ilişkinin düzeyi nedir?

1.2. MULTILOG 7.03 programı ile TIMSS 2011 Türkiye 4.sınıf matematik alt testindeki her bir maddeye ait MTK'nın 1PL, 2PL ve 3PL modeline göre kestirilen b parametreleri ile KTK'ya göre hesaplanan madde güçlük (p) katsayıları arasındaki ilişkinin düzeyi nedir?

1.3. BILOG MG programı ile TIMSS 2011 Türkiye 4.sınıf matematik alt testindeki her bir maddeye ait MTK'nın 1PL, 2PL ve 3PL modeline göre kestirilen b parametreleri ve bu parametreler ile KTK'ya göre hesaplanan madde güçlük (p) katsayıları arasındaki ilişkinin düzeyi nedir?

2. TIMSS 2011 Türkiye 4.sınıf matematik alt testindeki her bir maddenin MTK'ya göre elde edilen a parametresi ile KTK'ya göre elde edilen madde ayırt edicilik katsayısı (r) arasındaki ilişkinin düzeyi nedir?

2.1. BILOG MG programı ile TIMSS 2011 Türkiye 4.sınıf matematik alt testindeki her bir maddeye ait MTK'nın 2PL ve 3PL modeline göre kestirilen a parametreleri ve bu parametreler ile KTK'ya göre hesaplanan madde ayırt edicilik (r) katsayıları arasındaki ilişkinin düzeyi nedir?

2.2. R programı ile TIMSS 2011 Türkiye 4.sınıf matematik alt testindeki her bir maddeye ait MTK'nın 2PL ve 3PL modeline göre kestirilen a parametreleri ve bu parametreler ile KTK'ya göre hesaplanan madde ayırt edicilik (r) katsayıları arasındaki ilişkinin düzeyi nedir?

2.3. MULTILOG 7.03 programı ile TIMSS 2011 Türkiye 4.sınıf matematik alt testindeki her bir maddeye ait MTK'nın 2PL ve 3PL modeline göre kestirilen a parametreleri ve bu parametreler ile KTK'ya göre hesaplanan madde ayırt edicilik (r) katsayıları arasındaki ilişkinin düzeyi nedir?

3. TIMSS 2011 Türkiye 4.sınıf matematik alt testi verilerinin BILOG-MG, R ve MULTILOG 7.03 programlarına göre hesaplanan 1PL, 2PL ve 3PL modeller ile uyumu ne düzeydedir?

4. TIMSS 2011 Türkiye 4.sınıf matematik testinin KTK'ya göre hesaplanan Cronbach- α güvenirlik katsayısı ve MTK'ya göre kestirilen güvenirlik katsayıları nedir ve aralarında anlamlı farklılık var mıdır?

4.1. KTK'ya göre hesaplanan Cronbach- α güvenirlik katsayısı ve MTK'ya göre kestirilen marjinal ve görgül güvenirlik katsayıları nedir?

4.2. BILOG MG, R ve MULTILOG 7.03 programlarında 1PL, 2PL ve 3PL model için kestirilen görgül ve marjinal güvenirlik katsayıları arasında anlamlı fark var mıdır?

5. TIMSS 2011 Türkiye 4.sınıf matematik alt testi verilerinin R, BILOG MG, MULTILOG 7.03 programlarıyla 1PL, 2PL ve 3PL modellere göre yapılan analizler sonucunda elde edilen a,b ve c parametrelerinin standart hatalar arasında manidar farklılık var mıdır?

Araştırmanın Önemi

Bir teste ait psikometrik özelliklerin farklı kuramlara göre karşılaştırılması konusunda çok sayıda çalışma mevcuttur. Psikometrik özelliklerin karşılaştırılması kuramların uyumlarını belirleme açısından önem taşımaktadır. Bu araştırmada MTK parametrelerini kestirebilen BILOG MG, MULTILOG 7.03 ve R programları kullanılarak yapılan madde parametresi kestirimlerinin, KTK'ya göre elde edilen madde istatistiklerinden farklılaşıp farklılaşmadığı incelenmiştir. Çalışma parametre kestiriminde farklı programlar arasındaki standart hataların karşılaştırılmasına olanak sağlamaktadır. Araştırmada, MTK'nın 1PL, 2PL ve 3PL modellerine göre model veri uyumları incelenerek programların model veri uyumu analizlerinde farklı sonuçlar verip vermediği incelenmiştir.

Daha önce alan yazında bu programların parametre kestirimlerinin ilişkisinin incelendiği bir çalışma bulunmamaktadır. Ayrıca R yazılımı paketlerine göre elde edilen parametre, standart hata ve güvenilirlik hesaplamaları daha önce herhangi bir program ile karşılaştırılmadığından çalışmanın alan yazına katkı sağlayacağı düşünülmektedir.

Sınırlılıklar

1. MTK'ya dayalı farklı programlarla yapılmak istenen analizler BILOG MG, MULTILOG 7.03 ve R programları ile sınırlandırılmıştır.
2. MULTILOG 7.03 programının 3PL modelde yapılan hata kestirimlerini vermemesinden dolayı, MULTILOG 7.03'e göre kestirilen 3PL modelin hata kestirimleri KTK ve MTK karşılaştırılmasına dâhil edilmemiştir.
3. R programında yapılan madde analizleri 'lrm (1PLM için) ve sirt' paketleri ile sınırlıdır.
4. Araştırmanın örneklemi TIMSS 2011 Türkiye 4.sınıf uygulamasından seçilen 11. matematik kitapçığını alan 543 kişi ile sınırlıdır.
5. Araştırma verisi ikili (0-1) puanlanan maddeler ile sınırlıdır.

Tanımlar

İki seçenekli (dichotomous) madde: Soru cevabının evet-hayır, katılıyorum-katılmıyorum, doğru-yanlış gibi iki seçenekli olduğu soru yapısıdır (DeMars, 2010).

a Parametresi: MTK'ya göre hesaplanan madde ayırt edicilik parametresidir. Madde karakteristik eğrisinin eğimine denk gelir (DeMars, 2010).

b Parametresi: MTK'ya göre hesaplanan güçlük parametresidir. Bir maddenin belli bir olasılıkla doğru yanıtlanması için gerekli yetenek düzeyini gösterir (DeMars, 2010).

c Parametresi: MTK'ya göre hesaplanan şans parametresidir. Düşük yetenek düzeyindeki bireylerin bir maddeyi doğru yanıtlama olasılığını temsil eder. Madde karakteristik eğrisinin y eksenini kestiği noktadır (DeMars, 2010).

Ayırt Edicilik Katsayısı: Madde puanları ile test puanları arasındaki korelasyondur (Baykul, 2015).

Madde Güçlüğü Katsayısı: Testin uygulandığı grupta, o maddeye doğru cevap veren öğrencilerin yüzdesi olarak ifade edilir (DeMars, 2010).





BÖLÜM II

İLGİLİ ARAŞTIRMALAR

Bu bölümde KTK'ya göre kestirilen madde istatistikleri ve MTK'ya göre kestirilen madde parametrelerinin karşılaştırıldığı çalışmalar ve MTK kestirimi yapan farklı programların karşılaştırıldığı çeşitli araştırmalara yer verilmiştir. Alanyazında MTK ile KTK'nın karşılaştırıldığı çok sayıda çalışma mevcuttur. Çalışmalar çoğunlukla madde ve test istatistiklerinin kestirilmesi ve test geliştirilmesine ilişkindir.

Bu bölümde KTK ve MTK'ya göre kestirim karşılaştırması yapılan, model veri uyumu incelenen ve güvenirlik katsayısı karşılaştırması yapılan araştırmalara yer verilmiştir.

KTK'ya Göre Kestirilen Madde İstatistikleri ve MTK'ya Göre Kestirilen Madde Parametrelerinin Karşılaştırıldığı Çalışmalar

Türkiyede MTK üzerine yürütülen ilk araştırma Baykul'un (1979) çalışmasıdır. Baykul, araştırmasında KTK ile üç parametrelili MTK'ya göre geliştirdiği matematik testlerinin güvenirlik ve geçerliklerini karşılaştırmasını yapmıştır. Çalışma sonucunda MTK'ya göre kestirilen puanlardan elde edilen KR-20 güvenirlik katsayısının KTK'ya göre kestirilen puanlardan hesaplanan KR-20 güvenirlik katsayısından yüksek olduğunu sonucuna ulaşmıştır.

Ndalichako ve Rogers (1997) çalışmalarında son sınıflara uygulanan okuduğunu anlama çoktan seçmeli testinden elde edilen yetenek kestirimlerini KTK, MTK ve sonlu puanlama (finite state score) teorisine göre karşılaştırmıştır. Araştırmada klasik test skorları ITEMAN programı ile elde edilmiş olup, 1PL modele göre yetenek kestirimleri RASCAL programı aracılığıyla elde edilmiştir. 2PL ve 3PL modeller için ise ASCAL programı ile yetenek kestirimleri yapılmıştır. Analizler sonucunda KTK'den elde edilen yetenek parametreleri ile 1PL, 2PL ve 3PL modellerden kestirilen parametreler arasında

mükemmele yakın ($<0,964$) korelasyon bulunmuştur. Modellerin yetenek kestirimleri arasındaki korelasyonlara bakılmış ve 1PL ile 2PL modeller ($0,77$) ve 2PL ile 3PL modellerin ($0,79$) yakın uyumlu olduğu tespit edilmiştir. KTK ile MTK modelleri karşılaştırıldığında ise KTK'nın 3PL model ile sağladığı uyumun 2PL modele ve 1PL modele göre daha yüksek olduğu gözlenmiştir.

Berberoğlu (1988), Öğrenci Seçme Sınavı (ÖSS)'ndaki alt testlerden elde edilen ham puanlar ile Rasch modelinden kestirilen yetenek parametrelerini karşılaştırmıştır. Araştırmaya göre yetenek ölçülerinin farklı grupların karşılaştırılmasında kullanılabileceği görülmüştür. Rasch modelinin güvenilirliğe kesresreatkılarını incelemek için Rasch yeterli ölçülerinin standart hatası ile ölçmenin standart hatalarını karşılaştırmıştır. Çalışma sonuçlarına göre Rasch modelini kullanarak elde edilen standart hata, ölçmenin standart hatasına göre daha küçük bulunmuştur ve Rasch modeline göre elde edilen kestirimlerin güvenilirlik katsayıları, KTK'ya göre elde edilen güvenilirlik katsayılarından daha yüksek bulunmuştur.

Fan (1998) çalışmasında Teksas Akademik Beceri Değerlendirme Ölçeğinin matematik ve Türkçe alt testlerinden elde edilen puanları cinsiyet ve yetenek düzeyleri için KTK'ye göre ve MTK'nın 1PL, 2PL ve 3PL'ye göre hesaplanan madde istatistiklerini karşılaştırmıştır. Analizler sırasında ilk olarak KTK'ya göre yapılan yetenek tahminleriyle ile MTK'nın üç modeline göre elde edilen yetenek tahminlerinin ortalama korelasyon katsayılarını karşılaştırmıştır. Korelasyon katsayıları $0,9$ 'dan büyük olup, manidar ve oldukça yüksek ilişki gözlenmiştir. Daha sonra KTK'ya dayalı kestirilen madde güçlük indeksi ile 1PL, 2PL ve 3PL modellerine göre kestirilen b parametresinin korelasyonu incelenmiştir. Korelasyon katsayıları $0,9$ 'dan yüksek ve manidardır. Ayrıca KTK'ya dayalı olarak kestirilen madde ayırtıcılık katsayısı ile 2PL ve 3PL modellere göre kestirilen a parametresinin korelasyonu incelenmiştir. Sonuç olarak iki kuramın madde ayırt edicilik korelasyonlarında düşük ($0,192$) ve yüksek ($0,950$) değerler elde edilmiştir. Araştırma sonucunda KTK'ye göre elde edilen puanlarla, bir, iki ve üç parametrelili MTK modellerine göre kestirilen yetenek parametreleri arasında yüksek ve manidar düzeyde ilişki saptanmıştır olup madde güçlük parametrelerinin farklı örneklemelerde benzer sonuç verdiği ancak madde ayırt edicilik parametresinin farklı örneklemelerde farklılaştığı sonucuna ulaşılmıştır.

Stage (1998), çalışmasında İsveç Akademik Yetenek Testi'nin ön ve asıl uygulamasının kelime alt testi verilerinin KTK ve MTK'ya göre kestirilen madde parametrelerini kendi içlerinde karşılaştırmıştır. Araştırmacı, analizler sırasında ön uygulama ve asıl uygulamadan elde edilen madde parametrelerinin kararlılığını incelemek amacıyla, uygulamaların ortak maddelerini kullanmıştır. KTK ve MTK'ya göre yapılan analizler sonucunda, ön ve asıl uygulama için her madde için kestirilen parametreler arasındaki korelasyonun yüksek korelasyonlu olduğu görülmüştür. Tüm maddeler MTK'nın 3PL modeli ile uyumlu çıkmıştır. Araştırmacı büyük örneklemelerde uygulanacak testler hazırlandığı takdirde, madde analizinde KTK veya MTK çerçevesinde yapılmasının herhangi bir farklılık yaratmayacağını belirtmiştir.

Anderson (1999) çalışmasında British Columbia Diploma Sınavı'nın matematik alt testi verilerini KTK ve 3PL MTK modele göre karşılaştırmıştır. Çalışmanın amacı MTK kullanımı ile ortaya çıkan karmaşıklık ve zorluk karşısında önemli bir faydası olup olmadığını tespit etmektir. Model uyumunu en çok 3PL model sağladığı için bu modelle devam edilmiştir. Analiz sırasında test iki yarıya bölünmüş olup, aynı yarılar için KTK ve MTK'dan elde edilen ortalama puanlar hemen hemen aynıdır. Farklı iki alt test için KTK'ye göre elde edilen test puanları ile MTK'ye göre kestirilen yetenek parametreleri arasında 0,9'dan yüksek korelasyonlar elde edilmiştir.

Gelbal (1994), Rasch modeline göre elde edilen madde ve yetenek parametreleri ile KTK'ya göre elde edilen parametrelerin benzerliklerini incelemiştir. Araştırma verileri ilköğretim 5.sınıflar için hazırlanan Türkçe ve matematik testlerinden elde edilmiştir. Verilerin Rasch modelinin varsayımlarını sağladığı tespit edildikten sonra, testlerden Rasch modeline göre elde edilen b parametreleri ile KTK'ya göre elde edilen madde güçlük parametrelerinin birim normal değerlerinin korelasyonuna bakılmıştır. Analizler sonucu parametreler arasında oldukça yüksek korelasyongözlenmiştir. Ayrıca Türkçe testinin değişik yöntemlerle kestirilen yetenek ölçüleri arasındaki korelasyon, matematik testinden elde edilen korelasyona göre yüksektir. Bu durum, matematik testinin normallik ve tek boyutluluk sayılıtlarının daha az karşılanması olarak yorumlanmıştır.

Kelecioğlu (2001) çalışmasında Milli Eğitim Bakanlığının Anadolu Lisesi Giriş Sınavının Türkçe ve Matematik alt testlerini kullanarak KTK'ya göre hesaplanan madde parametrelerini, geçiş formülleri kullanarak MTK parametreleri elde etmiştir. Analizler sırasında parametreler boş bırakılan cevapların yanlış ve atlanmış olarak ayrı ayrı ele

alınıp, bu iki hesaplamanın MTK'ya göre hesaplanan b ve a parametrelerinin, KTK'ya göre hesaplanan p ve r katsayılarından etkileyip etkilemediği incelenmiştir. MTK analizleri 2PL modele göre yapılmış olup, BILOG programından faydalanılmıştır. Analizlere göre puanlama farkı gözetmeksizin, b parametreleri arasındaki korelasyonlar a parametreleri arasındaki korelasyonlardan daha yüksek çıkmıştır. Test bazında ise Türkçe testinden elde edilen korelasyonlar Matematik testinden elde edilen korelasyonlardan daha yüksek çıkmıştır.

Çıkrıkçı Demirtaşlı (2002)'nin çalışmasında, Raven Standart İlerlemeli Matrisler Testini KTK ve MTK'ya göre incelenmiştir. İki kurama göre kestirilen madde istatistikleri ve madde parametrelerinin karşılaştırılabilirlikleri ile düşük ve yüksek yetenek grupları için parametrelerin değişmezlik özelliği gösterip göstermediği incelenmiştir. Medel veri uyumunun en çok 3PL model ile sağlandığı çalışmada, madde analizleri sonucunda kestirilen madde güçlük değerlerinin MTK ve KTK'ya göre karşılaştırılabilir olduğu görülmüştür. KTK'ya göre kestirilen madde güçlük katsayıları ile MTK'nın 1PL, 2PL ve 3PL modeline göre kestirilen b parametreleri arasında 0,89 ve daha yüksek korelasyonlar elde edilmiştir. KTK ile MTK'ya göre kestirilen ayırt edicilik güçleri ise daha az karşılaştırılabilir bulunmuştur. İki kurama göre kestirilen ayırt edicilik güçleri arasında madde güçlük korelasyonlarına göre daha düşük korelasyon katsayıları elde edilmiştir. 3PL modele göre kestirilen ayırt edicilik parametresi ile KTK'ya göre kestirilen ayırt edicilik katsayısı arasındaki korelasyon ise 0,03 olarak bulunmuştur.

Doğan ve Tezbaşaran (2003) MTK'nın KTK'ya üstünlüğü olarak kabul gören madde parametrelerinin değişmezlik özelliğini sağlayıp sağlamadığını araştırmışlardır. Araştırmada Milli Eğitim Bakanlığı Orta Öğretim Kurumları Öğrenci Seçme ve Yerleştirme Sınavı verilerinden seçkisiz, sağa çarpık, sola çarpık, basık ve normal dağılım gösteren beş farklı örneklem seçilmiştir. Verilerin analizi sırasında ilk olarak evren ve belirlenen örneklem için betimsel istatistikler hesaplanmıştır. Daha sonra evrendeki verilerin MTK varsayımlarını karşılayıp karşılamadığı incelenip, varsayımların sağlandığı sonucuna ulaşılmıştır. Araştırmacılar son olarak, madde istatistik ve parametrelerinin hesap ve kestirimlerini yaparak, madde istatistikleri, madde parametreleri ve evren değeri arasındaki ilişkiyi incelemiştir. Analizler sonucunda madde istatistikleri ve parametrelerinin örneklem değişikliğinden etkilendiği görülmüştür. Dolayısıyla, kuramlar parametre değişmezliği açısından bir diğerine üstünlük sağlamamaktadır. Madde güçlük

parametreleri arasındaki korelasyon, madde ayıricılık parametrelerinin korelasyonlarından daha yüksek bulunmuştur. Korelasyon manidarlıklarına bakarak, MTK'nın a parametrelerinde değişmezliği sağlayamadığı korelasyonlararası farklara bakarak ise iki kuramın da değişmezliği sağlamadığı sonucuna ulaşılmıştır.

Courville (2005), İngilizce, matematik, okuma ve fen çoktan seçmeli alt testlerinden oluşan ACT Değerlendirme Testinden elde edilen veriler üzerinde yaptığı çalışmasında, madde ve birey istatistiklerini karşılaştırılabilirliğini ve parametrelerin farklı örneklemelerde değişmezliğini KTK ve MTK'ya göre araştırmıştır. Bu amaçla, yüz ve bin kişilik farklı örneklem büyüklüklerinde KTK ve MTK'ya göre elde ettiği madde ve test istatistikleri ile yetenek parametrelerini karşılaştırmıştır. KTK ve MTK'ya göre hesaplanan birey yetenekleri korelasyon katsayıları, 0.90'dan yüksek bulunmuştur. KTK'ya göre hesaplanan madde güçlük katsayıları ile 2PL ve 3PL modeller için hesaplanan madde güçlük parametreleri arasında genellikle 0,95'ten yüksek korelasyonlar elde edilmiştir. Sonuç olarak, iki kurama göre elde edilen madde ve birey istatistikleri benzerlik göstermektedir ve karşılaştırılabilir. KTK ile 2PL modele ait madde ayırt edicilik değerleri oldukça yüksek korelasyonlar vermektedir. Ancak 3PL model için daha düşük korelasyonlar elde edilmiştir.

MacDonald ve Paunonen (2002) yaptıkları araştırmada Monte Carlo simülasyon tekniğiyle üretilen verilerin madde ve birey kestirimlerini Madde Tepki Kuramı ve Klasik Test Kuramına göre karşılaştırmıştır. Analizler sırasında madde istatistikleri, madde parametreleri ve birey parametreleri kuramlara göre elde edilmiştir. Çalışmada 1000'er kişilik cevaplayıcı grubu için 100 adet simülasyon testi üzerinden parametreler kestirilmiş ve bu parametrelerin karşılaştırılabilirlikleri, değişmezliklikleri ve doğruluk dereceleri incelenmiştir. MTK'ya göre kestirim yapılırken kestirim yöntemi MML olan PARSCALE programı kullanılmıştır. Yapılan analizler sonucunda MTK ve KTK'ya göre elde edilen madde güçlük kestirimleri ortalama 0,964 korelasyonla yüksek derecede karşılaştırılabilir çıkmıştır. Birey parametreleri ortalama 0,985 korelasyon ile yüksek derecede karşılaştırılabilir. Ancak madde ayırt edicilik kestirimleri ortalama 0,676 korelasyon ile iki kuram için daha düşük karşılaştırılabilir nitelikte çıkmıştır. KTK'dan elde edilen katsayıların ve MTK'dan elde edilen madde parametrelerin değişmezlik özelliğine sahip olduğu gözlenmiştir. Ayrıca kestirimlerin doğruluklarının incelenmesi sonucunda birey parametrelerinin doğru kestirildiği, güçlük kestirimlerinin ise MTK'ya göre doğru,

KTK'ya göre belli deneysel kořullar altında doęru olduęu sonucuna ulařılmıştır. Arařtırma sonucunda MTK'nın birey ve madde kestirimlerinde üstünlük saęladıęı belirtilmiştir.

Nartgün (2002) çalışmasında, biri likert tipi dięeri metrik olarak düzenlenmiş ölçek uygulamasının madde istatistiklerinin iliřkisini ve farklı örneklemlerden kestirilen madde istatistiklerinin deęişmezlięini KTK ve MTK'ya göre karşılařtırmıştır. Analizler sırasında verilerin MTK varsayımlarını saęladıęı görülmüş olup, MTK analizleri için MULTILOG 7.03 programı kullanılmıştır. İki kurama göre kestirilen madde ayırt edicilik istatistikleri arasındaki korelasyonlar sonucunda yüksek derecede iliřki görülmüřtür. Maddelerin güçlük istatistikleri açısından karşılařtırılması sonucunda ise negatif yönde yüksek iliřki saptanmıştır. Maddelerin ölçtükleri tutum düzeylerinin kuramlara göre korelasyonları yüksek çıkmış olup, uygulanan iki ölçeęin de ölçüt grup geçerlięine sahip olduęu ve ölçme sonuçlarının güvenilir olduęu gözlenmiştir.

Özdemir (2004) çoktan seçmeli testlerin KTK ve MTK'ya göre hesaplanan psikometrik özelliklerini iki kategorili ve aęırlıklı puanlama yöntemlerine göre karşılařtırmıştır. Arařtırmada iki kategorili ve aęırlıklı puanlama yöntemlerinin güvenilirlik ve geçerlikleri iki kurama göre karşılařtırılmıştır. Veriler ilköęretim seviyesinde okuyan öęrencilere uygulanan çokta seçmeli testten elde edilmiştir. Analiz ařamasında testin MTK varsayımlarını saęladıęı görülmüřtür. İkili puanlama için KTK'ya göre grubun geneline iliřkin elde edilen güvenilirlik katsayısı 0,49 ve MTK'ya göre elde edilen güvenilirlik katsayısı 0,78'dir. Aęırlıklı puanlamaya dayalı elde edilen güvenilirlik katsayıları incelendięinde ise MTK için 0,68 ve KTK için 0,54 deęerleri bulunmuřtur. Her iki puanlama için de MTK daha güvenilir sonuçlar vermiştir. Ayrıca ölçüt grup geçerlięi incelendięinde KTK için sınıfların test puanları ortalamaları, MTK için ise sınıfların yetenek ortalamaları arasında manidar farklılık olduęu görülmüřtür.

Tomkowicz ve Rogers (2005) arařtırmalarında KTK, MTK'nın 1PL, 2PL, 3PL modelleri ve nominal tepki modelinin yetenek kestirimlerini karşılařtırmışlardır. Arařtırmacılar veri seti olarak on ikinci sınıfa uygulanan çoktan seçmeli Sosyal Bilgiler ve Kimya testinden alınan verileri kullanmıştır. Analizler sırasında öęrenciler, yüksek, orta ve düşük yetenek gruplarına ayrılarak, verilerin analizinde klasik yöntem için LERTAB programı, tepki modelleri için ise MULTILOG 7.03 programı kullanılmıştır. Madde analizleri sonucunda tüm yetenek gruplarında her iki alt test için, KTK ile MTK'nın 1PL modeli ve MTK'nın 2PL ie 3PL modellerine göre yapılan kestirimler benzerlik göstermiştir. Sosyal Bilgiler testindeki farklılaşma Kimya testindeki farklılaşmaya göre daha azdır. Arařtırmacılar bu

beş modelin iki gruba ayrılabilceğini vurgulamışlardır. Birinci grup KTK ve 1PL model, ikinci grup ise 2PL, 3PL ve nominal tepki modelidir.

Kan (2006) çalışmasında Orta Öğretim Kurumları Sınavı (OKS) Türkçe alt testine ait maddelerin KTK ve MTK'ya göre kestirimini yaparak madde istatistiklerini incelemiştir. İki kurama göre kestirilen madde parametreleri ve istatistikleri arasındaki ilişkiyi incelemek için Spearman Brown sıra farkları korelasyon katsayısı kullanılmıştır. Analizler sonucunda OKS Türkçe alt testinin KTK ve MTK'ya göre kestirilen madde parametreleri ve istatistikleri benzerlik göstermiştir. r parametresi ile a parametresi korelasyonu 0,961, p parametresi ile b parametresi korelasyonu 0,981 bulunmuştur.

Çelen (2008) çalışmasında geliştirdiği sözel yetenek testinden KTK ve MTK doğrultusunda madde seçimi yaparak oluşturduğu iki testin psikometrik özelliklerini karşılaştırmıştır. Test geliştirme aşamasında KTK ve MTK'nın gerekleri yerine getirilmeye çalışılmıştır. Deneme uygulaması verileri ile yapılan analiz sonuçlarına göre madde güçlük indeksleri ile b parametreleri arasında negatif yönde yüksek düzeyde (-0,97) ilişki, ayırt edicilik katsayıları ile a parametreleri arasında ise pozitif yönde ve orta düzeyde (0,56) ilişki gözlenmiştir. Ayrıca KTK'ya göre geliştirilen test puan ortalaması ile MTK'ya dayalı olarak geliştirilen testin yetenek kestirimleri arasında pozitif yönde yüksek ilişki (0,96) bulunmuştur. KTK'ya göre geliştirilen testin KR-20 katsayısı ile MTK'ya göre geliştirilen testin hesaplanan Lord'un güvenilirlik katsayısı oldukça benzer çıkmıştır.

Progar ve Sočan (2008) çalışmalarında TIMSS 1995 verilerini kullanarak, farklı bireyler ve maddeler üzerinde KTK ve MTK'nın karşılaştırılabilirliğini, bu değişmezliğin farklı bireyler üzerinde ne kadar olduğunu ve farklı madde setlerinde ne kadar olduğunu araştırmışlardır. Bu amaçla analizler esnasında TIMSS verileri madde ve birey gruplarına ayırmışlardır. Araştırma sırasında SPSS 21.0 ve BILOG MG programları kullanılmıştır. Model veri uyumu tüm MTK modelleri için gözlenmiştir. Araştırmacı, birey ve madde istatistiklerinin KTK ve MTK ile karşılaştırılabilirliğini her birey için doğru cevap oranı ve yetenek kestirimi değerlerinin korelasyonunu hesaplayarak değerlendirmiştir. Farklı katılımcı gruplarına göre madde parametrelerinin değişmezliğinin derecesini incelemek amacıyla madde parametreleri kestirimlerini iki veya daha fazla katılımcı gruplar arasında karşılaştırılmıştır. Farklı madde gruplarının madde parametrelerinin değişmezliğini incelemek amacıyla araştırmacılar, KTK ve MTK'ya göre hesaplanan madde istatistiklerini gözleyerek 4 test oluşturmuş ve orijinal parametrelerle yeni testlerden elde edilen

parametreler arasındaki korelasyon incelenmiştir. Araştırma sonucunda model veri uyumunu en çok 2PL modelin sağladığı tespit edilmiştir. Birey ve madde istatistiklerinin korelasyonları incelenerek karşılaştırılabilir oldukları belirlenmiştir. Çalışma sonucunda KTK ve MTK parametrelerinin karşılaştırılabilir olduğu ve değişmezliğin sağlandığı MTK'nın daha avantajlı olduğu vurgulanmıştır.

Jimelo ve Silvestre-Tipay (2009) KTK ile MTK arasındaki kuramsal olmasına farklılıkların sebebini ve ne ölçüde olduğu hususunda ampirik bilgi eksikliği bulunduğunu vurgulayarak araştırmalarında bu konuyu ele almışlardır. Veriler üniversite birinci sınıf öğrencilerine uygulanan biyoloji testinden elde edilmiştir. MTK ve KTK'ya göre madde güçlükleri kestirilmiş, kestirimlerin tutarlılığı incelenmiş ve MTK ile KTK'ye dayalı hesaplanan iç tutarlılıklar karşılaştırılmıştır. 326 kişiye uygulanan testten ikiye bölme yöntemi ile elde edilen iç tutarlık katsayısı 0,70 bulunmuş, Kuder-Richardson 20 yöntemi ile elde edilen iç tutarlık katsayısı ise 0,72 bulunmuştur. KTK ve MTK'ya göre kestirilen madde güçlükleri birbirinden önemli derecede farklı bulunmuş, KTK'ye göre toplam 60 sorunun 54'ü ortalama güçlükte çıkmış, MTK'ye göre 34 soru ortalama güçlükte çıkmıştır. İç geçerlikler iki kuram içinde yüksek ve yakın çıkmıştır. Araştırmacı iki kuramın birbirine üstünlüğü olmadığını vurgulamış, iki kuramın birbirinden bağımsız ele alınması gerektiğini belirtmiştir.

Atar (2011) çalışmasında TIMSS 2007 verilerini Rasch modeli, örtük regresyon Rasch modeli, doğrusal lojistik test modeli ve örtük regresyon doğrusal lojistik test modellerine göre incelemiştir. Araştırmanın verisi TIMSS 2007 Türkiye sekizinci sınıf matematik testinin bir kitapçığından alınan iki kategorili puanlanan maddelerden elde edilmiş, üç kategorili puanlanan madde analiz dışı bırakılmıştır. Öğrencilerin matematik başarılarındaki farklılığın belirlenmesi için birey özellikleri dört farklı yordayıcı değişken kullanılmıştır: Cinsiyet, matematiğe önem verme, matematiğe karşı olumlu tutum ve matematiği öğrenmede özgüven. Analizler sırasında MULTILOG 7.03, STATA ve GLLAMM yazılımı kullanılmıştır. MULTILOG 7.03 programı ile elde edilen madde güçlük parametreleri -1,28 ile 3,53 arasında değişmektedir ve ortalama madde güçlüğü yaklaşık 0,57 çıkmıştır. GLLAMM yazılımı ile elde edilen madde güçlük parametresi tahminleri ise -1,31 ile 3,67 arasında değişmektedir ve ortalama madde güçlüğü yaklaşık 0,58 çıkmıştır. Bu değer, ortalama matematik başarısına sahip bir öğrencinin testteki orta güçlükteki bir maddeyi doğru olarak cevaplama olasılığı %50'den düşük olduğu için testin zor bir test

olduđuna karar verilmiřtir. Rasch modelinde ğrencilerin matematik bařarısında anlamlı farklar bulunmuřtur.

Çelen ve Aybek (2013) alıřmalarında lme ve deęerlendirme dersindeki bařarıyı lmeyi amalayan bir test geliřtirip uygulamıř, test klasik test kuramı ve madde tepki kuramı yntemleriyle puanlayıp elde edilen puanları karřılařtırmıřlardır. Analiz ařamasında MTK kestirimleri iin BILOG MG programı kullanılmıř olup, daha iyi uyum saęladığı iin 3PL model kullanılmıřtır. Analizler sonucunda KTK ya da MTK'ya gre hesaplanan puanlar arasındaki iliřki pozitif ynde, yksek (0,951) ve manidardır. Model veri uyumunu saęlamayan sorular testten ıkarıldıđında ise KTK ve MTK'ya gre hesaplanan puanlar arasındaki iliřki pozitif ynde, yksek (0,953) ve manidar ıkmıřtır. Testin KTK'ya ya da MTK'ya gre geliřtirilmiř olmasının ğrencilerin bařarı sıralamalarında anlamlı bir deęiřikliğe yol amadığı tespit edilmiřtir.

Gler, Uyanık ve Teker (2014) alıřmalarında madde parametrelerini KTK ve MTK'ya gre karřılařtırmıřlardır. alıřmanın amacı iki kuramı kullanarak madde parametrelerinin kestirimlerindeki benzerlik ve farklılıkları belirlemektir. Arařtırmada Liselere Giriř Sınavına giren 5989 ğrenciden rastgele seilmiř 1250 ğrencinin Trke testine verdiđi yanıtlar kullanılmıřtır. Arařtırmanın ilk ařamasında modellerin varsayımları ve uyum iyiliđi incelenmiřtir. Model veri uyumunu en ok saęlayan model 3PL model olmuřtur. Daha sonra madde glkleri ve ayırt edicilikleri iki kurama gre hesaplanmıř ve karřılařtırılmıřtır. Madde glk parametreleri iin en yksek iliřki (0,99) KTK ile 1PL model arasında, madde ayırt edicilik parametreleri iin en yksek iliřki (0,96) KTK ile 2PL model arasında ıkmıřtır. 3PL model veri uyumunu en ok saęlayan model olmasına raęmen, KTK ile iliřkisi en dřk ıkan modeldir. Arařtırmacılar KTK ile 1PL veya 2PL model arasında ok fark olmadığını ve tahmin etme olasılıđının yksek olduđu durumlarda KTK ile 3PL model arasında nemli fark olduđunu belirtmiřlerdir.

zer zkan (2014) tarafından yapılan alıřmada ğrenci Bařarılarının Belirlenmesi Sınavı'nın (BBS) Trke ve matematik alt testi verilerini kullanarak elde edilen bařarı lleri kestiriminin doęruluđunun belirlemek ve KTK ile tek ve ok boyutlu MTK modellerinin testlere uygulanarak bařarı llerinin karřılařtırılması amalanmıřtır. Karřılařtırmalar sırasında bařarı llerini en az hata ile kestiren model arařtırılmıřtır. Analizler sonucunda Trke ve Matematik testinin ok boyutlu olduđu, ancak her iki testin de iki alt boyutu olduđu ve her bir alt boyutun tek bir yapıyı ltđ tespit edilmiř olup,

verinin alt boyutlar düzeyinde tek boyutlu MTK'ya göre kestirim yapılması için gereken tek boyutluluk varsayımını sağladığı belirtilmiştir. KTK'ya dayalı puanları elde etmek amacıyla ikili puanlama yapılmış olup, tek boyutlu MTK'ya dayalı dayalı yetenek kestirimlerinde her alt boyut için kestirim yapılmıştır. Yapılan analizler sonucunda Türkçe testi yetenek kestirimlerinden en yüksek hata (0,44) KTK'ya göre yapılan analizde ortaya çıkmıştır. Matematik testi yetenek kestirimlerinde en yüksek hata (0,47), Türkçe testinde olduğu gibi KTK'ya göre yapılan analizlerde ortaya çıkmıştır. KTK'ya göre elde edilen test puanları ile tek boyutlu MTK'ya göre elde edilen yetenek kestirimleri arasındaki korelasyonlar Türkçe testi için 0,71 ile 0,98 arasında, matematik testi için 0,66 ile 0,99 arasında değişmektedir. Sonuç olarak iki kuramın yetenek parametreleri arasında yüksek ve manidar ilişki bulunmuş, KTK'ya dayalı analizlerde tek boyutluluk varsayımı sağlanmadığında hata miktarının arttığı gözlenmiştir.

MTK Kestirimleri Yapan Programların Karşılaştırıldığı Araştırmalar

Bu bölümde paket programların karşılaştırmasının yapıldığı araştırmalara yer verilmiştir. Ancak literatürde MTK kestirimleri yapan programların karşılaştırıldığı sınırlı sayıda çalışma mevcuttur.

Ackerman (1989), 4 tamamlayıcı 4 de tamamlayıcı olmayan model için, boyutları farklı düzeylerde korelasyon gösteren 2 boyutlu yapılardan oluşan 8 veri kümesi üretmiştir. Her veri kümesinde güçlük, boyutla bağdaştırılmıştır. Her veri kümesi BILOG ve LOGIST kullanılarak analiz edilmiştir. Boyutluluğun güçlükle bağdaştırılması sonucu elde edilen yapılardaki parametre kestirimlerinin, BILOG programında LOGIST programından daha çok etkilendiği bulunmuştur.

Harris (1989) çalışmasında 1PL, 2PL ve 3PL modelleri tartışmış, her model için matematiksel formülleri vermiş ve üç model arasında karşılaştırmalar yapmıştır. 1PL modelin 2PL ve 3PL modele göre bazı avantajları vardır. Toplam test puanı yeteneği kestirmek için yeterli bir istatistiktir ve bir maddeye doğru yanıt veren öğrenci sayısı güçlük parametresini kestirmek için yeterli bir parametredir. Ayrıca aynı ham puana sahip cevaplayıcılar aynı yetenek düzeyine sahip olacaklardır. Bu açıdan diğer modellere göre basitlik sağlamaktadır ve testte maddeler arası ayırt ediciliklerinin eşitlenmesi varsayımı varsa 1PL modelin fayda sağlayacağı belirtilmiştir. 2PL model ise gerçek veri setlerinde c parametresinin tahmininde çekilen zorluklardan ötürü çok sık olmamakla birlikte tercih edilmektedir. Araştırmacı parametre tahminlerinde en çok LOGIST ve BILOG programlarının kullanıldığını, BILOG programının maksimum marjinal likelihood

yöntemiyle kestirim yaptığını belirtmiştir. Küçük örneklerde 3PL kullanmak yerine 1PL modelin kullanılması gerektiğini vurgulamıştır.

Kirisci, Hsu ve Yu (2001) çalışmalarında BILOG, MULTILOG 7.03 ve XCALIBRE programlarıyla kestirilen madde parametre tahminlerinin duyarlılığını araştırmayı amaçlamışlardır. Bu amaçla simülasyon verileri kullanılmış olup simüle edilmiş madde yanıtları programlar kullanılarak kalibre edilmiştir. Çalışmada tek boyutluluk ve normallik varsayımı ihlal edilmiş olup, 3PL modele göre kestirim yapılmıştır. Araştırmacılar her programın replikasyonundan edilen parametre tahminlerini kendi gerçek parametreleriyle eşitlemiştir. Analizler sonucunda tek faktörlü maddeler için, madde güçlükleri için RMSE ortalaması XCALIBRE programında BILOG'dan, MULTILOG 7.03 programında ise BILOG'dan yüksek sonuç vermiştir ancak 0,001 anlamlılık düzeyinde manidar farklılık gözlenmemiştir. Programlar ve boyutlar arası etkileşim anlamlı sonuç vermiştir ve madde ayırt edicilikleri için RMSE ortalaması BILOG'da en yüksek, MULTILOG'da en düşük çıkmıştır. Ayrıca tek boyutluluk varsayımını ihlal edilmesinden dolayı, madde parametre tahminleri XCALIBRE ve BILOG programlarında daha güvenilir sonuçlar vermiştir. Sonuç olarak, üç programda aynı kestirim yöntemiyle (MML) kestirim yapılsa da benzer sonuçlar elde edilmediği gözlenmiştir.

Rupp (2003) yaptığı araştırmada MTK modellerinin analizlerinde kullanımı yaygın olan BILOG MG ve MULTILOG 7.03 programlarını yorumlamayı amaçlamıştır. Çalışmada bu programlar ile kestirilebilen ya da kestirilemeyen modeller yorumlanmış, ayrıntılı matematiksel türevler verilmeden programların kestirim algoritmaları açıklanmış ayrıca programların kullanım kolaylığı değerlendirilmiştir. İkili puanlama yapılan testlerde iki programın da kullanıldığı, ancak gerekli varsayımların sağlanması gerektiğini vurgulamıştır. Bu varsayımlar yerel bağımsızlık, monotonluk ve tek boyutluluk olarak belirtilmiştir. İkili puanlamalarda MULTILOG 7.03 programında, BILOG MG'de ve diğer programlardan farklı olarak 0 ve 1 olarak puanlama yapılmayacağı, çünkü programda 0 olarak kodlanan cevapların kayıp veri olarak kabul edildiğine değinilmiştir. Ayrıca iki programında madde kestirimlerinde marjinal maksimumum likelihood (MML) kullandığı belirtilmiştir. İki programın kullanım kolaylığı karşılaştırıldığında BILOG MG programının sadece işaretleme ve seçim yönergelerinin olduğu ancak MULTILOG 7.03 programında uygulayıcının bazı syntax (söz dizimi) komutlarını bilmesi gerektiği, her iki programın grafik ara yüzlerinin profesyonel çıktı verdiğine değinilmiştir. Her iki

programın MTK modellerini geniş bir yelpazede kestirdiğini, ancak kullanıcıların MTK modelleri hakkında derinlemesine bilgi sahip olması gerektiğini vurgulamıştır.

Sonuç olarak, yapılan çalışmalarda ölçme araçlarının madde ve test parametrelerinin kestirilmesinde KTK ve MTK modellerinin kullanıldığı görülmektedir. Berberoğlu (1988) ÖSS sınavı için, Kelecioğlu (2001) Anadolu Lisesi Giriş Sınavı için, Doğan ve Tezbaşaran (2003) Liselere Giriş Sınavı için, Kan (2006) Ortaöğretim Kurumları Sınavı için, Güler, Uyanık ve Teker (2013) Liselere Giriş Sınavı için ve Özer Özkan (2014) Öğrenci Başarıları Belirleme sınavı için KTK ve MTK arasındaki ilişkiyi incelemişlerdir. Bu araştırmalarda genel olarak ikili puanlanan testlerde daha çok lojistik modellerin kullanıldığı, model veri uyumunun incelenerek daha çok 2PL ve 3PL modellerin uyum sağladığı gözlenmiştir. Progar ve Sočan (2008)'in çalışmasında TIMSS 1995 verilerinin KTK ve MTK'ya göre madde parametre değişmezliği incelenmiş, madde istatistikleri ile madde parametrelerinin korelasyonlarını inceleyerek benzerlik gösterdikleri sonucuna varmıştır. Ancak TIMSS 2011 projesi için KTK ile MTK'nın 1PL, 2PL ve 3PL modeli için karşılaştırma yapılmamıştır. Kirisci, Hsu ve Yu (2001) çalışmalarında BILOG, MULTILOG 7.03 ve XCALIBRE programlarıyla kestirilen madde parametre tahminlerinin duyarlılığını araştırmış, MULTILOG 7.03 programına göre hata katsayısı ortalamasını BILOG'dan daha düşük ancak güvenilirlik katsayısını MULTILOG 7.03 programında daha yüksek bulmuştur.

Farklı programlar ile yapılan parametre kestirimlerinde farklı sonuçlar elde edilebileceği görülmüştür. Alanyazında madde parametresi analizlerinde en çok BILOG MG programının kullanıldığı ancak BILOG MG, MULTILOG 7.03 ile R yazılımı karşılaştırmalarının yapıldığı çalışmalarda farklı sonuçlar elde edilebileceği, bu yüzden BILOG MG, MULTILOG 7.03 programları ve R yazılımlarının madde parametrelerinin karşılaştırılması araştırmaya değer görülmüştür.

BÖLÜM III

YÖNTEM

Bu bölümde araştırmanın deseni, evren ve örneklem, verilerin toplanması ve veri analizine yer verilmiştir.

Araştırmanın Deseni

TIMSS 2011 Türkiye 4.sınıf matematik alt testinin Klasik Test Kuramı'na göre hesaplanan madde istatistikleri ile Madde Tepki Kuramına göre kestirilen madde parametreleri arasında nasıl bir ilişki olduğunu incelemek amacıyla, bilinen yöntemler ile analiz edilip yorumlama içermesinden dolayı betimsel türde temel bir çalışmadır. Betimsel araştırmalar bilimin açıklayıcı işlevine hizmet eder. Betimleyici bir araştırmada herhangi bir ilişkinin varlık ya da yokluk durumu incelenir (Büyüköztürk, Çakmak, Akgün, Karadeniz ve Demirel, 2012). Bu kapsamla bu araştırmada KTK ile hesaplanan ve MTK'ya göre BILOG MG, MULTILOG 7.03 ve R paket programları ile elde edilen madde analizleri sonuçları karşılaştırmalı olarak incelenmiştir.

Evren ve Örneklem

Araştırmanın evrenini 2011 yılında Türkiye TIMSS uygulamasına katılan resmî ve özel ilköğretim okullarının 4. sınıfına devam eden öğrenciler oluşturmaktadır. TIMSS, ülkemizde 4. sınıflar düzeyinde 257 okula uygulanan bir örneklem çalışmasıdır. Araştırma kapsamında amaçlı örnekleme yöntemiyle 11. matematik kitapçığını alan ve tüm yanıtlarına ulaşılan 543 öğrenci çalışma grubunu oluşturmaktadır.

Veri Toplama Araçları

Araştırma kapsamında ele alınan veriler, TIMSS & PIRLS Uluslararası Çalışma Merkezine ait internet sitesindeki uluslararası veri tabanından SAS dosyaları şeklinde alınmıştır. SAS dosyasında 4. sınıf verilerinin bulunduğu dosyadan Türkiye katılımcılarının matematik

testine verdikleri yanıtlar alınarak veriler temin edilmiştir. Araştırmanın amacına yönelik yapılan analizlerde 26 matematik sorusu içeren 11. kitapçığı yanıtlayan 543 öğrencinin yanıtlarından elde edilen veriler kullanılmıştır. Seçilen kitapçık sadece ikili puanlanan maddelerden oluşmakta olup, yanıtlar doğru olan yanıtlara 1, yanlış yanıtlara 0 verilerek ikili puanlamaya çevrilerek analize dahil edilmiştir.

Verilerin Analizi

Analiz sırasında öncelikle verileri analizlere uygun hale getirmek için farklı kodlanmış olan soruların yanıtları 0-1 olarak ikili puanlamaya dönüştürülmüştür. Daha sonra SPSS 21.0 programı ile betimsel istatistikler hesaplanmıştır. Veriler ikili kodlanmış olduğundan yapı geçerliği tetrakorik açımlayıcı faktör analizi ile incelenmiş olup bu aşamada MPlus programı ile analiz yapılmıştır. Doğrulayıcı faktör analizi ise LISREL 8.8 programıyla incelenmiştir. Ardından MTK'nın normallik, tek boyutluluk, yerel bağımsızlık ve testin hız testi olup olmadığının kontrol edilmesi varsayımlarını sağlayıp sağlamadığı SPSS 21.0 ve Mplus programı ile kontrol edilmiştir. Ayrıca yerel bağımsızlık Yen (1984)'in Q_3 istatistiği ile test edilmiştir. Analiz aşamasında ikili puanlanan maddelerin parametreleri ve güvenilirlik katsayıları BILOG MG, MULTILOG 7.03 ve R programları ile ayrı ayrı kestirilmiştir. R, BILOG MG ve MULTILOG 7.03 programları ile parametre kestirimi yapılırken standart hatalar belirlenmiştir. Veri setine ilişkin betimsel istatistikler Tablo 1'de verilmiştir.

Tablo 1. Veri Setine İlişkin Betimsel İstatistikler

Betimsel istatistikler	
Birey Sayısı (n)	543
Madde Sayısı (k)	26
Std. Sapma (SS)	5,351
Ortalama ($\bar{X}$)	11,803
Medyan	12,0
Mod	6
Ranj	24

Testte 26 madde vardır ve en yüksek puan 25'tir. Analizlere dahil edilen birey sayısı 543'tür. Betimsel istatistiklere göre ortalama 11,803'tür. Ayrıca standart sapma 5,351 olarak bulunmuştur. Standart sapma ve bağıl değişkenlik katsayısı grupların homojenliğini betimleyen bir istatistiktir. Bağıl değişkenlik katsayısı 45,33 olarak hesaplanmış, bulunan değer 26'dan büyük olduğu için grubun heterojen olduğu sonucuna varılmıştır (Field,2009).

Araştırma kapsamında TIMSS 2011 4. sınıf matematik alt testi verilerinin normal dağılımı sağlayıp sağlamadığı, tek boyutluluk ve yerel bağımsızlık sayıtlarını karşılayıp karşılamadığı ve verilerin model-veri uyumu incelendikten sonra araştırma sorularının incelenmesine geçilmiştir. Aşağıda her bir araştırma sorusunun nasıl analiz edildiğine ilişkin bilgiler sırasıyla verilmiştir.

1. Araştırma Sorusuna İlişkin Veri Analizi

Araştırmanın 1. sorusu için test maddelerinin TIMSS 2011 Türkiye 4.sınıf matematik alt testindeki her bir maddeye ait KTK r (madde ayırt edicilik) indeksleri TAP (Test Analysis Program) programından elde edilmiştir.

1.1. Araştırma Sorusuna İlişkin Veri Analizi

1.1. Araştırma sorusu için TIMSS 2011 Türkiye 4.sınıf matematik alt testindeki her bir maddeye ait MTK a parametreleri 2PL ve 3PL model için BILOG MG programında kestirilmiştir. 2PL ve 3PL model için MTK a parametreleri ile KTK madde ayırt edicilik (r) katsayıları arasındaki ilişki SPSS 21.0 programında, madde sayısı 30'dan az olduğu için Spearman sıra farkları korelasyon katsayısı hesaplanarak incelenmiştir ve korelasyon katsayısının anlamlılığı Fisher'in z istatistiği ile test edilmiştir.

1.2. Araştırma Sorusuna İlişkin Veri Analizi

1.2. Araştırma sorusu için TIMSS 2011 Türkiye 4.sınıf matematik alt testindeki her bir maddeye ait MTK a parametreleri 2PL ve 3PL model için R programı ltm paketinde kestirilmiştir. Daha sonra 2PL ve 3PL model için MTK a parametreleri ile KTK madde ayırt edicilik (r) katsayıları arasındaki ilişki SPSS 21.0 programında, madde sayısı 30'dan az olduğu için Spearman sıra farkları korelasyon katsayısı hesaplanarak incelenmiştir ve korelasyon katsayısının anlamlılığı Fisher'in z istatistiğiyle test edilmiştir.

1.3. Araştırma Sorusuna İlişkin Veri Analizi

1.4. Araştırma sorusu için TIMSS 2011 Türkiye 4.sınıf matematik alt testindeki her bir maddeye ait MTK a parametreleri 2PL ve 3PL model için MULTILOG 7.03 programı ile kestirilmiştir. 2PL ve 3PL model için MTK a parametreleri ile KTK madde ayırt edicilik (r) katsayıları arasındaki ilişki SPSS 21.0 programında, madde sayısı 30'dan az olduğu için

Spearman sıra farkları korelasyon katsayısı hesaplanarak incelenmiştir ve korelasyon katsayısının anlamlılığı Fisher'in z istatistiği ile test edilmiştir.

2. Araştırma Sorusuna İlişkin Veri Analizi

Araştırmanın 2. sorusu için test maddelerinin TIMSS 2011 Türkiye 4.sınıf matematik alt testindeki her bir maddeye ait KTK madde güçlük (p) indeksleri TAP (Test Analysis Program) programı ile elde edilmiştir.

2.1. Araştırma Sorusuna İlişkin Veri Analizi

2.1. Araştırma sorusu için TIMSS 2011 Türkiye 4.sınıf matematik alt testindeki her bir maddeye ait MTK b parametreleri 1PL, 2PL ve 3PL model için R programı ltm paketinde kestirilmiştir. Parametreler elde edildikten sonra 1PL, 2PL ve 3PL model için MTK b parametreleri ile KTK madde güçlüğü (p) katsayıları arasındaki ilişki madde sayısı 30'dan az olduğu için SPSS 21.0 programında Spearman sıra farkları korelasyon katsayısı hesaplanarak incelenmiştir ve katsayının anlamlılığı Fisher'in z istatistiği ile sınanmıştır.

2.2. Araştırma Sorusuna İlişkin Veri Analizi

2.2. Araştırma sorusu için TIMSS 2011 Türkiye 4.sınıf matematik alt testindeki her bir maddeye ait MTK b parametreleri 1PL, 2PL ve 3PL model için MULTILOG 7.03 programı ile kestirilmiştir. Ardından 1PL, 2PL ve 3PL model için MTK b parametreleri ile KTK madde güçlüğü (p) katsayıları arasındaki ilişki madde sayısı 30'dan az olduğu için SPSS 21.0 programında Spearman sıra farkları korelasyon katsayısı hesaplanarak incelenmiştir ve korelasyon katsayısının anlamlılığı Fisher'in z istatistiğiyle test edilmiştir.

2.3. Araştırma Sorusuna İlişkin Veri Analizi

2.3. Araştırma sorusu için TIMSS 2011 Türkiye 4.sınıf matematik alt testindeki her bir maddeye ait MTK b parametreleri 1PL, 2PL ve 3PL model için BILOG MG programı ile kestirilmiştir. 1PL, 2PL ve 3PL model için MTK b parametreleri ile KTK madde güçlüğü (p) katsayıları arasındaki ilişki madde sayısı 30'dan az olduğu için SPSS 21.0 programında Spearman sıra farkları korelasyon katsayısı ile hesaplanmıştır. Korelasyon katsayısının anlamlılığı Fisher'in z istatistiğiyle test edilmiştir.

3. Araştırma Sorusuna İlişkin Veri Analizi

Araştırmanın 3. araştırma sorusu için model veri uyumu incelenmiş, TIMSS 2011 Türkiye 4.sınıf matematik alt testinin MTK'nın hangi modeli ile daha iyi uyum sağladığı test edilmiştir. Model veri uyumu incelenirken öncelikle BILOG MG, MULTİLOG 7.03 ve R yazılımı 'ltm' ve 'sirt' paketlerinden -2loglikelihood değerleri elde edilmiş ve bu değerler yorumlanmıştır. Bunun yanı sıra her program için χ^2 istatistiği hesaplanarak hangi modelin TIMSS 2011 Türkiye 4.sınıf matematik alt testi ile elde edilen verilere daha uyumlu olduğu incelenmiştir.

4. Araştırma Sorusuna İlişkin Veri Analizi

4.1. Araştırma Sorusuna İlişkin Veri Analizi

4.1. Araştırma sorusu için güvenirlik katsayısı SPSS 21.0 programı ile KTK'ye dayalı olarak Cronbach alfa iç tutarlık katsayısı hesaplanmıştır. Cronbach alfa iç tutarlık katsayısı formülü aşağıda verilmiştir:

$$\alpha = \frac{k}{k-1} \left(1 - \frac{\sum_{i=1}^k S_i^2}{S_x^2} \right)$$

MTK'nın 1PL, 2PL ve 3PL modellerine dayalı olarak yapılan güvenirlik katsayısı kestirimleri sırasında MULTİLOG 7.03 ve R programı 'ltm' ve 'sirt' paketinde yapılan kestirimlerden marjinal güvenirlik katsayıları elde edilmiştir. BILOG MG programında yapılan kestirimlerde ise görgül güvenirlik katsayısı elde edilmiştir.

4.2. Araştırma Sorusuna İlişkin Veri Analizi

Programlar aracılığı ile hesaplanan güvenirlik katsayıları birbirleriyle karşılaştırılarak yorumlanmıştır. Elde edilen güvenirlik katsayıları arasındaki farkın manidarlığı Feldt testi (Alsawalmeh ve Feldt, 1994) ile sınanmıştır. Bağımsız gruplar için güvenirlik katsayısı karşılaştırması yapılırken kullanılan Feldt testi, k, kullanılan ölçme araçlarındaki madde sayısını olmak üzere $(k - 1).(N - 1)$ değeri 1000'in üzerine çıktığı takdirde bağımlı gruplar için de bu formül kullanılabilmektedir.

$$W = \frac{1 - \alpha_2}{1 - \alpha_1}$$

Feldt testine göre α_1 , birinci teste ait güvenilirlik katsayısı ve α_2 , ikinci teste ait olan güvenilirlik katsayısı olmak üzere elde edilen W değeri, $F_{(N-1, N-1)}$ değerinden büyük ise farkın manidar olduğu söylenebilir (Alsawalmeh ve Feldt, 1994).

5. Araştırma Sorusuna İlişkin Veri Analizi

Beşinci araştırma sorusunda 1PL, 2PL ve 3PL modellere göre kestirilen a, b ve c parametreleri için standart hata değerleri MULTILOG 7.03, BILOG MG 3 ve R programlarından elde edilmiştir ve bu standart hata değerleri arasında manidar bir farklılık olup olmadığını incelenmiştir. Farklı programların kestirdiği parametrelerin standart hataları arasındaki manidar farklılığı incelemek için öncelikle verilerin tekrarlı ölçümler için tek yönlü ANOVA testinin sayıltılarını sağlayıp sağlamadığı kontrol edilmiştir. Ancak programlardan elde edilen parametrelere ait standart hatalar normallik ve küresellik sayıltılarını sağlamadığı için Friedman testi uygulanmıştır (Kilmen, 2015). Friedman testi sonuçlarına dayalı olarak programlar arasında farklılık olup olmadığı tespit edilmiştir. Ayrıca programlardan elde edilen standart hata değerleri arasında manidar bir farklılık varsa bu farklılığın hangi programlar arasında olduğunu tespit etmek amacıyla SPSS 21.0 programında 0,05 anlamlılık düzeyinde Friedman ikili karşılaştırması uygulanmıştır. Etki büyüklüğünün hesaplanmasında Wilcoxon İşaretli Sıralar Testi ile önce Z değerleri, daha sonra r etki büyüklüğü değerleri hesaplanmıştır.

KTK'ya İlişkin Geçerlik Çalışması

Geçerliği incelemenin bir yolu doğrulayıcı faktör analizidir (DFA). DFA yardımı ile yapı geçerliği kestirilir (Hambleton vd.,1991). Yapı geçerliğini belirlemek için açıklayıcı faktör analizi (AFA) ve doğrulayıcı faktör analizi (DFA) yapılmıştır. DFA'nın amacı farklı değişkenlerden oluşturulan faktörlerin gerçek verilerle ne derece uyum gösterdiğini değerlendirmektir (Büyüköztürk, 2009). İlk olarak örneklem uygunluğu KMO ve Barlett küresellik testleri yapılmıştır. Verilerin faktör analizine uygunluğu incelenirken KMO değerinin 0,60'dan yüksek ve Barlett testinin manidar çıkması beklenir (Çokluk, Şekercioğlu ve Büyüköztürk, 2012). Yapılan analizler sonucunda KMO örneklem uygunluk istatistiği 0,912 ve Barlett Sphericity testi anlamlı olup χ^2 değeri 2252,159 ($p < 0,001$) bulunmuştur. Veri setine Lisrel 8.8 yazılımı ile DFA yapılmıştır. Elde edilen değerler Tablo 2'de verilmiştir.

Tablo 2. Veri Setine Ait DFA Sonuçları

Uyum Ölçütleri	İyi Uyum Aralığı	Kabul Edilebilir Uyum Aralığı	Veri Seti için Bulunan Değerler
χ^2 /sd	$0 \leq \chi^2 /sd \leq 2$	$2 < \chi^2 /sd \leq 3$	1,25
p Value	$0,05 < p \leq 1,00$	$0,01 \leq p \leq 0,05$	1,00
RMSEA	$0 \leq RMSEA \leq 0,05$	$0,05 < RMSEA \leq 0,08$	0,013
NFI	$0,95 \leq NFI \leq 1,00$	$0,90 \leq NFI < 0,95$	0,98
NNFI	$0,97 \leq NNFI \leq 1,00$	$0,95 \leq NNFI < 0,97$	0,97
SRMR	$0 \leq SRMR \leq ,05$	$0,05 < SRMR \leq 0,10$	0,036
GFI	$0,95 \leq GFI \leq 1,00$	$0,90 \leq GFI < 0,95$	0,95
AGFI	$0,90 \leq AGFI \leq 1,00$	$0,85 \leq AGFI < 0,90$	0,94
CFI	$0,97 \leq CFI \leq 1,00$	$0,95 \leq CFI < 0,97$	0,99

Schermelleh, Engel ve Moosbrugger (2003)'den uyarlanan uyum ölçütleri ve aralıklarına göre, tüm ölçütler için iyi uyum sağlandığı tespit edilmiştir. Uyum indeksi değerleri model veri uyumunun iyi olduğunu göstermektedir.

Madde Tepki Kuramının Sayıtlarının Kontrolü

Test edilmesi gereken varsayımlar; normallik, tek boyutluluk, yerel bağımsızlık ve testin hız testi olup olmadığının kontrol edilmesidir. Varsayımlar sırasıyla incelenmiş ve kontrol edilmiştir.

Normallik

MTK'nın sayıtlarından biri örneklem üzerinden ölçülen özelliklerin evrende normal dağılım göstermesidir (Hambleton vd., 1991). Testten elde edilen puanların normal dağılım gösterip göstermediğini test etmek amacıyla yapılan Kolmogorov-Smirnov uyum iyiliği istatistiği sonucunda dağılımın 0,05 anlamlılık düzeyinde normal dağılmadığı gözlenmiştir ($p=0,00<0,05$). Yapılan test sonucu Tablo 3'de verilmiştir.

Tablo 3. Veri Setine İlişkin Kolmogorov-Smirnov ve Shapiro Wilk Testi Sonuçları

Kolmogorov-Smirnov ^a			Shapiro-Wilk		
İstatistik	sd	Sig.	İstatistik	sd	p
,084	543	0,000	0,975	543	0,000

SPSS 21.0 yardımıyla elde edilen betimsel istatistiklere göre basıklık katsayısı 0,77 bulunmuştur. Basıklık katsayısına göre dağılımın sivri şekilde olduğu söylenebilir. Çarpıklık katsayısı ise 0,207 bulunmuş olup puan dağılımı hafif sağa kayışıktır. Bulunan değerler Tablo 4'te verilmiştir.

Tablo 4. Veri Setine Ait Merkezi Dağılım Ölçüleri

Merkezi Dağılım Ölçüleri	Standart Hata	
Çarpıklık katsayısı	0,207	0,105
Basıklık katsayısı	-0,842	0,209

Tablo 4’te görüldüğü üzere veriler basıklık değerinden dolayı heterojen şekilde dağılmış olup, simetriğe yakın bir şekilde dağılmışlardır. Çarpıklık ve basıklık momentleri, standart hataları ile oranlandığında elde edilen değerler 1,96’dan büyüktür. Bu sonuca göre verinin normal dağılmadığı söylenebilir. Ancak bu durumun MTK’nın lojistik modellerine göre yapılacak kestirimlere önemli bir etkisi olmayacağı düşünülmüştür.

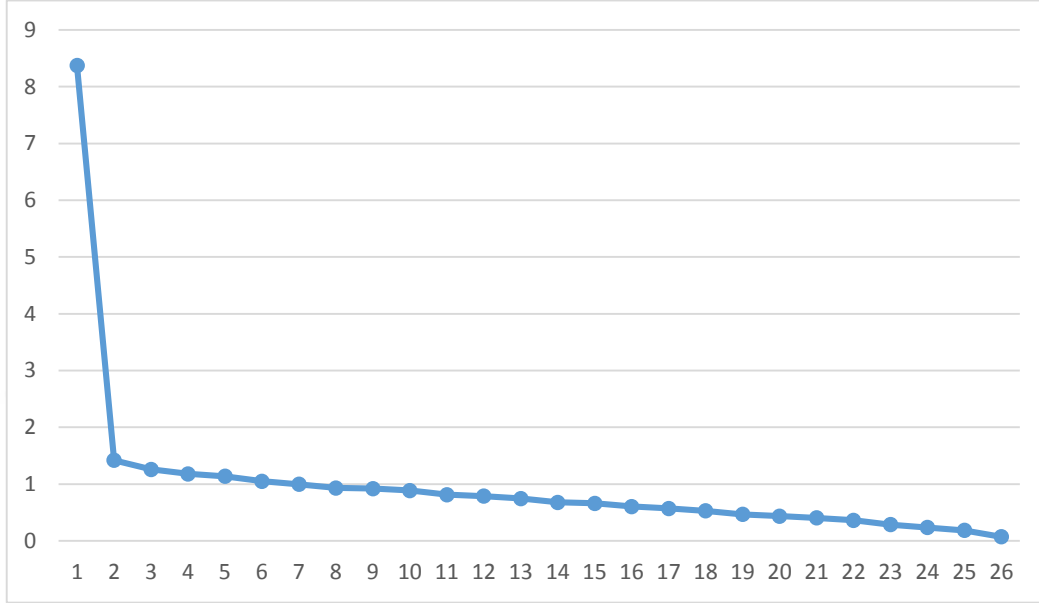
Tek Boyutluluk

MTK’nın sayıtlarından biri tek boyutluluktur. Bu sayıltıya göre ölçülen özellik tek boyutlu olmalıdır. Kullanılan veri iki kategorili olduğundan, sayıltının sağlanıp sağlanmadığı tetrakorik matrise dayalı açımlayıcı faktör analizi ile incelenmiş, yapılan faktör analizine ilişkin saçılma grafiği Şekil 4’de verilmiştir. Tetrakorik matrisin oluşturulmasında MPlus programından yararlanılmış, matrisin oluşturulmasından sonra açımlayıcı faktör analizi yapılmıştır. Analiz sonucunda faktör yük değerlerinin iki grupta toplandığı ve birinci faktörün öz değerinin 8,373, ikinci faktöre ait öz değerin ise 1,419 olduğu görülmüştür. Faktör analizi çıktılarına göre ilk faktör toplam varyansın % 32,295’ini açıklamaktadır. Birinci faktöre ilişkin özdeğer ikinci faktöre ilişkin özdeğerin 5 katından fazladır. İkinci faktöre ait öz değer diğer faktörlerin öz değerlerinden çok farklı olmamasından dolayı tek boyutluluk varsayımının sağlandığı kabul edilir (Lord, 1980). Ayrıca özdeğer grafiğine göre birinci boyuttan sonra keskin bir düşüş izlenmektedir. Bu göstergeler sonucunda veri setinin tek boyutlu olduğuna karar verilmiştir. Ayrıca birinci faktör açıklanan varyansın üçte ikisinden fazlasını oluşturduğu için tek boyutlu olduğu söylenebilir (Çokluk vd., 2012). Faktör çıkarma yöntemi olarak Mplus programında “robust weighted least squares estimator” kullanılmıştır. Tablo 5’te, yapılan tetrakorik AFA’ya ait özdeğer sonuçları verilmiştir.

Tablo 5. Tetrakorik AFA’ya Ait Özdeğerler

Faktör	Özdeğer	Varyans Yüzdesi	Yığılmalı Yüzde
1	8,373	32, 295	32,295
2	1,419	3,679	35,974

Şekil 3'te 26 maddeye ait Mplus programıyla elde edilen öz değer saçılım grafiği verilmiştir. İlk özdeğerden sonraki ani düşüş ve diğer öz değerlerin çok daha düşük olması testin tek boyutluluğuna işaret etmektedir.



Şekil 3. TIMSS 2011 4. sınıf matematik 11. alt testine saçılma grafiği

Yerel Bağımsızlık

Madde tepki kuramının bir diğer varsayımı yerel bağımsızlıktır. Yerel bağımsızlık bireylerin yetenekleri sabit tutulduğunda, testteki herhangi madde çiftine verdiği tepkilerin istatistiksel olarak bağımsız olması demektir. Bireylerin bir maddedeki performansı diğer maddedeki performansından bağımsız olmalıdır. Sabit bir yetenek düzeyinde bulunan bireylerin test puanları da aynı olmalıdır (Hambleton ve Swaminathan, 1985; Hambleton vd., 1991). Bu varsayıma göre test performansını etkileyen yetenekler sabit tutulursa, restgele seçilen iki test maddesine karşı bireylerin yanıtları istatistiksel olarak bağımsızdır (Emberson ve Reise, 2000). Eş deyişle, maddeler istatistiksel olarak birbirinden bağımsız değilse, aynı yetenek düzeyindeki bireylerin test puanları farklı sonuçlanmalıdır. Tek boyutluluğun sağlanması, o teste ait verilerin yerel bağımsızlığı da sağladığını göstermektedir. Çalışmada tek boyutluluk sağlandığı için, yerel bağımsızlığın da sağlandığı kabul edilebilir. Ayrıca yerel bağımsızlık Yen (1984)'ün Q_3 istatistiği hesaplanarak test edilmiştir. İstatistiğin hesaplanmasında R yazılımı 'sirt' paket programı kullanılmış olup, θ düzeyi $[-3,+3]$ aralığında alınmıştır. 1PL, 2PL ve 3PL modeller için ayrı ayrı Q_3 istatistikleri hesaplanmış, her 3 modelde de Q_3 istatistiği 0,2'den küçük

bulunmuştur. Dolayısıyla testin 1PL, 2PL ve 3PL modeller için yerel bağımsızlığı sağladığı söylenebilir (DeMars, 2010). Hesaplanan Q_3 istatistikleri Ek 5'te verilmiştir.

Testin Hız Testi Olup Olmadığının Kontrol Edilmesi

Bir diğer varsayım ise testin hız testi olup olmadığının test edilmesidir. Testin hız testi olup olmadığına, her bir bireyin boş bıraktığı madde sayılarının varyansının, yanlış yaptıkları madde sayılarının varyansına oranı alınarak karar verilir. Bu oran sıfıra yakın ise test hız testi olarak çalışmamaktadır (Berberoğlu, Çömlekoğlu ve Yıldırım, 2003). Veri setindeki 26 maddeden oluşan test ikili puanlama şeklinde kodlanmıştır. Maddeler 1-0 şeklinde kodlanmış olup yanıtına ulaşamayan bireyler analize dahil edilmemiştir. Dolayısıyla testin hız testi olmadığı söylenebilir.

Model Veri Uyumu

Madde ve birey parametreleri kestirildikten sonra her bir model için uyum incelenebilir. Madde veri uyumunu incelemek için çeşitli yöntemler mevcuttur ve bunlardan biri χ^2 istatistiğidir. χ^2 istatistiğinin anlamlı çıkması istenmeyen bir durumdur. Madde-veri uyumunun sağlanması için bu istatistiğin anlamsız çıkması gerekmektedir (Klein, 2005). İncelenecek diğer bir uyum ise model veri uyumudur. MTK'ya dayalı kestirim yapan programlar $-2 \log$ likelihood değerleri vermektedir. Bu değer büyüdükçe uyum kötüleşir. Örneğin 1PL model için hesaplanan $-2 \log$ likelihood değeri, 2PL model için hesaplanan $-2 \log$ likelihood değerinden büyükse, 2PL modelin uyumu veriye daha yüksektir denir. Yani $-2 \log$ likelihood değeri düştükçe uyum artar. (Klein, 2005). Ayrıca maksimum likelihood kestiriminde negatif çıkan log likelihood değeri, verinin modelden uzaklaşma derecesidir (Embretson ve Reise, 2000).

MTK'da model veri uyumu sağlandığında, testteki madde parametreleri testin uygulandığı grubun yetenek düzeylerinden bağımsız olarak kestirilir. Bir başka deyişle madde parametrelerinin, denek evreninden seçilen farklı örneklemdeki bireylerin maddelere verdiği yanıtlardan bağımsız bir şekilde kestirildiği kabul edilmektedir (Hambleton ve Swaminathan, 1985).

MTK modelleri ile veri uyumu tatmin edici olduğunda, MTK uygulamalarının avantajlı olduğu görülmektedir. Veri ile MTK modelinin uyumu zayıf olduğunda madde ve birey parametrelerinin değişmezliği elde edilemez. Bu özellik, MTK'nın KTK'ya tercih

edilmesinin en önemli sebebi olarak gösterilir; çünkü KTK’da madde parametreleri ile bireylerin karakteristik özellikleri ayrı olarak ele alınamaz (Hambleton vd., 1991).

Bu araştırmada üç ayrı program kullanıldığından, her bir programın hangi model ile iyi uyum sağladığı araştırılmıştır. Bu amaçla programların 1PL, 2PL ve 3PL modeller ile uyumları incelenip karşılaştırılmıştır. Model veri uyumları incelenirken MTK’nın her bir modeli için $-2 \log$ likelihood değerleri elde edilmiş olup, bulunan değerler negatiflik ve pozitifliklerine ve 0’dan uzaklaşma derecelerine göre incelenmiştir. R ve MULTILOG 7.03 programlarına göre en düşük $-2 \log$ likelihood değeri 3PL model; BILOG programı için ise 2PL model için elde edilmiştir.

Ayrıca tek boyutlu MTK modellerin uyumunu test etmenin diğer bir yolu kay-kare testidir. Kay-kare testi için $-2\log$ likelihood değerlerinin farkı alınarak serbestlik derecesine bölünür ve fark kay-kare tablosundaki değerden büyük çıkarsa farkın anlamlı olduğu kabul edilir (Hambleton vd., 1991). Yapılan işlemler sonucunda, kay-kare tablo değeri $S_d = 26$ için, 0,05 anlamlılık düzeyinde 38,885 dir. Her üç programa göre 0,05 anlamlılık düzeyinde 1PL ile 2PL; 2PL ile 3PL modelleri arasındaki uyum farkı manidar çıkmamıştır. Yani $-2 \log$ likelihood değerlerine göre farklı modeller ile uyum dereceleri farklılaşsa da, modeller arası uyum farkı manidar değildir. Model veri uyumunun yanında maddelerin model uyumuna da bakılmasına gerek duyulmaktadır.

Model-Madde Uyumu

MTK’da tüm maddelere göre incelenen model uyumunun yanı sıra, her bir maddenin model ile uyumu da incelenmelidir (Demars, 2010). Madde uyumunu incelemek amacıyla madde bazında kay kare değerleri ve manidarlıkları incelenmektedir. Araştırma kapsamındaki 26 maddeye ilişkin 1PL, 2PL ve 3PL modele ait kay kare değerleri Tablo 6’da verilmiştir.

Tablo 6. Model Madde Uyumuna İlişkin Kay Kare (KK) ve p değerleri

	1PLM		2PLM		3PLM	
	KK	p	KK	p	KK	p
Madde 1	37,2	0,000	16,5	0,006	46,9	0,000
Madde 2	4,8	0,780	10,7	0,216	11,3	0,257
Madde 3	21,1	0,007	21,7	0,010	9,0	0,440
Madde 4	30,6	0,000	5,1	0,534	23,6	0,001
Madde 5	14,6	0,042	6,1	0,527	11,6	0,172
Madde 6	27,2	0,001	25,3	0,001	17,8	0,037
Madde 7	14,3	0,046	14,0	0,052	12,4	0,135
Madde 8	5,3	0,619	4,7	0,791	4,9	0,773
Madde 9	16,8	0,032	4,8	0,775	8,7	0,462
Madde10	5,4	0,615	4,1	0,852	10,7	0,221
Madde11	14,0	0,083	7,7	0,357	29,4	0,000
Madde12	13,3	0,104	9,2	0,242	24,0	0,001
Madde13	44,8	0,000	10,7	0,099	28,9	0,000
Madde14	4,5	0,726	2,0	0,962	10,5	0,235
Madde15	24,1	0,001	5,2	0,634	26,4	0,001
Madde16	3,9	0,867	6,5	0,597	11,2	0,189
Madde17	56,4	0,000	3,5	0,623	15,0	0,020
Madde18	2,2	0,974	3,6	0,894	4,3	0,890
Madde19	21,3	0,003	4,8	0,680	11,8	0,107
Madde20	11,2	0,188	6,0	0,741	4,9	0,839
Madde21	19,1	0,008	8,5	0,291	12,6	0,084
Madde22	21,3	0,003	3,4	0,846	7,6	0,369
Madde23	16,3	0,039	4,8	0,776	8,0	0,530
Madde24	29,8	0,000	11,2	0,189	5,9	0,655
Madde25	13,2	0,0675	8,0	0,331	21,4	0,006
Madde26	16,4	0,0371	13,6	0,093	10,2	0,252

Tablo 6’da görüldüğü üzere $p=0,001$ düzeyinde 1PL model için 5, 3PL model için 3 madde uyum sağlamamıştır. 2PL modelde ise uyum sağlamayan madde bulunmamaktadır. Madde bazında bakıldığında, maddeler en çok 2PL model ile uyum sağlamaktadır.

Madde Parametrelerinin Değişmezliği

Parametre değişmezliği madde parametrelerinin gruptan bağımsız olmasıdır. MTK’ya dayalı olarak gerçekleştirilen parametre kestirimleri sonrasında parametre değişmezliğinin gösterilmesi gerekmektedir. Madde parametrelerinin değişmezliği evrenden çekilen iki veya daha fazla alt gruptan elde edilen madde parametrelerinin karşılaştırılması ile kontrol edilmektedir (Hambleton ve Swaminathan,1985). Madde parametrelerinin değişmezlik özelliği incelenirken veri setinde yer alan katılımcılar seçkisiz olarak ikiye bölünerek

parametre kestirimi tekrar gerçekleştirilmiştir. Ardından iki gruptan elde edilen parametreler arasındaki korelasyonlar hesaplanmıştır. Veri setleri ile yapılan analizler sonucunda 1PL model için b parametreleri arasındaki korelasyon 0,895; 2PL model için a parametreleri arasındaki korelasyon 0,839; b parametreleri arasındaki korelasyon 0,966 dır. 3PL model için a parametreleri arasındaki korelasyon 0,824; b parametreleri arasındaki korelasyon 0,962 ve c parametreleri arasındaki korelasyon 0,654 çıkmıştır. Bu sonuçlara göre değişmezlik, a parametreleri için yüksek, b parametreleri için mükemmel ve c parametresi için orta derecede sağlanmıştır. a ve c parametreleri b parametrelerine göre daha düşük değişmezlik özelliği sağlamaktadır.





BÖLÜM IV

BULGULAR ve YORUM

Bu bölümde araştırma soruları sırasıyla incelenerek bulgulara ve yorumlara yer verilmiştir.

1. Araştırma Sorusuna İlişkin Bulgular

Araştırmanın ilk sorusunda matematik alt testindeki her bir madde için KTK'ya göre elde edilen güçlük katsayısı olan p elde edilmiştir. 26 maddeden oluşan testin KTK'ya göre hesaplanan güçlük katsayıları Tablo 7'de verilmiştir.

Tablo 7. KTK'ya Göre Kestirilen p Katsayıları

Madde Numarası	p
Madde 1	0,84
Madde 2	0,52
Madde 3	0,52
Madde 4	0,58
Madde 5	0,43
Madde 6	0,48
Madde 7	0,60
Madde 8	0,19
Madde 9	0,33
Madde10	0,26
Madde11	0,73
Madde12	0,73
Madde13	0,41
Madde14	0,22
Madde15	0,47
Madde16	0,76
Madde17	0,51
Madde18	0,44
Madde19	0,31
Madde20	0,53
Madde21	0,17
Madde22	0,24
Madde23	0,40
Madde24	0,31
Madde25	0,39
Madde26	0,48

Tablo 7’de görüldüğü gibi, KTK’ya göre kestirilen madde güçlük katsayıları 0,17 (madde 21) ile 0,84 (madde 1) arasında değişmektedir. Bulgulara göre 4 madde çok kolay, 6 madde kolay, 11 madde orta güçlükte ve 5 madde ise zordur.

1.1. Araştırma Sorusuna İlişkin Bulgular

“BILOG programı ile TIMSS 2011 Türkiye 4.sınıf matematik alt testindeki her bir maddeye ait MTK’nin 1PL, 2PL ve 3PL modeline göre kestirilen b parametreleri ve bu parametreler ile KTK’ye göre hesaplanan madde güçlük (p) katsayıları arasındaki ilişkinin düzeyi nedir?”

Bu araştırma sorusunda her bir maddenin MTK’ya göre BILOG programından elde edilen b parametreleri incelenmiş olup, Tablo 8’de verilmiştir.

Tablo 8. BILOG Programında Kestirilen b Parametreleri

Madde Numarası	1PLM için b parametresi	2PLM için b parametresi	3PLM için b parametresi
Madde 1	-1,969	-1,433	-0,980
Madde 2	-0,081	-0,082	0,690
Madde 3	-0,099	-0,130	1,044
Madde 4	-0,369	-0,276	0,047
Madde 5	0,332	0,287	0,652
Madde 6	0,089	0,126	1,016
Madde 7	-0,497	-0,461	0,356
Madde 8	1,695	2,365	2,120
Madde 9	0,872	1,523	2,152
Madde10	1,264	1,254	1,380
Madde11	-1,181	-1,042	-0,445
Madde12	-1,170	-1,011	-0,362
Madde13	0,424	0,323	0,526
Madde14	1,526	1,437	1,513
Madde15	0,169	0,134	0,382
Madde16	-1,349	-1,581	-0,089
Madde17	-0,027	-0,015	0,209
Madde18	0,278	0,348	1,038
Madde19	0,943	0,702	0,848
Madde20	-0,135	-0,171	0,834
Madde21	1,865	1,587	1,557
Madde22	1,415	1,141	1,225
Madde23	0,516	0,806	1,483
Madde24	0,943	1,796	2,184
Madde25	0,553	0,530	0,882
Madde26	0,116	0,136	0,819

Tablo 8 incelendiğinde, BILOG programı ile yapılan MTK'ya göre b parametresi kestirimlerinin 1PL modelde -1,969 (madde 1) ile 1,865 (madde 21) arasında değiştiği görülmektedir. 2PL modelde -1,581 (madde 16) ile 2,365 (madde 8) arasında değişmektedir. 3PL modelde ise -0,980 (madde 1) ile 2,184 (madde 24) arasında değişmektedir.

KTK p katsayıları ile BILOG programında 1PL modele göre kestirilen b parametreleri arasında negatif yönde yüksek düzeyde manidar ilişki bulunmuştur ($r=-0,999$, $p< 0,01$). Değişkenler arasındaki korelasyon Fisher'in z istatistiği ile test edilmiş ve 0,01 manidarlık düzeyinde anlamlı çıkmıştır. Maddelerin güçlük özelliğine ilişkin değerler KTK ve MTK'nın 1PL modeline göre negatif yönde ve yüksek düzeyde benzerlik göstermiştir.

KTK p katsayıları ile BILOG programında 2PL modele göre kestirilen b parametreleri arasında negatif yönde yüksek düzeyde manidar ilişki bulunmuştur ($r=-0,973$, $p< 0,01$).

Değişkenler arasındaki korelasyon Fisher'in z istatistiği ile test edilmiş ve 0,01 manidarlık düzeyinde anlamlı çıkmıştır.

KTK p katsayıları ile BILOG programında 3PL modele göre kestirilen b parametreleri arasında negatif yönde yüksek düzeyde manidar ilişki bulunmuştur ($r=-0,844$, $p<0,01$). Değişkenler arasındaki korelasyon Fisher'in z istatistiği ile test edilmiş ve 0,01 manidarlık düzeyinde anlamlı çıkmıştır. BILOG programından elde edilen bulgulara göre maddelerin güçlük özelliğine ilişkin değerler KTK ve MTK'nın 3PL modeline göre negatif yönde yüksek derecede benzerlik göstermektedir.

1.2. Araştırma Sorusuna İlişkin Bulgular

“MULTILOG 7.03 programı ile TIMSS 2011 Türkiye 4.sınıf matematik alt testindeki her bir maddeye ait MTK'nın 1PL, 2PL ve 3PL modeline göre kestirilen b parametreleri ile KTK'ye göre hesaplanan madde güçlük (p) katsayıları arasındaki ilişkinin düzeyi nedir?”

Araştırmanın bu sorusunda her bir maddenin MTK'ya göre MULTILOG 7.03 programından elde edilen b parametreleri incelenmiş olup, Tablo 9'da verilmiştir.

Tablo 9. MULTILOG 7.03'e Göre Elde Edilen b Parametreleri

Madde Numarası	1PLM için b parametresi	2PLM için b parametresi	3PLM için b parametresi
Madde 1	-1,950	-1,390	-1,420
Madde 2	-0,080	-0,090	0,550
Madde 3	-0,100	-0,140	0,960
Madde 4	-0,370	-0,270	-0,280
Madde 5	0,330	0,280	0,260
Madde 6	0,090	0,120	0,950
Madde 7	-0,490	-0,460	0,320
Madde 8	1,670	2,470	1,980
Madde 9	0,860	1,690	1,930
Madde10	1,250	1,240	1,190
Madde11	-1,170	-1,020	-1,010
Madde12	-1,160	-0,990	-1,010
Madde13	0,420	0,310	0,300
Madde14	1,500	1,410	1,320
Madde15	0,160	0,130	0,120
Madde16	-1,330	-1,590	0,030
Madde17	-0,030	-0,020	-0,020
Madde18	0,260	0,340	0,900
Madde19	0,930	0,680	0,650
Madde20	-0,140	-0,180	0,360
Madde21	1,840	1,550	1,390
Madde22	1,390	1,110	1,050
Madde23	0,510	0,850	1,290
Madde24	0,930	2,110	2,130
Madde25	0,540	0,500	0,470
Madde26	0,110	0,130	0,110

Tablo 9 incelendiğinde, MULTILOG 7.03 programı ile yapılan MTK'ya göre b parametresi kestirimlerinin 1PL modelde -1,95 (madde 1) ile 1,840 (madde 21) arasında değiştiği görülmektedir. 2PL modelde -1,59 (madde 16) ile 2,47 (madde 8) arasında değişmektedir. 3PL modelde ise -1,420 (madde 1) ile 2,130 (madde 24) arasında değişmektedir.

KTK p katsayıları ile MULTILOG 7.03 programında 1PL modele göre kestirilen b parametreleri arasında negatif yönde yüksek düzeyde manidar ilişki bulunmuştur ($r=-0,999$, $p<0,01$). Değişkenler arasındaki korelasyon Fisher'in z istatistiği ile test edilmiş ve 0,01 manidarlık düzeyinde anlamlı çıkmıştır. MULTILOG 7.03 programından elde edilen bulgulara göre maddelerin güçlük özelliğine ilişkin değerler KTK ve MTK'nın 1PL modeline göre negatif yönde yüksek derecede benzerlik göstermektedir.

KTK p katsayıları ile MULTILOG 7.03 programında 2PL modele göre kestirilen b parametreleri arasında negatif yönde yüksek düzeyde manidar ilişki ($r=-0,969$, $p<0,01$)

bulunmuştur. Değişkenler arasındaki korelasyon Fisher'in z istatistiği ile test edilmiş ve 0,01 manidarlık düzeyinde anlamlı çıkmıştır.

KTK p katsayıları ile MULTILOG 7.03 programında 3PL modele göre kestirilen b parametreleri arasında negatif yönde yüksek düzeyde manidar ilişki ($r=-0,815$, $p<0,01$) bulunmuştur. Değişkenler arasındaki korelasyon Fisher'in z istatistiği ile test edilmiş ve 0,01 manidarlık düzeyinde anlamlı çıkmıştır. MULTILOG 7.03 programından elde edilen bulgulara göre maddelerin güçlük özelliğine ilişkin değerler KTK ve MTK'nın 2PL modeline göre negatif yönde yüksek derecede benzerlik göstermektedir.

1.3. Araştırma Sorusuna İlişkin Bulgular

“R programı ile TIMSS 2011 Türkiye 4.sınıf matematik alt testindeki her bir maddeye ait MTK'nın 1PL, 2PL ve 3PL modeline göre kestirilen b parametreleri ve bu parametreler ile KTK'ye göre hesaplanan madde güçlük (p) katsayıları arasındaki ilişkinin düzeyi nedir?”

Araştırmadaki her bir maddenin MTK'ye göre R programından elde edilen b parametreleri incelenmiş olup, Tablo 10'da verilmiştir.

Tablo 10. R'ye göre elde edilen b parametreleri

Madde Numarası	1PLM için b parametresi	2PLM için b parametresi	3PLM için b parametresi
Madde 1	-1,944	-1,383	-1,404
Madde 2	-0,079	-0,082	0,603
Madde 3	-0,097	-0,134	1,005
Madde 4	-0,363	-0,269	-0,249
Madde 5	0,328	0,280	0,302
Madde 6	0,088	0,128	0,984
Madde 7	-0,490	-0,452	0,240
Madde 8	1,675	2,457	2,054
Madde 9	0,861	1,681	1,973
Madde10	1,249	1,233	1,257
Madde11	-1,164	-1,016	-1,004
Madde12	-1,154	-0,986	-0,988
Madde13	0,418	0,314	0,338
Madde14	1,506	1,403	1,389
Madde15	0,168	0,130	0,155
Madde16	-1,331	-1,580	0,150
Madde17	-0,026	-0,015	0,014
Madde18	0,268	0,339	0,953
Madde19	0,931	0,681	0,700
Madde20	-0,132	-0,174	0,391
Madde21	1,842	1,542	1,465
Madde22	1,398	1,109	1,114
Madde23	0,510	0,846	1,373
Madde24	0,931	2,101	2,104
Madde25	0,540	0,506	0,522
Madde26	0,115	0,135	0,150

Tablo 10 incelendiğinde, R programı ile yapılan MTK'ya göre b parametresi kestirimleri 1PL modelde -1,944 (madde 1) ile 1,842 (madde 21) arasında değişmektedir. 2PL modelde -1,580 (madde 16) ile 2,457 (madde 8) arasında değişmektedir. 3PL modelde ise -1,404 (madde 1) ile 2,104 (madde 24) arasında değişmektedir.

MTK'ya göre elde edilen b parametreleri ile KTK'ya göre elde edilen güçlük katsayıları arasındaki ilişki Spearman Sıra Farkları korelasyon katsayıları hesaplanarak elde edilmiştir. Bu aşamada üç programdan göre kestirilen parametreler ve güçlük indekslerinin ilişkisi SPSS 21.0 programı aracılığı ile incelenmiştir.

KTK p katsayıları ile R programında 1PL modele göre kestirilen b parametreleri arasında negatif yönde yüksek düzeyde manidar ilişki bulunmuştur ($r = -0,999$, $p < 0,01$). Değişkenler arasındaki korelasyon Fisher'in z istatistiği ile test edilmiş ve 0,01 manidarlık düzeyinde anlamlı çıkmıştır. BILOG MG programından elde edilen bulgulara göre

maddelerin güçlük özelliğine ilişkin değerler KTK ve MTK'nın 1PL modeline göre negatif yönde yüksek derecede benzerlik göstermektedir.

KTK p katsayıları ile R programında 2PL modele göre kestirilen b parametreleri arasında negatif yönde yüksek düzeyde manidar ilişki bulunmuştur ($r=-0,968$, $p< 0,01$). Değişkenler arasındaki korelasyon Fisher'in z istatistiği ile test edilmiş ve 0,01 manidarlık düzeyinde anlamlı çıkmıştır. R programından elde edilen bulgulara göre maddelerin güçlük özelliğine ilişkin değerler KTK ve MTK'nın 2PL modeline göre negatif yönde yüksek derecede benzerlik göstermektedir.

KTK p katsayıları ile BILOG MG programında 3PL modele göre kestirilen b parametreleri arasındaki korelasyon gösterilmiş olup, negatif yönde yüksek düzeyde manidar ilişki bulunmuştur ($r=-0,827$, $p< 0,01$). Değişkenler arasındaki korelasyon Fisher'in z istatistiği ile test edilmiş ve 0,01 manidarlık düzeyinde anlamlı çıkmıştır. R programından elde edilen bulgulara göre maddelerin güçlük özelliğine ilişkin değerler KTK ve MTK'nın 3PL modeline göre negatif yönde yüksek derecede benzerlik göstermektedir.

2.Araştırma Sorusuna İlişkin Bulgular

Araştırmanın ikinci sorusunda matematik alt testindeki her bir madde için KTK'ya göre elde edilen ayırt edicilik indeksi olan r katsayıları Tablo 11'de verilmiştir.

Tablo 11. KTK'ya Göre Kestirilen r Katsayıları

Madde Numarası	r
Madde 1	0,450
Madde 2	0,440
Madde 3	0,350
Madde 4	0,590
Madde 5	0,540
Madde 6	0,380
Madde 7	0,490
Madde 8	0,300
Madde 9	0,270
Madde10	0,440
Madde11	0,450
Madde12	0,470
Madde13	0,600
Madde14	0,440
Madde15	0,580
Madde16	0,350
Madde17	0,640
Madde18	0,410
Madde19	0,570
Madde20	0,370
Madde21	0,460
Madde22	0,510
Madde23	0,320
Madde24	0,250
Madde25	0,480
Madde26	0,420

Tablo 11’de verilen KTK’ya göre hesaplanan ayırt edicilik katsayıları incelendiğinde 18 maddenin çok iyi ayırt edici, 6 maddenin ise iyi derecede ayırt edicilik özelliği taşıdığı görülmüştür. 2 madde ise düşük ayırt edicilik özelliği göstermiştir (madde 9 ve madde 24). r katsayıları 0,25 ile 0,64 arasında değişmektedir.

2.1. Araştırma Sorusuna İlişkin Bulgular

“BILOG MG programı ile TIMSS 2011 Türkiye 4.sınıf matematik alt testindeki her bir maddeye ait MTK’nin 2PL ve 3PL modeline göre kestirilen a parametreleri ve bu parametreler ile KTK’ye göre hesaplanan madde ayırt edicilik (r) katsayıları arasındaki ilişkinin düzeyi nedir?”

MTK’ya ilişkin a parametreleri BILOG MG programı kullanılarak 2PL ve 3PL model için kestirilmiştir. Bulunan değerler Tablo 12’de verilmiştir.

Tablo 12. BILOG MG Programında Kestirilen a Parametreleri

Madde Numarası	2PLM için a parametresi	3PLM için a parametresi
Madde 1	1,735	1,977
Madde 2	0,906	1,756
Madde 3	0,653	1,875
Madde 4	1,635	1,960
Madde 5	1,295	1,910
Madde 6	0,674	2,632
Madde 7	1,117	2,478
Madde 8	0,653	1,488
Madde 9	0,501	1,010
Madde10	1,019	1,707
Madde11	1,216	1,466
Madde12	1,259	1,627
Madde13	1,644	1,908
Madde14	1,100	1,743
Madde15	1,589	1,926
Madde16	0,808	1,394
Madde17	2,007	2,401
Madde18	0,754	1,781
Madde19	1,695	2,160
Madde20	0,695	1,268
Madde21	1,287	1,976
Madde22	1,427	2,233
Madde23	0,571	1,447
Madde24	0,454	1,261
Madde25	1,085	1,404
Madde26	0,839	1,239

BILOG MG programı ile yapılan MTK'ya göre a parametresi kestirimleri 2PL modelde 0,501 (madde 9) ile 2,007 (madde 17) arasında değişmektedir. 3PL modelde ise 1,010 (madde 9) ile 2,632 (madde 6) arasında değişmektedir.

MTK'ya göre BILOG programı kullanılarak elde edilen a parametreleri ile KTK'ya göre elde edilen ayırt edicilik katsayıları arasındaki ilişki Spearman Sıra Farkları korelasyon katsayıları hesaplanarak elde edilmiştir. KTK ayırt edicilik katsayıları ile BILOG programında 2PL modele göre kestirilen a parametresi arasında pozitif ve yüksek düzeyde ilişki bulunmuştur ($r=0,930$). Değişkenler arasındaki korelasyon Fisher'in z istatistiği ile test edilmiş ve 0,01 manidarlık düzeyinde anlamlı çıkmıştır. Maddelerin ayırt edicilik özelliğine ilişkin değerler KTK ve MTK'nın 2PL modeline göre yüksek düzeyde benzerlik göstermiştir.

MTK'ya göre BILOG programı ile elde edilen 3PL modele göre kestirilen a parametresi ile KTK'ya göre elde edilen ayırt edicilik katsayısı arasındaki ilişkiye bakıldığında ise ikisi

arasında pozitif yönde orta düzeyde manidar bir ilişki bulunmuştur ($r=0,639$; $p<0,01$). Maddelerin ayırt edicilik özelliğine ilişkin değerler KTK ve MTK 3PL modele göre orta düzeyde benzerlik göstermiştir.

2.2. Araştırma Sorusuna İlişkin Bulgular

“MULTILOG 7.03 programı ile TIMSS 2011 Türkiye 4.sınıf matematik alt testindeki her bir maddeye ait MTK'nin 2PL ve 3PL modeline göre kestirilen a parametreleri ve bu parametreler ile KTK'ye göre hesaplanan madde ayırt edicilik (r) katsayıları arasındaki ilişkinin düzeyi nedir?”

MULTILOG 7.03 programına göre maddelerin a parametreleri kestirilmiştir. 2PL ve 3PL modelde göre kestirilen a parametresi değerleri Tablo 13'te verilmiştir.

Tablo 13. MULTILOG 7.03 Programında Kestirilen a Parametreleri

Madde Numarası	2PLM için a parametresi	3PLM için a parametresi
Madde 1	1,84	1,07
Madde 2	0,91	1,03
Madde 3	0,64	1,47
Madde 4	1,70	0,98
Madde 5	1,33	0,80
Madde 6	0,66	3,19
Madde 7	1,14	1,78
Madde 8	0,62	0,77
Madde 9	0,45	0,38
Madde10	1,03	0,83
Madde11	1,25	0,77
Madde12	1,30	0,76
Madde13	1,71	1,05
Madde14	1,13	0,70
Madde15	1,65	1,02
Madde16	0,81	0,96
Madde17	2,10	1,26
Madde18	0,75	1,23
Madde19	1,77	1,07
Madde20	0,68	0,55
Madde21	1,33	1,00
Madde22	1,48	1,06
Madde23	0,54	0,83
Madde24	0,38	0,59
Madde25	1,13	0,69
Madde26	0,84	0,50

Tablo 13'te gösterilen MULTILOG 7.03 ile yapılan kestirimler incelendiğinde, MTK'ya göre a parametresi kestirimlerinin 2PL modelde 0,38 (madde 24) ile 2,10 (madde 17) arasında değişmekteği, 3PL modelde ise 0,38 (madde 9) ile 3,19 (madde 6) arasında değıştiğı gözlenmiştir.

KTK r katsayıları ile MULTILOG 7.03 programında 2PL modele göre kestirilen a parametreleri arasındaki korelasyon pozitif yönde yüksek düzeyde manidar ilişki bulunmuştur ($r=0,932$ $p< 0,01$). Değişkenler arasındaki korelasyon Fisher'in z istatistiğı ile test edilmiş ve 0,01 manidarlık düzeyinde anlamlı çıkmıştır. Maddelerin ayırt edicilik özelliğine ilişkin değerler KTK ve MTK'nın 2PL modeline göre yüksek düzeyde benzerlik göstermiştir.

KTK r katsayıları ile MULTILOG 7.03 programında 3PL modele göre kestirilen a parametreleri arasındaki korelasyon pozitif yönde orta düzeyde ve manidardır ($r=0,370$; $p< 0,01$). Değişkenler arasındaki korelasyon Fisher'in z istatistiğı ile test edilmiş ve 0,01 manidarlık düzeyinde anlamlı çıkmıştır. Maddelerin ayırt edicilik özelliğine ilişkin değerler KTK ve MTK'nın 3PL modeline göre orta düzeyde benzerlik göstermiştir.

2.3. Araştırma Sorusuna İlişkin Bulgular

“R programı ile TIMSS 2011 Türkiye 4.sınıf matematik alt testindeki her bir maddeye ait MTK'nın 2PL ve 3PL modeline göre kestirilen a parametreleri ve bu parametreler ile KTK'ye göre hesaplanan madde ayırt edicilik (r) katsayıları arasındaki ilişkinin düzeyi nedir ?”

Araştırmanın bu sorusunda R programına göre maddelerin a parametreleri incelenmiştir. 2PL ve 3PL modelden elde edilen a parametreleri aşağıdaki Tablo 14'te verilmiştir.

Tablo 14. R Programında Kestirilen a Parametreleri

Madde Numarası	2PLM için a parametresi	3PLM için a parametresi
Madde 1	1,845	1,818
Madde 2	0,915	1,742
Madde 3	0,638	2,562
Madde 4	1,702	1,658
Madde 5	1,336	1,354
Madde 6	0,659	5,253
Madde 7	1,142	2,968
Madde 8	0,627	1,306
Madde 9	0,452	0,646
Madde10	1,039	1,413
Madde11	1,259	1,301
Madde12	1,304	1,283
Madde13	1,713	1,768
Madde14	1,131	1,188
Madde15	1,656	1,725
Madde16	0,809	1,640
Madde17	2,105	2,126
Madde18	0,757	2,103
Madde19	1,779	1,813
Madde20	0,686	0,937
Madde21	1,337	1,695
Madde22	1,483	1,800
Madde23	0,542	1,416
Madde24	0,387	1,034
Madde25	1,131	1,167
Madde26	0,844	0,847

Tablo 14 incelendiğinde, R programı ile yapılan MTK'ya göre a parametresi kestirimlerinin 2PL modelde 0,387 (madde 24) ile 2,105 (madde 17) arasında değiştiği görülmektedir. 3PL modelde ise 0,646 (madde 9) ile 5,253 (madde 6) arasında değişmektedir.

Analiz sonuçlarına göre, KTK r katsayısı ile R programında 2PL modele göre kestirilen a parametresi arasında pozitif yönde yüksek düzeyde manidar ilişki bulunmuştur ($r=0,930$, $p<0,01$). Değişkenler arasındaki korelasyon Fisher'in z istatistiği ile test edilmiş ve 0,01 manidarlık düzeyinde anlamlı çıkmıştır. Maddelerin ayırt edicilik özelliğine ilişkin değerler KTK ve MTK'nin 2PL modeline göre yüksek düzeyde benzerlik göstermiştir.

KTK r katsayıları ile R programında 3PL modele göre kestirilen a parametreleri arasında pozitif yönde orta düzeyde manidar ilişki bulunmuştur ($r=0,361$ $p<0,01$). Değişkenler arasındaki korelasyon Fisher'in z istatistiği ile test edilmiş ve 0,01 manidarlık düzeyinde

anlamli çıkmıřtır. Maddelerin ayırt edicilik özelliđine iliřkin deđerler KTK ve MTK'nın 3PL modeline göre orta düzeyde benzerlik göstermiřtir.

3. Arařtırma Sorusuna İliřkin Bulgular

“TIMSS 2011 Türkiye 4.sınıf matematik testi verilerinin 1PL, 2PL, 3PL modellerden hangisine uyumu yüksektir?”

Bu alıřmada üç ayrı program kullanıldıđından, her bir programın hangi model ile daha iyi uyum sađladıđı arařtırılmıřtır. Bu amaçla programların 1PL, 2PL ve 3PL modeller ile uyumları incelenip karřılařtırılmıřtır. Model veri uyumları incelenirken MTK'nın her bir modeli için –2 log likelihood deđerleri elde edilmiř olup, bulunan deđerler negatiflik ile pozitifliklerine göre ve 0'dan uzaklařma derecelerine göre incelenmiřtir. Tablo 15'de her bir model için, farklı programlardan elde edilen –2 log likelihood deđerleri verilmiřtir.

Tablo 15. BILOG MG, R ve MULTILOG 7.03 Programlarında Lojistik Modellere Göre Hesaplanan –2 log likelihood Deđerleri

	1PL	2PL	3PL
BILOG MG	16102.7978	15832.7276	15898.5699
R	16097,016	15825,12	15755,56
MULTILOG 7.03	9258,4	8987,0	8917,9

Tablo 15 incelendiđinde BI LOG MG programı için en düşük –2 log likelihood deđerinin 2PL için, R ve MULTILOG 7.03 programları için en düşük –2 log likelihood deđerinin 3PL model için elde edildiđi görölmüřtür. –2 log likelihood deđeri pozitif çıktıđında, deđerin 0'a yakınlařma derecesi model-veri uyumunu artırdıđından en uyumlu model BILOG programı için 2PL model, R ve MULTILOG 7.03 programları için 3PL model olmuřtur.

Ayrıca tek boyutlu MTK modellerin uyumunu test etmenin diđer bir yolu kay-kare testidir. Kay-kare testi için -2loglikelihood deđerlerinin farkı alınarak serbestlik derecesine bölünür ve fark kay-kare tablosundaki deđerden büyük çıkarsa fark anlamlıdır denir (Hambleton vd., 1991). Testin model veri uyumu üç farklı program için χ^2 ile incelendiđinde, Tablo 16'daki sonuçlar elde edilmiřtir.

Tablo 16. BILOG MG, R ve Programları için χ^2 Sonuçları

	1PL-2PL için χ^2	2PL-3PL için χ^2
BILOG MG	10,38	2,53
R	9,71	2,68
MULTILOG 7.03	10,43	2,66

Kay-kare tablo değeri $sd = 26$ için, 0,05 anlamlılık düzeyinde 35,712 dir. Her üç programa göre 0,05 anlamlılık düzeyinde modeller arasındaki fark manidar değildir. -2loglikelihood değerine göre model veri uyumları farklılaşsa da, bu farklar manidar çıkmamıştır. Test için 1PL, 2PL ve 3PL modeller kullanılabilir.

4. Araştırma Sorusuna İlişkin Bulgular

4.1. Araştırma Sorusuna İlişkin Bulgular

“Klasik test Kuramına göre hesaplanan Cronbach- α güvenirlik katsayısı ve MTK’ya göre kestirilen marjinal ve görgül güvenirlik katsayıları nedir?”

TIMSS verilerinin KTK’ya göre güvenirliğini elde etmek amacıyla SPSS 21.0 programı aracılığı ile Cronbach alfa iç tutarlık katsayısı hesaplanmış ve 0,858 bulunmuştur. Bu değere göre testten elde edilen puanların KTK’ya göre güvenilir olduğu sonucuna varılabilir.

MTK’ya göre güvenirlik katsayıları hesaplanırken ise BILOG MG, R, MULTILOG 7.03 programlarında her bir model için ayrı güvenirlik katsayısı kestirimleri yapılmıştır. BILOG MG programında görgül, MULTILOG 7.03 ve R programlarından marjinal güvenirlik katsayıları elde edilmiş olup, bulunan katsayılar aşağıdaki Tablo 17’de verilmiştir.

Tablo 17. Görgül ve Marjinal Güvenirlik Katsayıları

	1PL	2PL	3PL
BILOG	0,835	0,863	0,890
MULTILOG 7.03	0,826	0,849	0,857
R	0,829	0,851	0,858

Tablo 17’de görüldüğü üzere MTK’ya göre incelenen güvenirlik katsayıları her model için KTK’ya göre kestirilen Cronbach alfa iç tutarlık katsayısına (0,833) yakın sonuçlar vermiştir. Analiz sonuçlarına göre her üç program 1PL, 2PL ve 3PL model için yüksek güvenirlik katsayısı vermiştir. En düşük güvenirlik katsayısı MULTILOG 7.03 programında 1PL için (0,826), en yüksek güvenirlik katsayısı ise BILOG programında 3PL model için (0,890) elde edilmiştir. Her üç model için de en yüksek güvenirlik katsayısını BILOG programı vermiştir.

4.2. Araştırma Sorusuna İlişkin Bulgular

“BILOG MG, R ve MULTILOG 7.03 programlarında 1PL, 2PL ve 3PL model için kestirilen görgül ve marjinal güvenirlik katsayıları arasında anlamlı fark var mıdır?”

Bu araştırma sorusunda, bulunan güvenilirlik katsayıları arasındaki farkın manidarlığı Feldt Testi (Alsawalmeh ve Feldt, 1994) ile sınanmıştır. BILOG MG, MULTILOG 7.03 ve R programlarında 1PL, 2PL ve 3PL modellere göre hesaplanan güvenilirlik katsayıları için W değerleri hesaplanmış, bulunan W değerleri, $F_{(542,542)} = 1,152$ ile karşılaştırılmıştır ($p=0,05$). Her üç program için ayrı ayrı hesaplanan W değerlerine göre 1PL modele göre hesaplanan güvenilirlik katsayıları ile Cronbach Alfa güvenilirlik katsayısı arasındaki fark manidar değildir ($W < 1,152$).

2PL modele göre sadece BILOG programında hesaplanan güvenilirlik katsayısı ile Cronbach Alfa katsayısı arasında manidar fark vardır. 3PL modele göre ise, her üç programda hesaplanan güvenilirlik katsayıları ile Cronbach Alfa katsayısı arasında manidar fark vardır ($W > 1,152$).

5. Araştırma Sorusuna İlişkin Bulgular

“TIMSS 2011 Türkiye 4.sınıf matematik alt testi verilerinin R, BILOG MG, MULTILOG 7.03 programlarıyla 1PL, 2PL ve 3PL modellere göre yapılan analizler sonucunda elde edilen a,b ve c parametrelerinin standart hataları arasında anlamlı farklılık var mıdır?

Beşinci araştırma sorusunda BILOG, MULTLOG ve R programlarında kestirilen 1PL modelin b parametrelerine ait standart hatalar Friedman testi ile analiz edilerek, aralarında anlamlı farklılık olup olmadığına bakılmıştır. SPSS 21.0 analizi sonucu Tablo 18’de verilmiştir.

Tablo 18. 1PL Model İçin BILOG MG, R ve MULTILOG 7.03 Programlarında Kestirilen b Parametresi Standart Hataları için Friedman Testi Sonucu

	Sıralama Ortalamaları	Sd	χ^2	p
R	1,96	2	44,087	0,000
BILOG	1,04			
MULTILOG	3,00			

Friedman testi sonucuna göre manidar fark bulunmuştur ($p < 0,05$). Bu sonuç, maddelerin standart hataları arasında programlara göre farklılık olduğunu göstermektedir.

Hangi programlar arasında farklılık olduğu SPSS 21.0 programında, söz dizimi yardımı ile Friedman testinde ikili karşılaştırmalar yapılarak incelenmiştir. 1PL model için R ile BILOG programlarından elde edilen b parametrelerine ait standart hatalar arasında manidar farklılık vardır ($p < 0,05$). R programından elde edilen standart hataların sıra ortalaması,

BILOG programından elde edilen standart hata sıra ortalamasından daha yüksektir. Ayrıca BILOG ve R prgramları için Wilcoxon İşaretli Sıralar Testi ile Z değeri elde edilerek etki büyüklüğü hesaplanmıştır. $Z=-4,318$ ve $n=26$ için etki büyüklüğü $r=-0,85$ bulunmuştur. Bulunan değer etki büyüklüğünün büyük olduğunu göstermektedir (Cohen,1988). R programından elde edilen standart hatalar ile BILOG programından elde edilen standart hatalar arasındaki fark istatistiksel olarak önemli kabul edilebilecek düzeyde ve anlamlı bir farktır.

R ile MULTILOG 7.03 programlarından elde edilen standart hatalar arasında manidar farklılık vardır ($p<0,05$). R programında elde edilen 1PL modelin b parametresine ait standart hataların sıra ortalaması, MULTILOG 7.03 programından elde edilen standart hata sıra ortalamasından daha düşüktür. Etki büyüklüğü Wilcoxon İşaretli Sıralar Testi ile incelenmiş olup $Z=-4,458$ için $r=-0,87$ bulunmuştur. R programından elde edilen standart hatalar ile MULTILOG 7.03 programından elde edilen standart hatalar arasındaki fark önemli ve anlamlıdır.

BILOG ile MULTILOG 7.03 programlarından elde edilen standart hatalar arasında da manidar farklılık vardır ($p<0,05$). MULTILOG 7.03 programından elde edilen standart hata sıra ortalaması, BILOG programından elde edilen standart hata sıralama ortalamasından yüksektir. Etki büyüklüğü için Wilcoxon İşaretli Sıralar Testi ile Z değeri elde edilmiştir. $Z= -4,466$ için etki büyüklüğü $r=-0,88$ bulunmuştur. BILOG'dan elde edilen standart hatalar ile MULTILOG 7.03'dan elde edilen standart hatalar arasındaki fark önemli kabul edilebilecek derecede büyük ve anlamlıdır.

Bu bulgulara göre 1PL model için, b parametresi kestirimlerinin BILOG programında daha düşük hata ile kestirilebileceği söylenebilir. Ayrıca üç program arasından en düşük standart hata ortalamasını veren program BILOG'dur (0,103).

2PL modelin b parametrelerine ait standart hatalar üç ayrı programa göre Friedman testi ile analiz edilerek, aralarında anlamlı farklılık olup olmadığı incelenmiştir. SPSS çıktısı Tablo 19'da verilmiştir:

Tablo 19. 2PL Model İçin BILOG MG, R ve MULTILOG 7.03 Programlarında Kestirilen b Parametresi Standart Hataları için Friedman Testi Sonucu

	Sıralama Ortalamaları	Sd	χ^2	p
R	2,00	2	36,364	0,000
BILOG	1,09			
MULTILOG	2,91			

Tablo 19’da verilen Friedman testi sonucuna göre R, BILOG ve MULTILOG 7.03 programlarında kestirilen 2PL modelin b parametrelerine ait standart hatalar arasındaki fark manidardır ($p < 0,05$). Programlardan elde edilen standart hataların sıralamaları arasında anlamlı fark vardır.

SPSS 21.0 programında, söz dizimi yardımı ile Friedman testinde ikili karşılaştırmalar yaparak hangi programlar arasında manidar fark olduğu belirlenmiştir. R ile BILOG programlarından 2PL modele göre kestirilen b parametrelerine ait elde standart hatalar arasında manidar farklılık vardır ($p < 0,05$). R programından elde edilen standart hataların sıra ortalaması, BILOG programından elde edilen standart hata sıra ortalamasından daha yüksektir. Etki büyüklüğünü hesaplamak için Wilcoxon İşaretli Sıralar Testi ile Z değeri elde edilmiş olup, $Z = -4,381$ için $r = -0,86$ bulunmuştur. R programından elde edilen standart hatalar ile BILOG programından elde edilen standart hatalardan arasındaki fark istatistiksel olarak önemli ve anlamlıdır.

R ile MULTILOG 7.03 programlarından elde edilen standart hatalar arasında da manidar farklılık vardır ($p < 0,05$). R programından elde edilen standart hataların sıra ortalaması, MULTILOG 7.03 programından elde edilen standart hata sıra ortalamasından daha düşüktür. Etki büyüklüğü Wilcoxon İşaretli sıralar Testi ile incelenmiş olup, $Z = -4,457$ için $r = -0,87$ bulunmuştur. R programından elde edilen standart hatalar ile MULTILOG 7.03 programından elde edilen standart hatalar arasındaki fark önemli ve anlamlıdır.

BILOG ile MULTILOG 7.03 programlarından elde edilen standart hatalar arasında da manidar farklılık vardır ($p < 0,05$). MULTILOG 7.03 programından elde edilen standart hata sıra ortalaması, BILOG programından elde edilen standart hata sıralama ortalamasından yüksektir. Wilcoxon İşaretli Sıralar Testi ile Z değeri $-4,458$ olarak bulunmuş etki büyüklüğü r değeri ise $-0,87$ bulunmuştur. BILOG programından elde edilen standart hatalar ile MULTILOG 7.03 programından elde edilen standart hatalar arasındaki farkın önemli derecede büyük ve anlamlı olduğu söylenebilir. Bulgulara göre 2PL model

için, b parametresi kestirimleri, en düşük standart hata ortalaması ile R programından elde edilmiştir (0,109).

Daha sonra 3PL model için b parametrelerine ait standart hatalar MULTILOG 7.03 programı 3PL modele göre kestirilen parametrelere ait standart hataları vermediğinden, BILOG ve R programlarından elde edilmiş, elde edilen standart hatalar arasında anlamlı farklılık olup olmadığı Friedman testi ile incelenmiştir. SPSS çıktısı Tablo 20’de verilmiştir:

Tablo 20. 3PLM için BILOG MG ve R Programlarında Kestirilen b Parametresi Standart Hataları için Friedman Testi Sonucu

	Sıralama Ortalamaları	Sd	χ^2	p
R	1,00	1	26,00	0,000
BILOG	2,00			

Tablo 20’de gösterilen Friedman testi sonucuna göre R ve BILOG programlarından elde edilen 3PL modelin b parametrelerinin standart hata kestirimleri arasında manidar fark vardır ($p < 0,05$). 3PL modele göre kestirilen b parametrelerinin standart hataların sıra ortalaması, R programında BILOG programından daha düşük çıkmıştır. Etki büyüklüğü ise Wilcoxon İşaretli Sıralar Testi ile incelenmiş olup Z değeri -4,457, etki büyüklüğü r değeri ise -0,87 olarak bulunmuştur. Bu sonuca göre R programından elde edilen standart hata değerleri ile BILOG programından elde edilen standart hata arasındaki fark önemli derecede büyük ve anlamlıdır.

2PL modelin a parametrelerine ait standart hatalar üç programdan elde edilmiş, elde edilen standart hatalar arasında anlamlı farklılık olup olmadığı Friedman testi ile sınanmıştır. SPSS çıktısı Tablo 21’de verilmiştir:

Tablo 21. 2PL Model için BILOG MG, R ve MULTILOG 7.03 Programlarında Kestirilen a Parametresi Standart Hataları için Friedman Testi Sonucu

	Sıralama Ortalamaları	Sd	χ^2	p
R	1,00	2	52,00	0,000
BILOG	2,00			
MULTILOG	3,00			

Friedman testi sonucuna göre R, BILOG ve MULTILOG 7.03 programlarından elde edilen 2PL modelin a parametrelerinin standart hata kestirimleri arasında manidar fark vardır.

Friedman testi ile ikili karşılaştırmalar yaparak hangi programlar arasında farklılık olduğu incelenmiştir. R ile BILOG programlarından elde edilen standart hatalar arasında manidar farklılık vardır ($p<0,05$). R programından elde edilen standart hataların sıra ortalaması, BILOG programından elde edilen standart hata sıra ortalamasından daha düşüktür. Etki büyüklüğü için Wilcoxon İşaretli Sıralar Testi uygulanmış olup, $Z=-4,457$ için etki büyüklüğü $r=-0,87$ bulunmuştur. R programından elde edilen standart hata değerleri ile BILOG programından elde edilen standart hata arasındaki fark önemli kabul edilebilecek düzeyde ve anlamlıdır.

R ile MULTILOG 7.03 programlarından elde edilen standart hatalar arasında da manidar farklılık vardır ($p<0,05$). R programında elde edilen 1PL modelin b parametresine ait standart hataların sıra ortalaması, MULTILOG 7.03 programından elde edilen standart hata sıra ortalamasından daha düşüktür. Etki büyüklüğü Wilcoxon İşaretli Sıralar Testi ile incelenmiş olup $Z=-4,457$ için etki büyüklüğü değeri $r=-0,87$ bulunmuştur. R programından elde edilen standart hata değerleri ile MULTILOG 7.03 programından elde edilen standart hata değerleri arasındaki farkın önemli ve anlamlı olduğu söylenebilir.

BILOG ile MULTILOG 7.03 programlarından elde edilen standart hatalar arasında da manidar farklılık vardır ($p<0,05$). MULTILOG 7.03 programından elde edilen standart hata sıra ortalaması, BILOG programından elde edilen standart hata sıralama ortalamasından yüksektir. Etki büyüklüğü Wilcoxon İşaretli Sıralar Testi ile incelenmiş olup $Z=-4,465$ için etki büyüklüğü değeri $r=-0,88$ bulunmuştur. BILOG programından elde edilen standart hata değerleri ile MULTILOG 7.03 programından elde edilen standart hata değerleri arasındaki fark büyük ve anlamlıdır. Bu bulgulara göre 2PL model için, a parametresi kestirimlerinin R programında daha düşük hata ile kestirilebileceği söylenebilir. Üç program arasından en düşük standart hata ortalamasını veren program R'dır (0,108).

3PL modelin a parametrelerine ait standart hatalar iki programdan elde edilmiş, elde edilen standart hatalar arasında anlamlı farklılık olup olmadığı Friedman ile incelenmiştir. SPSS çıktısı Tablo 22'de verilmiştir:

Tablo 22. 3PLM için BILOG MG, R Programlarında Kestirilen a Parametresi Standart Hataları için Friedman Testi Sonucu

	Sıralama Ortalamaları	Sd	χ^2	P
R	1,04	1	22,154	0,000
BILOG	1,96			

Tablo 22’de verilen Friedman testi sonucuna göre R ve BILOG programlarından elde edilen 3PL modelin a parametrelerinin standart hata kestirimleri arasında manidar fark vardır ($p<0,05$). R programına göre elde edilen sıra ortalaması, BILOG programından elde edilen sıra ortalamasından düşüktür. Etki büyüklüğü Wilcoxon İşaretli Sıralar Testi ile incelenmiş olup $Z=-3,924$ için etki büyüklüğü değeri $r=-0,77$ bulunmuştur. BILOG programından elde edilen standart hata değerleri ile R programından elde edilen standart hata değerleri arasındaki farkın önemli derecede büyük ve anlamlı olduğu söylenebilir. Ayrıca R programından elde edilen standart hata ortalaması 0,233 olarak bulunmuş olup, BILOG programından elde edilen standart hata ortalamasından düşüktür.

3PL modelin c parametrelerine ait standart hatalar iki programdan elde edilmiş, elde edilen standart hatalar arasında anlamlı farklılık olup olmadığı Friedman testi ile incelenmiştir. SPSS çıktısı Tablo 23’de verilmiştir.

Tablo 23. 3PL Model için BILOG MG ve R Programlarında Kestirilen c Parametresi Standart Hataları için Friedman Testi Sonucu

	Sıralama Ortalamaları	Sd	χ^2	P
R	1,00	1	26,00	0,000
BILOG	2,00			

Tablo 23’de gösterilen Friedman testi sonucuna göre R ve BILOG programlarından elde edilen 3PL modelin c parametrelerinin standart hata kestirimleri sıra ortalamaları arasında manidar fark vardır ($p<0,05$). Etki büyüklüğü Wilcoxon İşaretli Sıralar Testi ile incelenmiş olup $Z=-4,457$ için etki büyüklüğü değeri $r=-0,87$ bulunmuştur. BILOG programından elde edilen standart hata değerleri ile R programından elde edilen standart hata değerleri arasındaki fark önemli olmakla birlikte, anlamlı ve yüksektir. En düşük standart hata ortalamasını veren program ise R programı olmuştur (0,027).



BÖLÜM V

SONUÇ ve TARTIŞMA

Bu bölümde araştırmanın bulguları doğrultusunda elde edilen sonuçlara yer verilmiş ve bu sonuçlar diğer çalışmaların sonuçlarıyla karşılaştırılarak tartışılmıştır. Araştırma soruları için elde edilen sonuçlar sırasıyla aşağıda verilmiştir.

1. Araştırma Sorusuna İlişkin Sonuçlar

Bu bölümde birinci araştırma sorusuna ilişkin sonuçlar literatürle desteklenerek tartışılmıştır. 1. araştırma sorusuna ilişkin sonuçlar literatürle desteklenerek tartışılmıştır. KTK'ya göre elde edilen madde güçlük katsayılarının hesaplanması sonucunda testin orta güçlükte olduğu sonucuna varılmıştır. Atar (2011), çalışmasında TIMSS 2007 sınavından bir matematik kitapçığı seçilerek testin ortalama güçlüğü'nün orta düzeyde olduğu belirtilmiştir.

1.1. Araştırma Sorusuna İlişkin Sonuçlar

KTK p katsayısı ile BILOG programında 1PL, 2PL ve 3PL modele göre kestirilen b parametreleri arasında negatif yönde yüksek düzeyde manidar ilişki bulunmuş olup bu sonuç daha önce yapılan KTK'ya göre kestirilen madde istatistik ve MTK'ya göre kestirilen parametrelerin karşılaştırıldığı çalışmalar ile uyum göstermektedir. Daha önce yapılan çalışmalarda Gelbal (1994), Fan (1998), Stage (1998), Kelecioğlu (2001), Çıkrıkçı Demirtaşlı (2002), Nartgün (2002), Çelen (2008) çok kategorili ya da ikili puanlanan ölçeklerde KTK'ya göre kestirilen p katsayıları ile MTK'ya göre kestirilen b parametreleri arasında negatif yönde yüksek korelasyon bulunmuşlardır. b parametresi ile p istatistiğinin ilişkisinin negatif yönlü olmasının nedeni, b parametresinde yüksek değerin, p istatistiğinde ise düşük değerin maddenin zor olduğunu göstermesidir (Çelen, 2008). Bireyler maddelere daha yüksek kategoride tepki verdikçe, KTK'ya göre maddelerin ölçtükleri özellik

düzeyleri değeri olan madde puanı ortalaması artarken, MTK'ya göre maddelerin ölçtükleri özellik düzeyinin bir göstergesi kabul edilen sınır konum parametreleri değerinin ortalaması azalmaktadır. Yani iki değişken arasındaki negatif yöndeki ilişki iki kurama göre kestirilen değerlerin birbirine benzediği şeklinde yorumlanabilir.

1.2. Araştırma Sorusuna İlişkin Sonuçlar

MULTILOG 7.03 programı ile yapılan b parametresi kestirimleri ile KTK'ya göre elde edilen p katsayıları arasındaki korelasyonlar negatif yönde çıkmaktadır. En yüksek korelasyon ise 1PL ve 2PL modeller ile elde edilmiştir. Elde edilen bulgular daha önce yapılan KTK ve MTK'ya göre karşılaştırılma yapılan çalışmalar ile uyumludur. Bulgularla uyumlu olarak Gelbal (1994), Fan (1998) ve Çıkrıkçı Demirtaşlı (2002)'nin çalışmalarında 1PL modele göre kestirilen b parametresi ile p katsayısı arasında yüksek korelasyon bulunmuştur. Courville (2004), Fan (1998), Stage (1998), Kelecioğlu (2001), Nartgün (2002) ve Çelen (2008) ve Progar ve Sočan (2008)'in çalışmalarında 2PL modele göre kestirilen b parametreleri ile KTK'ya göre hesaplanan p katsayısı arasındaki ilişki negatif yönde ve yüksek bulunmuştur. KTK p katsayısı ile MTK b parametresi arasındaki korelasyonun çok yüksek çıkması Fan (1998) tarafından, iki kuramın neredeyse aynı madde güçlük bilgisi sağlaması olarak açıklanmıştır.

1.3. Araştırma Sorusuna İlişkin Sonuçlar

R programı ile 1PL, 2PL ve 3PL modele göre kestirilen b parametreleri ile KTK'ya göre hesaplanan p katsayıları birbirine yakın sonuçlar vermiştir. p katsayısı ile b parametresi arasında negatif yönde yüksek korelasyonlar elde edilmiştir. 1PL modele göre elde edilen bparametreleri ile KTK'ya göre elde edilen b katsayıları arasındaki korelasyon ise 1'e yakın çıkmıştır. Bu sonuç Gelbal (1994), Fan (1998) ve Çıkrıkçı Demirtaşlı (2002)'nin çalışmaları ile uyumludur. 2PL ve 3PL modele göre kestirilen b parametresi ile p katsayıları arasında negatif yönde yüksek ve manidar korelasyon bulunmuştur. Bulgular Courville (2004), Fan (1998), Stage (1998), Kelecioğlu (2001), Nartgün (2002) ve Çelen (2008) ve Progar ve Sočan (2008)'in çalışmaları ile uyumludur. p katsayısı ile b parametreleri a arasındaki negatif yönde yüksek düzeyde ilişki iki kurama göre kestirilen değerlerin birbirine benzediği şeklinde yorumlanmıştır.

2. Araştırma Sorusuna İlişkin Sonuçlar

KTK'ya göre elde edilen ayırt edicilik katsayıları ortalaması BILOG MG ve MULTILOG 7.03 programlarında 2PL ve 3PL'ye göre elde edilen a parametrelerinden oldukça düşüktür. R programına göre elde edilen a parametreleri ortalamasından ise yüksektir. Sonuç olarak KTK ve MTK 2PL modeline göre kestirilen madde ayırt edicilik güçlerinin korelasyonlarının daha yüksek olduğu görülmüştür. Bu bulgu alanyazında daha önce yapılan Kelecioğlu (2001), Courville (2004), Kan (2006), Progar ve Sočan (2008) ve Güler, Uyanık ve Teker (2013)'ün çalışmaları ile uyumludur.

2.1. Araştırma Sorusuna İlişkin Sonuçlar

BILOG MG programı ile yapılan analiz sonucunda KTK r katsayısı ile MTK a parametresi arasındaki ilişki 2PL modele göre pozitif yönde ve yüksek düzeyde manidar ilişki vermektedir. Fan (1998), Kelecioğlu (2001), Nartgün (2002), Stage(2001), Courville (2004) ve Kan (2006)'nın çalışmalarında da r katsayısı ile a parametresi arasında pozitif yönde yüksek korelasyon bulunmuştur.

Ayrıca çalışmada 3PL modele göre kestirilen a parametreleri ile r katsayısı arasındaki korelasyonu incelenerek, pozitif yönde ve orta düzeyde ilişki saptanmıştır. Bu sonuç ikili ya da çoklu puanlanan ölçme araçları ile KTK ve MTK çerçevesinde madde istatistik ve parametre karşılaştırmalarının yapıldığı çalışmaların bulgularıyla uyumludur. Genellikle, MTK'nın 3PL modeline göre kestirilen a parametresi ile KTK ya göre hesaplanan ayırt edicilik indekslerinin korelasyonu, 2PL modele göre kestirilen a parametreleri ile elde edilen korelasyondan düşük çıkmaktadır (Fan,1998). Fan (1998), Stage (2001), McDonald ve Paunonen (2002), Çelen (2008), Güler, Uyanık ve Teker (2014) 'in çalışmalarında 3PL modele göre kestirilen a parametreleri ile KTK'ya göre hesaplanan r katsayıları arasındaki ilişki pozitif ve orta düzeyde bulunmuştur.

2.2. Araştırma Sorusuna İlişkin Sonuçlar

MULTILOG 7.03 programında 2PL modele göre elde edilen a parametreleri ile KTK'ya göre hesaplanan r parametreleri ile korelasyon pozitif yönde, yüksek ve manidar ilişki bulunmuştur. Fan (1998), Kelecioğlu (2001), Nartgün (2002), Courville (2004) ve Kan (2006)'nın çalışmalarında da r katsayısı ile a parametresi arasında pozitif yönde ve yüksek korelasyon bulunmuştur.

MULTILOG 7.03 programından elde edilen bulgulara göre 3PL model için kestirilen a parametresi ile r katsayıları orta düzeyde korelasyon vermektedir. MULTILOG 7.03 programına göre en iyi model veri uyumu 3PL için sağlanmış olmasına rağmen, KTK madde istatistikleri ile korelasyonu 2PL modelden düşük çıkmıştır. Güler, Uyanık ve Teker (2014)'in çalışmasında benzer şekilde 3PL modelin model veri uyumu yüksek olmakla birlikte en düşük korelasyon değerleri 3PL model için bulunmuş ve a parametreleri ile r katsayıları arasında orta düzeyde ilişki saptanmıştır. Benzer şekilde McDonald ve Paunonen (2002) ile Çelen (2008) çalışmalarında da orta düzeyde pozitif yönde korelasyon bulunmuştur.

2.3. Araştırma Sorusuna İlişkin Sonuçlar

R programı ile 2PL modele göre elde edilen parametrelerin KTK'ya göre elde edilen r parametreleri ile korelasyonu incelenerek, pozitif yönde yüksek düzeyde manidar ilişki bulunmuştur. Bu bulgu Fan (1998), Kelecioğlu (2001), Nartgün (2002), Courville (2004) ve Kan (2006)'nın çalışmaları ile uyumludur. Bu çalışmalarda da r katsayısı ile a parametresi arasında pozitif yönde ve yüksek korelasyon bulunmuştur. KTK r parametresi ile R programında 3PL modele göre kestirilen a parametresi arasında ise pozitif yönde orta düzeyde manidar ilişki bulunmuştur. R programında en yüksek model veri uyumu 3PL ile sağlanmasına rağmen, KTK ile korelasyonu en düşük çıkan model de 3PL modelidir. Bu bulgu Güler, Uyanık ve Teker (2014)'ün çalışması ile benzerlik göstermiştir. Bulgularla uyumlu olarak, McDonald ve Paunonen (2002), Çelen (2008) çalışmalarında 3PL modele göre kestirilen a parametreleri ile KTK'ya göre hesaplanan r katsayıları arasında pozitif ve orta düzeyde ilişki bulunmaktadır.

3. Araştırma Sorusuna İlişkin Sonuçlar

Araştırma sorusunda model veri uyumu incelenmiş olup, BILOG programı ile R ve MULTILOG 7.03 programlarına göre model veri uyumu farklılaşmaktadır ancak modeller arası uyum farkları manidar değildir.

MULTILOG 7.03 ve R programlarına göre en uygun model 3PL model olarak belirlenmiştir. Özer Özkan (2012)'nin çalışmasında Öğrenci Başarılarının Belirlenmesi Sınavından (ÖBBS) elde edilen verilerin model uyumunun incelemesi sonucu 2PL modelin 1PL modele göre, 3PL modelin ise 2PL modele göre daha yüksek uyum verdiği görülmüştür. Ayrıca bulgu, Ndalichako ve Rogers (1997), Stage (1998), Anderson (1999),

Çıkrıkçı Demirtaşlı (2002), Courville (2004), Gültekin ve Demirtaşlı (2012), Çelen ve Aybek (2013) ile Güler, Uyanık ve Teker (2014)'in çalışmalarını desteklemektedir. Çakıcı Eser (2013) PISA verileriyle yaptığı çalışmada BILOG programı kullanmış ve model veri uyumunu en çok 2PL modelin kullandığı sonucuna ulaşmıştır. Bu çalışmada da BILOG programına göre en uygun model 2PL model olmuştur. Aynı veri ile farklı uyum sonuçlarının elde edilmesi tartışmalı bir sonuç olmakla birlikte, programların farklı formüller ile hesaplamalar yapması bu durumun sebebi olarak görülmüştür.

4. Araştırma Sorusuna İlişkin Sonuçlar

4. araştırma sorusuna ilişkin sonuçlar literatürle desteklenerek tartışılmıştır.

4.1. Araştırma Sorusuna İlişkin Sonuçlar

Ölçümlerin güvenirliği KTK'ya göre Cronbach alfa iç tutarlık katsayısı ile hesaplanmış olup, bu sonuca göre testten alınan puanların KTK'ya göre güvenilir olduğu sonucuna varılmıştır. MTK'ya göre incelenen güvenirlik katsayıları her model için KTK'ya göre kestirilen Cronbach alfa iç tutarlık katsayısına yakın sonuçlar vermiştir. En yüksek güvenirlik katsayıları her üç model için BILOG programı ile elde edilmiştir. KTK ve MTK kuramlarına göre kestirilen güvenirlik katsayıları Stage (1998), Nartgün (2002), Çelen ve Aybek (2013)' in çalışmalarını desteklemektedir. Bu çalışmalarda da her iki kuram için elde edilen güvenirlik katsayıları birbirine yakın sonuçlar vermiştir.

4.2. Araştırma Sorusuna İlişkin Sonuçlar

Analizler sonucunda her üç programda 1PL, 2PL ve 3PL model için yüksek ve benzer güvenirlik katsayıları elde edilmiştir. Elde edilen güvenirlik katsayıları ile KTK'ya göre hesaplanan Cronbach alfa güvenirlik katsayısı arasındaki farkın manidarlığı Feldt testi ile sınanmıştır. 1PL model için güvenirlik katsayıları arasında anlamlı fark olmadığı, 2PL için BILOG programı ile kestirilen ve 3PL model için her üç programda kestirilen güvenirlik katsayıları ile Cronbach alfa güvenirlik katsayısı arasında manidar fark olduğu sonucuna varılmıştır. Literatürde böyle bir karşılaştırmaya rastlanmamıştır.

5. Araştırma Sorusuna İlişkin Sonuçlar

Bulgulara göre BILOG MG ve R programlarının, MULTILOG 7.03 programına göre daha düşük standart hata ile çalıştığı söylenebilir. 1PL, 2PL ve 3PL model için kestirilen

parametrelere ait olan standart hatalar arasında manidar fark bulunmuştur. 1PL ve 2PL model için kestirilen parametreler için en yüksek standart hata değerleri MULTILOG 7.03 programı ile elde edilmiştir. MUTILOG programının 3PL modelde c parametresine ait standart hataları vermemesi programın bir sınırlılığıdır. 3PL model için parametre kestirimlerinin standart hataları BILOG ve R programları için karşılaştırılmıştır. R programı ile elde edilen standart hatalar ile BILOG programından elde edilen standart hatalar arasında anlamlı fark vardır ve BILOG programı için standart hata ortalaması daha yüksektir. Literatürde farklı programlardan elde edilen parametrelere ait standart hataların karşılaştırıldığı bir çalışmaya rastlanmamıştır.

Özet olarak, incelenen araştırma sorularının sonuçlarına göre KTK ve MTK'nın 1PL, 2PL ve 3PL modellerine göre madde güçlük katsayısı ile b parametresi arasında ise negatif yönde yüksek korelasyon bulunmuştur. Ayırt edicilik katsayıları ve a parametreleri arasında 1PL ve 2PL için pozitif yönde yüksek; 3PL için pozitif yönde orta düzeyde korelasyon bulunmuştur. Genel olarak a parametresi ve r katsayısı korelasyonu için 3PL model 2PL modele göre daha düşük korelasyon vermektedir. Ayrıca iki kurama göre, ayırt edicilik değerleri madde güçlük değerlerine göre daha düşük korelasyon vermektedir (Fan,1988). Model veri uyumu incelendiğinde, BILOG programının R ve MULTILOG 7.03 programlarından farklı modele daha çok uyum gösterdiği ancak modeller arası uyum farkının üç program için de manidar olmadığı görülmüştür. KTK ile MTK 3PL modele göre kestirilen güvenirlik katsayıları arasında her üç programa göre manidar fark bulunmuştur. Standart hatalar karşılaştırıldığında ise her model için hesaplanan parametrelere ait en yüksek standart hatayı MULTILOG 7.03 programının verildiği tespit edilmiştir.

BÖLÜM VI

ÖNERİLER

Araştırmada incelenen alt problemlerin sonuçlarına göre MTK'ya göre kestirilen madde parametreleri ve KTK göre hesaplanan madde istatistikleri benzer sonuçlar vermiştir.

1. Çalışmada kullanılan BILOG MG, R ve MULTILOG 7.03 programları birbirine yakın güvenilirlik katsayısı vermiştir. Ancak MULTILOG 7.03 programı 3PL model için c parametrelerinin standart hata değerlerini vermemektedir. Bu sebeple 3PL modele göre çalışacak araştırmacılara BILOG MG veya R programı ile çalışmaları önerilebilir.

2. BILOG MG ve MULTILOG 7.03 yazılım ve veri formatları olarak benzemekte, kullanım kolaylığı açısından iki program da araştırmacılara önerilebilir. BILOG MG, R ve MULTILOG 7.03 programları madde parametre kestirimlerinde sıkça kullanılan programlardır. BILOG MG programı hem KTK hem de MTK'ya göre kestirimler yapmaktadır (Phase 1 KTK'ya göre yapılan kestirimleri vermektedir). MULTILOG 7.03 ve R programında ise KTK'ya göre kestirim yapılamamaktadır. Ayrıca R programı farklı paket programları ile çalışmakta, diğer programlara göre daha çok vakit gerektiren bir programdır.

3. BILOG MG sadece işaretleme ve seçim yönergesi ile çalışan bir programdır. MULTILOG 7.03 ve R ise söz dizimi komutları bilinmesi gereken programlardır. Bu yönden araştırmacılara BILOG MG kullanmaları önerilebilir.

Gelecekte Yapılacak Araştırmalar İçin Öneriler

Bu bölümde araştırmanın bulgularına dayalı olarak, araştırmanın kapsamına dahil edilmeyen ve diğer araştırmalara dahil edilebilecek hususlar konusunda araştırmacılara bazı önerilerde bulunulmuştur.

1. Arařtırmada BILOG MG, MULTILOG 7.03 ve R programları kullanılmıřtır. MTK kestirimleri yapılabilen diğerk programlar (PARSCALE, CONQUEST, LOGIST vs.) kullanılarak karřılařtırma yapılabilir.
2. Arařtırmada MTK tek boyutlu modelleri olan 1PL, 2PL ve 3PL modelleri incelenmiřtir. Çok boyutlu MTK modelleri ile kestirimler yapılarak lojistik modeller ve çok boyutlu modeller arasında karřılařtırma yapılabilir.
3. Arařtırmada MTK madde parametreleri MML kestirim yöntemine dayalı olarak yapılmıřtır. Diğerk kestirim yöntemleri (Joint Maximum Likelihood, Marginal Maximum Likelihood Estimation, Bayesian, Markov Chain Monte Carlo) kullanılarak kestirimler yapılabilir. Elde edilen kestirimler arasında manidar fark olup olmadıđı incelenebilir.
4. Çok kategorili maddeler ile alıřacak arařtırmacılar MULTILOG 7.03 ve R programları önerilebilir. alıřmada ikili puanlanan oktan semeli sorular kullanılmıřtır. BILOG MG, MULTILOG 7.03 ve R programı ikili kategorili maddeler iin analiz yapmakta, ancak ok kategorili maddelerde analizlerinde BILOG MG kullanılamamaktadır.

KAYNAKÇA

- Ackerman, T.A. (1989). Unidimensional IRT calibration of compensatory and non-compensatory multidimensional items. *Applied Psychological Measurement*, 13(1), 113–127.
- Alsawalmeth, Y.M. & Feldt, L.S. (1994). A modification of Feldt's test of the equality of two dependent alpha coefficients. *Psychometrika*, 59(1), 49-57.
- Anastasi, A. (1988). *Psychological testing*. New York: McMillan.
- Anderson, J.O. (1999). Does complex analysis (IRT) pay any dividends in achievement testing?. *The Alberta Journal of Educational Research*, XLV, 344-352.
- Atar, B. (2011). Tanımlayıcı ve açıklayıcı madde tepki modellerinin TIMSS 2007 Türkiye matematik verisine uyarlanması. *Eğitim ve Bilim*, 36, 255-269.
- Baker, F. B. (2001). *The basics of item response theory* (2. baskı). USA: Eric Clearinghouse on Assessment Evaluation.
- Baker, F. B. & Kim, S. (2004). *Item response theory parameter estimation techniques* (2. baskı). New York: Marcel Dekker.
- Baykul, Y. (1979). *Örtük özellikler ve klasik test kuramları üzerine bir araştırma*. Doktora Tezi, Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: Pegem.
- Bock, R. D., Thissen, D. & Zimowski, M. F. (1997). IRT estimation of domain scores. *Journal of Educational Measurement*, 37(3), 197–211.
- Bock, R. D. & Aitkin, M. (1981). Marginal maximum likelihood estimation of item parameters: *Application of an EM algorithm*. *Psychometrika*, 46, 443-459.

- Büyüköztürk, Ş. (2006). *Sosyal bilimler için veri analizi el kitabı: istatistik, araştırma deseni SPSS uygulamaları ve yorum* (6. baskı). Ankara: Pegem.
- Büyüköztürk, Ş. (2010). *Sosyal bilimler için veri analizi el kitabı*. Ankara: Pegem.
- Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş., ve Demirel, F. (2014). *Bilimsel araştırma yöntemleri* (14. baskı). Ankara: Pegem.
- Childs, R. A., & Chen, W. H. (1999). Software note obtaining comparable item parameter estimates in MULTILOG and PARSCALE for two polytomous IRT models. *Applied Psychological Measurement*, 23(4), 371-379.
- Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Ohio, USA: Cengage Learning.
- Courville, G.T. (2006). *An empirical comparison of item response theory and classical test theory item/person statistics*. Doktora Tezi, The Office of Graduate Studies of Texas A&M University, Texas.
- Çelen, Ü. (2008). *Klasik test kuramı ve madde tepki kuramı yöntemleriyle geliştirilen iki testin psikometrik özelliklerinin karşılaştırılması*. Doktora Tezi, Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- Çelen, Ü. (2008). Klasik test kuramı ve madde tepki kuramı yöntemleriyle geliştirilen iki testin geçerlik ve güvenirliğinin karşılaştırılması. *Eğitim Online*, 7(3). 758-768.
- Çelen, Ü. & Aybek, E.C. (2013). Öğrenci başarısının öğretmen yapımı bir testle klasik test kuramı ve madde tepki kuramı yöntemleriyle elde edilen puanlara göre karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 4(2). 64-75.
- Çokluk, Ö., Şekercioğlu, G. & Büyüköztürk, Ş. (2012). *Sosyal bilimler için çok değişkenli istatistik SPSS ve LISREL uygulamaları* (2. Baskı). Ankara: Pegem.
- De Ayala, R.J. (1993). An introduction to polytomous item response theory models. *Measurement and Evaluation in Counseling and Development*, 5(3), 172-189.
- De Ayala, R.J. (2009). *The theory and practice of item response theory*. U.S.A: Guilford.
- De Mars, C. (2010). *Item response theory series in understanding statistic: measurement*. New York: Oxford University.

- Çıkrıkçı Demirtaşlı, N. (2002). A study of raven standard progressive matrices test's item measures under classic and item response models: an empirical comparison. *Ankara University, Journal of Faculty of Educational Sciences*, 35(1-2).
- Doğan, N. & Tezbasaran, A. (2003). Klasik test kuramı ve örtük özellikler kuramının örneklemeler bağlamında karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 58-67.
- Du Toit, M. (2003). *IRT from SSI: BILOG-MG, MULTILOG, PARSCALE, TESTFACT*. Lincolnwood, IL: Scientific Software International.
- Embretson, S.E. & Reise, S.P. (2000). *Item response theory for psychologists*. New Jersey: Lawrence Erlbaum Associates.
- Erkuş, A. (2003). *Psikometri üzerine yazılar*. Ankara: Türk Psikologlar Derneği Yayınları.
- Eser, D. Ç. (2013). Pısa 2009 okuma testinden elde edilen iki kategorili verilerin bılog programı ile incelenmesi. *Eğitim ve Öğretim Araştırmaları Dergisi*, 2(4). 135-144.
- Fan, X. (1998). Item response theory and classical test theory: An empirical comparison of their item/person statistics. *Educational and psychological measurement*, 58(3), 357-381.
- Field, A. (2009). *Discovering statistics using SPSS* (3. baskı). London: Sage.
- Gelbal, S. (1994). P madde güçlük indeksi ve Rasch modelinin b parametresi ve bunlara dayalı yetenek ölçüleri üzerine bir karşılaştırma. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*. 10, 85-94.
- Güler, N. (2008). *Klasik test kuramı genellenebilirlik kuramı ve Rasch modeli üzerine bir araştırma*. Doktora Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Güler, N., Teker G. & Uyanık G. (2013). Comparison of classical test theory in terms of item parameters. *European Journal of Research on Education*, 2(1), 1-6.
- Gültekin, S., & Çıkrıkçı-Demirtaşlı, N. (2012). Comparing the test information obtained through multiple-choice, open-ended and mixed item tests based on item response theory. *Elementary Education Online*, 11(1), 251-263.

- Hambleton, R.K. & Jones, R.W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38-47.
- Hambleton, R.K. & Swaminathan H. (1985). *Item response theory: principals and applications*. Norwell, MA: Kluwer.
- Hambleton, R.K., Swaminathan, H. & Rogers, H.J. (1991). *Fundamentals of item response theory*. California: Sage.
- Harris, D. (1989). Comparison of 1-, 2-, and 3-parameter IRT models. *Educational Measurement: Issues and Practice* 8(1), 35-41.
- Harvey, R., & Hammer, A. (1999). Item Response Theory. *The Counseling Psychologist*, 27(3), 353-383.
- İlk, Ö. (2011). *R Yazılımına Giriş* (1. baskı). Ankara: ODTÜ Geliştirme Vakfı.
- Kan, A. (2006). Klasik test teorisine ve örtük özellikler teorisine göre kestirilen madde parametrelerinin karşılaştırılması üzerine ampirik bir çalışma. *Mersin University Journal of the Faculty of Education* 2(6), 227-235.
- Karasar, N. (2013). *Bilimsel araştırma yöntemi* (25. baskı). Ankara: Nobel.
- Kelecioğlu, H. (2001). Örtük özellikler teorisindeki b ve a parametreleri ile klasik test teorisindeki p ve r istatistikleri arasındaki ilişki. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 20(2), 104-110.
- Kirisci, L., Hsu, T. & Yu, L. (2001). Robustness of item parameter estimation programs to assumptions of unidimensionality and normality. *Applied Psychological Measurement*, 25(2), 146-162.
- Kilmen, S. (2015). *Eğitim araştırmacıları için SPSS uygulamalı istatistik*. (1.baskı). Ankara: Edge Akademi.
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking*. New York: Springer.
- Lord, F. M. (1968). An analysis of the verbal scholastic aptitude test using Birnbaum's three-parameter logistic model. *Educational and Psychological Measurement*, 28(2), 989-1020.

- Lord, F. (1980). *Applications of item response theory to practical testing problems*. New Jersey: Lawrence Erlbaum.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- Lord F. M. ve Stocking, M.L. (1997). *Item Response Theory, Educational Research Methodology and Measurement: An International Handbook*. Oxford: Pergamon Press.
- Macdonald, P. & Paunonen, S. V. (2002). A monte carlo comparison of item and person statistics based on item response theory versus classical test theory. *Educational and Psychological Measurement*, 62, 921-943.
- Milli Eğitim Bakanlığı EARGED. (2003). *Üçüncü uluslararası matematik ve fen bilgisi çalışması: Ulusal rapor*. 21/05/14 tarihinde <http://yegitek.meb.gov.tr/earged/arasayfa.php?g=25> sayfasından erişilmiştir.
- Nartgün, Z. (2002). *Aynı tutumu ölçmeye yönelik Likert tipi ölçek ile metrik ölçeğin madde ve ölçek özelliklerinin klasik test kuramı ve örtük özellikler kuramına göre incelenmesi*. Doktora Tezi, Hacettepe Üniversitesi Sosyal Bilimleri Enstitüsü, Ankara.
- Nartgün, Z. (2002). Klasik test kuramındaki Cronbach alfa güvenirlik katsayısı ile örtük özellikler kuramındaki bilgi fonksiyonları ve marjinal güvenirlik katsayısının karşılaştırılması. *A.İ.B.Ü. Eğitim Fakültesi Dergisi*, 1(2), 143-161.
- Ndalichako, J. L. & Rogers, W. T. (1997). Comparison of finite state theory, classical test theory and item response theory in scoring multiple-choice items. *Educational and Psychological Measurement*, 57(4), 580-589.
- Ostini, R. & Nering, M. L. (Eds.). (2006). *Polytomous item response theory models*. U.S.A: Sage.
- Özçelik, D. A. (1981). *Okullarda Ölçme ve Değerlendirme*. Ankara: ÜSYM-Eğitim.
- Özdemir, D. (2004). Çoktan seçmeli testlerin klasik test teorisi ve örtük özellikler teorisine göre hesaplanan psikometrik özelliklerinin iki kategorili ve ağırlıklandırılmış puanlanması yönünden karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 26, 117-123.

- Özer, Y. & Anıl, D. (2011). Öğrencilerin fen ve matematik başarılarını etkileyen faktörlerin yapısal eşitlik modeli ile incelenmesi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 41, 313-324.
- Özer Özkan, Y. (2014). A comparison of estimated achievement scores obtained from student achievement assessment test utilizing classical test theory, unidimensional and multidimensional IRT. *International Journal of Human Sciences*, 11(1), 20-44.
- Progar, S. & Sočan, G. (2008). An empirical comparison of item response theory and classical test theory, *Horizons of Psychology*, 17(3), 5–24.
- Rizopoulos, D. (2012). "Package 'ltm'". 02/06/14 tarihinde <http://wiki.r-project.org/rwiki/doku.php> sayfasından erişilmiştir.
- Robitaille, D. F. & Beaton, A. E. (2002). *Secondary analysis of the TIMSS data*. Netherlands: Kluwer Academic.
- Rupp, A. A. (2003). Item response modeling with BILOG-MG and MULTILOG for Windows. *International Journal of Testing*, 3(4), 365-384.
- Silvestre-Tipay, J. L. (2009). Item response theory and classical test theory: an empirical comparison of item/person statistics in a biological science test. *The International Journal of Educational and Psychological Assessment*, 1(1), 19-31.
- Schermelleh-Engel, K., Moosbrugger, H. & Müller, H. (2003). Evaluating the fit of structural equation models: tests of significance and descriptive goodness-of-fit. *Measures Of Psychological Research Online*, (8), 2, 23-74.
- Stage, C. (1998). A comparison between item analysis based on item response theory and classical test theory: A study of the SweSAT test WORD. (Educational Measurement No 29). Umea University, Department of Educational Measurement.
- Stage, C. (1997). A Comparison Between Item Analysis Based on Item Response Theory and Classical Test Theory: A Study of the SweSAT Subtest READ. *Educational Research*, 29, 2-22.
- Şişman, M., Acat, M.B., Aypay, A. & Karadağ, E. (2011). *TIMSS 2007 ulusal matematik ve fen raporu 8. sınıflar*. Ankara: Milli Eğitim Bakanlığı.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics* (5. baskı). California: Pearson.

- Tekin, H. (2014). *Eğitimde ölçme ve değerlendirme*. Ankara: Yargı.
- Turgut, M. F. & Baykul, Y. (2010). *Eğitimde ölçme ve değerlendirme*. Ankara: Pegem.
- Thissen, D. (1991). *Multilog user's guide*. Chicago: Scientific Software.
- Tomkiewicz, J.& Rogers, W.T. (2005). The use of one-, two-, and three-parameter and nominal item response scoring in place of number-right scoring in the presence of test-wisness, *The Alberta Journal of Educational Research*, 51(3),200-215.
- Van der Linden, W. J. & Hambleton, R.K. (1997). *Handbook of modern item response theory*. New York: Springer-Verlag.
- Wiberg, M. (2012). Can a multidimensional test be evaluated with unidimensional item response theory? *Educational Research and Evaluation*, 18(4), 307-320. 20
- Yen, W. M. (1987). A comparison of the efficiency and accuracy of BILOG and LOGIST. *Psychometrika*, 52(3), 275-291.
- Yıldırım, H.H., Çömlekoğlu & G. Berberoğlu, G. (2003). Milli eğitim bakanlığı özel okullar sınavı verilerinin madde tepki kuramı modellerine uyumu. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 23, 159-168.





EKLER

EK 1: Mplus Programı Tetrakorik AFA Syntax

Title: Tetrakorik Korelasyon Matrisi

Data:

File is C:\Users\gul\Desktop\GUL111.DAT;

Variable:

Names are

m1 m2 m3 m4 m5 m6 m7 m8 m9 m10 m11 m12 m13 m14 m15 m16 m17 m18 m19 m20
m21 m22 m23 m24 m25 m26;

Usevariables are

m1 m2 m3 m4 m5 m6 m7 m8 m9 m10 m11 m12 m13 m14 m15 m16 m17 m18 m19 m20
m21 m22 m23 m24 m25 m26;

Categorical are

m1 m2 m3 m4 m5 m6 m7 m8 m9 m10 m11 m12 m13 m14 m15 m16 m17 m18 m19 m20
m21 m22 m23 m24 m25 m26;

Analysis: Type = efa 1 3;

EK 2: LISREL Programı DFA Syntax
GUL11DFA

Observed Variables

M1-M26

Covariance Matrix from File GUL11.COV

Sample Size: 543

Latent Variables: Matematik

Relationships:

M1 M2 M3 M4 M5 M6 M7 M8 M9 M10 M11 M12 M13 M14 M15 M16 M17 M18 M19
M20 M21 M22 M23 M24 M25 M26= Matematik

Path Diagram

End of Problem

EK 3: BILOG Programında Kullanılan Syntax GUL

>COMMENT

>GLOBAL DFName = 'C:\Users\gul\Desktop\TIMSS11.dat',

 NPArm = 3,

 LOGistic,

 SAVe;

>SAVE PARm = 'C:\Users\gul\Desktop\TIMSS11.PAR',

 SCOr = 'C:\Users\gul\Desktop\TIMSS11.SCO';

>LENGTH NITems = (26);

>INPUT NTot = 26,

 NALt = 1000,

 NIDchar = 3;

>ITEMS ;

>TEST1 TNAme = 'TEST0001',

 INUmber = (1(1)26);

(3A1, 26A1)

>CALIB CYCles = 25,

 NEWton = 10,

 CRIt = 0.0010,

 PLOt = 1.0000,

 ACCEl = 0.0000,

 CHIsquare = (25, 9);

>SCORE METHod = 1,

 RSCtype = 3,

 POP;

Not: Parametre sayısı Nparm komutu ile 1,2 ve 3 olarak değiştirilebilmektedir.

EK 4: MULTILOG Programında Kullanılan Syntax

>PROBLEM RANDOM,

INDIVIDUAL,

DATA = 'C:\Users\gül\Desktop\gul\timss11.dat',

NITEMS = 26,

NGROUPS = 1,

NEXAMINEES = 543,

NCHARS = 3;

>TEST ALL,

L2;

>END ;

2

01

11111111111111111111111111111111

N

(3A1,26A1)

EK 5: R Programı ile Elde Edilen Q3 İstatistikleri

	1PLM	2PLM	3PLM
1	-0.0125540038	-0.1008800922	-0.0556855687
2	-0.0706004599	-0.1046534316	-0.0637740344
3	0.0272473519	0.0348129394	0.0273009192
4	0.1284053160	0.0102281478	0.0032558507
5	-0.0108663147	-0.0053499184	0.0459341457
6	0.0016695123	-0.1399497536	-0.0710511670
7	-0.0402164794	-0.0479132057	-0.0272746510
8	-0.0566855369	-0.0462388388	-0.0399432471
9	-0.0287400164	-0.1185037224	-0.1036705298
10	0.1388179271	-0.0484749308	-0.0968674766
11	0.1729728109	0.0009865096	-0.0273258904
12	0.1251834196	-0.0505969951	-0.0552224218
13	0.0905779850	0.0671058900	0.0637948182
14	0.0903163059	-0.1141039816	-0.1296755318
15	0.0571037881	0.0096096036	0.0255515118
16	0.1684276591	-0.0619018215	-0.0709935860
17	0.0107378899	-0.0140802004	0.0207092420
18	0.1347285517	0.0210125590	0.0220140963
19	0.0278976233	0.0019572522	-0.0034687018
20	0.0215972158	-0.0980500232	-0.0735200833
21	0.0257704916	-0.0789233183	-0.0699083860
22	-0.0164572218	-0.0033940382	0.0122457894
23	-0.0445315833	-0.0036259857	0.0065007656
24	0.0637707207	-0.0425292414	-0.0502900952
25	0.0466742867	-0.0308595634	-0.0366232048
26	-0.0850990949	-0.0332846721	-0.0379943813
27	0.1085022198	0.0401604821	0.0648958872
28	-0.0623381183	-0.1381236770	-0.1293947328
29	-0.0491417639	0.0023942784	-0.0142521169
30	0.0232186696	0.0015734800	-0.0128977609
31	-0.0017057321	0.0490178849	0.0453639081
32	-0.0516110557	0.0346090758	0.0376168412
33	0.0467376978	0.0403147615	0.0368250676
34	-0.0555422349	-0.1077837417	-0.0856565434
35	-0.0012542711	-0.0442195254	-0.0195756141
36	0.0966934752	0.0141767283	0.0242337286
37	-0.0394859291	-0.0715721317	-0.0716310593
38	0.0170826809	-0.0827305345	-0.0731122036
39	-0.0699110470	-0.0537437475	-0.0538999119
40	0.0679524546	-0.0601143322	-0.0462522087
41	-0.0085437239	0.0363504309	0.0253742416
42	0.1057134692	0.0215027012	0.0267398800

43	-0.1076690342	-0.0721255277	-0.0708777784
44	0.0451714768	0.0529326226	0.0475946070
45	0.0470659575	-0.0156898584	-0.0181715455
46	-0.1095361726	-0.0441050721	-0.0543880063
47	-0.0838973315	0.0147467877	0.0156041565
48	-0.0216736216	-0.0306863142	-0.0248432132
49	-0.0674125618	-0.0561779008	-0.0427881573
50	-0.0183578453	-0.0258646944	-0.0001773228
51	-0.0001549131	0.0198404732	0.0217385975
52	-0.0861388442	-0.0002592000	-0.0412973866
53	-0.0785378853	-0.0515134939	-0.0606963994
54	-0.0618785331	0.0090430139	-0.0041408901
55	-0.1436338716	-0.0401498740	-0.0385863521
56	-0.0781928383	-0.0537912235	-0.0679292545
57	-0.0257959765	-0.0022666286	0.0249803694
58	-0.0889738449	-0.0819024880	-0.0588610093
59	-0.0197698912	-0.0375569001	-0.0263449404
60	-0.0158670851	-0.0198103292	-0.0282703719
61	0.0101001166	0.0089457573	0.0216559591
62	-0.1346144577	-0.0825099294	-0.0814860258
63	0.0368205318	0.0221686927	0.0407725244
64	-0.0607432371	0.0270808429	0.0068121401
65	0.0002857102	-0.0239477096	-0.0269366222
66	-0.1154175198	-0.0369590932	-0.0391740500
67	0.0117500215	0.0528291824	0.0359144307
68	0.0070739703	0.0009742756	-0.0161490478
69	-0.0773408445	0.0231457103	0.0057968971
70	-0.1365625869	-0.0221627605	-0.0275118009
71	-0.0660909043	-0.0170330666	-0.0169828039
72	-0.1190428696	-0.0633396445	-0.0529471868
73	0.1419034982	-0.0181901498	-0.0092197428
74	-0.0328232198	-0.0510991177	-0.0194952893
75	0.1055997663	-0.0339889954	-0.0092150769
76	-0.0448475674	-0.0552087540	-0.0415124805
77	-0.0656632485	-0.0388762471	-0.0296398271
78	0.0874522691	0.0054754900	0.0187984872
79	0.1260802178	-0.0365193285	-0.0491439685
80	0.0413657340	-0.1606329032	-0.1603291653
81	0.1345864145	-0.1209587239	-0.1094618428
82	0.0505745343	-0.0539584417	-0.0499487092
83	0.2113459127	-0.0085846216	-0.0043070434
84	0.0621409842	0.0159226390	0.0218253711
85	0.1698599975	-0.1575991275	-0.1464797862
86	0.0183301056	-0.0114554190	0.0063215643
87	0.1840462110	-0.0173858936	0.0010743648

88	-0.0128053159	-0.0468954811	-0.0446170293
89	0.0455069199	-0.0701877259	-0.0561299540
90	0.0681594271	-0.0973818221	-0.0803978601
91	-0.0483212557	-0.0350667253	-0.0289829315
92	-0.0625608624	0.0004277481	0.0096129979
93	0.0682257500	-0.0495596360	-0.0448656865
94	-0.0194439572	-0.1094167286	-0.1030794113
95	-0.0159876556	-0.0116395103	-0.0150780756
96	0.0889643682	0.0002328414	-0.0007197001
97	-0.0054128730	0.0145315109	0.0179095188
98	-0.0687923152	-0.0140713511	-0.0092200443
99	0.0089026738	-0.0659737460	-0.0620053899
100	0.0049237766	-0.1102176741	-0.1214687138
101	0.0774475685	-0.0230558037	-0.0240275068
102	0.0985311233	-0.0887229449	-0.0812130328
103	0.0601320506	-0.0201793680	-0.0157354817
104	0.0931347909	-0.0871660040	-0.0850839944
105	-0.0404949036	-0.0784561901	-0.0787303611
106	0.1626467356	-0.0521445790	-0.0457988769
107	0.0114639082	0.0129273179	0.0122049933
108	0.0835929133	-0.1133698779	-0.0940954924
109	0.0178453142	0.0245288364	0.0242121703
110	0.0500373998	-0.0478968838	-0.0408020524
111	0.0895190748	-0.0419668554	-0.0311942402
112	-0.0473090478	-0.0117205393	-0.0170562526
113	-0.0660248919	0.0126237756	0.0165446160
114	0.0814842209	0.0137596561	0.0169987376
115	-0.0289972084	-0.0682888686	-0.0642016982
116	-0.0428327789	-0.0129878541	-0.0393462718
117	-0.0730974978	-0.0076867094	-0.0315462603
118	-0.0912622725	0.0141250820	0.0170644588
119	-0.0512168898	-0.0344521079	-0.0627168727
120	-0.0473112219	-0.0236079868	0.0103686095
121	-0.0354806105	-0.0125757839	0.0170799023
122	-0.0386962433	-0.0808070148	-0.0685128451
123	0.0537672694	0.0926849736	0.0799272465
124	-0.0380982395	-0.0680984231	-0.0528089716
125	-0.0827086342	-0.0245172394	-0.0334602849
126	0.0027074529	-0.0415928166	-0.0161177470
127	-0.0288808523	0.0524946450	0.0130795495
128	-0.0049456001	-0.0555517006	-0.0651868881
129	-0.0703555130	0.0093754564	0.0011115501
130	-0.0246000387	0.0165229240	-0.0108143284
131	-0.0315491188	-0.0674078861	-0.1015687269
132	-0.1068141293	-0.0123505853	-0.0470578743

133	-0.1257714048	-0.0122512741	-0.0246917513
134	-0.0353437442	0.0043938651	-0.0008401616
135	-0.0375246576	0.0239122227	0.0379590601
136	0.0002824555	0.0304902569	0.0195644772
137	-0.0586219704	0.0039674391	-0.0003664831
138	0.1237845155	0.0963785570	0.0815849128
139	-0.0316439276	-0.1309524013	-0.1016374578
140	0.0600933605	-0.0166158162	0.0146448560
141	0.1166591398	-0.0285927316	-0.0336794828
142	0.0358219912	-0.0381121120	-0.0453857216
143	0.0708387225	-0.0816666767	-0.0838649613
144	-0.0926936643	-0.1080825309	-0.1161254467
145	0.1409434171	-0.0419960985	-0.0413124558
146	-0.0339897429	-0.0161143180	-0.0389881603
147	0.0545377937	-0.1080511440	-0.1243503390
148	-0.0509792166	-0.0362685096	-0.0415346192
149	0.0681761598	0.0493748859	0.0320358007
150	0.0797996429	-0.0170563348	-0.0378885453
151	0.0383151080	0.1025945265	0.0880836179
152	-0.1436100400	-0.0722693150	-0.0747538911
153	0.0206999346	-0.0338269963	-0.0309574069
154	-0.0210448689	-0.0406437943	-0.0278733877
155	-0.1266719120	-0.0298333454	-0.0282626573
156	-0.0433919514	-0.0078905620	-0.0141975872
157	-0.0449284543	-0.0258214455	-0.0139061822
158	-0.0539042733	-0.0414210260	-0.0293485061
159	-0.0632236563	-0.0830139314	-0.0759290116
160	-0.1090650030	-0.0735443559	-0.0762076274
161	0.0284985159	0.0400215798	0.0455730644
162	-0.0466983411	-0.0054886223	-0.0084928375
163	0.0374109426	0.0327469075	0.0425273233
164	-0.0265851275	0.0065832647	-0.0071452708
165	-0.0311389547	-0.0521835323	-0.0479977675
166	-0.0873317773	-0.0223515483	-0.0228713810
167	0.0004274717	-0.0007220719	-0.0080315563
168	-0.0869216426	-0.1037628065	-0.1102807594
169	-0.0537340698	0.0330595988	0.0212927586
170	-0.0920014739	0.0121106174	0.0077060293
171	-0.0708163257	0.0130289834	0.0121135036
172	-0.0772833421	-0.0232779020	-0.0183476101
173	-0.1181640615	-0.0492837406	-0.0477498990
174	-0.1483010669	-0.1169784801	-0.1159384083
175	-0.0307541470	0.0192013307	0.0249997347
176	-0.0252882225	0.0173302719	0.0228021429
177	-0.0688420567	-0.0169592343	-0.0159601910

178	-0.0361764569	0.0052066939	0.0089373246
179	-0.0406326015	0.0341453343	0.0331634643
180	0.0037388026	0.0349861429	0.0419180167
181	-0.0415501129	0.0802041962	0.0802094569
182	-0.1000406117	-0.0821958480	-0.0737983203
183	-0.1770695221	-0.0815183200	-0.0807941180
184	-0.0160659848	0.0120374420	0.0138343027
185	-0.0544232627	-0.0073677337	-0.0038896039
186	-0.1434667390	-0.0268436717	-0.0287382978
187	-0.1764183535	-0.0468296642	-0.0450417869
188	-0.1053576736	0.0008894174	0.0029342278
189	-0.1547035184	-0.0702933888	-0.0662571409
190	-0.0236965485	-0.0739148259	-0.0685578253
191	0.0263133142	-0.0203406402	-0.0100812012
192	0.0776379934	-0.0201682205	-0.0154503071
193	0.0057779918	-0.0311113958	-0.0319992098
194	0.0466384983	-0.0535704171	-0.0516015433
195	-0.0126991076	-0.0152412937	-0.0184370990
196	0.0596382874	-0.0813860387	-0.0742054016
197	0.0384792197	0.0110752429	-0.0045797304
198	0.0766612644	-0.0376976688	-0.0308076939
199	-0.0438089991	-0.0212349789	-0.0227729610
200	0.0944565783	0.0217110768	0.0161651515
201	0.0267008437	-0.0653276526	-0.0691505301
202	-0.0925435418	-0.0465802291	-0.0608985291
203	-0.0361888277	0.0527003416	0.0519856046
204	-0.0571075264	-0.1270874573	-0.1276861682
205	-0.0533241086	-0.0537877297	-0.0489850134
206	0.0727102558	-0.0493596953	-0.0740037125
207	0.1431971196	0.0183347096	0.0085252219
208	0.0131223423	0.0174347442	0.0129296185
209	0.1410544867	0.0114950638	-0.0046238358
210	-0.0397133523	-0.0667997591	-0.0647643180
211	0.1750047999	0.0020004039	-0.0121889233
212	-0.0575610476	-0.0493095837	-0.0319033499
213	0.0907680697	-0.0211665250	-0.0245235697
214	-0.0166450453	-0.0101612837	-0.0174389307
215	0.0276449010	-0.0537930689	-0.0412611882
216	0.0395170430	-0.0359309417	-0.0330622460
217	-0.0367098112	0.0062130908	0.0130686975
218	-0.0651202286	0.0048682273	0.0113493781
219	0.0613196412	-0.0017863196	-0.0105261268
220	0.0138345539	-0.0160353091	-0.0242458933
221	0.1005624064	-0.0528914886	-0.0528635618
222	0.0044921962	0.0029440568	0.0015749586

223	0.1082621643	-0.0433156962	-0.0492292939
224	0.0459749125	0.0351012273	0.0406997194
225	0.1328669677	-0.0694501137	-0.0716910888
226	-0.0544034730	-0.0501498942	-0.0333880544
227	0.0554362352	-0.0875515991	-0.0812536511
228	-0.0381337258	-0.0399036068	-0.0429926723
229	0.0627978154	-0.0322157889	-0.0174735969
230	0.0816761300	0.0055940179	0.0140818728
231	-0.0851330854	-0.0552609065	-0.0499013423
232	-0.1075340626	-0.0476272002	-0.0414657954
233	0.1089432747	0.0370018972	0.0352064330
234	0.0429171199	0.0140254249	0.0133948374
235	0.1209006190	0.0123754598	0.0161746926
236	0.1671172243	-0.0725787388	-0.0721438367
237	0.0289354282	-0.0325618152	-0.0346006797
238	0.1916765123	-0.1255715369	-0.1200196499
239	0.0403561699	0.0084259500	0.0118943726
240	0.1841257412	-0.0499779441	-0.0311892841
241	-0.0243513338	-0.0666502184	-0.0669642580
242	0.0170331899	-0.1027171464	-0.0955702305
243	0.1225839141	-0.0503406096	-0.0377507369
244	0.0324652615	0.0617437617	0.0595521018
245	-0.0967793275	-0.0436062842	-0.0357240297
246	0.0825278795	-0.0268537747	-0.0242181348
247	0.0171988245	-0.0525677717	-0.0490494823
248	0.0323019778	-0.0807183297	-0.0792795560
249	-0.0234245578	-0.0369115341	-0.0386476435
250	0.0544814478	-0.0952755549	-0.0894874363
251	-0.0113462535	0.0278641809	0.0190221192
252	0.0164616097	-0.1320949236	-0.1141353285
253	0.0042617683	0.0233884334	0.0231239028
254	-0.0481433039	-0.1014734316	-0.0981622884
255	0.1575021921	0.0813204672	0.0891705054
256	-0.0469102335	-0.0065815876	-0.0160535212
257	-0.0131959534	0.0682772712	0.0692482135
258	0.0127396897	0.0079939997	0.0105981976
259	-0.0254489082	-0.0306506612	-0.0272379877
260	-0.0050735697	-0.0705733879	-0.0743852278
261	0.0279726699	0.0119083758	0.0108746700
262	-0.0363281586	-0.0671183050	-0.0636650701
263	0.0208850626	0.0006882737	0.0119418061
264	0.0038921315	-0.0234410886	-0.0268646892
265	0.0286745281	-0.0869661360	-0.0819772591
266	0.0959458262	-0.0733447958	-0.0660259158
267	-0.1027036118	-0.1066143309	-0.1110595499

268	-0.1435535343	-0.1004712086	-0.0952091569
269	0.0714480742	-0.0370985543	-0.0382142207
270	0.0183664244	-0.0484899663	-0.0487266033
271	0.0851546361	0.0187855783	0.0162099693
272	-0.0187893017	0.0266818095	0.0215538515
273	0.0500727141	-0.0045881997	-0.0072290752
274	-0.0660782579	-0.0167978699	-0.0202704470
275	0.0081068128	-0.0572947225	-0.0603474547
276	-0.0057749951	-0.0480277242	-0.0519189180
277	-0.1316921580	-0.0695995744	-0.0767352207
278	-0.0349777582	0.0532824659	0.0529240114
279	-0.0321918887	-0.0330633844	-0.0330572441
280	-0.0105702393	0.0161722950	0.0203917748
281	-0.0268059218	-0.0944106438	-0.0860483006
282	-0.2010954870	-0.0839453031	-0.0656317760
283	0.0138650295	-0.0442866819	-0.0456409474
284	0.1359894916	-0.0153268763	-0.0065543021
285	0.1051968990	-0.1161882700	-0.1032656973
286	-0.0844215230	-0.1138339342	-0.1150740182
287	-0.0934800185	-0.0556062843	-0.0476178416
288	0.0998040072	-0.0543462759	-0.0520363439
289	0.0609564333	-0.0349773775	-0.0317299897
290	-0.0213711651	-0.0986533643	-0.1049601598
291	-0.0962999358	-0.0352879907	-0.0400940891
292	0.0408366712	0.0664902233	0.0488902581
293	0.0296255860	-0.0055000216	-0.0266055448
294	-0.0942007821	-0.0086539380	-0.0297060703
295	-0.1143725512	-0.0072708252	-0.0139155336
296	-0.0576998163	-0.0307941547	-0.0338865126
297	-0.0491274768	-0.0054116721	0.0034000079
298	-0.0094058074	-0.0527522394	-0.0490247768
299	0.0974830975	-0.0469488696	-0.0371449496
300	0.1401429407	-0.0592407182	-0.0364442472
301	-0.0702018164	-0.0830266745	-0.0906622201
302	-0.0271819766	0.0359420130	0.0418410072
303	0.0708083455	-0.0338833752	-0.0239006072
304	0.0357029842	-0.0259637372	-0.0165350459
305	0.0645342770	0.0243029575	0.0256749150
306	-0.0802525750	-0.1145389299	-0.1143373205
307	-0.1041620467	-0.0117890265	-0.0179026821
308	-0.1062081948	0.0065900249	0.0069709186
309	0.0076922843	0.0386514920	0.0372855598
310	-0.0508698649	0.0030154397	0.0036604780
311	0.1327662832	0.0120653857	0.0107317664
312	0.0048167562	0.0303035919	0.0117145877

313	-0.0334057736	0.0272930661	0.0244232821
314	0.0172250231	-0.0020977164	-0.0037456213
315	0.0203556655	0.0005691817	0.0038418702
316	0.0335352737	0.0604131279	0.0438143167
317	-0.0654229449	-0.0084926665	-0.0082896132
318	0.0711757084	-0.0647181494	-0.0600945046
319	0.0372427448	0.0076520350	0.0152945041
320	-0.1762448479	-0.0538819959	-0.0607925350
321	-0.0765475798	0.0016188989	-0.0046690799
322	-0.0826530000	-0.0041092656	-0.0034220798
323	-0.1171082040	-0.0562272112	-0.0550413070
324	-0.1423381810	-0.0425667294	-0.0382526209
325	0.0334005502	0.0332305289	0.0347454410

