

**T.C.**  
**AKDENİZ ÜNİVERSİTESİ**  
**EĞİTİM BİLİMLERİ ENSTİTÜSÜ**  
**EĞİTİM BİLİMLERİ ANABİLİM DALI**  
**EĞİTİMDE ÖLÇME VE DEĞERLENDİRME**  
**YÜKSEK LİSANS PROGRAMI**

**MADDE TEPKİ KURAMINA DAYALI TEST EŞİTLEME**  
**YÖNTEMLERİNİN KARŞILAŞTIRILMASI: PISA 2012 FEN**  
**TESTİ ÖRNEĞİ**

**YÜKSEK LİSANS TEZİ**

**Burcu AKSEKİOĞLU**

**Antalya**

**Ocak, 2017**

**T.C.**  
**AKDENİZ ÜNİVERSİTESİ**  
**EĞİTİM BİLİMLERİ ENSTİTÜSÜ**  
**EĞİTİM BİLİMLERİ ANABİLİM DALI**  
**EĞİTİMDE ÖLÇME VE DEĞERLENDİRME**  
**YÜKSEK LİSANS PROGRAMI**

**MADDE TEPKİ KURAMINA DAYALI TEST EŞİTLEME**  
**YÖNTEMLERİNİN KARŞILAŞTIRILMASI: PISA 2012 FEN**  
**TESTİ ÖRNEĞİ**

**YÜKSEK LİSANS TEZİ**

**Burcu AKSEKİOĞLU**

**Danışman: Doç. Dr. Bayram BIÇAK**

**Antalya**

**Ocak, 2017**

## DOĐRULUK BEYANI

Yüksek lisans tezi olarak sunduĐum bu çalıřmayı, bilimsel ahlak ve geleneklere aykırı düřecek bir yol ve yardıma bařvurmaksızın yazdıĐımı, yararlandıĐım eserlerin kaynakçalardan gösterilenlerden oluřtuĐunu ve bu eserleri her kullarıřımda alıntı yaparak yararlandıĐımı belirtir; bunu onurumla doĐrularım. Enstitü tarafından belli bir zamana baĐlı olmaksızın, tezimle ilgili yaptıĐım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçlara katlanacaĐımı bildiririm.

06.01.2017

Burcu AKSEKİOĐLU



T.C.  
AKDENİZ ÜNİVERSİTESİ  
EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Burcu AKSEKİOĞLU'nun bu çalışması 06.01.2017 tarihinde jürimiz tarafından Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Tezli Yüksek Lisans Programında **Yüksek Lisans Tezi** olarak **oy birliği** ile kabul edilmiştir.

İMZA

**Başkan** : Yrd. Doç. Dr. Mustafa KILINÇ



(Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü)

**Üye** : Doç. Dr. Bayram BIÇAK (Danışman)



(Akdeniz Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü)

**Üye** : Yrd. Doç. Dr. Hakan KOĞAR



(Akdeniz Üniversitesi Eğitim Fakültesi Eğitim Bilimleri Bölümü)

**YÜKSEK LİSANS TEZİNİN ADI:** Madde Tepki Kuramına Dayalı Test Eşitleme Yöntemlerinin Karşılaştırılması: PISA 2012 Fen Testi Örneği

**ONAY:** Bu tez, Enstitü Yönetim Kurulunca belirlenen yukarıdaki jüri üyeleri tarafından uygun görülmüş ve Enstitü Yönetim Kurulunun .....tarihli ve .....sayılı kararıyla kabul edilmiştir.

Doç. Dr. Mehmet CANBULAT

Enstitü Müdürü

## ÖNSÖZ

Öncelikle bu çalışma süresince tezimin tamamlanmasına önemli katkılar sağlayan ve bana daima destek olan değerli tez danışmanım Doç. Dr. Bayram BIÇAK'a çok teşekkür ederim.

Bu süreçte bana sağladığı kolaylıklardan ve yardımlarından dolayı Ölçme ve Değerlendirme Anabilim Dalı Başkanı Yrd. Doç. Dr. Mustafa KILINÇ'a, test eşitleme ve madde tepki kuramı konusunda değerli bilgilerinden faydalandığım Yrd. Doç. Dr. Neşe ÖZTÜRK GÜBEŞ ve Yrd. Doç. Dr. Şeyma UYAR'a ve desteğini benden hiçbir zaman esirgemeyen oda arkadaşım Arş. Gör. Elif Süreyya KANYILMAZ'a çok teşekkür ederim.

Son olarak bugünlere gelmemde büyük emeği olan annem ve babama sonsuz teşekkür ederim.

Burcu AKSEKİOĞLU

**ÖZET**  
**MADDE TEPKİ KURAMINA DAYALI TEST EŞİTLEME**  
**YÖNTEMLERİNİN KARŞILAŞTIRILMASI: PISA 2012 FEN TESTİ**  
**ÖRNEĞİ**

Aksekioğlu, Burcu  
Yüksek Lisans, Eğitim Bilimleri Anabilim Dalı  
Tez Yöneticisi: Doç. Dr. Bayram BIÇAK  
Ocak 2017, 109 sayfa

Bu çalışmanın amacı, madde tepki kuramına dayalı test eşitleme yöntemlerini gerçek veriler üzerinde karşılaştırmak ve en az hata veren yöntemi tespit etmektir. Bu amaçla PISA 2012 fen okuryazarlığı testinden seçilen iki kitapçık (kitapçık-1 ve kitapçık-8) ölçek dönüştürme (ortalama-ortalama, ortalama-standart sapma, Stocking-Lord, Haebara) ve test eşitleme yöntemleri (MTK gerçek-puan eşitleme, MTK gözlenen-puan eşitleme) kullanılarak eşitlenmiş ve farklı yöntemlerden elde edilen hata miktarları incelenmiştir.

Araştırmanın çalışma grubunu, 1 numaralı kitapçığı cevaplayan 350 ve 8 numaralı kitapçığı cevaplayan 351 olmak üzere toplam 701 öğrenci oluşturmaktadır. Denk olmayan gruplarda ortak madde deseninde yürütülen çalışmada her iki kitapçıkta 17'si ortak olmak üzere ikili puanlanan toplam 33 madde yer almaktadır. Verilerin analizinde ilk olarak madde tepki kuramının varsayımları test edilmiştir. İkinci aşamada madde ve yetenek parametreleri iki parametrelili lojistik model (2PLM) kullanılarak BILOG-MG programı ile kestirilmiştir. Parametre kestiriminde EAP (Beklenen A Posteriori) yöntemi kullanılmıştır. Ardından STUIRT programı ile dört farklı yöntem için ölçek dönüştürme işlemi gerçekleştirilmiştir. Son aşamada ise iki kitapçıktan elde edilen puanlar POLYEQUATE programı ile eşitlenmiş ve farklı yöntemlerden elde edilen hata miktarları için WMSE (ağırlıklandırılmış hata kareleri ortalaması) katsayısı hesaplanmıştır.

Araştırmadan elde edilen sonuçlara göre, en az hata miktarını ortalama-standart sapma yönteminin verdiği görülmüştür. Bu yöntemi sırasıyla ortalama-ortalama, Stocking-Lord ve Haebara yöntemleri takip etmektedir. Test eşitleme yöntemlerinden elde edilen sonuçlar incelendiğinde ise her iki yöntem birbirine

benzer sonuçlar vermekle birlikte MTK gözlenen-puan eşitleme yönteminin MTK gerçek-puan eşitleme yönteminden daha az hata miktarına sahip olduğu bulunmuştur.

**Anahtar kelimeler:** Test Eşitleme, Madde Tepki Kuramı, Eşitleme Yöntemleri, Eşitleme Hatası



## ABSTRACT

### COMPARISON OF TEST EQUATING METHODS BASED ON ITEM RESPONSE THEORY: PISA 2012 SCIENCE TEST SAMPLE

Aksekioglu, Burcu

Post Graduate, Department of Educational Sciences

Supervisor: Assoc. Prof. Dr. Bayram BIÇAK

January 2017, 109 pages

The objective of this study is to compare test equating methods based on item response theory over real data and to detect the method that gives the least error. For this purpose, two booklets selected from PISA 2012 science literacy test (booklet-1 and booklet-8) were equated using scale linking (mean-mean, mean-sigma, Stocking-Lord, Haebara) and test equating methods (IRT true-score equating, IRT observed-score equating), and error rates obtained from various methods were analysed.

The research study group consisted of a total of 701 students including 350 students answering booklet-1 and 351 students answering booklet-8. In the study that was carried out in common-item nonequivalent groups design, there were a total of 33 dichotomous scored items including 17 common items in each booklet. For the data analysis, firstly item response theory assumptions were tested. At the second stage, item and ability parameters were estimated using two-parameter logistic model (2PLM) via BLOG-MG software. In parameter estimation, EAP (Expected A Posteriori) method was used. Then, scale linking process was carried out for four different methods via STUIRT software. In the final stage, scores obtained from two booklets were equated with POLYEQUATE software and WMSE (weighted mean square error) index was calculated for error rates that were obtained from different methods.

According to the results obtained from the research, it was observed that the least error rate was obtained with mean-sigma method. This method was followed by mean-mean, Stocking-Lord and Haebara methods respectively. When the results obtained from the test equating methods were analyzed, it was found that, while both methods yielded similar results, IRT observed-score equating method gave less rate compared to IRT true-score equating method.



**Keywords:** Test Equating, Item Response Theory, Equating Methods, Equating Error.



## İÇİNDEKİLER

ÖNSÖZ .....	i
ÖZET.....	ii
ABSTRACT .....	iv
İÇİNDEKİLER .....	vi
TABLolar LİSTESİ.....	ix
ŞEKİLLER LİSTESİ .....	x
KISALTMALAR LİSTESİ.....	xi

### BİRİNCİ BÖLÜM

#### GİRİŞ

1.1 Problem Durumu.....	1
1.2 Araştırmanın Amacı.....	2
1.3 Araştırmanın Önemi.....	4
1.4 Varsayımlar .....	4
1.5 Sınırlılıklar .....	5
1.6 Tanımlar .....	5

### İKİNCİ BÖLÜM

#### KAVRAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

2.1 Uluslararası Öğrenci Değerlendirme Programı (PISA).....	6
2.2 Test Eşitleme.....	8
2.2.1 Yatay Eşitleme .....	9
2.2.2 Dikey Eşitleme.....	9
2.3 Eşitleme Varsayımları.....	10
2.4 Eşitleme Basamakları.....	10
2.5 Eşitleme Desenleri .....	11
2.5.1 Tek Grup Deseni .....	11

2.5.2 Dengelenmiş Tek Grup Deseni .....	12
2.5.3 Eşdeğer Gruplar Deseni .....	12
2.5.4 Denk Olmayan Gruplarda Ortak Madde Deseni .....	13
2.6 Test Eşitleme Yöntemleri.....	15
2.6.1 Madde Tepki Kuramına Dayalı Eşitleme.....	15
<i>Bir Parametrelili Lojistik Model (1PLM)</i> .....	16
<i>İki Parametrelili Lojistik Model (2PLM)</i> .....	17
<i>Üç Parametrelili Lojistik Model (3PLM)</i> .....	18
2.7 Madde Tepki Kuramına Dayalı Kalibrasyon Yöntemleri.....	18
2.7.1 Eş Zamanlı Kalibrasyon.....	19
2.7.2 Ayrı Kalibrasyon .....	19
2.7.2.1 Moment Yöntemleri.....	20
2.7.2.2 Karakteristik Eğri Dönüştürme Yöntemleri.....	22
2.8 Madde Tepki Kuramına Dayalı Eşitleme Yöntemleri .....	23
2.8.1 MTK Gerçek-Puan Eşitleme .....	23
2.8.2 MTK Gözlenen-Puan Eşitleme .....	24
2.9. Eşitleme Hatası.....	25
2.10 İlgili Araştırmalar.....	26
2.10.1 Ölçek Dönüştürme Yöntemlerinin Karşılaştırıldığı Araştırmalar.....	26
2.10.2 Test Eşitleme Yöntemlerinin Karşılaştırıldığı Araştırmalar .....	30

## ÜÇÜNCÜ BÖLÜM

### YÖNTEM

3.1 Araştırmanın Türü.....	34
3.2 Çalışma Grubu .....	34
3.3 Araştırma Verileri .....	34
3.4 Veri Toplama Deseni .....	35
3.5 Verilerin Analizi.....	36
3.5.1 Betimsel İstatistikler.....	36
3.5.2 Madde Tepki Kuramı Varsayımlarının Test Edilmesi .....	37
3.5.3 Madde ve Yetenek Parametrelerinin Kestirilmesi .....	45
3.5.4 Eşitleme Sürecinin Uygulanması .....	48
3.5.5 Eşitleme Hatası .....	49

## DÖRDÜNCÜ BÖLÜM

### BULGULAR

4.1 Birinci Alt Probleme İlişkin Bulgular .....	50
4.2 İkinci Alt Probleme İlişkin Bulgular .....	53
4.3 Üçüncü Alt Probleme İlişkin Bulgular.....	57
4.4 Dördüncü Alt Probleme İlişkin Bulgular .....	60
4.5 Beşinci Alt Probleme İlişkin Bulgular .....	63

## BEŞİNCİ BÖLÜM

### SONUÇ VE TARTIŞMA

5.1 Sonuç ve Tartışma.....	66
5.2 Öneriler .....	69
5.2.1 Uygulayıcılara Yönelik Öneriler .....	69
5.2.2 Araştırmacılara Yönelik Öneriler.....	69

<b>KAYNAKÇA</b> .....	71
-----------------------	----

<b>EKLER</b> .....	83
--------------------	----

EK-1 Kız ve Erkek Öğrencilerden Elde Edilem Madde Parametreleri .....	83
---	----

EK-2 Kitapçık-1 İçin Tek Numaralı Madde Setinden Elde Edilen Yetenek Parametreleri .....	84
--	----

EK-3 Kitapçık-1 İçin Çift Numaralı Madde Setinden Elde Edilen Yetenek Parametreleri .....	85
---	----

EK-4 Kitapçık-8 İçin Tek Numaralı Madde Setinden Elde Edilen Yetenek Parametreleri .....	86
--	----

EK-5 Kitapçık-8 İçin Çift Numaralı Madde Setinden Elde Edilen Yetenek Parametreleri .....	87
---	----

EK-6 Kitapçık-1'den Elde Edilen Madde Karakteristik Eğrileri .....	88
--	----

EK-7 Kitapçık-8'den Elde Edilen Madde Karakteristik Eğrileri.....	90
---	----

EK-8 Farklı Yöntemlerden Elde Edilen Dönüştürülmüş Parametreler.....	92
--	----

<b>BİLDİRİM</b> .....	93
-----------------------	----

<b>ÖZGEÇMİŞ</b> .....	94
-----------------------	----

## TABLolar LİSTESİ

Tablo 2.1 Tek Grup Deseni .....	11
Tablo 2.2 Dengelenmiş Tek Grup Deseni .....	12
Tablo 2.3 Eşdeğer Gruplar Deseni (Random Grup Deseni) .....	13
Tablo 2.4 Denk Olmayan Gruplarda Ortak Madde Deseni .....	15
Tablo 3.1 Araştırmaya İlişkin Denk Olmayan Gruplarda Ortak Madde Deseni .....	35
Tablo 3.2 Kitapçıklara İlişkin Betimsel İstatistikler .....	36
Tablo 3.3 Kitapçıklara İlişkin Açıklayıcı Faktör Analizi Sonuçları .....	38
Tablo 3.4 Doğrulayıcı Faktör Analizi Sonucu Elde Edilen Uyum İndeksleri .....	39
Tablo 3.5 Kitapçık-1 İçin Model-Veri Uyumunun İncelenmesi .....	40
Tablo 3.6 Kitapçık-1 İçin Maddelere Ait Ki-Kare Ve Anlamlılık Değerleri .....	41
Tablo 3.7 Kitapçık-8 İçin Model-Veri Uyumunun İncelenmesi .....	42
Tablo 3.8 Kitapçık-8 İçin Maddelere Ait Ki-Kare Ve Anlamlılık Değerleri .....	43
Tablo 3.9 Cinsiyete Göre Elde Edilen Madde Parametreleri Arasındaki Korelasyon .....	44
Tablo 3.10 Madde Setlerinden (Tek-Çift) Elde Edilen Yetenek Parametreleri Arasındaki Korelasyon .....	45
Tablo 3.11 Kitapçık-1 İçin Elde Edilen Madde Parametrelerine İlişkin Betimsel İstatistikler.....	46
Tablo 3.12 Kitapçık-1'e Ait Madde Parametre Değerleri .....	47
Tablo 3.13 Kitapçık-8 İçin Elde Edilen Madde Parametrelerine İlişkin Betimsel İstatistikler.....	47
Tablo 3.14 Kitapçık-8'e Ait Madde Parametre Değerleri .....	48
Tablo 4.1 Ortalama-Ortalama Yöntemi İçin Gerçek-Puan ve Gözlenen-Puan Eşitlemeden Elde Edilen Eşitlenmiş Puanlar .....	51
Tablo 4.2 Ortalama-Standart Sapma Yöntemi İçin Gerçek-Puan ve Gözlenen-Puan Eşitlemeden Elde Edilen Eşitlenmiş Puanlar .....	55
Tablo 4.3 Stocking-Lord Yöntemi İçin Gerçek-Puan ve Gözlenen-Puan Eşitlemeden Elde Edilen Eşitlenmiş Puanlar .....	58
Tablo 4.4 Haebara Yöntemi İçin Gerçek-Puan ve Gözlenen-Puan Eşitlemeden Elde Edilen Eşitlenmiş Puanlar .....	61
Tablo 4.5 Farklı Yöntemlerden Elde Edilen WMSE Katsayıları .....	63

## ŞEKİLLER LİSTESİ

Şekil 2.1 İç Ortak Maddelere İlişkin Görsel .....	14
Şekil 2.2 Dış Ortak Maddelere İlişkin Görsel .....	14
Şekil 2.3 Gerçek-Puan Eşitleme Yönteminin Grafikselsel Gösterimi.....	24
Şekil 2.4 Gözlenen-Puan Eşitleme Yönteminin Grafikselsel Gösterimi.....	25
Şekil 3.1 Sırasıyla Kitapçık-1 Ve Kitapçık-8 İin Elde Edilen Puan Dağılımları ....	37
Şekil 3.2 Sırasıyla Kitapçık-1 ve Kitapçık-8'e İlişkin Özdeğer Grafikleri .....	38
Şekil 4.1 Ham Puanlar İle Ortalama-Ortalama Yönteminden Elde Edilen Eşitlenmiş Puanlar Arasındaki İlişki .....	50
Şekil 4.2 Ortalama-Ortalama Yöntemi İin Ham Puanlar İle Eşitlenmiş Puanlar Arasındaki Fark .....	53
Şekil 4.3 Ham Puanlar İle Ortalama-Standart Sapma Yönteminden Elde Edilen Eşitlenmiş Puanlar Arasındaki İlişki .....	54
Şekil 4.4 Ortalama-Standart Sapma Yöntemi İin Ham Puanlar İle Eşitlenmiş Puanlar Arasındaki Fark .....	56
Şekil 4.5 Ham Puanlar İle Stocking-Lord Yönteminden Elde Edilen Eşitlenmiş Puanlar Arasındaki İlişki .....	57
Şekil 4.6 Stocking-Lord Yöntemi İin Ham Puanlar İle Eşitlenmiş Puanlar Arasındaki Fark .....	59
Şekil 4.7 Ham Puanlar İle Haebara Yönteminden Elde Edilen Eşitlenmiş Puanlar Arasındaki İlişki .....	60
Şekil 4.8 Haebara Yöntemi İin Ham Puanlar İle Eşitlenmiş Puanlar Arasındaki Fark .....	62
Şekil 4.9 Farklı Yöntemlerden Elde Edilen Hata Miktarları (WMSE).....	64

## KISALTMALAR LİSTESİ

**MTK:** Madde Tepki Kuramı

**KTK:** Klasik Test Kuramı

**OO:** Ortalama-Ortalama Yöntemi

**OS:** Ortalama-Standart Sapma Yöntemi

**SL:** Stocking-Lord Yöntemi

**HA:** Haebara Yöntemi

**1PLM:** Bir Parametrelı Lojistik Model

**2PLM:** İki Parametrelı Lojistik Model

**3PLM:** Üç Parametrelı Lojistik Model

**WMSE:** Ağırlıklandırılmış Hata Kareleri Ortalaması

**RMSE:** Hata Kareleri Ortalamasının Karakökü

**BIAS:** Eşitleme Yanlılığı

**EAP:** Beklenen A Posteriori

**GRM:** Kısmi Kredi Modeli

**YDS:** Yabancı Dil Sınavı

**KPSS:** Kamu Personeli Seçme Sınavı

**PISA:** Uluslararası Öğrenci Değerlendirme Programı

**TIMSS:** Uluslararası Matematik ve Fen Bilgisi Çalışması

**ALES:** Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı

**OECD:** Ekonomik İşbirliği ve Kalkınma Teşkilatı

**MEB:** Milli Eğitim Bakanlığı

**EARGED:** Eğitimi Araştırma ve Geliştirme Daire Başkanlığı

# BÖLÜM I

## GİRİŞ

### 1.1 Problem Durumu

Uluslararası Öğrenci Değerlendirme Programı –PISA, Ekonomik İşbirliği ve Kalkınma Teşkilatı (OECD) tarafından düzenlenen dünyanın en kapsamlı eğitim araştırmalarından biridir. 2000 yılından itibaren yapılan bu araştırmayla zorunlu eğitimin sonuna gelen 15 yaş grubu öğrencilerin modern toplumda yerlerini alabilmeleri için gereken temel bilgi ve becerilere ne ölçüde sahip oldukları değerlendirilmektedir (Milli Eğitim Bakanlığı [MEB], 2013). Bu amaçla öğrencilerin okulda öğrendikleri bilgileri gerçek yaşamı yansıtan problem durumlarında ne derece kullanabildikleri ölçülmektedir. Uygulama ile aynı zamanda ülkelerin eğitim sistemleri de değerlendirilmektedir. Bu kapsamda ulusal boyutta alınan eğitsel kararların etkisi incelenerek elde edilen sonuçlara göre yeni politikalar geliştirilmektedir (MEB, 2011). Uluslararası anlamda ise eğitim sistemlerinin güçlü ve zayıf yönleri tespit edilerek ülkeler arası karşılaştırmalar yapılmaktadır.

PISA uygulamasında 13 farklı kitapçık kullanılmaktadır ve tüm sorular bu kitapçıklara harmanlanmış şekildedir. Bu nedenle öğrenciler soruların yalnızca bir kısmını görmektedir. Dolayısıyla kullanılan iki farklı kitapçıkta ortak maddelerin yanı sıra farklı maddeler de yer almaktadır.

PISA uygulaması gibi aynı anda birden fazla kitapçığın kullanıldığı ya da yıl içinde farklı dönemlerde uygulanan sınavlarda güvenlik nedeniyle her uygulamada farklı sorular sorulmaktadır. Bu sınavların farklı formlarını alan bireylere eşit davranması, yanlış davranmaması testin geçerliliği açısından son derece önemlidir (Kan, 2011). Ancak her formda kullanılan soruların değişmesi yansız olma ve eşitlik konularını gündeme getirmektedir (Cook ve Eignor, 1991). Çünkü test geliştiricilerin farklı sorulardan oluşan, fakat tamamıyla birbirine paralel test formları geliştirmeleri oldukça zordur ve bu nedenle test uygulamalarında oluşturulan formlar güçlük düzeyi bakımından bir miktar farklılık gösterebilmektedir (Kolen ve Brennan, 1995;



Kan, 2010). Örneğin bir sınav uygulamasında A formunu alan öğrenci B formunu alandan daha yüksek bir puan aldıysa bununla ilgili iki ihtimal vardır. Ya A formunu alan öğrenci daha başarılıdır ya da A formu B formundan daha kolaydır. Böyle bir durumda aynı yetenek düzeyinde olan iki öğrenciden kolay formu alan avantajlı olabilirken, zor formu alan dezavantajlı duruma düşebilmektedir. Benzer şekilde eğer belirli bir grubun performansını yıllara göre karşılaştırmak ve gelişim düzeyini izlemek istediğimizde öğrencilerin başarı düzeylerindeki artışın performanslarındaki gelişimden mi yoksa testin güçlük seviyesindeki değişiklikten mi kaynaklandığını bilemeyiz (Langer ve Swanson, 2010). Bu nedenle farklı soruları cevaplayan öğrenci performanslarını doğrudan karşılaştırabilmek için elde edilen puanların ortak bir ölçüğe yerleştirilmesi gerekmektedir. “Test eşitleme” olarak adlandırılan bu işlemle farklı formlardan doğrudan karşılaştırılabilir puanlar elde edilebilmektedir.

Test eşitleme çalışmaları ile sınav uygulamalarında daha adil ve yansız sonuçlar elde etmek mümkündür. Ancak eşitleme çalışmalarının yapılmadığı durumlarda yanıltıcı ve yorumlanması güç sonuçlar ortaya çıkabilmektedir. Örneğin aynı uygulamada birden fazla formun kullanıldığı bir sınavda birinci formdan 80 puan alan bir öğrenci ikinci formdan aynı puanı alamayabilir. Ya da aynı sınava farklı dönemlerde giren bir öğrenci sınavın güçlük seviyesine göre her uygulamada farklı puanlar alabilir. Bu da yanlış sonuçların ortaya çıkmasına neden olabilmektedir. Özellikle kesme puanının kullanıldığı test uygulamalarında bu durum adil bir değerlendirmenin yapılmasını engellemekte ve hatalı kararların alınmasına neden olabilmektedir. Bu nedenle özellikle önemli kararların alınacağı büyük ölçekli test uygulamalarında kullanılan farklı kitapçıkların istatistiksel eşitliğinin sınanması gerekmektedir.

## **1.2 Araştırmanın Amacı**

Birden fazla formun kullanıldığı test uygulamalarında farklı formlardan elde edilen puanlar test eşitleme çalışmaları ile aynı ölçüğe yerleştirilebilmektedir. Ancak gerçekleştirilen eşitleme çalışmalarında elde edilen sonuçlar çeşitli faktörlerden (örneklem büyüklüğü, ortak madde sayısı, test uzunluğu, yetenek dağılımı vb.) etkilenmektedir. Bu nedenle hangi şartlarda hangi eşitleme yönteminin en iyi sonucu verebileceği ile ilgili çeşitli simülasyon çalışmaları gerçekleştirilmektedir. Bu

alıřmalardan elde edilen sonuların gerek test durumlarıyla da tutarlılık gsterip gstermediĐinin incelenmesi gerekmektedir. Bu nedenle bu alıřmanın amacı madde tepki kuramına dayalı test eřitleme yntemlerini gerek veriler zerinde karřılařtırmak ve en az hata veren yntemi tespit etmektir.

### **Problem Cmlesi**

PISA 2012 Fen okuryazarlıĐı testinden seilen 1 ve 8 numaralı iki kitapıktan elde edilen puanların eřitlenmesinde madde tepki kuramına dayalı lek dnřtrme (ortalama-ortalama, ortalama-standart sapma, Stocking-lord, Haebara) ve test eřitleme yntemlerinden (MTK gerek-puan, MTK gzlenen-puan) en az eřitleme hatasına sahip yntem hangisidir?

Arařtırmanın problem cmlesi doĐrultusunda ařaĐıdaki alt problemlere cevap aranmıřtır.

PISA 2012 Fen okuryazarlıĐı testinden seilen 1 ve 8 numaralı iki kitapıktan elde edilen puanların MTK gerek-puan ve MTK gzlenen-puana gre eřitlenmesinde;

1. Ortalama-Ortalama ynteminin kullanılması sonucu elde edilen eřitlenmiř puanlar nasıldır?
2. Ortalama-Standart Sapma ynteminin kullanılması sonucu elde edilen eřitlenmiř puanlar nasıldır?
3. Stocking-Lord ynteminin kullanılması sonucu elde edilen eřitlenmiř puanlar nasıldır?
4. Haebara ynteminin kullanılması sonucu elde edilen eřitlenmiř puanlar nasıldır?
5. En az eřitleme hatasına sahip lek dnřtrme ve test eřitleme yntemi hangisidir?

### **1.3 Araştırmanın Önemi**

Büyük ölçekli sınavların yılda birden fazla yapılması ya da aynı uygulamada güvenlik sebebiyle farklı formların kullanılması ile test eşitleme çalışmaları önem kazanmıştır. Özellikle PISA gibi uluslararası karşılaştırmaların yapıldığı, önemli kararların alındığı bir uygulamada kullanılan farklı soruların istatistiksel eşitliğinin sınanması gerekmektedir. Fakat eşitleme çalışmalarında kullanılacak yöntemin seçimi kritik öneme sahiptir. Çünkü kullanılan yöntem eşitleme sonuçlarını değiştirebilmekte, yanlış yöntem seçimi yüksek eşitleme hatalarının elde edilmesine neden olabilmektedir. Bu nedenle olabildiğince hatadan arınık eşitleme çalışmalarının yapılması gerekmektedir. Literatür incelendiğinde hangi test eşitleme yönteminin en iyi sonucu verdiği ile ilgili ortak bir görüş bulunmamaktadır. Gerçekleştirilen birçok çalışmada eşitleme koşulları değiştikçe en az hata veren yöntemin de değiştiği görülmektedir. Alanyazın incelendiğinde test eşitleme ile ilgili genellikle simülasyon çalışmalarının (Gök, 2012; Wingersky, Cook ve Eignor, 1987; Kilmen, 2010; Hanson ve Beguin, 2002; Kim ve Cohen, 2002; Speron, 2009; Bastari, 2000) yapıldığı görülmektedir. Gerçek verilerle yapılan çalışmaların ise daha çok klasik test kuramına dayalı olduğu görülmektedir (Bozdağ, 2007; İnci, 2014; Çetin, 2009; Kan, 2011; Demir ve Güler, 2014). Türkiye’de madde tepki kuramına dayalı gerçek verilerle yapılan çalışmalar ise sınırlı sayıdadır. Bu nedenle söz konusu çalışmanın literatüre katkı sağlayacağı ve Türkiye’de yapılmayan test eşitleme çalışmalarına dikkat çekeceği düşünülmektedir.

### **1.4 Varsayımlar**

1. Öğrenciler PISA 2012 fen okuryazarlığı testinde soruları cevaplarken gerçek performanslarını göstermişlerdir.
2. PISA 2012 uygulamasında sorulara verilen cevaplar tarafsız ve doğru bir şekilde puanlanmıştır.

## 1.5 Sınırlılıklar

1. Bu çalışma, PISA 2012 fen okuryazarlığı testinde kullanılan 1 ve 8 numaralı kitapçıklardan alınan puanlarla sınırlıdır.
2. Araştırma 1-0 şeklinde puanlanan iki kategorili verilerle sınırlıdır.
3. Kullanılan yöntemler madde tepki kuramına dayalı ayrı kalibrasyon yöntemleri ile sınırlıdır.

## 1.6 Tanımlar

**Test Eşitleme:** Test formlarından alınan puanların ortak bir ölçüğe yerleştirilerek karşılaştırılabilir puanların elde edildiği istatistiksel süreç.

**Madde Tepki Kuramı:** Bireyin yetenek seviyesine göre bir maddeyi doğru cevaplama olasılığı arasındaki ilişkiyi açıklayan matematiksel kuram.

## BÖLÜM II

### KAVRAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

#### 2.1 Uluslararası Öğrenci Değerlendirme Programı (PISA)

Uluslararası Öğrenci Değerlendirme Programı –PISA (Programme for International Student Assessment) Ekonomik İşbirliği ve Kalkınma Teşkilatı (Organisation for Economic Co-Operation and Development- OECD) tarafından üç yılda bir düzenlenmektedir. İlk olarak 2000 yılında başlayan programa Türkiye 2003 yılından itibaren katılmaktadır. Uygulama, öğrencilerin okulda öğrendiklerini günlük yaşamda karşılaştıkları problem durumlarına ne derece uygulayıp uygulamadıklarını ölçmeyi amaçlamaktadır. Bu anlamda PISA, öğrencilerin ne bildiklerine değil, bildiklerini ne derece kullanabildiklerine odaklanmaktadır (Ekonomik İşbirliği ve Kalkınma Teşkilatı [OECD], 2014a).

Uygulama ile aynı zamanda politika belirleyiciler kendi öğrencilerinin performansını diğer ülke öğrencileriyle karşılaştırarak eğitim sisteminin güçlü ve zayıf yönlerini belirlemekte ve eğitim kalitesinin yükseltilmesi amacıyla standartlar oluşturmaktadırlar (MEB, 2011). PISA’da ülkeler arası karşılaştırmaların yanı sıra her bir ülkenin performansını belirleyen faktörler de incelenmektedir. Bu anlamda bilişsel maddelerin yanı sıra uygulanan anketlerle öğrencilerden kendileri, aileleri, okulları ve öğrenme deneyimleriyle ilgili; müdürlerden ise okul sistemi ve öğrenme ortamı ile ilgili bilgiler alınmaktadır. Böylece öğrenci başarısının bu anketlerden elde edilen sonuçlarla ilişkisine bakılarak başarı ya da başarısızlığın altında yatan sebepler incelenmektedir (Şirin ve Vatanartıran, 2014).

Uygulamada diğer değerlendirme uygulamalarından farklı olarak “okuryazarlık” kavramı kullanılmaktadır. Okuryazarlık kavramı, öğrencilerin bilgilerini günlük yaşamda kullanarak mantıksal çıkarımlar yapmak, çeşitli durumlarla ilgili problemleri yorumlamak ve çözmek için öğrendiklerinden çıkarımlar yapma kapasitesidir (Eğitimi Araştırma ve Geliştirme Daire Başkanlığı [EARGED], 2010).

PISA uygulaması okuma, matematik ve fen okuryazarlığı alanlarını içermektedir. Fen okuryazarlığı kavramı, bilimsel bilginin günlük yaşamda kullanımının

değerlendirilmesine odaklanmaktadır (MEB, 2013). PISA’da puanların kolay yorumlanabilmesi için soruların zorluk derecesine göre yeterlilik düzeyleri belirlenmiştir. 2012 uygulamasında fen alanında 7 düzey (1. Düzey altı, 1-6 düzeyler) belirlenmiş ve her düzeyde yer alan öğrencilerin hangi becerilere sahip oldukları açıklanmıştır (Şirin ve Vatanartıran, 2014). Fen alanındaki yeterlikler, öğrencilerin sahip oldukları bilimsel bilgi, bilim hakkındaki bilgileri ve fen alanına yönelik tutumları ile günlük yaşamda karşılaşılabilecekleri durumlarda bu becerileri kullanabilmelerine göre değerlendirilmektedir. 2012 uygulamasında fen alanındaki alt alanlar “sağlık”, “doğal kaynaklar”, “afetler” ve “bilim ve teknolojinin öncüleri” olarak belirlenmiştir (MEB, 2013).

PISA 2012 uygulamasına 34’ü OECD ülkesi olmak üzere toplam 65 ülkeden 15 yaşında yaklaşık 510 bin öğrenci katılmıştır (OECD, 2014a). Uygulamanın Türkiye örneklemini ise 15 yaş grubunda 170 okuldan seçilen toplam 4848 öğrenci oluşturmaktadır (MEB, 2013). Öğrenciler Türkiye İstatistik Kurumu (TUİK) tarafından belirlenen 12 istatistiki bölgeyi temsil edecek şekilde seçilmiştir.

Uygulamada 13 farklı soru kitapçığı kullanılmış ve farklı soru kümeleri kitapçıklara belirli bir döngü ile yerleştirilmiştir. Kitapçıklarda kullanılan madde türleri seçeneklerden birinin seçildiği çoktan seçmeli maddeler, bir dizi doğru/yanlış ya da evet/hayır seçeneklerinden birinin işaretlendiği karmaşık çoktan seçmeli maddeler, kısa yanıt gerektiren kapalı uçlu maddeler, kısa yanıt gerektiren açık uçlu maddeler ve uzun yanıt gerektiren açık uçlu maddelerden oluşmaktadır (MEB, 2011). Öğrencilere kağıt-kalem testi için 2 saat, bilgisayar tabanlı değerlendirme için ise 40 dakika süre verilmiştir (OECD, 2014b). Bilişsel maddelerin ardından öğrencilerden kendileri, aileleri ve okul çevresi ile ilgili birtakım soruları içeren anketi 30 dakika içinde doldurmaları istenmiştir. Cevapların puanlanması aşamasında açık uçlu sorular birbirinden bağımsız 4 uzman tarafından puanlanmış, belirli bir bölümü doğru olan sorulara ise kısmi puan verilmiştir (OECD, 2010).

## 2.2 Test Eşitleme

Test eşitleme, benzer içeriğe sahip farklı test formları arasındaki güçlük farklılıklarını düzenleyerek karşılaştırılabilir puanların elde edildiği istatistiksel bir süreçtir (Kolen ve Brennan, 2014). Kaskowitz ve De Ayala (2001), test eşitlemeyi farklı formlardan alınan puanları karşılaştırabilmek amacıyla gerçekleştirilen istatistiksel bir düzeltme işlemi olarak tanımlamışlardır. Angoff (1984) a göre ise eşitleme, bir formun birim sistemini diğer bir formun birim sistemine dönüştürmektir. Bu dönüşüm, inç i santimetreye, pound u grama ya da Fahrenheit’ı Celcius’a dönüştürmek olarak düşünülebilir. Ancak burada önemli olan formların aynı özelliği, yeteneği ya da beceriyi ölçmesidir (Tsai, Hanson, Kolen ve Forsyth, 2001). Fahrenheit ve Celcius her ikisi de sıcaklığı ölçer, ancak farklı ölçeklerde yer aldıkları için elde edilen sıcaklıklar ancak dönüştürüldükten sonra karşılaştırılabilir. Test eşitleme çalışmalarında da farklı formlardan elde edilen puanlar doğrusal bir dönüşümle ortak bir ölçeğe yerleştirilerek birbirinin yerine kullanılabilir.

Test eşitleme ile bir testin farklı formlarını alan iki kişinin performansı doğrudan karşılaştırılabilir (Angoff, 1984). Örneğin başarılı bir eşitlemeden sonra bir öğrenci eşitlenmiş bir testten 26 puan aldıysa, bu öğrencinin farklı bir eşitlenmiş testten 26 puan alan biriyle ortalama aynı başarı düzeyinde olduğu söylenebilir (Kolen ve Brennan, 2014). Doğru bir eşitleme ile büyümenin, gelişimin ölçülmesi ve eğilimlerin tespit edilmesi sağlanabilmektedir (Crocker ve Algina, 1986; Kan, 2011). Ayrıca test geliştirme, puanlama ve yorumlama gibi durumlarda da eşitleme çalışmaları önem kazanmakta ve sürecin daha adil olmasını sağlamaktadır (Felan, 2002). Test eşitleme ihtiyacı iki durumda ortaya çıkmaktadır:

1. Eşitlenecek testlerin güçlük düzeyi ile testi alan grupların yetenek dağılımlarının benzer olduğu durumlarda,
2. Eşitlenecek testlerin güçlük düzeyi ile testi alan grupların yetenek dağılımlarının farklı olduğu durumlarda ortaya çıkmaktadır.

Yukarıdaki durumlarda gerçekleştirilecek eşitleme çalışmaları “yatay” ve “dikey” eşitleme olmak üzere iki başlık altında incelenmektedir (Hambleton ve Swaminathan, 1985).

### **2.2.1 Yatay Eşitleme**

Bir testin farklı formlarından alınan puanların eşitlenmesi yatay eşitleme olarak adlandırılır (Crocker ve Algina, 1986; Woldbeck, 1998). Yatay eşitlemeye bir testin güvenlik sebebiyle birden fazla formunun kullanılması gerektiği durumlarda başvurulur (Hambleton ve Swaminathan, 1985). Bu eşitlemede amaç, benzer psikometrik özelliklere sahip test formlarından doğrudan karşılaştırılabilir puanlar elde etmektir. Yaygın olarak kullanılan bu eşitleme türü, genellikle standartlaştırılmış testlerde test güvenliğini sağlamak için oluşturulan farklı formları eşitlemek için kullanılır (Skaggs ve Lissitz, 1986). Yatay eşitlemenin gerçekleştirilebilmesi için test formlarının içerik, güvenilirlik ve güçlük düzeyi bakımından olabildiğince birbirine paralel olması gerekmektedir. Ayrıca bireylerin yetenek dağılımlarının da benzer olması beklenir (Kolen, 1988). Eğer gruplararası yetenek farklılıkları fazlaysa geleneksel eşitleme yöntemleri uygun olmayabilir (Zhu, 1998). Bu eşitleme türüne, yılda birden fazla uygulanan ALES, YDS gibi sınavların farklı formlarını eşitlemek ya da PISA, TIMSS gibi aynı test uygulamasında kullanılan birden fazla kitapçığı eşitlemek örnek olarak verilebilir. Bu araştırmada kitapçıkları eşitlemek için yatay eşitleme yöntemi kullanılmıştır.

### **2.2.2 Dikey Eşitleme**

Dikey eşitleme, aynı bilgi ve becerileri ölçen ancak güçlük düzeyi bakımından farklılaşan iki ya da daha fazla testin eşitlenmesini kapsar (Loyd ve Hoover, 1980; Crocker ve Algina, 1986). Yatay eşitlemeden farklı olarak bu eşitlemede bireylerin yetenek dağılımları birbirinden farklılık gösterir (Felan, 2002).

Dikey eşitlemeye genellikle ilkokulda başarı testlerinin eşitlenmesinde başvurulur (Kolen ve Brennan, 2014). Bu eşitleme ile öğrencilerin zaman içerisinde akademik başarılarındaki gelişimi izlemek mümkündür. Ayrıca farklı sınıflarda öğrenim gören öğrencilerin yetenek düzeyleri karşılaştırılabilir. Ancak dikey eşitleme yatay eşitlemeye göre daha karmaşık bir süreç gerektirir (Eid, 2005). Çünkü farklı yetenek düzeyindeki gruplar için farklı güçlük düzeyinde hazırlanmış testler arasında bağlantı kurmayı gerektirir (Skaggs ve Lissitz, 1986).



### 2.3 Eşitleme Varsayımları

Test eşitlenmenin gerçekleştirilebilmesi için bazı varsayımların karşılanması gerekmektedir (Petersen, Kolen ve Hoover, 1989).

**1. Aynı Yapıyı Ölçme:** Bu özellik eşitlenecek iki testin aynı yeteneği, özelliği ya da beceriyi ölçmesini ifade eder (Angoff, 1984). Holmes (1986) başarı testi gibi birden fazla alanı kapsayan çok boyutlu testler için eşitlemenin alt test düzeyinde yapılmasını önermiştir.

**2. Simetri:** Bu özellik eşitleme dönüşümlerinin simetrik olmasını gerektirir (Livingston, 2004). Örneğin, form A'dan alınan 20 puan, form B'den alınan 22 puana karşılık geliyorsa, form B'den alınan 22 puan da form A'dan alınan 20 puana karşılık gelmelidir.

**3. Eşitlik:** Lord (1980) tarafından öne sürülen eşitlik özelliğine göre bir cevaplayıcının eşitlenen iki testten (X ve Y) herhangi birini alması fark oluşturmamalıdır (Hambleton ve Swaminathan, 1985). Ancak uygulamada bu koşulu sağlamak oldukça zordur ve kullanımı daha çok teoride kalmaktadır (Crocker ve Algina, 1986).

**4. Gruplararası Değişmezlik:** Bu özellik eşitleme ilişkisinin eşitlemenin yapıldığı gruptan bağımsız olmasını yani tüm altgruplarda aynı eşitleme ilişkisinin elde edilmesini ifade eder. Dorans ve Holland (2000) bu varsayımın alt populasyonlar arasındaki eşitlik ilişkisinin karşılaştırılarak incelenebileceğini belirtmişlerdir.

### 2.4 Eşitleme Basamakları

Test eşitlenme için gerekli koşullar incelendikten sonra eşitleme basamaklarına geçilmelidir. Kolen ve Brennan (2014)'a göre eşitleme çalışmaları için şu aşamalar izlenmelidir:

1. Eşitlemenin amacına karar verilmesi
2. Alternatif formların oluşturulması
3. Veri toplama desenine karar verilmesi

4. Veri toplama desenine uygun olarak testlerin uygulanması
5. Eşitlemenin bir ya da birden fazla işlevsel tanımının seçilmesi
6. Bir ya da daha fazla istatistiksel kestirim yönteminin seçilmesi
7. Eşitleme sonuçlarının değerlendirilmesi

## **2.5 Eşitleme Desenleri**

Eşitlemenin amacı belirlenip alternatif formlar oluşturulduktan sonra hangi veri toplama deseninin kullanılacağına karar verilir. Eşitlemede veri toplama deseni “eşitleme deseni” olarak ifade edilir (Kolen ve Brennan, 2014). Her desenin kendi içinde avantaj ve dezavantajları vardır. Desen seçiminde hem uygulanabilirlik hem de istatistiksel özellikler göz önünde bulundurulmalıdır. Literatürde yaygın olarak kullanılan eşitleme desenleri dört başlık altında sınıflandırılmıştır (Kolen ve Brennan, 2014; Holland ve Dorans, 2006).

### **2.5.1 Tek Grup Deseni**

En basit test eşitleme desendir. Bu desende test formları aynı katılımcı grubuna uygulanır. Her iki form da aynı gruba uygulandığı için testlerin güçlük seviyesinden kaynaklanan farklılıklar gruplar arasındaki yetenek farklılıklarından etkilenmez (Hambleton ve Swaminathan, 1985). Bu desende grupların hedef evreni temsil etmelerine gerek yoktur (Livingston, 2004). Ayrıca küçük örneklerde de düşük eşitleme hataları elde edilebilmektedir (Dorans, Moses ve Eignor, 2010).

Ancak bu desende grubun önce X daha sonra da Y formunu alması durumunda yorgunluk ve sıra etkisinden dolayı Y formu daha zor gibi görünebilir. Ya da katılımcılar X formunu aldıktan sonra Y formunu alırlarsa teste aşinalıktan dolayı Y formu daha kolay gibi görünebilir. Bu gibi sıra etkilerinden dolayı bu desen nadir olarak kullanılmakta ve genellikle iki formu da alan katılımcıların olduğu özel durumlarda tercih edilmektedir (Kolen ve Brennan, 2014; Von Davier, Holland, ve Thayer, 2004).

Tablo 2.1 Tek Grup Deseni

Grup	Form X	Form Y
1	✓	✓

### 2.5.2 Dengelenmiş Tek Grup Deseni

Bu desen tek grup desenindeki sıra etkisini ortadan kaldırmak için geliştirilmiştir. Desende birinci gruba önce X, sonra Y formu; ikinci gruba ise önce Y sonra da X formu uygulanır. Bu sayede ilk olarak X formunu alan grup puanları ile ilk olarak Y formunu alan grup puanları karşılaştırılabilir. Ancak iki test arasındaki süre, cevaplayıcıların bilgi düzeylerinde değişim olmayacak şekilde ayarlanmalıdır (Kolen, 2007; Livingston, 2004). Tek grup deseninde olduğu gibi bu desende de örneklem grubunun hedef evreni temsil etme zorunluluğu yoktur, gruplar hedef evrenden daha başarılı ya da daha başarısız olabilirler. Aynı şekilde küçük örneklemelerde de eşitleme çalışması yapılabilir. Ancak bu desen uygulama için iki kat test süresi ve eşitleme için özel bir uygulama gerektirir (Dorans vd., 2010).

Tablo 2.2 Dengelenmiş Tek Grup Deseni

Grup	İlk Uygulama		İkinci Uygulama	
	X	Y	X	Y
1	✓			✓
2		✓	✓	

### 2.5.3 Eşdeğer Gruplar Deseni

Uygulama kolaylığı nedeniyle yaygın olarak kullanılan bir desendir. Desen, iki şekilde uygulanabilir. Birincisi büyük ve heterojen bir grup seçkisiz olara ikiye ayrılır ve her iki gruba farklı bir form verilir. İkincisi ise sarmal bir şekilde kitapçıklar dağıtılır. Birinci kişiye X formu, ikinciye Y formu, üçüncüye ise tekrar X formu verilir. Ancak burada eşitleme sonuçlarını etkilememesi için cevaplayıcıların belirli bir sıraya göre (kız-erkek; kız-erkek gibi.) oturmamaları gerekir (Angoff, 1984; Kolen 2007; Dorans vd., 2010).

Bu desende grup performansları arasındaki fark, formlar arasındaki güçlük farklılıklarının doğrudan göstergesidir. Bu nedenle örneklemeden kaynaklı yanlılığı önlemek için büyük ve heterojen örneklemlerle çalışmak gerekir (Livingston, 2004; Angoff, 1984). Bu desenin avantajı uygulama kolaylığı ve tek grup desenindeki sıra etkisinin ortadan kaldırmasıdır. Ayrıca bu desende her bir katılımcının sadece test formlarından birini almasıyla test süresi minimize edilmiş olur ve ikiden fazla test formu aynı anda eşitlenebilir (Von Davier, Holland, ve Thayer, 2004). Desenin dezavantajı ise dengelenmiş tek grup deseninde olduğu gibi veri toplamak için özel bir uygulama gerektirmesidir (Kolen, 2007).

*Tablo 2.3 Eşdeğer Gruplar Deseni*

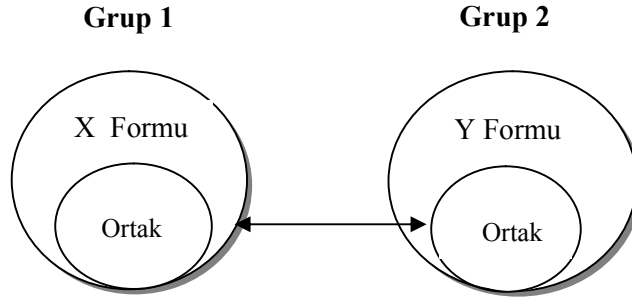
Grup	Form X	Form Y
1	✓	
2		✓

#### **2.5.4 Denk Olmayan Gruplarda Ortak Madde Deseni**

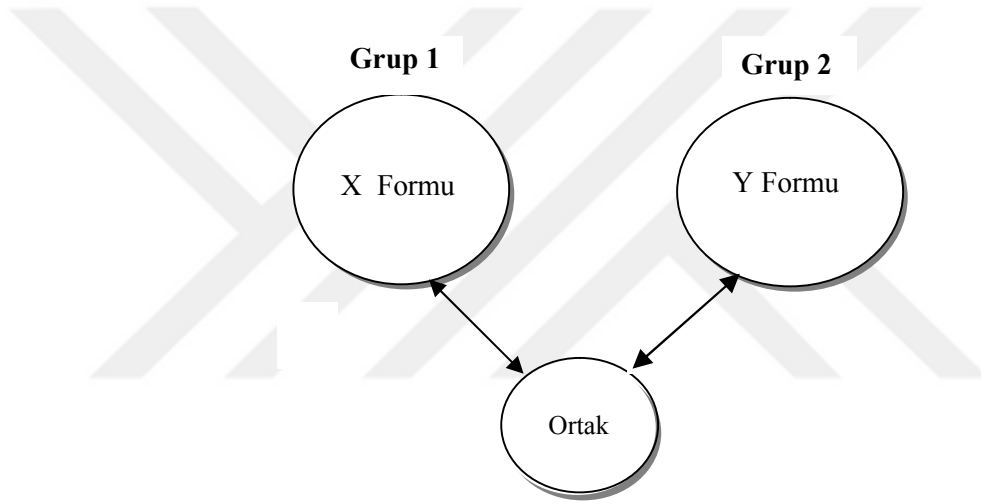
Literatürde en yaygın olarak kullanılan eşitleme desendir (Kim, Walker ve McHale, 2010; Felan, 2002; Kolen ve Brennan, 2014; Norman-Dvorak, 2009). Bu desen genellikle test güvenliği açısından birden fazla formu uygulamanın mümkün olmadığı durumlarda kullanılır (Zhu, 1998). Desende ortak maddelere sahip form X ve form Y iki farklı gruba uygulanır. Ortak maddeler sayesinde her iki formdan elde edilen madde parametreleri aynı ölçeğe yerleştirilir (Embretson ve Reise, 2000; Cook ve Eignor, 1991). Ortak maddelerden alınan puanlar iki grubun performansı arasındaki farklılığın doğrudan göstergesidir ve iki test arasındaki farklılıklar ortak maddelerin kullanımıyla düzeltilebilmektedir. Ayrıca ortak maddeler eşitleme fonksiyonunda gruplar arasındaki farklılıktan kaynaklanabilecek yanlılığı ortadan kaldırarak fonksiyonunun tahminindeki hassaslığı artırır (Powers ve Kolen, 2014; Zhu, 1998; Dorans vd., 2010).

Desen iki şekilde uygulanabilir. Ortak maddeler diğer test maddeleri ile birlikte uygulanıyorsa buna “iç ortak madde” adı verilir ve kişinin ortak maddelerden aldığı puan test puanına dahil edilir. Eğer ortak maddeler testten ayrı bir süre zarfında

uygulanıyorsa buna da “dış ortak madde” adı verilir ve ortak maddelerden alınan puan test puanına dahil edilmez (Kolen ve Brennan, 2014; Livingston, 2004).



Şekil 2.1. İç Ortak Maddelere İlişkin Görsel



Şekil 2.2. Dış Ortak Maddelere İlişkin Görsel

Bu desenin en önemli avantajı uygulamasının esnek olmasıdır (Petersen vd., 1989; Sinharay ve Holland, 2010). Desende eşitleme yapılacak iki grup denk olmak zorunda değildir. Bu sayede diğer desenlerde karşılaşılan güçlüklerin üstesinden gelinir (Hambleton ve Swaminathan, 1985). Ayrıca her gruba tek bir test formu uygulanarak test süresi kısaltılır. Ancak bu desende ortak madde sayısı eşitleme sonuçlarını etkilemektedir. Angoff (1984)'a göre bir testteki ortak madde sayısı 20'den az olmamalı ya da testin en az %20'si kadar olmalıdır. Woldbeck (1998)'e göre ise ortak madde sayısı test uzunluğunun en az %20-25'i kadar olmalıdır.

Yapılan arařtırmalar ortak madde sayısı arttıkça eřitleme hatasının dūřtūđünü gōstermiřtir (Kolen ve Brennan, 2014).

*Tablo 2.4 Denk Olmayan Gruplarda Ortak Madde Deseni*

Grup	Form X	Ortak	Form Y
1	✓	✓	
2		✓	✓

## 2.6 Test Eřitleme Yōntemleri

Eřitleme yōntemleri, test eřitlemeyi gerekleřtirmede kullanılan iřlem sūrecidir (Öztürk, 2010). Eřitleme yōntemleri dayandıđı kuramsal temele gōre klasik test kuramına dayalı yōntemler ve madde tepki kuramına dayalı yōntemler olmak üzere iki bařlık altında incelenir. Gerekleřtirilen birok alıřmada (Petersen, Cook ve Stocking, 1983; Yang ve Houang, 1996; Caldwell, 1984; Han, Kolen ve Pohlmann, 1997; von Davier ve Wilson, 2007; Yang, 1997; Chen, 2001; Meng, 2012) MTK'ya dayalı eřitleme yōntemleri KTK'ya dayalı yōntemlerden daha kararlı ve daha az hata oranı verdiđi iin bu alıřmada madde tepki kuramına dayalı eřitleme yōntemleri kullanılmıřtır.

### 2.6.1 Madde Tepki Kuramına Dayalı Eřitleme

Madde tepki kuramı, “gizil deđiřken teorisi” ya da “madde karakteristik eđrisi teorisi” olarak da bilinir (Embretson ve Reise, 2000). Bu kuram, farklı yetenek dūzeyindeki cevaplayıcıların ölçülen özellik aısından bir maddeye verebilecekleri cevabı tanımlayan matematiksel bir model sunar (Hambleton ve Jones, 1993).

Klasik test kuramına dayalı eřitlenmenin bazı zayıf noktalarından dolayı madde tepki kuramına dayalı test eřitleme önem kazanmıřtır (Hambleton, 1989). Klasik test kuramında bazı varsayımlarının (eřitlik, simetri, gruptan bađımsızlık) karřılanmasında sıkıntılar yařanmaktadır (Keleciođlu, 1994). Ayrıca MTK'da madde parametreleri katılımcı gruptan, yetenek parametreleri ise cevaplanan madde setinden bađımsız olarak kestirilebilmektedir. “Parametre deđiřmezliđi” olarak adlandırılan bu ilkeye gōre bireyin zor ya da kolay testi alması sonucu deđiřtirmemektedir (Hambleton ve Swaminathan, 1985). MTK ile eřitleme yetenek seviyesi bakımından

denk olmayan gruplara uygulanan, güvenilirlik ve güçlük düzeyi bakımından farklılaşan testlerin eşitlenmesinde etkili bir yöntemdir (Cook ve Eignor, 1991). Ancak MTK ile eşitleme yapılabilmesi için bazı varsayımların karşılanması gerekmektedir. Temel varsayımlardan ilki tek boyutluluktur. Tek boyutlu madde tepki kuramı bireyin test performansının altında tek bir gizil yeteneğin (örn. matematik yeteneği) olduğunu varsayar (De Ayala, 2009). Fakat bu varsayım tam olarak karşılanmayabilir. Çünkü bazı bilişsel ya da kişisel faktörler de (motivasyon düzeyi, test kaygısı vb.) test performansını etkileyebilir. Bu varsayımın karşılanması için test performansını etkileyen baskın bir faktörün olması yeterli görülmektedir. Bu baskın faktör testle ölçülmek istenen özellik olarak tanımlanmaktadır (Hambleton ve Swaminathan, 1985). İkinci varsayım ise yerel bağımsızlıktır. Yerel bağımsızlık yetenek değişkeni sabit tutulduğunda herhangi iki maddeye verilen cevapların istatistiksel olarak birbirinden bağımsız olması demektir (Lord, 1980; akt. Hambleton ve Swaminathan, 1985). Örneğin bir maddenin içeriği diğer bir maddenin cevabı hakkında ipucu veriyorsa yerel bağımsızlık ihlal edilmiş olur (Embretson ve Reise, 2000).

Madde tepki kuramı bireyin performansı ile performansı etkilediği düşünülen birtakım özellikler arasındaki ilişkiyi “madde karakteristik eğrisi” olarak adlandırılan bir fonksiyonla tanımlar (Hambleton ve Swaminathan, 1985). Madde karakteristik eğrisi, farklı yetenek düzeyindeki cevaplayıcıların bir maddeyi doğru cevaplama olasılığı arasındaki ilişkiyi gösterir (Baker, 2001). Madde karakteristik eğrisinin hangi matematiksel fonksiyonla gösterileceğini belirlemek için farklı modeller geliştirilmiştir. İki kategorili puanlanan maddeler için en yaygın olarak kullanılan modeller bir parametrelili (1PLM), iki parametrelili (2PLM) ve üç parametrelili (3PLM) lojistik modellerdir (Embretson ve Reise, 2000).

### ***1 Parametrelili Lojistik Model (1PLM)***

En basit MTK modelidir. Bu model tüm maddelerin eşit ayıricılığa sahip olduğunu ve düşük yetenek düzeyindeki bireylerin şansla doğru cevap verme olasılığının olmadığını varsayar (Crocker ve Algina, 1989). Modelde, bağımsız değişkenler kişinin yetenek düzeyi ve madde güçlük parametresidir. Ayıricılık parametresi ise tüm maddeler için sabittir (Embretson ve Reise, 2000). Matematiksel olarak bir

parametrelili model ile Rasch modeli birbirine eşittir. Aralarındaki fark Rasch modelinde ayırıcılık parametresi 1'e sabitlenmiştir. Ancak bir parametrelili modelde 1'e eşit olmak zorunda değildir (De Ayala, 2009). Rasch modelinde küçük örneklemlerde de eşitleme çalışması yapılabilmektedir (Baker, 2001). Yetenek düzeyine göre bir maddeyi doğru cevaplama olasılığı aşağıdaki eşitlikle hesaplanır:

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

$P_i\theta$  :  $\theta$  yeteneğindeki bir kişinin  $i$  maddesini doğru cevaplama olasılığı

$b_i$  :  $i$  maddesine ait madde güçlük indeksi

$D_i$  : ölçek sabiti

## **2 Parametrelili Lojistik Model (2PLM)**

Bu modelde güçlük parametresine ek olarak ayırıcılık parametresi de eklenir. Ayırıcılık parametresi teorik olarak  $-\infty$  ile  $+\infty$  arasında değer almaktadır. Ancak uygulamada genellikle 0 ile 2 arasında değişmektedir. Değer 2'ye yaklaştıkça maddenin ayırıcılığı yükselmektedir (Hambleton vd., 1991). Bu modelde de bir parametrelili modelde olduğu gibi şans başarısının olmadığı varsayılır (Hambleton ve Swaminathan, 1985). İki parametrelili model aşağıdaki eşitlik yardımıyla hesaplanır:

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

$P_i\theta$  :  $\theta$  yeteneğindeki bir kişinin  $i$  maddesini doğru cevaplama olasılığı

$b_i$  :  $i$  maddesine ait madde güçlük indeksi

$a_i$  :  $i$  maddesine ait madde ayırıcılık indeksi

$D_i$  : ölçek sabiti



### 3 Parametrelı Lojistik Model (3PLM)

Bu model iki parametrelı modele üçüncü bir parametre olan şans parametresinin eklenmesiyle elde edilir. Şans parametresi düşük yetenek düzeyindeki bireylerin bir soruyu şansla doğru cevaplama ihtimalini gösterir (Lord, 1980; akt. Hambleton ve Swaminathan, 1985). Şans parametresi teorik olarak 0 ile 1 arasında, uygulamada ise genellikle 0 ile .35 arasında değerler alır (Baker, 2001). Üç parametrelı model özellikle çoktan seçmeli testlerde önem taşımaktadır. Ancak bu model madde ve yetenek kestirimi için büyük örneklemeler ve uzun testler gerektirir (Hambleton ve Swaminathan, 1985). Modelde bireyin maddeyi doğru cevaplama olasılığı aşağıdaki eşitlikle hesaplanır:

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

$P_i\theta$  :  $\theta$  yeteneğindeki bir kişinin  $i$  maddesini doğru cevaplama olasılığı

$b_i$  :  $i$  maddesine ait madde güçlük indeksi

$a_i$  :  $i$  maddesine ait madde ayırıcılık indeksi

$c_i$  :  $i$  maddesine ait şans parametresi

### 2.7 Madde Tepki Kuramına Dayalı Kalibrasyon Yöntemleri

Madde parametreleri madde tepki kuramına dayalı modellerle kestirildikten sonra parametrelerin ortak bir ölçeğe yerleştirilmesi gerekir. Çünkü elde edilen parametreler farklı örneklemelerden elde edildiği için aynı ölçek üzerinde değildir. Parametre kestirimi ve dönüştürme yoluyla parametreleri ortak bir ölçeğe yerleştirme süreci “kalibrasyon” olarak adlandırılır (Kolen ve Brennan, 2004). Kalibrasyon süreci boyunca madde ve yetenek parametreleri aynı ölçekte kestirilir. Bu sayede farklı formları alan katılımcılar birbirleriyle karşılaştırılabilir (Crocker ve Algina, 1989). Parametreleri ortak bir ölçeğe yerleştirmek için çeşitli kalibrasyon yöntemleri kullanılmaktadır. Bunlar parametrelerin tek bir analizle kestirildiği “eşzamanlı kalibrasyon” ve parametrelerin ayrı ayrı kestirildiği “ayrı kalibrasyon” yöntemleridir.

### 2.7.1 Eş Zamanlı Kalibrasyon

Eş zamanlı kestirimde kalibrasyon işlemi tek bir aşamada gerçekleştirilir (Chu, 2002). Bu yöntemde her iki formdaki tüm maddeler için madde ve yetenek parametreleri tek bir analizle aynı anda kestirilir (Hanson ve Beguin, 2002). Bu şekilde iki veri seti eş zamanlı olarak kalibre edilerek ortak bir metrik oluşturulur (Kim ve Cohen, 2002). Bu yöntemde ortak maddelerin her iki formda da aynı parametrelere sahip olduğu varsayılır ve dönüştürme işlemi ortak madde parametreleri yardımıyla gerçekleştirilir (Nozawa, 2008; Wingersky vd., 1987). Ayrıca bu yöntemde tüm parametreler aynı ölçekte yer aldığı için doğrusal bir ölçek dönüştürme işlemine gerek kalmaz (Tsai vd., 2001; Kim ve Kolen, 2006).

### 2.7.2 Ayrı Kalibrasyon

Bu yöntemde kalibrasyon işlemi iki aşamada gerçekleştirilir. Öncelikle iki farklı test formu için madde ve yetenek parametreleri ayrı ayrı kestirilir. Ardından kestirilen parametreler için lineer bir dönüşüm işlemi gerçekleştirilir (Cook ve Eignor, 1991). Bu dönüşüm kestirilen parametreleri ortak bir ölçeye yerleştirmeyi sağlayacak doğrusal dönüşüm denklemi ile hesaplanır (Kim ve Cohen, 2002; Hanson ve Beguin, 2002). Yetenek parametresi için dönüşüm denklemi aşağıda verilmiştir (Cook ve Eignor, 1991).

$$\theta^* = A\theta + B$$

$\theta$  : Kişinin yetenek düzeyi

$\theta^*$  : Kişinin girmedığı sınavdan alacağı kuramsal yetenek düzeyi

$A$  : Eşitleme denkleminin eğimi

$B$  : Eşitleme denkleminin sabiti

Doğrusal dönüşüm denklemi aynı şekilde ayırıcılık (a) ve güçlük (b) parametreleri için de şu şekilde hesaplanır (Cook ve Eignor, 1991).

$$a_i^* = \frac{1}{A} a_i$$

$$b_i^* = Ab_i + B$$

$$c_i^* = c_i$$

$a_i$  : Ayırıcılık Parametresi

$a_i^*$  : Dönüştürülmüş ayırıcılık parametresi

$b_i$  : Güçlük parametresi

$b_i^*$  : Dönüştürülmüş güçlük parametresi

$c_i$  : Şans parametresi

$c_i^*$  : Dönüştürülmüş şans parametresi

Görüldüğü gibi şans parametresi ölçek dönüştürmeden bağımsızdır, sadece a ve b parametreleri dönüştürülür (Kolen ve Brennan, 2004).

Ayrı kalibrasyonda elde edilen puanları ortak bir ölçeğe yerleştirmek için çeşitli yöntemler geliştirilmiştir. Bu yöntemlerin amacı eşitleme için gerekli olan katsayıları (A ve B) elde etmektir (Embretson ve Reise, 2000). Literatürde bu yöntemler moment yöntemleri (ortalama-ortalama, ortalama-standart sapma) ve karakteristik eğri dönüştürme yöntemleri (Haebara, Stocking-Lord) olarak üzere iki başlık altında incelenir.

### 2.7.2.1 Moment Yöntemleri

Moment yöntemleri ortalama-ortalama (Lloyd ve Hoover, 1980) ve ortalama-standart sapma yöntemi (Marco, 1977) olmak üzere iki grupta incelenir.

**Ortalama-Ortalama (Mean-Mean) Yöntemi:** Lloyd ve Hoover (1980) tarafından öne sürülen bu yöntem A ve B katsayılarını ortak maddelerin ayırıcılık ve güçlük parametrelerinin ortalamasını kullanarak hesaplar.

Bu yöntem için katsayılar şu şekilde hesaplanır:

$$A = \frac{\mu(\alpha_i)}{\mu(\alpha_j)}$$

$$B = \mu(b_j) - A\mu(b_i)$$

$\mu(\alpha_i)$  : i testindeki ortak maddelerden elde edilen a parametrelerinin ortalaması

$\mu(\alpha_j)$  : j testindeki ortak maddelerden elde edilen a parametrelerinin ortalaması

$\mu(b_j)$  : j testindeki ortak maddelerden elde edilen b parametrelerinin ortalaması

$\mu(b_i)$  : i testindeki ortak maddelerden elde edilen b parametrelerinin ortalaması

**Ortalama-Standart Sapma (Mean-Sigma) Yöntemi:** Marco (1977) tarafından tanımlanan bu yöntem A ve B katsayılarını ortak maddelerin güçlük parametrelerinin ortalama ve standart sapmasını kullanarak hesaplar. Bu yöntem için katsayılar şu şekilde hesaplanır:

$$A = \frac{\sigma(b_j)}{\sigma(b_i)}$$

$$B = \mu(b_j) - A\mu(b_i)$$

$\sigma(b_i)$  : i testindeki ortak maddelerden elde edilen b parametrelerinin standart sapması

$\sigma(b_j)$  : j testindeki ortak maddelerden elde edilen b parametrelerinin standart sapması

$\mu(b_i)$  : i testindeki ortak maddelerden elde edilen b parametrelerinin ortalaması

$\mu(b_j)$  : j testindeki ortak maddelerden elde edilen b parametrelerinin ortalaması

### 2.7.2.2 Karakteristik Eğri Dönüştürme Yöntemleri

Moment yöntemlerinden farklı olarak bu yöntemlerde tüm ortak madde parametreleri hesaplama sürecine dahil edilir (Kaskowitz, 1998). Karakteristik eğri dönüştürme yöntemleri ilk olarak Haebara (1980) tarafından geliştirilmiş ardından Stocking ve Lord (1983) bu yönteme benzeyen farklı bir yöntem geliştirmişlerdir. Bu yöntemlerde öncelikle ölçek dönüşüm sabitleri kestirilir. Ardından test ya da madde karakteristik eğrileri arasındaki fark hesaplanarak bu farklar azaltılmaya çalışılır (Kilmen, 2010).

**Haebara Yöntemi:** Bu yöntem madde karakteristik eğrileri arasındaki farkı, her bir ortak maddenin madde karakteristik eğrileri arasındaki farkın karelerinin toplamını alarak hesaplar (Haebara, 1980). Yöntemin matematiksel ifadesi aşağıda verilmiştir.

$$Hdiff(\theta_i) = \sum_{j \in V} \left[ p_{i_j}(\theta_{j_i}; a_{j_j}, b_{j_j}, c_{j_j}) - p_{i_j}(\theta_{j_i}; \frac{a_{i_j}}{A}, Ab_{i_j} + B, c_{i_j}) \right]^2$$
$$Hcrit = \sum_i Hdiff(\theta_i)$$

$A$  : Eşitleme denkleminin eğimi

$B$  : Eşitleme denkleminin sabiti

$p_{i_j}(\theta_{j_i}; a_{j_j}, b_{j_j}, c_{j_j})$  : Madde karakteristik fonksiyonu

$p_{i_j}(\theta_{j_i}; \frac{a_{i_j}}{A}, Ab_{i_j} + B, c_{i_j})$  : Eşitlenmiş madde karakteristik fonksiyonu

**Stocking-Lord Yöntemi:** Haebara yöntemine benzer olarak bu yöntem ise madde karakteristik eğrileri arasındaki farkı, her bir ortak maddenin madde karakteristik eğrileri arasındaki farkın toplamının karesini alarak hesaplar.

Yöntemin matematiksel ifadesi şu şekildedir:

$$SLdiff = \left[ \sum_{j \in V} p_{i_j}(\theta_{j_i}, a_{j_j}, b_{j_j}, c_{j_j}) - p_{i_j}\left(\theta_{j_i}; \frac{a_{i_j}}{A}, Ab_{i_j} + B, c_{i_j}\right) \right]^2$$

$$SLcrit = \sum_i SLdiff(\theta_i)$$

$A$  : Eşitleme denkleminin eğimi

$B$  : Eşitleme denkleminin sabiti

$p_{i_j}(\theta_{j_i}; a_{j_j}, b_{j_j}, c_{j_j})$  : Madde karakteristik fonksiyonu

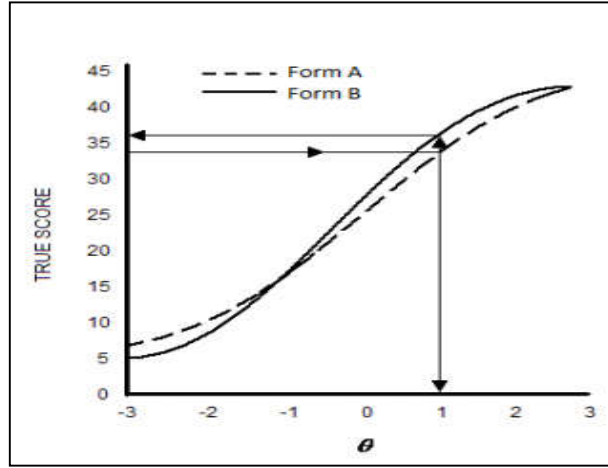
$p_{i_j}\left(\theta_{j_i}; \frac{a_{i_j}}{A}, Ab_{i_j} + B, c_{i_j}\right)$  : Eşitlenmiş madde karakteristik fonksiyonu

## 2.8 Madde Tepki Kuramına Dayalı Eşitleme Yöntemleri

Ölçek dönüştürme işlemi gerçekleştirildikten sonra artık aynı ölçekte yer alan parametreler için eşitleme işlemi uygulanır. Madde Tepki Kuramına dayalı eşitleme yöntemleri “MTK gerçek-puan eşitleme” ve “MTK gözlenen-puan eşitleme” olarak iki başlık altında incelenir.

### 2.8.1 MTK Gerçek-Puan Eşitleme

Gerçek-puan eşitleme yönteminde öncelikle farklı yetenek düzeylerine göre gerçek puanlar kestirilir ve grafikleştirilir (Hambleton ve Swaminathan, 1985). Ardından X formu üzerinde bir gerçek puan seçilir. Daha sonra bu puana karşılık gelen yetenek düzeyi kestirilir. Son aşamada ise bu yetenek düzeyine ilişkin Y formu üzerindeki gerçek puan hesaplanır (Kolen ve Brennan, 2004).



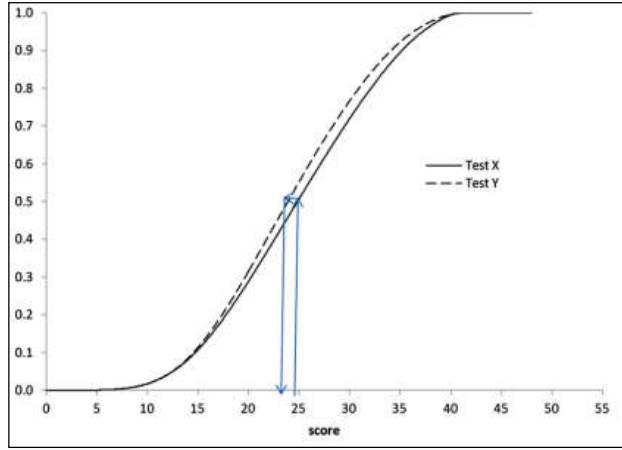
Şekil 2.3 Gerçek-Puan Eşitleme Yönteminin Grafıksel Gösterimi

Örneğin Şekil 1 incelediğinde A formundan 34 puan alan bir kişinin yetenek düzeyinin 1.0 olarak hesaplandığı görülmektedir. B formu üzerinde aynı yetenek düzeyine karşılık gelen puanın 36 olduğunu varsaydıığımızda A formundan alınan 34 puanın B formundan alınan 36 puana eşit olduğunu söyleyebiliriz.

Gerçek-puan eşitlemenin avantajı, her iki grubun yetenek dağılımının kestirilmesine gerek olmamasıdır. Ayrıca bu yöntemde eşitleme sonuçları teorik olarak gruptan bağımsızdır ve tüm alt gruplarda aynı eşitlik ilişkisi elde edilir (Han vd., 1997). Yöntemin dezavantajı ise gerçek test uygulamalarında elde edilen puanların gözlenen puan olmasıdır. Lord ve Wingersky (1984) bu yöntemin gerçek test uygulamalarında kullanımının teorik bir dayanağının olmadığını belirtmişlerdir.

### 2.8.2 MTK Gözlenen-Puan Eşitleme

Bu yöntemde öncelikle madde ve yetenek parametreleri kestirilir. Ardından elde edilen parametrelerle form X ve form Y için teorik gözlenen puan dağılımları oluşturulur. Son aşamada da ise her iki dağılımdan elde edilen puanlar eşit yüzdelli eşitleme yöntemi ile eşitlenir (Kolen ve Brennan, 2004). Bu yöntemde her iki formdan alınan puanlar için yığılmalı frekans dağılımları oluşturulur. Buna göre birinci formdaki A puanı ikinci formdaki A puanı ile aynı yüzdellik sırada yer alan B puanına eşittir (Han vd., 1997).



Şekil 2.4 Gözlenen-Puan Eşitleme Yönteminin Grafiksəl Gösterimi

Örneğin Şekil 2 incelendiğinde form X den 24 puan alan kişinin 50. yüzdalık sırada yer aldığı görülmektedir. Form Y de 50. Yüzdalık sıraya denk gelen puanın 23 puan olduğunu düşündüğümüzde form X'den alınan 24 puanın form Y'den alınan 23 puana eşit olduğunu söyleyebiliriz.

Bu yöntemin avantajı, gerçek test uygulamalarında kullanımının teorik bir dayanağının olmasıdır. Dezavantajı ise elde edilen eşitlik ilişkisinin gruba bağlı olmasıdır. Ayrıca bu yöntem daha karmaşık bir süreç gerektirir ve uygulamada gerçek-puan eşitlemeye göre daha nadir kullanılır (Lord ve Wingersky, 1984).

## 2.9 Eşitleme Hatası

Test eşitleme sürecinde genellikle tesadüfi hata ve sistematik hata olmak üzere iki tür eşitleme hatası görülür. Tesadüfi hata genellikle eşitleme çalışmasının yapıldığı örneklem büyüklüğünün yetersiz olduğu durumlarda oluşur ve kişi sayısı artırılarak hata oranı düşürülebilir (Felan, 2002). Sistematik hata ise eşitleme yanlılığı olarak ifade edilir ve uygun desen ve yöntemin kullanılmadığı durumlarda ya da kullanılan eşitleme yönteminin varsayımlarının karşılanmadığı durumlarda ortaya çıkar. Örneğin denk olmayan gruplar deseninde grupların yetenek dağılımları büyük oranda farklılaşıyorsa sistematik hata meydana gelebilir (Zeng, 1991). Bazen de seçilen desenden kaynaklı olabilir. Örneğin tek grup deseninde yorgunluk ve sıra etkisinden dolayı meydana gelen hata sistematik hatadır (Zhu, 1998).



Tesadüfi hatanın miktarı eşitlemenin standart hatasıyla belirlenebilirken, sistematik hata oranını belirlemek ise oldukça zordur. Eşitleme çalışmalarında her iki hata türünün de olabildiğince düşük olması sağlanmalıdır. Çünkü bazı eşitleme çalışmalarında hata miktarı çok büyük olabilmektedir. Bu gibi durumlarda testleri eşitlemek, eşitleme çalışmasını hiç yapmamaktan daha kötü sonuçlara neden olabilir (Kolen ve Brennan, 2004).

## **2.10 İlgili arařtırmalar**

### **2.10.1 Ölçek Dönüřtürme Yöntemlerinin Karşılaştırıldığı Arařtırmalar**

Stocking ve Lord (1983) tarafından gerçekleştirilen çalışmada yeni bir eşitleme yöntemi önerilmiş ve yaklaşık 2000 kişilik örneklem üzerinde Stocking-Lord ve ortalama standart sapma yöntemlerinden elde edilen sonuçlar karşılaştırılmıştır. 20 test çiftinin kullanıldığı arařtırmada ortak madde sayısı 40-60 arasında, toplam madde sayısı ise 85-125 arasında deęişmektedir. Her bir test çifti için kalibrasyonlar LOGIST programı ile gerçekleştirilmiştir. Çalışma sonucunda Stocking-Lord yönteminin ortalama-standart sapma yönteminden daha doğru kestirimler yaptığı sonucuna ulaşılmıştır.

Kilmen (2010) çalışmasında MTK'ya dayalı ortalama-ortalama, ortalama-standart sapma, Stocking-Lord ve Haebara yöntemlerini simülatif veriler üzerinde karşılařtırmıştır. Ortak maddeli denk olmayan gruplar deseninde yürütölen çalışmada 3PLM'ye uyumlu 600 adet 1-0 verisi üretilmiştir. Arařtırmada elde edilen eşitleme hataları yetenek daęılımı ve örneklem büyüklüğü deęişkenlerine göre incelenmiştir. Çalışmada deęerlendirme kriteri olarak yetenek parametresi deęişkeni için RMSD (en küçük fark kareleri ortalaması) katsayısı hesaplanmıştır. Çalışma sonucunda 500 ve 1000 kişilik benzer ve farklı yetenek daęılımlarında en az hata oranına sahip yöntemin Stocking-Lord yöntemi olduęu tespit edilmiştir.

Gök (2012), gerçekleştirdiği çalışmada 2PLM ve 3PLM'ye uyumlu simülatif veriler üretmek için madde tepki kuramına dayalı ölçek dönüştürme yöntemlerini (ortalama-ortalama, ortalama standart sapma, Stocking Lord, Haebara) örneklem büyüklüğü, yetenek dağılımı, test uzunluğu ve model türü değişkenlerine göre incelemiştir. Çalışmada denk olmayan gruplarda ortak madde deseni kullanılmış, elde edilen sonuçlar ise eşitleme hatası (RMSE) ve eşitleme yanlılığı (BIAS) kriterlerine göre değerlendirilmiştir. Araştırmada gerçek-puan eşitleme yöntemi kullanılmıştır. Çalışma sonucunda en iyi eşitlemenin, benzer yetenek dağılımına sahip 3000 kişilik örneklem, 80 maddelik testler, iki parametrelili lojistik model ve Stocking-Lord yöntemi ile gerçekleştirilebileceği ifade edilmiştir. Elde edilen hata oranları puan düzeyinde hesaplandığında ise 2PLM kullanıldığında en büyük hata ve yanlılığa sahip yöntemin SL olduğu, en az hata ve yanlılığa sahip yöntemin ise pek çok durumda OS olduğu, bazı durumlarda da OO olduğu görülmüştür. 3PLM kullanıldığında ise en az hata ve yanlılığa sahip yöntemin tüm koşullarda OS yöntemi olduğu bulunmuştur.

Uysal (2014), karma modelde yürüttüğü simülasyon çalışmasında madde tepki kuramına dayalı ölçek dönüştürme yöntemlerini karşılaştırmıştır. Araştırmada 40 maddelik formlar ve 1000 kişilik örneklemden elde edilen sonuçlar ortak madde oranı (%10, %20 ve %30) ve yetenek dağılımı (normal dağılım, sağa çarpık ve sola çarpık dağılım) değişkenlerine göre incelenmiştir. İkili puanlanan maddeler için 3PLM, çoklu puanlanan maddeler için ise genelleştirilmiş kısmi kredi modeli (GPCM) kullanılmış, sonuçlar eşitleme hatası (RMSD) ölçütüne göre değerlendirilmiştir. Çalışma sonucunda tüm koşullar incelendiğinde yetenek düzeyine göre sağa çarpık-sağa çarpık gruplarda ve %30 ortak maddenin yer aldığı testlerde en az hatayı Stocking Lord yönteminin verdiği bulunmuştur. Gruplardan birisi sağa çarpık diğeri sola çarpık olduğu durumda ise en az hata oranını ortalama-ortalama yöntemi verirken, bu yöntemi sırasıyla ortalama-standart sapma, Stocking-Lord ve Haebara yöntemleri takip etmiştir.

Skoupski, Jodoin, Swaminathan ve Keller (2003) tarafından yapılan çalışmada ikili puanlanan simülatif veriler üzerinde sabit ortak madde parametrelili eş zamanlı kestirim (FCIP), tekli ve çoklu grup eş zamanlı kestirim, ortalama-standart sapma ve Stocking Lord yönteminin performansı incelenmiştir. Üç parametrelili modelin kullanıldığı çalışmanın değişkenlerini ortak madde sayısı (25-50), test uzunluğu (30 ve 50 madde), örneklem büyüklüğü (500-1000-2000) ve yetenek dağılımı oluşturmaktadır. Çalışma sonucunda farklı koşullarda eş zamanlı kestirim, Stocking-Lord ve ortalama-standart sapma yönteminin benzer sonuçlar verdiği tespit edilmiştir.

Cohen ve Kim (1998) tarafından karma modelde yürütülen çalışmada dereceli tepki modeli kullanılarak, moment (ortalama-ortalama, ortalama standart sapma) ve karakteristik eğri dönüştürme (Stocking-Lord, Haebara) yöntemlerinden elde edilen sonuçlar incelenmiştir. Yatay ve dikey eşitlemenin birlikte gerçekleştirildiği çalışmada sonuçlar yetenek dağılımı, örneklem büyüklüğü ve ortak madde oranına göre değerlendirilmiştir. Araştırma sonucunda, kullanılan eşitleme yöntemlerinin benzer sonuçlar verdiği ve ortak madde sayısı ile örneklem büyüklüğü arttıkça eşitleme hatasının azaldığı sonucuna varılmıştır.

Lee ve Ban (2010) tarafından eşdeğer gruplar deseninde yürütülen çalışmada madde tepki kuramına dayalı üç yöntemden elde edilen sonuçlar incelenmiştir. Araştırmada ayrı kestirim yöntemlerinden Stocking-Lord ve Haebara yöntemleri ile çoklu grup eş zamanlı kestirim yöntemi kullanılmıştır. Elde edilen sonuçlar örneklem büyüklüğü ve madde sayısı değişkenlerine göre değerlendirilmiştir. Çalışma sonucunda genel olarak ayrı kestirimin eş zamanlı kestirimden daha iyi sonuçlar verdiği görülmüştür. Karakteristik eğri yöntemlerinden Haebara yönteminin ise Stocking-Lord yönteminden daha az eşitleme hatasına sahip olduğu bulunmuştur.

Keller (2007) çalışmasında, madde tepki kuramına dayalı gerçek-puan eşitleme ile yeni bir yöntem olan kestirilen gerçek-puan eşitleme yöntemlerini karşılaştırmıştır. Araştırmanın değişkenlerini ortak madde uzunluğu, yetenek dağılımı ve ölçek dönüştürme yöntemi (ortalama-ortalama, ortalama-standart sapma, Stocking-Lord, Haebara) oluşturmaktadır. Çalışma sonucunda 1PLM ve 3PLM'ye uygun verilerde en az hata veren yöntemin ortalama-ortalama yöntemi olduğu tespit edilmiştir. Bunu sırasıyla Stocking-Lord, Haebara ve ortalama-standart sapma yöntemleri takip etmektedir. Ancak çalışmada koşullar değiştikçe en az hata veren yöntemin de değiştiği ve bu nedenle kullanılacak yöntem seçilirken eşitleme çalışması yapılacak test özellikleri ve grubun yetenek dağılımının göz önüne alınması gerektiği belirtilmiştir.

Tate (2000), karma modelde gerçekleştirdiği çalışmasında ortalama-standart sapma ve Stocking-Lord yöntemlerini simülatif veriler üzerinde karşılaştırmıştır. Yapılandırılmış cevaplı ve çoktan seçmeli maddelerin bir arada kullanıldığı çalışmada sonuçlar çeşitli faktörlere (ortak maddelerin uzunluğu, yapılandırılmış cevaplı madde sayısı, yetenek metriği) göre incelenmiştir. Araştırma sonucunda ortalama-standart sapma ve Stocking-Lord yönteminin benzer sonuçlar verdiği bulgusuna ulaşılmıştır. Ayrıca ortak madde sayısı arttıkça test eşitleme hatasının azaldığı ve ikili puanlanan maddelerde de tek boyutluluk varsayımı karşılandığında geçerli eşitleme sonuçlarının elde edildiği görülmüştür.

Keller, Skorupski, Swaminathan ve Jodoin (2004) tarafından karma testler üzerinde yürütülen çalışmada ortalama-standart sapma, Stocking-Lord ve eş zamanlı kestirim ile sabit ortak madde parametrelili eş zamanlı kestirim (FCIP) yöntemlerinden elde edilen hata oranları incelemiştir. Çalışmada model türü (2PLM, 3PLM), yetenek dağılımı ve ortak madde sayısı değişken olarak ele alınmıştır. Simülasyon veriler üzerinden gerçekleştirilen çalışmada en hatalı sonuçları sabit ortak madde parametrelili eş zamanlı kestirimin (FCIP) verdiği görülmüştür. Araştırma sonucunda genel olarak tüm koşullarda ortalama-standart sapma yönteminin daha kararlı sonuçlar verdiği tespit edilmiştir.

Kim ve Lee (2006) tarafından gerçekleştirilen çalışmada, karışık formattaki testlerde ayrı kalibrasyon yöntemlerinden (ortalama-ortalama, ortalama-standart sapma, Stocking-Lord ve Haebara) elde edilen sonuçlar karşılaştırılmıştır. 3PLM ve karma modellerin birlikte kullanıldığı araştırmanın değişkenlerini örneklem büyüklüğü, grupların denkliliği, test türü ve eşitleme yöntemi oluşturmaktadır. Araştırma sonucunda Haebara yönteminin diğer dört yöntemden daha az hata miktarına sahip olduğu görülmüştür.

### **2.10.2 Test Eşitleme Yöntemlerinin Karşılaştırıldığı Araştırmalar**

Han ve diğerleri (1997) tarafından gerçekleştirilen çalışmada gerçek-puan eşitleme, gözlenen-puan eşitleme ve eşit yüzdelli eşitleme yöntemleri gerçek veriler üzerinde karşılaştırılmıştır. Random grup deseninde yürütülen çalışmada ikili puanlanan maddelerin yer aldığı 7 farklı form kullanılmıştır. Her formu yaklaşık 3.000 öğrenci cevaplamış ve uygulamaya toplam 22.848 öğrenci katılmıştır. Çalışma sonucunda gerçek-puan eşitleme yönteminin gözlenen-puan eşitlemeden daha kararlı sonuçlar verdiği bulunmuştur. En hatalı yöntemin ise eşit yüzdelli eşitleme olduğu tespit edilmiş ve formların güçlük seviyelerindeki farklılıklar arttıkça eşitleme sonuçlarının değiştiği sonucuna varılmıştır.

Tsai ve diğerleri (2001) tarafından yürütülen çalışmada gerçek veriler üzerinde madde tepki kuramına dayalı 5 farklı eşitleme yöntemi (ayrı ve eş zamanlı kalibrasyona dayalı gerçek ve gözlenen-puan eşitleme, zincirleme gerçek-puan eşitleme) karşılaştırılmıştır. Araştırmada eşitleme deseni olarak ortak maddeli test deseni kullanılmıştır. Çalışmada her formda ikili puanlanan 38'i ortak olmak üzere toplam 150 madde yer almaktadır. Araştırmada kullanılan birinci formu 1493 kişi, ikinci formu ise 1793 kişi yanıtlamıştır. Ayrıca küçük örneklerde de eşitleme sonuçlarını karşılaştırmak amacıyla 500 kişilik bir örneklem grubu oluşturulmuştur. Çalışma sonucunda beş yöntemin de 0.1 in altında standart hataya sahip olduğu tespit edilmiştir. Genel olarak gözlenen-puan eşitlemenin gerçek-puan eşitlemeden daha az hata verme eğiliminde olduğu bulunmuştur. Ayrıca küçük örneklerde de madde tepki kuramına dayalı eşitlemenin geçerli sonuçlar verdiği görülmüştür.

Wang (2012), denk olmayan gruplarda ortak test deseninde yürüttüğü çalışmasında gerçek-puan ve gözlenen-puan eşitleme yöntemlerini simülatif veride karşılaştırmıştır. Çalışma sonuçları test uzunluğu, test formatı, yetenek dağılımı, farklı güçlük seviyesindeki formlar ve parametre kestirim yöntemi değişkenlerine göre değerlendirilmiştir. Çalışma sonucunda gerçek-puan ve gözlenen-puan eşitleme yöntemlerinin benzer sonuçlar verdiği tespit edilmiştir. Ayrıca formların güçlük düzeyi ve grupların yetenek dağılımları farklılaştıkça eşitleme hatasının da arttığı bulgusuna ulaşılmıştır.

Kolen (1981), çalışmasında farklı modellerin (1PLM, 2PLM, 3PLM) kullanıldığı durumlarda gerçek-puan ve gözlenen-puan eşitleme yöntemlerinden elde edilen sonuçları incelemiştir. Random grup deseninde gerçek test verileri ile yürütülen çalışmada bir ve iki parametrelili lojistik model kullanıldığında genel olarak gözlenen puan eşitleme yöntemi daha iyi sonuçlar verirken, üç parametrelili model kullanıldığında ise her iki yöntemden elde edilen sonuçların farklı gruplarda değişiklik gösterdiği tespit edilmiştir. Çalışma sonucunda elde edilen bulgulara göre eşitleme koşulları değiştikçe en az hata veren yöntemin de değiştiği belirtilmiştir.

Hagge (2010), gerçekleştirdiği çalışmada geleneksel eşitleme yöntemleri (frekans tahmin, zincirleme eşit yüzdellikli) ile madde tepki kuramına dayalı eşitleme yöntemlerini (gerçek-puan, gözlenen-puan) karşılaştırmıştır. Ortak maddeli test deseninde gerçekleştirilen çalışmada çeşitli değişkenlerin (ortak maddelerin niteliği, yetenek dağılımı, test özellikleri) sonuçlar üzerindeki etkisi incelenmiştir. Araştırma sonucunda madde tepki kuramına dayalı yöntemlerin daha iyi sonuç verdiği, gözlenen puan eşitleme yönteminin ise gerçek-puan eşitleme yönteminden daha az hata oranı verdiği bulunmuştur. Ayrıca yetenek dağılımına göre gruplararası farklılık arttıkça yanlılığın da arttığı sonucuna ulaşılmıştır.

Cho (2007), çalışmasında eşit yüzdelikli eşitleme yöntemi ile madde tepki kuramına dayalı eşitleme yöntemlerini (gerçek-puan eşitleme, gözlenen-puan eşitleme) gerçek veriler üzerinde karşılaştırmıştır. Araştırmada farklı örneklem büyüklüklerinden (250, 500, 1000, 1680) elde edilen sonuçlar incelenmiştir. Çoklu puanlanan maddelerin kullanıldığı araştırmada, denk olmayan gruplarda ortak test deseni kullanılmıştır. Çalışma sonucunda genel olarak gözlenen-puan eşitlemenin gerçek-puan eşitlemeden daha düşük standart hataya sahip olduğu görülmüştür. Ancak 250 kişilik örneklem büyüklüğünde ise gerçek-puan eşitleme yönteminin gözlenen-puan eşitlemeden daha düşük hata oranı verdiği tespit edilmiştir.

Chen (2014) çalışmasında, gerçek ve gözlenen-puan eşitleme yöntemlerinden elde edilen sonuçları çeşitli faktörlere (tek boyutlu ve çok boyutlu testler, test uzunluğu, yerel bağımsızlık) göre incelemiştir. Karma modelde (3PLM ve GRM) yürütülen çalışmada gerçek ve simülasyon veriler birlikte kullanılmıştır. Araştırma sonucunda gerçek-puan eşitleme yönteminin yanlılık ve hata oranının daha yüksek olduğu görülmüştür. Ancak gerçek test uygulamalarında hangi yöntemin kullanılacağına dair her iki yöntemin avantaj ve dezavantajları göz önünde bulundurularak eşitlenecek testlerin ve eşitleme çalışması yapılacak grupların özelliklerine göre seçim yapılmasını önermiştir.

Brossman (2010) gerçekleştirdiği çalışmada tek boyutlu ve çok boyutlu madde tepki kuramına dayalı gerçek-puan ve gözlenen-puan eşitleme yöntemlerini karşılaştırmıştır. Araştırmada Iowa Eğitsel Gelişim Testi (ITED)'nin ikili puanlanan maddelerden oluşan A ve B formları eşitlenmiştir. Her bir formu 2500 kişi cevaplamıştır. Parametre kestirimi iki parametrelili lojistik model kullanılarak BILOG-MG programı ile yapılmıştır. Kalibrasyon işlemi TESTFACT programında, test eşitleme ise PIE programı ile gerçekleştirilmiştir. Çalışma sonucunda hem tek boyutlu hem de çok boyutlu modellerde gerçek-puan eşitleme yöntemi ile gözlenen puan eşitleme yönteminin birbirine benzer sonuçlar verdiği görülmüştür.

Lord ve Wingersky (1984), ortak test deseninde gerçekleştirdikleri çalışmada MTK gerçek-puan ve MTK gözlenen-puan eşitleme yöntemlerini karşılaştırmışlardır. 2670 kişilik gerçek veri üzerinde yürütülen çalışmada zincirleme eşitleme yöntemi kullanılarak ortak maddelerin yer aldığı 6 farklı form için kalibrasyon işlemi gerçekleştirilmiştir. Her formda 40 ortak maddenin yer aldığı ortalama 85-90 madde bulunmaktadır. Çalışma sonucunda gerçek-puan ve gözlenen-puan eşitleme yöntemlerinin benzer sonuçlar verdiği bulgusuna ulaşılmıştır.

Özetle ölçek dönüştürme ve test eşitleme yöntemlerinden elde edilen sonuçların eşitleme koşulları değiştikçe farklılık gösterdiği bazı durumlarda ise farklı yöntemlerin benzer sonuçlar verdiği görülmektedir.



## BÖLÜM III

### YÖNTEM

Bu bölümde araştırmanın türü, çalışma grubu, araştırma verileri, veri toplama deseni ve verilerin analizine yer verilmiştir.

#### 3.1 Araştırmanın Türü

Bu çalışmada, PISA 2012 Fen okuryazarlığı testinde madde tepki kuramına dayalı test eşitleme yöntemlerinden elde edilen hata miktarları karşılaştırılmıştır. Araştırmada var olan yöntem ve teknikler gerçek veriler üzerinde sınındığı için çalışma, betimsel araştırma niteliği taşımaktadır.

#### 3.2 Çalışma grubu

PISA 2012 uygulamasına 65 ülkeden yaklaşık 510 bin öğrenci, Türkiye'den ise 12 istatistiki bölgeden seçkisiz yöntemle belirlenen 15 yaş grubu toplam 4848 öğrenci katılmıştır (MEB, 2013). Çalışmada Türkiye örnekleminde kullanılan 13 farklı kitapçıktan en fazla fen bilimleri sorusunun bulunduğu 1 ve 8 numaralı iki kitapçık analize dahil edilmiştir. Bu nedenle araştırmanın çalışma grubunu, 1 numaralı kitapçığı cevaplayan 350 ve 8 numaralı kitapçığı cevaplayan 351 olmak üzere toplam 701 öğrenci oluşturmaktadır.

#### 3.3 Araştırma Verileri

Araştırma verilerini PISA 2012 uygulamasına katılan öğrencilerin fen bilimleri testinden aldıkları puanlar oluşturmaktadır. PISA uygulamasında her formda ortak maddelerin yer aldığı 13 farklı kitapçık kullanılmaktadır. Her kitapçıkta matematik, fen ve okuma alanlarında sorulan soru sayıları değişmektedir. Bu nedenle öncelikle kitapçıklarda yer alan 1-0 şeklinde puanlanmış fen bilimleri maddelerinin sayısı incelenmiştir. Daha sonra en fazla maddenin yer aldığı 1 ve 8 numaralı kitapçıklar

için varsayımların karşılanıp karşılanmadığı test edilmiştir. Analiz sonucunda çalışma için 33 maddenin yer aldığı 1 ve 8 numaralı kitapçıklar seçilmiştir.

Verilerin düzenlenmesi aşamasında ilk olarak kayıp değer içeren bireyler incelenmiştir. PISA uygulamasında öğrenci tarafından ulaşılamayan sorular “8” numara ile kodlanmaktadır. Ancak BILOG-MG programı parametre kestirimi yaparken boş cevapları yanlış (0) olarak kabul etmektedir. Bu nedenle kayıp değer içeren bireylerde hatalı yetenek kestirimine neden olmaması için her iki kitapçıkta da soruların tamamına ulaşamayan öğrenciler analize dahil edilmemiştir. Bu nedenle 1 numaralı kitapçığı cevaplayan 372 kişiden 22 kişi ve 8 numaralı kitapçığı cevaplayan 382 kişiden 31 kişi veri setinden çıkarılmıştır. Ardından tek yönlü uç değerler incelenmiş ve her iki kitapçık için de  $[+3,-3]$  Z puanı aralığında herhangi bir uç değere rastlanmamıştır. Çok yönlü uç değer analizinde ise  $p<.001$  düzeyinde manidar Mahalanobis uzaklığı ölçütü kabul edilmiştir (Çokluk, Şekercioglu ve Büyüköztürk, 2014). Buna göre her bir birey için Mahalanobis uzaklığı hesaplanmış ve her iki kitapçıktan da birer kişi veri setinden çıkarılmıştır. Sonuç olarak analize 1 numaralı kitapçığı cevaplayan 350 ve 8 numaralı kitapçığı cevaplayan 351 kişi ile devam edilmiştir.

### **3.4 Veri Toplama Deseni**

Çalışmada test eşitleme için ‘denk olmayan gruplarda ortak madde deseni’ kullanılmıştır. Desen, literatürde en yaygın kullanılan eşitleme desendir (Kim vd., 2010; Bastari, 2000; Felan, 2002; Kolen ve Brennan, 2014; Meng, 2012; Chulu, 2006; Keller, 2007; Tian, 2011; Norman-Dvorak, 2009; Michaelides, 2003). Bu desende her bir grup ortak maddelerin yer aldığı iki test formundan birini yanıtlar ve ortak maddeler sayesinde her iki formdan elde edilen parametreler aynı ölçeğe yerleştirilir. Araştırmaya ilişkin kullanılan eşitleme deseni Tablo 3.1’de verilmiştir.

*Tablo 3.1. Araştırmaya İlişkin Denk Olmayan Gruplarda Ortak Madde Deseni*

Grup	Kitapçık-1	Ortak	Kitapçık-8
1	33	17	
2		17	33

Tablo 3.1'e göre her iki kitapçıkta da 17'si ortak olmak üzere toplam 33 madde yer almaktadır.

### **3.5 Verilerin Analizi**

Verilerin analizinde ilk olarak her bir kitapçığa ilişkin betimsel istatistikler hesaplanmış ve madde tepki kuramının varsayımları test edilmiştir. İkinci aşamada madde ve yetenek parametreleri kestirilmiştir. Ardından ölçek dönüştürme işlemi gerçekleştirilerek elde edilen parametreler ortak bir ölçeğe yerleştirilmiştir. Son aşamada ise eşitleme işlemi uygulanarak farklı yöntemlerden elde edilen eşitleme hataları hesaplanmıştır.

#### **3.5.1 Betimsel İstatistikler**

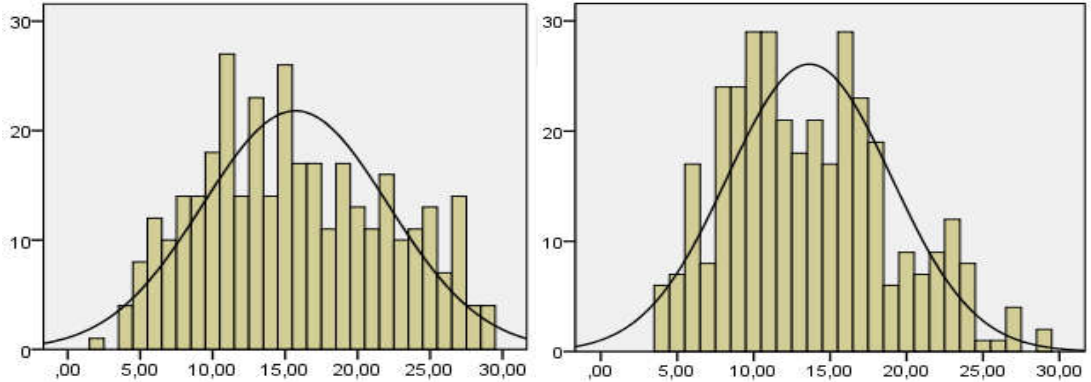
Verilerin analizinde öncelikle betimsel istatistikler hesaplanmış ve Tablo 3.2'de verilmiştir.

*Tablo 3.2 Kitapçıklara İlişkin Betimsel İstatistikler*

Test İstatistikleri	Kitapçık-1	Kitapçık-8
Madde Sayısı	33	33
Kişi Sayısı	350	351
Aritmetik Ortalama	15.71	13.65
Ortanca (Medyan)	15.00	13.00
Tepe Değer (Mod)	11.00	10.00
En Düşük	2.00	4.00
En Yüksek	29.00	29.00
Standart Sapma	6.40	5.37
Varyans	40.99	28.84
Çarpıklık Katsayısı	0.196	0.451
Basıklık Katsayısı	-0.886	-0.357

Tablo 3.2’teki betimsel istatistikler incelendiğinde 1 numaralı kitapçık 350 öğrenciye uygulanmıştır ve 33 sorudan alınan en düşük puan 2, en yüksek puan ise 29’dur. Merkezi eğilim ölçüleri incelendiğinde aritmetik ortalama 15.71, ortanca 15.00, tepe değer ise 11.00’dur. Dağılımın standart sapması ise 6.40 olarak elde edilmiştir. Puan dağılımına ilişkin çarpıklık katsayısı 0.196, basıklık katsayısı ise -0.886 olarak bulunmuştur. 351 öğrenciye uygulanan 8 numaralı kitapçık incelendiğinde ise toplam 33 sorudan alınan en düşük puan 4, en yüksek puan ise 29’dur. Dağılıma ilişkin aritmetik ortalama 13.65, ortanca 13.00 ve tepe değer 10.00 olarak elde edilmiştir. Dağılımın standart sapması 5.37’dir. Çarpıklık katsayısının 0.451, basıklık katsayısının ise -0.357 olduğu görülmektedir.

Genel olarak puan dağılımlarında merkezi eğilim ölçülerinin birbirine yakın olması, çarpıklık ve basıklık katsayılarının  $[-1, +1]$  aralığında olması ve histogramların (Şekil 3.1) da incelenmesi ile puanların normale yakın bir dağılım gösterdiği söylenebilir (Büyüköztürk, Çokluk ve Köklü, 2013).



Şekil 3.1 Sırasıyla Kitapçık-1 Ve Kirapçık-8 İçin Elde Edilen Puan Dağılımları

### 3.5.2 Madde Tepki Kuramı Varsayımlarının Test Edilmesi

#### Tek Boyutluluk

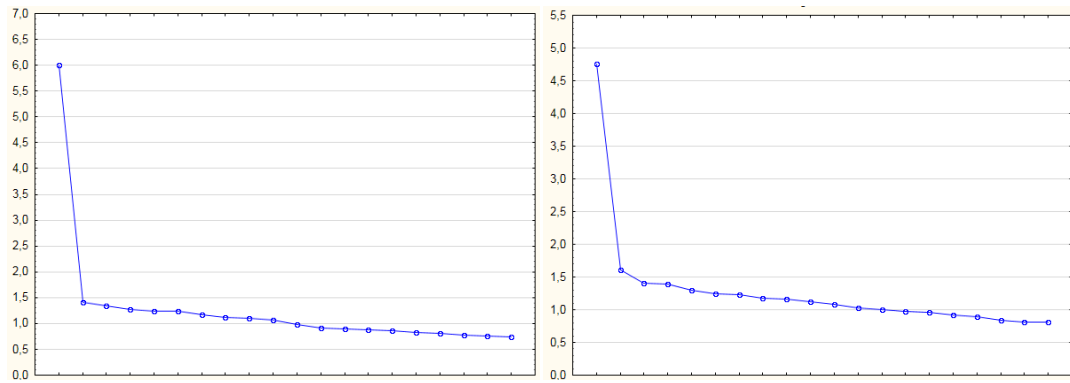
MTK’nın en önemli varsayımı tek boyutluluktur. Bu özellik eşitlenecek iki testin aynı yeteneği, özelliği ya da beceriyi ölçmesini ifade eder (Angoff, 1984; Felan, 2002; Zhu, 1998; Holmes, 1986). Tek boyutluluk varsayımının karşılanıp karşılanmadığı faktör analizi ile test edilebilir. Analiz sonucunda baskın bir faktörün

elde edilmesi tek boyutluluğun göstergesi olarak kabul edilmektedir. (Hambleton ve Swaminathan, 1985; Crocker ve Algina, 1986). Bu çalışmada araştırma verileri iki kategorili olduğu için açımlayıcı faktör analizi, ikili değişkenler arasındaki doğrusal ilişkiyi tanımlayan tetrakorik korelasyon matrisi oluşturularak STATISTICA 13.0 programı ile gerçekleştirilmiştir.

*Tablo 3.3 Kitapçıklara İlişkin Açımlayıcı Faktör Analizi Sonuçları*

Faktör	Özdeğer	Kitapçık-1		Özdeğer	Kitapçık-8	
		Varyans Yüzdesi	Yığılmalı Yüzde		Varyans Yüzdesi	Yığılmalı Yüzde
1	6,010982	18,21510	18,21510	4,756505	13,98972	13,98972
2	1,404071	4,25476	22,46986	1,611895	4,74087	18,73059

Tablo 3.3'e göre 1 numaralı kitapçık için birinci faktörün özdeğeri ikinci faktörün 4 katından fazladır. 8 numaralı kitapçık incelendiğinde ise birinci faktörün özdeğeri ikinci faktörün yaklaşık 3 katına eşittir. Şekil 3.2'deki yamaç-birikinti grafiklerine bakıldığında her iki grafikte de birinci faktörden sonra keskin bir düşüşün olduğu görülmektedir.



Şekil 3.2 Sırasıyla Kitapçık-1 ve Kitapçık-8'e İlişkin Özdeğer Grafikleri

Açımlayıcı faktör analizi ile elde edilen tek boyutlu yapının doğruluğunu sınamak için LISREL (Jöreskog ve Sorbön, 1986) programı ile doğrulayıcı faktör analizi yapılmış ve elde edilen değerler Tablo 3.4'te verilmiştir.

Tablo 3.4. Doğrulayıcı Faktör Analizi Sonucu Elde Edilen Uyum İndeksleri

Uyum İndeksi	Mükemmel Uyum Değerleri	İyi Uyum Değerleri	Model Değerleri	
			Kitapçık-1	Kitapçık-8
$\chi^2/sd$	$\chi^2/sd \leq 3$	$\chi^2/sd \leq 5$	1.10	1.25
RMSEA	$RMSEA \leq 0.05$	$RMSEA < 0.08$	0.02	0.03
GFI	$GFI \geq 0.95$	$GFI \geq 0.90$	0.91	0.90
AGFI	$AGFI \geq 0.95$	$AGFI \geq 0.90$	0.90	0.89
NFI	$NFI \geq 0.95$	$NFI \geq 0.90$	0.89	0.78
NNFI	$NNFI \geq 0.95$	$NNFI \geq 0.90$	0.99	0.94
CFI	$CFI \geq 0.95$	$CFI \geq 0.90$	0.99	0.94
RMR	$RMR \leq 0.05$	$RMR \leq 0.08$	0.01	0.01
SRMR	$SRMR \leq 0.05$	$SRMR \leq 0.08$	0.04	0.05

(Kaynak: Sümer, 2000; Brown, 2006; akt. Çokluk, Şekercioğlu ve Büyüköztürk, 2014)

Tablo 3.4'e göre Ki-Kare/Serbestlik Derecesi ( $\chi^2/sd$ ), oranının 3'ten küçük, Yaklaşık Hataların Ortalama Karakökü (RMSEA) değerlerinin ise 0.05'ten küçük olduğu görülmektedir. Uyum İyiliği İndeksi (GFI), her iki kitapçık için iyi uyuma karşılık gelirken, Düzeltilmiş Uyum İyiliği İndeksi (AGFI), kitapçık-1 için iyi, kitapçık-8 için ise zayıf uyuma işaret etmektedir. Normlaştırılmış Uyum İndeksi (NFI) değerleri incelendiğinde iki kitapçıkta da zayıf uyumun elde edildiği görülmektedir. Normlaştırılmamış Uyum İndeksi (NNFI) ve Karşılaştırmalı Uyum İndeksi (CFI) değerlerine bakıldığında kitapçık-1 için mükemmel, kitapçık-8 için iyi uyumun elde edildiği görülmektedir. Artık Ortalamaların Karakökü (RMR) ve Standardize Edilmiş Artık Ortalamaların Karakökü (SRMR) değerleri incelendiğinde ise 0'a yakın değerler elde edilmiştir.

Sonuç olarak tek boyutlu model için elde edilen uyum indeksleri incelendiğinde modelin veri ile iyi uyum gösterdiği görülmüştür. Bu nedenle tek boyutluluk varsayımının karşılandığı söylenebilir.

### Yerel Bağımsızlık

MTK'nın bir diğer varsayımı ise yerel bağımsızlıktır. Yerel bağımsızlık yetenek değişkeni sabit tutulduğunda bireyin farklı maddelere verdiği cevapların istatistiksel olarak birbirinden bağımsız olmasıdır (Lord, 1980; akt. Hambleton ve Swaminathan, 1985). Yerel bağımsızlık tek boyutlulukla ilişkili bir kavramdır. Lord (1980), yerel bağımsızlığı tek boyutluluğun bir getirisi olarak görmektedir. Genel olarak tek boyutluluğun karşılanması durumunda yerel bağımsızlığın da sağlandığı kabul

edilmektedir (Hambleton ve Swaminathan, 1985). Bunun yanında literatürde yerel bağımsızlık varsayımı farklı yöntemlerle test edilmektedir. Bu yöntemlerden birisi de artık korelasyon matrisinin incelenmesidir (Embretson ve Reise, 2000). Bu çalışmada yerel bağımsızlık varsayımını sınamak için artık korelasyon matrisi oluşturularak ikili çiftlere ilişkin korelasyon katsayıları incelenmiştir. Analiz sonucunda 1 numaralı kitapçığa ilişkin korelasyon katsayıları -0.19 ile 0.18 arasında bulunurken, 8 numaralı kitapçığa ait korelasyon katsayıları ise -0.16 ile 0.15 arasındadır. Analiz sonucunda artık korelasyonların düşük olmasından dolayı yerel bağımsızlık varsayımının da karşılandığı kabul edilmiştir.

### **Model-Veri Uyumu**

Verilerin hangi modele uygun olduğunu belirlemek amacıyla BILOG-MG programı ile her bir kitapçık için 1 ve 2 parametrelili lojistik modele göre veri seti analiz edilmiştir. Kitapçıklarda açık uçlu maddeler de yer aldığı için 3 parametrelili model kullanılmamıştır. Model veri uyumunu değerlendirmede -2loglikelihood uyum istatistiği ile ki-kare istatistiğinden yararlanılmıştır. -2loglikelihood değeri, verinin modelden uzaklaşma derecesini göstermektedir (Embretson ve Reise, 2000). Değerin düşük olması daha iyi model-veri uyumuna işaret etmektedir.

*Tablo 3.5 Kitapçık-1 İçin Model-Veri Uyumunun İncelenmesi*

	1PLM	2PLM
-2Loglikelihood	13423.0164	13245.5923
Parametre Sayısı	33	66

Tablo 3.5'e göre elde edilen değerler incelendiğinde en küçük -2loglikelihood değerinin 2 parametrelili model kullanıldığında elde edildiği görülmektedir. -2loglikelihood değerleri arasındaki fark ( $\chi^2 = -2\loglikelihood_{1PLM} - 2\loglikelihood_{2PLM} = 177.50$ ) ilgili serbestlik derecesi ve 0.05 önem derecesinde ki-kare tablosu ile karşılaştırıldığında verilerin 2 parametrelili modele uyum sağladığı tespit edilmiştir. Maddelere ait ki-kare ve anlamlılık değerleri Tablo 3.6'da verilmiştir.

Tablo 3.6 Kitapçık-1 İçin Maddelere Ait Ki-Kare Ve Anlamlılık Değerleri

Madde	1PLM		2PLM	
	$\chi^2$	p	$\chi^2$	p
O1	10.9	0.2046	2.9	0.9694
O2	4.0	0.8595	6.1	0.6409
O3	15.8	0.0271	8.1	0.3234
O4	11.9	0.1552	5.0	0.7603
O5	28.5	0.0001	5.4	0.4925
O6	8.3	0.4003	5.7	0.6826
O7	7.3	0.5053	7.4	0.3843
O8	39.6	0.0000	4.5	0.6069
O9	7.9	0.4450	4.8	0.7791
O10	6.0	0.4261	1.7	0.9724
O11	9.7	0.1394	4.7	0.4525
O12	10.4	0.2402	8.0	0.4346
O13	1.8	0.9858	0.7	0.9995
O14	7.4	0.4957	10.9	0.2046
O15	19.6	0.0121	12.8	0.1176
O16	24.6	0.0018	24.6	0.0018
O17	8.1	0.3223	9.4	0.3062
A18	14.6	0.0684	1.7	0.9892
A19	3.4	0.8460	1.3	0.9988
A20	8.2	0.4104	3.1	0.8802
A21	7.7	0.4646	3.9	0.8650
A22	5.7	0.6825	4.1	0.8471
A23	6.9	0.5529	4.1	0.8468
A24	6.0	0.6511	5.0	0.6641
A25	7.3	0.3934	11.2	0.1320
A26	5.5	0.7071	5.0	0.7613
A27	15.1	0.0575	10.6	0.1583
A28	10.8	0.1483	14.2	0.0470
A29	2.9	0.9430	2.5	0.9609
A30	26.8	0.0008	8.2	0.4113
A31	28.2	0.0001	11.5	0.0416
A32	3.1	0.8769	6.9	0.5423
A33	31.7	0.0000	12.5	0.0857

\*O = ortak maddeler, A = kitapçık-1'deki ayrı maddeler



Tablo 3.6 incelendiğinde 1 parametrelı lojistik model için 27 maddenin, 2 parametrelı lojistik model için ise 32 maddenin modele uyum gösterdiği görölmektedir ( $p > 0.01$ ). Aynı şekilde 8 numaralı kitapçık için de analiz tekrar edilmiştir.

*Tablo 3.7 Kitapçık-8 İin Model-Veri Uyumunun İncelenmesi*

	1PLM	2PLM
-2Loglikelihood	13570.9494	13382.1854
Parametre Sayısı	33	66

Tablo 3.7'ye göre model-veri uyumunun deęerlendirilmesinde aynı şekilde en küçük -2loglikelihood deęerinin 2 parametrelı modelde elde edildięi görölmektedir. Elde edilen deęerler arasındaki fark ( $\chi^2 = -2\loglikelihood_{1PLM} - 2\loglikelihood_{2PLM} = 200.54$ ) ilgili serbestlik derecesi ve 0.05 önem derecesinde ki-kare tablosu ile karşılaştırıldığında verilerin 2 parametrelı modele uyum gösterdiği görölmüştür. Maddelere ait ki-kare ve anlamlılık deęerleri Tablo 3.8'de verilmiştir.

Tablo 3.8 *Kitapçık-8 İçin Maddelere Ait Ki-Kare Ve Anlamlılık Değerleri*

Madde	1PLM		2PLM	
	$\chi^2$	p	$\chi^2$	p
O1	42.9	0.0000	4.6	0.3275
O2	8.9	0.2602	0.9	0.9962
O3	32.2	0.0001	20.5	0.0046
O4	25.8	0.0011	10.2	0.1787
O5	12.4	0.0894	7.0	0.4325
O6	39.2	0.0000	4.4	0.3567
O7	6.9	0.5426	4.7	0.7852
O8	7.3	0.3974	7.5	0.3768
O9	1.9	0.9657	3.4	0.9058
O10	19.9	0.0057	4.3	0.7455
O11	9.5	0.3021	10.0	0.2643
O12	1.7	0.9753	3.3	0.8550
O13	5.9	0.6632	13.5	0.0959
O14	2.1	0.9552	2.2	0.9746
O15	26.0	0.0005	20.6	0.0082
O16	6.0	0.5410	8.6	0.2843
O17	3.4	0.9038	1.3	0.9961
B18	14.1	0.0490	15.5	0.0505
B19	20.5	0.0045	5.8	0.4508
B20	4.2	0.7523	4.5	0.8113
B21	12.3	0.0924	12.6	0.1266
B22	5.2	0.6314	10.5	0.1642
B23	6.4	0.3826	2.1	0.9082
B24	6.4	0.4959	1.8	0.9704
B25	10.3	0.2455	9.0	0.3406
B26	3.8	0.8000	2.5	0.9603
B27	13.9	0.0844	11.1	0.1942
B28	5.1	0.6485	5.5	0.7049
B29	7.8	0.3549	10.5	0.1635
B30	11.5	0.1747	4.6	0.7955
B31	2.7	0.6167	2.7	0.8474
B32	14.8	0.0380	7.8	0.4523
B33	4.0	0.7848	7.1	0.4195

\*O = ortak maddeler, B = kitapçık-8'deki ayrı maddeler

Tablo 3.8 incelendiğinde 1 parametrelı lojistik model için 26 maddenin, 2 parametrelı lojistik model için ise 31 maddenin modele uyum gösterdiği görülmektedir ( $p > 0.01$ ).

Model-veri uyumunun incelenmesi sonucunda parametre kestirimi için 2 parametrelı lojistik model (2PLM) tercih edilmiştir.

### **Madde ve yetenek parametre deęişmezlięi**

Varsayımlara ek olarak MTK'nın temel ilkeleri de incelenmiştir. Hambleton ve Swaminathan (1985), madde parametre deęişmezlięini incelemek için testin uygulandıęı farklı gruplardan (kızlar-erkekler, üst ve alt yetenek grubundaki öğrenciler vb.) elde edilen madde parametrelerinin karşılaştırılabileceęini belirtmişlerdir. Bu nedenle veri seti cinsiyete göre kız ve erkek olmak üzere ikiye ayrılmış ve parametreler yeniden kestirilmiştir (EK-1).

*Tablo 3.9 Cinsiyete Göre Elde Edilen Madde Parametreleri Arasındaki Korelasyon*

Kitapçık	Madde Ayırıcılık Parametresi (a)	Madde Güçlük Parametresi (b)
1	0.71	0.93
8	0.87	0.93

\*\* $p < 0.01$

Analiz sonucunda kız ve erkeklerden elde edilen ayırıcılık ve güçlük parametreleri arasında pozitif, yüksek ve anlamlı bir ilişki vardır ( $r_{a1} = 0.71$ ,  $r_{b1} = 0.93$ ;  $r_{a8} = 0.87$ ,  $r_{b8} = 0.93$ ;  $p < 0.01$ ).

Yetenek parametre deęişmezlięini incelemek için ise Hambleton ve diğerleri (1991) testte yer alan maddelerin belirli bir ölçüte göre (tek-çift, kolay-zor vb.) iki ya da daha fazla gruba ayrılıp aralarındaki ilişkinin incelenebileceęini ifade etmişlerdir. Bu nedenle her bir kitapçıkta maddeler tek-çift olarak ikiye ayrılmış ve her bir madde seti için yetenek parametreleri yeniden kestirilmiştir (EK-2).

*Tablo 3.10 Madde Setlerinden (Tek-Çift) Elde Edilen Yetenek Parametreleri Arasındaki Korelasyon*

Kitapçık	Öğrenci Sayısı	Yetenek Parametresi
1	350	0.72
8	351	0.69

\*\*p < 0.01

Tablo 3.10'a göre 1 numaralı kitapçıkta her iki madde grubundan elde edilen yetenek parametreleri arasındaki korelasyon 0.72, 8 numaralı kitapçıkta ise 0.69 olarak bulunmuştur (p < 0.01). Elde edilen parametreler pozitif yönde, yüksek korelasyona yakın olduğu için yetenek parametrelerinin elde edildiği madde grubundan bağımsız olduğu söylenebilir.

Analiz sonucunda tek boyutlu madde tepki kuramı varsayımlarının sağlandığı görülmüştür.

### **3.5.3 Madde ve Yetenek Parametrelerinin Kestirilmesi**

Çalışmada varsayımlar test edildikten sonra her iki kitapçık için BILOG-MG (Zimowski, Muraki, Mislevy ve Bock, 2003) programı ile madde ve yetenek parametreleri kestirilmiştir. Parametre kestiriminde farklı yöntemler kullanılmaktadır. Bunlardan bazıları maksimum olabilirlik (Maximum Likelihood Estimates-MLE), maksimum a posteriori (Maximum A Posteriori -MAP) ve beklenen a posteriori (Expected A Posteriori -EAP) yöntemleridir. MLE, yansız bir kestirim yöntemidir, ancak geniş madde havuzu gerektirir ve tüm soruları doğru cevaplayanlarla tüm soruları yanlış cevaplayanlar için yetenek kestirimi yapamamaktadır (Embretson ve Reise, 2000). Ayrıca bu yöntem küçük örneklerde yanlış sonuçlar verebilmektedir (Wang, Jiao ve Xiang, 2013). MAP ve EAP yöntemi ise öncül dağılım bilgisini kullanır. Aralarındaki fark EAP yönteminin öncül dağılımın ortalamasını kullanarak, MAP yönteminin ise tepe değerini kullanarak kestirim yapmasıdır. Ancak MAP yöntemi de kısa testlerde yanlış sonuçlar verebilmektedir (Embretson ve Reise, 2000). Bu nedenle bu çalışmada parametre kestiriminde EAP yöntemi kullanılmıştır (Kim ve Nicewander, 1993).

1 numaralı kitapçıkta kestirilen a ve b parametrelerine ilişkin betimsel istatistikler Tablo 3.11’de verilmiştir.

*Tablo 3.11 Kitapçık-1 İçin Elde Edilen Madde Parametrelerine İlişkin Betimsel İstatistikler*

İstatistikler	Madde Ayırıcılık Parametresi (a)	Madde Güçlük Parametresi (b)
Ortalama	0.997	0.126
Standart Sapma	0.371	0.996
En Küçük	0.406	-2.103
En Büyük	1.996	1.631

Tablo 3.11 incelendiğinde a parametresinin en küçük değeri 0.406, en büyük değeri 1.996, ortalaması ise 0.997 olarak bulunmuştur. Ayırıcılık parametresi teorik olarak  $-\infty$  ile  $+\infty$  arasında değer almaktadır. Ancak uygulamada genellikle 0 ile 2 arasında değişmektedir. Değer 2’ye yaklaştıkça maddenin ayırıcılığı yükselmektedir (Hambleton, Swaminathan ve Rogers, 1991). Elde edilen değerden yola çıkarak genel olarak testin ayırtediciliğinin orta düzeyde olduğu söylenebilir. Güçlük parametresinin (b) ise ortalaması 0.126 olmakla birlikte -2.103 ile 1.631 arasında değerler aldığı görülmektedir. Bu parametre genel olarak [+2-2] arasında değerler almakta, değer -2’ye yaklaşması maddenin kolay, +2’ye yaklaşması ise zor olduğunu göstermektedir (Hambleton vd., 1991). Elde edilen değerlerden yola çıkarak testin orta güçlükte olduğu söylenebilir. Elde edilen madde parametreleri Tablo 3.12’de verilmiştir.

*Tablo 3.12 Kitapçık-1'e Ait Madde Parametre Değerleri*

Madde no	a	b	Madde no	a	b
1	0.547	-2.103	18	1.405	-0.076
2	0.773	-0.683	19	1.003	1.551
3	1.216	-0.068	20	1.177	-0.245
4	1.260	0.248	21	0.628	-0.323
5	1.835	1.097	22	0.852	-0.523
6	0.677	0.992	23	0.897	-1.004
7	0.950	-0.337	24	0.917	-0.729
8	1.996	0.673	25	1.142	0.277
9	0.665	0.785	26	0.745	-1.114
10	1.152	1.036	27	1.233	-0.927
11	1.286	1.631	28	0,842	1.504
12	0.896	-0.920	29	0,816	-0.539
13	0.787	0.549	30	0,406	1.355
14	0.798	1.016	31	1.577	-1.403
15	0.583	0.679	32	0.923	0.487
16	0.633	1.049	33	1.532	-0.085
17	0.740	1.297			

8 numaralı kitapçıkta elde edilen parametrelere ilişkin betimsel istatistikler Tablo 3.13'te verilmiştir.

*Tablo 3.13 Kitapçık-8 İçin Elde Edilen Madde Parametrelerine İlişkin Betimsel İstatistikler*

İstatistikler	Madde Ayırıcılık Parametresi (a)	Madde Güçlük Parametresi (b)
Ortalama	1.832	0.534
Standart Sapma	0.410	1.325
En Küçük	0.328	-2.884
En Büyük	2.033	3.660

Tablo 3.13 incelendiğinde a parametresinin en küçük değeri 0.328, en büyük değeri 2.033, ortalaması ise 1.832 olarak bulunmuştur.

Elde edilen değerlerden yola çıkarak testin ayırtediciliğinin yüksek olduğu görülmektedir. Güçlük parametresi incelendiğinde ise değerler -2.884 ile 3.660 arasında değişmektedir. Parametre ortalamasının 0.534 olarak bulunması testin orta güçlükte olması şeklinde yorumlanabilir. Elde edilen madde parametreleri Tablo 3.14’te verilmiştir.

*Tablo 3.14 Kitapçık-8’e Ait Madde Parametre Değerleri*

Madde no	a	b	Madde no	a	b
1	2.033	0.841	18	0.426	-2.884
2	0.886	1.828	19	1.253	-0.024
3	0.867	-0.964	20	0.601	1.237
4	1.124	-0.681	21	0.477	2.475
5	1.040	0.056	22	0.801	0.999
6	2.206	0.983	23	0.987	1.341
7	0.701	0.133	24	0.743	-1.444
8	0.839	2.051	25	0.462	-0.870
9	0.561	-2.009	26	0.476	0.975
10	1.290	0.478	27	0.478	-0.717
11	0.924	0.149	28	0.598	0.931
12	0.744	0.889	29	0.926	0.831
13	0.618	0.369	30	0.920	0.268
14	0.556	0.991	31	0.838	3.660
15	0.328	2.267	32	0.385	0.105
16	0.769	2.169	33	0.771	0.792
17	0.820	0.406			

#### **3.5.4 Eşitleme Sürecinin Uygulanması**

Parametre kestiriminde elde edilen parametreler farklı ölçeklerde olduğu için doğrusal bir dönüşüm işleminin yapılması gerekmektedir (Kim ve Hanson, 2000). Ölçek dönüştürme olarak adlandırılan bu işlemle parametreler ortak bir metriğe yerleştirilir. Çalışmada her bir kitapçık için parametreler ayrı ayrı kestirildikten sonra STUIRT (Kim ve Kolen, 2004) programı ile dört farklı yöntem (ortalama-ortalama,

ortalama-standart sapma, Stocking-Lord, Haebara) için ölçek dönüştürme işlemi gerçekleştirilmiştir (EK-4). Eşitleme çalışması için iki kitapçıktan birinin yeni form (X) diğerinin eski form (Y) olarak belirlenmesi gerekmektedir. Bu çalışmada 1 numaralı kitapçık “eski form”, 8 numaralı kitapçık ise “yeni form” olarak belirlenerek 8 numaralı kitapçıktan alınan puanlar (yeni form) 1 numaralı kitapçıktan (eski form) alınan puanlara eşitlenmiştir. Analiz POLYEQUATE (Kolen, 2004) programı ile MTK gerçek-puan ve MTK gözlenen-puan eşitleme yöntemleri kullanılarak gerçekleştirilmiştir. Son aşamada ise farklı yöntemlerden elde edilen sonuçlar için hata miktarları hesaplanmıştır.

### 3.5.5 Eşitleme Hatası

Çalışmada farklı yöntemlerden elde edilen hata miktarını belirlemek için aşağıdaki eşitlikten faydalanarak ağırlıklandırılmış hata kareleri ortalaması (WMSE-weighted mean square error) hesaplanmıştır.

$$WMSE = \frac{\sum_{i=1}^{k-1} f_i (X_E - X_{crit})^2}{\sum_{i=1}^k f_i S^2_Y}$$

- $k$  : Y testindeki madde sayısı  
 $S^2$  : Y testindeki ham puanların varyansı  
 $X_{crit}$  : Y testindeki i. ham puan  
 $X_E$  : X testindeki i. ham puana eşit olan puan  
 $f_i$  : Y testindeki i. ham puanın frekansı

Eşitlikte yer alan Y testi eski formu (Kitapçık-1), X testi ise yeni formu (Kitapçık-8) temsil etmektedir. WMSE katsayısının 0'a yaklaşması eşitleme hatasının azaldığı anlamına gelmektedir. Eşitlenmiş puanlar ile ham puanlar arasındaki farklar ne kadar azsa eşitleme hatası da o oranda azalmaktadır.



## BÖLÜM IV

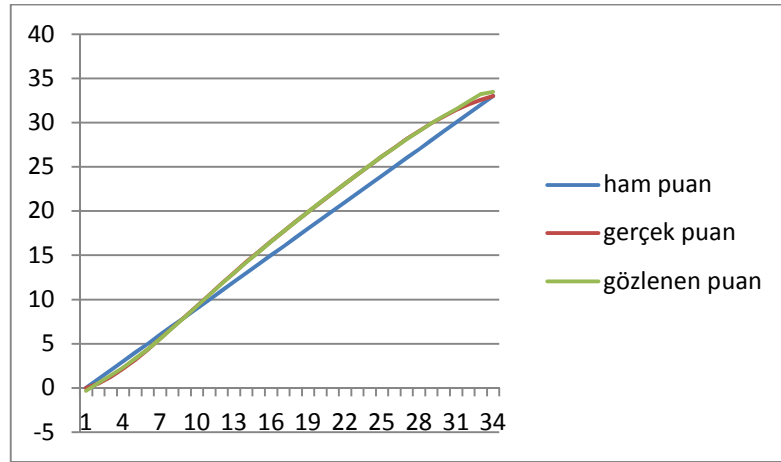
### BULGULAR

Bu bölümde araştırma sonunda elde edilen bulgulara ve yorumlara yer verilmiştir.

#### 4.1 Birinci Alt Probleme İlişkin Bulgular

*PISA 2012 Fen okuryazarlığı testinden seçilen 1 ve 8 numaralı iki kitapçıktan elde edilen puanların MTK gerçek-puan ve MTK gözlenen-puana göre eşitlenmesinde ortalama-ortalama yönteminin kullanılması sonucu elde edilen eşitlenmiş puanlar nasıldır?*

Bu alt probleme çözüm bulmak için öncelikle iki kitapçıktan elde edilen ham puanlar üzerinde ortalama-ortalama yöntemine göre ölçek dönüştürme işlemi gerçekleştirilmiş, ardından MTK gerçek-puan ve gözlenen-puan eşitleme yöntemleri kullanılarak 8 numaralı kitapçıktan elde edilen puanlar 1 numaralı kitapçıktan elde edilen puanlara eşitlenmiştir. Ham puanlar ile eşitlenmiş puanlar arasındaki ilişki Şekil 4.1’de verilmiştir.



Şekil 4.1 Ham Puanlar İle Ortalama-Ortalama Yönteminden Elde Edilen Eşitlenmiş Puanlar Arasındaki İlişki

Şekil 4.1 incelendiğinde, her iki yöntemden elde edilen eşitlenmiş puanların birbiriyle paralellik gösterdiği görülmektedir. Dağılıma bakıldığında düşük puanlarda eşitlenmiş puanlar ham puanlardan küçük, yüksek puanlarda ise eşitlenmiş puanlar ham puanlardan daha büyüktür ve dağılım uç noktalarda birleşmektedir. Ortalama-ortalama yöntemi için ham puanlara karşılık gelen eşitlenmiş puanlar Tablo 4.1’de verilmiştir.

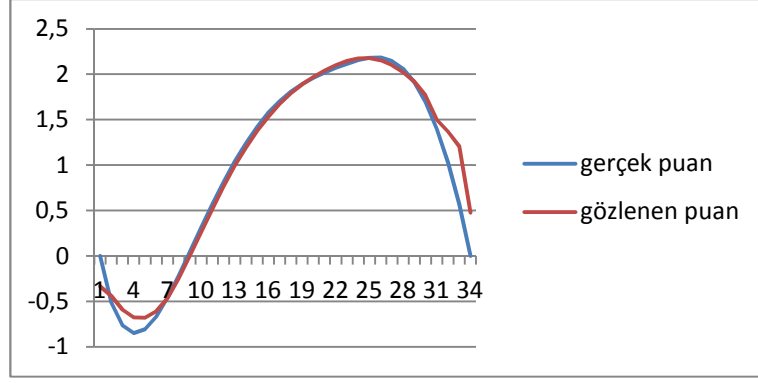
*Tablo 4.1 Ortalama-Ortalama Yöntemi İçin Gerçek-Puan Ve Gözlenen-Puan Eşitlemeden Elde Edilen Eşitlenmiş Puanlar*

MTK Gerçek-Puan Eşitleme			MTK Gözlenen-Puan Eşitleme		
Ham Puan	Eşitlenmiş Puan	Fark	Ham Puan	Eşitlenmiş Puan	Fark
0	0.00	0.00	0	-0.33	-0.33
1	0.49	-0.51	1	0.56	-0.44
2	1.24	-0.76	2	1.41	-0.59
3	2.15	-0.85	3	2.32	-0.68
4	3.19	-0.81	4	3.32	-0.68
5	4.33	-0.67	5	4.39	-0.61
6	5.54	-0.46	6	5.53	-0.47
7	6.78	-0.22	7	6.75	-0.25
8	8.05	0.05	8	8.00	0.00
9	9.31	0.31	9	9.26	0.26
10	10.57	0.57	10	10.51	0.51
11	11.81	0.81	11	11.76	0.76
12	13.04	1.04	12	12.99	0.99
13	14.24	1.24	13	14.20	1.20
14	15.42	1.42	14	15.38	1.38
15	16.58	1.58	15	16.54	1.54
16	17.71	1.71	16	17.68	1.68
17	18.81	1.81	17	18.79	1.79
18	19.89	1.89	18	19.89	1.89
19	20.96	1.96	19	20.97	1.97
20	22.02	2.02	20	22.04	2.04
21	23.07	2.07	21	23.10	2.10
22	24.11	2.11	22	24.15	2.15
23	25.15	2.15	23	25.17	2.17
24	26.18	2.18	24	26.18	2.18
25	27.18	2.18	25	27.15	2.15
26	28.15	2.15	26	28.10	2.10
27	29.06	2.06	27	29.02	2.02
28	29.91	1.91	28	29.92	1.92
29	30.69	1.69	29	30.77	1.77
30	31.40	1.40	30	31.50	1.50
31	32.03	1.03	31	32.37	1.37
32	32.57	0.57	32	33.20	1.20
33	33.00	0.00	33	33.48	0.48

Tablo 4.1’de 8 numaralı kitapçıktan elde edilen ham puanların 1 numaralı kitapçıktaki karşılığı olan eşitlenmiş puanlar görülmektedir. Ham puanların 0 ile 33 arasında değerler aldığı, eşitlenmiş puanların ise gerçek-puan eşitleme yöntemi için 0 ile 33 arasında, gözlenen-puan eşitleme yöntemi için ise -0.33 ile 33.48 arasında değerler aldığı görülmektedir. Ham puanlar ile eşitlenmiş puanlar arasındaki ilişki incelendiğinde gerçek-puan eşitlemede 0 ile 7 arasındaki ham puanların eşitlenmiş puanlardan büyük, 8 ile 32 arasında küçük, 33 puanda ise eşit olduğu görülmektedir. Gözlenen-puan eşitleme yönteminde ise gerçek-puan eşitlemeye benzer olarak 0 ile 7 arasındaki ham puanların eşitlenmiş puanlardan büyük, 8. puanda eşit, 9 ile 32 puan arasında küçük, 33 puanda ise yine eşit olduğu görülmektedir.

Tablo 4.1’de gerçek-puan eşitleme yöntemi kullanıldığında 8 numaralı kitapçıktan alınan 16 ham puanın 1 numaralı kitapçıkta 17.71 puana karşılık geldiği görülmektedir. Buradan yola çıkarak 8 numaralı kitapçığın daha zor gibi görünmektedir. Ancak 8 numaralı kitapçıktan alınan 2 ham puan 1 numaralı kitapçıkta 1.24 puana karşılık geldiği için kitapçıkların güçlük seviyelerinin düşük puanlarda farklılık gösterdiğini söyleyebiliriz.

Eşitleme hatası ham puanlarla eşitlenmiş puanların farkı alınarak hesaplanmaktadır ve farkın azalması hata oranının azaldığının göstergesidir. Tablo 4.1’e göre her iki eşitleme yöntemi için de ham puanlar ile eşitlenmiş puanlar arasındaki farklar 0 ile 2.18 arasında değişmektedir. Gerçek-puan eşitlemede farkların 0-3 puanları arasında arttığı, 4-8 arasında azaldığı, 13-27 puanları arasında artarken, 28-32 arasında ise tekrar azaldığı görülmektedir. Gözlenen-puan eşitleme yönteminde ise puanlar arasındaki farklar 0-4 puanları arasında artarken, 5-8 arasında azalmakta, 9-24 puanları arasında artarken, 25-33 puanları arasında ise tekrar azalmaktadır. Farklar incelendiğinde puanlar arttıkça dalgalanmaların olduğu görülmektedir. Ham puanlar ile eşitlenmiş puanlar arasındaki farklar Şekil 4.2’de verilmiştir.



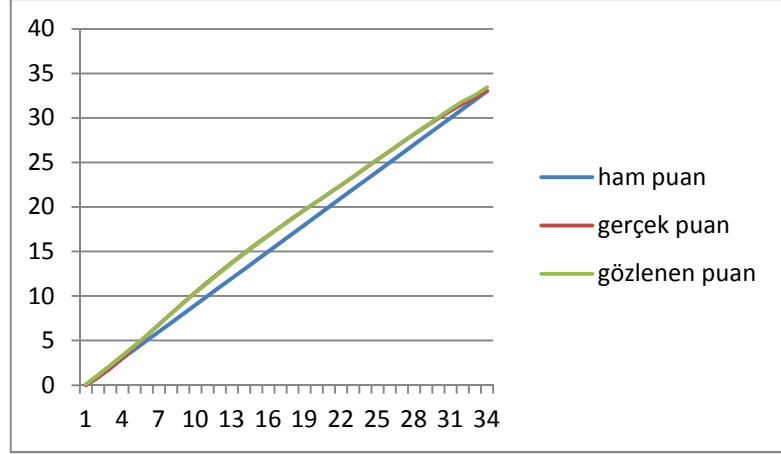
Şekil 4.2 Ortalama-Ortalama Yöntemi İçin Ham Puanlar İle Eşitlenmiş Puanlar Arasındaki Farklar

Şekil 4.2 incelendiğinde düşük ham puanlar için gözlenen-puan eşitleme yönteminin, yüksek ham puanlar için ise gerçek-puan eşitleme yönteminin ham puanlara daha yakın değerler verdiği görülmektedir. Ham puan dağılımının ortalarında ise her iki yöntem benzer sonuçlar vermektedir. Bu nedenle en az hata veren yöntemin düşük ve yüksek puanlarda farklılık gösterdiği söylenebilir.

#### 4.2 İkinci Alt Probleme İlişkin Bulgular

*PISA 2012 Fen okuryazarlığı testinden seçilen 1 ve 8 numaralı iki kitapçıktan elde edilen puanların MTK gerçek-puan ve MTK gözlenen-puana göre eşitlenmesinde ortalama-standart sapma yönteminin kullanılması sonucu elde edilen eşitlenmiş puanlar nasıldır?*

Bu alt problemin çözümünde ortalama-standart sapma yöntemi kullanılarak eşitleme işlemi gerçekleştirilmiştir. Şekil 4.3'te ham puanlar ile eşitlenmiş puanlar arasındaki ilişki verilmiştir.



Şekil 4.3 Ham Puanlar İle Ortalama-Standart Sapma Yönteminden Elde Edilen Eşitlenmiş Puanlar Arasındaki İlişki

Şekil 4.3 incelendiğinde her iki yöntemden (gerçek-puan, gözlenen-puan) elde edilen eşitlenmiş puanların aynı şekilde paralel olduğu görülmektedir. Eşitlenmiş puanlarla ham puanlar arasındaki farklar incelendiğinde farkın, düşük ve yüksek puanlarda azalırken orta puanlarda arttığı görülmektedir. Ayrıca ortalama-ortalama yöntemi ile karşılaştırıldığında eşitlenmiş puanların ham puanlara daha yakın değerler aldığı dikkat çekmektedir. Elde edilen eşitlenmiş puanlar Tablo 4.2’de verilmiştir.

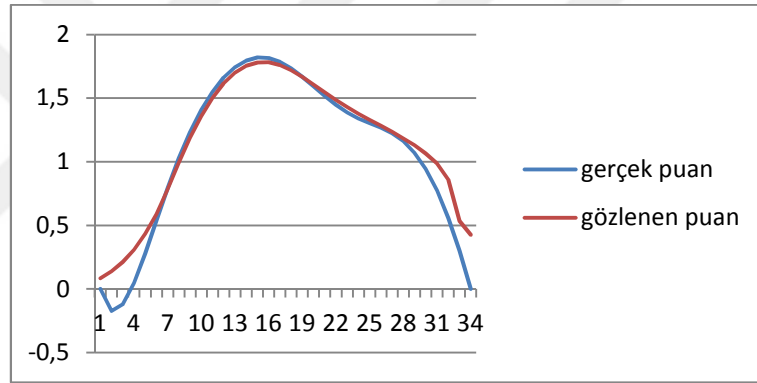
Tablo 4.2. Ortalama-Standart Sapma Yöntemi İçin Gerçek-Puan Ve Gözlenen-Puan Eşitlemeden Elde Edilen Eşitlenmiş Puanlar

MTK Gerçek-Puan Eşitleme			MTK Gözlenen-Puan Eşitleme		
Ham Puan	Eşitlenmiş Puan	Fark	Ham Puan	Eşitlenmiş Puan	Fark
0	0.00	0.00	0	0.08	0.08
1	0.83	-0.17	1	1.14	0.14
2	1.88	-0.12	2	2.21	0.21
3	3.05	0.05	3	3.31	0.31
4	4.28	0.28	4	4.43	0.43
5	5.54	0.54	5	5.59	0.59
6	6.79	0.79	6	6.78	0.78
7	8.03	1.03	7	7.99	0.99
8	9.23	1.23	8	9.19	1.19
9	10.41	1.41	9	10.36	1.36
10	11.55	1.55	10	11.50	1.50
11	12.66	1.66	11	12.62	1.62
12	13.74	1.74	12	13.70	1.70
13	14.80	1.80	13	14.75	1.75
14	15.82	1.82	14	15.78	1.78
15	16.81	1.81	15	16.78	1.78
16	17.78	1.78	16	17.76	1.76
17	18.73	1.73	17	18.72	1.72
18	19.67	1.67	18	19.67	1.67
19	20.59	1.59	19	20.61	1.61
20	21.52	1.52	20	21.55	1.55
21	22.45	1.45	21	22.49	1.49
22	23.39	1.39	22	23.43	1.43
23	24.34	1.34	23	24.38	1.38
24	25.30	1.30	24	25.33	1.33
25	26.27	1.27	25	26.28	1.28
26	27.22	1.22	26	27.24	1.24
27	28.16	1.16	27	28.13	1.13
28	29.07	1.07	28	29.13	1.13
29	29.94	0.94	29	30.07	1.07
30	30.77	0.77	30	30.99	0.99
31	31.56	0.56	31	31.86	0.86
32	32.30	0.30	32	32.53	0.53
33	33.00	0.00	33	33.43	0.43

Tablo 4.2 incelendiğinde eşitlenmiş puanların gerçek-puan eşitleme yöntemi için 0 ile 33 arasında, gözlenen-puan eşitleme yöntemi için ise 0.08 ile 33.43 arasında değerler aldığı görülmektedir. Ham puanlar ile eşitlenmiş puanlar arasındaki ilişki incelendiğinde gerçek-puan eşitlemede 0 ile 3 arasındaki ham puanların eşitlenmiş

puanlardan büyük, 4 ile 32 arasında küçük, 33 puanda ise eşit olduğu görülmektedir. Gözlenen-puan eşitleme yönteminde ise tüm ham puanların eşitlenmiş puanlardan küçük olduğu görülmektedir. Bunun sebebinin formlar arasındaki güçlük farklılıklarından kaynaklandığını söyleyebiliriz.

Tablo 4.2’de ham puanlar ile eşitlenmiş puanlar arasındaki farklar incelendiğinde farkın gerçek-puan eşitleme için 0 ile 1.82, gözlenen-puan eşitleme için ise 0 ile 1.78 arasında değiştiği görülmektedir. Gerçek-puan eşitlemede farkların 1-3 puanları arasında azaldığı, 4-14 arasında attığı ve 15 -33 arasında ise tekrar azaldığı görülmektedir. Gözlenen-puan eşitlemede ise 0-15 arasında artarken, 16-33 ham puan arasında ise azaldığı dikkat çekmektedir. Her iki yöntemden elde edilen puanlar arasındaki farklar Şekil 4.4’te görülmektedir.



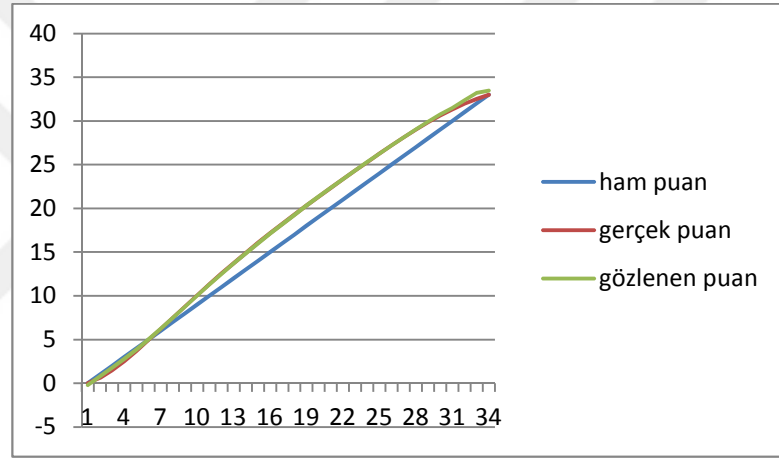
Şekil 4.4 Ortalama-Standart Sapma Yöntemi İçin Ham Puanlar İle Eşitlenmiş Puanlar Arasındaki Fark

Şekil 4.4’e göre her iki yöntemden elde edilen farkların uç noktalarda arttığı, düşük ve yüksek ham puanlarda gerçek-puan eşitleme yönteminin daha iyi sonuç verdiği görülmektedir. Bu bulgu Roos (2002) un gerçekleştirdiği çalışma bulgusu ile tutarlılık göstermektedir.

### 4.3 Üçüncü Alt Probleme İlişkin Bulgular

*PISA 2012 Fen okuryazarlığı testinden seçilen 1 ve 8 numaralı iki kitapçıktan elde edilen puanların MTK gerçek-puan ve MTK gözlenen-puana göre eşitlenmesinde Stocking-Lord yönteminin kullanılması sonucu elde edilen eşitlenmiş puanlar nasıldır?*

Bu alt probleme ilişkin ölçek dönüştürme işlemi Stocking-Lord yöntemi için tekrar edilmiş ve elde edilen puanlar arasındaki ilişki Şekil 4.5'te verilmiştir.



Şekil 4.5 Ham Puanlar İle Stocking-Lord Yönteminden Elde Edilen Eşitlenmiş Puanlar Arasındaki İlişki

Şekil 4.5'e göre elde edilen puan dağılımı moment yöntemleri ile karşılaştırıldığında Stocking-Lord yönteminde eşitlenmiş puanların ham puanlardan daha fazla uzaklaştığı ve aradaki farkların arttığı göze çarpmaktadır. Bu yönteme ilişkin elde edilen puanlar Tablo 4.3'te verilmiştir.



Tablo 4.3 Stocking-Lord Yöntemi İçin Gerçek-Puan Ve Gözlenen-Puan Eşitlemeden Elde Edilen Eşitlenmiş Puanlar

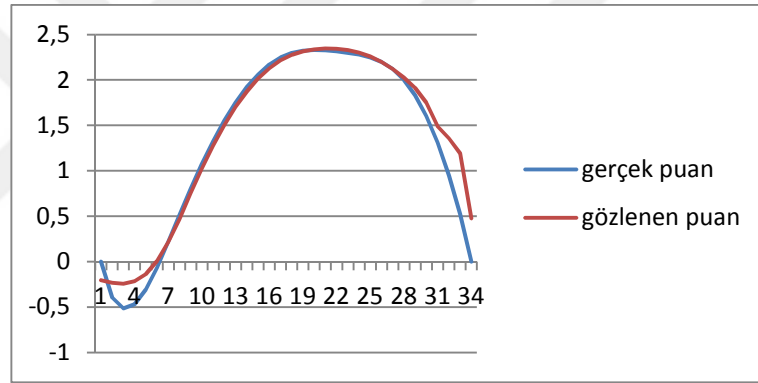
MTK Gerçek-Puan Eşitleme			MTK Gözlenen-Puan Eşitleme		
Ham Puan	Eşitlenmiş Puan	Fark	Ham Puan	Eşitlenmiş Puan	Fark
0	0.00	0.00	0	-0.21	-0.21
1	0.60	-0.40	1	0.77	-0.23
2	1.49	-0.51	2	1.76	-0.24
3	2.53	-0.47	3	2.79	-0.21
4	3.70	-0.30	4	3.86	-0.14
5	4.94	-0.06	5	5.01	0.01
6	6.22	0.22	6	6.22	0.22
7	7.51	0.51	7	7.47	0.47
8	8.80	0.80	8	8.75	0.75
9	10.08	1.08	9	10.02	1.02
10	11.33	1.33	10	11.27	1.27
11	12.55	1.55	11	12.50	1.50
12	13.75	1.75	12	13.70	1.70
13	14.92	1.92	13	14.88	1.88
14	16.06	2.06	14	16.02	2.02
15	17.17	2.17	15	17.13	2.13
16	18.25	2.25	16	18.22	2.22
17	19.30	2.30	17	19.28	2.28
18	20.32	2.32	18	20.32	2.32
19	21.33	2.33	19	21.34	2.33
20	22.32	2.32	20	22.35	2.35
21	23.31	2.31	21	23.34	2.34
22	24.30	2.30	22	24.33	2.33
23	25.28	2.28	23	25.30	2.30
24	26.25	2.25	24	26.26	2.26
25	27.20	2.20	25	27.20	2.20
26	28.12	2.12	26	28.12	2.12
27	29.00	2.00	27	29.03	2.03
28	29.83	1.83	28	29.91	1.91
29	30.60	1.60	29	30.75	1.75
30	31.31	1.31	30	31.49	1.49
31	31.95	0.95	31	32.36	1.36
32	32.52	0.52	32	33.19	1.19
33	33.00	0.00	33	33.48	0.48

Tablo 4.3'e göre eşitlenmiş puanların gerçek-puan eşitleme yöntemi için 0 ile 33 arasında, gözlenen-puan eşitleme yöntemi için ise -0.21 ile 33.48 arasında değerler aldığı görülmektedir. Moment yöntemleri ile benzer olarak gözlenen-puan eşitlemede eşitlenmiş puanların puan ranjının dışına çıktığı dikkat çekmektedir. Ham puanlar ile eşitlenmiş puanlar arasındaki ilişki incelendiğinde gerçek-puan eşitlemede 0 ile 5

arasındaki ham puanların eşitlenmiş puanlardan büyük, 6 ile 32 arasında küçük, 33 puanda ise eşit olduğu görülmektedir.

Gözlenen-puan eşitleme yönteminde ise ham puanların eşitlenmiş puanlardan 0-4 arasında büyük, 5-33 arasında ise küçük olduğu görülmektedir. Puan dağılımlarındaki bu dalgalanmalar moment yöntemleri ile benzerlik göstermektedir.

Tablo 4.3'e göre her iki eşitleme yöntemi için ham puanlar ile eşitlenmiş puanlar arasındaki farklar gerçek-puan eşitlemede 0 ile 2.33 arasında, gözlenen-puanda ise 0 ile 2.35 arasında değişmektedir. Gerçek-puan eşitlemede farkların genel olarak 5-19 puanları arasında arttığı, 20-33 arasında ise azaldığı görülmektedir. Gözlenen-puan eşitleme yönteminde ise aynı şekilde puan farklılıkları 5-20 arasında artarken 21-33 arasında ise azalmaktadır.



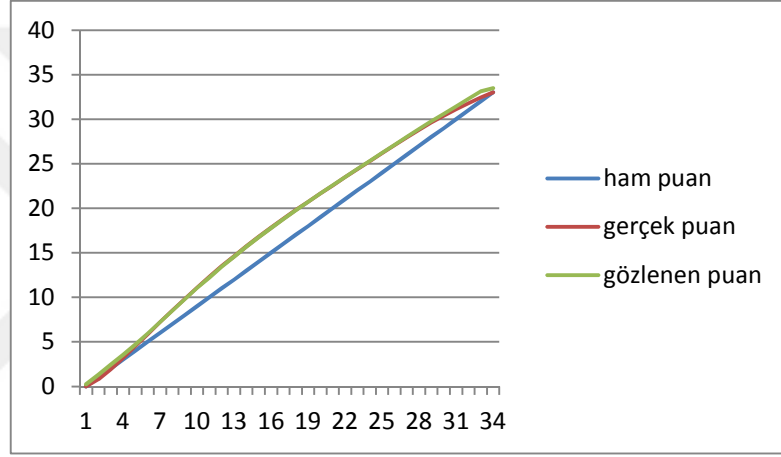
Şekil 4.6 Stocking-Lord Yöntemi İçin Ham Puanlar İle Eşitlenmiş Puanlar Arasındaki Fark

Şekil 4.6'ya göre düşük ham puanlarda her iki yöntemden elde edilen farkların arttığı görülmektedir. Bu bulgu Roos (2002) un çalışma bulgusu ile örtüşmektedir. Düşük ham puanlarda en az hata veren yöntem ortalama-ortalama yönteminde olduğu gibi gözlenen-puan eşitleme olurken, yüksek puanlarda ise gerçek-puan eşitleme olmuştur.

#### 4.4 Dördüncü Alt Probleme İlişkin Bulgular

*PISA 2012 Fen okuryazarlığı testinden seçilen 1 ve 8 numaralı iki kitapçıktan elde edilen puanların MTK gerçek-puan ve MTK gözlenen-puana göre eşitlenmesinde Haebara yönteminin kullanılması sonucu elde edilen eşitlenmiş puanlar nasıldır?*

Bu alt problemin çözümünde ölçek dönüştürme işlemi son olarak Haebara yöntemi için gerçekleştirilmiştir. Puanlar arasındaki ilişki Şekil 4.7’te verilmiştir.



Şekil 4.7 Ham Puanlar İle Haebara Yönteminden Elde Edilen Eşitlenmiş Puanlar Arasındaki İlişki

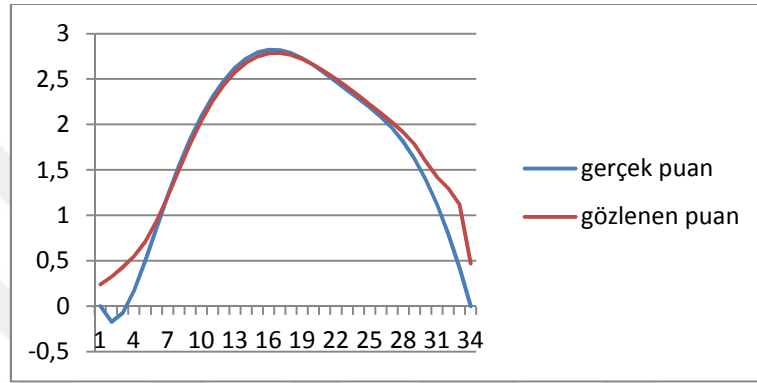
Şekil 4.7 incelendiğinde ham puanlar ile eşitlenmiş puanlar arasındaki farkın Haebara yönteminde diğer yöntemlerden daha fazla olduğu görülmektedir. Bu yöntemin kullanılması sonucu elde edilen puanlar Tablo 4.4’te verilmiştir.

Tablo 4.4 Haebara Yöntemi İçin Gerçek-Puan Ve Gözlenen-Puan Eşitlemeden Elde Edilen Eşitlenmiş Puanlar

MTK Gerçek-Puan Eşitleme			MTK Gözlenen-Puan Eşitleme		
Ham Puan	Eşitlenmiş Puan	Fark	Ham Puan	Eşitlenmiş Puan	Fark
0	0.00	0.00	0	0.24	0.24
1	0.83	-0.17	1	1.33	0.33
2	1.93	-0.07	2	2.43	0.43
3	3.17	0.17	3	3.55	0.55
4	4.50	0.50	4	4.71	0.71
5	5.86	0.86	5	5.93	0.93
6	7.21	1.21	6	7.20	1.20
7	8.54	1.54	7	8.49	1.49
8	9.83	1.83	8	9.78	1.78
9	11.09	2.09	9	11.04	2.04
10	12.30	2.30	10	12.25	2.25
11	13.48	2.48	11	13.43	2.43
12	14.62	2.62	12	14.57	2.57
13	15.73	2.73	13	15.68	2.68
14	16.80	2.80	14	16.75	2.75
15	17.82	2.82	15	17.78	2.78
16	18.82	2.82	16	18.79	2.79
17	19.79	2.79	17	19.77	2.77
18	20.73	2.73	18	20.72	2.72
19	21.65	2.65	19	21.66	2.66
20	22.57	2.57	20	22.58	2.58
21	23.47	2.47	21	23.50	2.50
22	24.38	2.38	22	24.41	2.41
23	25.29	2.29	23	25.32	2.32
24	26.19	2.19	24	26.22	2.22
25	27.08	2.08	25	27.12	2.12
26	27.96	1.96	26	28.02	2.02
27	28.81	1.81	27	28.91	1.91
28	29.62	1.62	28	29.78	1.78
29	30.39	1.39	29	30.60	1.60
30	31.12	1.12	30	31.42	1.42
31	31.79	0.79	31	32.30	1.30
32	32.42	0.42	32	33.12	1.12
33	33.00	0.00	33	33.47	0.47

Tablo 4.4'e göre eşitlenmiş puanların gerçek-puan eşitleme yöntemi için 0 ile 33 arasında, gözlenen-puan eşitleme yöntemi için ise 0.24 ile 33.47 arasında değerler aldığı görülmektedir. Ham puanlar ile eşitlenmiş puanlar arasındaki ilişki incelendiğinde diğer yöntemlerle benzer olarak gerçek-puan eşitlemede 0 ile 2 arasındaki ham puanların eşitlenmiş puanlardan büyük, 3 ile 32 arasında küçük, 33 puanda ise eşit olduğu görülmektedir.

Gözlenen-puan eşitleme yönteminde ise tüm ham puanların eşitlenmiş puanlardan küçük olduğu görülmektedir. Her iki puan türü arasındaki farklar incelendiğinde elde edilen farklar gerçek-puan eşitlemede 0 ile 2.82 arasında, gözlenen-puan eşitlemede ise 0 ile 2.79 arasında değişmektedir. Gerçek-puan eşitlemede farkların 3-16 puanları arasında arttığı, 17-33 arasında ise azaldığı göze çarpmaktadır. Gözlenen-puan eşitleme yönteminde ise puanlar arasındaki farklar genel olarak 0-16 puanları arasında artarken 17-33 puanları arasında ise azalmaktadır.



Şekil 4.8 Haebara Yöntemi İçin Ham Puanlar İle Eşitlenmiş Puanlar Arasındaki Fark

Şekil 4.8 incelendiğinde ortalama-standart sapma yöntemi ile benzer olarak düşük ve yüksek puanlarda gerçek-puan eşitleme yöntemi daha az hata miktarına sahiptir. Ancak elde edilen farkların diğer yöntemlerden yüksek olduğu dikkat çekmektedir.

Genel olarak elde edilen bulgulara bakıldığında gerçek-puan ve gözlenen-puan eşitleme yöntemlerinin genellikle birbiriyle paralellik göstermekle birlikte düşük ve yüksek ham puanlarda en az hata veren yöntemin değiştiği tespit edilmiştir. Ayrıca gözlenen-puan eşitleme yöntemi kullanıldığında elde edilen eşitlenmiş puanların genel olarak puan ranjının dışına çıktığı görülmüştür. Ölçek dönüştürme yöntemleri incelendiğinde, yöntemlerin tamamında eşitlenmiş puanların ham puanlardan genel olarak daha yüksek olduğu görülmüştür. Bu bulgu 8 numaralı kitapçıkta yer alan soruların 1 numaralı kitapçıkta sorulardan daha zor olduğu şeklinde yorumlanabilir.

#### 4.5 Beşinci Alt Probleme İlişkin Bulgular

*PISA 2012 Fen okuryazarlığı testinden seçilen 1 ve 8 numaralı iki kitapçıktan elde edilen puanların MTK gerçek-puan ve MTK gözlenen-puana göre eşitlenmesinde en az eşitleme hatasına sahip ölçek dönüştürme ve test eşitleme yöntemi hangisidir?*

Bu alt probleme ilişkin farklı yöntemlerden elde edilen puanlar için WMSE katsayısı (Ağırlıklandırılmış Hata Kareleri Ortalaması) hesaplanmış ve Tablo 4.5'te verilmiştir.

*Tablo 4.5 Farklı Yöntemlerden Elde Edilen WMSE Katsayıları*

	MTK Gerçek-Puan	MTK Gözlenen-Puan
Ortalama-Ortalama	0.058	0.057
Ortalama-Standart Sapma	<b>0.055</b>	<b>0.054</b>
Stocking-Lord	0.088	0.087
Haebara	0.137	0.136

Tablo 4.5'e göre gerçek-puan eşitleme yöntemi için elde edilen WMSE katsayıları ortalama-ortalama yönteminde 0.058, ortalama-standart sapma yönteminde 0.055, Stocking-Lord yönteminde 0.088, Haebara yönteminde ise 0.137 olarak bulunmuştur. Gözlenen-puan eşitleme yöntemi kullanıldığında ise ortalama-ortalama yönteminde hata miktarının 0.057, ortalama-standart sapmada 0.054, Stocking-Lord yönteminde 0.087, Haebara yönteminde ise 0.136 olarak elde edildiği görülmektedir.

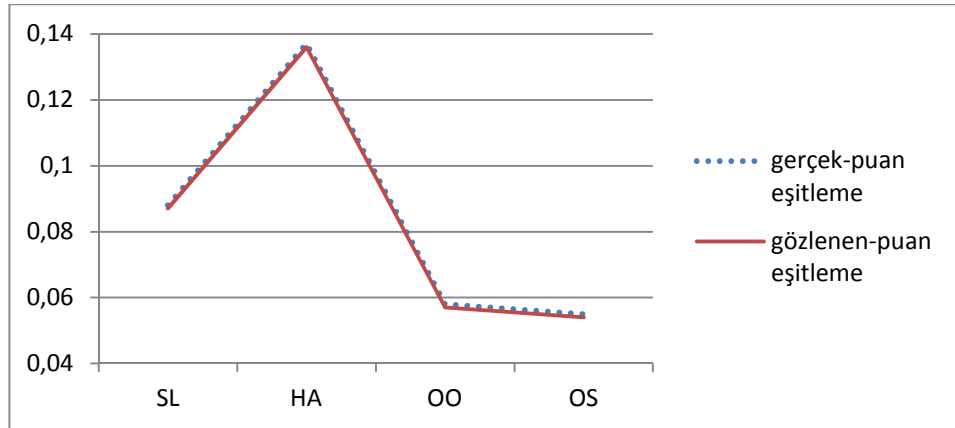
Tablo 4.5 incelendiğinde en az hata miktarına sahip yöntemin her iki eşitleme metodu için de ortalama-standart sapma olduğu görülmektedir. Bu yöntemi sırasıyla ortalama-ortalama, Stocking-Lord ve Haebara yöntemleri takip etmektedir. Elde edilen bu bulgu literatürde yapılan benzer çalışmalarla (Keller vd., 2004; Gök, 2012) tutarlılık göstermektedir.

Çalışmada son olarak farklı yöntemlerden elde edilen sonuçların önemli olup olmadığına standartlaştırılmış DTM (Standardized Difference That Matters-SDTM) değeri ile karşılaştırılarak karar verilmiştir (Öztürk-Gübeş ve Kelecioğlu, 2017).

Dorans ve Feigenbaum (1994) tarafından önerilen DTM (Difference That Matters) katsayısı, eşitleme sonuçlarını değerlendiren bir ölçüttür ve sonuçları yorumlamak amacıyla kullanılır (Dorans ve Cook, 2016; Kolen ve Brennan, 2014; Shin, 2015). DTM değeri eski formun standart sapmasına bölünerek standartlaştırılmış DTM değeri (SDTM) elde edilir (Kolen ve Brennan, 2014; Kim, 2006). Eşitleme sonucunda elde edilen değerler arasındaki farkın SDTM değerinden büyük olması farkın önemli olduğu şeklinde yorumlanmaktadır (Dorans ve Cook, 2016).

Bu çalışmada “0.5” DTM ölçütü kullanılmıştır. SDTM değeri ise DTM değerinin (0.5) eski formun (Kitapçık-1) standart sapmasına bölünmesiyle (0.5/6.40) “0.08” olarak elde edilmiştir. Tablo 4.5 incelendiğinde hem gerçek-puan hem de gözlenen puan eşitleme yöntemlerinden elde edilen değerler arasındaki farkın ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemi için 0.08’den küçük, Haebara yöntemi için ise 0.08’den büyük olduğu görülmektedir. Bu nedenle bu çalışmada Haebara yöntemi hariç, ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri de tercih edilebilir. Ancak en küçük WMSE değerine sahip olduğu için ortalama-standart sapma yönteminin kullanılması önerilmektedir.

Farklı yöntemlerinden elde edilen hata miktarları Şekil 4.9’da verilmiştir.



Şekil 4.9: Farklı Yöntemlerden Elde Edilen Hata Miktarları (WMSE)

Şekil 4.9'a göre MTK gerçek-puan ve MTK gözlenen-puan eşitleme yöntemlerinden elde edilen sonuçların birbirine yakın olduğu görülmektedir. Elde edilen sonuçlar SDTM değeri (0.08) ile karşılaştırıldığında kitapçıkların eşitlenmesi için her iki yöntemin de kullanılabilmesi görülmektedir. Fakat karşılaştırılan yöntemlerin tamamında daha az hata miktarına sahip olduğu ve gerçek verilerin eşitlenmesinde teorik bir dayanağı olduğu için gözlenen-puan eşitleme yönteminin kullanılması önerilmektedir.





## BÖLÜM V

### SONUÇ, TARTIŞMA VE ÖNERİLER

Bu bölümde bulgular doğrultusunda elde edilen sonuçlara ve önerilere yer verilmiştir.

#### 5.1 Sonuç ve Tartışma

Bu çalışmada madde tepki kuramına dayalı ölçek dönüştürme (ortalama-ortalama, ortalama-standart sapma, Stocking-Lord, Haebara) ve test eşitleme yöntemlerini (MTK gerçek-puan eşitleme, MTK gözlenen-puan eşitleme) gerçek veriler üzerinde karşılaştırmak ve en az hata veren yöntemi tespit etmek amaçlanmıştır. Bu amaçla PISA 2012 fen okuryazarlığı testinden seçilen 1 ve 8 numaralı iki kitapçık farklı yöntemler kullanılarak eşitlenmiş ve elde edilen hata oranları puan düzeyinde hesaplanan WMSE katsayısı ile incelenmiştir.

Çalışma sonucunda farklı ölçek dönüştürme yöntemlerinden elde edilen hata miktarları incelendiğinde en az hata oranına sahip yöntemin moment yöntemlerinden ortalama-standart sapma yöntemi olduğu bulunmuştur. Bunu sırasıyla ortalama-ortalama, Stocking-Lord ve Haebara yöntemleri takip etmektedir. Gök (2012), gerçek-puan eşitleme yöntemini kullanarak gerçekleştirdiği simülasyon çalışmasında iki parametrelili modele uygun verilerde farklı yöntemlerin puan düzeyinde verdiği hata miktarını incelemiştir. Çalışma sonucunda en az hata oranına sahip yöntemin ortalama-standart sapma yöntemi olduğunu bulmuş ve küçük örneklem, kısa testlerin kullanıldığı gerçek test uygulamalarında ortalama-standart sapma yönteminin kullanılmasını önermiştir. Keller ve diğerleri (2004) tarafından gerçekleştirilen çalışmada ise tüm koşullarda en az hata veren yöntemin ortalama-standart sapma olduğu tespit edilmiştir. Elde edilen bu bulgular çalışma sonucuyla tutarlılık göstermektedir. Bunun yanında Stocking ve Lord (1983) gerçekleştirdikleri çalışmada Stocking-Lord yönteminin ortalama-standart sapma yönteminden daha az hata miktarı verdiğini belirtmişlerdir. Bunun sebebi örneklem büyüklüğü ve kullanılan madde sayısı ile açıklanabilir.

Kilmen (2010), ise gerçekleştirdiği simülasyon çalışmasında tüm koşullarda Stocking-Lord yönteminin ortalama-standart sapma yönteminden daha kararlı sonuçlar verdiğini tespit etmiştir. Bunun nedeni kullanılan değerlendirme ölçütünün (yetenek parametresi) farklılığından kaynaklanıyor olabilir. Bunun yanında gerçekleştirilen bazı çalışmalarda (Tate, 2000; Skoupski vd., 2003) ise her iki yöntemin benzer sonuçlar verdiği bulgusuna ulaşılmıştır.

Eşitleme çalışması yapılan grubun yetenek dağılımının farklılaşması da sonuçları değiştirebilmektedir. Uysal (2014)'ın çalışmasında, her iki grubun yetenek dağılımının sağa çarpık olduğu durumlarda karakteristik eğri dönüştürme yöntemleri (Stocking-Lord, Haebara) daha iyi sonuç verirken, gruplardan birinin sağa, diğerinin sola çarpık olduğu durumda moment yöntemleri (ortalama-ortalama, ortalama-standart sapma) daha az hata miktarı vermiştir. Gök (2012) ün gerçekleştirdiği çalışmada da benzer yetenek dağılımlarında karakteristik eğri dönüştürme yöntemleri daha iyi sonuç verirken, yetenek farklılıklarının fazla olduğu pozitif ve negatif çarpık dağılımlarda moment yöntemlerinin daha az hata miktarı verdiği görülmüştür.

Çalışmada moment yöntemlerinde elde edilen sonuçlar incelendiğinde ortalama-standart sapma yönteminin, yakın olmakla birlikte ortalama-ortalama yönteminden daha az hata miktarına sahip olduğu görülmüştür. Bu bulgu Kim ve Kolen (2006) in gerçekleştirdiği çalışma bulgusu ile örtüşmektedir. Bu sonuca ilişkin Kolen ve Brennan (2004) ortalama-standart sapma yönteminin ortalama-ortalama yönteminden bazen daha tercih edilebilir olduğunu, çünkü güçlük parametresinin (b) ayırıcılık parametresinden (a) daha kararlı kestirimler sağladığını ifade etmişlerdir.

Karakteristik eğri dönüştürme yöntemleri karşılaştırıldığında ise Stocking-Lord yönteminin Haebara yönteminden daha düşük hata miktarı verdiği tespit edilmiştir. Keller (2007), gerçek-puan eşitleme yöntemini kullandığı simülasyon çalışmasında Stocking-Lord yönteminin Haebara yönteminden daha kararlı sonuçlar verdiğini bulmuştur. Bu bulgu Uysal (2014) ve Kilmen (2010)'ın çalışma sonuçlarıyla da tutarlılık göstermektedir. Ancak Lee ve Ban (2010)'ın eşdeğer gruplar deseninde gerçekleştirdikleri çalışmada Haebara yönteminin Stocking-Lord yönteminden daha iyi sonuç verdiğini belirtmişlerdir. Kim ve Lee (2006) ise karışık formattaki testlerde farklı ölçek dönüştürme yöntemlerini karşılaştırmış ve Haebara yönteminin

diğer yöntemlerden daha az hata oranı verdiği sonucuna ulaşmışlardır. Bu durumun sebebi seçilen eşitleme deseni ve test türünün farklılaşması ile açıklanabilir. Kim ve Kolen (2007) ile Way ve Tang (1991) tarafından yapılan çalışmalarda ise her iki yöntemin benzer sonuçlar verdiği görülmüştür.

Elde edilen sonuçlar MTK gerçek ve MTK gözlenen-puan eşitleme yöntemlerine göre incelendiğinde ise genel olarak her iki yöntemden elde edilen sonuçların benzer olduğu, ancak gözlenen-puan eşitleme yönteminin daha az hata miktarı verdiği tespit edilmiştir. Tsai ve diğerleri (2001) tarafından gerçek veriler üzerinde yapılan çalışmada gözlenen-puan eşitlemenin daha az hata verme eğiliminde olduğu görülmüştür. Chen (2014), karma modelde gerçekleştirdiği çalışmasında da gözlenen-puan eşitleme yönteminin daha iyi sonuç verdiğini belirtmiştir. Bu bulgu Hagge (2010)'nin gerçekleştirdiği çalışma sonucu ile de örtüşmektedir. Ancak Han ve diğerleri (1997) tarafından random grup deseninde gerçekleştirilen çalışmada, gerçek-puan eşitleme yönteminin daha kararlı sonuçlar verdiği görülmüştür. Bu durumun sebebi kullanılan eşitleme deseninin farklılaşması ile açıklanabilir.

Cho (2007) gerçek veriler üzerinde yürüttüğü çalışmasında büyük örneklerde gözlenen-puan eşitlemenin daha az hata oranı verdiğini bulurken, küçük örneklerde ise gerçek-puan eşitlemenin daha iyi sonuç verdiğini tespit etmiştir. Aynı şekilde Kolen (1981) tarafından gerçek veriler üzerinde yapılan çalışmada bir ve iki parametrelili model kullanıldığında gözlenen-puan eşitlemenin, üç parametrelili model kullanıldığında ise en az hata veren yöntemin farklı gruplarda değişiklik gösterdiği bulunmuştur. Bazı çalışmalarda (Lord ve Wingersky, 1984; Brossman, 2010; Wang 2012) ise her iki yöntemin benzer sonuçlar verdiği görülmektedir. Chen (2014) bu konuya ilişkin gerçek test uygulamalarında yöntem seçiminde, eşitlenecek testlerin ve eşitleme çalışması yapılacak grupların özelliklerine göre karar verilmesini önermiştir.

## 5.2 Öneriler

Araştırmadan elde edilen bulgulara dayanılarak gelecekte yapılacak çalışmalar için bir takım öneriler verilmiştir.

### 5.2.1 Uygulayıcılara Yönelik Öneriler

Bu araştırmanın sonucunda en az hata veren yöntemin gözlenen-puan eşitleme yöntemi kullanıldığında ortalama-standart sapma olduğu bulunmuştur. Bu nedenle küçük örneklemelerin ve kısa testlerin kullanıldığı gerçek test uygulamaları için bu yöntemlerin kullanılması önerilebilir. Bunun yanında farklı yöntemlerden elde edilen sonuçlar SDTM değeri ile karşılaştırıldığında ortalama-ortalama, ortalama-standart sapma ve Stocking-Lord yöntemleri de tercih edilebilir. Ancak en küçük WMSE değerine sahip olduğu için ortalama-standart sapma yönteminin kullanılması önerilmektedir. Çalışmada MTK gerçek-puan ve MTK gözlenen-puan eşitleme yöntemlerinden birbirine yakın sonuçlar elde edilmiştir. Fakat karşılaştırılan yöntemlerin tamamında daha az hata miktarına sahip olduğu ve gerçek verilerin eşitlenmesinde teorik bir dayanağı olduğu için gözlenen-puan eşitleme yönteminin kullanılması önerilmektedir.

### 5.2.2 Araştırmacılara Yönelik Öneriler

Gelecekte yapılacak çalışmalar için aşağıdaki öneriler dikkate alınabilir:

1. Bu çalışmada madde tepki kuramına dayalı ayrı kalibrasyon yöntemleri karşılaştırılmıştır. İleride yapılacak çalışmalarla eş zamanlı ve ayrı kalibrasyondan elde edilen sonuçlar karşılaştırılabilir.
2. Araştırma gerçek veriler üzerinde yürütülmüştür. Gerçek ve simülasyon veriden elde edilen sonuçlar incelenebilir.
3. Çalışma sadece madde tepki kuramına dayalı yöntemler kullanılarak gerçekleştirilmiştir. Klasik test kuramına dayalı eşitleme çalışmaları da yapılarak her iki kuramdan elde edilen hata oranları karşılaştırılabilir.
4. Araştırmada denk olmayan gruplarda ortak madde deseni kullanılmıştır. Farklı test eşitleme desenlerinin kullanıldığı durumlarda elde edilen sonuçların değişip değişmeyeceği incelenebilir.

5. Bu çalışmada değerlendirme ölçütü olarak WMSE katsayısı kullanılmış olup ileride yapılacak çalışmalarda farklı ölçütler kullanılabilir.
6. Araştırma sadece ikili puanlanan veriler üzerinde yürütülmüştür. Karma modeller kullanılarak elde edilen sonuçlar incelenebilir.
7. Bu çalışmada eşitleme yöntemleri kısa testler ve küçük örneklemeler üzerinde karşılaştırılmıştır. Yapılacak çalışmalarla uzun testler ve büyük örneklemeler kullanıldığında elde edilen sonuçlar incelenebilir.



## KAYNAKÇA

- Angoff, W. H. (1984). *Scales, norms and equivalent scores*. Princeton, New Jersey: Educational Testing Service.
- Baker, F. B. (2001). *The basics of item response theory*. USA: Clearinghouse on Assessment and Evaluation.
- Bastari, B. (2000). *Linking multiple-choice and constructed-response items to a common proficiency scale*. Doktora Tezi, University of Massachusetts, Massachusetts, USA.
- Bozdağ, S. (2007). *Şans başarısının test eşitlemeye etkisi*. Yüksek Lisans Tezi. Mersin Üniversitesi, Sosyal Bilimler Enstitüsü, Mersin.
- Brossman, B. G. (2010). *Observed score and true score equating procedures for multidimensional item response theory*. Doktora Tezi, University of Iowa, Iowa.
- Büyüköztürk, Ş., Çokluk, Ö. ve Köklü, N. (2013). *Sosyal bilimler için istatistik*. (12. baskı). Ankara: Pegem Akademi.
- Caldwell, L. J. (1984). *A comparison of equating error in linear and rasch model test equating method*. Doktora Tezi, Florida State University, Florida.
- Chen, H. W. (2001). *Calibration of the ITBS test battery to the complete test battery: a comparison five linking methods*. Doktora Tezi, University of Iowa, Iowa.
- Chen, J. (2014). *Model selection for IRT equating of testlet-based tests in the random group design*. Doktora Tezi, University of Iowa, Iowa.

- Cho, Y. (2007). *Comparison of bootstrap standard errors of equating using IRT and equipercentile methods with polytomously-scored items under the common-item-nonequivalent-group design*. Doktora Tezi, University of Iowa, Iowa.
- Chu, K-L. (2002). *Equivalent group test equating with the presence of differential item functioning*. Doktora Tezi, The Florida State University, Florida.
- Chulu, B. W. (2006). *Equating high stakes educational measurement: a study of design and consequences*. Doktora Tezi, University of Massachusetts, Massachusetts, USA.
- Cohen, A. S. ve Kim, S. H. (1998). An investigation of linking methods under the graded response model. *Applied Psychological Measurement*, 22(2), 116-130.
- Cook L. ve Eignor D. R. (1991). NCME instructional module: IRT equating methods. *Educational Measurement: Issues and Practices*, 10(3), 37-45.
- Crocker, L. ve Algina, J. (1986). *Introduction to classical & modern test theory*. USA: Cengage Learning.
- Çetin, E. (2009). *Dikey ölçeklemede klasik test ve madde tepki kuramına dayalı yöntemlerin karşılaştırılması*. Doktora Tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Çokluk, Ö., Şekercioğlu, G. ve Büyüköztürk, Ş. (2014). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları*. (3. baskı). Ankara: Pegem Yayıncılık.
- De Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: The Guilford Press.
- Demir, S. ve Güler, N. (2014). Ortak maddeli denk olmayan gruplar desenine ilişkin bir test eşitleme çalışması. *International Journal of Human Sciences*, 11(2), 190-208.

- Dorans, N. J. ve Cook, L. L. (Editörler). (2016). *Fairness in educational assessment and measurement*. New York: Routledge.
- Dorans, N. J. ve Holland, P. W. (2000). Population invariance and the equatability of tests: basic theory and the linear case. *Journal of Educational Measurement*, 37(4), 281-306.
- Dorans, N. J., Moses, T. P. ve Eignor, D. R. (2010). *Principles and practices of test score equating*. Princeton, New Jersey: Educational Testing Service.
- EARGED (2010). *Uluslararası öğrenci değerlendirme programı PISA 2009 ulusal ön raporu*. Ankara: Milli Eğitim Bakanlığı Eğitimi Araştırma ve Geliştirme Daire Başkanlığı.
- Eid, G. K. (2005). Item parameter estimates onto a one-parameter scale the effects of sample size on the equating of test items. *Education*, 126(1), 165-180.
- Embretson, S. E. ve Reise, S. P. (2000). *Item response theory for psychologists*. London: Lawrence Erlbaum Associates Publishers.
- Felan, G. D. (2002, 14-16 Şubat). *Test equating: meani linear, equipercntile and item response theory*. Southwest Educational Research Association (SERA)'da sunulmuştur. Austin, TX.
- Gök, B. (2012). *Denk olmayan gruplarda ortak madde deseni kullanılarak madde tepki kuramına dayalı eşitleme yöntemlerinin karşılaştırılması*. Doktora Tezi. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Haebara, T. (1980). Equating lojistic ability scales by a weighted least squares method. *Japanese Psychological Research*, 22(3), 144-149.
- Hagge, S. L. (2010). *The impact of equating method and format representation of common items on the adequacy of mixed-format test equating using nonequivalent groups*. Doktora Tezi, The University of Iowa, Iowa.



- Hambleton, R. K. (1989). *Item response theory: introduction and bibliography*. (Rapor no:196) Amherst: University of Massachusetts.
- Hambleton, R. K. ve Jones, R. W. (1993). An NCME instructional module on comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practices*, 12, 38-47.
- Hambleton, R. K. ve Swaminathan, H. (1985). *Item response theory: principles and applications*. Boston: Kluwer, Nijhoff Publishing.
- Hambleton, R. K., Swaminathan, R. ve Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, USA: Sage Publication.
- Han, T., Kolen, M. J. ve Pohlmann, J. (1997). A comparison among IRT true-and observed score equating and traditional equipercentile equating. *Applied Measurement in Education*, 10(2), 105-121.
- Hanson, B. A. ve Beguin, A. A. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common item equating design. *Applied Psychological Measurement*, 26(1), 3-24.
- Holland, P. W. ve Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Editör) *Educational Measurement (4th ed.)* (s.87-220). Westport, CT: American Council on Education and Praeger.
- Holmes, S. E. (1986). Test equating and credentialing examinations. *Evaluation and the Health Professions*, 9, 230-249.
- İnci, Y. (2014). *Örneklem büyüklüğünün test eşitlemeye etkisi*. Yüksek Lisans Tezi. Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

- Jöreskog, K. G., Sorböm, D. (1986). Prells a program for multivariate data screening and data summarization: A preprocessor for LISREL. Mooresville, Ind: Scientific Software Inc.
- Kan, A. (2010). Test eşitleme: aynı davranışları ölçen farklı madde formlarına sahip testlerin istatistiksel eşitliğinin sınanması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(1), 16-21.
- Kan, A. (2011). Test Eşitleme: OKS testlerinin istatistiksel eşitliğinin sınanması. *Eğitim ve Bilim*, 36(160), 38-51.
- Kaskowitz, G. S. (1998). *The effect of error in item parameter estimates on linking and equating with the IRT test characteristic curve method*. Doktora Tezi, University of Maryland, USA, Maryland.
- Kaskowitz, G. S. ve De Ayala R. J. (2001). The effect of error in item parameter estimates on the test response function method of linking. *Applied Psychological Measurement*, 25 (1), 39–52.
- Kelecioğlu, H. (1994). *Öğrenci seçme sınavı puanlarının eşitlenmesi üzerine bir çalışma*. Doktora Tezi. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Keller, R. R. (2007). *A comparison of item response theory true score equating and item response theory-based local equating*. Doktora Tezi, University of Massachusetts, Massachusetts, USA.
- Keller, L. A., Skorupski, W. P., Swaminathan, H. ve Jodoin, M. G. (2004, Nisan). *An evaluation of the effect of the linking method on capturing changes in examinee distributions with mixed-format tests in an item response theory context*. National Council on Measurement in Education'da sunulmuştur. San Diego, CA.

- Kilmen, S. (2010). *Madde tepki kuramına dayalı test eşitleme yöntemlerinden kestirilen eşitleme hatalarının örneklem büyüklüğü ve yetenek dağılımına göre karşılaştırılması*. Doktora Tezi. Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Kim, S. H. ve Cohen, A. S. (2002). A comparison of linking and concurrent calibration under the graded response model. *Applied Psychological Measurement*, 26, 25-41.
- Kim, J. S. ve Hanson B. A. (2000). Test equating under the multiple-choice model. *Applied Psychological Measurement*, 26(3), 255-270.
- Kim, H. K. (2006). *The effect of repeaters on equating: A population invariance approach*. Doktora Tezi, The University of Iowa, Iowa.
- Kim, S. ve Kolen, M. J. (2004). STUIRT [computer program]. Iowa City IA: The Center for Advanced Studies in Measurement and Assessment, The University of Iowa, Iowa.
- Kim, S. ve Kolen, M. J. (2006). Robustness to format effects of IRT linking methods for mixed-format tests. *Applied Measurement in Education*, 19(4), 357-381.
- Kim, S. ve Kolen, M. J. (2007). Effects on scale linking of different definitions of criterion functions for the IRT characteristic curve methods. *Journal of Educational and Behavioral Statistics* 32(4), 371-397.
- Kim, S. ve Lee, W. (2006). An extension of four IRT linking methods for mixed-format tests. *Journal of Educational Measurement*, 43(1), 53-76.
- Kim, J. K. ve Nicewander, W. A. (1993). Ability estimation for conventional tests, *Psychometrika*, 58(4), 587-599.

- Kim, S., Walker, M. E., ve McHale, F. (2010). Comparisons among designs for equating mixed-format tests in large scale assessments. *Journal of Educational Measurement*, 47(1), 36-53.
- Kolen, M. J. (1981). Comparison of traditional and item response theory methods for equating tests. *Journal of Educational Measurement*, 18, 1-11.
- Kolen, M. J. (1988). An NCME instructional module on traditional equating methodology. *Educational Measurement: Issues and Practice*, 7(4), 29-36.
- Kolen, M. J. (2004). POLYEQUATE [computer program]. Iowa City IA: The Center for Advanced Studies in Measurement and Assessment, The University of Iowa.
- Kolen, M. J. (2007). Data collection designs and linking procedures. N. J. Dorans, M. Pommerich, ve P. W. Holland, (Editörler), *Linking and aligning scores and scales* (s. 31-55). Springer: New York.
- Kolen, M. J. ve Brennan, R. L. (1995). *Test equating: methods and practices*. New York: Springer.
- Kolen, M. J. ve Brennan, R. L. (2004). *Test equating, scaling and linking: methods and practices* (2nd.ed.). New York: Springer.
- Kolen, M. J. ve Brennan, R. L. (2014). *Test equating, scaling and linking: methods and practices* (3rd. ed.). New York: Springer.
- Langer, M. M. ve Swanson, D. B. (2010). Practical considerations in equating progress tests. *Medical Teacher*, 32(6), 509-512.
- Lee, W. ve Ban, J. (2010). A comparison of IRT linking procedures. *Applied Measurement in Education*, 23(1), 23-48.

- Livingston, S. A. (2004). *Equating test scores (without IRT)*. Princeton, New Jersey: Educational Testing Service.
- Lord, F. M. ve Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed-score equatings. *Applied Psychological Measurement*, 8(4),453-461.
- Loyd, B. H. ve Hoover, H. D. (1980). Vertical equating using the rasch model. *Journal of Educational Measurement*, 17(3), 179-193.
- Marco, G. L.(1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement*, 14(2), 139-160.
- MEB (2011). *PISA Türkiye*. Ankara: Eđitek.
- MEB (2013). *PISA 2012 ulusal ön raporu*. Ankara: Sebit.
- Meng, Y. (2012). *Comparison of kernel equating andı tem response theory equating methods*.Doktora Tezi, Massachusetts, USA.
- Michaelides, M. P. (2003, 21-25 Nisan). *Sensitivity of IRT equating to the behavior of test equating items*. American Educational Research Association'da sunulmuřtur. Chicago.
- Nozawa, Y. (2008). *Comparison of parametric and nonparametric IRT equating methods under the common item nonequivalent group design*. Doktora Tezi, Universtiy of Iowa, Iowa.
- Norman-Dvorak, R. L. (2009). *A comparison of kernel equating to the test characteristic curve method*. Doktora Tezi, University of Nebraska, Lincoln, USA.

- OECD (2010). *PISA 2009 results: What students know and can do: student performance in reading, mathematics and science* (Volume I). Paris: OECD Publishing.
- OECD (2014a). *PISA 2012 results in focus: what 15-year-olds know and what they can do with what they know*. Paris: OECD Publishing.
- OECD (2014b). *PISA 2012 results: what students know and can do-student performance in mathematics, reading and science* (Volume I, Revised edition, February, 2014), Paris: OECD Publishing.
- Öztürk, N. (2010). *Akademik personel ve lisansüstü eğitimi giriş sınavı puanlarının eşitlenmesi üzerine bir çalışma*. Yüksek Lisans Tezi. Hacettepe Üniversitesi Sosyal Bilimler Enstitüsü, Ankara.
- Öztürk-Gübeş, N. ve Kelecioğlu, H. (2017). Eşitleme sonuçlarının gruptan bağımsızlığının incelenmesi. *İlköğretim Online*, 16(1), 217-227.
- Petersen, N. S., Cook, L. L. ve Stocking, M. L. (1983). IRT versus conventional equating methods: a comparative study of scale stability. *Journal of Educational Statistics*, 8(2), 137-156.
- Petersen, N. S., Kolen, M. J. ve Hoover, H. D. (1989). Scaling, norming and equating. In: Linn, R. L. (Ed.). *Educational Measurement* (pp. 221-262). New York: Macmillan.
- Powers, S. ve Kolen, M. J. (2014). Evaluating equating accuracy assumptions for groups that differ in performance. *Journal of Educational Measurement*, 51(1), 39-56.
- Roos, L. L. (2002). *Testing the robustness of item response theory equating to violation of the condition of equivalent item difficulty parameters in both the extant pool and infused items: a comparison of the effect on true-score and observed-score equating*. Doktora Tezi, University of Nebraska, Lincoln.

- Shin, M. (2015). *An investigation of subtest score equating methods under classical test theory and item response theory frameworks*. Doktora Tezi, University of Massachusetts, Amherst.
- Sinharay, S. ve Hollland, P. W. (2010). A new approach to comparing several equating methods in the context of the NEAT design. *Journal of Educational Measurement*, 47(3), 261-285.
- Skaggs, G. ve Lissitz, R. W. (1986). IRT test equating: relevant issues and a review of recent research. *Review of Educational Research*, 56(4), 495-529.
- Skoupski, W. P., Jodoin, M. G., Keller, L.A., Swaminathan, H. (2003, 21-25 Nisan). *An evaluation of equating procedures for capturing growth*. National Council on Measurement in Education'da sunulmuştur. Chicago, IL.
- Speron, E. (2009). *A comparison of metric linking procedures in item response theory*. Doktora Tezi, Illinois Institute of Technology, Chicago, Illinois.
- Stocking, M. L. ve Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7(2), 201-210.
- Şirin, S. R. ve Vatanartıran, S. (2014). *PISA 2012 değerlendirmesi: Türkiye için veriye dayalı eğitim reformu önerisi* (Yayın no: TÜSİAD-T/2014-2/549). İstanbul: TÖDER.
- Tate, R. (2000). Performance of a proposed method for the linking of mixed-format tests with constructed response and multiple choice items. *Journal of Educational Measurement*, 37(4), 329-346.
- Tian, F. (2011). *A comparison of equating/linking using the Stocking-Lord method and concurrent calibration with the mixed-format tests in the non-equivalent groups common item design under IRT*. Doktora Tezi, Boston College, Massachusetts.

- Tsai, T. H., Hanson, B. A., Kolen, M. J. ve Forsyth, R. A. (2001). A comparison of bootstrap standard errors of IRT equating methods for the common-item nonequivalent groups design. *Applied Measurement in Education*, 14(1), 17-30.
- Uysal, İ. (2014). *Madde tepki kuramına dayalı test eşitleme yöntemlerinin karma modeller üzerinde karşılaştırılması*. Yüksek Lisans Tezi. Abant İzzet Baysal Üniversitesi, Eğitim Bilimleri Enstitüsü, Bolu.
- Von Davier, A. A., Holland, P. W. ve Thayer, D. T. (2004). *The kernel method of test equating*. New York: Springer.
- Von Davier, A. A. ve Wilson, C. (2007). IRT true-score test equating: a guide Through assumptions and applications. *Educational and Psychological Measurement*, 67(6), 940-957.
- Yang, W. L. (1997). *The effects of content homogeneity and equating method on the accuracy of common item test equating*. Doktora Tezi, Michigan State University, Michigan.
- Yang, W. L. ve Houang, R. T. (1996, 8-12 Nisan). *The effect of anchor length and equating method on the accuracy of test equating comparisons of linear and IRT-based equating using an anchor-item design*. American Educational Research Association'da sunulmuştur. New York.
- Wang, X. (2012). *Effect of sample size on IRT equating of uni-dimensional tests in common item non-equivalent group design: a monte carlo simulation study*. Doktora Tezi, Virginia State University, USA.
- Wang, S., Jiao, H. ve Xiang, Y. (2013, 28-30 Nisan). *A comparison of different reduction methods for bias of maximum likelihood estimator of student ability based on graded response and generalized partial credit models*. National Council on Measurement in Education'da sunulmuştur. San Francisco, CA.



Way, W. D. ve Tang, K. L. (1991, Nisan). *A comparison of four logistic model equating methods*. American Educational Research Association'da sunulmuştur. Chicago.

Wingersky, M. S., Cook, L. L. ve Eignor, D. R. (1987). *Specifying the characteristics of linking item used for item response theory item calibration*. Princeton, New Jersey: Educational Testing Service.

Woldbeck, T. (1998, Nisan). *Basic concepts in modern methods of test equating*. Southwest Psychological Association'da sunulmuştur. New Orleans, LA.

Zeng (1991). *Standard errors of linear equating for the single-group design*. (ACT-RR-91-4). Iowa city: American College Testing, Inc.

Zimowski, M. F., Muraki, E., Mislevy, R. J. ve Bock, R. D. (2003). BILOG-MG 3 For Windows: Multiple-group IRT analysis and test maintenance for binary items [computer program]. Skokie, IL: Scientific software international, Inc.

Zhu, W. (1998). Test equating: what, why, how? *Research Quarterly for Exercise and Sport* 69(1), 11-23.

## EKLER

EK-1: Kız ve Erkek Öğrencilerden Elde Edilen Madde Parametreleri

Madde	Kitapçık-1				Kitapçık-8				
	Kız		Erkek		Madde	Kız		Erkek	
	a	b	a	b		a	b	a	b
O1	0.535	-2.219	0.629	-1.767	O1	1.458	0.957	2.351	0.814
O2	0.600	-0.941	1.049	-0.485	O2	0.903	1.956	0.887	1.685
O3	1.738	0.318	0.974	0.156	O3	1.046	-0.864	0.952	-0.619
O4	1.963	1.012	1.558	1.240	O4	0.944	-0.001	1.119	0.100
O5	0.590	0.955	0.829	0.951	O5	2.011	0.878	2.283	1.100
O6	0.807	-0.301	1.150	-0.359	O6	0.444	0.282	0.986	0.058
O7	1.966	0.693	1.868	0.672	O7	0.658	2.433	1.022	1.794
O8	0.778	0.308	0.616	1.289	O8	0.452	-2.536	0.694	-1.588
O9	1.042	1.297	1.284	0.813	O9	1.356	0.523	1.261	0.432
O10	1.238	1.473	1.223	1.881	O10	1.020	0.233	0.881	0.065
O11	0.926	-1.330	0.904	-0.534	O11	0.677	1.372	0.890	0.491
O12	0.716	1.136	0.916	0.883	O12	0.633	0.938	0.574	0.906
O13	0.659	0.430	0.574	0.874	O13	0.378	2.397	0.391	1.556
O14	0.548	1.460	1.027	1.124	O14	0.744	0.184	0.954	0.550
O15	1.367	-0.192	1.073	0.067	O15	0.318	-2.172	0.291	-2.398
O16	0.738	0.575	0.875	0.502	O16	0.425	0.766	0.346	0.338
O17	0.550	1.355	0.795	0.739	O17	0.548	2.610	0.434	3.711
A1	1.560	-0.158	1.231	0.011	B1	0.513	-2.684	0.456	-2.432
A2	0.907	1.548	1.085	1.560	B2	1.013	-0.001	1.420	-0.042
A3	1.159	-0.262	1.199	-0.230	B3	0.624	1.178	0.625	1.204
A4	0.642	-1.062	0.679	-1.444	B4	0.464	3.425	0.572	1.523
A5	0.897	-0.562	0.829	-0.472	B5	0.691	1.215	0.902	0.842
A6	0.929	-0.837	0.879	-1.154	B6	1.104	1.059	0.872	1.666
A7	0.923	-0.840	0.921	-0.616	B7	0.788	-1.259	0.805	-1.453
A8	0.918	0.521	1.377	0.096	B8	0.540	-0.868	0.478	-0.725
A9	0.712	-1.437	0.813	-0.804	B9	0.651	1.137	0.439	0.574
A10	1.084	-1.129	1.361	-0.778	B10	0.550	-0.613	0.506	-0.691
A11	0.642	1.374	1.180	1.517	B11	0.592	0.713	0.693	0.994
A12	0.978	-0.560	0.707	-0.495	B12	0.752	0.895	1.101	0.785
A13	0.525	1.052	0.413	1.328	B13	0.870	0.097	0.932	0.419
A14	1.632	-1.334	1.398	-1.537	B14	0.783	3.842	0.813	3.662
A15	0.703	0.949	1.251	0.175	B15	0.434	0.116	0.407	0.079
A16	1.509	-0.160	1.438	-0.016	B16	0.817	0.695	0.731	0.881

EK-2: Kitapçık-1 İçin Tek Numaralı Madde Setinden Elde Edilen Yetenek Parametreleri

1	-0.221	60	0.7106	119	1.6084	178	1.5658	237	0.5657	296	-0.3727
2	0.5906	61	-0.9302	120	1.1898	179	1.3529	238	0.7010	297	0.0514
3	0.0664	62	-0.1955	121	-0.3795	180	-0.4158	239	1.6496	298	-0.3621
4	-1.0318	63	-0.6567	122	0.1884	181	0.6237	240	0.8266	299	-0.5700
5	-0.8152	64	0.2697	123	-1.1848	182	1.5649	241	1.3327	300	-0.3823
6	-0.6892	65	-0.9000	124	0.6335	183	-0.1630	242	-0.0262	301	-0.3816
7	0.1727	66	-0.0230	125	0.0529	184	-1.1591	243	1.1756	302	0.2236
8	-0.4338	67	-0.1322	126	-0.8682	185	-0.3808	244	0.1871	303	-0.3473
9	0.1554	68	-1.3099	127	0.3360	186	1.1525	245	0.3381	304	-1.1635
10	-0.7381	69	0.6944	128	0.1968	187	1.0557	246	0.0611	305	1.7411
11	0.0989	70	-0.6059	129	-2.0374	188	-0.1814	247	-0.2181	306	1.0058
12	-1.7345	71	0.5839	130	-0.6153	189	1.7938	248	0.3395	307	1.0678
13	-0.3143	72	0.5010	131	-0.4802	190	0.6832	249	-0.4529	308	0.3554
14	-1.0254	73	-0.5953	132	1.6129	191	2.1617	250	-1.1659	309	-0.6485
15	-0.4535	74	-1.3287	133	1.2444	192	-0.2600	251	-0.1141	310	-0.7601
16	0.1755	75	0.7039	134	0.9183	193	0.3459	252	-0.3494	311	-0.3041
17	-0.0148	76	-0.9390	135	-0.1016	194	-0.6928	253	0.3404	312	-0.7434
18	-1.6713	77	-0.9535	136	-0.3737	195	0.1555	254	-1.3287	313	-0.9286
19	-0.4775	78	-0.0956	137	-1.4242	196	-0.7334	255	-1.2649	314	-0.3791
20	0.7400	79	-0.8817	138	0.9902	197	0.3209	256	0.5544	315	-0.8856
21	0.0059	80	-0.4160	139	0.7684	198	-0.6683	257	-0.1536	316	-0.7401
22	0.3116	81	-0.0064	140	0.1985	199	-0.8792	258	1.5198	317	-0.6010
23	-0.2472	82	-0.3104	141	-0.0110	200	0.6947	259	1.0033	318	-0.5364
24	0.2926	83	-1.7484	142	0.0886	201	1.5649	260	1.1695	319	-0.4157
25	-0.7134	84	0.9079	143	-0.8777	202	1.3581	261	0.4114	320	-0.7313
26	-0.3529	85	1.2555	144	0.3675	203	-0.6538	262	0.6866	321	-1.8579
27	-0.6286	86	-0.5408	145	-1.8336	204	0.0054	263	1.2612	322	-1.6749
28	-1.5388	87	0.1679	146	0.2659	205	0.6593	264	1.2375	323	0.0446
29	-0.7927	88	-0.8330	147	-1.5559	206	1.1607	265	0.8571	324	1.4730
30	-1.0274	89	0.4672	148	-0.0087	207	0.5260	266	-1.3195	325	1.1707
31	0.1319	90	0.3082	149	-1.4006	208	1.2379	267	-0.7163	326	0.3601
32	-1.3990	91	0.7734	150	-0.7781	209	1.4599	268	0.2710	327	-0.1187
33	-0.6718	92	0.7871	151	-0.7848	210	1.1646	269	0.3633	328	0.7787
34	-1.5033	93	0.3681	152	-0.9239	211	-0.0238	270	0.3664	329	0.1242
35	-1.3465	94	-0.9362	153	-0.8078	212	0.1658	271	-1.6334	330	0.3361
36	0.8616	95	-0.4355	154	-1.6796	213	1.5603	272	-0.5926	331	-0.7212
37	-0.8047	96	-1.1210	155	0.2062	214	0.3994	273	-1.0947	332	0.3089
38	-0.6666	97	-1.9687	156	-0.9479	215	0.9375	274	-1.5998	333	1.4730
39	-0.3470	98	-1.1474	157	1.0279	216	0.8164	275	0.5113	334	0.6763
40	0.7029	99	-0.7992	158	-0.6655	217	0.6768	276	-0.9084	335	0.4248
41	-0.4274	100	-0.5817	159	-1.0054	218	1.3426	277	0.8297	336	0.1139
42	-1.7450	101	-0.9601	160	0.2764	219	-0.2545	278	1.1541	337	0.2121
43	0.0354	102	0.2218	161	-0.8532	220	-0.6755	279	1.0896	338	0.7240
44	0.0876	103	-0.0487	162	0.8774	221	-0.0904	280	0.0882	339	0.7474
45	-0.3582	104	0.25690	163	1.3352	222	-0.3092	281	-0.3168	340	0.7547
46	-1.5881	105	-0.5291	164	1.5861	223	1.1406	282	-0.5231	341	-0.5017
47	0.1850	106	-0.8050	165	0.5549	224	0.8218	283	-0.7310	342	0.7468
48	-0.8095	107	-1.6492	166	-0.3168	225	1.3628	284	0.8610	343	0.4306
49	-0.8714	108	1.9353	167	0.3608	226	0.8929	285	-0.8927	344	1.4730
50	-1.0187	109	1.0684	168	-1.1047	227	1.9348	286	0.9306	345	0.8616
51	0.5306	110	0.4058	169	0.6476	228	1.1882	287	0.8532	346	-0.5220
52	0.4306	111	-0.3972	170	-0.4578	229	-0.5413	288	1.5198	347	-0.1971
53	0.5014	112	0.2176	171	0.2329	230	1.9088	289	0.716	348	0.0667
54	0.7257	113	-1.7577	172	-0.6485	231	-0.7134	290	-1.1587	349	-0.1805
55	1.2347	114	0.7892	173	-0.9655	232	0.2415	291	0.0146	350	0.2719
56	1.0382	115	-0.0776	174	0.1107	233	-0.7566	292	1.2348		
57	-0.5591	116	-0.5963	175	1.0873	234	-1.7667	293	0.0928		
58	-0.1767	117	0.8227	176	2.1617	235	-1.0359	294	-0.5114		
59	0.2755	118	1.4243	177	0.3629	236	-0.9623	295	0.0142		

EK-3: Kitapçık-1 İçin Çift Numaralı Madde Setinden Elde Edilen Yetenek Parametreleri

1	0.6176	60	-0.4398	119	0.4115	178	-0.1007	237	0.4107	296	0.2273
2	-0.4928	61	-0.7207	120	1.8507	179	0.9038	238	1.8126	297	0.5303
3	0.8665	62	-0.1643	121	-0.6932	180	0.0513	239	1.5127	298	-0.8300
4	0.2208	63	-1.3783	122	0.3275	181	1.0491	240	1.4562	299	-1.6598
5	-1.4898	64	-0.3204	123	-0.0050	182	1.2840	241	1.8507	300	0.0156
6	-0.8793	65	0.0947	124	1.4410	183	0.8311	242	0.5545	301	-0.5796
7	-0.2268	66	-0.5022	125	-0.0718	184	-0.3446	243	0.2139	302	0.2722
8	-0.4673	67	0.4143	126	-0.7344	185	0.4950	244	-0.8553	303	-0.0673
9	0.0108	68	-0.6496	127	-0.2205	186	0.4643	245	-1.0003	304	-0.3365
10	-0.3073	69	1.6128	128	-0.8787	187	0.5733	246	-0.8219	305	0.6489
11	0.5120	70	-0.1782	129	-1.2549	188	-0.0357	247	-0.5976	306	0.9935
12	-0.9892	71	0.2678	130	-1.3776	189	0.4134	248	0.0863	307	0.3695
13	-1.0946	72	-0.7278	131	-1.2580	190	-0.3185	249	-0.5058	308	0.0257
14	-0.7173	73	-0.1516	132	1.1540	191	1.2880	250	-0.6072	309	-0.5476
15	-1.6605	74	-1.3689	133	1.1168	192	-0.4201	251	-0.3150	310	-0.5495
16	0.1808	75	1.0215	134	0.2405	193	0.5986	252	-0.6165	311	-0.9732
17	-0.5372	76	0.2779	135	0.0960	194	-0.0066	253	-0.4158	312	-0.4297
18	-1.0238	77	-0.0781	136	-0.6962	195	-0.2545	254	-1.3931	313	-0.1943
19	0.4170	78	0.3699	137	-1.6605	196	-1.2830	255	-1.1599	314	-1.3042
20	0.2257	79	-0.9253	138	1.4050	197	-0.1283	256	-0.8150	315	-1.2765
21	0.3234	80	-0.8114	139	1.4818	198	-1.5359	257	-0.0936	316	0.0244
22	0.6382	81	-0.8252	140	1.3092	199	-0.8293	258	1.4370	317	0.0027
23	-0.0188	82	0.0606	141	-0.6706	200	1.0521	259	-0.2498	318	0.9643
24	0.1683	83	-1.0410	142	0.9314	201	0.4184	260	0.9388	319	-0.7372
25	-1.9626	84	0.7416	143	-0.9918	202	1.4247	261	1.0739	320	-0.3429
26	-1.3774	85	0.5271	144	-0.6710	203	-0.4710	262	0.9184	321	0.4115
27	-1.1807	86	0.4367	145	-0.3585	204	0.4803	263	1.5296	322	-0.8095
28	-0.8327	87	0.5686	146	0.9314	205	0.4028	264	1.4769	323	-0.4322
29	-1.3567	88	-0.7805	147	-0.6925	206	1.1859	265	0.9292	324	1.6502
30	-0.5927	89	0.5621	148	-0.6775	207	-0.2183	266	-0.5699	325	1.5834
31	-0.4801	90	0.8734	149	-1.0238	208	0.8680	267	-0.7673	326	0.3980
32	-0.7812	91	0.8610	150	-0.0999	209	0.7576	268	-0.4701	327	0.0849
33	-0.9014	92	0.5082	151	-0.0609	210	0.3964	269	0.6235	328	1.1520
34	-1.7963	93	-0.2660	152	-0.5156	211	0.7909	270	-0.1938	329	0.6347
35	-0.6230	94	-0.9339	153	-0.7770	212	-0.6950	271	-1.1240	330	1.6585
36	0.6233	95	-0.0748	154	-1.0238	213	0.6097	272	-0.4736	331	-0.4402
37	-0.4205	96	-1.6466	155	-0.4661	214	0.2057	273	-1.3900	332	-0.2466
38	-1.2952	97	-1.5359	156	-0.4728	215	1.0597	274	-0.7237	333	0.8026
39	-0.6362	98	-0.3264	157	0.4938	216	0.3132	275	-0.3857	334	2.1081
40	0.0136	99	-0.6235	158	0.4117	217	0.9535	276	-1.7963	335	1.2079
41	-0.2623	100	-1.1648	159	-0.1178	218	1.4818	277	0.9553	336	-0.7624
42	-1.3945	101	-0.1449	160	-0.6712	219	-1.1098	278	1.8881	337	0.1425
43	0.8816	102	-0.2063	161	-1.9626	220	-0.5259	279	1.5193	338	0.5761
44	-0.2313	103	-0.2263	162	0.6642	221	0.1809	280	-0.1727	339	1.5883
45	-1.0032	104	-1.0779	163	1.1227	222	-0.8121	281	-0.1515	340	1.2241
46	-0.5978	105	-0.7372	164	1.4614	223	0.6383	282	0.3467	341	-0.9014
47	-0.3237	106	-0.5008	165	-0.0700	224	1.6239	283	-0.607	342	1.5025
48	-0.4145	107	-0.9371	166	0.2943	225	1.5029	284	-0.1085	343	0.5711
49	-1.3776	108	1.2407	167	0.0262	226	0.6121	285	-0.0926	344	1.2118
50	-0.7796	109	0.6185	168	-0.8046	227	0.9410	286	0.4435	345	0.0592
51	0.5042	110	1.2534	169	-0.0974	228	0.8989	287	0.4118	346	0.0021
52	0.3024	111	-0.0525	170	-1.0290	229	-0.4209	288	1.4170	347	-0.7433
53	-0.5407	112	0.2887	171	0.4223	230	0.4721	289	0.4834	348	-0.1228
54	-0.8537	113	-1.0489	172	0.6576	231	-0.4643	290	-1.0383	349	-0.0070
55	1.0296	114	0.2722	173	-0.5143	232	0.6165	291	-0.8039	350	0.4253
56	2.1572	115	0.5485	174	0.4557	233	-0.4168	292	1.6738		
57	0.1132	116	0.5588	175	0.5830	234	-1.0946	293	-0.6996		
58	-0.3442	117	1.0867	176	1.4979	235	0.8314	294	-0.7433		
59	0.1659	118	0.9390	177	0.2046	236	-0.2066	295	0.4107		

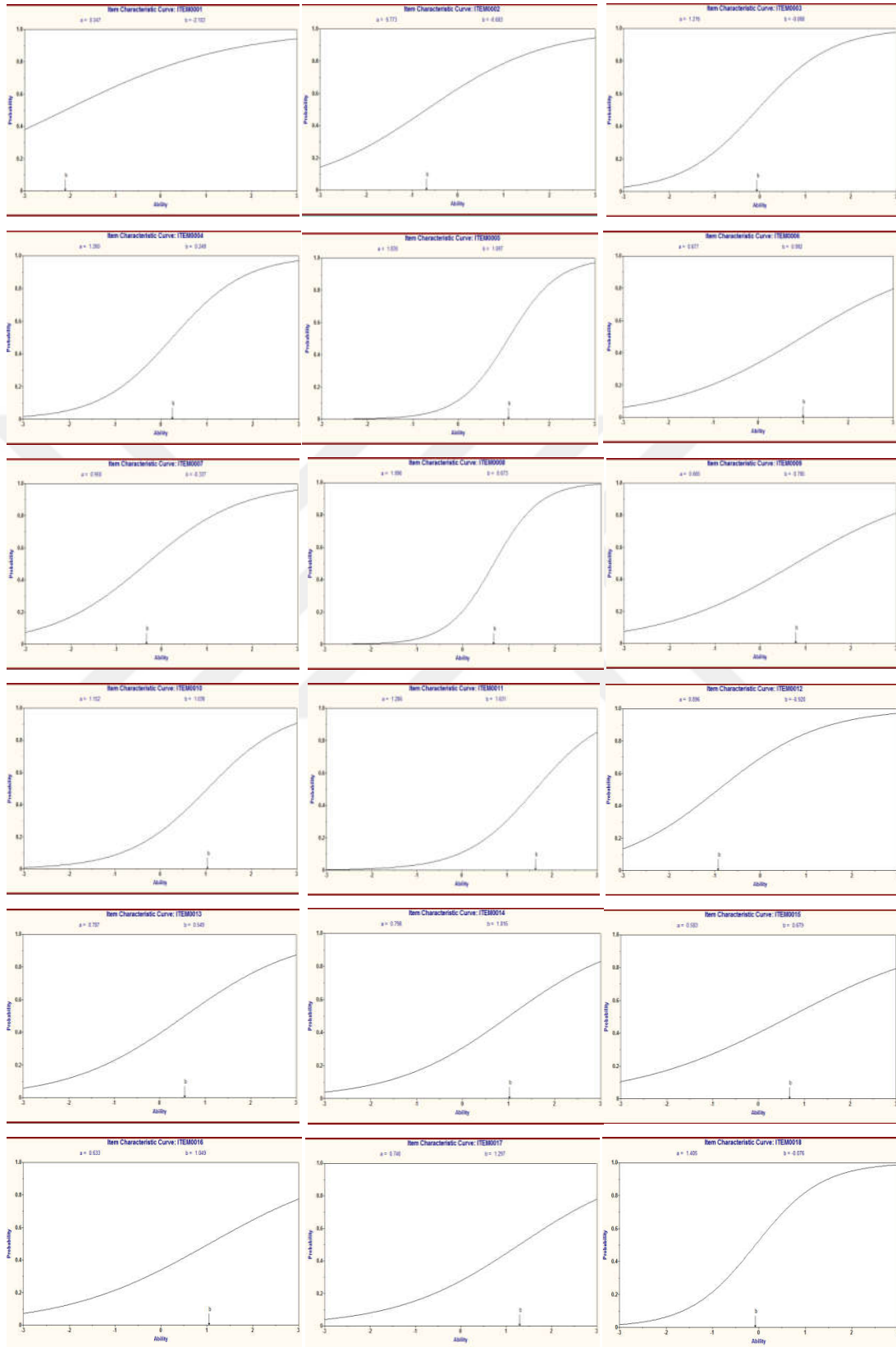
EK-4: Kitapçık-8 İçin Tek Numaralı Madde Setinden Elde Edilen Yetenek Parametreleri

1	-1.1241	60	1.2808	119	-0.4575	178	-0.2017	237	-1.0251	296	1.4765
2	-1.5277	61	1.8627	120	-0.2015	179	-0.5702	238	-0.0593	297	-0.7207
3	-1.0825	62	-0.2648	121	-0.5950	180	-0.4605	239	-0.7943	298	1.1447
4	-0.6820	63	-0.0624	122	0.2725	181	-0.2975	240	-0.4244	299	0.1424
5	-1.5225	64	0.6088	123	-0.1923	182	1.4772	241	0.5973	300	0.3243
6	-1.3468	65	-0.3143	124	-0.1531	183	1.3746	242	0.0170	301	-0.3280
7	-1.1386	66	-1.5277	125	-0.2207	184	-0.7076	243	0.8458	302	1.3578
8	-1.4167	67	0.2908	126	-0.1101	185	0.2427	244	1.5617	303	1.3786
9	-1.6090	68	-1.3242	127	0.2661	186	0.1825	245	0.9827	304	0.0860
10	-1.3325	69	-1.1261	128	1.1342	187	-0.9464	246	0.2427	305	-0.2261
11	-1.7579	70	-1.1290	129	-0.0217	188	-0.0077	247	0.5720	306	0.5505
12	-0.5074	71	-1.2764	130	-0.2176	189	-0.3535	248	0.8641	307	-0.3402
13	-0.6844	72	0.0712	131	0.2548	190	-1.2022	249	0.1367	308	0.4757
14	-1.0942	73	0.3042	132	-0.1525	191	1.6425	250	-0.2180	309	-0.6003
15	0.4054	74	0.3138	133	-0.7626	192	-0.1564	251	1.0817	310	0.7051
16	-1.0647	75	-0.7080	134	-0.7288	193	-0.9415	252	0.3953	311	-0.1675
17	-0.9288	76	-0.7671	135	-0.7901	194	-0.1834	253	0.0869	312	0.4153
18	-0.7637	77	-0.5433	136	0.0416	195	0.2830	254	1.0231	313	-0.1514
19	-1.4082	78	-0.9957	137	-0.2643	196	0.5278	255	-0.2777	314	0.7707
20	-1.1356	79	-0.0640	138	-0.1810	197	0.9742	256	0.3425	315	0.1018
21	-0.7905	80	-1.1547	139	1.3668	198	-1.2071	257	0.3414	316	-0.5191
22	-1.3325	81	-0.7087	140	0.0183	199	0.9000	258	1.1715	317	-0.6590
23	-1.0375	82	-1.1481	141	0.9922	200	0.9809	259	0.1026	318	-0.2953
24	-0.6895	83	-0.0823	142	-0.5567	201	-0.1045	260	2.1380	319	0.8435
25	-1.2578	84	-0.8368	143	1.2265	202	0.3795	261	-0.1592	320	-0.4058
26	0.1047	85	-0.1602	144	0.0024	203	0.6023	262	0.7130	321	0.4050
27	-0.8052	86	-1.2085	145	-0.3067	204	-0.5576	263	0.4919	322	2.3828
28	-0.6868	87	-0.5196	146	-0.5040	205	0.7314	264	-0.9660	323	-0.5447
29	0.5707	88	-0.2645	147	-0.5351	206	0.8256	265	-0.1765	324	0.6982
30	-1.5277	89	1.4475	148	-0.1118	207	1.1667	266	1.1014	325	-0.0553
31	-0.7605	90	-1.2483	149	-1.0936	208	-0.7364	267	0.2961	326	-0.0772
32	-0.4791	91	-0.4700	150	0.1016	209	2.3828	268	0.5593	327	0.4463
33	-1.0153	92	-0.3107	151	0.4439	210	0.8082	269	-0.5521	328	0.5013
34	-0.8223	93	1.0360	152	-1.1217	211	1.1832	270	1.1596	329	-0.9800
35	-0.5936	94	-0.2756	153	-0.4353	212	0.8662	271	1.2133	330	-0.0986
36	-0.8581	95	-0.5843	154	0.7625	213	1.4861	272	0.2389	331	0.4367
37	-0.9732	96	-0.8437	155	1.4738	214	-0.6311	273	0.9684	332	-0.0466
38	-1.3263	97	1.6400	156	-0.6713	215	-0.2732	274	0.6804	333	0.1383
39	-0.7655	98	-0.5615	157	0.5984	216	-0.9571	275	0.7891	334	1.5777
40	-0.6820	99	-0.2145	158	-0.8466	217	-0.8852	276	0.6306	335	-0.2863
41	-0.6844	100	-0.7411	159	0.0419	218	0.7798	277	1.8459	336	0.0755
42	-0.5069	101	-0.0708	160	0.5541	219	1.6462	278	-0.0754	337	1.5665
43	-0.7701	102	-0.2809	161	0.6298	220	0.4488	279	0.9401	338	1.0294
44	-1.3005	103	0.3480	162	-0.5615	221	-0.7024	280	0.9456	339	0.8570
45	0.2368	104	0.9809	163	-0.5566	222	-0.8425	281	0.4503	340	0.3783
46	-0.4129	105	-0.2884	164	0.6171	223	0.4496	282	-0.2226	341	1.2575
47	0.3980	106	0.2767	165	0.0267	224	0.5791	283	0.8175	342	0.4174
48	-1.2075	107	0.5573	166	0.9328	225	1.3322	284	-0.1514	343	-0.2414
49	0.1047	108	-0.3582	167	-0.2320	226	-0.5880	285	0.0710	344	0.9922
50	-0.2652	109	-0.4587	168	-0.9241	227	0.2220	286	0.8795	345	1.1753
51	0.8894	110	-0.7326	169	-0.4456	228	1.4155	287	0.0367	346	-0.0936
52	0.7562	111	-0.4178	170	-0.0756	229	0.4742	288	1.3006	347	-0.9677
53	-0.7783	112	-0.1787	171	-0.5428	230	0.3830	289	-0.1841	348	0.8705
54	-1.4558	113	-0.7740	172	2.0313	231	1.2160	290	0.1219	349	1.3643
55	-0.4591	114	-0.4428	173	-0.1030	232	0.7028	291	0.2370	350	0.0749
56	-0.4419	115	-0.8727	174	-0.8608	233	-1.1935	292	0.9013	351	0.8588
57	-0.8268	116	0.2127	175	0.0925	234	-0.4545	293	0.4402		
58	-0.5981	117	-0.2871	176	1.2743	235	0.8399	294	0.9000		
59	-0.4068	118	0.1567	177	-0.6721	236	1.1879	295	0.2254		

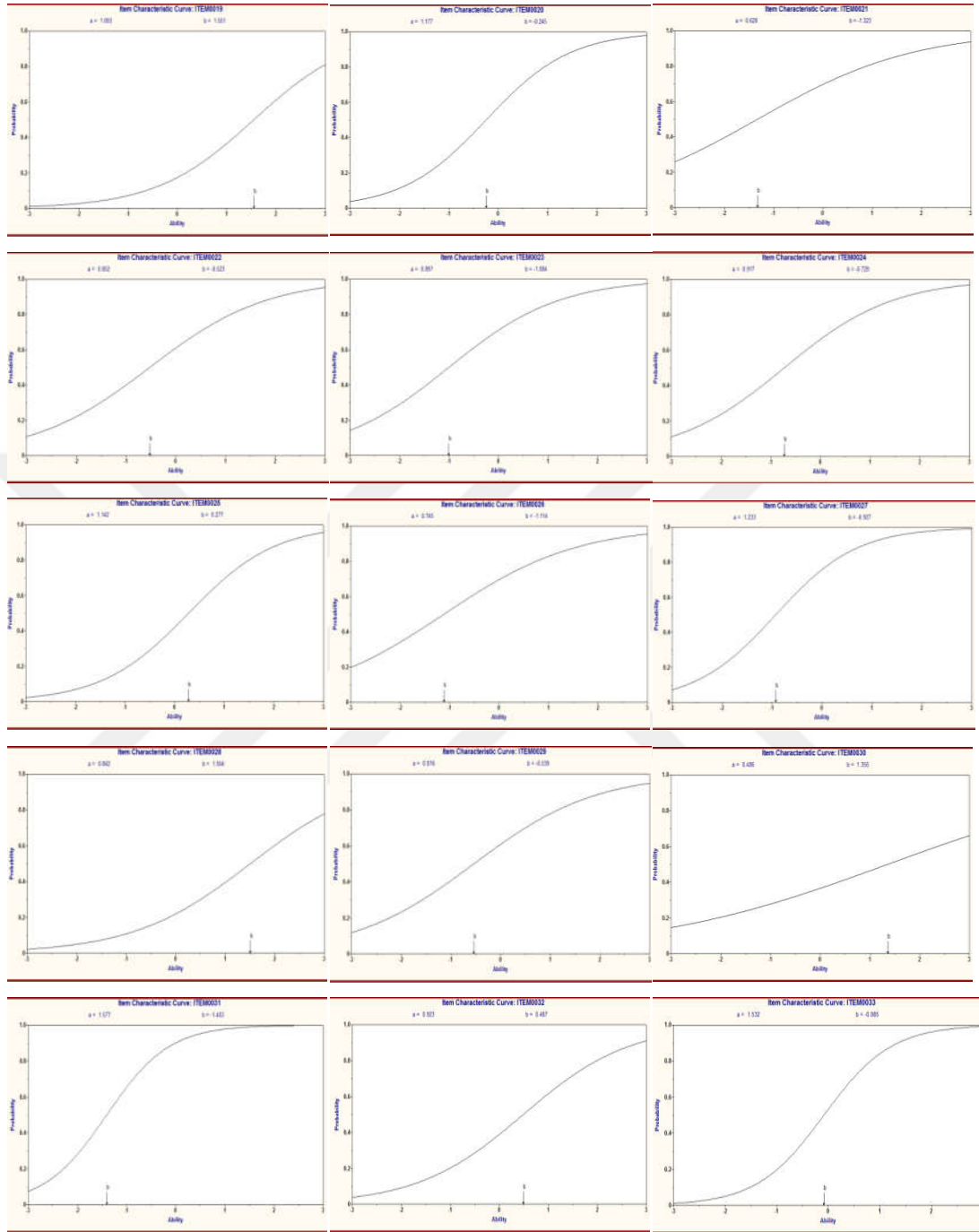
EK-5: Kitapçık-8 İçin Çift Numaralı Madde Setinden Elde Edilen Yetenek Parametreleri

1	-0.5186	60	1.3089	119	-0.2835	178	0.1919	237	-0.3880	296	1.0419
2	-1.4837	61	2.0041	120	-0.6301	179	-1.2099	238	-0.5297	297	0.3403
3	-1.6628	62	-0.5967	121	0.3703	180	-1.1849	239	-0.0523	298	0.7897
4	-0.3497	63	-0.0339	122	0.2471	181	-0.1693	240	-0.2446	299	0.4995
5	-0.9411	64	0.6178	123	0.1577	182	1.2571	241	0.1434	300	0.3631
6	-0.7134	65	-0.3040	124	0.2297	183	2.0041	242	0.9381	301	0.7027
7	-1.3810	66	-0.8556	125	-0.4818	184	-0.4399	243	0.3105	302	1.4593
8	-0.8548	67	-0.4809	126	0.4208	185	-0.0795	244	1.1857	303	0.7464
9	-0.8181	68	-1.2622	127	0.0350	186	0.1590	245	0.5444	304	0.5087
10	-1.1229	69	-1.4127	128	1.6273	187	0.6117	246	-0.7944	305	-0.1086
11	-1.0374	70	-0.9498	129	0.7668	188	0.6178	247	0.2886	306	0.8554
12	0.1510	71	-0.7178	130	-0.3661	189	0.5169	248	1.8401	307	0.0426
13	0.2330	72	-1.1229	131	-0.1966	190	0.2322	249	0.6690	308	0.8185
14	-1.0667	73	-0.1100	132	-0.6435	191	1.6362	250	-0.0950	309	0.0664
15	0.2232	74	0.1677	133	-0.3332	192	-0.4601	251	-0.4960	310	0.7885
16	-0.7467	75	-0.5107	134	-1.5256	193	-0.2926	252	-0.9968	311	0.3197
17	-0.8548	76	-0.3999	135	-0.8358	194	0.0350	253	-0.6859	312	0.9564
18	-1.4648	77	-0.5838	136	-1.1567	195	0.4092	254	-0.7327	313	0.2357
19	-0.4194	78	-0.4473	137	-0.0100	196	0.6554	255	-0.5984	314	0.4214
20	-1.1610	79	0.0475	138	-0.6945	197	-0.2335	256	0.5403	315	-0.1290
21	-0.8749	80	0.2313	139	0.5383	198	-0.3966	257	1.0413	316	-0.0095
22	-0.7291	81	-0.5500	140	0.1179	199	1.2375	258	0.9546	317	0.5502
23	-1.1707	82	-0.3829	141	-0.1474	200	0.8702	259	0.4701	318	0.2869
24	-0.3669	83	-1.0341	142	-0.6462	201	-0.4303	260	1.8670	319	0.3032
25	-0.6634	84	-0.5830	143	1.7823	202	0.7414	261	0.4806	320	0.6863
26	0.1881	85	-0.9349	144	-0.5893	203	0.4563	262	0.5834	321	0.3596
27	-0.0215	86	-0.7944	145	-0.7472	204	-0.2276	263	0.6138	322	1.3164
28	-1.8707	87	-0.8181	146	-0.7562	205	0.4358	264	-0.5586	323	-0.6863
29	0.9766	88	-1.0433	147	-1.5108	206	0.3311	265	-0.0480	324	-0.0171
30	-1.1057	89	2.2593	148	0.5946	207	0.8152	266	0.9745	325	0.4189
31	-0.9875	90	-0.5713	149	-1.2353	208	-1.0584	267	0.7647	326	0.9186
32	-0.3907	91	-0.8486	150	0.1260	209	1.3218	268	1.0975	327	-0.1680
33	-0.0495	92	-0.2996	151	-0.4583	210	1.0501	269	0.6755	328	1.0130
34	-0.1767	93	0.2681	152	-0.8604	211	1.3880	270	0.3317	329	-0.7250
35	-0.7110	94	-0.5757	153	-0.9753	212	0.6437	271	0.2747	330	1.3614
36	-0.4473	95	-0.3763	154	-0.1487	213	1.3559	272	-0.2294	331	0.6533
37	-0.3419	96	-0.0054	155	2.0584	214	-0.0171	273	0.2741	332	0.8957
38	-1.1394	97	0.5651	156	-1.0341	215	-0.0958	274	-0.3126	333	0.2773
39	-1.2269	98	-0.4603	157	0.0344	216	0.1203	275	-0.0571	334	0.5417
40	-1.3537	99	-0.7784	158	-0.3146	217	-0.4615	276	1.6547	335	0.2360
41	-0.1049	100	0.0258	159	-0.2327	218	0.1077	277	0.6342	336	-0.0581
42	-0.2399	101	-1.3810	160	-0.5536	219	0.1530	278	-0.5469	337	1.2670
43	-0.5777	102	-1.3350	161	1.2804	220	0.5013	279	0.6166	338	1.1477
44	-0.7155	103	-0.0361	162	-0.1284	221	-0.6078	280	1.7145	339	-0.1930
45	-0.0108	104	-0.6272	163	-0.1422	222	-0.3348	281	0.6375	340	0.3587
46	-0.6540	105	-0.7097	164	-0.1100	223	-0.3387	282	1.7397	341	0.4549
47	-0.6378	106	-0.2294	165	0.8223	224	0.1866	283	1.3424	342	0.3460
48	-1.0374	107	-0.1218	166	1.5112	225	1.0108	284	-0.4330	343	-0.6457
49	0.2372	108	-0.4934	167	-0.0274	226	-0.2912	285	0.6503	344	0.5716
50	-0.0562	109	-0.2593	168	0.0045	227	-0.7527	286	-0.3880	345	1.3069
51	0.2941	110	-0.1202	169	0.1152	228	-0.3611	287	0.5209	346	0.7073
52	1.3165	111	-0.5801	170	-0.6383	229	1.1149	288	0.7122	347	0.5610
53	-0.8978	112	-0.5227	171	-1.0668	230	-0.8695	289	0.7623	348	1.0395
54	-1.3025	113	-0.5420	172	1.0682	231	1.4215	290	0.3974	349	0.7398
55	-1.1567	114	-0.8303	173	-0.3720	232	0.6909	291	0.3417	350	1.3815
56	-1.1882	115	-0.1775	174	-0.9574	233	-0.5813	292	0.4189	351	0.9587
57	-0.3237	116	-0.3829	175	-1.0916	234	-0.2731	293	1.9357		
58	-0.8435	117	-1.4837	176	0.9502	235	1.2780	294	0.1131		
59	-1.2945	118	-0.5500	177	-1.0396	236	1.5399	295	0.3122		

## EK-6: Kitapçık-1'den Elde Edilen Madde Karakteristik Eğrileri

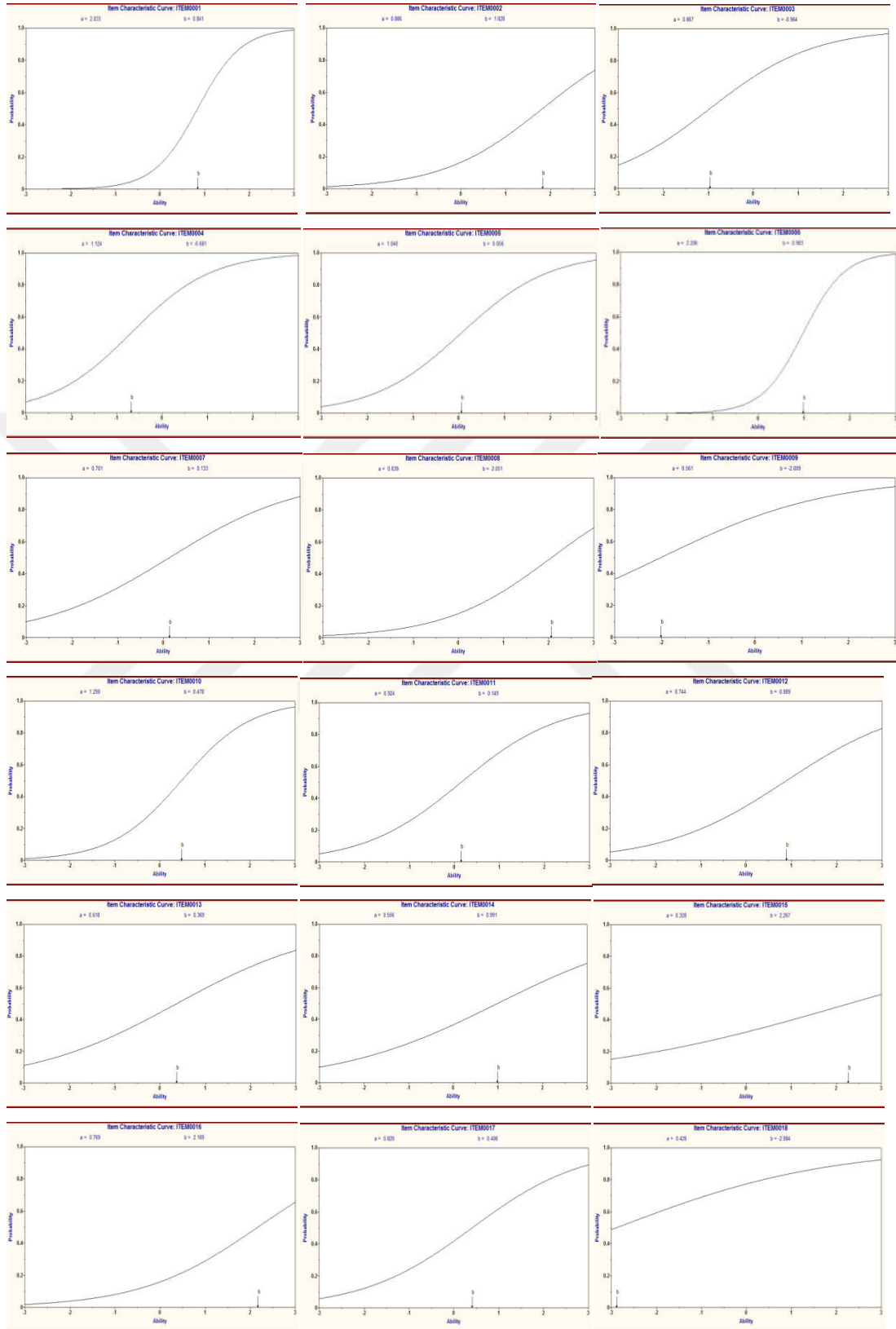


## EK-6 (DEVAMI): Kitapçık-1’den Elde Edilen Madde Karakteristik Eğrileri

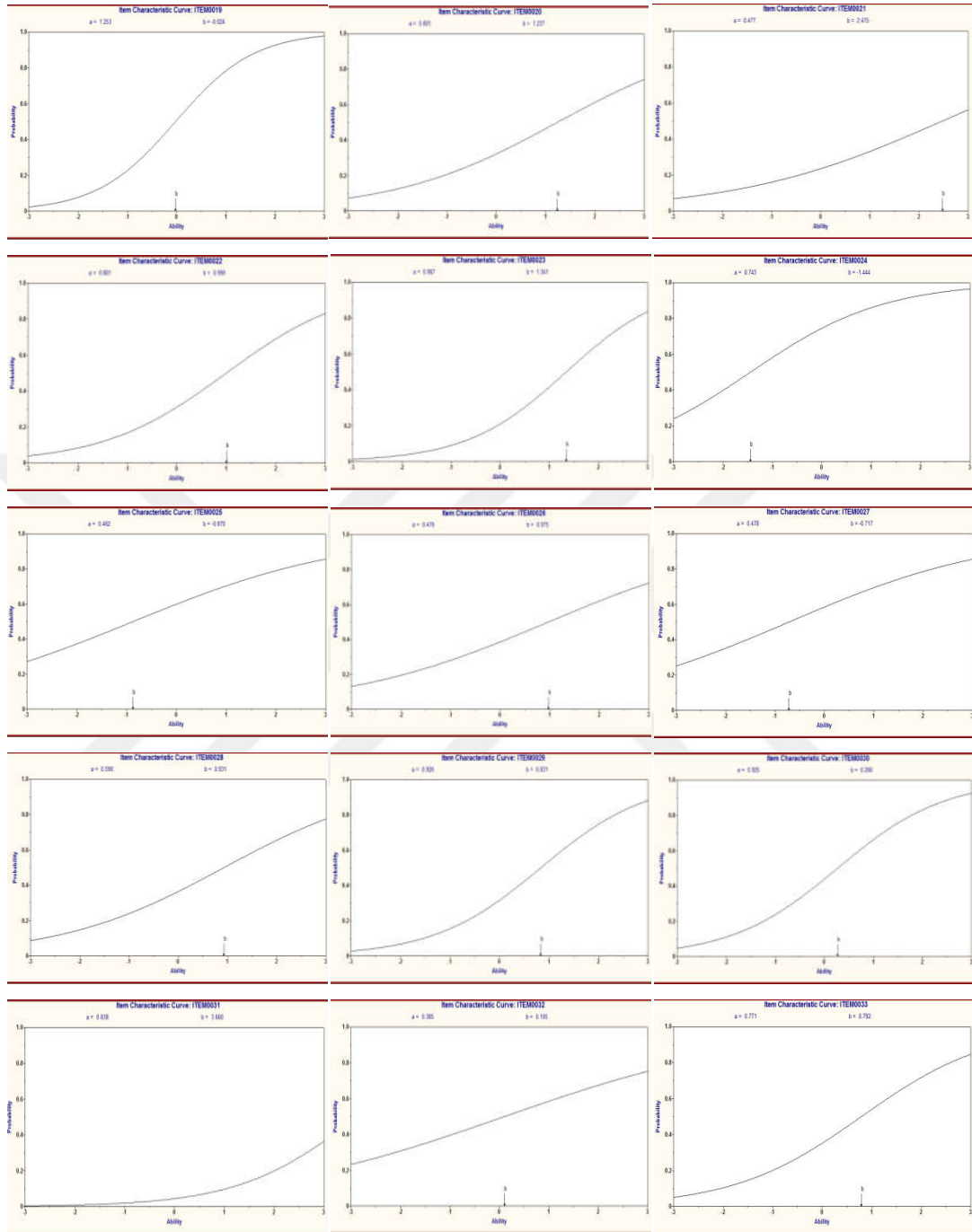




## EK-7: Kitapçık-8'den Elde Edilen Madde Karakteristik Eğrileri



## EK-7 (DEVAMI): Kitapçık-8'den Elde Edilen Madde Karakteristik Eğrileri



EK-8: Farklı Yöntemlerden Elde Edilen Dönüştürülmüş Parametreler

	OO		OS		SL		HA	
	a	b	a	b	a	b	a	b
M1	2.094	0.656	2.434	0.622	2.192	0.723	2.361	0.786
M2	0.913	1.615	1.061	1.446	0.955	1.638	1.029	1.636
M3	0.893	-1.096	1.038	-0.885	0.935	-0.951	1.007	-0.768
M4	1.158	-0.821	1.346	-0.649	1.212	-0.689	1.306	-0.524
M5	1.072	-0.106	1.246	-0.034	1.122	-0.006	1.208	0.110
M6	2.273	0.794	2.642	0.740	2.379	0.854	2.563	0.908
M7	0.722	-0.031	0.839	0.030	0.756	0.066	0.814	0.176
M8	0.864	1.831	1.005	1.633	0.905	1.845	0.975	1.828
M9	0.578	-2.110	0.672	-1.758	0.605	-1.920	0.652	-1.667
M10	1.328	0.304	1.544	0.319	1.390	0.386	1.497	0.474
M11	0.951	-0.015	1.106	0.044	0.996	0.081	1.073	0.190
M12	0.766	0.703	0.891	0.661	0.802	0.767	0.864	0.827
M13	0.636	0.198	0.740	0.227	0.666	0.285	0.718	0.379
M14	0.572	0.802	0.665	0.747	0.599	0.861	0.645	0.915
M15	0.338	2.040	0.393	1.812	0.354	2.044	0.381	2.013
M16	0.792	1.945	0.920	1.730	0.829	1.954	0.893	1.929
M17	0.844	0.234	0.981	0.259	0.884	0.319	0.952	0.412
M18	0.439	-2.960	0.510	-2.489	0.459	-2.732	0.495	-2.421
M19	1.291	-0.184	1.501	-0.101	1.351	-0.080	1.456	0.041
M20	0.619	1.041	0.720	0.953	0.648	1.090	0.698	1.127
M21	0.491	2.243	0.571	1.986	0.514	2.238	0.554	2.193
M22	0.825	0.810	0.959	0.754	0.864	0.869	0.930	0.922
M23	1.016	1.142	1.181	1.039	1.064	1.186	1.146	1.216
M24	0.765	-1.562	0.890	-1.287	0.801	-1.397	0.863	-1.181
M25	0.476	-1.004	0.553	-0.807	0.498	-0.864	0.537	-0.687
M26	0.490	0.786	0.570	0.734	0.513	0.847	0.553	0.901
M27	0.492	-0.856	0.572	-0.679	0.515	-0.722	0.555	-0.555
M28	0.616	0.743	0.716	0.696	0.645	0.805	0.695	0.863
M29	0.954	0.647	1.109	0.614	0.999	0.714	1.076	0.778
M30	0.948	0.100	1.102	0.143	0.992	0.191	1.069	0.293
M31	0.863	3.393	1.003	2.975	0.904	3.336	0.973	3.212
M32	0.396	-0.058	0.460	0.007	0.415	0.040	0.446	0.152
M33	0.795	0.608	0.924	0.580	0.832	0.677	0.896	0.743

## BİLDİRİM

Hazırladığım tezin tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin kağıt ve elektronik kopyalarının Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

- Tezimin tamamı her yerden erişime açılabilir.
- Tezim sadece Akdeniz Üniversitesi yerleşkelerinden erişime açılabilir.
- Tezimin 2 yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin tamamı her yerden erişime açılabilir.

06.02.2017

  
Burcu AKSEKİOĞLU

## ÖZGEÇMİŞ

### Kişisel Bilgiler

Adı Soyadı : Burcu AKSEKİOĞLU  
Doğum Yeri ve Tarihi : Karaman-Ermenek / 14.12.1990

### Eğitim Durumu

Lisans Öğrenimi : Akdeniz Üniversitesi, Eğitim Fakültesi, Sınıf Eğitimi Bölümü  
Yüksek Lisans Öğrenimi : Akdeniz Üniversitesi, Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı  
Bildiği Yabancı Diller : İngilizce  
Bilimsel Faaliyetler : 4th International Conference on Education (ICED-2015), San Petersburg / V. Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi (EPOD-2016), Antalya.

### İş Deneyimi

Çalıştığı Kurumlar : Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi

### İletişim

E-Posta Adresi : baksekioglu@mehmetakif.edu.tr

Tarih : 06.01.2017

## İNTİHAL RAPORU

Turnitin Orijinallik Raporu

TEST Burcu Burcu tarafından



EĞİTİMDE TESTLER (YÜKSEK LİSANS) den

- 02-Ara-2016 15:13 EET' de işleme konu
- NUMARA: 746298010
- Kelime Sayısı: 13238

Benzerlik Endeksi

%8

Kaynağa göre Benzerlik

İnternet Sources:

%6

Yayınlar:

%4

Öğrenci Ödevleri:

N/A

Doç. Dr. Bayram BİÇME