



**KLASİK TEST KURAMINA VE MADDE TEPKİ KURAMINA
DAYALI TEST EŞİTLEME YÖNTEMLERİNİN KARŞILAŞTIRMASI:
ULUSLARARASI ÖĞRENCİ DEĞERLENDİRME PROGRAMI (PISA)
2012 MATEMATİK TESTİ ÖRNEĞİ**

Ceren Mutluer

**DOKTORA TEZİ
EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME BİLİM DALI**

**GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

OCAK, 2021

TELİF HAKKI VE TEZ FOTOKOPİ İZİN FORMU

Bu tezin tüm hakları saklıdır. Kaynak göstermek koşuluyla tezin teslim tarihinden itibaren 12 ay sonra tezden fotokopi çekilebilir.

YAZARIN

Adı : Ceren

Soyadı : Mutluer

Bölümü : Eğitimde Ölçme ve Değerlendirme

İmza :

Teslim tarihi :

TEZİN

Türkçe Adı: Klasik Test Kuramına ve Madde Tepki Kuramına Dayalı Test Eşitleme Yöntemlerinin Karşılaştırması: Uluslararası Öğrenci Değerlendirme Programı (PISA) 2012 Matematik Testi Örneği

İngilizce Adı: Comparison of Test Equating Methods Based on Classical Test Theory And Item Response Theory: International Student Assessment Program (PISA) 2012 Mathematics Test Case

ETİK İLKELERE UYGUNLUK BEYANI

Tez yazma sürecinde bilimsel ve etik ilkelere uyduđumu, yararlandıđım tüm kaynakları kaynak gösterme ilkelerine uygun olarak kaynakçada belirttiđimi ve bu bölümler dıřındaki tüm ifadelerin řahsıma ait olduđunu beyan ederim.

Yazar Adı Soyadı: Ceren Mutluer

İmza:

JÜRİ ONAY SAYFASI

Ceren MUTLUER tarafından hazırlanan “Klasik Test Kuramına ve Madde Tepki Kuramına Dayalı Test Eşitleme Yöntemlerinin Karşılaştırması: Uluslararası Öğrenci Değerlendirme Programı (PISA) 2012 Matematik Testi Örneği” adlı tez çalışması aşağıdaki jüri tarafından oy birliği ile Gazi Üniversitesi Eğitim Bilimleri Anabilim Dalı Eğitimde Ölçme ve Değerlendirme Bilim Dalı’nda Doktora tezi olarak kabul edilmiştir.

Danışman: Prof. Dr. Mehtap Çakan

.....

(Eğitim Bilimleri Anabilim Dalı, Gazi Üniversitesi)

Başkan: Prof. Dr. Şeref Tan

.....

(Eğitim Bilimleri Anabilim Dalı, Gazi Üniversitesi)

Üye: Prof. Dr. Zekeriya Nartgün

.....

(Eğitim Bilimleri Anabilim Dalı, Bolu Abant İzzet Baysal Üniversitesi)

Üye: Prof. Dr. Nuri Doğan

.....

(Eğitim Bilimleri Anabilim Dalı, Hacettepe Üniversitesi)

Üye: Prof. Dr. Hakan Yavuz Atar

.....

(Eğitim Bilimleri Anabilim Dalı, Gazi Üniversitesi)

Tez Savunma Tarihi: 24/12/2020

Bu tezin Eğitim Bilimleri Anabilim Dalı’nda Doktora Tezi olması için şartları yerine getirdiğini onaylıyorum.

Prof. Dr. Yücel GELİŞLİ

Eğitim Bilimleri Enstitüsü Müdürü

.....

Sadık dostum Vanilya'ya ve Bolu'nun babası merhum yardımsever İzzet Baysal'a ithafen...

TEŞEKKÜR

Lisans öğrenciliğimde aldığım ‘Eğitimde Ölçme ve Değerlendirme’ dersi sayesinde bu alana yönelmemi sağlayan, doktora tez sürecimde birlikte çalışmaktan büyük mutluluk duyduğum, desteğini her zaman hissettiğim değerli danışmanım Prof. Dr. Mehtap ÇAKAN’a çok teşekkür ederim.

Tez yazım sürecinde bilgi ve görüşleri ile çalışmamı yönlendiren, tez izleme ve savunma jürimde yer alan Prof. Dr. Şeref TAN’a teşekkür ederim.

Çalıştığım kurum Bolu Abant İzzet Baysal Üniversitesi Eğitim Fakültesi’nde görev yapan yüksek lisans danışmanım ve Ana bilim dalı başkanım Prof. Dr. Zekeriya NARTGÜN’e teşekkür ederim.

Tez jürimde yer alan ve süreçte bilgileri ile çalışmamı yönlendiren değerli hocalarım Prof. Dr. Nuri DOĞAN ve Prof. Dr. Hakan Yavuz ATAR hocalarıma teşekkür ederim.

Doktora eğitimi sürecinde tanıdığım ve her zaman yanımda olan değerli dostlarım Arş. Gör. Tuba GÜNDÜZ ve Arş. Gör. Dr. Elif SEZER’e, hayatımın her döneminde manevi desteklerini hissettiğim dostlarım Şef Tuncay ALP, Büşra TEKİN, Olcay GEDİKLİ, Hayal ÖZÇİM, Dr. Zarife TAŞTAN, Âdem FIRAT, Nurdan DAMARLI’ya teşekkürü borç bilirim.

Hayatımın her noktasında varlıklarının bana destek verdiği, sevgilerini her zaman hissettiğim, benim mutluluğumla mutlu olan başta güzeller güzeli yeğenim Mira, canım annem, babam ve ablama çok teşekkür ederim.

**KLASİK TEST KURAMI'NA VE MADDE TEPKİ KURAMI'NA
DAYALI TEST EŞİTLEME YÖNTEMLERİNİN KARŞILAŞTIRMASI:
ULUSLARARASI ÖĞRENCİ DEĞERLENDİRME PROGRAMI (PISA)
2012 MATEMATİK TESTİ ÖRNEĞİ**

(Doktora Tezi)

**Ceren MUTLUER
GAZİ ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

OCAK,2021

ÖZ

Bu araştırmada Klasik Test Kuramı'ndaki ve Madde Tepki Kuramı'ndaki eşitleme yöntemleri kullanılarak en az eşitleme hata ile en uygun eşitleme yönteminin belirlenmesi amaçlanmıştır. Çalışmanın verileri için PISA 2012 Matematik testi, kitapçık 1 ve kitapçık3 puanları eşitlenmede kullanılmıştır. Araştırmaya Türkiye, Endonezya, Şangay/Çin ve Finlandiya verileri seçilmiştir. Araştırmada Klasik Test Kuramı doğrusal eşitleme için Tucker ($w_1=1$, $w_1=0,5$), Levine gözlenen puan ($w_1=1$, $w_1=0,5$), Levine gerçek puan, klasik konjenerik ve Braun-Holland kullanılmıştır. Belirtilen bu yöntemlerden Tucker eşitlemenin en az hata ile eşitleme yaptığı, Levine gerçek puan eşitleme yönteminin ise en fazla hata ile

eşitleme yaptığı bulunmuştur. Klasik Test Kuramında yer alan eşit yüzdelikli eşitleme yöntemlerinden önce düzgünleştirme yapılmıştır. Ön düzgünleştirmede C 6 polinom derecesi ve beta4 binomial fonksiyonu, son düzgünleştirme için ise kübik spline S 0,05 derecesi analitik ve grafiksel süreç sonunda en uygun düzgünleştirme olarak seçilmiştir. Hata bant değeri içerisindeki dağılımları ve hata miktarlarına göre kendi aralarında kıyaslandığında beta4 binomial fonksiyonun en iyi, C 6 polinom derecesinin ise en fazla hataya sahip olarak düzgünleştirme yöntemi olduğu bulunmuştur. Eşit yüzdelikli eşitleme sürecinde frekans kestirim ($w_1=1$, $w_1=0,5$) eşit yüzdelikli eşitleme yöntemi kullanılmıştır. Eşit yüzdelikli eşitleme için en az hata $w_1=0,5$ frekans kestirim yönteminden, en fazla hata $w_1=1$ frekans kestirim eşitleme yönteminden elde edilmiştir. Klasik Test Kuramına ilişkin doğrusal ve eşit yüzdelikli eşitleme yöntemleri kıyaslandığında eşit yüzdelikli eşitlemede daha az hata ile eşitlenmiş puanların elde edildiği görülmüştür. Madde Tepki Kuramına dayalı eşitleme sürecine kalibrasyon yaparak başlanmıştır. Madde karakteristik eğri kalibrasyon yöntemlerinden Stocking-Lord yönteminin daha iyi sonuç sağladığı görülmüştür. Madde Tepki Kuramı gözlenen ve gerçek puan eşitleme yöntemlerinden elde edilen eşitlenmiş puanlar karşılaştırıldığında Madde Tepki Kuramı gerçek puan eşitleme, daha az hata barındırdığı sonucuna ulaşılmıştır. Kuramlardaki eşitleme yöntemleri kıyaslandığında Madde Tepki Kuramı yöntemlerinin daha az hata ile eşitleme yaptığı bulunmuştur. Yapılacak eşitleme çalışmalarında Madde Tepki Kuramı gerçek puana dayalı eşitleme yapılması önerilmiştir.

Anahtar Kelimeler : Test Eşitleme, Klasik Test Kuramına Dayalı Eşitleme Yöntemleri, Madde Tepki Kuramına Dayalı Eşitleme Yöntemleri, Ortak Maddeli Eşdeğer Olmayan Grup Deseni

Sayfa Adedi :

Danışman : Prof. Dr. Mehtap Çakan

**COMPARISON OF TEST EQUATING METHODS BASED ON
CLASSICAL TEST THEORY AND ITEM RESPONSE THEORY:
INTERNATIONAL STUDENT ASSESSMENT PROGRAM (PISA) 2012
MATHEMATICS TEST CASE
(Doctoral Dissertation)**

**Ceren MUTLUER
GAZI UNIVERSITY
GRADUATE SCHOOL OF EDUCATIONAL SCIENCES
JANUARY, 2021**

ABSTRACT

In this study, it was aimed to determine the most appropriate equating method with the least equating error by using the equating methods in Classical Test Theory and Item Response Theory. The scores of the booklet 1 and the booklet 3 in the PISA 2012 Mathematics test were used to be equated for the data of the research. The data of Turkey, Indonesia, Shanghai / China and Finland were selected for the research. In the study, Tucker ($w_1 = 1$, $w_2 = 0.5$), Levine observed score ($w_1 = 1$, $w_2 = 0.5$), Levine true score, classical congeneric and Braun-Holland methods were used. Among these methods, Tucker equalization was found to equate with the fewer error, and Levine true score equating method equated with the highest error. The smoothing procedure was completed before the equipercetile equating methods in the Classical Test Theory. The C 6 polynomial degree and beta4 binomial function in the pre-

smoothing, and the cubic spline S 0,05 degree for the final smoothing were selected as the most appropriate smoothing at the end of the analytical and graphical process. When compared to each other according to their distribution within the error band value and the amount of error, it was found that the beta4 binomial function was the best smoothing method, while the C 6 polynomial degree had the highest error. Frequency estimation ($w_1 = 1$, $w_1 = 0.5$) equating method were used in the equipercentile equating process. For equipercentile equating, the least error was obtained from the $w_1=0,5$ frequency estimation method, and the highest error was obtained from the $w_1=1$ frequency estimation equipercentile equating method. When the linear and equipercentile equating. Methods related to Classical Test Theory were compared, it was seen that scores equated with less errors were obtained in equipercentile equating. The equating process based on Item Response Theory started with making calibration. It was seen that the Stocking-Lord method provided better results among item characteristic curve calibration methods. When the equated scores obtained from the Item Response Theory observed and real point equating methods were compared, it was concluded that the Item Response Theory actual score equating contained less errors. When the equating methods in the theories were compared, it was found that the Item Response Theory methods made equalization with less error. It is recommended to make equating based on Item Response Theory true score in equating studies.

Key Words : Test Equating, Test Equating Methods Based on Classical Test Theory, Test Equating Methods Based on Item Response Theory, Common Item Non-Equivalent Groups Design,

Page Number :

Supervisor : Prof. Dr. Mehtap Çakan

İÇİNDEKİLER

ÖZ.....	vii
ABSTRACT.....	ix
İÇİNDEKİLER.....	xi
TABLolar LİSTESİ.....	xv
ŞEKİLLER LİSTESİ.....	xviii
BÖLÜM I.....	1
GİRİŞ.....	1
Problem Durumu.....	1
Test Eşitleme.....	4
Test Eşitlemenin Koşulları.....	7
Test Eşitleme Aşamaları.....	9
Test Eşitleme Desenleri.....	9
Klasik Test Kuramına Dayalı Eşitleme Yöntemleri.....	15
Ortalama Eşitleme Yöntemi.....	16
Doğrusal (Lineer) Eşitleme.....	17
<i>Ortak Maddeli Eşdeğer Olmayan Gruplar- Doğrusal Eşitleme Yöntemleri</i>	18
<i>Tucker Doğrusal Eşitleme Yöntemi.....</i>	20
<i>Levine Gözlenen Puan Eşitleme Yöntemi.....</i>	21
<i>Levine Gerçek Puan Eşitleme.....</i>	21
<i>Klasik Konjenerik Modele Dayalı Doğrusal Eşitleme.....</i>	22
<i>Braun-Holland Yöntemi.....</i>	23
Eşit Yüzdelikli Eşitleme (EYE).....	24
<i>Ön Düzgünleştirme (Pre-Smoothing).....</i>	27
<i>1-Log-linear yöntem:.....</i>	28
<i>2-Beta 4 yöntemi:.....</i>	28

<i>Son Düzgünleştirme (Post- Smoothing).....</i>	<i>28</i>
<i>Ortak Maddeli Eşdeğer Olmayan Gruplar – Eşit Yüzdelikli Eşitleme Yöntemleri.....</i>	<i>29</i>
<i>Frekans Tahmin Yöntemi.....</i>	<i>29</i>
Madde Tepki Kuramı ve Madde Tepki Kuramına İlişkin Test Eşitleme Yöntemleri	30
Madde Tepki Kuramı Ölçeğinin Dönüştürümü.....	35
Test Eşitlemede Hata	44
Araştırmanın Amacı	44
Problem İfadesi	45
Araştırma Soruları.....	45
Araştırmanın Önemi.....	45
Varsayımlar	45
Sınırlılıklar.....	47
BÖLÜM II	48
İLGİLİ ARAŞTIRMALAR	48
Klasik Test Kuramı Temelli Eşitleme Yöntemlerine İlişkin Çalışmalar	48
Madde Tepki Kuramına Yönelik Eşitleme Yöntemlerine İlişkin Çalışmalar.....	55
Klasik Test Kuramı ve Madde Tepki Kuramının Karşılaştırıldığı Eşitleme Yöntemlerine İlişkin Çalışmalar	60
BÖLÜM III.....	64
YÖNTEM.....	64
Araştırmanın Deseni.....	64
Evren ve Örneklem	64
Verilerin Toplanması ve Düzenlenmesi	66
PISA 2012 Hakkında Kısa Bilgi.....	66
Eşitleme Deseni (Veri Toplama Deseni).....	70
Verilerin Analizi ve Test Eşitleme Varsayımlarının Test Edilmesi.....	70
Verilerin Analizi	87
BÖLÜM IV	90
BULGULAR VE YORUMLAR	90
Birinci Araştırma Sorusuna Yönelik Bulgular ve Yorumları:	90
İkinci Araştırma Sorusuna Yönelik Bulgular ve Yorumları:.....	108
Üçüncü Araştırma Sorusuna Yönelik Bulgular ve Yorumları:.....	127

Dördüncü Araştırma Sorusuna Yönelik Bulgular ve Yorumları:	134
BÖLÜM V.....	137
TARTIŞMA VE SONUÇ	137
Birinci Alt Problemine İlişkin Sonuçlar	137
İkinci Alt Problemine İlişkin Sonuçlar	139
Üçüncü Alt Problemine İlişkin Sonuçlar	141
Dördüncü Alt Problemine İlişkin Sonuçlar	143
BÖLÜM VI.....	145
ÖNERİLER	145
Uygulayıcılara Yönelik Öneriler	145
Araştırmacılara Yönelik Öneriler	146
KAYNAKLAR	148
EKLER.....	158
EK 1. Kitapçıklardaki Madde Parametrelerinin Dağılımları	158
EK 2. FACTOR Programındaki Kitapçık 1'e İlişkin Özdeğerlerin Çıktısı	159
EK 3. FACTOR Programındaki Kitapçık 3'e İlişkin Özdeğerlerin Çıktısı	160
EK 4. Kitapçık 1'deki Madde Çiftlerinin Yerel Bağımsızlık Sonuçları	161
EK 5. Kitapçık 3' teki Madde Çiftlerinin Yerel Bağımsızlık Sonuçları	166
EK 6. Kitapçık 1 Maddelerine İlişkin Madde Karakteristik Eğrileri.....	171
EK 7. Kitapçık 3 Maddelerine İlişkin Madde Karakteristik Eğrileri.....	173
EK 8. MTK İçin Model Veri Uyum Sonucu	175
EK 9. Tucker Doğrusal Yöntemine İlişkin Equating Error 2.0Program Çıktısı.....	176
EK 10. Levine Gözlenen Puan İçin Eşitleme Hata Değerleri.....	180
EK 11. Levine Gerçek Puan Eşitlemenin Hata Değeri.....	184
EK 12. Klasik Konjenerik Eşitleme Hatası	186
EK 13. Braun-Holland Yöntemi için Eşitleme Hatası	188
EK 14. Ön Düzgünleştirme İçin Çeşitli C Ponimomial Dereceler İçin Eşitlenmiş Puanlar	190
EK 15-.Polynomial Log-Linear (C) ve Beta 4 ile Elde Edilen Grafikler.....	191
EK 16. C 6 ve Beta4 için Hata Bant Grafiği Oluşturmak için Hesaplanan İstatistikler	193
EK 17. Farklı Kübik Spline Derecelerinin Hata Bant Değerleri Arasındaki Dağılımları	195
EK 18. Ön Düzgünleştirme İçin C6 Polinom Fonksiyonun Eşitleme Hatası	198
EK 19. Ön Düzgünleştirme Yöntemlerinden Beta4 Fonksiyona İlişkin Eşitleme Hatası	200

EK 20. Son Düzgünleştirme Yöntemlerinden S 0,05 Fonksiyona İlişkin Eşitleme Hatası	202
EK 21. EYE Yöntemlerden Frekans Kestirim için İstatistiksel Süreç	204
EK 22. Frekans Kestirim Yöntemi için Eşitleme Hata Değerleri.....	208
EK 23. Kalibrasyon Yöntemlerine İlişkin ST 2.0 Program Çıktısı.....	210
EK 24. MTK Gerçek ve Gözlenen Puan Eşitleme Süreci	211



TABLÖLAR LİSTESİ

Tablo 1 Ülkelere Göre Kitapçıkları Alan Öğrenci Dağılımları	66
Tablo 2 Kitapçıklarda Yer Alan Maddeler ve Kodları	69
Tablo 3 Kitapçık 1 ve Kitapçık 3 İçin Betimsel İstatistikler	71
Tablo 4 Kitapçık 1 ve Kitapçık 3 İçin Ortalamalar Arası Farka İlişkin İlişkisiz Örneklemeler <i>t</i> Testi Sonucu	72
Tablo 5 Kitapçık 1 ve Kitapçık 3 için Ortak Madde Puanları ve Testin Bütününe Arasındaki Korelasyon	73
Tablo 6 Kitapçık 1 ve 3'e İlişkin Kestirilen Parametre Kestirimleri.....	74
Tablo7 Kitapçıklara Ait Ayırteçilik, Madde Güçlüğü ve Şans Parametrelerinin Ortalamları	75
Tablo 8 Kitapçıklardaki <i>a</i> Parametresine Yönelik İlişkisiz Örneklemeler <i>t</i> Testi Sonuçları.	75
Tablo 9 Kitapçıklardaki <i>b</i> Parametresine Yönelik İlişkisiz Örneklemeler <i>t</i> Testi Sonuçları.	76
Tablo 10 Kitapçıklardaki <i>c</i> Parametresine Yönelik İlişkisiz Örneklemeler <i>t</i> Testi Sonuçları	76
Tablo 11 Kitapçık 1 ve Kitapçık 3'e İlişkin Güvenirlik Katsayıları	77
Tablo 12 Kitapçık 1 ve Kitapçık 3'e İlişkin Fisher Z Katsayıları.....	77
Tablo 13 Kitapçık 1 ve 3'e İlişkin KMO ve Barlett Testi Sonuçları	78
Tablo 14 Model uyum İndekslerinin Kriterleri ve Kabulü İçin Kesme Noktaları	79
Tablo 15 Kitapçıklara İlişkin Model-Veri Uyumu İndeks Değerleri	81

Tablo 16 <i>Model-Veri Uyum Karşılaştırması</i>	87
Tablo 17 <i>Form X'i Y'ye Eşitlemede Kullanılan Veri İçin Doğrudan Gözlemlenen İstatistikler</i>	90
Tablo 18 <i>Tucker İç Ortak Eşitleme Yöntemi Kullanılarak Elde Edilen Eşitlenmiş Puanlar ve Fark Değerleri</i>	92
Tablo 19 <i>Levine Gözlenen Eşitleme Yöntemi Kullanılarak Elde Edilen Eşitlenmiş Puanlar ve Fark Değerleri</i>	95
Tablo 20 <i>Levine Gerçek Puan Eşitleme Yöntemi Kullanılarak Elde Edilen Eşitlenmiş Puanlar ve Fark Değerleri</i>	98
Tablo 21 <i>Klasik Konjenerik Eşitleme Yöntemi Kullanılarak Elde Edilen Eşitlenmiş Puanlar ve Fark Değerleri</i>	100
Tablo 22 <i>Braun-Holland Yönteminden Elde Edilen Eşitlenmiş Puanlar...</i>	102
Tablo 23 <i>KTK'daki Doğrusal Eşitleme Yöntemi Kullanılırken Dikkate Alınan Parametre Değerleri</i>	103
Tablo 24 <i>Doğrusal Eşitleme Yöntemlerinden Elde Edilen WMSE Değerleri</i>	107
Tablo 25 <i>Momentler ve Ön Düzgünleştirmenin Uyum İstatistikleri</i>	109
Tablo 26 <i>Ön Düzgünleştirmeye Dayalı Eşit Yüzdelikli Eşitleme Yönteminden Elde Edilen Değerler</i>	111
Tablo 27 <i>Ön-Düzgünleştirme için Ham Puan Momentleri</i>	113
Tablo 28 <i>Son-Düzgünleştirme için Ham Puan Dönüşümleri</i>	114
Tablo 29 <i>Son-Düzgünleştirme için Ham Puan Momentleri</i>	117
Tablo 30 <i>Düzgünleştirme Yöntemlerinden Elde Edilen Ham Puan Momentleri</i>	118
Tablo 31 <i>Sentetik Evrenin Farklı Ağırlık Değerlerine Göre Frekans Kestirim Yöntem Sonuçları</i>	121
Tablo 32 <i>KTK'daki EYE Kullanılırken Dikkate Alınan Parametre Değerleri</i>	123
Tablo 33 <i>KTK EYE Yöntemlerinin Hata Katsayıları</i>	125
Tablo 34 <i>KTK Bünyesindeki Doğrusal ve EYE Yöntemlerinin Hata Katsayıları</i>	126

Tablo 35 Kalibrasyon Yöntemlerinden Elde Edilen Dönüşüm Katsayıları ve Dönüşüm Sabitleri.....	127
Tablo 36 Kalibrasyon Yöntemleri için Hesaplanan Hata Katsayıları.....	129
Tablo 37 MTK Gerçek ve Gözlenen Puan Eşitleme Yöntemlerinden Elde Edilmiş Puanlar	130
Tablo 38 MTK Eşitleme Yöntemleri ve Bu Yöntemlerin Hata Katsayıları	134
Tablo 39 KTK ve MTK'ya İlişkin Eşitleme Yöntemlerinin Hata Katsayıları	135



ŞEKİLLER LİSTESİ

<i>Şekil 1</i> Test eşitlemede kullanılan formların güçlük düzeylerine göre sınıflandırılması.....	5
<i>Şekil 2</i> Test eşitleme desenlerinin sınıflandırılması.....	10
<i>Şekil 3</i> Rastgele grup deseni	11
<i>Şekil 4</i> Tek grup deseni	11
<i>Şekil 5</i> Dengelenmiş tek grup deseni	12
<i>Şekil 6</i> Ortak maddeli eşdeğer olmayan gruplar deseni	14
<i>Şekil 7</i> Klasik Test Kuramına dayalı eşitleme yöntemleri	16
<i>Şekil 8</i> Eşit yüzdelikli eşitlemenin grafiksel gösterimi	25
<i>Şekil 9</i> Düzgünleştirme yöntemleri	27
<i>Şekil 10</i> Ölçek dönüştürme yöntemleri	37
<i>Şekil 11</i> Madde Tepki Kuramına dayalı test eşitleme yöntemleri	43
<i>Şekil 12</i> Yıllara göre PISA’da odaklanılan alanlar	67
<i>Şekil 13</i> Kitapçık 1 ve Kitapçık 3 için ölçme modelleri	80
<i>Şekil 14</i> Kitapçık 1’deki maddelere ilişkin özdeğer-saçılım grafiği	83
<i>Şekil 15</i> Kitapçık 1’deki maddelere ilişkin özdeğer-saçılım grafiği	84
<i>Şekil 16</i> PISA 2012 Ham puanların Tucker iç ortak eşitleme yöntemiyle eşitlenmesi	94
<i>Şekil 17</i> PISA 2012 Ham puanların Levine gözlenen eşitleme yöntemiyle eşitlenmesi	96
<i>Şekil 18</i> PISA 2012 Ham puanların Levine gerçek eşitleme yöntemiyle eşitlenmesi	99
<i>Şekil 19</i> PISA 2012 Ham puanların konjenerik eşitleme yöntemiyle eşitlenmesi	101

<i>Şekil 20</i> Braun-Holland yönteminden elde edilen EYE puanları.....	103
<i>Şekil 21</i> Doğrusal eşitleme yöntemlerine göre eşitlenmiş puanların dağılımı.....	105
<i>Şekil 22</i> Doğrusal eşitleme yöntemlerine göre fark puanların dağılımı.....	106
<i>Şekil 23</i> C=6 Polinomial derece için ön-düzgünleştirme ve ham puan dağılımı	110
<i>Şekil 24</i> Standart hata bantına göre ön-düzgünleştirme yöntemlerinden elde edilen puanlar	112
<i>Şekil 25</i> Standart hata bantına göre S 0,05 için son-düzgünleştirme yönteminden elde edilen puanlar.....	116
<i>Şekil 26</i> Standart hata bantına göre düzgünleştirme yöntemleri fark değerlerinin dağılımları	119
<i>Şekil 27</i> Farklı sentetik evren ağırlıklarına göre frekans kestirim yönteminden elde edilen EYE puanları.....	122
<i>Şekil 28</i> EYE Yönteminden elde edilen puanların dağılımı	124
<i>Şekil 29</i> EYE Yönteminden elde edilen fark puanların dağılımı.....	124
<i>Şekil 30</i> MTK gerçek puan eşitleme yönteminden elde edilen eşitlenmiş puanlar.....	131
<i>Şekil 31</i> MTK gözlenen puan eşitleme yönteminden elde edilen eşitlenmiş puanlar.....	132
<i>Şekil 32</i> MTK gözlenen ve gerçek puan eşitleme yöntemlerindeki fark puanlarının dağılımları	133

BÖLÜM I

GİRİŞ

Bu bölümde problem durumuna, araştırmanın amacına, önemine, problem ifadesine, alt problem cümlelerine, sınırlılıklarına ve kuramsal temeline yer verilmiştir.

Problem Durumu

Bir ülkenin gelişmesi ülkesinin eğitime verdiği önemle ilişkilidir. Gelişmekte olan ülkeler arasındaki Türkiye, bilimin ilerlemesi için ulusal, ekonomik, sosyal, kültürel değerlerini, dünyanın evrensel değerlerini dikkate alarak eğitim-öğretim aşamalarında reform niteliği taşıyan değişimler yapmıştır. Bu değişim sürecinde gelişme katetmiştir. Bu değişimin asıl felsefesi öğrenciyi, bilgiyi ezberleyen, gerektiğinde ezberlediği bilgiyi aktaran, süreçte bilgiyi kâğıda aktaran olmaktan çıkartıp, süreçte bilgiyi kendisi kazanan, kazandığı bilgiyi sınayan ve bilgiyi farklı durumlarda kullanan etkin kişi yapmaktır. Süreçte eğitim sistemindeki öğelerdeki değişikliklerin etkileri testlerle, anketlerle, ölçeklerle ve gerekli envanterlerle ölçülmektedir.

Ulusal boyutta öğrencilerin başarı düzeylerinin tespiti için Seviye Belirleme Sınavı (SBS), Lisans Yerleştirme Sınavı (LYS), Yükseköğretime Geçiş Sınavı (YGS), Kamu Personeli Seçme Sınavı (KPSS), Yabancı Dil Sınavı (YDS), Akademik Personel ve Lisansüstü Eğitimi

Giriş Sınavı (ALES) gibi sınavlar yapılmaktadır. Süreç içinde eğitim programlarında ve diğer öğelerdeki reform hareketlerin yansımaları olarak, Türkiye'nin eğitim politikasının diğer ülkeler arasındaki başarı sırasındaki yeri de görülmek istenmiştir. Bunun için Uluslararası Fen ve Matematik Çalışması (TIMSS), Okuma Becerilerinde Gelişim (PIRLS) ve uygulaması Ekonomik İşbirliği ve Kalkınma Örgütü (OECD) tarafından yapılan Uluslararası Öğrenci Değerlendirme Programı (PISA) sınavlarına ülkelerimizden katılım gerçekleştirilmektedir.

Yapılan sınavların sonuçlarının, kişilerin performansı ve bir ülkenin genel eğitim düzeyi hakkında karar verilmesi, iş yaşamına geçişin sağlanması, bir üst kuruma yerleşmenin sağlanması gibi durumlarda kullanıldığı düşünüldüğünde sınavların, katılan bireylerin yeteneklerini daha az hatalı, daha doğru ve objektif bir şekilde ölçmesi beklenmektedir. Belirtilen objektif ölçme için ölçme sürecine öncelikle testlerin uygulama koşullarını standart koşullara dönüştürmekle başlanmalıdır.

Standartlaştırılmış ortamlarda elde edilen test puanlarının, testin ölçmek istediği beceri veya yeteneklerin adil ve eşitlenebilir bir şekilde değerlendirilmesi şarttır. Nitekim adil ve eşitlenebilir bir psikolojik veya eğitimsel değerlendirme aracı sağlanması, standartlaştırılmış testin varlığının ana nedenlerinden biridir (Cook ve Eignor, 1991). Merkezi sınavlar ya da geniş ölçekli testler, standartlaştırma amacı taşımasıyla test edilmeye çalışılarak sınavların güvenilirliği yükseltilmeye çalışılmaktadır. Crocker ve Algina'nın (1986) ve Lord'un (1968) belirttiği gibi yapılan uygulamalarda ölçme sonuçlarına her zaman birtakım hatalar karışmaktadır. Bu nedenle uygulama süreci, testin kendisi ya da puanlayıcılar gibi dinamikler düşünüldüğünde elde edilen ölçümlerde hata payının olması kaçınılmazdır. Bu durum, standart koşulların oluşmasına engel olmaktadır. İşte bu sebepten merkezi ya da geniş ölçekli sınavlarda aynı soruların tekrar-tekrar sorulması, kişide beklenen performansın ölçülmesi yerine kişinin daha önce çözdüğü bir soruyu tekrar hatırlamasına sebep olacaktır. Bu durumu önlemek amacıyla aynı özelliği ölçecek olan testin farklı formlarının geliştirilmesi uygun olacaktır.

Ülkemizde yapılan KPSS, YDS, KPDS, ÜDS, ALES, YÖKDİL gibi sınavlarda elde edilen sonuçlar birkaç yıl için geçerliğini korumaktadır. Bu sınav sonuçlarının çeşitli kurumlara yerleşme, görevde yükselme gibi durumlar için önemli bir katkısı vardır. Bu uygulamalarda hatırlamayla işaretlemenin önüne geçilip, test gizliliği sağlamada aynı özellikleri ölçtüğü iddia edilen farklı formlar kullanılmaktadır. Testin güvenilirliği farklı formlar uygulanarak sağlanmış olsa da elde edilen sonuçların eşitliği hakkında birtakım şüphelerin oluşması kaçınılmazdır. Formların paralel test mantığıyla hazırlanmasına rağmen testi alan grubun yeteneklerinin farklı olması nedeniyle formlardan alınan aynı puan, farklı formlar için aynı yetenek düzeyinin göstergesi olmayabilir. Aynı içeriği baz alarak aynı amaç için hazırlanmış test formları paralellik iddiasıyla hazırlanmış olsa da testin güvenilirliği, madde güclüğü, testin standart sapması gibi indeksler her uygulamada farklı değer almaktadır. Haliyle farklı formu alan kişiler arasında testin zorluk ve kolaylık durumuna göre puanlar değişecektir. Aynı birey farklı formlardan aynı puanı alsa bile performansı için aynı yorum yapılması doğru olmayacaktır. İşte bu yüzden formlardan alınan puanların karşılaştırılması gerekmektedir. Farklı zamanlarda, farklı testlerle yapılan uygulamalarda gerek aynı kişinin puanları gerekse de farklı kişilerin puanları doğrudan karşılaştırılamaz. Dorans ve Holland'nın da (2000) belirttiği gibi farklı durumlarda, farklı yöntemlerle yapılan testlerden elde edilen ölçümlerin karşılaştırılması tüm bilimlerin temel ön koşulu olmuştur. Bu bağlamda aynı içeriğe, aynı performansa yönelik hazırlanmış test formlarından elde edilen puanların birbirlerine dönüşümünü sağlayan istatistiksel bir sürece ihtiyaç vardır.

Farklı formlardan elde edilen sonuçların birbiri yerine kullanılabilmesi için öncelikle bu puanlar ortak ölçek üzerinden oluşturmalı ya da daha spesifik olarak verilen iki testin puanları arasında karşılıklı ilişki kurulmalıdır. İşte bu ilişki test eşitleme ile gerçekleştirilebilir (Zhu, 1998).

Test Eşitleme

Ulaşılabilen literatür tarandığında ‘test eşitleme’ kavramına yönelik çok sayıda tanımla karşılaşılmiştir. Angoff (1987)’e göre test eşitleme, bir formun birim sistemini diğer bir formun birim sistemine dönüştürmektir. Bir başka tanımla, iki ya da daha fazla testteki puanlar arasındaki ilişkileri bir istatistiksel yöntem ile kurma ya da basit bir şekilde, ortak ölçek üzerinde bu test sonuçlarının yer almasına “test eşitleme” denir (Hambleton ve Swaminathan, 2013). Kolen ve Brennan (2014) ise test eşitleme kavramı, benzer içerik ve benzer güçlük düzeyindeki test formları arasındaki farklılıkları düzenleyerek bu formlardan elde edilen puanların birbiri yerine kullanılmasını sağlayan istatistiksel bir süreç olarak tanımlanmıştır. Eşitleme ile birbirinin yerine kullanılan formlar için mümkün olduğunca güçlük farklılıklarının düzeltilmesi amaçlanmaktadır (Brennan ve Kolen, 1987). Bu nedenle eşitleme süreci, testlerin alternatif formlarının olduğu ve farklı formlardan elde edilen puanların karşılaştırıldığı durumlarda kullanılmaktadır (Livingston, 2014). Test eşitlemeyle, aynı amaç için hazırlanmış test formlarının güçlüklerinden kaynaklanan farklılıkların giderilmesi sonucunda, öğrencilerin daha kolay veya daha zor bir test formunu almış olmasından dolayı bir avantaj veya dezavantaj sağlaması önlenmiş olunur (Tan, 2015). Özellikle bağıl değerlendirmenin esas alındığı ve sıralamalarının bir kuruma yerleşme, görevde yükselme gibi önemi düşünüldüğünde eşitleme sayesinde adil bir ölçme işlemi gerçekleştirilmiş olacaktır.

Test puanlarının eşitlenme nedenleri çeşitlilik göstermektedir. Bu nedenler genellikle üç kategoriye ayrılır (Barnard, 1996; Dorans, Moses ve Eignor, 2010):

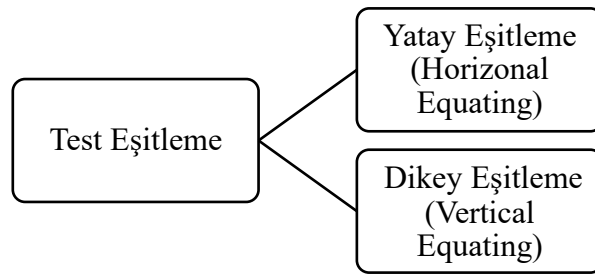
1- Eşitlemenin en genel nedeni, aynı testin tekrar uygulanması yerine paralel testlerden elde edilen puanların doğruluğunu değerlendirmek ve test geliştirmektir. Uygulama etkilerini devre dışı bırakmak ya da bir testin doğruluğunu sağlamak için çeşitli formlar sıklıkla gerekmektedir. Bu alınan puanlardan yola çıkılarak iki test formu arasında ilişki kurmaktır.

2- Test formları için birbirleri ile deęişebilir olmasını saęlamaktır. Çeşitli testler uygulamada benzer deęişkenleri ölçmek için kullanılırken, bu testlerden alınan puanlar farklı ölçeklere yerleştirildiğinden genellikle deęiştirilemezler. Katılımcıların performanslarını ya da testler arasında eşdeğer kriterler oluşturmak için genellikle test puanlarının aynı ölçek üzerine dönüştürülmesi gerekmektedir. Bu sayede farklı test formlarını alan bireyler arasında oluşabilecek yanlılığın önüne geçilebilmektedir.

3- Test eşitlemenin üçüncü nedeni ise test formları için sürekliliğin saęlanmasıdır. Uygulamada çeşitli testler farklı yetenek aşamalarındaki ölçülen gelişimde ya da bir yetenek ya da özelliğin deęişiminde kullanılmaktadır. Bu durumda, testin zorluğu amaçsal olarak farklılaşarak hazırlanmaktadır. Özellikle bu eşitleme çalışmasında kişilerin gelişimsel süreçleri izlenmek istenmektedir.

Yatay ve Dikey Eşitleme

Eşitleme yöntemleri genel başlık olarak iki kategoriye ayrılmaktadır. Bu sınıflandırma Şekil-1'de belirtilmiştir (Cook ve Eignor, 1983; Cook ve Eignor, 1989; Hambleton ve Swaminathan, 2013; Kolen, 1988).



Şekil 1 Test eşitlemede kullanılan formların güçlük düzeylerine göre sınıflandırılması

Aynı içerikteki bir testin farklı formlarında test ölçümlerinin güvenilirliğini saęlamak gerektiğinde yatay eşitleme kullanılır. Bu testin formlarının tüm maddelerce özdeş olmamasına rağmen içerik ve madde güçlükleri bakımından paralel olduğu düşünülmektedir (Cook ve Eignor, 1991). Buna ek olarak, testin yöneltileceği kişilerin yetenek dağılımlarının

oldukça benzer olacağı beklenmektedir. Yetenek dağılımları arasında büyük farklılıklar olduğunda geleneksel eşitleme yöntemleri (ortalama, doğrusal, eşit yüzdelikli) uygun olmayacaktır (Kolen, 1988; Woldbeck, 1998).

Yatay eşitleme için Türkiye’de yapılan merkezi sınavlardan ALES, YDS, YÖKDİL gibi uygulamalarda benzer güçlük düzeyleri oluşturularak hazırlanmış farklı formlar uygulanmaktadır. Ülkelerin performanslarını karşılaştırmak için ise geniş ölçekli TIMSS, PISA ve PIRLS gibi sınavlarda, aynı zaman diliminde, birden fazla kitapçık ile aynı içerik ve benzer güçlük düzeyindeki maddeler kullanılmaktadır. Bu kitapçıklarda benzer güçlük düzeylerinde hazırlanmış farklı maddeler dışında aynı maddeler de yer almaktadır. Bu bağlamda, aynı güçlük düzeyi dikkate alınarak hazırlanmış farklı zamanlarda uygulanan formların sonuçları ya da aynı zamanda uygulanan farklı kitapçıkların sonuçları yatay eşitleme ile karşılaştırılabilir.

Bu araştırmada da benzer içerik ve benzer güçlük düzeyi esas alınarak hazırlanmış PISA kitapçıklardan elde edilen puanların eşitlenmesi amaçlandığı için bu bir yatay eşitleme çalışmasıdır.

Dikey eşitlemede, aynı özelliği ölçen ancak farklı eğitim düzeylerinde olan testleri eşitlemek için kullanılan ortak ölçek dönüştürme işleminde ham puanlar kullanılır (Holmes, 1982; Kolen, 1988).

Ülkemizde Milli Eğitim Bakanlığı (MEB) tarafından en son uygulamanın 2013 yılında gerçekleştirildiği SBS’ye 6., 7. ve 8. sınıf öğrencileri katılmıştır. Bu sınav sisteminde, öğretim programının sarmal oluşundan ötürü her ünite, farklı yıllarda, öğrencilerin karşısına farklı güçlük düzeyleri ile çıkmaktadır. SBS’ye ilişkin matematik, Türkçe, fen bilimleri ve sosyal bilimler için alt testler her yıl sınıf düzeyi arttıkça farklılaşan güçlüklerle öğrencilere uygulanmaktadır. İşte bu yüzden, SBS sonuçları puanları dikey eşitlemeye uygundur. Bu sayede sınıflarlar arası öğrenci başarılarını karşılaştırmak mümkündür.

Test Eşitlemenin Koşulları

Eşitleme sürecine başlamadan önce birtakım koşulların sağlanması gerekmektedir. Ulaşılabilen alanyazın tarandığında sağlanması gereken beş koşul olduğu tespit edilmiştir (Angoff, 1987; Dorans ve Holland, 2000; Kolen ve Brennan, 1995; Kolen ve Brennan, 2014; Lord, 1980; Petersen, Kolen ve Hoover, 1989). Bu beş koşul şunlardır; simetri, aynı özellikleri ölçme , eşit güvenirlik, gruptan bağımsızlık, eşitlik özelliğidir.

Simetri Özelliği

Form X'deki puanların form Y ölçeğine dönüşümünü sağlayan bir fonksiyon, form Y'den alınan puanların form X'e dönüşümü için kullanılan fonksiyonun alınmasını gerektirir. Örneğin; X formundan alınan 60 puan eşitlik denklemi ile Y formundaki 57 puana dönüştürülüyorsa aynı denklem ile Y formundaki 57 puan X formundaki 60 puana dönüştürülebilir ya da dönüştürülebilir olması, simetri özelliğinden kaynaklanmaktadır. Eşitlemede simetri özelliği vardır ama bu özellik regresyon denklemi değildir. Çünkü genelde X formunun Y formundaki regresyonu ile Y formunun X formundaki regresyonu farklıdır. Eğer X ve Y formu arasında hesaplanan korelasyon 1 olursa regresyon ile simetri özelliğindeki denklem aynı sonucu verir (Kolen ve Brennan, 2014).

Aynı Özellikleri Ölçme Özelliği

Eşitleme sürecinde kullanılacak olan testlerin benzer yetenekleri, özellikleri ya da yapıyı ölçmesidir. Bu nedenle farklı içerikteki testler eşitlenmez (Dorans ve Holland, 2000).

Eşit Güvenirlik Özelliği

Eşitlenecek olan testlerin güvenirliğinin aynı düzeyde olması gerekmektedir.

Gruptan Bağımsızlık Özelliği

Eşitleme yapmak için kurulacak olan eşitleme ilişkisini, sınava giren gruplardan bağımsız olmasıdır. Bu koşul sağlandığında kız ve erkek grupları için aynı eşitleme ilişkisi bulunur (Kolen ve Brennan, 1995).

Eşitlik Özelliği

Kolen ve Brennan (2014), bu eşitlik özelliğini ‘Lord’un eşitlik özelliği’ ve ‘zayıf eşitlik özelliği’ olarak iki başlık altında ele almıştır.

Lord’un Eşitlik Özelliği

Eşitlemenin temel bir gerekliliği hangi testin X, hangisinin Y olarak adlandırıldığına bakılmaksızın sonucun aynı olması gerektiğidir. Belirtilen bu özellik, testleri alan bireyler ile belirlenen gerçek puanların, gözlenen puanların ortalamasının ve standart sapmasının hem X formunda hem de Y formundaki dönüştürülmüş puanların dağılım şekillerinin aynı olmasını belirtmektedir. Bu özellik, form X ve form Y’nin parametreleri aynı olduğunda mümkündür. Böyle bir durumda da eşitlemeye gerek kalmaz. Örneğin, W akademik not ortalaması ise, Y testinde y puan alan bir kişi ve X testinde x puanı alan bir kişi, tahmin edilen not ortalamaları eşit olduğunda eşit kabul edilmektedir.

Zayıf Eşitlik Özelliği

Sınavı alan öğrenciler, gerçek puanları ile X ve Y formlarında aynı puanlara sahiptirler. Bu özellik ile sınava giren bireylerin, form X’de olduğu gibi form Y’de de aynı eşitlenmiş puanı elde etmeleri beklendiği anlamına gelmektedir.

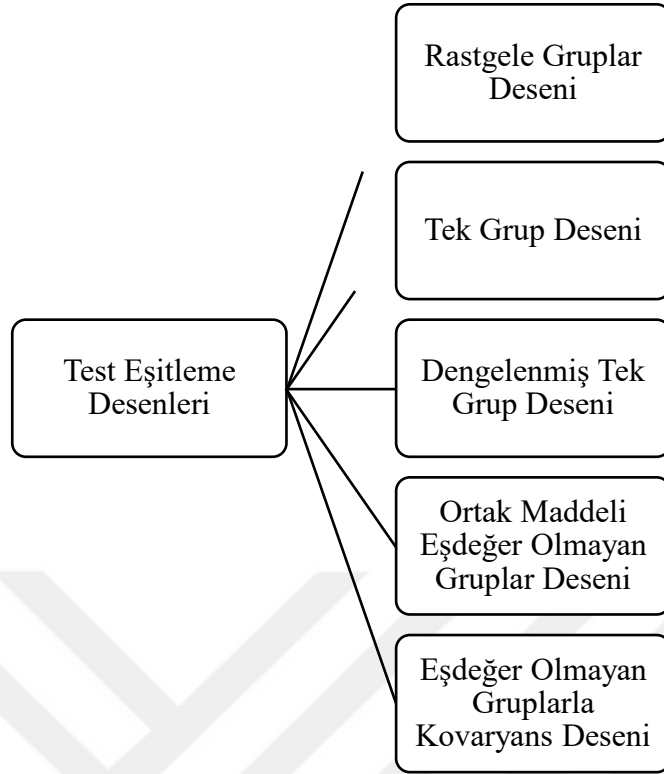
Test Eşitleme Aşamaları

Yukarıda belirtilen koşullar sağlandığında eşitleme süreci başlayabilir. Eşitleme süreci yedi aşamadan oluşmaktadır (Kolen ve Brennan, 2014):

1. Eşitleme için öncelikle amaç saptama
2. Testin aynı içeriğine ve benzer istatistiksel özelliklerine uygun alternatif formlarını oluşturma
3. Eşitleme için uygun desen seçme
4. Seçilen uygun desenin uygulamasını yapma. Bu desene uygun formlar uygulanır ve veri toplanır.
5. Eşitleme için bir ya da daha fazla operasyonel tanım seçme. Eşitleme ile kestirim yapılacak formlar arasında ne tür bir ilişki olduğunun belirtilmesi gerekir.
6. Bir ya da daha fazla istatistiksel tahmin yöntemi seçme. Eşitlemedeki ilişkiler için çeşitli kestirim yöntemleri vardır; örneğin eşit yüzdelikli eşitleme için Braun-Holland yönteminin seçilmesi gibi.
7. Eşitleme sonuçlarını değerlendirme.

Test Eşitleme Desenleri

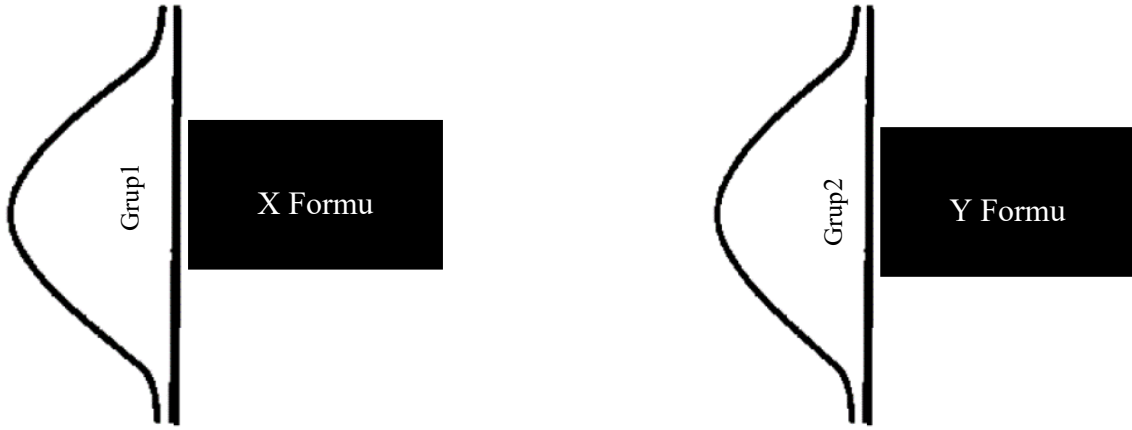
Eşitleme yapabilmek için veri toplama sürecine başlamak gerekmektedir. Eşitlemedeki veri toplama işlemine ‘eşitleme deseni’ denir (Kolen ve Brennan, 2014). Seçilen desen test eşitleme sürecinin başarılı bir şekilde neticelenmesi için çok önemlidir. Bu sebepten dolayı eşitleme için seçilecek desenin ekonomik ve yansız olması beklenmektedir (Thorndike, 1982). Veri toplama deseni seçilirken ‘uygulanabilirlik’ ve ‘istatistiksel özellikler’ göz önünde bulundurulması gereken unsurlardır (Öztürk, 2010). Test formlarını alan gruplar, formların uygulanış süreçleri, ekonomiklik ve ortak maddeler bulunuş durumları dikkate alınarak test eşitleme desenlerine göre Şekil-2’de sınıflandırılmıştır.



Şekil 2 Test eşitleme desenlerinin sınıflandırılması

Rastgele (Random) Gruplar Deseni

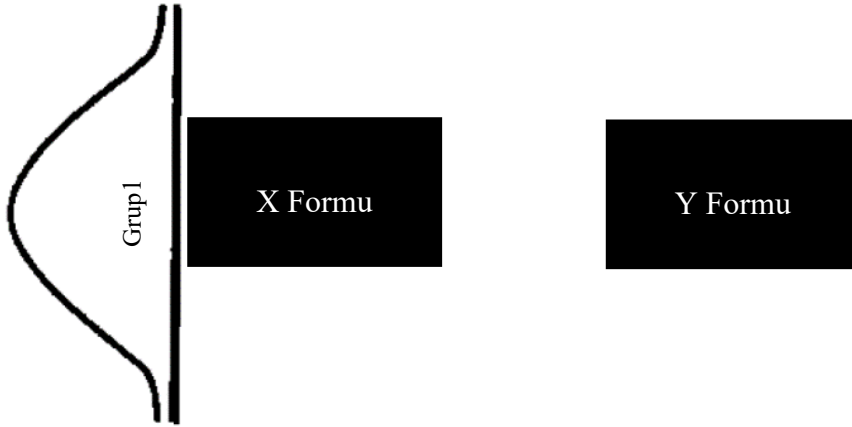
Bu desende, formlar sınava giren bireylere rastgele (random) olarak verilir. Her katılımcı test formunun yalnız bir tanesini alır. Bu sayede uygulama süreci kısaltılmıştır. İlk alınan form kolay, sonra alınan form ise yorgunluk gibi nedenlerden zor gibi görünebilir. Farklı bir durum olarak, ilk alınan form zor ise bu durum öğrenci motivasyonunu düşüreceğinden ikinci form kolay olsa bile zor olarak algılanır. Bu duruma “sıra etkisi” denir (Kolen ve Brennan, 2014). Bu desende sıra etkisi yoktur. Bu desen kullanıldığında iki formdaki gruplar arasındaki performans farklılıkları, formlar arasındaki güçlük farklılığının doğrudan bir göstergesidir. Bu desenin, oldukça büyük bir örneklemi gerektirmesine rağmen sayıltısı azdır. Simetri özelliğini sağlamaktadır. Sistemik hata azdır. (Kolen ve Brennan, 2014).



Şekil 3 Rastgele grup deseni

Tek Grup Deseni

Tek grup deseninde, iki test formu katılımcıların aynı grubuna uygulanır (Zhu, 1998). Yani sınava giren tüm adaylara hem X hem de Y formu verilir. Bu desen, rastgele grup desenine nazaran daha küçük örneklemelerde kullanılabilir.



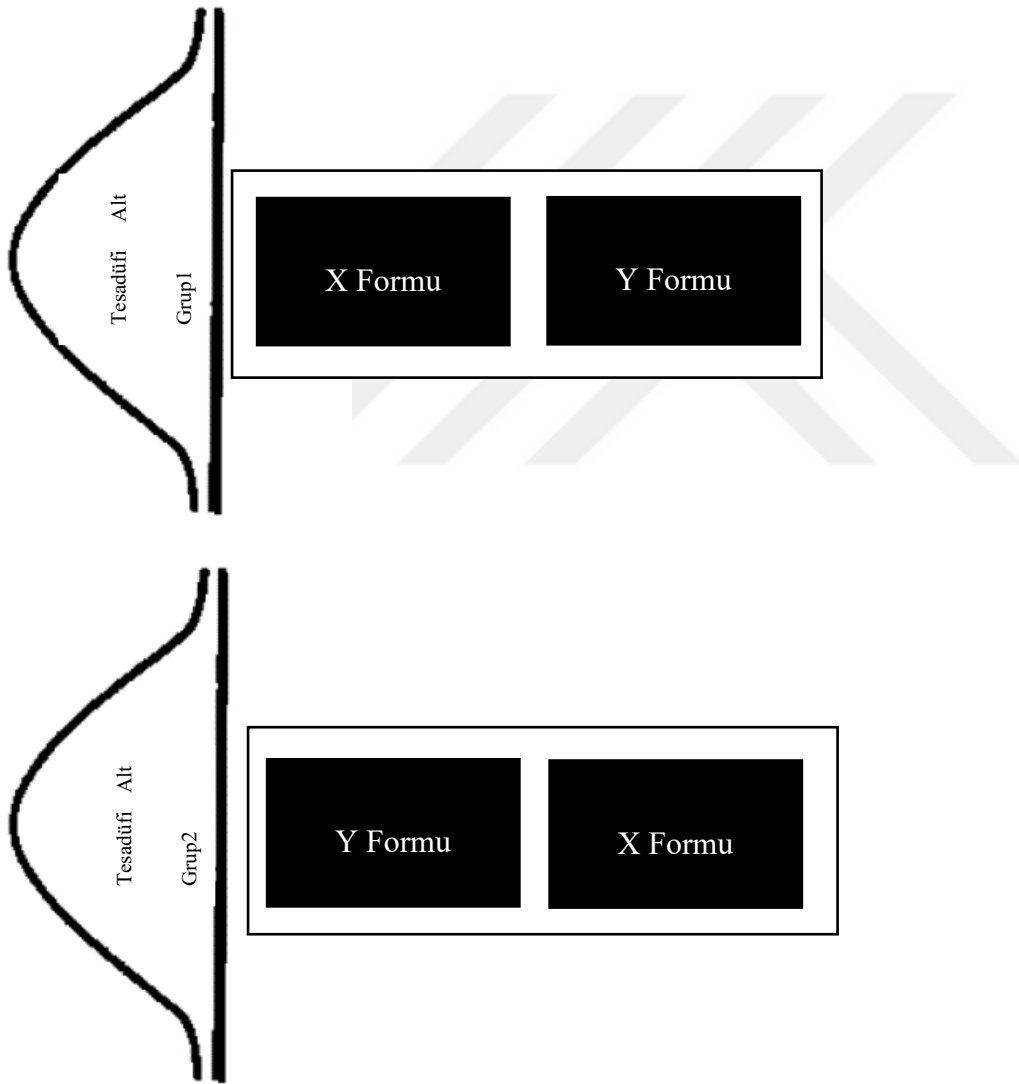
Şekil 4 Tek grup deseni

Dengelenmiş Tek Grup Deseni

Tek grup deseninde formların uygulanış sırasından kaynaklanan sıra etkisi, bu desende ortadan kalkmıştır. Çünkü dengelenmiş tek grup deseninde test kitapçıları hem X formunu

hem de Y formunu içermektedir. Belirtilen bu test kitapçıklarından birinde önce X formu sonra Y formu yer alırken; diğerinde önce Y formu sonra da X formu vardır (Kolen ve Brennan, 2014; Kolen, 1988).

Bu desenle sıra etkisi ortadan kaldırılırken katılımcılar, her iki formu da aynı anda cevapladığı için veri toplama sürecinde rastgele gruplar deseninden daha çok avantaj sağlamaktadır. Denek sayısı az olduğu ve test formları çok uzun olmadığına bu desenin kullanımı önerilmiştir (Tan, 2015).



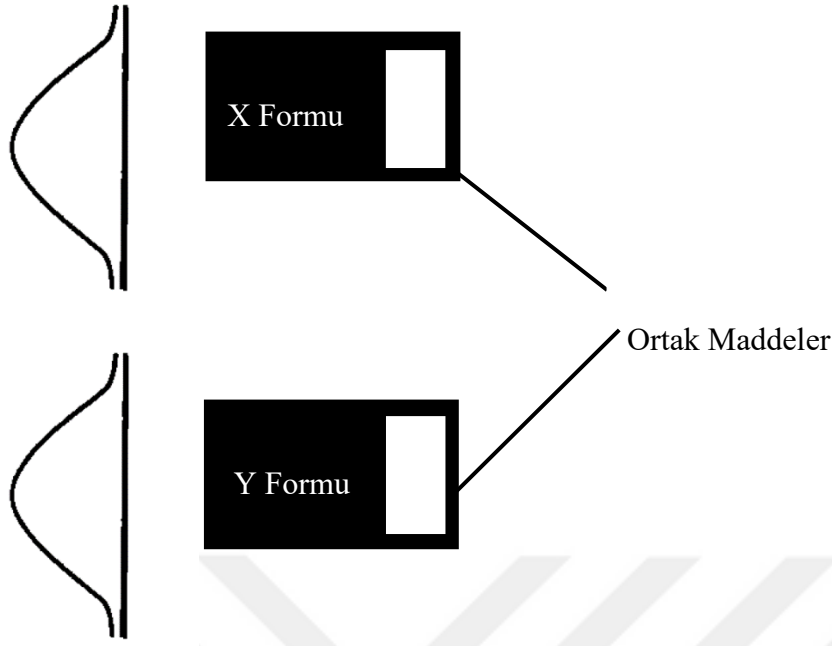
Şekil 5 Dengelenmiş tek grup deseni

Ortak Maddeli Eşdeğer Olmayan Grup Deseni

Testler, katılımcıların iki farklı grubuna uygulanmak uygulanmaktadır. Katılımcılara verilen formlarda aynı maddeler yer almaktadır. Bunlara, ankor (ortak) maddeler denir. Bu ortak maddeler, iki formda da uygulanan tamamen aynı maddelerdir. Bu maddelerin testteki sıraları da aynıdır. Ortak madde sayısı testin tüm madde sayısının en az %20'si kadar olmalıdır. Ortak maddeler testteki tüm madde grubunu temsil edecek şekilde olmalıdır (Kolen ve Brennan, 2014; Petersen ve diğerleri, 1989).

Oluşturulan ortak maddeler her bir testte yer almaktadır. Bu nedenle iki form arasındaki farklılıklar, ortak madde istatistiklerine bağlı olarak ayarlanabilmektedir. Çünkü formları alan iki grup eşdeğer olmak zorunda değildir. Bu desen, testin güvenilirliği ve uygulama etkisinden dolayı bu iki formun uygulanması güç olduğu durumlarda oldukça kullanışlıdır. Ayrıca bu desende ortak ölçekte yer alan test maddelerinin olduğu bir soru bankası geliştirilmesi gerekmektedir. Süreçte yer alan istatistiksel varsayımlar, grup özelliklerinin etkisini ortadan kaldırmak için gerekmektedir (Petersen ve diğerleri, 1989).

Uygulanabilirlik, veri seçim yöntemleri için oldukça önemli bir kriterdir. Tekrar etmek gerekirse tek grup, rastgele grup ve dengelenmiş tek grup desenleri uygulamada yerine getirilmesi oldukça zor olan, çok az ya da araya zaman girmeden yöneltile test ile iki test uygulamasını gerektirir. Bu sebeple ortak test deseni uygulamada daha çok tercih edilir. (Kolen, 1995; Kolen ve Brennan, 2014).



Şekil 6 Ortak maddeli eşdeğer olmayan gruplar deseni

Eşdeğer Olmayan Gruplarla Kovaryans Deseni

Test puanını eşitleme, farklı test formlarından farklı test puanlarını karşılaştırmak için kullanılır (Kolen ve Brennan 2014; Gonzalez ve Wiberg 2017). Farklı test formlarını alan test grupları benzer olarak değerlendirilebiliyorsa, eşdeğer gruplar tasarımı kullanılabilir. Burada ortaya çıkabilecek olan bir sorun, farklı test formları alan test gruplarının eşdeğer olamayacağı ve bu nedenle, bir ortak testi kullanıldığında eşdeğer olmayan gruplarda ortak test deseni kullanılması gerektiğidir (Lyren ve Hambleton 2011). Ortak test, ortak maddeli eşdeğer olmayan gruplar deseninde belli yetenek farklılığı olan grupları ayarlamak için kullanılır. Ortak maddeli eşdeğer olmayan gruplar deseninin her zaman kullanımı mümkün değildir. Farklı yetenekteki grupların puanları, ortak test aracılığı ile eşitlenmek istenir. Tüm testlerin ortak test mantığı ile oluşturulması mümkün değildir. Bir yapının zaman içerisindeki değişimini incelemek için ortak test kullanılıyorsa ve bu ortak testler sınırlı sayıda test konu alanını kapsıyorsa yine problem teşkil etmektedir. Bu sorunların her ikisi

de test güvenilirliğini ve geçerliğini sağlama girişiminden kaynaklanmaktadır. Ortak testin yokluğunda, eşdeğer olmayan gruplarda kovaryans desen önerilmektedir (Wiberg ve Branberg, 2015). Bu desende, ortak test yerine test puanları ile yüksek korelasyon gösteren değişken kullanılmaktadır. Örneğin, bir j maddesine doğru cevap verme olasılığı ve kodeğişken olarak o soruyu çözebilme kabiliyeti göz önüne alınarak j maddesinden elde edilen puanlar eşitlenmek istendiğinde bu desen uygun olacaktır.

Önerilen yöntemlerle ilgili bir sorun, her zaman test puanları ile yüksek korelasyonlu değişkenler değil, cinsiyete göre değişen puanları veya eğitim geçmişi gibi test puanları etkileyen ortak değişkenlerin olmasıdır.

Veri toplama sürecine yönelik desen seçildikten sonra eşitleme yöntemleri belirlenmelidir.

Test eşitleme yöntemleri, test teorilerine bağlı olarak iki genel kategoriye ayrılır:

- Geleneksel eşitleme (Kolen, 1988)
- Madde Tepki Kuramına (MTK) dayalı eşitleme (Cook ve Eignor, 1991)

Klasik Test Kuramına Dayalı Eşitleme Yöntemleri

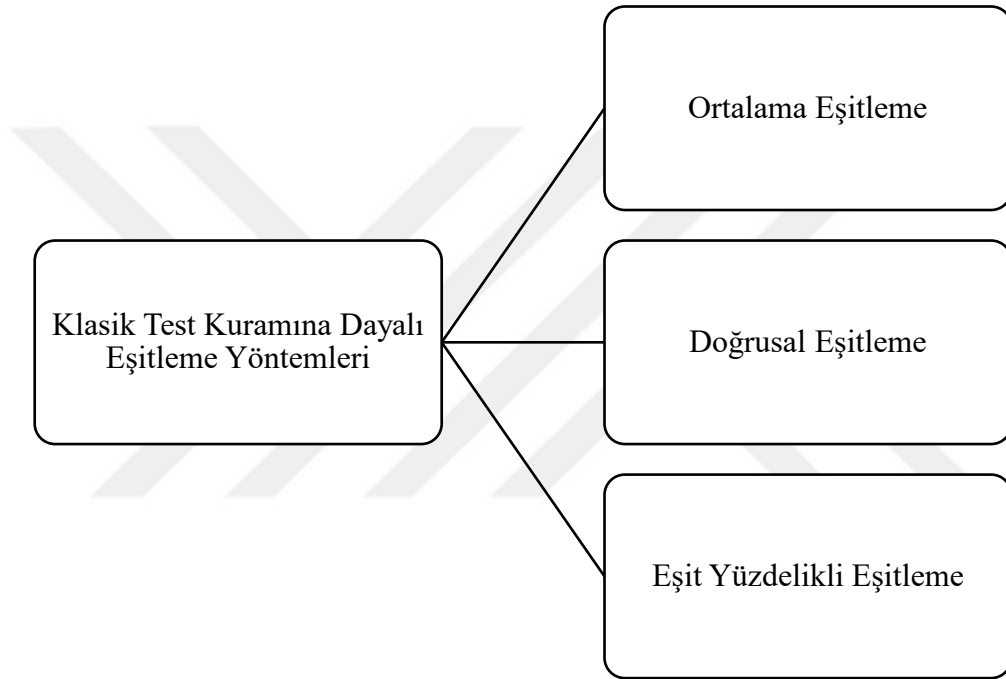
Klasik Test Kuramı (KTK), İngiliz psikolog Charles Spearman'ın "korelasyon" kavramına olan yoğun ilgisi sonucu ortaya çıkan "Klasik Gerçek Puan Modeline" dayanmaktadır. Spearman (1904), test puanlarının, bireylerin özelliklerine ilişkin "hata içeren ölçmeler" olduğunu ve bu hatalı puanlar arasında gözlenen korelasyonun bireylere ilişkin "gerçek objektif puanlar" arasındaki korelasyondan daha düşük olduğunu ifade etmiştir. Spearman (1907, 1913), "hata içeren ölçmeler" ve "gerçek objektif puanlar" kavramları arasındaki farkı açıklayabilmek için klasik gerçek puan modelini oluşturmuştur.

Spearman'ın modelinin iki ana noktası vardır ve test puanları bu iki hipotetik parçadan oluşmaktadır. Bu parçalar: Gerçek puan ve rastgele hatadır. Model, aşağıdaki eşitlikte açıklanmıştır.

$$X=T+E$$

- X : Gözlenen test puanı
T : Gerçek test puanı
E : Hata

Geleneksel eşitleme yöntemi, gerçek puan ve hatanın gözlenen puanı oluşturduğuna inanılan Klasik Test Kuramına dayanır (Hambleton ve Jones, 2012). Geleneksel eşitleme yöntemleri ortalama eşitleme, doğrusal eşitleme ve eşit yüzdelikli eşitleme (EYE) olmak üzere üçe ayrılır (Peterson, 2014). Şekil 7’de geleneksel eşitleme yöntemleri belirtilmiştir.



Şekil 7 Klasik Test Kuramına dayalı eşitleme yöntemleri

Ortalama Eşitleme Yöntemi

İki form için kendi ortalamalarından eşit uzaklıktaki puanların eşit olduğu düşüncesi üzerine kurulmuş bir eşitleme yöntemidir. Diğer bir deyişle, grubun kendi aritmetik ortalamasından aynı mesafede olan puanların eşit olduğu söylenebilir (Kolen ve Brennan, 2014). Puanlardaki tek farklılık, test formlarının güçlük farkından kaynaklanır. Örneğin, ortalamaya göre daha yüksek puan alan öğrenciler için form X puanı Y puanından 2 puan kolaysa, düşük

puan alan öğrenciler için de X formu Y formundan 2 puan daha kolay olacaktır. Bu bağlamda iki test arasındaki güçlük farkının, tüm puan ölçeği boyunca sabit olduğu söylenebilir. Ortalama eşitleme yöntemi kullanılarak eski form yeni forma eşitlenmek istendiğinde aşağıda belirtilen eşitlik kullanılmaktadır (Kolen ve Brennan, 2014):

$$X - \mu(X) = Y - \mu(Y)$$

X : X testinden (eski formdan) alınan puan

$\mu(X)$: X testinin ortalaması

Y : Y testinden (yeni formdan) alınan puan

$\mu(Y)$: Y testinin ortalaması

X formu, bu eşitlik kullanılarak Y formu için düzenlendiğinde $\mu_Y(X)$, eşitlik aşağıda belirtilen şekilde kullanılabilir (Kolen ve Brennan, 2014).

$$\mu_Y(X) = Y - \mu(X) + \mu(Y)$$

Doğrusal (Linear) Eşitleme

Bu eşitleme yöntemi, iki form arasındaki değişikliklerin ölçek boyunca belirlenmesine imkân verir. Bunun nedeni, ortalama eşitleme gibi iki form arasında sabit bir farklılığın olmamasıdır. (Kolen, 1995; Kolen ve Brennan, 2014). Örneğin, X ve Y formlarının güçlük farklılıkları yetenek düzeyi düşük, orta ve yüksek yetenek düzeylerindeki öğrenciler için farklılık göstermektedir. Bu durum, bireylerin yeteneklerinin ölçek boyunca aynı olmamasıyla ilişkilendirilebilir.

Doğrusal eşitleme, puanların kendi ortalamalarından farkının standart sapmalarına bölümü üzerine kurulmuştur. Ortalama eşitlemeden farkı, aşağıdaki eşitlikteki standart sapma değeridir (Kolen ve Brennan, 2014).

$$\frac{X - \mu(X)}{\sigma(X)} = \frac{Y - \mu(Y)}{\sigma(Y)}$$

X : X testinden (eski formdan) alınan puan

$\mu(X)$: X testinin ortalaması

Y : Y testinden (yeni formdan) alınan puan

$\mu(Y)$: Y testinin ortalaması

$\sigma(X)$: X testinin standart sapması

$\sigma(Y)$: Y testinin standart sapması

X formu bu eşitlik kullanılarak Y formu için düzenlendiğinde, $l_Y(X)$, eşitlik aşağıda belirtilen şekilde kullanılabilir.

$$l_Y(X) = y = \sigma(Y) \left[\frac{x - \mu(X)}{\sigma(X)} \right] + \mu(Y)$$

Yeniden düzenlendiğinde,

$$l_Y(X) = y = \frac{\sigma(Y)}{\sigma(X)} x + \left[\mu(Y) - \frac{\sigma(Y)}{\sigma(X)} \mu(X) \right]$$

Eşitliği ile eğim ve sabitler elde edilir.

Bu araştırmada ortak maddeli eşdeğer olmayan grup deseni altında doğrusal eşitleme yöntemleri ayrıca ele alınmıştır.

Ortak Maddeli Eşdeğer Olmayan Gruplar- Doğrusal Eşitleme Yöntemleri

Bu desen için farklı örneklemelerden katılımcılar iki gruba ortak madde içeren farklı testler uygulanmıştır. Bu desen, genellikle verilen test zamanında testin sadece bir formu uygulandığında kullanılır. Böylece ortak maddelerin bir testte, aynı sırada, aynı içerikte, aynı istatistik özelliklerde hazırlanması gerektiği vurgulanmıştır (Kolen, 1988).

Ortak maddeli eşdeğer olmayan grup desenleri için iki özel durum vardır. Bunlardan ilki ortak madde, tüm formlar için test puanlarına yansıtılarak hesaplanırsa internal (iç ankor)

madde olarak belirtilir. İkincisi ise bu ortak maddeler test puanında hesaba katılmıyorsa bunlara da external (dış ankor) maddeler denir. Sembollerle gösterilecek olursa yeni form ve random değişkenlik puanları, X formu gibi; eski form ve random değişkenlik, Y formu gibi gösterilir ve ortak madde oluşturulur. Ayrıca ortak maddeler üzerindeki random değişkenlik puanları, V olarak kabul edilir (Kolen ve Brennan, 2014).

Örneklem 1’den katılımcıların grupları tarafından alınan X ve V’nin alındığı; örneklem 2’den katılımcı grupları tarafından alınan Y ve V’nin alındığı varsayılır. Eğer V, ortak maddelerden oluşan iç ankor ise X ve V üzerinden puanları içerir. Eğer V dış ankor ise X, Y ve V üzerinden puanları içermez. Örneğin 10 maddesi ortak olan, 40 maddesi ortak olmayan maddeleri doğru cevaplamış olan katılımcıyı düşünelim; eğer V iç ankor madde ise X puanı 50’dir. Eğer V dış ankor ise X puanı 40 olur.

Genel olarak ortak maddeler, örneklem farklılıklarını düzeltmek için kullanılır. Her bir katılımcı sadece bir örneklemden ve bir formdan geldiği için güçlü istatistiksel varsayımları gerektirir. Ortak maddeli eşdeğer olmayan gruplar deseni altında eşitleme performansı için çeşitli yöntemler, bu istatistiksel varsayımlar açısından tanımlanabilir (Kolen ve Brennan, 2014).

Bu desen, iki evren içermesine rağmen bir eşitleme fonksiyonu, tipik olarak, tek evren için tanımlanıyor gibi görünmektedir. Bundan dolayı, evren 1 ve evren 2 bir ilişkisinin tanımlaması için tek bir evrenden elde ediliyor gibi birleştirilmelidir. Bu evren “sentetik evren” olarak tanımlanır (Braunve Holland, 1982). Sentetik evren, form X ve form Y’nin uygulandığı evrenlerin farklı ağırlıklandırılmaları ile tanımlanır (Zhu, 1998).

Evren 1’in ağırlığı w_1 , evren 2’nin ağırlığı w_2 olduğu düşünüldüğünde $w_1 + w_2 = 1$ kuralına uyarak $w_1, w_2 \geq 0$ olmalıdır. Bu desen için uygulanacak doğrusal eşitleme yöntemlerinde bu ağırlıklar için olası durumlar şu şekildedir (Kolen ve Brennan, 2014):

- 1- $w_1=1, w_2=0$
- 2- $w_1=0, w_2=1$

3- $w_1=0,5$, $w_2=0,5$

Aynı zamanda form X ve form Y'yi alan örneklemelerin büyüklüklerine göre bir ağırlıklandırılma da yapılabilir:

$$w_1 = \frac{N_1}{N_1 + N_2} \quad w_2 = \frac{N_2}{N_1 + N_2}$$

N_1 : Evren 1'in örneklem büyüklüğü

N_2 : Evren 2'nin örneklem büyüklüğü

Formlardan ilki evren 2'ye ve formlardan ikincisi ise evren 1'e uygulanmadığı için doğrudan kestiremeyeceğimiz değerler ortaya çıkacaktır. İşte bu durumda, ortak maddeli eşdeğer olmayan gruplar için uygun doğrusal eşitleme yöntemleri kullanılmalıdır. Aşağıda doğrusal yöntemler özetlenmiştir.

Tucker Doğrusal Eşitleme Yöntemi

Tucker yönteminde, Ledyard Tucker'a ithafen Gulliksen (1950) tarafından tanımlanmıştır. Ortak maddeli eşdeğer olmayan gruplar deseninde farklı gruplara uygulanan ortak maddelerin olduğu farklı formlar uygulandığı bir desendir (Angoff, 1971). Tucker doğrusal yöntemde, yukarıda sunulan sentetik evren ağırlıklarını dikkate alarak gruplar farklı sentetik evren ağırlıkları dikkate alınarak eşitlenmeye çalışılır. Kolen ve Brennan'a (2014).ve Gulliksen' e (1950) göre V ortak test olarak kabul edildiğinde, X'in V üzerindeki regresyonu hem evren 1 hem de evren 2 için aynı doğrusal fonksiyon varsayılır. Benzer varsayım Y'nin V üzerindeki regresyonu için de yapılır. Tucker yöntemi için V'de verilen X'in koşullu varyansı ile V'de verilen Y'nin koşullu varyansı, evren 1 ve evren 2 için aynıdır. Bu bilgiler ışığında, iç ankor V puanların yardımı ve sentetik evren ağırlıkları kullanılarak ortalama ve varyans değerleri kestirilmeye çalışılmaktadır.

Levine Gözlenen Puan Eşitleme Yöntemi

Bu yöntem, Levine (1955) tarafından geliştirilmiştir. Bu yöntemde sentetik evren kavramı açık olarak ele alınmamıştır. Bu, X üzerindeki gözlenen puanları Y üzerindeki gözlenen puan ölçeğiyle ilişkilendiren bir gözlenen puan yöntemidir. Bu yöntem için varsayımlar KTK' ya göre gözlenen puanlarla ilişkili olduğu varsayılan T_x , T_y ve T_v gerçek puanları ile ilişkilidir.

$$X = T_x + E_x$$

$$Y = T_y + E_y$$

$$V = T_v + E_v$$

E_x , E_y ve E_v sıfır beklentiye sahip ve gerçek puanlarla ilişkisiz hatalardır. Levine yöntemi X, Y ve V evren 1 ve evren 2'nin her ikisinde T_x ve T_v arasında, T_y ve T_v arasında korelasyonun mükemmel olması durumunda X, Y ve V' nin aynı şeyleri ölçtüğünü belirtmektedir. Yani ortak madde puanlarıyla hem T_x hem de T_y mükemmel korelasyonlu olduğunda T_x ve T_y de ilişkili olur (Kolen ve Brennan, 2014).

Levine Gerçek Puan Eşitleme

Levine (1955) tarafından geliştirmiş olup Levine gözlenen puan eşitleme yöntemi ile aynı varsayımları içermektedir. Gözlenen puan ve gerçek puan yöntemleri arasındaki uygulama farklılığı, X üzerindeki gözlenen puanları Y üzerindeki gözlenen puanlar ölçeğine dönüştüren

$$l_{ys}(X) = \frac{\sigma(Y)}{\sigma(X)} [X - \mu_s(X)] + \mu_s(Y)$$

denkleminde gerçek puanların kullanılması oluşturmaktadır. KTK'da gözlenen puan ortalaması ile gerçek puan ortalamalarının eşit olduğu varsayımı, bu yöntemde kullanılmaktadır. Hesaplanan eğim ve regresyon kesimleri sentetik evren ağırlıklarına bağlı

değildir. Bu nedenle Levine gerçek puan eşitleme yöntemi, sentetik evrenin kavramsal çerçevesini gerektirmemektedir.

Klasik Konjenerik Modele Dayalı Doğrusal Eşitleme

Bu yönteme geçmeden önce bu bölümde Konjenerik Test Kuramı özetlenmek istenmiştir. Klasik Test Kuramı, üç önemli kavram üzerine inşa edilmiştir. Bunlar; hata (E), gözlenen puan (X) ve gerçek puan (T)'dir. KTK'da güvenilirlik kavramı, tesadüfi hatalardan arınık olma derecesi olarak tanımlanmıştır. Bu kuramdaki güvenilirlik indeksi KTK'daki terimlerle ifade edilir. Buna göre güvenilirlik indeksi, gerçek puan varyansının gözlenen puanlar varyansına oranı olarak formüle edilir. Gerçek puanlar, bir testin sonsuz defa uygulanmasından elde edilen sonuçların ortalamasıdır. Bu sebeple doğrudan gözlenemediğinden KTK'nın ve güvenilirlik indeksinin "platonik" bir yapısı vardır (Lord, 1968). Bu bağlamda ölçme sonuçlarının güvenilirliklerini elde edebilmek için aynı özelliği ölçmeye yönelmiş ek ölçmelere/maddelere ihtiyaç duyulur. Birleşik testlerde, ölçme aracının güvenilirlik indeksi, "güvenirlik katsayıları" olarak adlandırılan birtakım formüller sayesinde kestirilebilmektedir (Baykul, 2015). Konjenerik test kuramı KTK'nın bir alt boyutudur (Lucke, 2005). Bu kuramdaki gözlenen puan eşitliği KTK'daki doğrusal modelin madde karakteristiklerini kapsayan geliştirilmiş bir biçimdir. Konjenerik ölçme modeli;

$$X_i = \mu_i + \lambda_i F + E_i$$

μ_i : i. ölçmenin/maddenin güçlük indeksini

λ_i : i. maddenin ayırtıcılık indeksi

F : Tüm ölçmelerin ortak faktör değeri

Şeklinde (Joreskog, 1969; Raykov, 1997). McDonald (2013) bu modele dayalı madde puanını;

$$T_i = \mu_i + \lambda_i F$$

olarak tanımlamıştır.

Konjenerik test kuramı; daha öncede bahsedildiği gibi, birleşik testte ($Y=X_1+...+X_k$) yer alan maddelerin;

$$X_1=T_1+E_1$$

$$X_2=T_2+E_2$$

:::

$$X_k=T_k+E_k$$

ölçmeye yöneldiği özelliği aynı nicel büyüklükte ölçmediği durumlarda gerçek puanlar arasındaki fonksiyonel bir bağıntının varlığı ile açıklanmaktadır (Joreskog, 1969).

Konjenerik modele dayalı eşitleme, T_x ve T_v 'nin mükemmel korelasyon veren varsayımla uyumlu olduğu doğrusal ilişki bulunan T_x ve T_v 'nin bir adım daha iyileştirilmiş halidir. Bu yöntemde λ 'ler eğimler ve δ 'lar sabit kesimler ile T_x ve T_v doğrusal olarak ilişkiyi temsil etmenin yolu $T_x = \lambda_x T + \delta_x$ ve $T_v = \lambda_v T + \delta_v$ ayarlamaktır (Feldt ve Brennan 1989).

Braun-Holland Yöntemi

Bu yöntemde, frekans kestirim yönteminin varsayımları kullanılarak ortaya çıkan ortalama ve standart sapmaları kullanılarak eşitleme yapılır. Frekans kestirim varsayımı ile X formunun ortalama ve standart sapma puanları aşağıdaki eşitlik sayesinde kestirilebilir.

$$\mu_s(X) = \sum x f(X)$$

$$\sigma_s^2(X) = \sum [X - \mu_s(X)]^2 f_s(x)$$

Bu sayede X formunun sentetik evrendeki ortalama ve varyans değerleri, Y formuyla benzer hale getirilmiş olur. Braun-Holland yöntemi Tucker yöntemi ile yakından ilişkilidir. Aşağıdaki koşullar sağlandığında bu iki yöntemin sonuçları özdeş olur:

*V üzerideki X'in ve V üzerindeki Y'nin regresyonu lineer olduğunda

*V üzerinde X'in ve V üzerinde Y'nin regresyonu homoscedasticity olduğunda Tucker yöntemi gibi bu yöntemde de eğim (slope) ve regresyon kesim (intercept) değerleri hesaplanmalıdır. Y formunun standart sapması ile X formunun standart sapması oranlandığında eğim; Y formunun ortalamasından, bulunan eğim değerinin X formu ortalaması ile çarpımından elde edilen değeri çıkarıldığında ise regresyon kesim değeri elde edilmiş olur.

Braun-Holland yöntemi kullanılarak eşitlenmiş puanlar aşağıdaki denklem kullanılarak elde edilir.

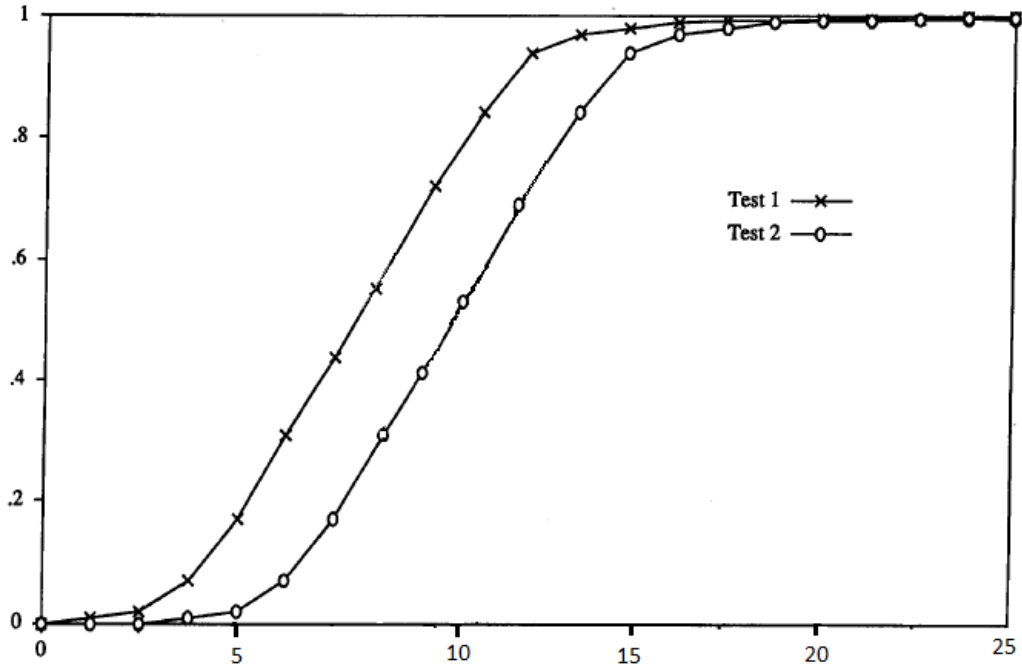
$$I_Y(X) = \frac{\sigma(Y)}{\sigma(X)} X + \left[\mu(Y) - \frac{\sigma(Y)}{\sigma(X)} \mu(X) \right]$$

Geleneksel eşitlemenin barındırdığı pek çok yöntemin eşdeğer olmayan gruplar desenine ilişkin problemlerle başa çıkmak için geliştirilmesine rağmen, genelde umulduğu gibi çalışmaz (Kolen, 1981). Madde tepki kuramına (MTK) dayalı eşitleme yöntemlerinin bu gibi durumlar için daha iyi alternatifleri vardır (Lord, 1975).

Eşit Yüzdelikli Eşitleme (EYE)

Eşit yüzdelikli eşitleme, eğri bir formdan diğer bir forma değişen güçlükleri tanımlamak için kullanılmaktadır. Bu da eşit yüzdelikli eşitlemeyi doğrusal eşitlemeden daha kullanışlı hale getirmektedir. Eşitleme fonksiyonunda, form Y ölçeğine dönüştürülmüş form X puanlarının dağılımı evrendeki form Y puanlarının dağılımına eşitse bu bir EYE fonksiyonudur. EYE fonksiyonu, form Y'deki puanlar gibi aynı yüzdelik sıralara sahip form X'deki puanlar tanımlanarak geliştirilmiştir. Yani EYE'de esas olan, farklı evrenlerden elde edilen puan dağılımlarını aynı yüzdelik sıraya denk gelen eşdeğerlerine dönüştürmektir. Angoff (1987)'a göre, eşit güvenilirlik derecesi ile X ve Y formlarından aynı özelliğin ölçülmesinde elde edilen puanların eşit yüzdelik sıralara sahip olanları eşdeğer olarak kabul edilir.

Bu eşitleme süreci grafiksel ve analitik süreçleri içermektedir.



Şekil 8 Eşit yüzdelikli eşitlemenin grafiksel gösterimi

EYE’de grafiksel ve analitik süreçte de aşağıdaki genel adımlar izlenir:

- Form X’de belirli bir X puanı için bu puan ve altında değer elde eden bireylerin yüzdeliği bulunur
- Bulunan yüzdeliğin, aynı yüzdeliğe sahip olan form Y’deki puan dengi bulunur.
- Bulunan Y form puanı, form X’deki puanın eşdeğeridir.

EYE’deki yüzdelik sırayı bulabilmek için aşağıdaki eşitlikten yararlanılır.

$$P(X)=100[F(X-1) +f(X)/2]$$

P(X): Yüzdelik sıra

F(X): Kesikli değerler için kümülatif yoğunluk fonksiyonu

f(X): Kesikli değerler için yoğunluk fonksiyonu

Ancak her zaman eşit puanlar ve yüzdeler iki form için bire bire denk gelmeyebilir. Bu nedenle kesikli olan test puanları, yüzdelikler ve yüzde aralıkları kullanarak işlem yapılabilir. Örneğin 20 puan, 19.5 ile 20.5 arasındaki puanlara karşılık geldiği düşünülerek bu değerler işleme alınabilir.

Grafiksel gösterimin uygulamada büyük sayılarda kullanılması oldukça zordur. Buna ek olarak, grafiksel yöntemi kullanarak yapılan eşitleme kusurlu olabilir. Yüzdelik sıraların daha uygun dağılımı için EYE’de formül kullanılması daha doğru sonuçlar verecektir.

$$e_y(X) = ((P(X))/100 - G(y_u^* - 1)) / (G(y_u^*) - G(y_u^* - 1)) + (y_u^* - 0,5)$$

$e_y(X)$: X formu puanından Y formuna dönüştürülen simetrik eşitleme fonksiyonu

$P(X)$: X için yüzdelik sıra fonksiyonu

y_u^* : Q^* 'dan daha büyük kümülatif yüzdeli en küçük tam sayı

$G(y)$: y'nin ayrı kümülatif dağılımı

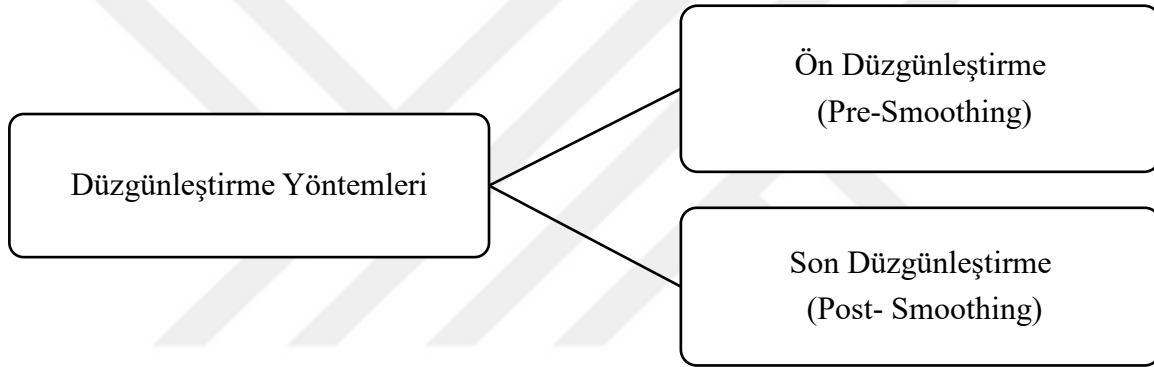
EYE’de testi alan bireylerin puan dağılımları düzensizdir. Puanı alan bireylerin yüzdesinin, kademeli olarak puan arttıkça değişme göstermediği ve bu yüzdelilerle dalgalanmaların olduğu görülmüştür. İşte bu düzensizlik EYE için sorun teşkil etmektedir. Bu nedenle puan dağılımındaki bu dalgalanmalar, oldukça büyük örneklem kullanılarak önüne geçilmeye çalışılsa da bilhassa örneklem küçük olduğunda bu eğriler, genellikle analitik düzgünleştirme yöntemleri ile düzgünleştirilir (Kolen, 1988; Livingston, 2014). Düzgünleştirme yöntemleri, evren karakteristiğinde düzgünlük özelliklerine sahip eşit yüzdelik ilişkisi ve puan dağılımlarının tahminini üretmek için geliştirilmiştir. Düzgünleştirme ile evrendeki ilişkileri bulmaya ve örneklemdeki kesikli dağılımların sürekli bir fonksiyona çevrilmesine çalışılmaktadır.

Düzgünleştirme yöntemleri, düzgünleştirilmemiş EYE’ den elde edilen daha az rastgele hata içeren düzgünleştirme fonksiyonlarını üretmek için tasarlanmıştır (Hanson, Zeng ve Colton, 1994). Düzgünleştirme yöntemlerinin kullanılmasıyla toplam hata ve rastgele hata azalır. Ancak sistematik hatayı artırabilir (Felan, 2002). Düzgünleştirme yönteminin amacı random hatayı azaltmaktan ziyade sistematik hatayı etkilemektir. Bu bağlamda, düzgünleştirme ile puan noktaları üzerinde evren eşit yüzdelikli eşdeğerlerinin tahmini suretiyle toplam hatayı azaltmak amaçlanmıştır. Eşitlemedeki hata türleri, bölümün sonunda daha detaylı bir şekilde belirtilmiştir.

Düzgünleştirme yöntemleri uygulamada kullanışlılık sağlanması açısından dört önemli özelliğe sahip olmalıdır (Kolen ve Brennan, 2014):

- 1- Dağılımların ve eşit yüzdelikli ilişkilerin çeşitliğini ele almada yeterli esnekliğe sahip olma
- 2- Evren dağılımlarının tahmininde doğru sonuçlar üretebilmeli
- 3- Çalışmaya uygun istatistiksel çerçeveye ilgili olmalı
- 4- Deneysel araştırmalara dayalı olarak gösterildiği gibi tahminler geliştirebilmeli

Analitik düzgünleştirme yöntemleri aşağıdaki Şekil 9’da sınıflandırılmıştır (Hanson ve diğerleri, 1994; Kolen, 1988; Kolen ve Brennan, 2014).



Şekil 9 Düzgünleştirme yöntemleri

Ön Düzgünleştirme (Pre-Smoothing)

Bu yöntemde öncelikle puan dağılımları düzgünleştirilir. Puan dağılımların doğru şekilde kestirilmesi, düzgünleştirme işleminde dikkat edilmesi gereken önemli bir noktadır. EYE işlemi daha sonra yapılır. Dağılımların düzgünleştirilmesinde, dağılımların tahminindeki doğruluk çok önemlidir. Doğrulukla yakından ilişkili önemli bir özellik, moment korunmasıdır. Moment korunmasında, düzgünleştirilmiş dağılım gözlenen dağılım gibi aynı merkezi momentlerden (ortalama, standart sapma, çarpıklık, basıklık) en az birine sahiptir. Örneğin, düzgünleştirilmiş dağılım ile düzgünleştirilmemiş dağılımın basıklık ve standart sapma değerleri eşitse bu yöntem sayesinde iki momentin korunmuş olduğu söylenebilir.

Ön düzgünleştirme yöntemleri ikiye ayrılır (Kolen ve Brennan, 2014):

1-Log-lineer yöntem:

Ön düzgünleştirme yöntemleri arasında en çok kullanılan yöntemdir (Liu, 2011). Bu yöntemde, örneklem yoğunluğunun logaritmasının polinomial fonksiyona uygunluğu dikkate alınarak oluşturulmuştur (Darroch ve Ratcliff, 1972; Haberman, 1974). Polinomial log-lineer yönteminde ne kadar düzgünleştirme yapılacağı polinomial derecesinin (C) seçimi ile ilişkilidir. Belirtilen yöntemde C değeri 1'den 10'a kadar değer alır (Kolen ve Brennan, 2014; Liu, 2011). Düzgünleştirilmiş model ile gözlenen modelin uygunluğu ki-kare dağılım tablosu ile karşılaştırılarak belirlenir (Kolen ve Brennan, 2014; Rosenbaum ve Thayer, 1987).

2-Beta 4 yöntemi:

Bu yöntem gerçek puanları kestirebilmek için kullanılır. Bu nedenle güçlü gerçek puan yöntemi olarak adlandırılmıştır. Bu yöntem, gerçek puanların düzgünleştirilmiş dağılımlarını verir. Aynı polinomial log-lineer yöntem gibi model uyumu incelenir. Bunun dışında, düzgünleştirilmemiş puan dağılımları ve beta4 yönteminden elde edilen puan dağılımları için grafik çizilerek model uyumu hakkında yorum yapılabilir (Kolen ve Brennan, 2014).

Son Düzgünleştirme (Post- Smoothing)

Son düzgünleştirme yönteminde eşit yüzdelikli eşdeğerler elde edildikten sonra düzgünleştirme yapılır. Tan'ın (2015) belirttiği gibi son düzgünleştirme yönteminde, test puanlarının dağılımı değil, dönüştürülen puanların düzgünleştirilmesi yapılır. Bu yöntemde log-lineer yöntemindeki polinom değerleri yerine kübik ara değerler kullanılmaktadır. Bu

değerler, dağılımdaki eğrilik düzeylerini minimum yapmak için esneklik sağlamaktadır (Kolen ve Brennan, 2014). Uygunluk fonksiyonunun kare farkı ve gözlenen fonksiyon, S düzgünleştiren parametresi tarafından kontrol edilmektedir. S değeri ne kadar yüksek olursa kareler farkı o kadar büyük olmakta; böylece uygunluk fonksiyonu o kadar düzgün olmaktadır. S parametresi 0'dan sonsuza kadar değişebilir. $S=0$ olduğunda, yöntem sanki hiç düzgünleştirme yapmamış gibi görünmektedir. Çünkü bu durumda fit dağılımı ve orijinal dağılım arasında hiç fark yokmuş gibi görünmektedir (Hanson ve diğerleri 1994; Kolen, 1988).

Ortak Maddeli Eşdeğer Olmayan Gruplar – Eşit Yüzdelikli Eşitleme Yöntemleri

Toplam puanların ve ortak maddelerden elde edilen puanların dağılımına dikkat etmek önemlidir. Bu yöntem, sentetik evrenin dikkate alınmasını gerektirir. Ortak maddeli eşdeğer olmayan grup deseninde elde edilen toplam puan ve ortak madde puanları, EYE ile eşitlenmeye çalışılırken frekans kestirim yönteminin farklı sentetik evren ağırlıklarına göre eşitleme fonksiyonları kullanılmıştır.

Frekans Tahmin Yöntemi

Angoff (1971), Braun ve Holland (1982) tarafından açıklanan frekans tahmini eşit yüzdelikli eşitleme yöntemi, ortak maddeli eşdeğer olmayan grup deseni kullanılarak toplanan verilerden, bir sentetik evren için Form X ve form Y üzerindeki kümülatif puan dağılım tahmini için bir ortalama sağlar. Yüzdelik sıralar kümülatif dağılımlardan elde edilir ve formlar EYE yöntemi ile eşitlenir.

Frekans tahmininde, eşit yüzdelikli eşitlemeye yol göstermek amacıyla sentetik evren için dağılımların ifade edilmesi gereklidir. Bu dağılımlar her bir evren için dağılımların ağırlıklandırılmış kombinasyonu olduğu düşünülür.

$$f_s(X)=w_1.f_1(X)+w_2.f_2(X) \text{ ve } g_s(Y)=w_1.g_1(Y)+w_2.g_2(Y)$$

w_1 : Evren 1 için ağırlık

w_2 : Evren2 için ağırlık

f_1 : Evren 1’deki X formun frekans değeri

f_2 : Evren 2’deki X formun frekans değeri

g_1 : Evren 1’deki Y formun frekans değeri

g_2 : Evren 2’deki Y formun frekans değeri

X formu evren 2’de Y formu da evren 1’de uygulanmadığı için bu değerler doğrudan gözlenemez. Bu durumda, bu yöntem için hem form X’de hem form Y’de her bir puanın verilen toplam koşullu dağılımının her iki evren için de aynı olduğu varsayımı kabul edilmiştir.

Frekans kestirim yöntemi ile EYE’de evrenlerin birbiriyle benzer olduğu durumda yapılması uygundur. Evrenler farklılaştıkça gerçek puana dayalı yöntemlere (modifiye edilmiş frekans yöntemini gibi) yönelmek daha uygun olacaktır. Bu araştırmada, evrenler arasında farklılıklar fazla olmadığından frekans kestirim yöntemi uygulanmıştır.

Madde Tepki Kuramı ve Madde Tepki Kuramına İlişkin Test Eşitleme Yöntemleri

İlk temeli Thurstone’ nun (1925) ‘Psikolojik ve Eğitsel Testlerin Ölçeklenmesi’ adlı eseri ile ortaya atılan Madde Tepki Kuramı (MTK), KTK’ nın zayıf varsayımlarına karşı geliştirilmiştir (Embretson ve Reise, 2013). Klasik madde analizi istatistiklerinde cevaplayıcının, maddeyi ölçtüğü özelliğin farklı beceri düzeylerinde nasıl olduğu hakkında bilgi sağlanamamaktadır (Crocker ve Algina, 1986). Maddelere doğru cevap verme

olasılığını kontrol eden sürekli bir değişkenin vardır. Bu değişkenin bilinen bir değişken değil örtük (gizil) bir değişken olduğu MTK'nın çıkış noktasıdır (Embretson ve Reise, 2013). Kuramın temelinde, ölçülen özellik için farklı yetenek düzeylerindeki bireylerin bir maddeye nasıl yanıt vereceğini gösteren matematiksel bir model vardır (Baker, 1992; Crocker ve Algina, 1986). Baker (2001), Crocker ve Algina (1986), Doğan ve Tezbaşaran (2003), Embretson ve Reise (2013), Hambleton ve Swaminathan (2013) ve Lord (1968) çalışmalarında aşağıda belirtilen özelliklerle MTK'yı kullanmanın avantajlı olduğunu savunmuşlardır. Bu avantajların birkaçı şu şekildedir:

- Testteki madde güçlük düzeyleri ve elde edilen yetenek düzeyleri aynı ölçekte yer alır.
- Bireyin örtük özellikteki yeri, model parametreleri göz önünde bulundurularak tahmin edilebilmektedir.
- Cevaplayıcının yeteneği, daha az ölçme hatası ile tespit edilebilir. Ve bu ölçme hatası tüm cevaplayıcılarda aynı değildir.
- Madde parametreleri, aynı evrenden gelmiş farklı örneklerde de değişmezlik özelliği göstermektedir.

Bu bilgi, farklı testleri yanıtlayan bireylerin performanslarını birbirleriyle karşılaştırmaya imkân sağlamaktadır.

MTK, KTK' daki doğru-yanlış, var-yok gibi ikili puanlamanın dışında likert tipi, doğru cevabın derecelendirildiği çoklu puanlamaya da izin vermektedir. Fakat bu araştırma, MTK'nın sadece ikili puanlaması üzerine kurulmuştur.

Bir araştırmayı MTK bünyesinde inceleyebilmek için üç önemli varsayımı sağlamak gerekir. Bunlardan ilki "*tek boyutluluk*" varsayımdır. Tek boyutluluk varsayımı, test maddelerinin tek bir örtük özellik için istatistiksel bağımlı olmasıdır. Bu da bir testin maddeleri tüm evren için istatistiksel bağlıysa ve örtük özelliğe göre homojen kişilerden oluşuyorsa bir alt grupta

maddeler tek bir örtük özellik ortaya çıkarmışsa testin tek boyutlu olduğu anlamına gelmektedir (Crocker ve Algina, 1986). Kısacası, testteki maddelerle tek bir özellik ölçülüyorsa o testte tek boyutluluk sağlanmıştır. Belirtilen bu tek boyutluluk varsayımı MTK tek boyutlu eşitleme için aranan bir varsayımdır. Bu varsayımla birlikte MTK kapsamında çok boyutlu eşitleme çalışmaları, alanyazında pek çok çalışmanın konusunu olmuşturmaktadır.

İkinci varsayım, “*yerel bağımsızlık*” varsayımdır. MTK’ya ilişkin pek çok dokümanda tek boyutluluğun sağlanması, bu varsayımda sağlandığına bir kanıt oluştursa da yerel bağımsızlık testi alanın, farklı test maddelerine verdiği tepkilerin istatistiksel olarak bağımsızlığıdır. Yani madde içeriğinin, diğer maddelerin cevaplanmasında ipucu olabilecek ifadeleri kapsamaması gerekir. Bu varsayımın doğrulanması için testi alanın bir maddedeki performansının, iyi veya kötü olarak testteki diğer maddeleri etkilememesi gerekmektedir.

Üçüncü varsayım ise “*madde karakteristik eğrisinin (MKE) monotonik artışı*” varsayımdır. MKE testi alanın örtük özelliği ile madde parametreleri arasında ilişki kuran, olasılıklı bir modeldir. MKE, belli bir yetenek düzeyindeki cevaplayıcının maddeyi doğru cevaplama olasılığını veren matematiksel bir fonksiyondur. Bu maddede, karakteristik eğrisiyi belirten fonksiyonu temel alan iki model bulunmaktadır. Bu modeller “Normal Ogive modeli” ve “lojistik modeller”dir. MTK için ilk çalışmalar Normal Ogive modeli üzerine kuruluyken zaman içerisinde yerlerini lojistik modellere bırakmıştır. Çoklu puanlama için Muraki’nin (1992) genelleştirilmiş kısmi puanlama modeli, Samejima’nın (1969) ağırlıklı tepki modeli ya da Andrich’in (1978) dereceleme ölçeği, Master’ın (1983) kısmi kredi modeli gibi modeller kullanılmaktadır (Choi, 2009; Embretson ve Reise, 2013; Hambleton ve Swaminathan, 2013). Ama bu araştırmada ikili puanlamaya yer verildiği için üç lojistik model tanıtılmaya çalışılmıştır.

Bir Parametrelili Lojistik Model (1 PLM)

Bir parametrelili modelde, bireylerin performanslarını etkileyen tek madde özelliğinin madde güçlüğü (b parametresi) olduğu varsayılmaktadır. Bu parametre dışında ayırıcılık parametresi (a parametresi), şans parametresi (c parametresi) bütün maddeler için eşit olduğu için bu modelde yer almaz. b parametresi, maddenin yanıtlayıcılarının %50'si tarafından doğru yanıtlanan yetenek düzeyini ifade etmektedir. Dolayısıyla b parametresinin artması, madde güçlüğünün artması (maddenin zorlaşması); azalması ise madde güçlüğünün azalması (maddenin kolaylaşması) anlamına gelmektedir. 1 PLM temel eşitliği aşağıda verilmiştir (Embretson ve Reise, 2013).

$$P_i(\theta) = \frac{e^{D(\theta-b_i)}}{1+e^{D(\theta-b_i)}} \quad i=1,2,3, \dots, n$$

θ : i kişinin yetenek düzeyi

b_i : i maddesinin güçlüğü

D : 1.7 değerindeki sabit bir sayı

e : 2,718 değerindeki sabit bir sayı

Rasch modeli, 1 PLM modelinin özel bir türüdür. Rasch modelinde aynı eşitlik kullanılmasına rağmen ayırıcılık parametresinin değeri “1” olarak kabul edilir (Crocker ve Algina, 1986).

İki Parametrelili Lojistik Model (2 PLM)

2 PLM'nin özelliği, bireylerin maddeyi yanıtlama olasılıklarında b parametresinin yanı sıra a parametresinin de kullanılmasıdır. 1 PLM' de olduğu gibi şans parametresi bu modele dahil edilmez ve sıfır olarak kabul edilir (Hambleton ve Swaminathan, 2013). 2 PLM temel eşitliği aşağıda verilmiştir (Embretson ve Reise, 2013).

$$P(\theta_j) = \frac{e^{Da(\theta_j-b_i)}}{1 + e^{Da(\theta_j-b_i)}}$$

- θ_j : j kişinin yetenek düzeyi
 a_i : i maddesinin ayırtıcılık parametresi
 b_i : i maddesinin güçlüğü
 D : 1.7 değerindeki sabit bir sayı
 e : 2,718 değerindeki sabit bir sayı

Üç Parametrelili Lojistik Model (3 PLM)

Geleneksel ölçme ve değerlendirme tekniklerinden ister doğru-yanlış tekniği ister öncüller ve ifadelerin eşit sayıda olarak hazırlanmış eşleştirme tipi testler isterse de çoktan seçmeli testler olsun şans ile doğru cevap işaretlenebilir. Bu nedenle kişilerin gerçek yeteneklerini belirlemek güçleşir. Üç parametrelili lojistik modelde bu gibi durumlar düşünülerek modele şans başarısı parametresi de eklenmiştir. Birnbaum'un (1968) geliştirdiği bu lojistik modellerden biri olan 3 PLM'de c parametresi şansla doğru yanıt bulma olasılığı aşağıdaki eşitlikte yer almıştır.

$$P(\theta_j) = c_i + (1 - c_i) \frac{e^{Da(\theta_j - b_i)}}{1 + e^{Da(\theta_j - b_i)}}$$

- θ_j : j kişinin yetenek düzeyi
 a_i : i maddesinin ayırtıcılık parametresi
 b_i : i maddesinin güçlüğü
 c_i : i maddesinin şans parametresi
 D : 1.7 değerindeki sabit bir sayı
 e : 2,718 değerindeki sabit bir sayı

Yukarıda açıklanan modeller ile MTK'da test geliştirme, madde bankalarının hazırlanması, değişen madde fonksiyonları, bireylere uyarlanmış testler, test eşitleme ve ölçekleme

alışmaları tasarlanmaktadır. Madde tepki kuramına dayalı eşitleme işlemi belirli aşamalar ile tanımlanmıştır. Kolen’e (1988) ve Hambleton’e (1985) göre bu aşamalar:

1. Testin yapısına ve testi alacak gruba uygun bir eşitleme deseninin seçilmesi.
2. Uygun bir madde tepki kuramı modelinin seçilmesi. Dikey eşitleme yollarında bir parametrelili model uygun değilken iki ve üç parametrelili model, yatay eşitleme için uygun olabilmektedir.
3. Yetenek ve madde parametrelerinin ortak bir ölçeğe dönüştürülmesi. Bu aşama, eşitleme sabitinin belirlenmesi amacıyla önemlidir.
4. Test puanlarının sunulacağı ölçeğe karar verilmesi.

Bu araştırma için de seçilen eşdeğer olmayan gruplarla eşitleme yapıldığında farklı test formlarındaki parametrelerin, aynı MTK ölçeğine dönüştürülmesi gerekmektedir

Madde Tepki Kuramı Ölçeğinin Dönüştürümü

MTK parametre kestirim işlemlerinden sonuçlanan parametre kestirimleri, genellikle farklı MTK ölçekleri üzerindedir. Örneğin, MTK modelleri için parametreler örneklem 1’deki katılımcı örneklemin dayandığı X formu için ve örneklem 2’deki katılımcı örneklemin dayandığı Y formu için tahmin edilir ve bu iki örneklem eşit değildir. Gösterildiği gibi bilgisayar programları genellikle standart sapması 1, ortalaması 0 olan analiz edilmiş veriler olarak θ ölçeği tanımlanır. Bu durumda, her bir grup için yetenek ortalama 0 ve standart sapması 1 olacak şekilde yetenek kestirimleri yapılır. Bu nedenle MTK ölçeklerine dönüşüm gereklidir.

Test eşitlemede, parametre kestirimlerine göre ölçek dönüştürme (kalibrasyon), kendi arasında “eş zamanlı kalibrasyon” ve “eş zamanlı olmayan kalibrasyon” olmak üzere ikiye ayrılır. Eş zamanlı kalibrasyonda formların parametreleri birlikte kestirilirken eş zamanlı olmayan kalibrasyonda formlardaki parametreler ayrı ayrı kestirilir ve lineer denklem ile

aynı ölçek üzerine konumlanır. Her formdan alınan puanların a ve b parametrelerini aynı ölçeğe dönüştürmek için lineer bir denklem kullanılır. Doğrusal dönüşüm kullanarak üç parametrelili lojistik MTK modeli, form J ve form I gibi farklı ölçeklerdeki θ 'lar arasındaki ilişki şu şekilde olur;

$$\theta_{Ji} = A\theta_{Ii} + B$$

θ_{Ji} : i bireyinin J formundaki yetenek düzeyi

θ_{Ii} : i bireyinin I formundaki yetenek düzeyi

A : Dönüştürme denkleminin eğimi

B : Dönüştürme denkleminin sabiti

Bu denklem sayesinde J formunu almayan i kişinin yetenek düzeyi elde edilmiş olur. Bu süreçte yetenek düzeylerinin yanında madde parametreleri de dönüştürülür. Bunun için aşağıdaki denklemden yararlanılır:

$$a_{Jj} = \frac{a_{Ij}}{A}$$

$$b_{Jj} = Ab_{Ij} + B$$

$$c_{Jj} = c_{Ij}$$

a_{Jj} : Form J'deki j maddenin ayıricılık parametresi

a_{Ij} : Form I'deki j maddenin ayıricılık parametresi

b_{Jj} : Form J'deki j maddenin güçlük parametresi

b_{Ij} : Form I'deki j maddenin güçlük parametresi

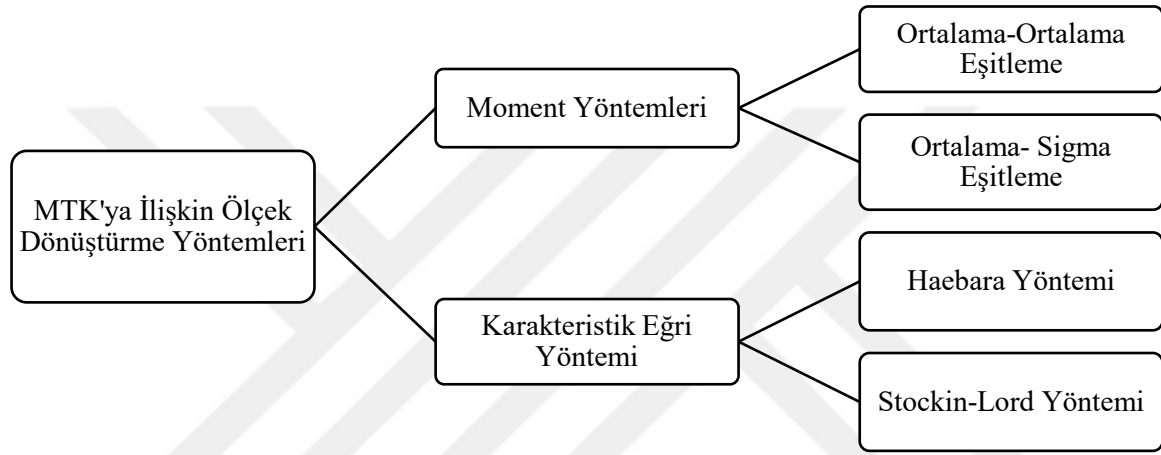
c_{Jj} : Form J'deki j maddenin şans parametresi

c_{Ij} : Form I'deki j maddenin şans parametresi

A: Dönüştürme denkleminin eğimi

B: Dönüştürme denkleminin sabiti

Form J ya da form I fark etmeksizin iki durumda da c parametresi sabittir. c parametresinin bu madde parametre dönüşümünden etkilenmediği söylenebilir. Yukarıda verilen eşitlikteki ölçek dönüştürme sabitine ve eğim değerlerine karar vermede, ortak ölçek üzerinde parametre kestirimleri aşaması önemli bir süreçtir. Dönüşümdeki bu değişkenlere karar vermede birtakım yöntemler geliştirilmiştir. Bu yöntemler Şekil 10’da sunulmuştur (Haebara, 1980; Kolen ve Brennan, 2014; Loyd ve Hoover, 1980; Marco, 1977; Stocking ve Lord,1982).



Şekil 10 Ölçek dönüştürme yöntemleri

Moment Yöntemleri

Moment yöntemi, ortalama-ortalama ve ortalama-sigma olmak üzere iki ölçek dönüştürme yöntemini içermektedir. Bu yöntemlerde madde parametreleri momentlerinden yararlanılarak A ve B değişkenleri hesaplanmaya çalışılır.

Ortalama-Ortalama Eşitleme

Bu dönüştürme yönteminde, ölçekleri dönüştürmek için ortak maddelerin madde parametre kestirimlerinin ortalaması kullanılır. Dönüşüm denklemleri için a ve b parametreleri kullanılır. Aşağıda verilen dönüşüm denklemi, form J ve form I’ dan hesaplanmış momentler

ile A ve B terimlerinin hesaplanması için sunulmuştur (Kolen ve Brennan, 2014; Loyd ve Hoover, 1980).

$$A = \frac{\mu(a_I)}{\mu(a_J)}$$

$$B = \mu(b_J) - A\mu(b_I)$$

$\mu(a_I)$: Form I'nın ayırıcılık parametrelerinin ortalaması

$\mu(a_J)$: Form J'nin ayırıcılık parametrelerinin ortalaması

$\mu(b_I)$: Form I'nın güçlük parametrelerinin ortalaması

$\mu(b_J)$: Form J'nin güçlük parametrelerinin ortalaması

A = Dönüştürme denklemindeki eğim

B = Dönüştürme denkleminde sabit

Ortalama-ortalama dönüşümünü veren denklemde öncelikle A terimi hesaplanmalıdır. A terimi, I formu uygulandıktan sonra hesaplanan a parametresinin ortalamasının J formundan hesaplanan a parametresi ortalamasına bölünmesinden elde edilir. Bu ölçek dönüştürme yöntemi ile elde edilen A terimi ikinci denkleme eklenir. B sabitini bulabilmek için J formu uygulanıp elde edilen b parametrelerinin ortalamasından, I formundaki b parametreleri ortalamasının A eğim değeri ile çarpılarak elde edilen sonuç çıkartılır. Bu sayede B sabiti de bulunmuş olur.

Ortalama-Sigma Eşitleme

Ölçekleri dönüştürmek için ortak maddelerin madde parametre kestirimlerinin ortalama ve standart sapması kullanılmaktadır. Ortalama-sigma dönüşüm yönteminde sadece b parametresi kullanılır.

$$A = \frac{\sigma(b_J)}{\sigma(b_I)}$$

$$B = \mu(b_J) - A\mu(b_I)$$

$\sigma(b_J)$: Form J'nin güçlük parametrelerinin standart sapması

$\sigma(b_I)$: Form I'nin güçlük parametrelerinin standart sapması

$\mu(b_J)$: Form J'nin güçlük parametrelerinin ortalaması

$\mu(b_I)$: Form I'nin güçlük parametrelerinin ortalaması

A = Dönüştürme denklemindeki eğim

B = Dönüştürme denkleminde sabit

Ortalama-sigma dönüşümünde A terimini bulmak ile sürece başlanır. A terimi, I formu uygulandıktan sonra hesaplanan a parametresinin standart sapmasının J formundan hesaplanan a parametresi standart sapmasına bölünmesinden elde edilir. Bu ölçek dönüştürme yöntemi ile elde edilen A terimi ikinci denkleme eklenir. B sabitini bulabilmek için J formu uygulanıp elde edilen b parametrelerinin ortalamasından, I formundaki b parametreleri ortalamasının A eğim değeri ile çarpılarak elde edilen sonuç çıkartılır. Bu sayede B sabiti de bulunmuş olur.

Karakteristik Eğri Yöntemi

Momentler yöntemlerine ilişkin potansiyel bir sorun, katılımcı puanlarının oluşturduğu yetenek aralığında a , b ve c parametrelerinin çeşitli kombinasyonlarının neredeyse aynı madde karakteristik eğrilerini oluşturmasıdır. Örneğin, iki farklı b parametresiyle kestirilen madde karakteristik eğrileri oldukça benzer olabilir. Bu durumda, ortalama/sigma yöntemi madde karakteristik eğrileri oldukça benzer olsa da b parametrelerinin kestirimlerindeki farklardan fazlasıyla etkilenebilir. Bu sorunun oluşma nedeni, şu ana kadar açıklanan yöntemlerin, parametreleri eş zamanlı kestirememesidir. Bu soruna bir yanıt olarak Haebara

(1980), tüm parametrelerin eş zamanlı kestirildiği bir yöntem geliştirmiş, ardından Stocking ve Lord (1983) bu yönteme benzer bir başka yöntem geliştirmişlerdir. Stocking ve Lord (1983) kendi geliştirdikleri yöntem ve Haebara'nın yöntemine birlikte "karakteristik eğri yöntemleri" adını vermişlerdir. Bu yöntemleri geliştirmek için I ve J yetenek ölçekleri için,

$$p_{ij}(\theta_i; a_{ji}, b_{ji}, c_{ji}) = p_{ij}(A\theta_i; B; \frac{a_{ji}}{A}, Ab_{ji} + B, c_{ji})$$

İ katılımcısı ve j maddesi için geçerlidir. Yukarıdaki eşitlikte verilen bir yetenek düzeyindeki bir katılımcının belirli bir maddeye doğru yanıt verme olasılığının, puanların raporlandığı ölçekten bağımsız olduğunu ifade eder.

Eşitlikte parametreler yerine kestirimler kullanılırsa herhangi bir A ve B üzerindeki tüm maddeler ve katılımcılar için eşitliğin sağlanacağı garanti edilemeyecektir. Bu durum, karakteristik eğri yöntemlerinin bir eksikliğidir.

Karakteristik eğri yöntemi üç aşamadan oluşmaktadır:

1-Ölçek dönüşüm sabitlerinin kestirimi

2-Test karakteristik eğrileri arasındaki farkın hesaplanması

3-Bu farklılıkların minimize edilip edilmediğinin saptanması

Bu aşamalar doğrultusunda Haebara ya da Stocking-Lord ölçek dönüşüm yöntemleri uygulanmaktadır.

Haebara Yaklaşımı

Haebara (1980) tarafından madde karakteristik eğrileri arasındaki farkı ifade etmek için kullanılan fonksiyon, belirli bir yetenek düzeyinde bulunan katılımcıların her biri için madde karakteristik eğrileri farklarının kareleri toplamına eşdeğerdir. Verilen bir θ_i için maddeler üzerinde farkların kareleri toplamı şu şekilde gösterilebilir:

$$H_{diff}(\theta_i) = \sum_{j=V} \left[p_{ij}(\theta_{ji}; \hat{a}_{jj}, \hat{b}_{jj}, \hat{c}_{jj}) - p_{ij} \left(\theta_{ji}; \frac{\hat{a}_{ij}}{A}, A\hat{b}_{ij} + B, \hat{c}_{ij} \right) \right]^2$$

A : Dönüştürme denklemindeki eğim

B : Dönüştürme denkleminde sabit

$p_{ij}(\theta_{ji}; \hat{a}_{jj}, \hat{b}_{jj}, \hat{c}_{jj})$: Madde karakteristik eğrisi

$p_{ij} \left(\theta_{ji}; \frac{\hat{a}_{ij}}{A}, A\hat{b}_{ij} + B, \hat{c}_{ij} \right)$: Kestirilen madde karakteristik eğrisi

$H_{diff}(\theta_i)$: Verilen bir θ_i yetenek düzeyi için testteki her bir madde karakteristik eğrilerinin farkların karelerinin toplamı

Bu eşitlikte, iki ölçekteki her bir madde karakteristik eğrisi arasındaki farkın karesi alınmış ve toplanmıştır. Ardından H_{diff} katılımcılar üzerinden toplanmıştır. Kestirim sürecinde, aşağıdaki kritere göre A ve B 'yi minimum yapan değer hesaplanır:

$$H_{crit} = \sum_i H_{diff}(\theta_{ji})$$

Stocking ve Lord Yaklaşımı

Haebara yaklaşımın aksine, Stocking ve Lord (1983), kareleri alınmış farkların toplamalarını maddeler üzerinde hesaplarlar.

$$SL_{diff}(\theta_i) = \left[\sum_{j=V} p_{ij}(\theta_{ji}; \hat{a}_{jj}, \hat{b}_{jj}, \hat{c}_{jj}) - \sum_{j=V} p_{ij} \left(\theta_{ji}; \frac{\hat{a}_{ij}}{A}, A\hat{b}_{ij} + B, \hat{c}_{ij} \right) \right]^2$$

A = Dönüştürme denklemindeki eğim

B = Dönüştürme denkleminde sabit

$p_{ij}(\theta_{ji}; \hat{a}_{jj}, \hat{b}_{jj}, \hat{c}_{jj})$: Madde karakteristik eğrisi

$p_{ij}(\theta_{ji}; \frac{\hat{a}_{ij}}{A}, A\hat{b}_{ij} + B, \hat{c}_{ij})$: Kestirilen madde karakteristik eğrisi

$SLdiff(\theta_i)$: Her bir θ_i yetenek düzeyi için testteki toplam tüm madde karakteristik eğrilerinin toplamlarının farkının karesi

Stocking ve Lord'un (1983) yaklaşımında, her bir parametre kestirimi seti için toplamlar kareler alınmadan önce yapılır.

$$\tau(\theta_{ji}) = \sum_{j=V} p_{ij}(\theta_i)$$

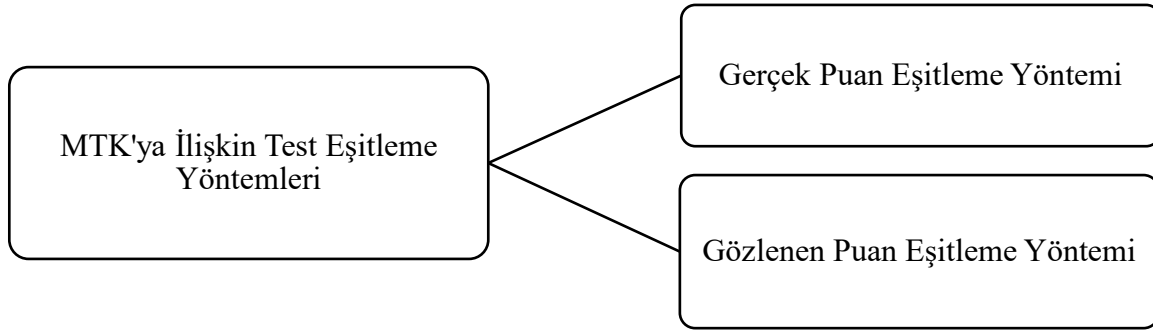
MTK'da test karakteristik eğrisi olarak adlandırılmaktadır. Dolayısıyla, $SLdiff(\theta_i)$ ifadesi, verilen bir (θ_i) düzeyinde madde karakteristik eğrileri arasındaki farkların karelerinin toplamına eşittir. Kestirim süreci, A ve B'yi minimum yapan aşağıdaki kriterin bulunmasını içerir.

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

Yukarıdaki eşitlik toplam katılımcılar üzerinden hesaplanmıştır.

MTK'ya Dayalı Test Eşitleme Yöntemleri

MTK bünyesinde gerçekleşen test eşitleme sürecine başlamadan önce moment yöntemleri ya da karakteristik eğri dönüştürme yöntemleri ile formlardan elde edilen puanlar aynı ölçeğe dönüştürülmüş olur. Bundan sonraki süreç, MTK test eşitleme yöntemlerinin denenmesidir. MTK eşitleme yöntemleri, Şekil 11'de belirtildiği üzere ikiye ayrılmaktadır. Sunulan şekilden sonra da bu eşitleme yöntemleri tanıtılmaya çalışılmıştır.



Şekil 11 Madde Tepki Kuramına dayalı test eşitleme yöntemleri

Gerçek Puan Eşitleme

Madde parametreleri aynı ölçeğe dönüştürüldükten sonra MTK gerçek puan eşitleme, X formu ve Y formu üzerindeki doğru cevap puanlarını ilişkilendirmek için kullanılabilir. Bu süreçte belirli bir θ ile ilgili bir form üzerindeki gerçek puanın, bu θ ile ilgili diğer bir form üzerindeki gerçek puanlara eşdeğer olmasına dikkat edilir.

Gerçek puan eşitleme süreci şu şekilde izlenmelidir.

- 1-Form X' deki bir τ_x gerçek puanı belirle
- 2-Belirlenen gerçek puana karşılık gelen θ_i değerini bul
- 3-Bu θ_i 'ye karşılık gelen form Y' deki gerçek puanı bul

Bireylerin gerçek puan ilişkisi asla bilinemez. Çünkü onlar parametredir. Uygulamada gerçek puan eşitleme ilişkisi, form X' deki doğru cevap gözlenen puanları form Y' deki doğru cevap gözlenen puanlara dönüştürmede kullanılır. Uygulamada MTK gerçek puan eşitleme, madde parametreleri kestirimlerinden bir kestirilmiş gerçek puan ilişkisi üretilerek kullanılır. Daha sonra kestirilen gerçek puan dönüşümü, gözlenen puanlar için uygulanır. Gerçek puan eşitlemede non-lineer fonksiyonların kökenini bulmak için Newton-Raphson yöntemi kullanılmaktadır. Bu yöntem iteratif bir süreci içermektedir. Bu yöntemde en önemli nokta, ilk değer seçimidir. Çünkü yanlış bir seçim hatalı çözümlere neden olmaktadır.

Gözlenen Puan Eşitleme

MTK gözlenen puan eşitleme, her bir form üzerindeki gözlenen doğru madde sayısı dağılım kestirimidir. X formu için bileşik binom dağılımı, verilen yetenek seviyesinde katılımcılar için gözlenen doğru yanıtlanmış madde puanı dağılımını üretmek için kullanılır.

Sunulan bilgiler ışığında çeşitli desen ve kuramlara yönelik yöntemlerle testler eşitlenmeye çalışılmaktadır. Bu test eşitleme çalışmalarında, PISA 2012 matematik alt testine yönelik hazırlanmış 13 kitapçık arasından seçilen iki kitapçık ile PISA sonuçları eşitlenmek istenmiştir.

Test Eşitlemede Hata

Eşitlemenin standart hatası, eşitleme işlemlerinin tekrarları üzerindeki puan benzerliklerinin standart sapmasıdır. Eşitleme hatası, random ve sistematik hata olmak üzere ikiye ayrılır. Random eşitleme hatası, eşitleme ilişkilerinin tahminini içeren eşitleme parametreleri kullanıldığı zaman ortaya çıkar. Random hata, geniş örneklem büyüklüğü kullanılarak ve hatayı azaltacak eşitleme tasarımı seçilerek kontrol altına alınabilir. Sistematik eşitleme hatası, sayıltıların ve eşitleme koşullarının ihlal edilmesinden kaynaklanır. Random hata genellikle eşitlemenin standart hatası olarak belirtilir. Sistematik hata miktarını belirlemek için belirli bir yöntem yoktur.

Araştırmanın Amacı

Bu araştırmada ortak maddeli denk olmayan gruplar deseni kullanılarak Klasik Test Kuramı (KTK) ve Madde Tepki Kuramına (MTK) dayalı eşitleme yöntemleri ile PISA 2012'nin kitapçık 1 ve kitapçık 3'teki matematik puanlarını eşitlemek ve kullanılan eşitleme yöntemlerinden en uygun olanını belirlemek amaçlanmıştır.

Problem İfadesi

Ortak maddeli denk olmayan gruplar deseni kullanılarak KTK ve MTK eşitleme yöntemlerinin PISA 2012 matematik testi verileri ile karşılaştırılarak en uygun eşitleme yöntemi hangisidir?

Araştırma Soruları

Araştırma için oluşturulmuş problem ifadesi dikkate alınarak aşağıdaki sorulara cevap aranmıştır.

1-Klasik Test Kuramı'nda yer alan Tucker eşitleme yöntemi, Levine gerçek puan, Levine gözlenen puan, konjenerik ve Braun-Holland doğrusal eşitleme yöntemleri kullanılarak PISA 2012'ye ait farklı kitapçıklardan elde edilen eşitlenmiş puanlar için hangi eşitleme yöntemi en az eşitleme hatası içermektedir?

2- Klasik Test Kuramı'nda yer alan frekans kestirim eşit yüzdelikli eşitleme yöntemleri kullanılarak PISA 2012'ye ait farklı kitapçıklardan elde edilen eşitlenmiş puanlar için hangi eşitleme yöntemi en az eşitleme hatası içermektedir?

3-Madde Tepki Kuramı'na dayalı eşitleme yöntemlerinden gerçek ve gözlenen puan eşitleme yöntemleri kullanılarak PISA 2012'ye ait farklı kitapçıklardan elde edilen eşitlenmiş puanlar için hangi eşitleme yöntemi az eşitleme hatası içermektedir?

4- Klasik Test Kuramı'nda en az eşitleme hatasına sahip olan eşitleme yöntemi ile Madde Tepki Kuramı'ndan en az eşitleme hatasına sahip olan eşitleme yöntemi karşılaştırıldığında hangi kurama ait eşitleme yöntemi en az eşitleme hatası içermektedir?

Araştırmanın Önemi

Farklı test formlarından alınan puanların karşılaştırılabilmesi ve birbiri yerine kullanılması, test eşitleme çalışmalarının yapılması için bir gerekçe sunmaktadır. PISA gibi geniş ölçekli sınavlarda aynı anda farklı formların kullanıldığı uygulamalarda başarı durumların tespiti,

başarı sıralamasının doğru yapılması için test eşitleme çalışmalarına vurgu yapılmaktadır. Araştırma verisi olarak seçilen PISA 2012 verileri, ölçek puan dönüşümleri yapılarak eşitleme yapılmaktadır. 13 kitapçık için elde edilen puanlar ortak maddelerce eşitlenmeye çalışılmaktadır. PISA 2012 verilerine yönelik eşitleme sürecinde MTK bünyesinde yetenek düzeyleri kestirilip, aynı yıl verileri için Rasch model kullanılmaktadır (OECD, 2014). PISA 2012 verileri, PISA 2003, 2006 ve 2009 verileri ile puanların karşılaştırılmasına olanak sağlamak için bir bağlantı (linking) ölçeği hazırlanmıştır. Özellikle PISA verilerinin ülke başarı durumlarının karşılaştırıldığı uygulamalardaki eşitleme yöntemlerine karar verilmesi ve en az hataya sahip olan eşitleme yönteminin belirlenmesi kritik bir süreçtir. Süreçte kullanılan veri yapısına göre en uygun eşitleme yönteminin belirlenmeli ve en az hata ile eşitleme sürecinin tamamlanması gerekmektedir. Literatür incelendiğinde en uygun eşitleme yöntemine ilişkin ortak bir görüş bulunmamaktadır. Kullanılan desen, ortak madde oranı, verilerin simülasyon ya da gerçek veri olma durumları, örneklem büyüklüğü ve bu örneklemin dağılımı gibi koşullarda en az hata ile eşitleme yöntemi değişkenlik göstermektedir. Literatürde farklı örneklem büyüklükleri ve gerçek veri kullanılan araştırmaların genellikle KTK bünyesinde eşitleme çalışmalarının (Von Davier ve Kong, 2005; Skaggs, 2005; Wang, Lee, Brennan ve Kolen, 2008; Tan, 2015; Özdemir, 2017; Kazanç 2019) yer aldığı görülmüştür. Kuramdaki varsayımların ve eşitleme koşulların sağlanmasının daha zor olduğu için simülasyon verisi kullanılan ya da daha büyük örneklemle çalışılan eşitleme çalışmalarının (Kilmen, 2010; Brossman ve Lee, 2013; Uysal, 2014; Gündüz, 2015; Yurtçu, 2018) ise MTK kapsamında yapılmıştır. KTK ve MTK'ya dayalı eşitleme çalışmaları ise oldukça sınırlıdır. Bu araştırma PISA 2012 raporunda belirtilen Rasch modeli kullanılmayıp 3PLM altında inceleme yapması, KTK ve MTK kapsamındaki eşitleme yöntemlerinin detaylı olarak ele alınıp eşitlemiş puanların en uygun olanın belirlendiği, gerçek veri kullanıldığı bir çalışma olarak alana katkı sağlayacağı düşünülmektedir.

Varsayımlar

1. Sınavı katılan bütün 15 yaş grubu öğrencilerinin aynı koşullar altında kitapçıkları cevaplandırıdıkları varsayılmıştır.

Sınırlılıklar

Araştırmanın sınırlılıkları aşağıda verilmiştir.

1. Bu çalışma PISA 2012 matematik alt test puanları ile sınırlıdır.
2. PISA 2012 için hazırlanmış 13 kitapçıktan sadece kitapçık 1 ve kitapçık 3 puanları eşitleme çalışmasına alınmıştır.
3. Araştırmada yer alan kitapçıklarda bazı maddeler kısmi doğru olarak puanlansa da tam doğru dışındaki tüm yanıtlar 0 olarak yeniden kodlanmıştır.
4. MTK için lojistik regresyon modelleri incelediğinde 3 PLM uygun görülüp bu model kullanılarak eşitleme gerçekleştirilmiştir.
5. Bu araştırma PISA 2012'ye katılan ülkelerden Endonezya, Türkiye, Finlandiya ve Çin (Singapur) verileri ile sınırlıdır.

BÖLÜM II

İLGİLİ ARAŞTIRMALAR

Araştırmanın bu bölümünde, Klasik Test Kuramı ve Madde Tepki Kuramı temelli eşitleme yöntemlerinin kullanıldığı çeşitli araştırmalara yer verilmiştir.

Klasik Test Kuramı Temelli Eşitleme Yöntemlerine İlişkin Çalışmalar

Klasik Test Kuramı'na dayalı eşitleme yöntemleri ortalama, doğrusal ve eşit yüzdelikli eşitlemedir. Araştırmanın bu kısmında, doğrusal ve eşit yüzdelikli eşitleme yöntemleri kullanılarak gerçekleştirilen araştırmalar raporlaştırılmıştır.

Klein ve Jarjoura (1985) tarafından yapılan araştırmada, sertifika sınavı için geliştirilmiş testlerin puanları eşitlenmeye çalışılmıştır. Bu sınav, yılda iki kez yapılmaktadır. Madde havuzunda toplam 250 madde bulunmaktadır. Aynı özelliği ölçmek için hazırlanmış üç testin ikili eşitlenmesiyle zincirleme bir süreç tasarlanmıştır. Birinci testte 60'ı ortak 105 madde, ikinci testte 60'ı ortak 101 madde, üçüncü testte ise 60'ı ortak 104 madde bulunmaktadır. Çalışmada Angoff- Levine yönteminin, test ve ortak arasında içerik dengesi eksikliği nedeniyle Tucker'dan daha duyarlı olduğu belirlenmiştir. Bu araştırmada ayrıca konjenerik yöntemin performansının, Tucker ve Angoff yöntemleri kıyaslandığında, Tucker yönteminde daha yakın güvenilirlik değerlerine sahip olduğu tespit edilmiştir. Araştırmacılar, test ile ortak arasındaki kovaryansın ikili uygulamalarda giderek azaldığını ortaya koyan bir açıklama yapmıştır.

Kolen (1985), Hanson, Zeng ve Kolen (1993)'ün yaptıkları çalışmada, 30 maddesi ortak madde olan 125 çoktan seçmeli madde ile delta yöntemini kullanarak Tucker eşitlemenin standart hatasını üretmek istemişlerdir. Bu araştırmada ayrıca normallik varsayımı

sağlanarak ve sağlanmadan elde edilen standart hatalar karşılaştırılmıştır. Normallik varsayımı sağlamadan elde edilen standart hataların daha doğru sonuçlar verdiği sonucuna ulaşılmıştır.

Cope (1987), eşdeğer olmayan gruplar ortak maddeli desen kullanarak lineer eşitleme yöntemlerinden Angoff V, Tucker ve Levine yöntemlerinin güvenilirlik katsayılarını karşılaştıran bir çalışma gerçekleştirmiştir. Araştırmada kullanılan veriler, sertifika programı sınavı puanıdır. Bu puanlar çoktan seçmeli test formatında hazırlanmış olup 0 ile 280 arasında ham puan aralığındadır. Tucker yöntemi, iki formu alan grup için ortak test puanları üzerinden, test form puanlarının lineer regresyonunu tanımlayan tekli bağlantıyı test eden varsayımı sağlayamadığından en kötü işleyen eşitleme yöntemi olarak bulunmuştur. Diğer iki yöntem arasında ise en yüksek güvenilirlik değeri, Angoff V yöntemi ile elde edilmiştir.

Woodruff (1989), ortak maddeli eşdeğer olmayan grup deseni ile Tucker, Angoff-Levine ve konjenerik eşitleme yöntemlerini karşılaştırmıştır. Bu araştırmada 100 ortak, toplamda 300 madde yazmış ve iki forma da uygulanmıştır. Araştırma iki grubun ortak test üzerinde ortalama ya da varyans değerleri arasında fark bulunursa Tucker yönteminin uygun olmayacağı hipotezi ile başlanmıştır. Ayrıca bu yöntemlerin hangisinin daha uygun olacağını belirlemek için minimum eşitleme hatası kriteri yerine dolaylı olarak güvenilirlik katsayısını baz alınmıştır. Test ve ortak maddeler arasındaki içerik dengesinde Angoff-Levine yönteminin Tucker yönteminden daha hassas olduğunu tespit edilmiştir. Gruplar arasındaki yetenek farklılıkları fazla olmadığından konjenerik eşitleme yönteminden elde edilen güvenilirlik katsayısı en düşüktür. Tucker ve Angoff-Levine yöntemlerinin konjenerik yöntemle kıyasla daha uygun eşitleme yöntemleri olduğu ve bu yöntemlerin uygulamalarının da daha kolay olduğu belirtilmiştir.

Hanson, Zeng ve Kolen (1993), Levine gözlenen ve Levine gerçek puan doğrusal eşitleme yöntemleri arasındaki standart hata farkını delta yöntemi ile kestirmeye çalışmıştır. Bootstrap yöntemine (20.000 bootstrap) dayalı bir standart hata hesaplanmıştır. Bu araştırmada Kolen'in (1985) çalışmasındaki maddeler kullanılmıştır. Bu veri seti, 30 maddesi ortak madde olacak şekilde 125 maddeden oluşmaktadır. Yeni form (X) 773 kişiye, eski form da (Y) 795 kişiye uygulanmıştır. Ortak maddeli eşdeğer olmayan gruplar deseni kullanılarak yeni form eski forma eşitlenmeye çalışılmıştır. Bu araştırmada eşitleme yöntemlerinden hangisinin etkili olduğunun tespiti dışında, bir de grupların normal dağılıp dağılmadığı durumunun test puanlarına etkisi incelenmek istenmiştir. Dağılımlar incelendiğinde, çarpık

bir dağılım elde edildiği için önceden incelenmesi amaçlanan normallik sağlanması ve sağlanamaması durumları karşılaştırılamamıştır. Eşitleme yöntemleri kıyaslandığında Levine gerçek puan eşitleme yönteminde daha az hata bulunduğu raporlanmıştır.

Livingstone (1993), ortak maddeli eşdeğer grup desenini kullanarak “25”, “50”, “100” ve “200” katılımcı ile log-lineer ön düzgünleştirme yöntemlerini karşılaştırmıştır. Bu katılımcılara, lisedeyken üniversite derslerini alma ve üniversiteye geçince bu dersleri saydırabilecekleri 24’ü ortak olmak üzere 58 maddenin yer aldığı AP History testi uygulanmıştır. Ön düzgünleştirme yönteminin, standart hatayı küçük örneklem için azalttığı ortaya konulmuştur. Ön düzgünleştirme kullanılarak elde edilen eşitleme hatasının, örneklem büyüklükleri arttıkça düzgünleştirme yapılmayan eşitleme kadar etkili olduğu bulunmuştur.

Hanson, Zeng ve Colton (1994) araştırmalarında, beş ACT (American College Testing) değerlendirme testi için düzgünleştirilmemiş, ön ve son düzgünleştirme yöntemleri ile eşitleme yapılmamış puanları incelemiştir. Klasik gerçek puan teorisi ile Monte Carlo simülasyon tekniğini kullanarak tekrar örneklem yöntemlerinin eşitleme doğrulukları incelenmiştir. Simülasyon ile 100’den 3000’e kadar olan örneklem büyüklükleri ile random grup deseni kullanılmıştır. Küçük grup ile düzgünleştirmenin manidar bir şekilde EYE’yi geliştirdiği bulunsu da açıkça tercih edilebilir ön ya da son düzgünleştirme yönteminden bahsedilmemiştir.

Skaggs (2005), Amerika’da lise düzeyindeki öğrenciler için GED (Genel eğitimin gelişmesi için testler) formunun 2002 verilerini kullanmıştır. 110000 katılımcıya uygulanan A formunun, 112996 kişiye uygulanan B formuna göre daha kolay olduğu belirtilmiş ve bu amaçla puanlar arasındaki adaletsizliği gidermek için ortak maddeler yardımı ile puanlar eşitlenmeye çalışılmıştır. Skaggs (2005) çalışmada ortalama eşitleme, doğrusal eşitleme, düzgünleştirilmemiş EYE, log-lineer ön düzgünleştirmenin 6 momentini kullanmıştır. Oldukça kalabalık bir katılımcı grubu olmasına rağmen yeniden örneklem yoluna gidilerek “25”, “50”, “75”, “100”, “150” ve “200” gruplar oluşturulmuştur. Bu çalışmanın en büyük örneklem büyüklüğü olarak 200 kullanıldığında ham puan ölçeğinin geniş bölümünde önemli büyüklükte eşitleme hatası üretilmiştir. Z puan ölçeğine dönüştürülmüş puanlar incelendiğinde Z puanı 0 ile 1.0 arasında değişen ham puanlar için doğrusal eşitleme en uygun yöntem olarak seçilirken, Z puanı 1.0’ın üstündeki değerler için EYE en uygun eşitleme yöntemi olarak seçilmiştir. Yüksek dereceli (C:6, C:7 gibi) log-lineer EYE’de kötü

değerler üretilmiştir. C:2 derecesi log-lineer için uygun görülmüştür. 25 kişilik katılımcı grubu için eşitleme yapmanın doğru olmadığına karar verilmiş ve bunun yerine özdeş eşitlemenin daha uygun olacağı belirtilmiştir. Örneklem büyüklüğü arttıkça eşitleme hatasının azaldığı tespit edilmiştir.

Von Davier ve Kong (2005), ortak maddeli eşdeğer olmayan desen kullanarak iki bölgede uygulanmış geniş kitleli test programının İlkbahar 2001 dönemi ve Kış 2000 dönemi puanlarını eşitlemek istemiştir. Araştırmada 35'i dış ankor olacak şekilde 78 madde kullanılmış ve toplamda 10634 kişiye uygulanmıştır. Araştırmada Tucker, Levine gözlenen ve zincirleme doğrusal eşitleme yöntemleri kullanılmıştır. Ortak puanların standartlaştırılmış ortalama farklılıklarının iki grupta da 0.25'ten az olduğu ve iki grupta da ortak maddelerin varyans oranlarının 0,80 ve üzerinde değer aldığı Tucker yöntemi en uygun eşitleme yöntemi olarak seçilmiştir.

Wang, Lee, Brennan ve Kolen (2008) tarafından yapılan çalışmada, ortak maddeli eşdeğer olmayan grup deseni kullanılarak matematik testi sonuçları eşitlenmeye çalışılmıştır. Araştırmadaki ana amaç, KTK bünyesindeki EYE yöntemlerinden frekans kestirim yöntemi ile zincirleme EYE yöntemini karşılaştırmak olsa da süreçte, doğrusal eşitleme yöntemlerinden Tucker, zincirleme doğrusal yöntemleri, EYE'den ise Braun-Holland eşitleme yöntemlerinin en uygun olanı da tespit edilmeye çalışılmıştır. KTK'nın dışında, eşitleme koşullarını sağlayabilmek için IRT'den de yararlanılmıştır. Bu çalışmada dört form kullanılmıştır. Birinci formda 20'si ortak olacak şekilde 60 madde, ikinci formda 12'si ortak olacak şekilde 60 madde, üçüncü formda 40'i ortak olacak şekilde 120 madde, dördüncü formda 24'ü ortak olacak şekilde 120 madde bulunmaktadır. Her form, 3000 kişi tarafından yanıtlanmıştır. Bu dört form için kullanılan eşitleme yöntemlerinin en iyisi için bileşenlerinin yanlılık ve standart hatalarının olduğu RMSE değeri kullanılmıştır. Ölçek puanları için değer büyüdükçe Tucker yönteminin daha az yanlılık içerdiği bulunmuştur. Doğrusal yöntemler arasında en az yanlı sonucun, zincirleme doğrusal yöntemine ait olduğu görülmüştür. Eşit yüzdelikli eşitleme yöntemleri için düşük ölçek puanlarda ise Braun-Holland yönteminin daha yanlı sonuçlar verdiği, ölçek bütününde ise yanlılığın en fazla olduğu yöntemin EYE frekans kestirim yöntemi olduğu görülmüştür. Standart hatalar karşılaştırıldığında zincirleme eşitleme yönteminin EYE için, zincirleme eşitleme yönteminin de doğrusal eşitleme için uygun olduğu bulunmuştur. Doğrusal yöntemler ve eşit

yüzdelikli eşitlemeler kıyaslandığında, bu araştırma için standart hatalar doğrusal eşitleme yöntemlerinde daha az olduğu görülmüştür.

Kane, Mroch, Suh ve Ripkey (2009), araştırmalarında yüksek riskli test programları için kullanılan testin iki formundaki puanların eşitlenmesini istemişlerdir. İç ankor maddeler kullanılarak ortak maddeli eşdeğer olmayan grup deseni ile eşitleme çalışması gerçekleştirilmiştir. Bu desenle beş lineer eşitleme yöntemi denenmiştir. Bu yöntemler; Tucker, Levine gerçek puan, Levine gözlenen puan, zincirleme lineer, Tucker-Live (Angoff V)'dir. Araştırmada Levine yöntemlerinin zincirleme yöntemlerinden daha çok düzeltme gerektirdiği, zincirleme doğrusal eşitlemenin ise Tucker yönteminden daha çok düzeltme gerektirdiği bulunmuştur. Testi alan iki grup arasındaki farklılaşmanın az olmasından kaynaklı Tucker yöntemi, en iyi çalışan eşitleme yöntemi olarak seçilmiştir.

Kahraman (2012), yaptığı yüksek lisans tezinde, düzgünleştirilmiş puanların eşitleme hatasına etkisini araştırmıştır. Araştırmasında eşdeğer gruplar (rastgele) deseni kullanarak Hacettepe Üniversitesi İngilizce Yeterlik Sınavı 2010 ve 2011 dönemi verileri eşitlenmeye çalışılmıştır. Bu sınavda dinleme becerisi 14 madde ile, temel bileşenler (kelime bilgisi, dilbilgisi) 28 madde ile, okuduğunu anlama ise 20 madde ile ölçülmeye çalışılmıştır. Bu sınav 810 kişiye uygulanmıştır. Araştırmada eşit yüzdelikli eşitleme yöntemi, log-lineer ve kübik spline yöntemleri arasında en az hata içeren yöntem seçilmeye çalışılmıştır. Log lineer için sınavın üç alt boyutunda (dinleme becerisi, temel bileşenler, okuduğunu anlama) C:9 derecesinde model veri uyumu sağlanırken son düzgünleştirme için S:0,01 de model veri uyumu sağlanmıştır. Çalışmada kullanılan üç eşitleme yöntemi arasında en az hata içeren yöntem ise log-lineer ön düzgünleştirme yöntemi olmuştur.

Eşdeğer olmayan ortak maddeli deseni kullanılarak yapılan araştırmaların çoğunda, doğrusal gözlenen puan eşitlemenin standart hatasının çoklu normallik varsayımından elde edildiği görülmüştür. Zu ve Yuan (2012), çalışmalarında normallik varsayımının ihlal edildiği eşitleme koşullarını incelemek istemişlerdir. Üç geniş ölçekli veri seti kullanılmıştır. Bu verilerden yeniden örneklem oluşturulmuştur. Araştırmada “50”, “100”, “250” ve “500” kişilik örneklem büyüklükleri kullanılmıştır. Eşitleme yöntemleri olarak zincirleme eşitleme, Tucker ve Levine gözlenen puan eşitleme yöntemleri kullanılmıştır. Testler, aday ilköğretim öğretmenleri için oluşturulmuş olup üç kısımdan oluşmaktadır. Birinci bölümde program planlama, öğretim yöntemleri, öğrenci öğrenmelerini değerlendirme bilgi alanına; ikinci bölümde okuma/ dil bilgisi, matematik, sosyal bilgiler ve fen bilimleri bilgi alanına;

üçüncü bölümde ise ikincil okulda (ülkemizdeki ortaokul kademesi olarak düşünülebilir) yeni başlayan öğretmenler için sosyal araştırmalara yönelik içeriklere bilgi ve beceri alanlarına hitap eden sorular bulunmaktadır. Hazırlanan testlerde 35 madde, ortak madde olarak kullanmıştır. Eski form ve yeni formdaki örneklem büyüklükleri benzer olacak şekilde ayarlandığı için sentetik evren ağırlığı için $w_1=0.5$ ve $w_2=0.5$ olacak şekilde seçilmiştir. En az eşitleme hatası 100 kişilik örneklemde elde edilmiştir. En uygun eşitleme yöntemi sırasıyla Tucker, zincirleme lineer ve Levine yöntemleridir.

Kelecioğlu ve Öztürk Gübeş (2013), yaptıkları çalışmada 2009 Öğrenci Başarılarının Belirlenmesi Sınavının (ÖBBS) sosyal bilimler alt testlerinden, C ve A kitapçıklarının puanlarını eşitlemek istemişlerdir. Doğrusal eşitleme ve üç eşit yüzdelikli eşitleme (düzgünleştirilmemiş, ön-düzgünleştirilmiş ve son-düzgünleştirilmiş eşit yüzdelikli eşitleme) yöntemleri kullanılmıştır. Eşitleme yöntemlerinden en uygun olanı, RMSD ve bootstrap standart hatalarının ortalaması (BSHO) ölçütleri kullanılarak karar verilmiştir. Doğrusal eşitleme yöntemi sonucu olarak C formunun A formundan daha kolay olduğu belirlenmiştir. Ön düzgünleştirme için C:5 derecesinde, son düzgünleştirme için ise S:0.01 düzeyinde model uyumu yakalanmıştır. Hatalar incelendiğinde en fazla hata, düzgünleştirilmemiş EYE’de rastlanırken en az hata, doğrusal eşitleme yöntemi ile elde edilmiştir.

Mutluer (2013), KTK’ya dayalı doğrusal ve EYE yöntemlerini karşılaştırmak istemiştir. Bu süreçte ALES 2011 ilkbahar ve sonbahar dönemlerinin her ikisine de katılan 21860 kişiden %10’luk bir örneklem alınmıştır. Tek grup desen kullanılarak ALES sayısal, sözel ve eşit aralık puanları eşitlenmiştir. Süreçte eşitleme hatası, WMSE değeri bakılarak incelenmiş ve en uygun eşitleme yönteminin eşit yüzdelikli eşitleme olduğu bulunmuştur.

İnci (2014), yapmış olduğu araştırmasında örneklem büyüklüğünün test eşitlemeye etkisini incelemek istemiştir. Çalışmanın verisi 2011-2012 yılı SBS 8. Sınıf matematik alt testi sonuçlarıdır. 100, 500, 1500 ve 3000 kişilik örneklem grupları araştırmanın bağımsız değişkenidir. Farklı örneklem büyüklükleri için ağırlıklandırılmış hata kareleri ortalaması (WMSE) eşitleme hatası niceliksel büyüklüklerine göre karşılaştırılmıştır. 3000 kişilik örneklem büyüklüğünde eşitlenmiş puanlar ham puanlardan olabildiğinde yüksek değerlere dönüşürken en az hata değerini doğrusal eşitleme yöntemi üretmiştir. Diğer örneklem büyüklükleri için ise EYE yönteminin doğrusal yöntemle göre daha az hata ürettiği bulunmuştur.

Tan (2015), çalışmasında, Gazi Üniversiteleri Eğitim Fakültesine kayıtlı 233 öğrenciye dengelenmiş tek grup deseni kullanarak iki farklı formdaki puanlarını EYE'deki beta 4 ve polinomial loglineer ön düzgünleştirme ve kübik eğri son düzgünleştirme yöntemleri ile eşitlemeye çalışmıştır. Araştırmacı, 20 hedef davranışı ölçmek için eşdeğer iki form hazırlamıştır. Polinomial loglineer ön düzgünleştirme için C düzeyleri incelendiğinde C:1 düzeyi hariç diğer C düzeylerinde eşitleme sonuçları için model uyumu sağlanırken son düzgünleştirme için de S:0.30 düzeyinde en uygun model uyumunu sağlamıştır. Bootstrap yöntemi kullanılarak eşitleme hata değerleri hesaplanmıştır. Bu çalışmada hata kavramı, momentlerin korunumu ile ilişkilendirilmiştir. Küçük örneklemelerde ön düzgünleştirme yöntemleri, kübik eğri son düzgünleştirme yöntemlerinden daha az hata içermektedir.

Pektaş ve Kılınç (2016), araştırmalarında KTK'ya dayalı ortalama eşitleme, doğrusal eşitleme ve eşit yüzdelikli eşitleme yöntemlerini kullanarak PISA 2012 matematik alt test puanlarını eşitlemek istemişlerdir. PISA 2012 uygulamasına katılan ülkelerden Türkiye örneklemini seçilmiş ve 7 ile 10 numaralı kitapçıklar eşitlenecek veriler olarak belirlenmiştir. 7 numaralı kitapçığı 370 kişi, 10 numaralı kitapçığı ise 372 kişi cevaplamıştır. 7. Kitapçıkta 12 ortak madde olacak şekilde 35 madde bulunurken 10. kitapçıkta, 12'si ortak olacak şekilde 36 madde bulunmaktadır. Kullanılan eşitleme yöntemleri sonucunda, ortalama eşitleme yönteminde ortalama momenti dışında diğer momentlerin iki form için eşitleme sonrasında eşit çıkmadığı bulunmuştur. Lineer (doğrusal) eşitleme yöntemi kullanılarak dört momente ait eşitlenmiş puanlar ile eski form puanları karşılaştırıldığında ortalama ve standart sapmaları aynı değerleri verirken diğer iki momentin farklı değerler aldığı bulunmuştur. Eşit yüzdelikli eşitleme yönteminde ise eşitlenmiş ve eski formdaki dört momentin tamamının eşit çıktığı görülmüştür. Bu üç KTK yöntemi kendi içinde karşılaştırıldığında puan dağılımlarını en doğru şekilde eşit yüzdelikli eşitleme yönteminin verdiği sonucuna ulaşılmıştır.

Özdemir (2017), ortak maddeli eşdeğer olmayan grup desenini kullanarak TIMSS 2011 ve TIMSS 2007 matematik sonuçlarını eşitlenmeye çalışmıştır. Araştırmada KTK'da yer alan frekans kestirim, zincirleme EYE, Braun-Holland, Tucker, Levine ve zincirleme doğrusal eşitleme yöntemleri karşılaştırılmıştır. 2007 TIMSS 8. sınıf öğrencilerinden 323 kişiye 23'ü ortak toplamda 115 maddeli bir test ile 2011 TIMSS 8. sınıf öğrencilerinden 495 kişiye 23'ü ortak toplamda 119 maddeli bir test eşitlenmiştir. Eşitleme süreci sonunda en uygun eşitleme

yöntemi Tucker olurken en fazla eşitleme hatası içeren yöntem ise zincirleme EYE yöntemi olmuştur.

Kazanç (2019), çalışmasında, şans başarısının farklı eşitleme yöntemleri kullanarak test eşitlemedeki etkisini araştırmayı amaçlamıştır. 2015-2016 TEOG (Temel Eğitimden Ortaöğretime Geçiş Sınavı) 8 sınıf İngilizce I ve İngilizce II sınavları kullanılmıştır. Araştırmada, 22 okula giden 1681 kişinin verileri kullanılmıştır. Çalışmada eşitleme yöntemleri, şans başarısından arındırılmış ve arındırılmamış puanlarına uygulanmış ve eşitleme hataları, ağırlıklandırılmış hata kareleri ortalaması (WMSE) katsayısına göre incelenmiştir. Şans başarısından arındırılan testlerde EYE yönteminin daha çok eşitleme hatasına sahip olduğu, şans başarısı arındırılmamış testte ise EYE yönteminin daha az eşitleme hatasına sahip olduğu görülmüştür.

Karagül (2020) yapmış olduğu araştırmasında, küçük örneklerde çok kategorili puanlama maddelerinden oluşan testleri KTK bünyesindeki eşitleme yöntemlerine göre incelemiştir. KTK kapsamında tüm eşitleme yöntemleri kullanılsa da ortalama, Tucker, Braun-Holland eşitleme yöntemlerini, eşitlemenin standart hatası (ESH), yanlılık ve ortalama hata karekökü (OHKK) kavramları bağlamında değerlendirmek istemiştir. Araştırmada madde sayısı, örneklem büyüklüğü, formlar arası madde eşik parametresi farkı değişkenlerine göre belirtilen bu eşitleme yöntemlerinden uygun olanının belirlenmesi amaçlanmıştır. Simülasyon çalışması ile eşdeğer olmayan gruplar deseni kullanılmıştır. Araştırma bulgularında 100'den küçük örneklerde ortalama eşitlemenin daha az eşitleme hatasına sahip olduğu, 100'den büyük örneklerde ise tüm eşitleme yöntemlerinin benzer sonuçlar verdiği görülmüştür. 3 kategorili 5 maddeli bir veri için yanlılık, OHKK, ESH'nin daha düşük olduğu bulunmuştur. Madde eşik parametresi arttıkça OHKK'nın arttığı, maddeler arası güçlük farkı arttıkça daha düşük ESH değerinin elde edildiği; eşik parametresi farklı koşulu arttıkça yanlılığın arttığı görülmüştür. En düşük ESH, OHKK ve yanlılık değerlerinin ortalama eşitlemesi ile sağlanmıştır.

Madde Tepki Kuramına Yönelik Eşitleme Yöntemlerine İlişkin Çalışmalar

Madde Tepki Kuramı'na dayalı eşitleme yöntemleri gerçek ve gözlenen eşitlemedir. Bu eşitleme yöntemlerinin kullanılması için kalibrasyon yöntemlerine ihtiyaç duyulur.

Araştırmanın bu bölümünde hem kalibrasyon yöntemlerine hem de MTK'da kullanılan eşitleme yöntemlerine yer verilmiştir.

Karkee ve Wright (2004), araştırmalarında iki moment (ortalama-ortalama, ortalama-sigma), iki karakteristik eğri (Stocking-Lord, Haebara) ve teta regresyon kalibrasyon yöntemlerini kullanarak bir eşitleme çalışması gerçekleştirmişlerdir. 8. ve 12. sınıfta öğrenim gören 31.813 öğrenciye 78'i çoktan seçmeli madde olan İngilizce testi uygulanmıştır. Eşitlemede 3 PLM ve 1 PLM ile parametre kestirimleri kullanılmıştır. En uygun eşitlemeyi değerlendirmek için test karakteristik eğrileri, frekans dağılımı eğrileri, test bilgi fonksiyonları, ölçek puanları ortalama ve standart sapmaları ölçüt olarak belirlenmiştir. 1 PLM ölçek puan dağılımında, Stocking-Lord kalibrasyon yöntemi ile benzer ortalama ve standart sapmalar ortaya çıkmıştır. Dolayısıyla ölçek dönüşüm için Stocking-Lord yöntemi önerilmiştir. Haebara yönteminde ise performans seviyeleri, sınıflama yüzdeleri önemli görüldüğünden daha doğru sonuçlara ulaşılmıştır. 1 PLM için moment ve teta regresyon yöntemleri, dağılımın ortasında daha büyük hatalar üretmiştir. Ayrıca tüm değerlendirme kriterleri için karakteristik eğrisi yöntemleri, 3 PL ölçeğin 1 PL ölçeğe dönüştürülmesinde en iyi sonuçları vermiştir. Puan dağılımının iki ucunda ise Ortalama-Ortalama ve teta regresyon yöntemlerinin daha küçük hata içerdiği saptanmıştır. Araştırmada, karakteristik eğri yöntemlerinin, daha küçük eşitleme hataları ile moment yöntemlerine üstünlük sağladığı belirlenmiştir.

Lee ve Ban (2009), MTK'ya dayalı yetenek dönüşüm yöntemlerinden eş zamanlı kalibrasyon, Stocking-Lord, Haebara ve yeterlik dönüşümü yöntemlerini karşılaştırmıştır. Yetenek dağılımı da araştırma için diğer bir değişken olarak ele alınmıştır. Araştırma bulgularına göre farklı yetenek dağılımına sahip grupların eşitlenmesinde karakteristik eğri yöntemlerinin daha iyi çalıştığı ve Haebara yönteminin Stocking Lord yönteminden daha az hata ortaya çıkardığı bulunmuştur.

Speron (2009), ortak maddeli eşdeğer olmayan gruplar deseni ile gerçekleştirdiği araştırmasında verileri simülasyon ile üretmiştir. Bu araştırmada, MTK eşitleme sürecinde yer alan kalibrasyon süreçlerini karşılaştırmak istemiştir. İkili puanlama ile oluşturulan veriler, 2 PLM ile uyumlu olarak oluşturulup parametreleri kestirilmiştir. Eşitleme sürecinde örneklem büyüklüğü, ortak madde sayısı, ortak maddelerin testi temsil etme gücü ve yetenek dağılımları da ele alınmıştır. Araştırmada, örneklem büyüklüğü, ortak madde sayısı, ortak maddeler ile testin tümü arasındaki korelasyon arttıkça benzer yetenek dağılımına sahip

gruplarda eşitleme hatasının daha az olduğu bulunmuştur. Araştırma sonucunda karakteristik eğri yöntemlerinin, çok az bir farkla ortalama-ortalama yönteminden daha güçlü bir yöntem olduğu kanıtlanmıştır. Ortalama-sigma yöntemi ise en fazla hata üreten kalibrasyon yöntemi olarak bulunmuştur.

Kilmen (2010), araştırmasında ortak maddeli eşdeğer olmayan grup deseni kullanarak farklı örneklem büyüklükleri (500, 1000) ve benzer ya da farklı yetenek dağılımına göre kullanılan ortalama-ortalama, ortalama-sigma, Stocking Lord ve Haebara yöntemlerinin (kalibrasyon) en uygun olanını seçilmeye çalışmıştır. Araştırmada, MTK kapsamında 3 PLM kullanılmıştır. Eşitleme sürecinde en uygun kalibrasyon yönteminin seçilmesi için RMSD değeri ölçüt olarak belirlenmiştir. Araştırmada 10'u ortak olmak üzere 50 maddeli test formları kullanılmıştır. Örneklem büyüklükleri karşılaştırıldığında 1000 kişilik örnekleme daha az eşitleme hatası tespit edilmiştir. Benzer yetenek dağılımına sahip gruplara ait eşitleme hatasının, farklı yetenek dağılımlarına sahip gruplardan elde edilen RMSD değerlerine kıyasla daha düşük olduğu saptanmıştır. Kalibrasyon yöntemleri içerisinde ise en uygununun, Stocking-Lord yöntemi olduğu görülmüştür.

Brossman ve Lee (2013), MTK'ya dayalı gerçek ve gözlenen puan eşitleme yöntemlerini karşılaştırmışlardır. Ayrıca bu araştırmada formların tek boyutlu olma ve çok boyutlu olma durumları da karşılaştırılmıştır. Bu karşılaştırma için SAT verisi kullanılmıştır. Uygulanan sınavda 30 maddeli matematik testi, 50 maddeli sosyal testi, 48 maddeli fen testi kullanılmıştır. Çok boyutlu ve tek boyutlu testlerde kullanılan eşitleme yöntemlerinde EYE sonuçları ölçüt olarak kabul edilmiştir ve eşitleme hatası için MARA (Mean Absolute Difference on The Average of The Equating Functions) değerleri baz alınmıştır. Eşitleme sonucunda çok boyutlu gözlenen eşitleme yönteminin en az hata ürettiği belirlenmiştir.

Gök ve Kelecioğlu (2014), ortak maddeli eşdeğer olmayan grup desenini kullanarak bir eşitleme çalışması gerçekleştirmişlerdir. Araştırmada 2 PLM ve 3 PLM'ye uygun veriler, WINGEN3 programı ile üretilmiştir. Örneklem büyüklüğü, kalibrasyon yöntemi, yetenek dağılımı ve test uzunluğu değişkenleri test eşitleme süreci için incelenmiştir. Eşitleme yöntemleri RMSE değerlerince kıyaslanmıştır. Bu araştırmada ise 3000 kişilik örneklem grubunda, benzer yetenek dağılımına sahip grupta, ortalama-ortalama yöntemi ve 80 maddelik test kullanıldığında en az RMSE değeri elde edilmiştir.

Kabasakal (2014), çalışmasında DMF'li maddelerin testten çıkartılıp çıkartılmama durumunu, çok düzeyli MTK ve geleneksel MTK kuramı altında incelemek istemiştir. Eşitleme sürecinde ayrıca örneklem büyüklüğü, test uzunluğu, DMF etki büyüklüğü ve DMF'li maddelerin değişkenler eşitleme hatası üzerindeki etkileri incelemiştir. Simülasyonla 20 ve 40 maddeli test sonuçları 500 ve 2000 kişilik örneklem için oluşturulmuştur. Çok düzeyli MTK'ye dayalı yapılan eşitleme sürecinin, DMF için daha başarılı sonuçlar verdiği görülse de örneklem büyüklüğü ve test uzunluğu dikkate alındığında geleneksel MTK'ye dayalı eşitlemenin daha az hata içerdiği görülmüştür. DMF'li maddeler olduğunda ayrı kalibrasyon yöntemlerinin eşzamanlı kalibrasyona göre daha çok etkilendiği, DMF'li maddeler çıkartıldığında ise küçük örneklemelerde ve kısa testlerde eşitleme hatasını arttırdığı, büyük örneklem ve uzun testlerde ise eşitleme hatasını azalttığı sonucuna varılmıştır.

Uysal (2014), çalışmasında her ne kadar MTK'ya dayalı eşitleme yöntemlerini karşılaştırmak istese de araştırmasında kalibrasyon yöntemlerini incelemiştir. Simülasyon verisi ile 40 madde oluşturup ortak maddenin testteki oranını %10, %20 ve %30 olarak değiştirmiştir. Ayrıca yetenek dağılımlarını normal dağılım, sağa çarpık ve sola çarpık olarak üç farklı grupta değerlendirmek istemiştir. 3 PLM kullanılan 1000 kişilik simülasyon verisinin MTK 1000 kişilik simülasyon verisinde %30 ortak madde oranı ile en iyi eşitleme sonucunu verdiği, %10 ortak madde oranı için en fazla eşitleme hatasına sahip olduğu, sağa çarpık yetenek dağılımının ise en düşük eşitleme hatasını barındırdığı görülmüştür. MTK eşitleme yöntemleri yerine incelenen kalibrasyon yöntemlerinden Stocking-Lord'un, en az eşitleme hatasına sahip olan bir ölçek dönüştürme yöntemi olduğu sonucuna ulaşılmıştır.

Demirus (2015), araştırmasında MTK bünyesinde gerçekleştirdiği eşitlemeye, değişen madde fonksiyonu (DMF) durumunu da dâhil etmiştir. Araştırmasında ortak maddeli eşdeğer olmayan grup desenine uygun olarak hazırlanan fen testi, 8. Sınıf okuyan 1350 öğrenciye uygulanmıştır. Hazırlanan iki formda 8'i DMF'li, 6'sı DMF'siz ortak madde olmak üzere 29 madde kullanılmıştır. DMF analizleri için Mantel-Haenszel ve lojistik regresyon yöntemleri kullanılmıştır. Araştırmada ortak maddelerin erkekler lehine DMF'li olduğu, en büyük RMSD eşitleme hatasının ortalama-ortalama yöntemi, en küçük hatanın ise ortalama-sigma yöntemi ile üretildiği görülmüştür. Ortak maddeler DMF'siz olduğunda ise en büyük hatanın ortalama-sigma yönteminde, en küçük RMSD eşitleme hatasının

karakteristik eğrisi yöntemlerinde (Stocking-Lord ve Haebara) birbirine eşit olduğu bulunmuştur.

Gündüz (2015), araştırmasında MTK altında dayalı ortalama-ortalama, ortalama-standart sapma, Stocking-Lord ve Haebara ölçek dönüştürme yöntemlerini karşılaştırmayı amaçlamıştır.

Araştırmacı, 3PLM kullanarak TIMSS 2011 dördüncü sınıf düzeyine uygulanan iki kitapçıkta elde edilen puanları eşitlemeye çalışmıştır. İki matematik formunu alan toplamda 1064 Türk öğrenci, çalışma grubunu oluşturmaktadır. Testlerde 14'ü ortak olan 26 madde bulunmaktadır. Eşitleme deseni olarak ortak maddeli eşdeğer olmayan gruplar deseni kullanılmıştır. Kalibrasyon yöntemlerinden ortalama-ortalama yönteminin en az hata içeren yöntem olduğu raporlaştırılmıştır. MTK gerçek puan eşitlemesi yapılmıştır. Araştırma sonucunda Form Y'nin, Form X'e göre daha zor olduğu sonucuna ulaşılmıştır.

Aksekioglu (2017), MTK'ya dayalı test eşitleme yöntemlerini kıyaslamak istemiştir. Bunun için PISA 2012 fen okuryazarlığını ölçen kitapık 1 ve kitapık 8 puanlarını eşitlemek istemiştir. Bu süreçte, 701 kişinin verileri 2 PLM altında incelenmiştir. MTK'daki eşitleme yöntemlerinden önce kalibrasyon yöntemlerinden ortalama-standart sapma yönteminin daha uygun sonuçlar verdiği bulunmuştur. Yapılan bu araştırma sonucunda MTK gözlenen puan eşitleme yönteminin, gerçek puan eşitleme yönteminden daha iyi bir eşitleme yöntemi olduğu sonucu bulunmuştur.

Yurtçu (2018), parametrik olmayan Bayes yöntemiyle ortak değişkenlere göre yapılan test eşitleme yöntemlerini karşılaştırmak istemiştir. Yurtçu'nun (2018) araştırmasında, ortak değişken olarak cinsiyet, matematik öz yeterlilik puanları ve ortak madde puanları belirlenmiştir. Araştırmada, PISA 2012 kitapık 6 ve kitapık 5'teki Kanada ve İtalya verileri kullanılmıştır. Ortak maddeler dış ankor olarak değerlendirilmiştir. Dış ankor puanları ile yapılan eşitleme çalışmasında, ortalama-standart sapma kalibrasyon yönteminin en az hataya sahip olduğu sonucuna ulaşılmıştır. Cinsiyet ve matematik öz yeterlik puanlarına göre MTK gerçek puanlarında Stocking-Lord yöntemi ile en uygun eşitleme yapıldığı görülmüştür.

Klasik Test Kuramı ve Madde Tepki Kuramının Karşılaştırıldığı Eşitleme Yöntemlerine İlişkin Çalışmalar

Araştırmının bu kısmında da KTK ve MTK bünyesinde kullanılan eşitleme yöntemleri karşılaştırılmasının konu edildiği araştırmalara yer verilmiştir.

Kolen (1981), rastgele gruplar desenini kullanarak doğrusal, EYE ve MTK eşitleme yöntemlerini karşılaştırmıştır. Araştırmada kullanılacak veriler olarak IOWA Eğitimsel Gelişim Testleri ilişkin test bataryası verilerini kullanmıştır. Bu test bataryasının birinci formunda X-6 ve X-7 formları kullanılırken ikinci formunda Y-7 ve X-7 formları kullanılmıştır. Çalışma kapsamında 9, 10, 11, ve 12. sınıfa giden 10728 kişiye ulaşılmıştır. MTK kapsamında ele alınan eşitleme yöntemleri için 1 PLM, 2 PLM ve 3 PLM sonuçları kendi arasında kıyaslanmıştır.

Petersen, Cook ve Stocking (1983), MTK ve KTK’da yer alan eşitleme yöntemlerini karşılaştırmak istemişlerdir. Araştırmada 1974 Sonbahar dönemindeki SAT sözel ve matematik verileri kullanılmıştır. KTK’da doğrusal eşitleme yöntemlerinden Tucker, Levine eşit güvenirlik ve Levine eşit olmayan güvenirliğe bağlı eşitleme yöntemlerine yer verilmiştir. MTK ise kalibrasyon süreçleri (eş zamanlı- ayrı zamanlı) açısından incelenmiştir. SAT 1974 matematik formlarında 25’i ortak toplamda 60 madde bulunmaktadır. Sözel formlarda ise 10’u ortak toplamda 65 madde bulunmaktadır. Çalışmada 3 PLM kullanılmıştır. MTK kalibrasyon süreci için en kararlı eşitleme sonucu, 3 PLM kullanılarak eş zamanlı kalibrasyon ile sağlanmıştır. KTK temelli eşitleme yöntemlerinde ise en kararlı Levine, eşit güvenirliğe bağlı eşitleme yöntemi olmuştur.

Lord ve Wingersky (1984), orta maddeli eşdeğer olamayan grup deseni ile SAT puanlarını eşitlemek istemişlerdir. Bu süreçte MTK eşitleme yöntemleri ile EYE gözlenen puan eşitleme yöntemlerini de karşılaştırmayı hedeflemişlerdir. Eşitlemede kullanılan formlarda, 40’ı ortak olacak şekilde 85 madde bulunmaktadır. Bu formlar, 2670 kişiye uygulamıştır. Araştırmada en az hataya sahip eşitleme yöntemi, MTK gözlenen puan eşitleme ile elde edilirken en fazla hata EYE yöntemi ile elde edilmiştir.

Kelecioğlu’nun (1994) çalışması, Türkiye’de eşitleme konusunda yapılan ilk çalışma olduğu için önem arz etmektedir. ÖSS 1990, 1991 ve 1992 sınav sonuçlarının ardışıklı yıllar için eşitlenmesi amaçlanmıştır. Süreçte, ortak maddeler aracılığıyla puanlar eşitlenmeye çalışılmıştır. Ortak maddeler faktör analizi sonucu, aynı konuyu ilişkin benzer güçlüklerde

hazırlanmış maddeler eşitleme çalışmasında rol oynamıştır. 2000 kişilik bir örneklem kullanılmıştır. KTK'ya dayalı doğrusal ve EYE yöntemi, MTK için Rasch modeli ve 2 PLM ile eşitleme yapılmıştır. Yapılan eşitleme çalışmasında, Türkçe alt testinde doğrusal eşitleme, fen alt testinde EYE, sosyal bilimler ve matematik içinse Rasch modeline dayalı eşitlemenin en uygun eşitleme yöntemleri olduğu kanıtlanmıştır.

Han, Kolen ve Pohlmann (1997), araştırmalarında MTK gerçek puan eşitleme, MTK gözlenen puan eşitleme ve düzgünleştirme yapılamadan uygulanan EYE yöntemlerini karşılaştırmak istemişlerdir. Araştırmada, ACT sınavının Matematik ve İngilizce formları kullanılmıştır. Kullanılan matematik formlarının 60'ı çoktan seçmeli, 5'i de alternatif seçenekli maddelerden oluşmaktadır. Diğer İngilizce formunda ise 75'i çoktan seçmeli, 4'ü alternatif olmak üzere 79 madde bulunmaktadır. Rastgele grup deseni kullanılarak 22848 kişiden veri toplanmıştır. Araştırma sonucunda MTK gözlenen eşitleme yöntemi, MTK gerçek puan eşitleme yönteminden; MTK gerçek puan eşitleme yöntemi de EYE yönteminden daha kararlı sonuçlar oluşturmuştur.

Hagge, Liu, He, Powers, Wang ve Kolen (2011) araştırmalarında, karma test formatında puanlama yapılan test sonuçlarını KTK ve MTK'ya dayalı eşitlemek istemişlerdir. KTK'ya dayalı eşitleme yöntemi olarak EYE yöntemlerinden, frekans kestirim ve zincirleme EYE yöntemlerini kullanmıştır. MTK'ya dayalı eşitleme için ise gerçek puan eşitleme ve gözlenen puan eşitleme yöntemlerini seçmişlerdir. Bu çalışma ile seçilen dört eşitleme yönteminden en az eşitleme hatası içeren yöntem tespit edilmeye çalışılmıştır. Ayrıca süreçte kullanılan EYE yöntemlerinde düzgünleştirme yapılmaması, ön düzgünleştirme ve son düzgünleştirme süreçleri de ayrıca incelenmiştir. Düzgünleştirme sürecinde ön düzgünleştirme için C:6'da, son düzgünleştirmede ise S:0,1'de model veri uyumu sağlanmış olup son düzgünleştirme sonuçlarının daha az hata içerdiği bulunmuştur. Belirtilen dört eşitleme yöntemi kendi arasında eşitleme hataları bakımından en azdan en çoğa doğru şu şekilde sıralanmıştır: MTK gözlenen puan eşitleme yöntemi, MTK gerçek puan eşitleme yöntemi, KTK frekans kestirim yöntemi, KTK zincirleme EYE yöntemi. Düzgünleştirilmiş EYE yöntemi ile düzgünleştirilmemiş EYE sonuçları kıyaslandığında düzgünleştirme yapılmasının gerekli olduğu görülmüştür.

Liu ve Kolen (2011), ortak maddeli eşdeğer olmayan gruplar desenine uygun MTK gerçek puan eşitleme, MTK gözlenen puan eşitleme, KTK zincirleme EYE ve KTK frekans EYE yöntemlerini karşılaştırılmışlardır. AP (College Board Advanced Placement Program) testi

birinci formunda 70 çoktan seçmeli, 3 tane serbest cevaplı (0-9 arası puanlama yapılan) maddelerden oluşurken ikinci formda 98 çoktan seçmeli, 4 tane serbest cevaplı (0-9 arası puanlama yapılan) maddelerden oluşmaktadır. Her bir formu 2000 erkek 1000 kadın cevaplamıştır. EYE için ön düzgünleştirme yöntemi için C:6 derecesinde, son düzgünleştirme için S:0.1 derecesinde model veri uyumu sağlanmıştır. Kullanılan eşitleme yöntemleri yanlılık, eşitleme hatası değerlerine bakılarak en uygun yöntem seçilmeye çalışılmıştır. Araştırma sonucunda en az hata ve yanlılık içeren eşitleme yöntemi, MTK gözlenen puan eşitleme yöntemi olurken en fazla hata ve yanlılık, zincirleme EYE yöntemi ile elde edilmiştir.

Tanberkan Suna (2018), çalışmasında grup değişmezliğinin, farklı eşitleme yöntemlerinde elde edilen eşitleme fonksiyonlarına etkisini araştırmak istemiştir. Planlanan çalışmanın verileri için insan kaynakları yönetimi alanında işe alım ve terfi süreçlerinde kullanılan genel yetenek testi FVAT 2015-2016 V4 ve V13 kitapçık puanları kullanılmıştır. 40 maddeli bu test, kendi içerisinde dört bölüme ayrılmaktadır. Ortak maddeli eşit olmayan grup deseni kullanılan bu çalışmada, MTK ve KTK bünyesindeki eşitleme yöntemleri incelenmiştir. KTK doğrusal eşitleme yöntemlerinden Tucker, Levine, Braun-Holland yöntemleri kullanılırken EYE eşitleme için düzgünleştirme yapılmayan ve son düzgünleştirme S: 0,10 derecesi ile düzgünleştirme yapılan yöntemler kullanılmıştır. Araştırma bulguları incelendiğinde grup değişmezliği ihlaline karşı Tucker, Braun-Holland ve ortalama-ortalama kalibrasyon yöntemlerinin uygun olduğu görülmüştür. MTK gözlenen puan eşitleme, gerçek puan eşitlemeden daha iyi sonuç vermiştir. Cinsiyet değişkenine göre grup değişmezliğinde MTK gözlenen eşitleme için ortalama-ortalama kalibrasyon yönteminin; yaş değişkenine göre grup değişmezliğinde, MTK gözlenen eşitleme için Stocking-Lord kalibrasyon yönteminin uygun olduğu bilgisine ulaşılmıştır.

Alanyazında yapılmış çalışmalar incelendiğinde KTK ve MTK'ya yönelik eşitleme yöntemlerinden birkaçına odaklanılarak çalışmaların gerçekleştirildiği görülmektedir. MTK eşitleme sürecinde ise alanyazında en az hataya sahip eşitleme yöntemi dikkate alınıp eşitlenmiş puanlar elde edilirken, kalibrasyon süreci dahil edilip gerçek ve gözlenen puan eşitleme yöntemlerinin karşılaştırmalı kullanıldığı çalışmalar azınlıktadır. Bu araştırmada KTK'da yer alan doğrusal ve eşityüzdelikli eşitleme yöntemleri, farklı sentetik evren değerleri dikkate alınarak eşitlenmiş puanlar hesaplanılmış ve MTK kalibrasyon yöntemleri gerçek ve gözlenen puan eşitleme yöntemleri karşılaştırılması yapılmak istenmiştir.

Alanyazında EYE alıřmalarından nce dzgnleřtirmenin nemi ve doėrusal yntemlerin teker teker denendiėi alıřmaların azınlıkta olduėu grlmřtr. Bu arařtırmada alıřmanın amacı kapsamında KTK ve MTK bnyesindeki eřitli eřitleme yntemleri detaylı olarak ele alınarak bir eřitleme yntemleri karřılařtırılmıřtır.



BÖLÜM III

YÖNTEM

Bu bölümde araştırma modeli, çalışma grubu, veri toplama deseni, verilerin analizi ve test eşitleme varsayımlarının sağlanmasına ilişkin bilgilere yer verilmiştir.

Araştırmanın Deseni

Bu araştırmada, ortak maddeli denk olmayan gruplar deseni kullanılarak KTK ve MTK'ya dayalı eşitleme yöntemleri ile PISA 2012'nin matematik puanlarına dayalı olarak kullanılan eşitleme yöntemlerinden en uygun olanının seçilmesi amaçlanmıştır. Bu araştırmada kullanılan farklı eşitleme yöntemlerinden en az hata vereni bulmak amaçlandığından bu çalışma, betimsel araştırma niteliği taşımaktadır. Betimsel araştırmalar var olan durumu olduğu gibi ortaya çıkarılmasının amaçladığı araştırmalar için uygundur (Karasar, 2005).

Evren ve Örneklem

Çalışmanın evrenini PISA 2012'ye katılan 15 yaş grubu öğrenciler oluşturmaktadır. Çalışma grubu için PISA uygulamasına katılan 65 ülke arasından matematik okuryazarlığı puanları incelendiğinde en başarılı, en başarısız, ortalamanın altında ve ortalamanın üstünde bulunan toplamda dört ülke seçilmiştir. PISA 2012'ye katılan ülkeleri temsil ettiği için bu ülkelerin seçilmesi uygun görülmüştür. Çalışma grubundaki ülkelerin seçimi şu şekilde gerçekleştirilmiştir.

*PISA 2012 fen, matematik ve okuma becerisi alanında en iyi performans gösteren ülke Şangay/ Çin seçilmiştir.

*PISA 2012 için en başarısız ülke Peru olmasına rağmen Peru ile diğer ülkelerdeki kitapçıklar eşleşemediğinden başarı sıralamasında Peru'nun bir üstündeki ülke, Endonezya, belirlenmiştir.

*PISA 2012 uygulamasına katılan Türkiye, ortalamanın altında yer aldığı için uygun görülmiştir.

* PISA 2012'de genel başarı düzeyi bir önceki PISA uygulamasına göre düşen Finlandiya ise ortalamanın üstünde yer almaktadır. Ortalamanın altında yer alan Türkiye'ye karşılık ortalamanın üstünde olan Finlandiya, başarılı ülkeleri temsil ettiği için seçilmiştir.

Araştırmanın amacına bağlı olarak bilgi açısından zengin durumlar seçilerek eşitleme sürecinde derinlemesine araştırma yapılmasına olanak tanıdığından amaçsal örnekleme kullanılmıştır (Büyüköztürk, Çakmak, Akgün, Karadeniz ve Demirel, 2008).

Eşitleme sürecinde örneklem büyüklüğü önemlidir. Örneklem büyüklüğü rastgele hata üzerine doğrudan etki yapmaktadır. Böylece geniş örneklemle eşitleme hatasının az olması, eşitleme sonuçlarının daha iyi çıkmasını sağlamaktadır. Geleneksel eşitlemede doğrusal eşitleme, her bir form ya da test için 400 örneklem büyüklüğü gerektirirken EYE için 1500 ve üstü bir örneklem büyüklüğü yeterli olmaktadır. MTK eşitleme sürecinde Rasch model için genellikle 400'lük örneklem büyüklüğüne ve 3 PLM için en az 1500'lük örneklem büyüklüğüne ihtiyaç vardır (Kolen ve Brennan, 1995; Livingston, 1993; Zu ve Yuan, 2012). Bu çalışmada, amaçsal örnekleme kapsamında kitapçık 1'i 1921 kişi, kitapçık 3'ü ise 1900 kişi almıştır. Kitapçıkları alan kişilerin ülkelere göre dağılımları Tablo.1'de belirtilmiştir. Tablo 1 doğrultusunda, eşitlemede toplam 3821 kişiden elde edilen puanlar kullanılmıştır.

Tablo 1

Ülkelere Göre Kitapçıkları Alan Öğrenci Dağılımları

	Kitapçık 1	Kitapçık 3
Çin (QCN)	442	439
Endonezya (IDN)	424	417
Finlandiya (FIN)	683	669
Türkiye (TUR)	372	375
Toplam	1921	1900
Genel Toplam	3821	

Verilerin Toplanması ve Düzenlenmesi

Bu araştırmada PISA 2012 matematik alt testi puanları kullanılmıştır. Seçilen kitapçıkta maddeler ve PISA hakkında kısa bir bilgi aşağıda sunulmuştur.

PISA 2012 Hakkında Kısa Bilgi

Uluslararası Öğrenci Değerlendirme Programı- PISA (Programme for International Student Assessment) Ekonomik İşbirliği ve Kalkınma Teşkilatı (Organisation for Economic Co-Operation and Development-OECD) tarafından düzenlenen dünyanın en kapsamlı eğitim araştırmalarından biridir (MEB, 2013). Dünya genelinde, politika belirleyicileri kendi ülkelerindeki öğrencilerin bilgi ve beceri düzeylerini, projeye katılan diğer ülkelerdeki öğrencilerin bilgi ve beceri düzeyleriyle karşılaştırmak, eğitim düzeyinin yükseltilmesi amacıyla standartlar oluşturmak (örneğin, ülkeler tarafından elde edilen ortalama puanlar, ülkelerin eğitim çıktıları ve eğitim fırsatlarında eşitliği en yüksek düzeyde sağlama kapasiteleri) ve eğitim sistemlerinin güçlü ve zayıf yönlerini belirlemek için PISA sonuçlarını kullanmaktadırlar (MEB, 2015).

2000 yılından itibaren üç yılda bir yapılan bu araştırmayla OECD üyesi ülkeler ve diğer katılımcı ülkelerdeki (dünya ekonomisinin yaklaşık olarak %90'ı) 15 yaş grubu öğrencilerin, modern toplumda yerlerini alabilmeleri için gereken temel bilgi ve becerilere ne ölçüde sahip oldukları değerlendirilmektedir (MEB, 2013).

PISA ilk uygulamasını 2000 yılında gerçekleştirmiştir. Bundan sonraki uygulamalarını ise üçer yıl ara ile yapmıştır. PISA uygulaması okuma becerisi, matematik okuryazarlığı ve matematik okuryazarlığını ele almaktadır. Her üç yılda bir bu üç alandan bir tanesine odaklanmaktadır. Yıllara göre ölçülen beceriler aşağıdaki Şekil 12’de özetlenmiştir.

2000	2003	2006	2009	2012	2015	2018
Okuma Becerisi	Okuma Becerisi	Okuma Becerisi	Okuma Becerisi	Okuma Becerisi	Okuma Becerisi	Okuma Becerisi
Matematik Okuryazarlığı	Matematik Okuryazarlığı	Matematik Okuryazarlığı	Matematik Okuryazarlığı	Matematik Okuryazarlığı	Matematik Okuryazarlığı	Matematik Okuryazarlığı
Fen Okuryazarlığı	Fen Okuryazarlığı	Fen Okuryazarlığı	Fen Okuryazarlığı	Fen Okuryazarlığı	Fen Okuryazarlığı	Fen Okuryazarlığı

Şekil 12 Yıllara göre PISA’da odaklanılan alanlar

Şekil 12 incelendiğinde 2000 ve 2009 yıllarında okuma becerisine; 2003 ve 2012 yıllarında matematik okuryazarlığına; 2006 ve 2015 yıllarında ise fen okuryazarlığına odaklanıldığı görülmektedir. Bu eğitim araştırmasına Türkiye 2003 yılından itibaren fiilen katılmıştır. PISA Projesi kapsamında geliştirilen başarı testleri ve anketleri uygulanmaktadır (OECD, 2013).

PISA 2012 uygulamasında ağırlıklı alan, matematik okuryazarlığıdır. Bundan dolayı bu çalışmada analizler tamamen matematik sonuçlarına göre yürütülmüştür. Matematik, okuma ve fen alanlarında kullanılan puanlama ölçekleri için gerekli metrik, söz konusu yılda, OECD ülkelerinin ilgili alandaki ortalama puanı 500 ve standart sapması 100 olacak şekilde inşa edilmiştir.

PISA uygulamasında 13 tane kitapçık bulunup bu kitapçıklardan doğru-tam doğru-geçersiz-boş gibi yanıtlar, ortalama 500 ve standart sapma 100 kuralına uygun olarak puanlara dönüştürülmektedir. Bu 13 kitapçıkta ortak olan bazı maddeler sayesinde puanların birbirleri ile karşılaştırılması ve birbiri yerine kullanılmasına fırsat verilmiştir. Bu çalışmada test eşitleme çalışması için ortak maddeli eşdeğer olmayan grup deseni uygun görülmüştür. Bu

bağlamda bu araştırmada, PISA 2012 için seçilen iki kitapçıktan elde edilen puanların eşitlenme süreci anlatılmıştır (MEB, 2015).

Bu çalışma, PISA 2012 kapsamında 15 yaş grubu öğrencilere uygulanan bilişsel test verileri üzerinden yürütülmüştür. Veriler <http://www.oecd.org/pisa/data/> adresinden indirilmiştir. Bu veri setinde Çin-Şangay, Finlandiya, Türkiye ve Endonezya'ya ait 13 kitapçık incelenmiştir. Araştırma için matematik okuryazarlığına ilişkin veriler kullanılmıştır. Daha sonra her kitapçık için frekans analizi ile hangi soruların hangi kitapçıklarda uygulandığı belirlenmiştir. Ortak madde dizimi ve diğer maddelerin örtüşmeleri incelendiğinde eşitlenecek kitapçıklar, kitapçık 1 ve kitapçık 3 olarak belirlenmiştir. Kitapçıklardaki maddeler ve kodlamaları Tablo 2'de sunulmuştur.

Tablo 2

Kitapçıklarda Yer Alan Maddeler ve Kodları

Madde Türü	Kitapçık Türü	Maddeler	Madde Kodları
Ortak Olmayan Maddeler	Kitapçık 1	MATH-P2012- An advertising Column Q1	PM00GQ01
		MATH-P2012-Speeding Fines Q1	PM909Q01
		MATH-P2012-Speeding Fines Q2	PM909Q02
		MATH-P2012-Speeding Fines Q3	PM909Q03
		MATH-P2012-Roof Truss Design Q1	PM949Q01T
		MATH-P2012-Roof Truss Design Q2	PM949Q02T
		MATH-P2012-Roof Truss Design Q3	PM949Q03T
		MATH-P2012-Migration Q1	PM955Q01
		MATH-P2012-Migration Q2	PM955Q02
		MATH-P2012-Migration Q3	PM955Q03
	Kitapçık 3	MATH-P2012-Bike Rental Q2	PM998Q02T
		MATH-P2012-Bike Rental Q4	PM998Q04T
		MATH-P2000-Pipelines Q1	PM273Q01T
		MATH-P2003-Lotteries Q1	PM408Q01T
		MATH-P2003-Transport Q1	PM420Q01T
		MATH-P2003-TheThermometer Cricket Q1	PM446Q01
		MATH-P2003-TheThermometer Cricket Q2	PM446Q02
		MATH-P2003-Tile Arrangement Q1	PM447Q01
		MATH-P2003-The Fence Q1	PM464Q01T
		MATH-P2003-Telephone Rates Q1	PM559Q01
Ortak Maddeler	Kitapçık 1 ve Kitapçık 3	MATH-P2003-Computer Game Q1	PM800Q01
		MATH-P2003-Carbon Dioxide Q1	PM828Q01
		MATH-P2003-Carbon Dioxide Q2	PM828Q02
		MATH-P2003-Carbon Dioxide Q3	PM828Q03
		MATH-P2012-Apartment Purchase Q1	PM00FQ01
		MATH-P2012-Drip Rate Q1	PM903Q01
		MATH-P2012-Drip Rate Q3	PM903Q03
		MATH-P2012-Charts Q1	PM918Q01
		MATH-P2012-Charts Q2	PM918Q02
		MATH-P2012-Charts Q5	PM918Q05
		MATH-P2012-Sailing Ships Q1	PM923Q01
		MATH-P2012-Sailing Ships Q3	PM923Q03
		MATH-P2012-Sailing Ships Q4	PM923Q04
		MATH-P2012-Sauce Q2	PM924Q02
		MATH-P2012-Revolving Door Q1	PM995Q01
		MATH-P2012-Revolving Door Q2	PM995Q02
		MATH-P2012-Revolving Door Q3	PM995Q03

Tablo 12’de yer alan maddeler ülke ve kitapçıklara göre ayıklanarak oluşturulmuştur. Tablo 12 incelendiğinde her kitapçıkta 25 madde bulunmaktadır. Bu maddelerden 13’ü ortak, 12’si ortak olmayan maddedir. Bilişsel testte doğru yanıtlar “1”, kısmen doğru, yanlış ve diğer yanıtlamalar için “0” olacak şekilde kodlanmıştır. Bu şekilde ikili puanlama modeline dayalı kodlama tercih edildiğinden kısmen doğru yanıtlar da yanlış yanıt olarak değerlendirilmiştir.

Eşitleme süreci, yeni formun eski forma eşitlenmesi temeline dayanmaktadır. Bu araştırmada ise kitapçık 1, yeni form (X formu); kitapçık 3, eski form (Y formu) olarak belirlenmiştir. Bu durumda Kitapçık 1’den elde edilen puanlar, kitapçık 3’ten elde edilen puanlara eşitlenmek istenmiştir.

Eşitleme Deseni (Veri Toplama Deseni)

Araştırmada Çin, Finlandiya, Türkiye ve Endonezya ülkelerinin PISA 2012 kitapçık 1 ve kitapçık 3 puanları eşitlenmiştir. Eşitlemede kullanılan kitapçıklarda 12 madde, her iki grup için farklı olacak şekilde hazırlanmıştır. Kitapçıklardaki 13 ortak madde ise aynı sıra ile her iki gruba da uygulanmıştır. Bu sebeple kitapçık 1 ve kitapçık 3 puanları ortak maddeler (13 madde) kullanılarak eşitlenebilmektedir. Veriler, araştırma sürecinde bu şekilde toplandığı için ortak maddeli eşdeğer olmayan gruplar deseni kullanılarak eşitleme süreci planlanmıştır.

Verilerin Analizi ve Test Eşitleme Varsayımlarının Test Edilmesi

Bu kısımda verilerin analizine geçmeden önce eşitleme sürecinde test edilmesi gereken varsayımların analizine yer verilmiştir.

Yatay eşitlemeye dayalı eşitleme sürecinde, benzer içerik ve benzer güçlük düzeyindeki formların birbirleri yerine kullanılması esastır. Bu nedenle öncelikle kitapçık 1 ve kitapçık 3 için güçlük düzeylerini gösteren temel istatistikler Tablo 3’te sunulmuştur.

Tablo 3

Kitapçık 1 ve Kitapçık 3 İçin Betimsel İstatistikler

	Kitapçık 1			Kitapçık 3		
			Ortak			Ortak
	Tüm Maddeler	Ortak Maddeler	Olmayan Maddeler	Tüm Maddeler	Ortak Maddeler	Olmayan Maddeler
Kişi Sayısı (n)	1921	1921	1921	1900	1900	1900
Soru Sayısı (K)	25	13	12	25	13	12
$\bar{X}$	11,31	5,97	5,35	11,42	6,02	5,4
Mod	6	4	5	5	3	2
Medyan	11	6	5	11	6	5
Standart Sapma	6,08	3,41	2,98	6,35	3,47	3,25
Varyans	36,92	11,64	8,90	40,34	12,03	10,55
Minimum Değer	0	0	0	0	0	0
Maximum Değer	25	13	12	25	13	12
Basıklık	-0,96	-1,01	-0,81	-1,06	-1,07	-1,00
Çarpıklık	0,23	0,15	0,26	0,19	0,16	0,14

Tablo 3'te iki kitapçığı da benzer örneklem büyüklüğündeki gruplar almıştır. Her iki kitapçıkta da aynı sayıda ortak ve ortak olmayan maddeler bulunmaktadır. Tabloya göre kitapçıkların güçlük düzeyleri incelendiğinde iki kitapçığın da benzer güçlük düzeyinde ($\bar{X}_{\text{kitapçık 1}}=11,31$; $\bar{X}_{\text{kitapçık 2}}=11,42$) oldukları sonucuna ulaşılabilir. Yine de veri seti normal dağıldığından iki kitapçık arasında güçlük düzeyleri bakımından manidar bir farkın olup olmadığı, ilişkisiz örneklem t testi ile belirlenmiştir. Sonuçlar Tablo 4'te verilmiştir.

Tablo 4

Kitapçık 1 ve Kitapçık 3 İçin Ortalamalar Arası Farka İlişkin İlişkisiz Örneklem t Testi Sonucu

	n	$\bar{X}$	ss	sd	t	p
Kitapçık 1	1921	11,31	6,075913	3819	0,554	0,58
Kitapçık 3	1900	11,43	6,351331			

$p < 0,05$

Tablo 4 incelendiğinde kitapçık türüne göre kitapçıkların ortalamalarında anlamlı bir farklılığın olmadığı görülmektedir [$t(3819) = 0,554$, $p > 0,05$]. Kitapçık 1'in ortalaması ($\bar{X}=11,31$) ile kitapçık 3'ün ortalamasının ($\bar{X}=11,43$) benzer olduğu söylenebilir. Bu durum, yatay eşitleme için gerekli olan bir ön koşuldur. Puanların ortalamaya göre dağılımları incelendiğinde bu dağılımların iki kitapçık için de benzer olduğu görülmektedir. Kitapçıklardaki puanların dağılımları, çarpıklık ve basıklık değerleri dikkate alınarak incelenmiştir. Büyüköztürk, Çokluk ve Köklü (2010)'ye göre çarpıklık katsayısının +1 ile -1 sınırları içinde kalması, puanların normalden aşırı bir sapma göstermediği şeklinde yorumlanmaktadır. Kitapçıklardaki çarpıklık katsayıları incelendiğinde +1, -1 arasında değer aldığı görüldüğü için normal dağılımdan bahsedilebilir. Basıklık katsayısı ise dağılımın şekline yönelik bir bilgi vermekte ve bu katsayıda "0" baz alınarak incelenmektedir. Verilerden elde edilen basıklık katsayısı negatif çıkarsa dağılım "basık (platykurtic)", pozitif çıkarsa "sivri (leptokurtic)" bir dağılıma sahiptir (Tan, 2016; de Wslope, 2013). Tablodaki basıklık değerleri değerlendirildiğinde ise dağılımın basık olduğu görülmektedir.

Eşitleme sürecinde ortak maddelerin testin tümünü temsil etmesi beklenir. Bu sebeple ortak maddelerden elde edilen puanlar ile testin bütününden elde edilen puanlar arasında bir korelasyon katsayısı hesaplanmıştır. Tablo 5'te, kitapçık 1 ve kitapçık 3'e ilişkin korelasyon katsayıları yer almaktadır.

Tablo 5

Kitapçık 1 ve Kitapçık 3 için Ortak Madde Puanları ve Testin Bütününe Arasındaki Korelasyon

			Ortak Maddeler	Testin Bütünü
Kitapçık 1	Ortak Maddeler	r	1	0,956**
		p		0
		n	1921	1921
	Testin Bütünü	r	0,956**	1
		p	0	
		n	1921	1921
Kitapçık 3	Ortak Maddeler	r	1	0,949**
		p		0
		n	1900	1900
	Testin Bütünü	r	0,949**	1
		p	0	
		n	1900	1900

Kitapçıklardaki veriler eşit aralık ölçeği düzeyinde olduğu için Pearson korelasyon katsayısı her iki kitapçık için hesaplanmıştır. Tablo 5'te hesaplanan korelasyon katsayısının 1'e oldukça yakın olduğu görülmektedir. Ortak madde puanları ile kitapçıklardan alınan puanlar arasında pozitif yönlü oldukça güçlü bir korelasyon mevcuttur. Bu sonuç doğrultusunda, ortak maddelerin kitapçıkta maddeleri temsil etme gücünün yüksek olduğu sonucuna ulaşılabilir.

MTK'da formlardan elde edilen ortalama güçlük değeri, iki formun zorluk bakımından birbirine eş olduğuna ilişkin yeterli bir kanıt sağlamamaktadır. Bunun için öncelikle BILOGMG 3.0 programı ile parametreler kestirilmiştir. Parametre kestiriminde tüm sorulara doğru yanıt veren ve hiçbir soruyu yanıtlamayan öğrenciler bulunduğu EAP (Expected a Posteriori) yöntemi uygun görülmüştür. Sonuçlar Tablo 6'da verilmiştir.

Tablo 6

Kitapçık 1 ve 3'e İlişkin Kestirilen Parametre Kestirimleri

Kitapçık 1'e İlişkin Kestirilen Madde Parametreleri				Kitapçık 3'e İlişkin Kestirilen Madde Parametreleri			
Maddeler	a	b	c	Maddeler	a	b	c
1	2,102	2,045	0,028	1	1,691	0,361	0,212
2	1,805	-0,864	0,221	2	1,535	0,746	0,042
3	1,917	0,274	0,151	3	1,535	0,499	0,076
4	2,864	0,826	0,038	4	2,55	-0,147	0,067
5	1,974	-0,268	0,205	5	3,664	1,362	0,016
6	2,957	0,764	0,081	6	1,958	0,007	0,196
7	2,851	0,586	0,06	7	2,701	0,82	0,03
8	1,181	-0,548	0,233	8	2,097	0,018	0,245
9	1,55	0,422	0,068	9	1,298	-1,218	0,324
10	2,885	1,328	0,014	10	2,633	0,781	0,025
11	1,783	-0,124	0,227	11	1,702	0,144	0,088
12	2,039	1,103	0,276	12	1,82	0,977	0,033
13	2,662	0,402	0,07	13	2,613	0,444	0,066
14	3,718	0,987	0,02	14	3,372	0,945	0,02
15	3,222	0,761	0,045	15	2,812	0,836	0,054
16	1,029	-1,252	0,364	16	0,874	-1,652	0,391
17	1,889	-0,656	0,117	17	2,041	-0,561	0,141
18	1,6	-0,678	0,202	18	1,55	-0,469	0,257
19	2,114	0,176	0,144	19	2,462	0,165	0,142
20	2,735	0,497	0,184	20	2,703	0,509	0,167
21	2,727	1,214	0,017	21	2,592	1,109	0,02
22	2,499	-0,016	0,153	22	2,316	-0,087	0,124
23	2,69	0,027	0,067	23	3,035	0,084	0,095
24	1,939	2,317	0,014	24	2,343	2,049	0,012
25	1,582	0,493	0,063	25	1,899	0,367	0,059

Tablodaki her kitapçığın ilk 12 sorusu, ortak olmayan maddelere ilişkin parametre değerleridir. Madde 13'ten madde 25'e kadar ortak maddelerin parametreleri yer almaktadır.

Tablo 6’da belirtilen parametrelerin ayrı ayrı ortalamaları alınmış oluş Tablo 7’de sunulmuştur.

Tablo 7

Kitapçıklara Ait Ayırtedicilik, Madde Güçlüğü ve Şans Parametrelerinin Ortalamaları

Kitapçıklar	n	$\bar{X}_a$	$\bar{X}_b$	$\bar{X}_c$
Kitapçık 1	25 madde (12 ortak olmayan+13 ortak)	2,253	0,393	0,123
Kitapçık 3	25 madde (12 ortak olmayan+13 ortak)	2,232	0,324	0,116

Tablo 7 incelendiğinde kitapçık 1 ve kitapçık 3’teki parametreler için aynı yorum yapılabilir. Bunun nedeni hesaplanan madde parametrelerinin her iki kitapçıkta da oldukça benzer olmasıdır. Test eşitleme sürecinde de formların benzer madde parametrelerine sahip olması beklenmektedir. Formlardaki maddelere ilişkin kestirilen madde parametrelerinin iki form için manidar bir farklılık oluşturup oluşturmadığı ilişkisiz örneklem t testi ile test edilmiştir. Bu test istatistiği, $n > 30$ kuralını sağlanmasa da EK-1’de sunulan Kolmogorov, Shapiro-Wilk’s testi sonuçları dikkate alındığı ve ilişkisiz örneklem t testine ilişkin diğer varsayımlar sağlandığı için kullanılması uygun görülmüştür. İki kitapçık arasında a parametresine göre manidar bir farklılık olup olmadığı Tablo 8’de verilmiştir.

Tablo 8

Kitapçıklardaki a Parametresine Yönelik İlişkisiz Örneklem t Testi Sonuçları

Kitapçıklar	n	$\bar{X}$	ss	sd	t	p
Kitapçık 1	13	2,198	0,583	24	0,243	0,769
Kitapçık 3	13	2,138	0,661			

$p < 0,05$

Grafikle ortak maddelerin her iki grup için a değerlerini sunup yorumlamak da faydalı olur. Tablo 8’de görüldüğü gibi kestirilen ayrıcalık indeksinin kitapçık 1 ortalaması 2,198 iken, kitapçık 3’teki ortalaması 2,138’dir. Bu sonuçlara göre testteki maddelerin ayırt ediciliğinin yüksek olduğu söylenebilir. Ayırt edicilik parametresine göre iki kitapçık karşılaştırıldığında

a parametresi kitapçıklarda 0,05 manidarlık düzeyinde anlamlı bir farklılık göstermediği Tablo 8’de görülmektedir [$t(24) = 0,243$, $p > 0,05$]. Tablo 9’da kitapçıklardaki güçlük parametresine yönelik t testi sonuçlarına yer verilmiştir.

Tablo 9

Kitapçıklardaki b Parametresine Yönelik İlişkisiz Örneklem t Testi Sonuçları

Kitapçıklar	n	$\bar{X}$	ss	sd	t	p
Kitapçık 1	13	0,458	0,796	24	0,312	0,468
Kitapçık 3	13	0,369	0,644			

$p < 0,05$

Güçlük parametreleri her iki kitapçık için kestirilmiştir. Kitapçık ortalamaları sırası ile 0,458 ve 0,369 şeklindedir. Analiz sonucuna göre soruların zor olmadığı söylenebilir. İki kitapçıkta madde güçlüğü sonuçları arasındaki farklılığın manidar olup olmadığı, ilişkisiz örneklem t testine göre belirlenmiştir. Analiz sonuçlarına göre kitapçık 1 ve kitapçık 3’teki maddelerin madde güçlük parametreleri arasındaki farklılığın, 0,05 manidarlık düzeyinde farklı olmadığı sonucu bulunmuştur [$t(24) = 0,312$, $p > 0,05$]. Bu sonuç eşitleme türlerinden yatay eşitlemeye uygun bir sürecin işlendiğinin kanıtıdır. Çünkü yatay eşitlemede de benzer güçlük düzeylerine vurguda bulunulmuştur.

Tablo 10’da ise şans parametresinin her iki kitapçıkta ortalamaları arasındaki farklılığın karşılaştırması yapılmıştır.

Tablo 10

Kitapçıklardaki c Parametresine Yönelik İlişkisiz Örneklem t Testi Sonuçları

Kitapçıklar	n	$\bar{X}$	ss	sd	t	p
Kitapçık 1	13	0,129	0,093	24	0,513	0,991
Kitapçık 3	13	0,109	0,100			

$p < 0,05$

Tablo 10’da görüldüğü üzere şans parametreleri ortalamaları ($\bar{X}_{\text{kitapçık1}}=0,129; \bar{X}_{\text{kitapçık3}}=0,109$) birbirlerine oldukça benzerdir. İlişkisiz örneklem t testi

ile iki kitapçıktaki farklılığın manidarlığı test edildiğinde bu farklılığın, 0,05 düzeyinde anlamlı olmadığı görülmüştür [$t(24) = 0,513$, $p > 0,05$].

Eşitleme sürecinde, benzer güvenilirlik düzeyindeki formların kullanılması gerekmektedir. Testlerin güvenilirlikleri tespit edilmiş ve Tablo 11’de belirtilmiştir.

Tablo 11

Kitapçık 1 ve Kitapçık 3’e İlişkin Güvenirlik Katsayıları

	Kitapçık 1	Kitapçık 3
KR-20	0,902	0,910

Tablo 11 incelendiğinde her iki kitapçığın güvenilirliklerinin yüksek, ölçme sonuçlarına karışan hatanın ise az olduğu raporlaştırılmıştır. Kitapçıktan elde edilen değerlerin birbirine oldukça yakın olduğu görülmektedir.

Korelasyon katsayısı için Zr dönüşümünün kullanılması, r’lerin özellikle çok yüksek değerlerde (+1,00 ya da -1,00’e yakın) olması durumunda elverişlidir (Akhun, 1984). Bu sebeple güvenilirlik katsayıları arasındaki fark ile iki korelasyon arasındaki farkın manidarlığı için Fisher Zr dönüşümü yöntemi kullanılmıştır. Her bir kitapçık için hesaplanan Zr değerleri Tablo 12’de verilmiştir.

Tablo 12

Kitapçık 1 ve Kitapçık 3’e İlişkin Fisher Z Katsayıları

	Kitapçık 1	Kitapçık 3
Fisher Zr	1,472	1,528

İki Zr arasındaki fark, iki Zr değeri arasındaki farkın standart sapmasına bölündüğünde $Z=1,729$ elde edilmiştir. Bu değer, Z kritik değer ile karşılaştırıldığında bu Z bölgesi, H_0 ($r_{\text{kitapçık1}}=r_{\text{kitapçık3}}$) bölgesinde kalmaktadır. Eldeki verilere göre, bu iki kitapçığın güvenilirlikleri katsayıları arasında manidar bir farklılık yoktur. Her iki formun da güvenilirliklerinin benzer olduğu söylenmektedir.

Eşitleme sürecinde kullanılan kitapçık 1 ve kitapçık 3’te, matematik okuryazarlığına yönelik maddeler yer almaktadır. Eşitleme sürecinde her bir maddenin ölçtüğü tek bir başat yapının

olduğu varsayımı da bu araştırmada test edilmek istenmiştir. Tek boyutluluk varsayımının tespiti için yaygın olarak kullanılan açımlayıcı faktör analizi (AFA), bu araştırma için uygun görülmemiştir. Bunun nedeni, madde-yanıt matrisinin ikili kategoriye göre puanlanmasıdır. Bu durumda tetrakorik korelasyon temelli bir faktör analizi yapılmıştır. Tetrakorik korelasyon, iki kategorili değişkenin normal dağılımına (bivariate normal distribution) sahip değişkenler arasındaki ilişkiyi göstermek için kullanılır (Lord ve Novick, 1968; Ritchie-Scott, 1918). Bu çalışmada ise değişken olarak belirtilen ifadeler, iki kategorili kodlanan maddelerdir.

Faktör analizine başlamak için öncelikle veri yapısının uygunluğu, Kaiser-Meyer-Olkin (KMO) ve Barlett Küresellik testleri ile sınanmalıdır (Tabachnick ve Fidell, 2001). Kitapçık 1 ve 3'ün KMO ve Barlett testi sonuçları Tablo 13'te yer almaktadır.

Tablo 13

Kitapçık 1 ve 3'e İlişkin KMO ve Barlett Testi Sonuçları

	Kaiser-Mayer-Olkin (KMO)	Barlett Küresellik Testi Ki-Kare Değeri	sd	p
Kitapçık 1	0,96	13137,1	300	0,00
Kitapçık 3	0,97	13622,0	300	0,00

$p < 0,05$

KMO testi sonucu 0,60 ve bu değerden büyük olduğunda ve Barlett küresellik testi 0,05 düzeyinde manidar çıktığında yapının faktör analizine uygun olduğu kanıtlanmış olur (Büyüköztürk, 2011a). Tablo 12'de sunulan bulgular doğrultusunda kitapçık 1 için hesaplanan KMO değeri, 0,96 ve kitapçık 3 için hesaplanan KMO değeri, 0,97 olarak bulunmuştur. Barlett küresellik testi, her iki kitapçık için $p < 0,05$ koşulunu sağladığından veri yapısının faktör analizi için uygunluğu kanıtlanmıştır.

Araştırmada, tek boyutluluk tetrakorik korelasyona dayalı faktör analizi ile test edilmiştir. Bu analiz için LISREL 9.1 programından yararlanılmıştır. Araştırmada kitapçık 1 ve kitapçık 3 için tek boyutluluğun sağlandığına ilişkin kurulan modeli test etmek için kullanılacak

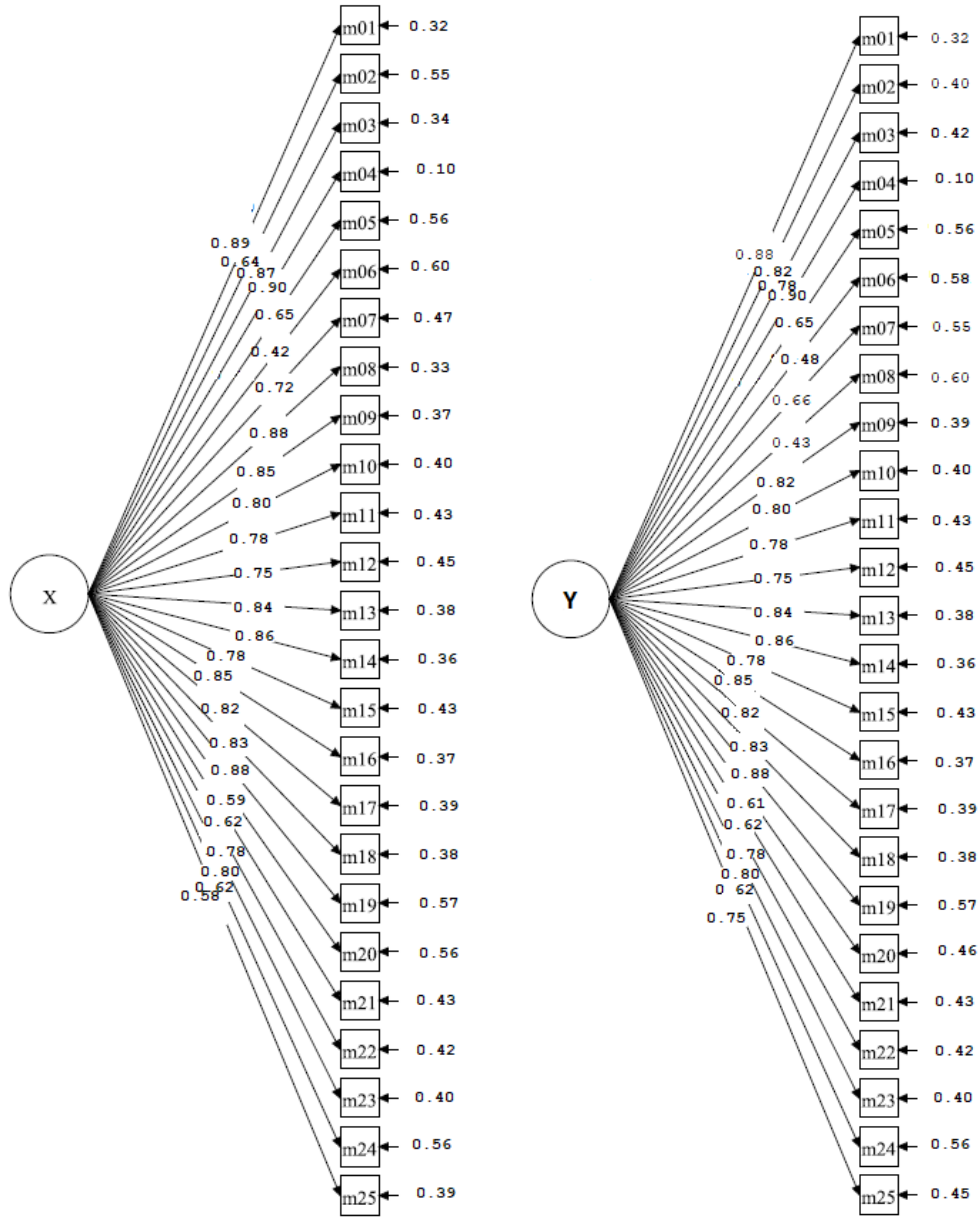
model uyum indeksleri kriterleri ve kabul için kesme noktaları, Tablo 14’te verilmiştir (Schumacker ve Lomax, 1996; Hu ve Bentler, 1999; Tabachnick ve Fidell, 2001).

Tablo 14

Model uyum İndekslerinin Kriterleri ve Kabulü İçin Kesme Noktaları

Uyum İndeksleri	Kriterler	Kesme Noktaları
χ^2	$p > 0,05$	
χ^2/sd		≤ 2 = Mükemmel Uyum $\leq 2,5$ = Mükemmel Uyum (küçük örneklemelerde) ≤ 3 = Mükemmel Uyum (büyük örneklemelerde) ≤ 5 = Orta Uyum Düzeyi
GFI	0 (uyum yok) 1 (mükemmel uyum)	$\leq 0,90$ = Kabul Edilebilir Uyum $\leq 0,95$ = Mükemmel Uyum
AGFI	0 (uyum yok) 1 (mükemmel uyum)	$\leq 0,90$ = Kabul Edilebilir Uyum $\leq 0,95$ = Mükemmel Uyum
CFI	0 (uyum yok) 1 (mükemmel uyum)	$\leq 0,90$ = Kabul Edilebilir Uyum $\leq 0,95$ = Mükemmel Uyum
NNFI(TLI)	0 (uyum yok) 1 (mükemmel uyum)	$\leq 0,90$ = Kabul Edilebilir Uyum $\leq 0,95$ = Mükemmel Uyum
NFI	0 (uyum yok) 1 (mükemmel uyum)	$\leq 0,90$ = Kabul Edilebilir Uyum $\leq 0,95$ = Mükemmel Uyum
RMSEA	0 (mükemmel uyum) 1 (uyum yok)	$\leq 0,08$ = Kabul Edilebilir Uyum $\leq 0,05$ = Mükemmel Uyum
RMR/SRMR	0 (mükemmel uyum) 1 (uyum yok)	$\leq 0,08$ = Kabul Edilebilir Uyum $\leq 0,05$ = Mükemmel Uyum

Ölçülen özellik, tek boyutlu bir model önseli ile Mplus 7.4 programı kullanılarak model uyumu incelenmiştir. Analiz sonucunda elde3 eidlen ölçme modeli Şekil 13’te sunulmuştur.



Şekil 13 Kitapçık 1 ve Kitapçık 3 için ölçme modelleri

Yukarıdaki şekilde kitapçık 1 ve kitapçık 3 için tek boyutluluğun sağlandığı görülmektedir. Görsel olarak incelenen ölçme modelinin hesaplanan model veri uyum indeksleri Tablo 15'te sunulmuştur.

Tablo 15

Kitapçıklara İlişkin Model-Veri Uyumu İndeks Değerleri

Kitapçıklar	Boyut	χ^2 (sd)	RMSEA	GFI	CFI	NNFI	SRMR	λ	ε
Kitapçık 1	Tek Boyutlu	11890,546 (275)	0,05	0,991	0,969	0,952	0,0531	0,42-0,90	0,10-0,60
Kitapçık 3	Tek Boyutlu	11821,59 (275)	0,05	0,990	0,965	0,950	0,0517	0,43-0,90	0,10-0,60

Araştırmada ele alınan PISA 2012 kitapçık 1 ve kitapçık 3 için tek boyutluluğun sağlanıp sağlanmadığına yönelik kurulan model için hesaplanan uyum indeksleri, yukarıdaki tabloda sunulmuştur. Analiz sürecinde ‘*Relative Multiariate Kurtosis*’ değerleri incelenmiştir. Bu değer, çok değişkenli normallik için bir göstergedir. Bu değer 1’den büyük olduğunda çok değişkenliğin normalliğine bir kanıt sunmaktadır (Hu ve Bentler, 1999). Kurulan modelin testi için Robust ML yöntemi kullanılmıştır. Genel olarak, analizde model uyumu değerlendirilirken model uyum indeksleri bütüncül olarak değerlendirilmiştir. Birinci model uyum indeksi olarak ki-kare değeri incelenmiştir. Ki-kare değeri örneklem büyüklüğünden etkilendiği için serbestlik derecesine bölünüp Tablo 14’teki kriterlere göre incelendiğinde bu değerler (χ^2 /sd) kitapçık 1 için 43,24; kitapçık 3 için 42,99 olarak bulunmuştur. Kriter değerleri ile karşılaştırıldığında kabul edilebilir uyum değerinin üstünde olduğundan model veri uyumu ki-kare değerleri için sağlanamamıştır. Örnek büyüklüğü arttıkça ki-kare değeri büyümektedir. Ki kare değeri, örneklem büyüklüğüne bağlılığından iyilik uyumunun değerlendirilmesi çok ideal değildir. Ancak geleneksel olarak rapor edilir ve diğer uyum indeksleriyle desteklenir.

Kurulan modelin uygunluğu için önemli bir indeks olarak RMSEA incelenmiştir. Bu indeks, “*kötülük indeksi*” indeksi olarak dikkate alınmaktadır. “0” değeri, en iyi uyumu belirtir. RMSEA indeksinin yüksek değerleri kötü uyumu belirtir. Ancak RMSEA’nın sıfıra eşit çıkması,

mükemmel uyumu ifade etmez. Yukarıda verilen tabloda, RMSEA değerleri kitapçık 1 için 0,05 ve kitapçık 3 için 0,05 olarak bulunmuştur. Bu değer, kriterler değeri ile kıyaslandığında mükemmel uyum gösterdiği görülmüştür.

GFI, AGFI, CFI ve NNFI (TLI) indeksleri için kriter değerleri aynı olduğu için bunlar birlikte değerlendirilmiştir. Kurulan modele ilişkin hesaplanan bu değerlerin 0,90'dan büyük olması, mükemmel uyumu göstermektedir. (Han, Kolen ve Pohlmann, 1997; Hu ve Bentler, 1999).

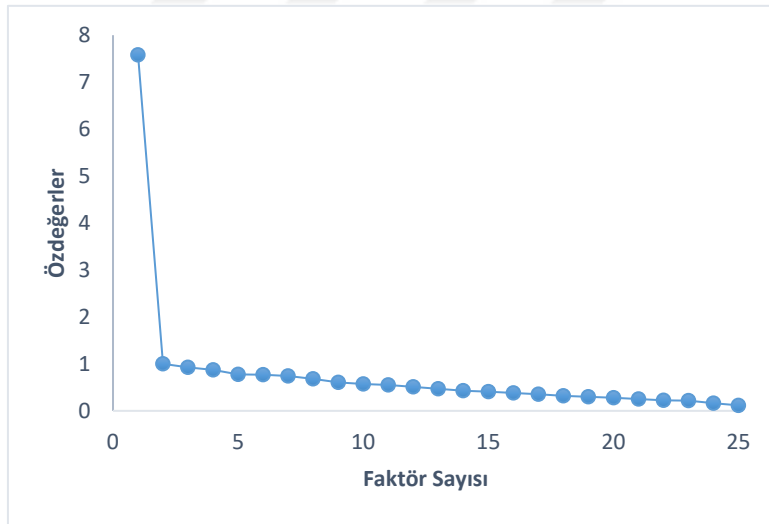
Kitapçık 1 ve kitapçık 3 için kurulan tek boyutluluk modeli için hesaplanan değerler, 0,90 kriterinden oldukça yüksektir. Model veri uyumu sağlanmıştır. Diğer bir model veri uyum indeksi olarak SRMR değerleri incelendiğinde kitapçık 1 ve kitapçık 3 değerleri sırasıyla 0,0531 ve 0,517 olarak bulunmuştur. Bu indeks değerinin, 0,05 ile 0,08 arasında bir değer aldığı görülmüştür. Bunun için modelin kabul edilebilir uyum düzeyinde olduğu söylenebilir.

Bir gizil değişkenin göstergesi olabilmek için faktör yük değerinin en az 0,30 olması ve hata varyansının ise en fazla 0,90 olması şartı bulunmaktadır (Şencan, 2005). Bu kural, standartlaştırılmış çözümdeki faktör yük değerleri için geçerlidir. Faktör yük değerleri incelendiğinde bu değerlerin 0,42 ile 0,90 arasında değer aldığı görülmüştür. İncelenen 25 maddenin, hepsi için kitapçıkta ölçülen özellik açısından iyi birer temsilcisi oldukları söylenebilir. Hata varyansları incelendiğine ise bu değerlerin 0,10 ve 0,60 arasında değer aldığı bulunmuştur. 0,90 ve üstü bir hata varyansı, o maddenin analizden çıkartılması gerektiğini belirtmektedir. Aynı faktör yükleri için yapılan yorum gibi hata varyansının da 0,90 ve üstü olan bir madde için belirtilen boyutun iyi birer göstergesi olamayacağı açıktır. Hata varyansları incelendiğine tüm maddelerinin istenen sınır değeri içinde olduğu görülmektedir.

Model veri uyumunu yorumlayabilmek ve doğru karar verebilmek için bütüncül yaklaşım seçilmelidir. Bu nedenle tüm model veri uyum indeksleri de incelenmiştir. İnceleme

sonucunda genel kanı, model veri uyumunun sağlandığı yönündedir. Bu bağlamda, tek boyutlu olarak belirlenen araştırma modellerinin, belirtilen aralıklar arasında yer aldığı ve mükemmel uyum gösterdiği söylenebilir. Bu sonuçlar doğrultusunda kitapçık 1 ve kitapçık 3 için tek boyutlu model sağlanmıştır.

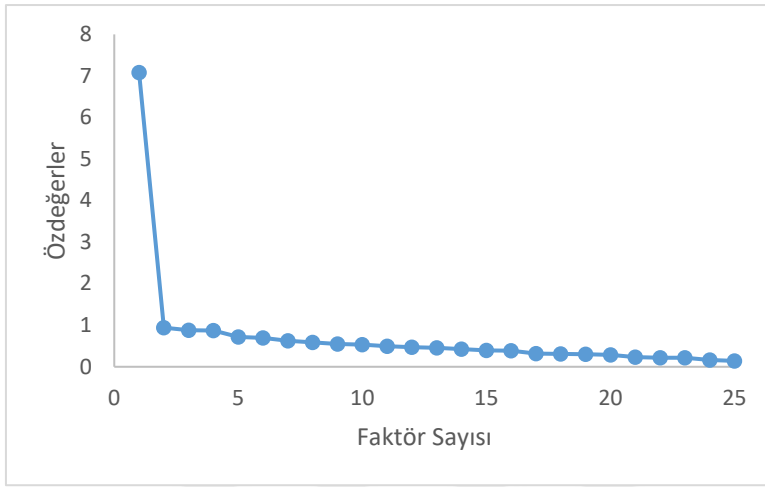
Ölçülen özelliğin tek boyutlu olmasının bir göstergesi, birinci faktöre ait özdeğerin ikinci faktöre ait özdeğerden büyük olması ve diğer faktörlerin büyük farklılıklar göstermeyen özdeğerlere sahip olmasıdır (Lord, 1980). Benzer bir şekilde Büyüköztürk (2011), tek boyutluluğun kanıtı olarak birinci faktöre ait özdeğerin, ikinci faktörün özdeğerinin 3 katından fazla olması gerektiğini belirtmiştir. Bunun dışında, birinci faktörün açıkladığı varyansın dikkate değer olmasının, tek boyutluluk için önemli olduğunu vurgulamıştır. Bu bilgiler ışığında araştırmada kullanılan kitapçıklardaki maddelerin özdeğerleri ve bu özdeğerlere ilişkin açıklanan varyans değeri EK 2’de sunulmuştur. Kitapçık 1’in maddelerine ilişkin özdeğer grafiği Şekil 14’te, kitapçık 3’ün özdeğer grafiği ise Şekil 14’te verilmiştir.



Şekil 14 Kitapçık 1’deki maddelere ilişkin özdeğer-saçılım grafiği

EK-2’te sunulan özdeğer tablosuna dayalı olarak Şekil 14’teki grafik çizilmiştir. Şekil 14’e göre birinci faktörün özdeğeri, 7,583 olarak elde edilmişken ikinci faktörün özdeğeri, 0,997 olarak hesaplanmıştır. Birinci özdeğerden ikinci özdeğere ivmeli bir düşüş gözlenmiştir.

Birinci ve ikinci özdeğer arasındaki fark ile ikinci ve üçüncü özdeğerler arasındaki fark, benzer değildir. Diğer faktörlerin özdeğerlerinin de 1'den büyük olmadığı görülmüştür. Ayrıca birinci faktörün özdeğeri ile ikinci faktörün özdeğer arasındaki oran, yaklaşık 7,60 olarak bulunmuştur. Bu bilgiler doğrultusunda, bu kitapçık için tek boyutluluk varsayımının sağlandığı görülmüştür. Şekil 15'te ise kitapçık 3'e ilişkin özdeğer ve bu özdeğerlerin faktörlere göre saçılım grafiği verilmiştir.



Şekil 15 Kitapçık 1'deki maddelere ilişkin özdeğer-saçılım grafiği

Şekil 14'te olduğu gibi Şekil 15, kitapçık 3'e ilişkin EK-3'te verilen tabloya göre çizilmiştir. Grafikte en yüksek özdeğer (7,082), birinci faktörde tespit edilmiştir. İkinci faktörün özdeğeri ise yaklaşık 0,940 olarak bulunmuştur. Diğer faktörlerin, ikinci faktörden daha büyük bir değer alamamakla birlikte benzer değerler aldığı görülmüştür. Bu kitapçıkta da birinci faktörden ikinci faktöre doğru ivmeli düşüşe rastlanmıştır. Birinci özdeğerden ikinci özdeğere ivmeli düşüşle özdeğerler arasındaki oran, 7,54 olarak hesaplanmıştır. Bulgular doğrultusunda kitapçık 3'ün de tek boyutlu bir yapıya sahip olduğu kanıtlanmıştır.

Yapılan analiz sonucu, tek boyutluluk varsayımının sağlandığı görülmüştür. Tek boyutluluğun sağlanması, eşitleme sürecinin önemli bir adımıdır. Ayrıca MTK için sağlanması gereken önemli bir varsayımdır.

MTK'daki diğ er bir varsayım ise yerel bağımsızlıktır. Yerel bağımsızlık, genellikle tek boyutluluk ile ilişkilendirilir. Hambleton ve Swaminathan (2013)'ın ve Lord (1980)'un belirttikleri üzere tek boyutluluğun sağlanması, yerel bağımsızlığın sağlandığına ilişkin veri oluşturmaktadır. Yani bu araştırmada yer alan maddelerin tek bir özelliğı ölçtüğü düşünülüyorsa herhangi bir maddenin yanıtına ilişkin hiçbir ipucunun, diğ er maddelerde yer almadığı kanatine varılabilir. Ancak bu düşünce, yerel bağımsızlığı sağlamada yeterli kanıt oluşturmamaktadır.

Tek boyutluluk ve yerel bağımsızlık için yapılan testler, artık (koşullu) kovaryanslara ve korelasyonlara odaklanıldığını göstermektedir. Tek boyutluluk için yapılan testler ortalama ve toplama odaklanırken Yen'in yerel bağımsızlık testi, ayrı ayrı madde çiftlerine odaklanmaktadır. Yerel bağımsızlığı ihlal eden iki madde, başka bir boyutu oluşturmak için yeterli olmayabilir; bu nedenle Yen'in testi, yerel bağımsızlık için daha güçlüdür (DeMars, 2010). Bu bağlamda, araştırmada yerel bağımsızlığı test etmek için Yen (1984)'in Q_3 istatistiğı kullanılmıştır. Q_3 istatistiğı için R yazılımı "sirt" paket programı kullanılmıştır. Belirtilen Q_3 istatistiğı, ölçülen özelliğı ilişkin her bir madde çifti için ürettiğı artık değ erlerin korelasyonudur. Eğ er madde çiftleri arasındaki artık korelasyon yüksek ise test geliştiriciler testten bu maddelerden birini çıkarmayı, bu maddeleri tek bir maddede birleştirmeyi ya da küçük madde grupları ile ikinci boyutu ölçmeye olanak sağlayan madde takımı modelini kullanmayı tercih edebilmektedir (DeMars, 2010). Artık değ erler arası korelasyona dayalı hesaplanan Q_3 değ erinin 0,20'yi aşmaması, yerel bağımsızlığın sağlanmasına kanıt oluşturmaktadır. Eğ er hesaplanan Q_3 değ eri 0,20'den büyük ise o zaman yerel bağımsızlık sağlanmamıştır (Yen, 1984; Zenisky, Hambleton, ve Sireci, 2001).

Araştırmada 25 maddeden oluşan madde çiftleri oluşturulmuştur. Toplamda 300 tane (25'in 2'li kombinasyonu) madde çifti elde edilmiştir. EK 4 ve EK 5'te belirtilen madde çiftlerine ilişkin Q_3 istatistiklerine yer verilmiştir. EK 4 te ve EK 5'te elde edilen değ erler, 0,20 kritik değ eri aşmadığı için her iki form için de yerel bağımsızlık sağlanmıştır. Bu varsayımdan

sonra MTK' ya ilişkin madde karakteristik eğrilerinin monoton artış varsayımları test edilmiştir.

Bir maddeyi doğru yanıtlama olasılığı, yetenek düzeyi ile doğrudan ilgilidir. Başka bir deyişle kişilerin yetenek düzeyleri ne kadar yüksekse bir maddeyi doğru yanıtlama olasılıkları o kadar yüksektir (Embretson ve Reise, 2013; Hambleton, Swaminathan, ve Rogers, 1991). Bu açıklama, MTK'da madde karakteristik eğrisinin monotonik artış varsayımını belirtmektedir. Bu varsayımı test etmek için her iki kitapçıktaki maddelerin madde karakteristik eğrileri BILOG-MG (3.0) programı ile incenmiş ve bu madde karakteris eğrileri EK 6 ve EK 7'de verilmiştir. PISA 2012 kitapçık 1 ve kitapçık 3'e ilişkin maddelerde, yetenek düzeyi arttıkça doğru yanıt verme olasılıklarının arttığı tespit edilmiştir.

MTK'nın bir diğer varsayımı ise model veri uyumudur. Seçilen modelin yanlış tanımlanmasını kontrol etmek için model ve veri uyumu belirlenmelidir. Model-veri uyumu için çoğunlukla kullanılan iki indeks bulunmaktadır. Bunlar: Pearson χ^2 ile genellikle G^2 şeklinde gösterilen log-olabilirlik χ^2 'dir. Bu araştırmada ise model veri uyumu log-olabilirlik değerlerine, ikinci durum dikkate alınarak bakılmıştır. Araştırmada 1, 2 ve 3 parametrelili modelin uyumu, R programının "sirt" paketi kullanılarak incelenmiştir. Modelin uyumunu test etmek için AIC ve BIC değerleri dikkate alınmış ve 2loglikelihood (-2ln) değerleri ANOVA ile karşılaştırılmıştır. Tablo 16' da 1 PL, 2 PL ve 3PL modeller karşılaştırılmıştır. R programına ilişkin çıktının detayı EK 8'de sunulmuştur.

Tablo 16

Model-Veri Uyum Karşılaştırması

Karşılaştırılan Modeller		-2ln	AIC	BIC
1 PLM ile 2 PLM	1 PLM	179270.8	358593.6	358804.4
	2 PLM	169223.6	338547.1	338952.4
2 PLM ile 3 PLM	2 PLM	169223.6	338547.1	338952.4
	3 PLM	169182.2	338514.3	339122.2

Tablo 16'ya bakıldığında 1 PLM' den 3 PLM ye doğru bir seçim yapıldığında -2ln, AIC ve BIC değerlerinin düştüğü gözlenmektedir. Kline (2005), bu durumu modelde parametre sayısının artışına buna bağlı olarak -2ln değerlerinin azalması halinde en uygun modelin, daha çok madde parametrelerini kullanan model olduğu şeklinde açıklamaktadır. Benzer bir şekilde Bock (1972), AIC ve BIC değerleri parametre sayısı arttıkça düşüyorsa parametre sayısının büyük olduğu modelin, model veri uyumu için daha uygun sonuç verdiğini belirtmiştir. Bu bilgiler ışığında, bu araştırma için 3 parametrelili lojistik model seçilmiştir.

Verilerin Analizi

Çalışmada öncelikle test eşitleme varsayımları test edilmiştir. KTK bünyesinde doğrusal ve EYE eşitlemede sentetik evrene yönelik evren ağırlıkları için üç farklı durum için hesaplamalar yapılmış ve eşitleme fonksiyonu kullanılarak eşitlenmiş puanlar hesaplanmıştır. EYE'de ise düzgünleştirme yöntemleri (ön düzgünleştirme- son düzgünleştirme) denenmiş ve uygun düzgünleştirme yöntemine karar verilmiş; EYE yöntemleri teker teker denenmiştir. Araştırmada, KTK ve MTK' ya dayalı eşitleme yöntemleri denenmiş ve aralarından en uygun olanı seçilmek istenmiştir. KTK bünyesindeki eşitleme yöntemlerini incelerken analitik, grafiksel çözümlerin yanında 'Equate-Error' programı ile hata değerleri incelenmiştir. Eşitleme yöntemlerinden KTK'daki doğrusal

eşitleme yöntemleri kendi aralarında kıyaslanırken EYE' deki eşitleme yöntemleri de kendi içinde kıyaslanmıştır. Eşitleme yöntemlerinde eşitlenmiş puanlar, ham puanlarla karşılaştırıldığından tüm yöntemlerdeki eşitleme hatası WMSE (Weighted Mean Square Error- Ağırlıklandırılmış Hata Kareleri Ortalaması-AHKO) katsayısı ile hesaplanmıştır. WMSE katsayısı, eşitleme çalışmalarında yaygın olarak kullanılan bir hata türüdür. Farklı eşitleme yöntemlerine göre eşitlenen puanların hataları bakımından en uygununun seçiminde kullanılır. Ağırlıklı ortalama kareler hatası (WMSE), aynı beceri düzeyindeki her bir ham skala karşılık gelen eşit puanların karşılaştırılmasıyla hesaplanır (Skaggs ve Lissitz, 1986). Eşitleme yöntemlerine ilişkin WMSE katsayısı aşağıda belirtilen eşitlik yardımı ile bulunmaktadır:

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

k : Y testindeki madde sayısı

$S^2 y$: Y testindeki ham puanların varyansı

X_{crit} : Y testindeki i. ham puan

X_E : Eşitleme yöntemleriyle elde edilen ve X testindeki i. ham puana eşit olan puan

f_i : Y testindeki i. ham puan frekansı

MTK' da ise öncelikle en uygun kalibrasyon yöntemi seçilip puanlar tek bir ölçeğe dönüştürüldükten sonra gerçek ve gözlenen puan yöntemleri karşılaştırılmıştır. Kalibrasyon sürecinde uygun ölçek dönüştürme yöntemini bulabilmek için RMSE (Root Mean Square E-Hata Kareleri Ortalamasının Karekökü) kullanılmıştır. Belirtilen RMSE indeksi gerçek yetenek kestirimi ve kestirilen yetenek arasındaki farka dayalı bir istatistik sunmaktadır (Wingersky ve Lord, 1984; Yen ve Candell, 1991; Fitzpatrick ve Skorupski, 2016). RMSE katsayısının hesaplanması için kullanılacak eşitlik aşağıda verilmiştir:

$$RMSE = \sqrt{\frac{\sum_i f_i (\theta^* - \theta)^2}{\sum_i f_i}}$$

θ : Gerçek yetenek düzeyi

θ^* : Kestirilen yetenek düzeyi

f : Frekans

Süreçte MTK'da kalibrasyon ve eşitleme yöntemlerinden en uygun eşitleme yöntemi seçmek için RMSE istatistiği kullanılmıştır.



BÖLÜM IV

BULGULAR VE YORUMLAR

Bu bölümde araştırma sorularına ilişkin bulgulara ve bu bulgulara yönelik yapılan yorumlara yer verilmiştir.

Birinci Araştırma Sorusuna Yönelik Bulgular ve Yorumları:

Klasik Test Kuramı'nda yer alan Tucker eşitleme yöntemi, Levine gerçek puan, Levine gözlenen puan, konjeneri ve Braun-Holland doğrusal eşitleme yöntemleri kullanılarak PISA 2012'ye ait farklı kitapçıklardan elde edilen eşitlenmiş puanlar için hangi eşitleme yöntemi en az eşitleme hatası içermektedir?

Aşağıda yer alan Tablo 17'de doğrusal eşitleme işlemine başlamadan önce X e Y formuna ilişkin ortalama ve standart sapma değerlerine yer verilmiştir.

Tablo 17

Form X'i Y'ye Eşitlemede Kullanılan Veri İçin Doğrudan Gözlemlenen İstatistikler

Gruplar	Formlar	$\hat{\mu}$	$\hat{\sigma}$	Kovaryans	Korelasyon
1	X	11,31442	6,0759	19,820084	0,956484
1	V	5,964602	3,4123		
2	Y	11,42579	6,3497	20,900286	0,949149
2	V	6,021579	3,4688		

$n_1=1921$ $n_2=1900$

Tablo 17 incelendiğinde X form puan ortalaması ($\hat{\mu}$) 11,31, Y form puan ortalaması ($\hat{\mu}$) 11,43 olarak bulunmuştur. Ortak maddelere ilişkin ortalamalar sırasıyla verildiğinde 5,97 ve 6,02'dir. X form puanları (toplam) ile X formundaki ortak maddeler arasındaki kovaryans 19,82 bulunurken, Y form puanları (toplam) ile Y formundaki ortak maddeler arasındaki kovaryans 20,90 olarak hesaplanmıştır.

Ortak madde ve ölçeğin toplam puanları arasında yüksek korelasyon elde edilmesi, ortak maddelerin iyi bir temsilci olduğuna kanıttır (Zu ve Yuan, 2012). X formu toplam puanları ile X formunun ortak maddelerinden elde edilen puanlar arasında korelasyon incelendiğinde yaklaşık +0,96 bir değer elde edilmiştir. Bu bulgu X ve V_X arasında yüksek pozitif bir korelasyon olduğunu ve V_X maddelerinin ölçme aracının toplam maddeleri için iyi birer temsilci olduğuna kanıttır. Bir diğer sonuç ise Y formu toplam puanları ile Y formunun ortak maddelerinden elde edilen puanlar arasında korelasyon ile elde edilmiştir. İncelenen katsayı yaklaşık +0,95 olarak bulunmuştur. Bu bulgu X ve V_X arasında yüksek pozitif bir korelasyon olduğunu ve V_X maddelerinin ölçme aracının toplam maddeleri için iyi birer temsilci olduğuna kanıttır.

Bu araştırmada doğrusal eşitleme yöntemi kullanılarak Tucker iç ortak puanları $w_1=1$, $w_1=0,50$ için eşitlenmiş puanlar hesaplanmıştır. Tablo 18'de Tucker iç ortak yöntemine göre bulgular sunulmuştur.

Tablo 18

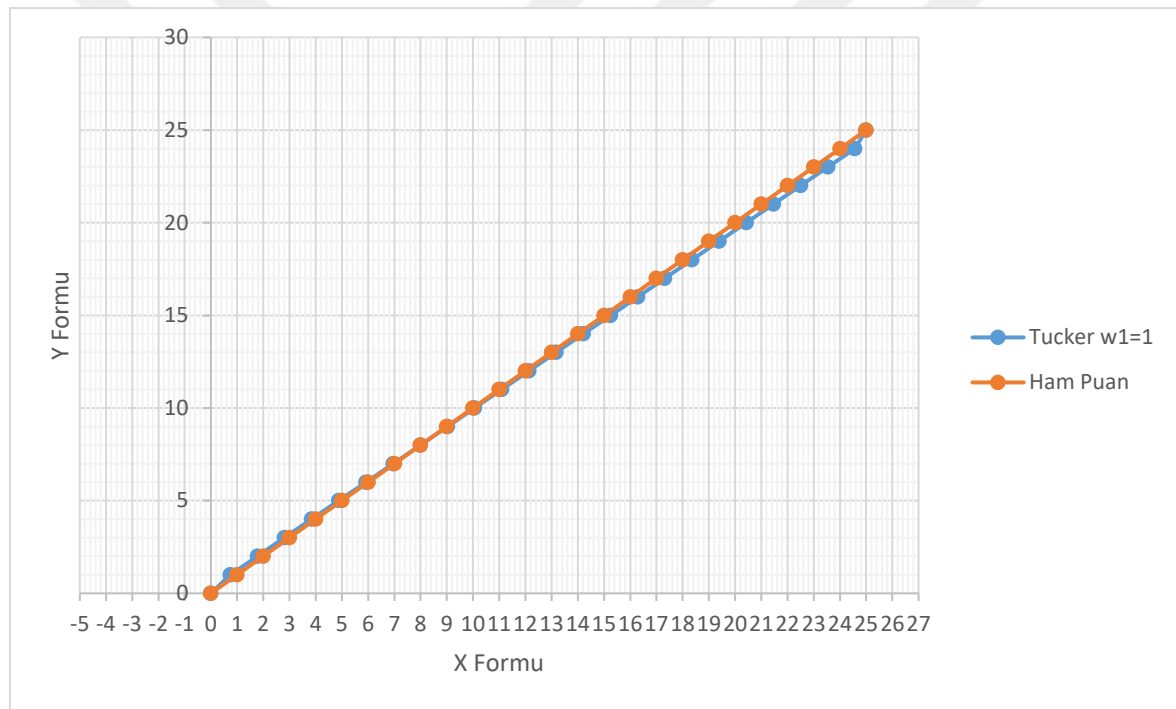
Tucker İç Ortak Eşitleme Yöntemi Kullanılarak Elde Edilen Eşitlenmiş Puanlar ve Fark Değerleri

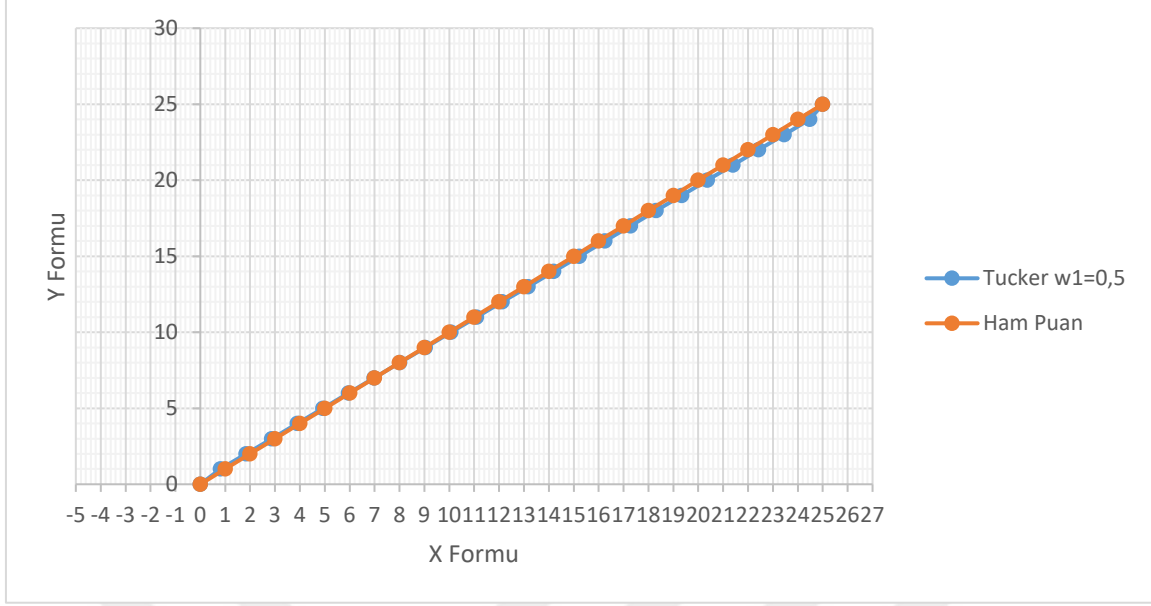
Tucker-İç Ortak				
Ham Puanlar	w1=1 için Eşitlenmiş Puanlar	Fark	w1=0,5 için Eşitlenmiş Puanlar	Fark
0	0		0	0
1	0,7374	0,2626	0,8071	0,1929
2	1,7737	0,2263	1,8366	0,1634
3	2,8099	0,1901	2,8661	0,1339
4	3,8462	0,1538	3,8956	0,1044
5	4,8824	0,1176	4,9251	0,0749
6	5,9187	0,0813	5,9546	0,0454
7	6,9549	0,0451	6,9841	0,0159
8	7,9912	0,0088	8,0136	-0,0136
9	9,0275	-0,0275	9,0431	-0,0431
10	10,0637	-0,0637	10,0726	-0,0726
11	11,1	-0,1	11,1021	-0,1021
12	12,1362	-0,1362	12,1316	-0,1316
13	13,1725	-0,1725	13,1611	-0,1611
14	14,2087	-0,2087	14,1906	-0,1906
15	15,245	-0,245	15,2201	-0,2201
16	16,2813	-0,2813	16,2496	-0,2496
17	17,3175	-0,3175	17,2791	-0,2791
18	18,3538	-0,3538	18,3086	-0,3086
19	19,39	-0,39	19,3381	-0,3381
20	20,4263	-0,4263	20,3676	-0,3676
21	21,4625	-0,4625	21,3971	-0,3971
22	22,4988	-0,4988	22,4266	-0,4266
23	23,5351	-0,5351	23,4561	-0,4561
24	24,5713	-0,5713	24,4856	-0,4856
25	25	0	25	0

Tablo 18’de öncelikle $w_1=1$ ağırlığı için bulgu sunulmuştur. Bu durumda $w_2=0$ olarak kabul edilmiştir. İlk sütunda ham puanlar ikinci sütunda eşitlenmiş puanlar, üçüncü sütunda ise fark puanları verilmiştir. Eşitlenmiş puanlar 0 değerinden daha düşük çıktığında 0’a, kitapçıklardan en yüksek alınacak 25 değerden daha büyük çıktığında ise 25’e denk getirilerek ölçek puanlarına dönüştürülmüştür. Yapılan bu işleme kesme adı verilir (Kolen ve Brennan, 2014). Kesme puanlar bu tablo ile birlikte diğer eşitlenmiş puanlar için de kullanılmıştır. Bu tablodan sonraki tüm eşitlenmiş puan tablolarında da fark puanları ham puandan eşitlenmiş puanlar çıkartılarak elde edilmiştir. $w_1=1$ için 0-8 puan aralığı için

eşitlenmiş puanlar ham puanlardan daha düşük, 9-25 arasındaki puan aralığındaki eşitlenmiş puanlar, ham puanlardan daha yüksek olduğu görülmektedir. Sentetik evren ağırlığı olarak seçilen 1 değeri için hesaplanan fark puanları -0,5713 ile 0,1929 arasında değişmektedir. Tucker $w_1=0,50$ için 0-7 puan aralığı için eşitlenmiş puanlar ham puanlardan daha düşük, 8-25 arasındaki puan aralığındaki eşitlenmiş puanlar, ham puanlardan daha yüksek olduğu görülmektedir. $w_1=0,5$ için fark puanları olarak eşitlenmiş puanlar ise -0,4856 ile 0,1929 arasında değiştiği görülmüştür. Bu tablodan sonraki tüm eşitlenmiş puan tablolarında da fark puanları ham puandan eşitlenmiş puanlar çıkartılarak elde edilmiştir.

PISA 2012 matematik ham puanlar ve Tucker iç ortak eşitleme yöntemi kullanılarak sentetik evren ağırlıklarına göre eşitlenmiş puanlar Şekil 16'da verilmiştir.





Şekil 16 PISA 2012 Ham puanların Tucker iç ortak eşitleme yöntemiyle eşitlenmesi

Şekil 16 incelendiğinde üç farklı sentetik evren ağırlıklarına göre eşitlenmiş puanları ham puandan ne kadar saptığı görülmektedir. Eşitlenmiş puanlar ve ham puanlar arasında doğrusal bir ilişki olduğu belirtilmiştir. Bu ilişki $w_1=1$ için $Y=0,965X+0,2884$; $w_1=0,5$ için $Y=0,9713X+0,216$ olarak ifade edilmiştir. Her üç sentetik evren ağırlığı için pozitif yönlü yüksek bir ilişki olduğu söylenebilir. Uç puanlara doğru ham puan ve eşitlenmiş puanların saptığı görülmektedir.

Tucker yönteminden sonra sentetik evren ağırlıklarını dikkate alan diğer bir doğrusal eşitleme yöntemi olan Levine gözlenen puan eşitleme yönteminden elde edilen eşitlenmiş puanlar aşağıdaki tabloda sunulmuştur.

Tablo 19

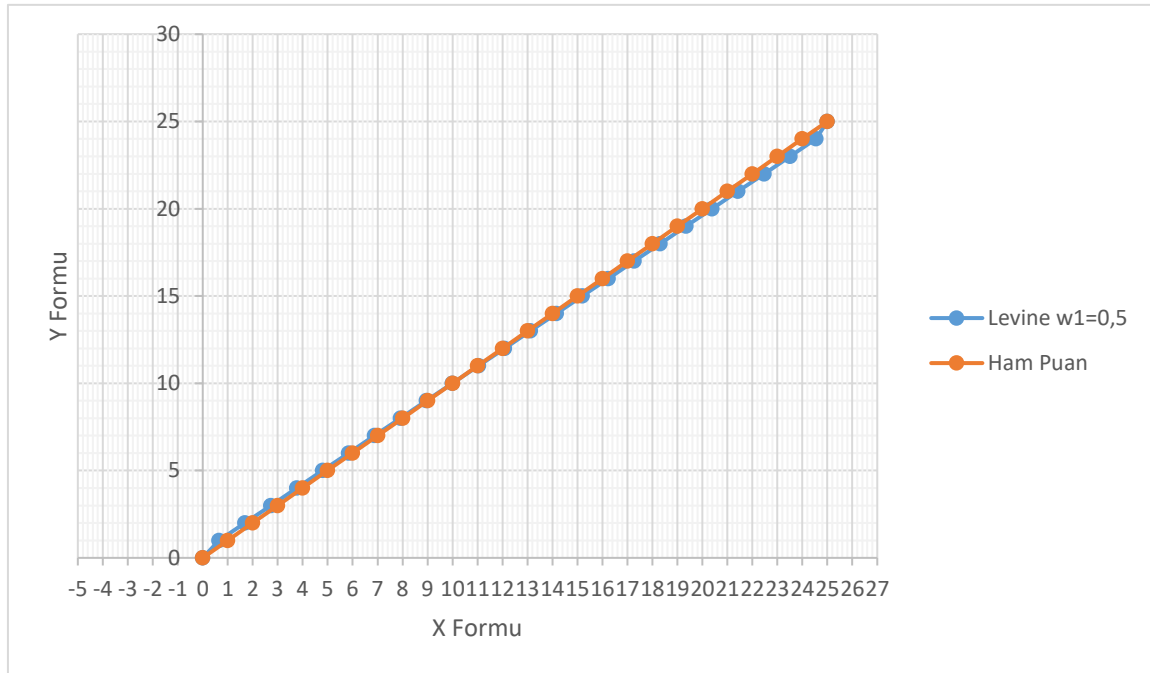
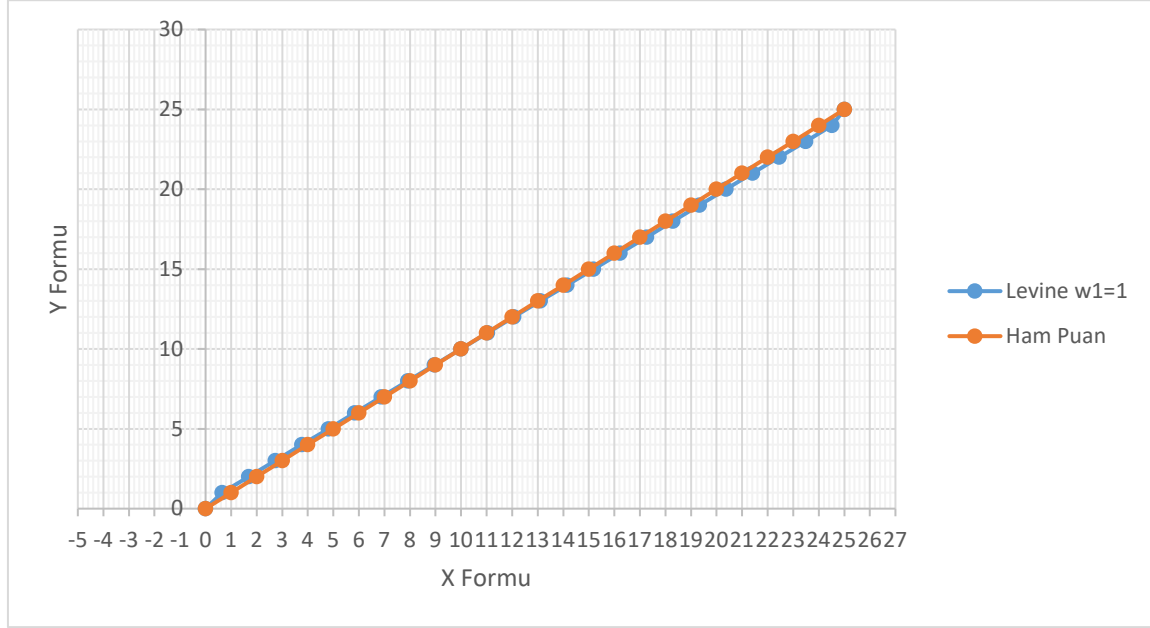
Levine Gözlenen Eşitleme Yöntemi Kullanılarak Elde Edilen Eşitlenmiş Puanlar ve Fark Değerleri

Levine Gözlenen				
Ham Puanlar	$w_1=1$ için Eşitlenmiş Puanlar	Fark	$w_1=0,5$ için Eşitlenmiş Puanlar	Fark
0	0	0	0	0
1	0,6535	0,3465	0,6459	0,3541
2	1,6913	0,3087	1,6849	0,3151
3	2,729	0,271	2,7239	0,2761
4	3,7668	0,2332	3,7629	0,2371
5	4,8045	0,1955	4,8019	0,1981
6	5,8423	0,1577	5,841	0,159
7	6,8801	0,1199	6,88	0,12
8	7,9178	0,0822	7,919	0,081
9	8,9556	0,0444	8,958	0,042
10	9,9933	0,0067	9,997	0,003
11	11,0311	-0,0311	11,036	-0,036
12	12,0688	-0,0688	12,075	-0,075
13	13,1066	-0,1066	13,114	-0,114
14	14,1444	-0,1444	14,153	-0,153
15	15,1821	-0,1821	15,192	-0,192
16	16,2199	-0,2199	16,2311	-0,2311
17	17,2576	-0,2576	17,2701	-0,2701
18	18,2954	-0,2954	18,3091	-0,3091
19	19,3331	-0,3331	19,3481	-0,3481
20	20,3709	-0,3709	20,3871	-0,3871
21	21,4087	-0,4087	21,4261	-0,4261
22	22,4464	-0,4464	22,4651	-0,4651
23	23,4842	-0,4842	23,5041	-0,5041
24	24,5219	-0,5219	24,5431	-0,5431
25	25	0	25	0

Tablo 19’da 0-25 arasında değişen ham puanlara karşılık, eşitlenmiş puanlar tabloda sunulmuştur. İlk olarak $w_1=1$ ve $w_2=0$ olarak kabul edilmiştir. Levine gözlenen eşitleme $w_1=1$ ağırlığı için 0-10 puan aralığına denk gelen eşitlenmiş puanlar ham puanlardan düşük, 11-25 arasındaki puanlar için eşitlenmiş puanlarham puanlardan daha yüksektir. $w_1=1$ için fark puanları ise en düşük -0,5219 ve en yüksek 0,3465 arasında değişmektedir. Levine gözlenen $w_1=0,50$ ağırlığı için 0-10 puan aralığına denk gelen eşitlenmiş puanlar ham

puanlardan düşük, 11-25 arasındaki puanlar için eşitlenmiş puanlar ham puanlardan daha yüksektir. Fark puanları ise -0,5431 ile 0,3541 arasında değer almaktadır.

Levine gözlenen puan eşitleme yönteminde $w_1=1$ ve $w_1=0,50$ ağırlıkları kullanılarak elde edilen eşitlenmiş puanlar ve fark değerleri Tablo 18’de verilmiştir. PISA 2012 matematik ham puanlar ve Levine gözlenen puan yöntemi kullanılarak eşitlenmiş puanlar Şekil 17’de verilmiştir.



Şekil 17 PISA 2012 Ham puanların Levine gözlenen eşitleme yöntemiyle eşitlenmesi

Şekil 17 incelendiğinde üç farklı sentetik evren ağırlıklarına göre eşitlenmiş puanları ham puandan ne kadar saptığı görülmektedir. Eşitlenmiş puanlar ve ham puanlar arasında doğrusal bir ilişki olduğu belirtilmiştir. Bu ilişki $w_1=1$ için $Y=0,9636X+0,3703$; $w_1=0,5$ için $Y=0,9625X+0,3783$ olarak ifade edilmiştir. Her üç sentetik evren ağırlığı için pozitif yönlü yüksek bir ilişki olduğu söylenebilir. Uç puanlara doğru ham puan ve eşitlenmiş puanların saptığı görülmektedir.

Levine gözlenen puan eşitlemeden sonra gerçek puanların kullanıldığı Levine gerçek puan eşitleme yönteminden elde edilen eşitlenmiş puanlar aşağıdaki tabloda sunulmaktadır.



Tablo 20

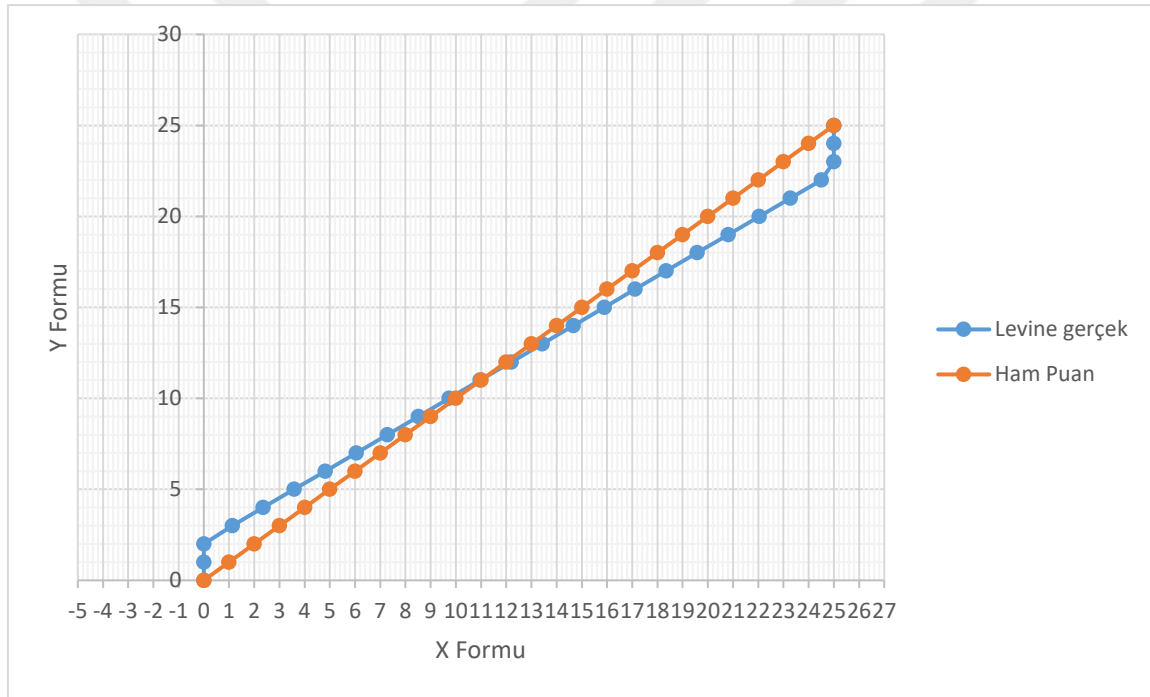
Levine Gerçek Puan Eşitleme Yöntemi Kullanılarak Elde Edilen Eşitlenmiş Puanlar ve Fark Değerleri

Ham Puanlar	Levine Gerçek Puan eşitleme	
	İçseler için	Fark
0	0	0
1	0	0
2	0	0
3	1,1301	1,8699
4	2,3602	1,6398
5	3,5902	1,4098
6	4,8203	1,1797
7	6,0504	0,9496
8	7,2804	0,7196
9	8,5105	0,4895
10	9,7406	0,2594
11	10,9706	0,0294
12	12,2007	-0,2007
13	13,4307	-0,4307
14	14,6608	-0,6608
15	15,8909	-0,8909
16	17,1209	-1,1209
17	18,3510	-1,3510
18	19,5811	-1,5811
19	20,8111	-1,8111
20	22,0412	-2,0412
21	23,2713	-2,2713
22	24,5013	-2,5013
23	25	0
24	25	0
25	25	0

Levine gerçek puan yönteminde sentetik evrendeki ağırlıklardan bağımsız olarak eşitlenmiş puanlar elde edilmiştir. Bu eşitleme yöntemi sayesinde 0-25 puan arasında değişen ham

puanlar, eşitlenmiş puanlara dönüşmüştür. Levine gerçek puan için 0-11 puan aralığı için eşitlenmiş puanlar ham puanlardan daha düşük, 12-25 puan aralığı için ise eşitlenmiş puanlar ham puanlardan daha yüksek olarak bulunmuştur. Fark puanları incelendiğinde ise en az -2,5013 ve en yüksek 1,8666 değerleri hesaplanmıştır. Bu yöntemle elde edilen eşitlenmiş puanlar ve fark puanları diğer Tucker ve Levine gözlenen puan yöntemlerinden elde edilen eşitlenmiş puanlardan ve ham puanlardan oldukça farklılaşmıştır.

Levine gerçek puan eşitleme yönteminde sentetik ağırlıkları kullanmamıştır. Bu yöntem ile elde edilen eşitlenmiş puanlar ve fark değerleri Tablo 20’de verilmiştir. PISA 2012 matematik ham puanlar ve Levine gerçek puan yöntemi kullanılarak eşitlenmiş puanlar Şekil 18’de verilmiştir.



Şekil 18 PISA 2012 Ham puanların Levine gerçek eşitleme yöntemiyle eşitlenmesi

Şekil 18 incelendiğinde Levine gerçek puanla eşitlenmiş puanların ham puandan ne kadar saptığı görülmektedir. Eşitlenmiş puanlar ve ham puanlar arasında doğrusal bir ilişki olduğu belirtilmiştir. Bu ilişki $Y=0,813X+2,0813$ olarak ifade edilmiştir. Eşitlenmiş puanlar ve ham puanlar arasında pozitif yönlü yüksek bir ilişki olduğu söylenebilir. Ham puan ve eşitlenmiş puanların diğer eşitleme yöntemlerine göre oldukça saptığı görülmektedir. Tucker, Levine gözlenen ve Levine gerçek puan eşitlemeden sonra Klasik konjenerik

modelin kullanıldığı eşitleme fonksiyonun kullanıldığı eşitleme yönteminin sonuçları aşağıdaki tabloda verilmiştir.

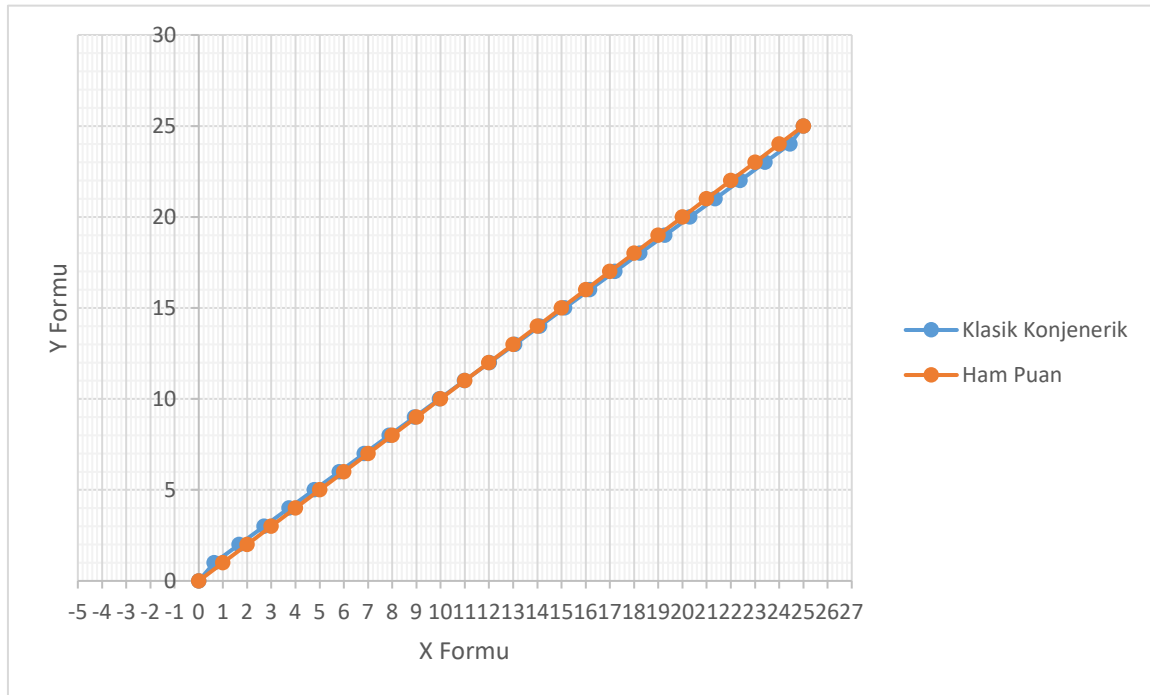
Tablo 21

Klasik Konjenerik Eşitleme Yöntemi Kullanılarak Elde Edilen Eşitlenmiş Puanlar ve Fark Değerleri

Ham Puanlar	Klasik Konjenerik	
	Gerçek	Fark
0	0	0
1	0,6333	0,3667
2	1,6690	0,3310
3	2,7047	0,2953
4	3,7404	0,2596
5	4,7761	0,2239
6	5,8118	0,1882
7	6,8475	0,1525
8	7,8832	0,1168
9	8,9188	0,0812
10	9,9545	0,0455
11	10,9902	0,0098
12	12,0259	-0,0259
13	13,0616	-0,0616
14	14,0973	-0,0973
15	15,1330	-0,1330
16	16,1687	-0,1687
17	17,2044	-0,2044
18	18,2401	-0,2401
19	19,2758	-0,2758
20	20,3115	-0,3115
21	21,3472	-0,3472
22	22,3829	-0,3829
23	23,4186	-0,4186
24	24,4542	-0,4542
25	25	0

Tablo 21’de klasik konjenerik eşitleme yöntemi ile elde edilen eşitlenmiş puanlar verilmiştir. 0 ve 25 puan arasında değişen ham puanlar bu eşitleme yöntemi kullanıldıktan sonra 0-11 puan aralığındaki puanlar için eşitlenmiş puanlar ham puanlardan daha düşük, 12-25 puan aralığında ise eşitlenmiş puanlar ham puanlardan daha yüksek olduğu görülmektedir. Ham puanlar ve eşitlenmiş puanlar arasındaki fark puanları ise -0,4542 ile 0,3667 arasındadır.

Konjenerik puan eşitleme yönteminden elde edilen eşitlenmiş puanlar ve fark değerleri Tablo 21’de verilmiştir. PISA 2012 matematik ham puanlar ve konjenerik eşitleme yöntemi kullanılarak eşitlenmiş puanlar Şekil 19’da verilmiştir.



Şekil 19 PISA 2012 Ham puanların konjenerik eşitleme yöntemiyle eşitlenmesi

Şekil 19’da ham puanlara karşılık konjenerik eşitlenmiş puanların dağılımı verilmiştir. Eşitlenmiş puanlar ve ham puanlar arasında doğrusal bir ilişki olduğu belirtilmiştir. Bu ilişki $Y=0,9655X+0,3885$ olarak ifade edilmiştir. Eşitlenmiş puanlar ve ham puanlar arasında pozitif yönlü yüksek bir ilişki olduğu söylenebilir. Genellikle uç değerlerin eşitlenmiş puanlarında sapma olduğu görülmektedir.

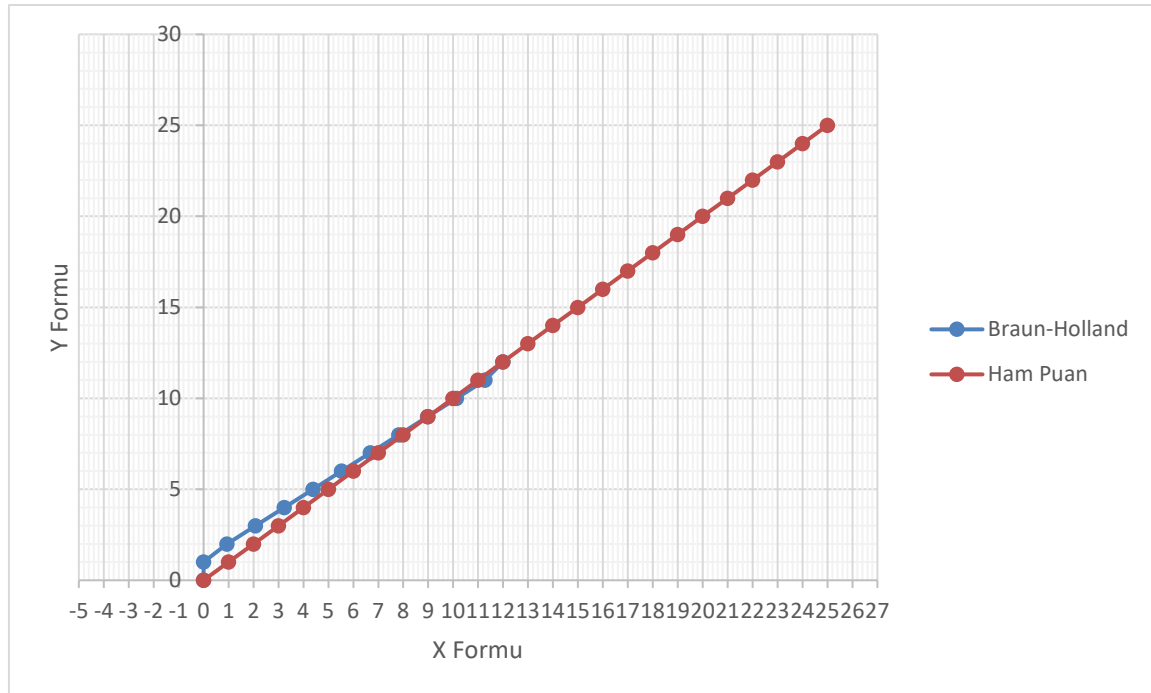
Kitapçıklarda yer alan 12 ‘si ortak olmayan, 13 ‘ü ortak toplamda 25 madde bulunmaktadır. Her bir maddenin ortak madde ile olan istatistiksel ilişkisinden yararlanarak Braun-Holland eşitleme yöntemi kullanılmıştır. Elde edilen eşitlenmiş puanlar aşağıdaki tabloda verilmiştir.

Tablo 22

Braun-Holland Yönteminden Elde Edilen Eşitlenmiş Puanlar

Braun-Holland		
Ham Puan	Eşitlenmiş puan	Fark
0	0	0
1	0	0
2	0,9335	1,0665
3	2,0835	0,9165
4	3,2334	0,7666
5	4,3833	0,6167
6	5,5333	0,4667
7	6,6832	0,3168
8	7,8331	0,1669
9	8,9831	0,0169
10	10,133	-0,133
11	11,2829	-0,2829
12	12	0

Tablo 32’de 0-12 ham puan aralığının Braun-Holland eşitleme yöntem ile elde edilen eşitlenmiş puanlar verilmiştir. Eşitlenmiş puan dağılımları incelendiğine 0-9 puan aralığında eşitlenmiş puanlar ham puanlardan daha düşük, 10-12 puan arasında ise ham puanlar eşitlenmiş puanlardan daha düşük değerler aldıkları görülmektedir. Braun-Holland için fark puanları en yüksek 1,0665 en düşük ise -0,2829 olarak hesaplanmıştır. Elde edilen eşitlenmiş puan ve ham puan dağılımları aşağıdaki Şekil 20’de verilmiştir.



Şekil 20 Braun-Holland yönteminden elde edilen EYE puanları

Şekil 28’de ham puanlara karşılık Braun-Holland eşitlenmiş puanların dağılımı verilmiştir. Eşitlenmiş puanlar ve ham puanlar arasında doğrusal bir ilişki olduğu belirtilmiştir. Bu ilişki $Y=1,1499X-1,3663$ olarak ifade edilmiştir. Eşitlenmiş puanlar ve ham puanlar arasında pozitif yönlü yüksek bir ilişki olduğu söylenebilir. Genellikle uç değerlerin eşitlenmiş puanlarında sapma olduğu görülmektedir.

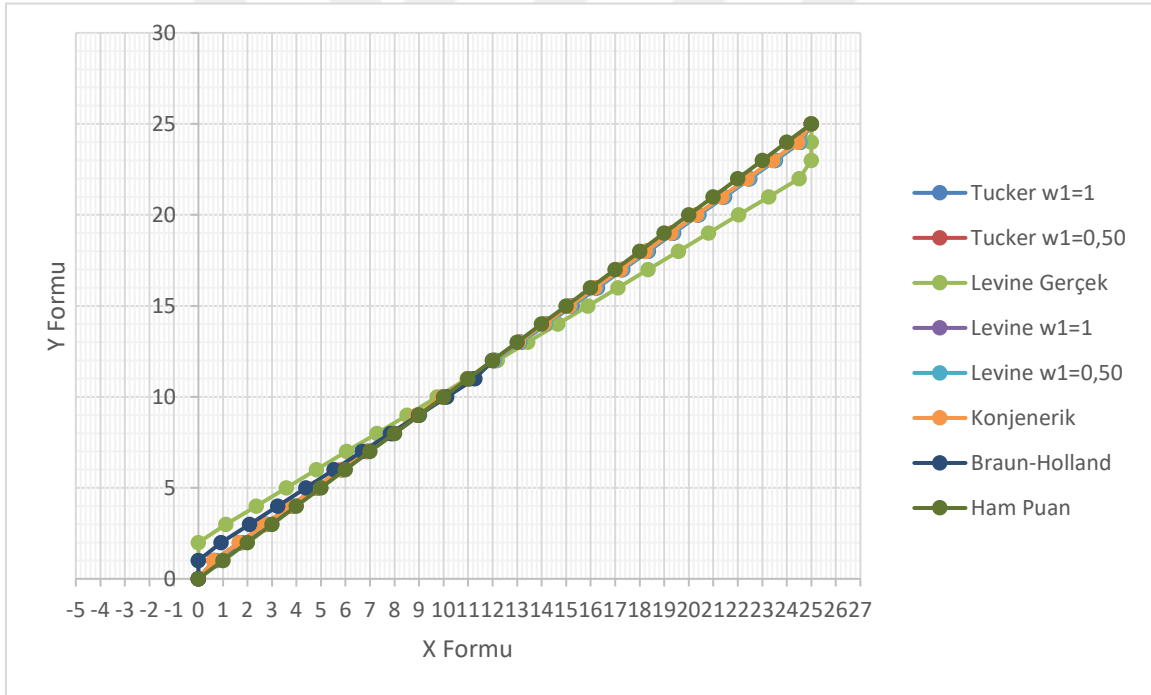
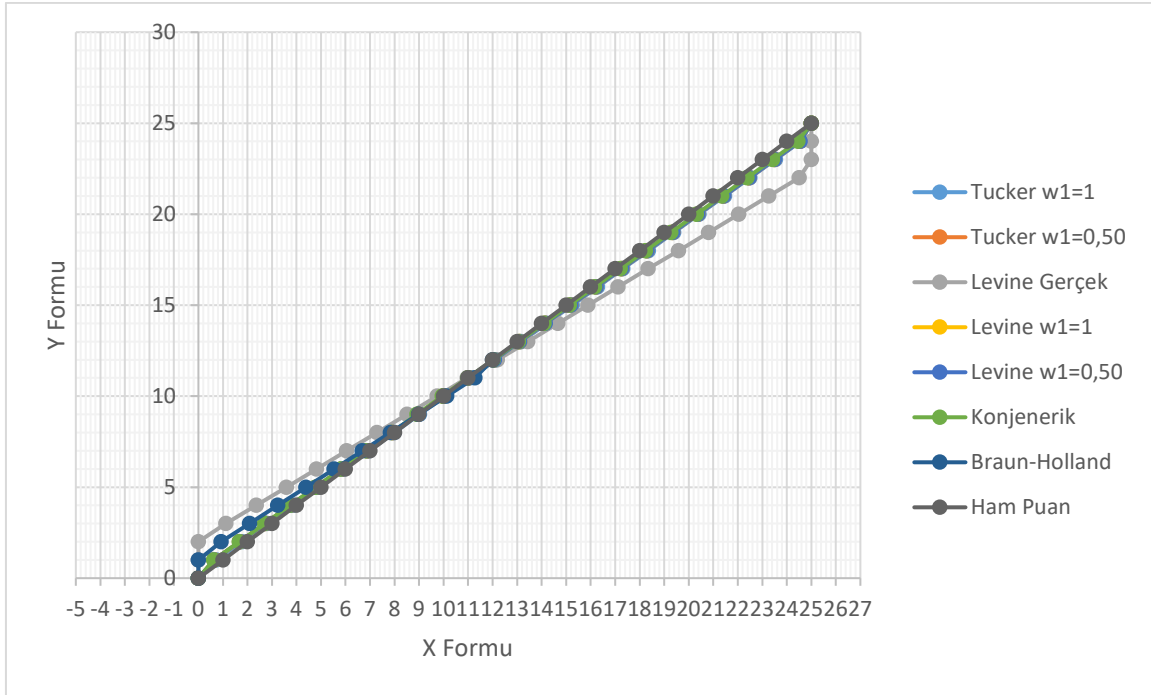
Araştırmada yer alan tüm doğrusal eşitleme yöntemlerine ilişkin eşitleme fonksiyonları elde edilip eşitlenmiş puanlar hesaplanmıştır. Hesaplanan eşitlenmiş puanlar için parametre değerleri bulunmuştur. Hesaplanan parametre değerler aşağıdaki tabloda verilmiştir.

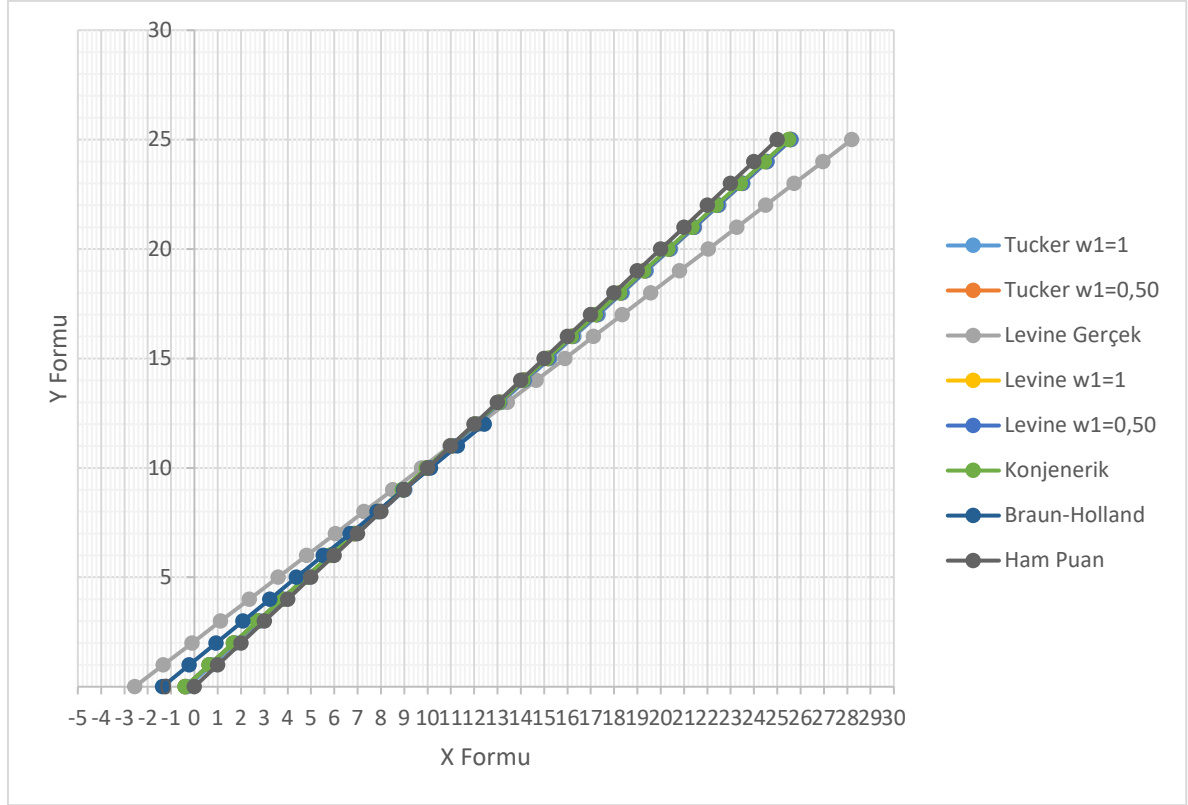
Tablo 23

KTk’deki Doğrusal Eşitleme Yöntemi Kullanılırken Dikkate Alınan Parametre Değerleri

w	Eşitleme Yöntemi	γ_1	γ_2	$ms(X)$	$ms(Y)$	$\sigma_s(X)$	$\sigma_s(Y)$
0,5	Tucker	1,702	1,737	11,363	11,376	6,1228	6,3035
0,5	Levine Gözlenen	0,976	1,201	11,342	11,391	6,0942	6,3317
1	Tucker	1,702	1,737	11,314	11,525	6,0759	6,2962
1	Levine Gözlenen	0,976	1,201	11,314	11,357	6,0759	6,3053
-	Levine Gerçek	0,976	1,201	11,365	11,356	6,1064	7,475
-	Konjenerik	1,863	1,929	11,316	11,314	6,2911	6,5763
-	Braun-Holland	1,1499	-1,3664	5,3706	4,8094	8,9733	11,8658

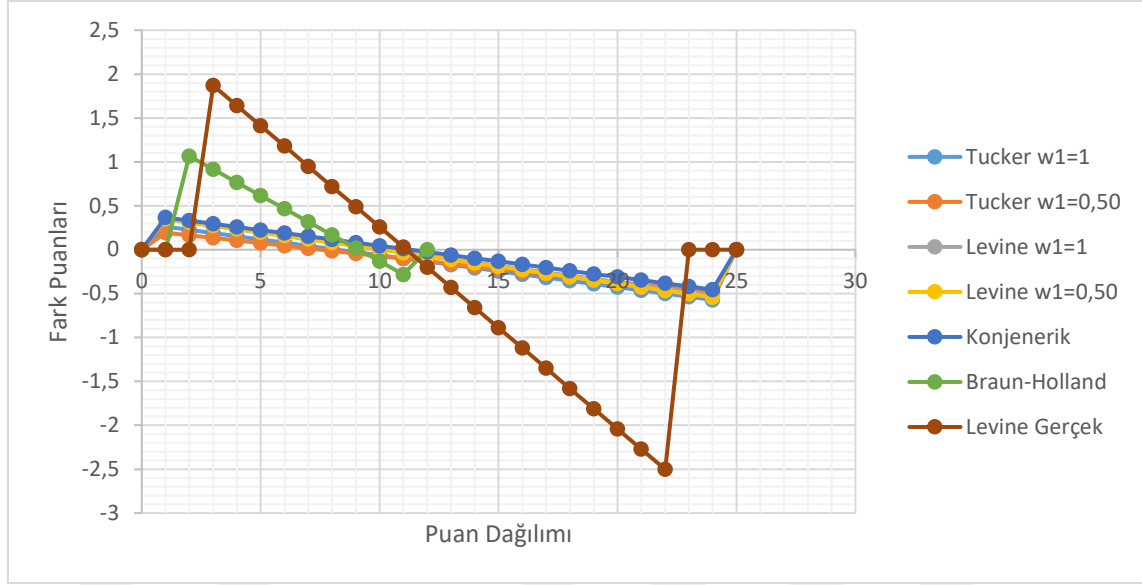
Ortak maddeli denk olmayan gruplar deseni kullanılarak, doğrusal eşitleme yöntemleri sırasıyla PISA 2012 verilerine uygulanmıştır. Eşitlenmiş puanların kullanıldığı doğrusal eşitleme yöntemlerin puan dağılımları Şekil 21’de verilmiştir.





Şekil 21 Doğrusal eşitleme yöntemlerine göre eşitlenmiş puanların dağılımı

Şekil 21 incelendiğinde ham puanlar, Tucker içsel ($w_1=1$, $w_1=0,50$), Levine gözlenen ($w_1=1$, $w_1=0,50$), Levine gerçek, Braun-Holland ve kojenerik eşitleme yöntemlerinden elde edilen eşitlenmiş puanların dağılımının belirtildiği görülmektedir. Ham puanlardan en fazla sapma, Levine gerçek puan yönteminden elde edilen eşitlenmiş puanlar tarafından elde edilmiştir. Şekil 21’de belirtilen eşitlenmiş puan ve ham puan arasındaki fark değerlerini daha iyi anlayabilmek için Şekil 22’de doğrusal eşitleme yöntemlerine ilişkin fark grafiği Şekil 22’de sunulmuştur.



Şekil 22 Doğrusal eşitleme yöntemlerine göre fark puanların dağılımı

Yukarıdaki Şekil 22’de eşitlenmiş puanlar ile ham puanlar arasındaki fark puanların dağılımları grafiksel olarak verilmiştir. Bütün doğrusal yöntemlerde eşitlenmiş puanlar, 12-13 ham puana kadar daha düşük değer alırken 13 ham puanından daha büyük değerler için eşitlenmiş puanlar, ham puanlardan daha yüksek değerler alma eğilimindedir. Grafik incelendiğinde Levine gerçek puan yönteminde eşitlenmiş puanlar, ham puanlardan sapma durumunun diğer yöntemlerden belirgin bir şekilde farklı ve fazla olduğu görülmektedir. Görsel olarak verilen eşitlenmiş puan dağılımların ve fark puanlarının dağılımlarına ilişkin grafikler incelense de analitik süreç ile eşitleme yönteminin hata değerleri hesaplanmıştır. Hata katsayıları belirlenmesi için Equating-Error_wg (v2.0) programı kullanılmıştır. Kullanılan programda bootstrap yöntemi kullanılarak hatalar hesaplanmıştır. Doğrusal yöntemlerden elde edilen hata değerleri EK9- EK13 arasında sunulmuştur. 500 replikasyon ile doğrusal yöntemlere ilişkin hata değerleri Tablo 24’te verilmiştir.

Tablo 24

Doğrusal Eşitleme Yöntemlerinden Elde Edilen WMSE Değerleri

Doğrusal Eşitleme Yöntemleri	Sentetik Evren Ağırlıkları	Bootstrap Yöntemi ile Hata Değerleri	Delta Yöntemi ile Hata Değerleri
Tucker Eşitleme	w1=1	0,160421	0,140345
	w1=0,50	0,176495	0,153678
Levine Gözlenen	w1=1	0,170873	0,151479
	w1=0,50	0,168369	0,150236
Levine Gerçek	-	0,344758	0,34028
Braun-Holland	-	0,22856	0,168854
Klasik Konjenerik	-	0,194246	0,163596

Yukarıdaki tabloda eşitleme sürecindeki yöntemler ve eşitleme hataları verilmiştir. Eşitleme hataları kestiriminde bootstrap yöntemi ve analitik süreç olan delta yöntemi karşılaştırılmıştır. Kolen ve Brennan'a (2014) göre bootstrap süreci, aynı verilerin içinden veri seti seçilip tekrar tekrar kullanılıp kestirilen bir hata değeridir. Standart hataları tahmin etmeye yönelik denklemler, hesaplama süresinin en aza indirilmesi gerektiğinde veya bir eşitleme çalışması için istenen örnek boyutlarını tahmin ederken yararlı olabilmektedir. Bu nedenle delta yöntemi de kullanılarak eşitleme hataları hesaplanmıştır. Kendall ve Stuart'a (1977) göre delta yöntemi, standart hata ifadelerini türetmek için yaygın olarak kullanılan bir istatistiksel yöntemdir. Delta yöntemi, standart hatalar için ifadelerin halihazırda mevcut olduğu bir istatistik fonksiyonu olan bir istatistiğin yaklaşık standart hatasını türetmek için kullanılır. Araştırmada delta yöntemi için Taylor açılımı kullanılarak eşitleme hataları hesaplanmıştır.

Tablo 24 incelendiğinde bootstrap yönteminde ve delta yönteminde hata değerlerinin niceliksel sıralamalarının değişmediği görülmektedir. Niceliksel olarak incelendiğinde delta yöntemindeki hata değerleri, bootstrap yöntemi ile elde edilen değerlerden daha düşük olduğu görülmektedir. Doğrusal eşitleme yöntemleri arasında en az hata değeri Tucker içsel ($w_1=1$) doğrusal eşitleme yönteminden elde edilmiştir. En fazla hata ise Levine gerçek puan eşitleme yönteminden hesaplanmıştır. Bu sonuç Şekil 20'deki grafiksel gösterim ile tutarlıdır. Tucker içsel eşitleme yöntemlerinde en az hata $w_1=1$; en fazla hata eşit ve $w_1=0,5$ olduğunda görülmektedir. Levine gözlenen eşitleme yönteminde belirtilen ağırlıklara göre eşitleme hataları sıralandığında en az hata $w_1=0,5$; en fazla hata ise $w_1=1$ olarak

bulunmuştur. En az hata içeren yöntemden en fazla hata içeren yönteme doğru bir sıralama yapıldığında sıralama şu şekildedir; Tucker içsel ($w_1=1$), Levine gözlenen ($w_1=0,50$), Levine gözlenen ($w_1=1$), Tucker içsel ($w_1=0,50$), klasik konjenerik, Braun-Holland ve Levine gerçek puan eşitlemedir.

İkinci Araştırma Sorusuna Yönelik Bulgular ve Yorumları:

Klasik Test Kuramı'nda yer alan frekans kestirim eşit yüzdelikli eşitleme yöntemleri kullanılarak PISA 2012'ye ait farklı kitapçıklardan elde edilen eşitlenmiş puanlar için hangi eşitleme yöntemi en az eşitleme hatası içermektedir?

Eşit yüzdelikli eşitleme yöntemine başlamadan önce düzgünleştirme yapmak gerekiyor. Düzgünleştirme yöntemleri ise kendi arasında ön düzgünleştirme ve son düzgünleştirme diye iki kategoriye ayrılır. Ön düzgünleştirmede puanlar öncelikle düzgünleştirilir sonra eşitleme yapılır. Bu bağlamda ön düzgünleştirme yapıldığında elde edilen momentler ve ön düzgünleştirme için uyum istatistikleri Tablo 25'te verilmiştir.

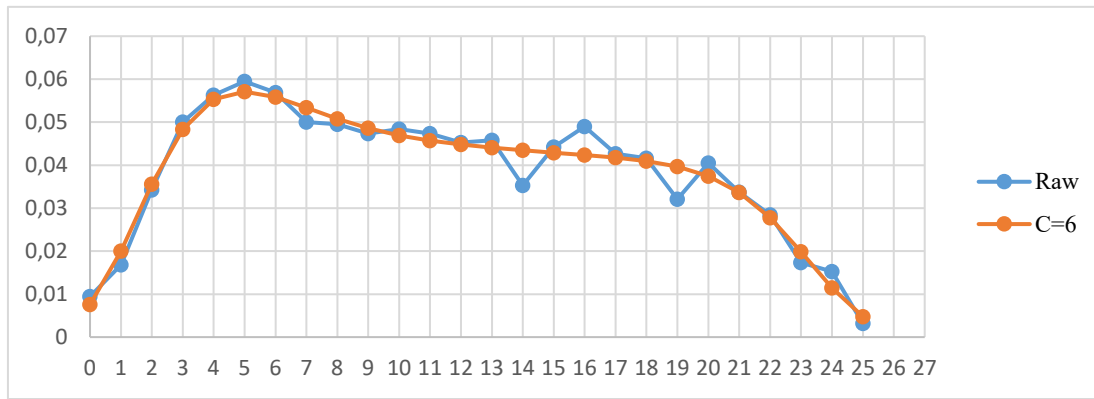
Tablo 25

Momentler ve Ön Düzgünleştirmenin Uyum İstatistikleri

Formlar	Yöntem	Ortalama	Standart Sapma	Çarpıklık	Basıklık	Chi- Squared	sd	Fark
X	Örnek	11,3144	6,074	0,233	2,037			
	Beta4 ^a	11,3144	6,07433	0,23325	2,0366	33,112	21	
	Log-linear	C=10	11,3144	6,07433	0,23325	2,0366	17,789	15
		C=9	11,3144	6,07433	0,23325	2,0366	17,782	16 0,003
		C=8	11,3144	6,07433	0,23325	2,0366	19,111	17 1,319
		C=7	11,3144	6,07433	0,23325	2,0366	19,289	18 0,178
		C=6	11,3144	6,07433	0,23325	2,0366	19,295	19 0,006
		C=5	11,3144	6,07433	0,23325	2,0366	34,468	20 12,17
		C=4	11,3144	6,07433	0,23325	2,0366	36,755	21 3,287
		C=3	11,3144	6,07433	0,23325	2,30885	123,167	22 88,41
		C=2	11,3144	6,07433	0,15049	2,25945	131,976	23 8,809
		C=1	11,3144	7,4435	0,19097	1,84675	426,087	24 294,11
Y	Örnek	11,4258	6,34966	0,186	1,935			
	Beta4	11,4258	6,34966	0,18627	1,93477	19,224	21	
	Log-linear	C=10	11,4258	6,34966	0,18627	1,93477	11,941	15
		C=9	11,4258	6,34966	0,18627	1,93477	11,985	16 0,044
		C=8	11,4258	6,34966	0,18627	1,93477	14,262	17 2,278
		C=7	11,4258	6,34966	0,18627	1,93477	14,295	18 0,033
		C=6	11,4258	6,34966	0,18627	1,93477	15,776	19 1,481
		C=5	11,4258	6,34966	0,18627	1,93477	21,123	20 5,346
		C=4	11,4258	6,34966	0,18627	1,93477	22,729	21 1,606
		C=3	11,4258	6,34966	0,18627	2,18753	120,968	22 98,23
		C=2	11,4258	6,34966	0,14582	2,1666	123,448	23 2,479
		C=1	11,4258	7,45364	0,17291	1,83768	315,546	24 192,09

Tablo 25 incelendiğinde X ve Y formlarına ilişkin dört moment (ortalama, standart sapma, basıklık, çarpıklık) değerleri ve ki kare uyum indeksleri verilmiştir. X ve Y formları için C polinomial dereceler incelendiğinde $C \geq 1$ durumlarında sadece bir moment (ortalama) değerleri, $C \geq 2$ durumlarında iki moment (ortalama ve standart sapma), $C \geq 3$ durumlarında üç moment (ortalama, standart sapma, çarpıklık), $C \geq 4$ durumlarında dört moment (ortalama,

standart sapma, çarpıklık ve basıklık) formların orijinal moment değerlerine eşit çıkmıştır. Ki-kare uyum indekslerine göre hangi polinomial derecenin uygun olduğuna bakılmak istenmiştir. $C + 1$ derecesine sahip modelin, C derecesi ile modelin üzerindeki uyumu iyileştirdiğini göstermektedir. Her bir C derecesi için 0.05'lik bir anlamlılık seviyesi kullanılarak $C=6$ polinomial derecede seçilmesinin uygun olduğu belirlenmiştir. Diğer C polinomial fonksiyon derecesi için eşitlenmiş puanlar EK 14'te, C grafikleri ise EK 15'te sunulmuştur. $C \geq 7$ durumunda ise ki-kare dağılımlarının oldukça büyüdüğü görülmüştür. İstatistiksel olarak kanıtlanmış $C=6$ derecesi grafiksel olarak aşağıda Şekil-23'te sunulmuştur.



Şekil 23 C=6 Polinomial derece için ön-düzgünleştirme ve ham puan dağılımı

Şekil 23 incelendiğinde ham puan ve bu ham puana karşılık $C=6$ derecesi kullanılıp ön-düzgünleştirme yönteminden elde edilen eşitlenmiş puanlar verilmiştir. 14, 16 ve 19 puanlarında eşitlenmiş puanlar ve ham puanlar arasında farklılaşma fazla olsada EK 15'te sunulan diğer polinomial derecelere göre $C=6$ 'nın ham puana daha yakın bir dağılım gösterdiği görülmektedir.

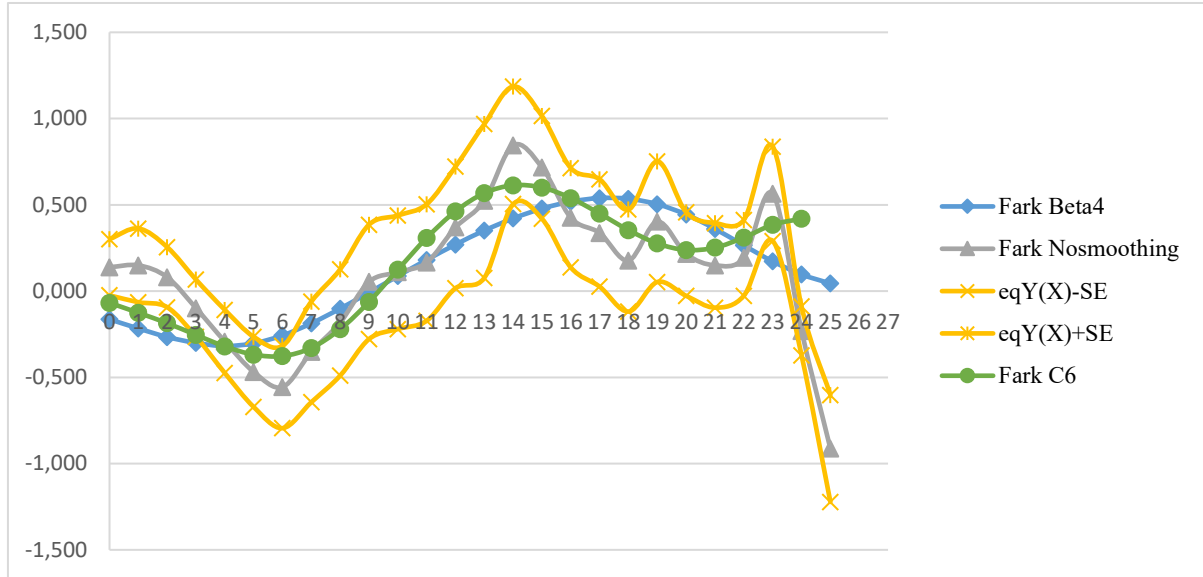
Tablo 26

Ön Düzgünleştirmeye Dayalı Eşit Yüzdelikli Eşitleme Yönteminden Elde Edilen Değerler

Form X Puanları	Standart Hata	Eşitleme Yöntemi Kullanılarak Y Formunun Eşdeğeri		
		Düzgünleştirme Yapılmadı	Log-Lineer (C=6)	Beta 4
0	0,163	0,138404	-0,06812	-0,1637
1	0,2137	1,149221	0,87427	0,78336
2	0,1742	2,080728	1,81406	1,73395
3	0,1655	2,900818	2,74600	2,69949
4	0,1831	3,707838	3,67946	3,68155
5	0,2035	4,531513	4,63179	4,69699
6	0,2384	5,443818	5,62346	5,74312
7	0,2916	6,647757	6,67088	6,81326
8	0,3081	7,818254	7,77910	7,89857
9	0,3316	9,053522	8,93798	8,99163
10	0,3292	10,10935	10,12430	10,0871
11	0,3379	11,16600	11,30846	11,1810
12	0,3525	12,36935	12,46263	12,2700
13	0,4455	13,52283	13,56775	13,3511
14	0,3409	14,84506	14,61318	14,4215
15	0,2978	15,71672	15,60005	15,4783
16	0,2881	16,42490	16,53916	16,5187
17	0,3112	17,33667	17,44920	17,5392
18	0,2956	18,17703	18,35349	18,5353
19	0,3496	19,40247	19,27609	19,5034
20	0,2425	20,21496	20,23841	20,4440
21	0,2452	21,14922	21,25196	21,3607
22	0,2197	22,19284	22,31062	22,2648
23	0,2737	23,56435	23,38476	23,1723
24	0,1414	23,76913	24,41969	24,0965
25	0,3099	24,08789	25,30893	25,0442

Ön düzgünleştirme sürecinde log-lineer yöntemlerden C 6 derecesi ve beta4 yöntemlerinden elde edilen eşitlenmiş puanlar yukarıdaki Tablo 26'da sunulmuştur. Bu tabloda X form ham puanları 0-25 arasında verilmiştir. Standart hata değerleri ve ön düzgünleştirme yapılmadan elde edilen eşit yüzdelikli eşitleme puanları verilmiştir. Ön düzgünleştirme yapılmadan elde edilen puanlar 0,138 ile 24,088 arasında değişmektedir. C=6 polinomial derecesine göre

hesaplanan log-lineer eşitlemeye göre eşitlenmiş puanlar -0,0068 ile 25,309 arasında değişirken, beta4 binominal eşitlemede ise -0,1637 ile 25,044 arasında puanlar hesaplanmıştır. EK 16’da sunulan hata bant değerleri doğrultusunda Tablo 26’da hesaplanan hata bantı ve bu bant içerisinde yer alıp yer almama durumuna göre ön-düzgünleştirme yöntemlerindeki puanlar Şekil-24’te verilmiştir.



Şekil 24 Standart hata bantına göre ön-düzgünleştirme yöntemlerinden elde edilen puanlar

Şekil 24 incelendiğinde beta4 binominal, C=6 polinomial dereceye sahip log-lineer ve ön-düzgünleştirme yapılmadan elde edilen eşit yüzdelikli eşitleme yöntemlerinden elde edilen puanları dağılımı sunulmuştur. Eşit yüzdelikli eşitleme denkleminde elde edilen standart hata bantı dikkate alınarak eşitleme yöntemlerindeki puanlardaki dağılım gösterilmiştir. Bantların üst kısmı, eşityüzdelikli eşitleme yöntemine standart hatanın eklenmesiyle oluşturulmuştur. Bantların alt kısmı bir standart hatanın çıkarılmasıyla oluşturulmuştur.

Beta4 yöntemi için gösterilen eşityüzdelikli ilişki, 0,1,2,3,14,18,24 ve 25'in Form X ham puanları haricinde standart hata bandında yer almaktadır. 20. ve 21. Puanlar hata bantın üzerine denk gelmiştir. Belirtilen bu skorlar aşırıdır. Bu puanlarda az sayıda inceleme yapıldığı ve standart eşitlikteki eşik değer denklemlerinin aşırı derecede kötü tahmin edildiği için, bu puanlar göz ardı edilebilmekte ve beta4 yönteminin uygunluğu yeterli görünmektedir. C = 6 ile loglineer yöntem için gösterilen eşityüzdelikli ilişki, 0,1,2,3, ve 24 puanı hariç tüm puanlar, standart hata bandının içinde yer almaktadır. Log-lineer yönteminin

dağılımı, beta4 yöntemi dağılımına göre daha az sapma ile standart hata bandı içerisinde yer almıştır.

Tablo 27

Ön-Düzgünleştirme için Ham Puan Momentleri

Test Formları	μ	σ	Çarpıklık	Basıklık
Form X	11,3144	6,0743	0,2333	2,0366
Form Y	11,4258	6,3497	0,1863	1,9348
Y Form Ölçeğine Eşitlenmiş X Formu				
Düzgünleştirilmemiş	11,4227	6,3455	0,1867	1,9343
Beta4	11,4256	6,3451	0,1873	1,9357
Log-Linear C=6	11,4235	6,3438	0,1849	1,9322

Tablo 27’de düzgünleştirilmemiş, ön düzgünleştirilme yapıp uygun olan polinom fonksiyonları özetlenmeye çalışılmıştır. Tablodaki parametreler ve özdüzgünleştirme sonrasında elde edilen parametreler incelendiğinde değerlerin birbirine olabildiğince yakın olduğu görülmektedir. İstatistiksel bulgularla birlikte Şekil 22 detaylı bir şekilde incelendiğinde C 6 polinom fonksiyonu kullanmanın daha uygun olduğu görülmektedir.

Düzgünleştirme yöntemlerinden ön düzgünleştirme yönteminde uygun polinom fonksiyonu tanımlandıktan sonra son düzgünleştirme için de S parametreleri test edilmiştir. Son düzgünleştirmede her bir puan için kübik eğri elde edilmektedir. Bu kübik eğriler sürekli. En düşük ve en yüksek ranjin eğri için uyumu incelenmektedir. Bunun için farklı S değerleri denenmiştir. Farklı S değerleri için 0 ile 25 puan arasında hesaplanan düzgünleştirilmiş puanlar EK 17’de ve Tablo 28’de sunulmuştur.

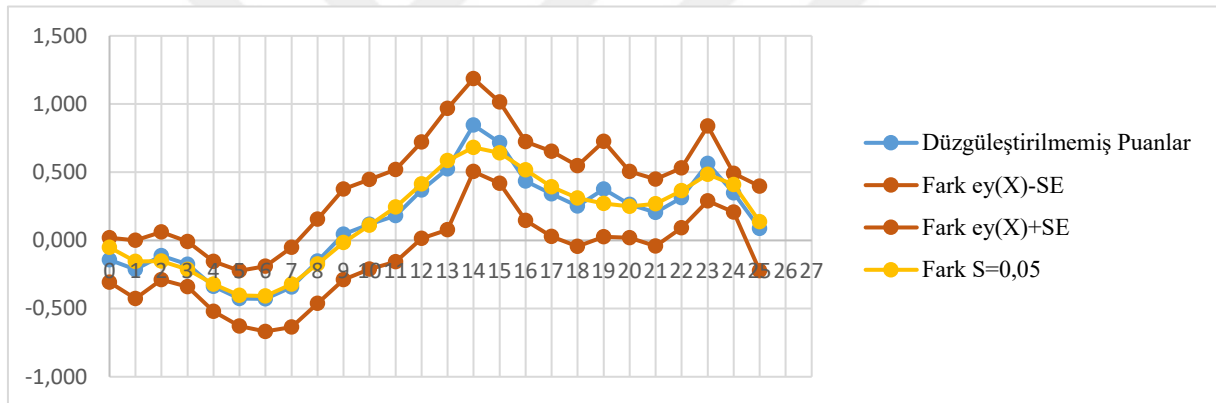
Tablo 28

Son-Düzgünleştirme için Ham Puan Dönüşümleri

Ham Puanlar	Standart Hata	Düzgünleştirilmemiş	S=0,01	S=0,05	S=0,10	S=0,20	S=0,30	S=0,40	S=0,50	S=0,75	S=1,00	Lineer
0	0,163	-0,143	-0,065	-0,052	-0,045	-0,054	-0,068	-0,081	-0,091	-0,092	-0,091	-0,401
1	0,2137	0,787	0,806	0,845	0,866	0,837	0,795	0,757	0,726	0,724	0,727	0,644
2	0,1742	1,887	1,874	1,847	1,823	1,787	1,759	1,735	1,715	1,726	1,728	1,689
3	0,1655	2,825	2,813	2,787	2,760	2,741	2,738	2,736	2,736	2,763	2,765	2,735
4	0,1831	3,663	3,667	3,679	3,684	3,702	3,723	3,742	3,76	3,801	3,802	3,780
5	0,2035	4,574	4,569	4,596	4,630	4,684	4,725	4,76	4,791	4,838	4,839	4,825
6	0,2384	5,571	5,571	5,592	5,631	5,701	5,752	5,795	5,831	5,876	5,875	5,870
7	0,2916	6,657	6,670	6,678	6,698	6,760	6,807	6,848	6,883	6,914	6,912	6,916
8	0,3081	7,847	7,848	7,828	7,817	7,855	7,888	7,917	7,943	7,951	7,949	7,961
9	0,3316	9,044	9,023	8,983	8,964	8,974	8,986	8,998	9,01	8,989	8,986	9,006
10	0,3292	10,118	10,116	10,112	10,117	10,104	10,091	10,083	10,079	10,026	10,023	10,052
11	0,3379	11,181	11,198	11,245	11,270	11,231	11,194	11,168	11,147	11,064	11,059	11,097
12	0,3525	12,368	12,364	12,413	12,417	12,344	12,287	12,245	12,21	12,101	12,096	12,142
13	0,4455	13,523	13,601	13,586	13,536	13,430	13,362	13,309	13,265	13,138	13,133	13,188
14	0,3409	14,845	14,778	14,682	14,595	14,480	14,411	14,358	14,309	14,175	14,17	14,233
15	0,2978	15,717	15,692	15,642	15,578	15,490	15,434	15,388	15,343	15,212	15,206	15,278
16	0,2881	16,435	16,485	16,517	16,507	16,468	16,435	16,402	16,366	16,248	16,243	16,324

17	0,3112	17,341	17,340	17,393	17,420	17,430	17,422	17,406	17,381	17,284	17,28	17,369
18	0,2956	18,251	18,289	18,311	18,345	18,391	18,405	18,404	18,391	18,32	18,317	18,414
19	0,3496	19,377	19,308	19,271	19,297	19,360	19,39	19,401	19,399	19,356	19,354	19,460
20	0,2425	20,263	20,262	20,249	20,278	20,346	20,383	20,401	20,407	20,392	20,39	20,505
21	0,2452	21,204	21,224	21,268	21,299	21,354	21,387	21,406	21,416	21,427	21,427	21,550
22	0,2197	22,312	22,360	22,364	22,365	22,386	22,404	22,417	22,427	22,463	22,464	22,596
23	0,2737	23,564	23,539	23,485	23,455	23,433	23,429	23,432	23,44	23,498	23,501	23,641
24	0,1414	24,349	24,403	24,409	24,404	24,383	24,372	24,371	24,379	24,44	24,444	24,686
25	0,3099	25,088	25,134	25,136	25,135	25,128	25,124	25,124	25,126	25,147	25,148	25,732

Yukarıdaki Tablo 28 incelendiğinde farklı S düzgünleştirme derecesine göre elde edilmiş puanlar görülmektedir. Hanson, Zeng ve Colton'un (1994) yapmış olduğu çalışma bulgularına göre S parametresinin çok küçük bir değer olarak seçilmesi ($S=0$) eşit yüzdelikli eşitleme fonksiyonunu gözlenen dağılımı hiç değiştirmeden eşitlemeyi yapmaktadır. S parametresinin çok büyük seçilmesi ($S=1$) eşitleme fonksiyonunun lineer bir fonksiyon olmasıyla sonuçlanır. Son düzgünleştirmede eşit yüzdelikli eşitleme sürecinden elde edilen puanlarla yapılmaktadır. Bu nedenle Segall (1997)'nda belirttiği gibi eşitleme sürecinde düzgünleştirme derecesine dikkatli bir şekilde karar vermek gerekir. Tablo sonuçları incelendiğinde S değeri küçüldükçe test puan dağılımına yakın değerler elde edilirken S 1'e yaklaştıkça da doğrusal eşitleme sonucunda elde edilen puan dağılımına yakın sonuçlar verdiği görülmüştür. Hangi S momentinin uygun olduğunu tespit etmek için öncelikle grafiksel çözüm incelenmiştir.



Şekil 25 Standart hata bantına göre S 0,05 için son-düzgünleştirme yönteminden elde edilen puanlar

Tüm S değerleri için hata bantı içindeki puan dağılımları incelenmiştir. Elde edilen grafiksel çizimler Ek 17'de verilmiştir. Tüm S değerleri için hata bantı içindeki düzgünleştirme puanlarının dağılımı incelendiğinde hata bantın dışına çıkmayan dağılım S 0,01 ve 0,05 değerlerinde elde edilmiştir. S 0,01'deki dağılım uygun olarak görülse de fonksiyon şekli pürüzlüdür. Zigzak şeklinde bir dağılım mevcuttur. S 0,05 için hata bantı dışına çıkan puan olmamakla birlikte düzgünleştirilmemiş puanlardaki sapmalara rağmen daha pürüzsüz bir dağılım görülmektedir. Grafiksel çözümlemeler incelendiğinde S 0,05 en uygun modeli yansıtmaktadır.

Uygun S momentine karar vermek için grafiksel çözüm yeterli değildir. Bütüncül bir değerlendirme için momentlerin de incelenmesi gereklidir. Bunun için aşağıdaki tablo değerleri incelenmelidir.

Tablo 29

Son-Düzgünleştirme için Ham Puan Momentleri

Test Formları	μ'	σ'	Çarpıklık	Basıklık
Form X	11,4258	6,3497	0,1863	1,9348
Form Y	11,3144	6,0743	0,2333	2,0366
Y Formu Ölçeğine Eşitlenmiş X Formu				
Düzgünleştirilmemiş	11,4227	6,3455	0,1867	1,9343
S=0,01	11,4255	6,3476	0,1874	1,9348
S=0,05	11,4257	6,346	0,1877	1,9351
S=0,10	11,4258	6,3443	0,1906	1,94
S=0,20	11,4249	6,3391	0,1992	1,9564
S=0,30	11,4234	6,333	0,2052	1,9704
S=0,40	11,4215	6,3255	0,2094	1,9832
S=0,50	11,4182	6,3163	0,2131	1,9959
S=0,75	11,388	6,2909	0,2325	2,0262
S=1,00	11,3859	6,2901	0,2336	2,0271
Lineer (Doğrusal)	11,4258	6,3497	0,2333	2,0366

Tablo 29 incelendiğinde X ve Y formlarına ilişkin momentler ve farklı S derecelerinden elde edile moment değerleri sunulmuştur. S değerlerine göre ortalama, standart sapma, çarpıklık ve basıklık değerlerine tek tek bakıldığında S 0,01 değerinin bu dört momente daha yakın değerler ürettiği görülmektedir. Hangi S derecesinin daha uygun bir model sağladığına bakmak için bütüncül bir değerlendirme yapılması gerekmektedir. S 0,01 uygun olarak görülsede grafiksel çözümü de dikkate aldığımızda daha pürüzsüz bir dağılım elde edilmektedir. Genel değerlendirme için S 0,05'in daha uygun bir model önerdiği görülmektedir. S = 0,05'in denenen değerlerin en uygun olduğunu göstermektedir. Ancak, bu örnek aşırı genelleştirilmemelidir. En küçük düzeltme değerleri her zaman en uygun denklemi üretmez. Özellikle kübik spline için, yüksek S değerleri genellikle daha yeterli sonuçlara yol açmaktadır.

Son düzgünleştirme yönteminde en uygun düzgünleştirme fonksiyonu belirlendikten sonra tüm düzgünleştirme yöntemleri birbirleri ile kıyaslanmıştır. Tek tek farklı son düzgünleştirme derecelerinin hata bant değerlerine göre dağılımları Ek 17’de verilmiştir. Kıyaslanırken analitik süreçte momentler incelenmiş grafiksel süreçte ise hata bant aralığına dikkat edilmiştir. Düzgünleştirme yöntemlerini kıyaslamak için analitik süreçler aşağıda yer alan Tablo 30’da sunulmuştur.

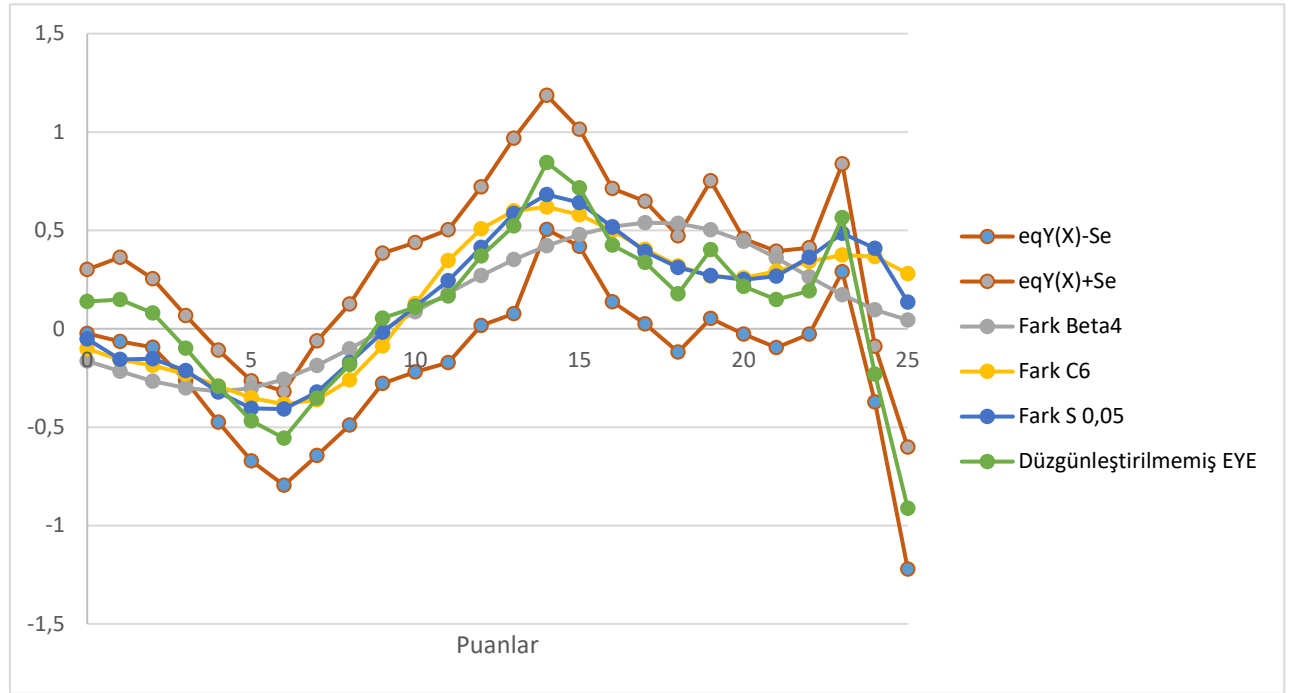
Tablo 30

Düzgünleştirme Yöntemlerinden Elde Edilen Ham Puan Momentleri

Test Formları	$\hat{\mu}$	$\hat{\sigma}$	Çarpıklık	Basıklık
Form Y	11,4258	6,3497	0,1863	1,9348
Form X	11,3144	6,0743	0,2333	2,0366
Y Form Ölçeğine Eşitlenmiş X Formu				
Düzgünleştirilmemiş	11,4227	6,3455	0,1867	1,9343
Beta4	11,4256	6,3451	0,1873	1,9357
Log-Lineer C=6	11,4235	6,3438	0,1849	1,9322
S=0,05	11,4257	6,346	0,1877	1,9351

Tablo 30’da form X, form Y, ön düzgünleştirilme yapılmamış, ön düzgünleştirme için kullanılan beta4 ve C 6 polinom fonksiyonu kullanılarak hesaplanan ve son düzgünleştirmede en uygun modeli sunan S 0,05 derecesi için hesaplanan değerler için hesaplanan dört moment (ortalama, standart sapma, basıklık ve çarpıklık) bilgileri sunulmuştur. Eski formun yeni forma eşitleme için geçen süreçte en uygun düzgünleştirmenin hangi yöntem olduğuna ilişkin karar vermek için momentler yukarıdaki tablo değerlerine bakılıp incelendiğinde beta4 ön düzgünleştirme yönteminin Y formundaki değerlere daha yakın çıktığı görülmektedir. Düzgünleştirme yöntemleri analitik olarak incelendiğinde beta4 ön düzgünleştirme yönteminden sonra kübik spline S 0,05 derecesinin dört moment için daha yakın değerler sağladığı görülmektedir. Momentlerdeki farklılaşma ise en fazla Log-Lineer C 6 ön düzgünleştirme fonksiyonunda elde edilmiştir. Analitik süreçten sonra grafiksel çözümlemede hata bant aralığında düzgünleştirme yöntemleri kıyaslanmıştır. 0-25

arasındaki puan dağılımına ilişkin puanların hata bantındaki dağılımları aşağıdaki Şekil 26’da verilmiştir.



Şekil 26 Standart hata bantına göre düzgünleştirme yöntemleri fark değerlerinin dağılımları

Yukarıdaki şekilde +,- eşit yüzdelikli hata değerleri arasında yer alan düzgünleştirme yöntemleri fark değerleri verilmiştir. 0,1,2,24 ve 25 ham puanına ilişkin kullanılan ön ve son düzgünleştirme yöntemleri hata bant değerinin dışına çıkmıştır. 3, 6,14 ve 18 ham puanlarında ise beta4 ön düzgünleştirme yönteminde fark puanları hata bant aralığının dışına çıkmıştır. Belirtilen dokuz puan için hata bant aralığının dışına çıkan puan dağılımına sahip olan beta 4 yöntemi, diğer düzgünleştirme yöntemlerine göre daha düzgün bir dağılım göstermektedir. Fark puan dağılımları incelendiğinde hata bant değerleri içerisinde yer alan fark puanları dağılımlarında düzgün dağılım gösteren ikinci yöntem ise log-lineer C 6 yöntemidir. Kübik spline S 0,05 derecesine göre fark dağılımı incelendiğinde ise dağılımın diğer düzgünleştirme yöntemlerine göre keskin olmasına rağmen en keskin dağılım düzgünleştirilmemiş yöntemde elde edilmiştir.

Hem analitik hem de grafiksel çözümlemeye ek olarak ham puanlar için 500 replikasyon ile yapılan düzgünleştirme yöntemlerinin bootstrap standart hataları Equating-Error_wg(v2.0) programı kullanılarak hesaplanmıştır. Ön düzgünleştirme C 6 derecesine göre hata değeri Ek 18’de, beta4 için hata değeri Ek 19’da ve son düzgünleştirme için S 0,05 derecesi için hata

değeri ise Ek 20’de sunulmuştur. Eşitlemenin ortalama standart hataları log lineer C 6 için 0,21389; beta4 için 0,14256; kübik spline S 0,05 için 0,20480 olarak bulunmuştur. Bütüncül olarak inceleniğinde PISA 2012 verilerinde düzgünleştirme yöntemlerinden beta 4 ön düzgünleştirme yöntemini belirtilen diğer düzgünleştirmelere göre daha az standart hataya sahip olduğu görülmüştür. Ön düzgünleştirme yöntemlerinden log-linner C 6 derecesinin ise en fazla eşitleme hatası barındırdığı belirlenmiştir. Bu sonuçlar doğrultusunda öncelikle eşit yüzdelikli eşitlemede puan dağılımlarının önce beta 4 düzgünleştirme yöntemi ile düzgünleştirilip daha sonra eşitleme yapılması önerilmektedir.

Düzgünleştirme yöntemine karar verildikten sonra eşit yüzdelikli eşitleme yöntemleri veri setine uygulanmıştır. İlk eşit yüzdelikli eşitleme yöntemi olarak frekans kestirim yönteminde kullanılmıştır. Frekans kestirim için form X ve form Y koşullu olasılık tablosu Ek 21’de sunulmuştur. Frekans kestirim yönteminde Tucker ve Levine gözlenen doğrusal yöntemler gibi senetik evren için $w_1=0,5$, iki farklı formu alanların kendi arasında kişi sayıları oranlanarak hesaplanan $w_1=1$ ağırlıkları kullanılarak eşit yüzdelikli eşitleme puanları hesaplanmış ve sonuçlar Tablo 31’de sunulmuştur.

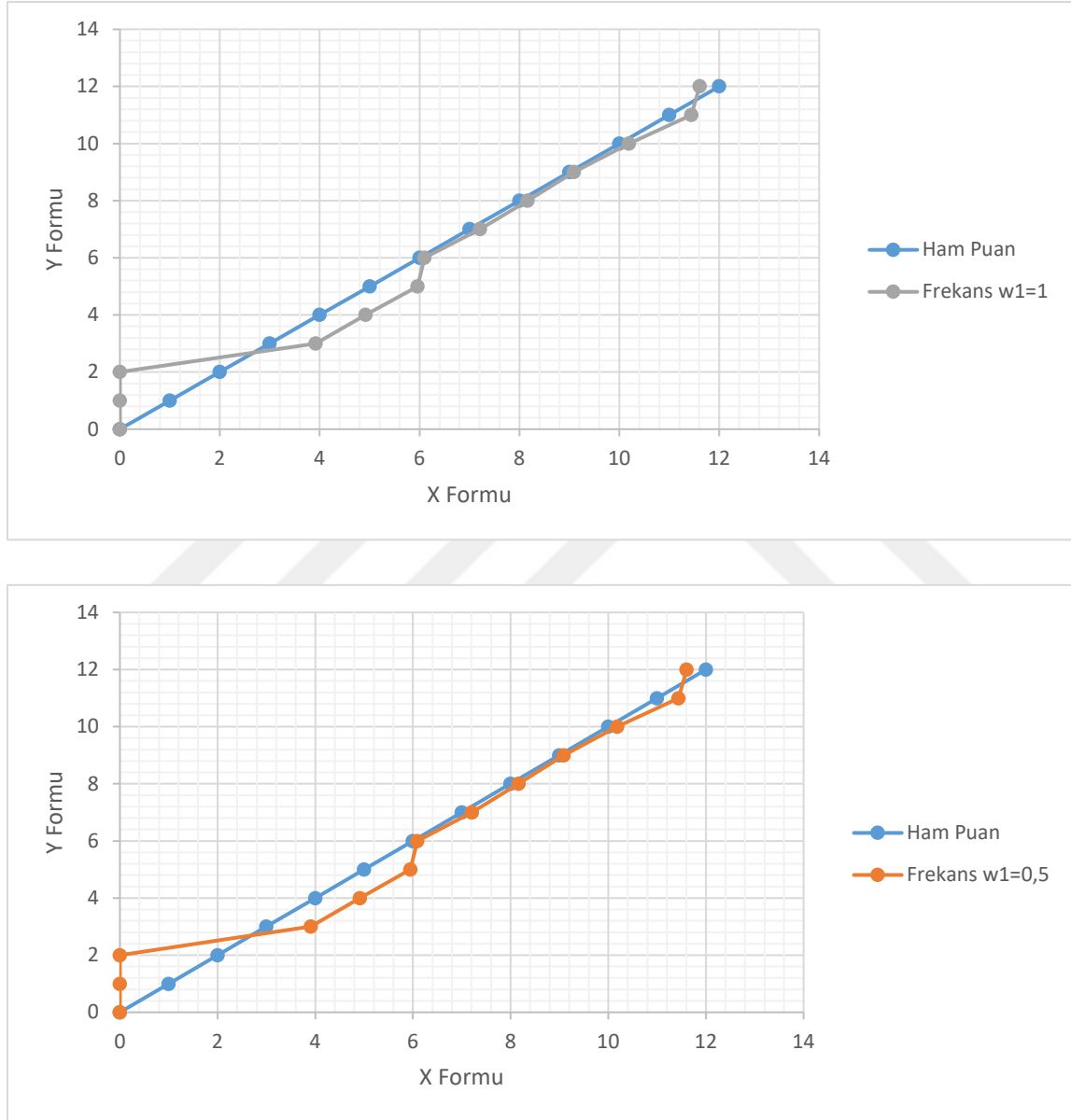
Tablo 31

Sentetik Evrenin Farklı Ağırlık Değerlerine Göre Frekans Kestirim Yöntem Sonuçları

Sentetik Evren Ağırlıkları	Ham Puanlar	Eşitlenmiş Puanlar	Fark	Sentetik Evren Ağırlıkları	Eşitlenmiş Puanlar	Fark
$w_1=0,5;$ $w_2=0,5$	0	0	0	$w_1=1;$ $w_2=0$	0	0
	1	0	0		0	0
	2	0	0		0	0
	3	3,9136	-0,9136		3,9172	-0,9172
	4	4,9173	-0,9173		4,9215	-0,9215
	5	5,9513	-0,9513		5,9593	-0,9593
	6	6,0923	-0,0923		6,0947	-0,0947
	7	7,2091	-0,2091		7,2133	-0,2133
	8	8,1629	-0,1629		8,1648	-0,1648
	9	9,0895	-0,0895		9,0912	-0,0912
	10	10,1874	-0,1874		10,1892	-0,1892
	11	11,4419	-0,4419		11,4434	-0,4434
	12	11,6069	0,3931		11,6065	0,3935

Tablo 31’de yer alan bulgulara frekans kestirim yöntemi kullanılarak eşitlenmiş puanları sunulmuştur. Bu tabloda 0-25 puan aralığı yerine, yöntemin özelliği dikkate alınarak, testte bulunan 12 ortak olmayan madde ve 13 ortak madde yanıtlarının kesişim matrisleri oluşturularak hesaplanması yönüyle Braun-Holland dışındaki diğer tablolardan farklıdır. EYE yöntemlerinden frekans kestirim yöntemi kullanılarak ortak olmayan maddelerin, ortak maddelere verilen doğru-yanlış cevap örüntüsüne bağlı oluşturulan matrisler doğrultusunda eşitlenmiş puanlar sentetik evren ağırlığına göre hesaplanmıştır. $w_1=0,5$ ağırlığında en düşük puan 0 en yüksek puan 11,6069; ağırlık $w_1=1$ olduğunda ise en düşük puan 0 en yüksek puan ise 11,60651 olarak bulunmuştur. Sentetik evren ağırlığı $w_1=0,5$; $w_1=0,1$ olduğunda ham puan 0,1, ve 2 değerine karşılık negatif değerler üretilirken ham puan 3’ten sonraki değerlerde (4,5,6,7,8,9,10 ve 11 için) belirtilen değerlerden daha yüksek değerler frekans

kestirim yöntemi ile elde edilmiştir. Ham puan 12 için sentetik evrenin farklı ağırlıkları kullanıldığında ise 12'den daha düşük değerler elde edilmiştir. Puanların farklı sentetik evren ağırlıklarına göre dağılımları Şekil 27'de sunulmuştur.



Şekil 27 Farklı sentetik evren ağırlıklarına göre frekans kestirim yönteminden elde edilen EYE puanları

Şekil 27’de ham puanlara karşılık farklı sentetik evren ağılıkları kullanılan frekans kestirim yöntemleri ile eşitlenmiş puanların dağılımı verilmiştir. Eşitlenmiş puanlar ve ham puanlar arasında doğrusal bir ilişki olduğu belirtilmiştir. Bu ilişki $w_1=0,5$ için $Y=1,1325X - 0,9731$; $w_1=1$ için $Y=1,1394X - 1,0899$ olarak ifade edilmiştir. Eşitlenmiş puanlar ve ham puanlar arasında pozitif yönlü yüksek bir ilişki ($r_{w1(0,5)}=0,9505$; $r_{w1(1)}=0,9591$) olduğu söylenebilir. Genellikle uç değerlerin eşitlenmiş puanlarında sapma olduğu görülmektedir.

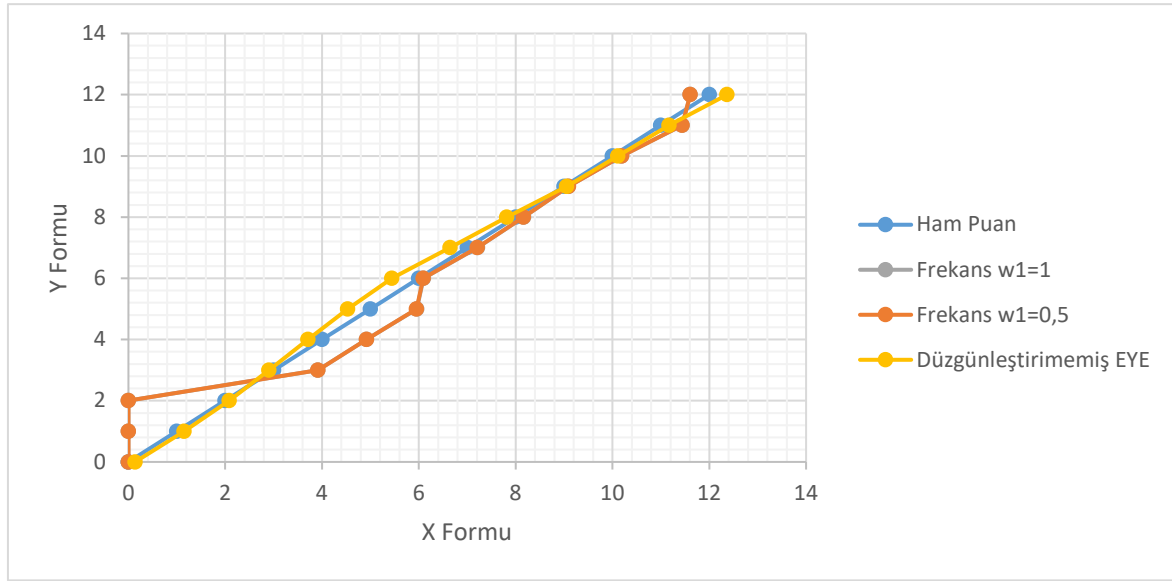
Hesaplanan bu iki EYE ve düzgünleştirme yapılmadan hesaplanan EYE eşitlenmiş puanlara ilişkin ortalama, varyans, slope ve intercept değerleri hesaplamıştır. Hesaplanan bu değerler Tablo 32’de sunulmuştur.

Tablo 32

KTK’deki EYE Kullanılırken Dikkate Alınan Parametre Değerleri

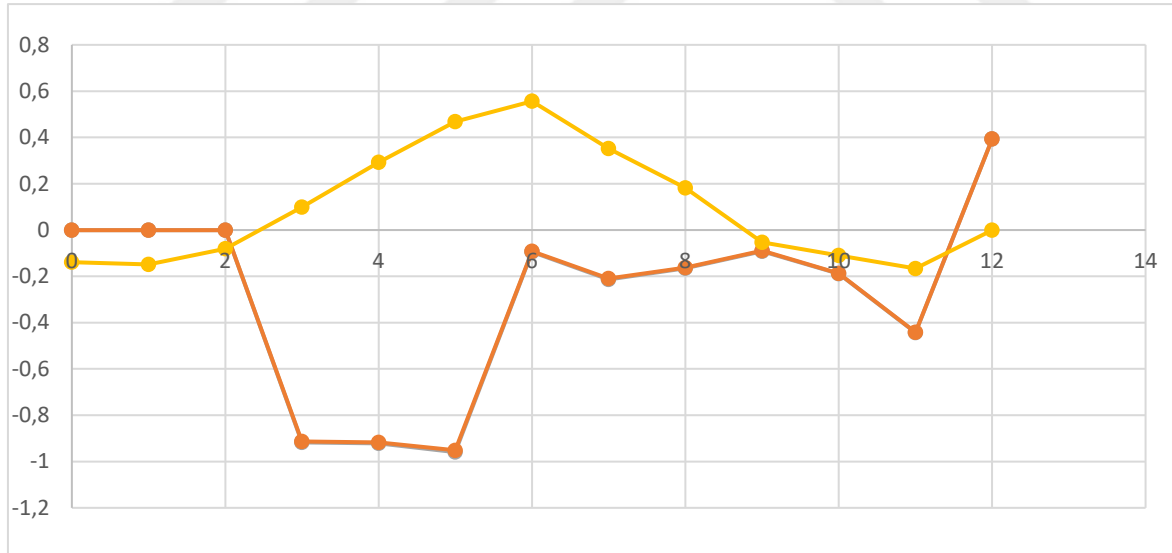
EYE Yöntemi	Sentetik	Slope	Intercept	$\mu(X)$	$\mu(Y)$	$\sigma(X)$	$\sigma(Y)$
	Evren Ağırlık						
Düzenleştirme	Yapılmamış						
EYE		0,9932	-0,3074	4,1561	3,8203	10,887	10,7383
Frekans Kestirim	$w_1=0,5$	1,1499	-1,3664	5,3706	4,8094	8,9733	11,8658
	$w_1=1$	1,1845	-2,1223	5,3498	4,2147	8,8948	12,4801

Eşit yüzdelikli eşitleme yönteminde öncelikle düzgünleştirme yöntemleri denenmiştir. Düzgünleştirme yöntemlerinden beta 4 yönteminin diğer yöntemlerden daha az hata barındırdığı bulunarak eşit yüzdelikli eşitleme yöntemleri denenmiştir. Kullanılan yöntemlerden elde edilen puanlar hesaplanmış ve grafiksel olarak Şekil 28 ve 29’da sunulmuştur. Belirtilen eşit yüzdelikli eşitleme yöntem dağılımları tek bir grafikte aşağıdaki görselde verilmiştir.



Şekil 28 EYE Yönteminden elde edilen puanların dağılımı

Yukarıdaki Şekil 30'daki ham puana göre EYE yöntemlerinin eşitlenmiş puan dağılımları sunulmuştur. EYE yöntemlerinden elde edilen eşitlenmiş puanlar ile ham puanlar arasındaki fark puanların dağılımları ise Şekil 29'da sunulmuştur.



Şekil 29 EYE Yönteminden elde edilen fark puanların dağılımı

Fark puan dağılımları için oluşturulan grafik incelendiğinde, düzgünleştirilmemiş EYE yöntemi dışındaki diğer yöntemlerde uç değerleri için hesaplanan eşitlenmiş puanlar, ham puandan yüksektir. Frekans kestirim yöntemindeki $w_1=1$ sentetik evren ağırlığı için sapma

değerleri fazla olarak görülmektedir. Grafiksel fark puan dağılımlarının yorumlanması dışında analitik süreç ile eşitleme yöntemleri incelenmiştir. Kullanılan EYE yöntemlerinden hangisinin daha uygun olduğunu kestirebilmek için standart hata katsayıları hesaplanmıştır. Bu nedenle Equating-Error_wg (v2.0) programı her bir yöntem için 500 replikasyonla kullanılmıştır. Program çıktıları Ek 22’de verilmiştir. Elde edilen hata değerleri Tablo 33’te sunulmuştur.

Tablo 33

KTK EYE Yöntemlerinin Hata Katsayıları

Eşityüzdelikli	Sentetik Evren	Bootstrap Yöntemi ile	Delta Yöntemi ile
Eşitleme Yöntemleri	Ağırlıkları	Hata Değerleri	Hata Değerleri
Düzgünleştirme			
Yapılmamış EYE	-	0,23299	0,044518
Frekans Kestirim	w ₁ =0,5	0,15891	0,038142
	w ₁ =1	0,19952	0,040117

Tablo 35 incelendiğinde düzgünleştirme yapılmadan EYE yöntemi için hesaplanan hata değeri 0,23298862 olduğu görülmektedir. Program çıktıları incelendiğinde bootstrap ile hesaplanan hatalar, w₁=0,5 ağırlığını kullanan frekans kestirim yöntemi için eşitlemenin standart hatası 0,15891377; w₁=1 ağırlığını kullanan frekans kestirim yöntemi için eşitemenin standart hatası 0,19951602 olarak bulunmuştur. Delta yönteminin Taylor seri fonksiyonuna göre hesaplanan yeni hata değerleri için ise düzgünleştirilmemiş EYE için 0,044518, frekans kestirim w₁=1 ağırlığı için 0,040117 ve w₁=0,5 için 0,038142 olarak bulunmuştur. Katsayılar incelendiğinde ortak maddeli denk olmayan gruplar deseni için uygulanan eşit yüzdelikli eşitleme yöntemlerinden w₁=0,5 ağırlığını kullanan frekans kestirim yönteminin daha az hata ile eşitleme yaptığı görülmüştür.

KTK’da varsayımlar denenerek genel bağlamda doğrusal ve EYE yöntemleri kullanılmıştır. PISA 2012 verilerinin bulunduğu kitapçık 1 ve kitapçık 3’ü alan katılımcıların puanlarına, doğrusal eşitleme yöntemlerinden Tucker içsel, Levine gözlenen puan, Levine gerçek puan, klasik konjenerik ve Braun-Holland eşitleme yöntemleri uygulanmıştır. Uygulanan doğrusal yöntemlerden Tucker ve Levine gözlenen lineer yöntemlerde farklı sentetik evren ağırlıkları ($w_1=0,5$; $w_1=1$) denenerek eşitleme yapılmıştır. EYE’de ise frekans kestirim yöntemleri ($w_1=0,5$; $w_1=1$) denenmiştir. Frekans kestirim yöntemi lineer eşitlemedeki Tucker ve Levine gözlenen eşitleme yöntemleri gini farklı sentetik evren ağırlıkları ($w_1=0,5$; $w_1=1$) bağlamında detaylı olarak incelenmiştir. Equating-Error_wg (v2.0) programı ile tüm KTK eşitleme yöntemlerinin eşitlemenin standart hatası hesaplanmıştır. Tüm yöntemler ve bu yöntemlerin eşitleme hataları aşağıdaki Tablo 34’te sunulmuştur.

Tablo 34

KTK Bünyesindeki Doğrusal ve EYE Yöntemlerinin Hata Katsayıları

	Eşitleme Yöntemleri	Sentetik Evren Ağırlıkları	Bootstrap Yöntemi ile Hata Değerleri	Delta Yöntemi ile Hata Değerleri
Doğrusal	Tucker Eşitleme	$w_1=1$	0,160421	0,140345
		$w_1=0,50$	0,176495	0,153678
	Levine Gözlenen	$w_1=1$	0,170873	0,151479
		$w_1=0,50$	0,168369	0,150236
	Levine Gerçek	-	0,344758	0,34028
	Braun-Holland	-	0,22856	0,168854
	Klasik Konjenerik	-	0,194246	0,163596
Eşit Yüzdeli	Düzgünleştirme			
	Yapılmamış EYE	-	0,23299	0,044518
	Frekans Kestirim	$w_1=0,5$	0,15891	0,038142
		$w_1=1$	0,19952	0,040117

Tablo 34’te yer alan bulgulara KTK’nın eşitleme yöntemleri ve eşitleme yöntemlerine ait eşitleme hatalarına yer verilmiştir. Eşitleme hataları incelendiğinde en az eşitleme hatası EYE yöntemi ile en fazla eşitleme hatasının ise lineer eşileme yöntemi ile elde edildiği görülmektedir. Eşitleme hata değerleri detaylı olarak incelendiğinde, en az hata içeren yöntemden en fazla hata içeren yönteme doğru bir sıralama yapıldığında sıralama şu şekildedir; frekans kestirim yöntemi ($w_1=0,5$), Tucker içsel ($w_1=1$), Levinegözlünen ($w_1=0,5$), Tucker içsel ($w_1=0,5$), klasik konjenerik, frekans kestirim ($w_1=1$), Braun-Holland, düzgünleştirme yapılmadan EYE ve Levine gerçek puan eşitlemedir.

Üçüncü Araştırma Sorusuna Yönelik Bulgular ve Yorumları:

Madde Tepki Kuramı’na dayalı eşitleme yöntemlerinden gerçek ve gözlenen puan eşitleme yöntemleri kullanılarak PISA 2012’ye ait farklı kitapçıklardan elde edilen eşitlenmiş puanlar için hangi eşitleme yöntemi az eşitleme hatası içermektedir?

MTK’nda eşitleme yapmadan önce ortak maddeler kullanılarak ölçek dönüşümü yapılmıştır. Ölçek dönüşüm (kalibrasyon) için momentler ve karakteristik eğri dönüşüm yöntemleri için dönüştürme katsayıları ST 2.0 ile hesaplanmış ve EK 23’te sunulmuştur. Dönüştürme sabitlerine ait bulgular Tablo 35’te verilmiştir.

Tablo 35

Kalibrasyon Yöntemlerinden Elde Edilen Dönüşüm Katsayıları ve Dönüşüm Sabitleri

Kalibrasyon Yöntemleri	A	B
Ortalama-Ortalama	0,993271	-0,038712
Ortalama-Standart Sapma	0,985476	-0,036151
Stocking-Lord	0,956823	0,009559
Haebara	0,95389	0,014415

Hesaplanan dönüşüm katsayıları ve dönüşüm sabitleri için kalibrasyon yöntemleri aşağıdaki eşitliklerde gösterilmiştir.

Ortalama-ortalama için;

$$\Theta_{Xi} = 0,993\Theta_{Yi} - 0,039$$

Ortaama-standart sapma için;

$$\Theta_{Xi} = 0,986\Theta_{Yi} - 0,036$$

Stocking-Lord için;

$$\Theta_{Xi} = 0,957\Theta_{Yi} + 0,010$$

Haebara için;

$\Theta_{Xi} = 0,954\Theta_{Yi} + 0,014$ şeklindedir. Ortalama-ortalama kalibrasyon yöntemi kullanılarak Form X'i cevaplandıran bir kişinin yeteneğinin 0,993 katından 0,039 çıkarılmasıyla hesaplanan yetenek, Form Y 'i cevaplandıran birey ile aynı ölçek değerine konumlandırılacaktır. Ortalama-standart sapma kalibrasyon yöntemi kullanılarak Form X'i cevaplandıran bir kişinin yeteneğinin 0,986 katından 0,036 çıkarılmasıyla hesaplanan yetenek, Form Y 'i cevaplandıran birey ile aynı ölçek değerine konumlandırılacaktır. Stocking-Lord kalibrasyon yöntemi kullanılarak Form X'i cevaplandıran bir kişinin yeteneğinin 0,957 katından 0,010 eklenmesiyle hesaplanan yetenek, Form Y 'i cevaplandıran birey ile aynı ölçek değerine konumlandırılacaktır. Haebara kalibrasyon yöntemi kullanılarak Form X'i cevaplandıran bir kişinin yeteneğinin 0,954 katından 0,014 eklenmesiyle hesaplanan yetenek, Form Y 'i cevaplandıran birey ile aynı ölçek değerine konumlandırılacaktır. Belirtilen kalibrasyon yöntemleri içinde hangi kalibrasyon yönteminin daha az hata ölçek değerlerinin konumlanacağına bakılmıştır. Hesaplanan hata değerleri Tablo 36'da sunulmuştur.

Tablo 36

Kalibrasyon Yöntemleri için Hesaplanan Hata Katsayıları

Kalibrasyon Yöntemleri	Hata Katsayısı
Ortalama-Ortalama	0,038706
Ortalama-Standart Sapma	0,036137
Stocking-Lord	0,035230
Hebara	0,037856

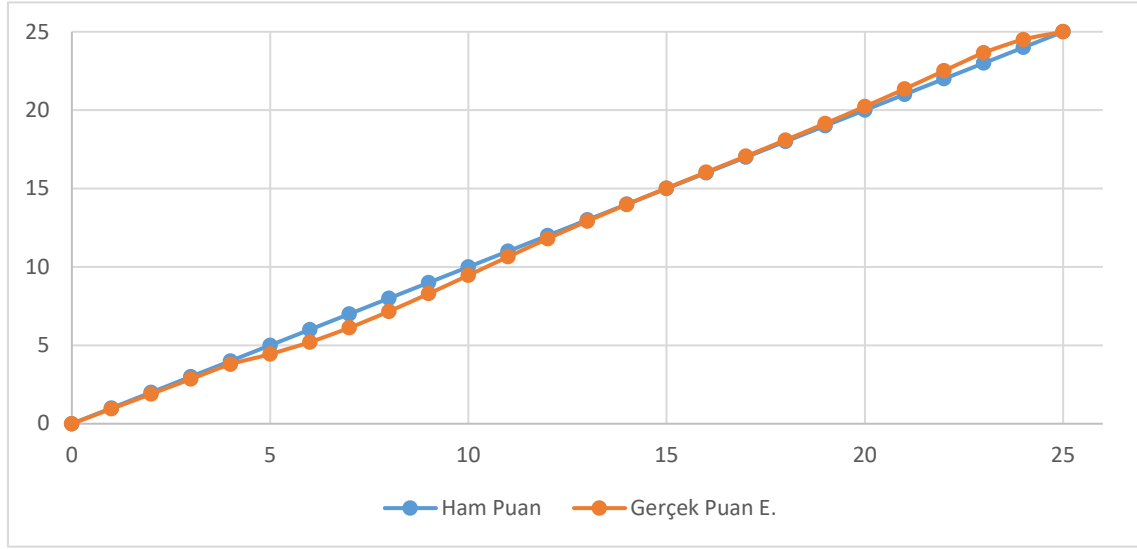
Tablo 36'daki hata değerleri incelendiğinde kalibrasyon yöntemlerinden Stocking-Lord yönteminin en az hata ile aynı ölçek üzerinde yeteneklerin konumlanmasına izin vermektedir. En fazla hata ise ortalama- ortalama kalibrasyon yöntemi ile elde edilmiştir. Kalibrasyon yöntemine karar verildikten sonra Stocking-Lord kalibrasyon yöntemi ile yetenekler aynı ölçeğe getirilmiş ve MTK'da kullanılan eşitleme yöntemlerinden elde edilen puanlar aşağıdaki Tablo 37'de verilmiştir. Hesaplanan eşitlenmiş puanlar ise Ek 26'da sunulmuştur.

Tablo 37

MTK Gerçek ve Gözlenen Puan Eşitleme Yöntemlerinden Elde Edilmiş Puanlar

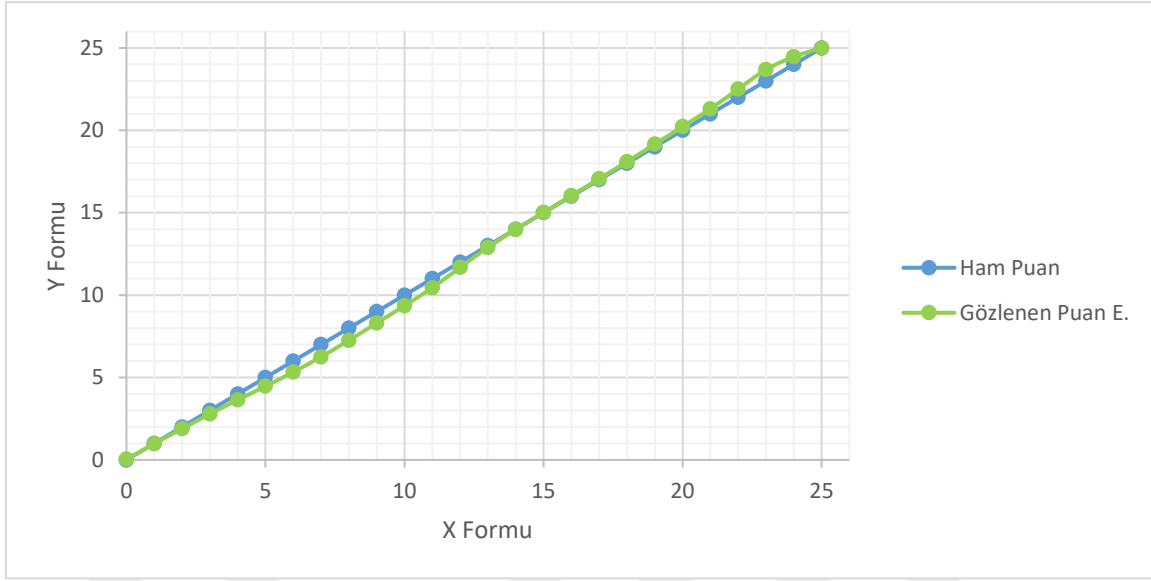
Form X	Gerçek Puan	Gözlenen Puan			
Puanı	Teta Eşdeğeri	Eşitlenmiş Puan	Fark	Eşitlenmiş Puan	Fark
0		0	0	0,0489	-0,0489
1		0,9477	0,0523	0,9824	0,0176
2		1,8955	0,1045	1,8965	0,1035
3		2,8462	0,1538	2,78	0,22
4	-1,3124	3,7859	0,2141	3,6408	0,3592
5	-0,9112	4,4421	0,5579	4,4699	0,5301
6	-0,647	5,205	0,795	5,3319	0,6681
7	-0,439	6,1223	0,8777	6,2561	0,7439
8	-0,2673	7,1676	0,8324	7,2557	0,7443
9	-0,1206	8,2952	0,7048	8,3046	0,6954
10	0,0103	9,4658	0,5342	9,362	0,638
11	0,131	10,645	0,355	10,4504	0,5496
12	0,2444	11,8017	0,1983	11,6826	0,3174
13	0,3516	12,9166	0,0834	12,8915	0,1085
14	0,4535	13,9877	0,0123	13,9777	0,0223
15	0,5513	15,0248	-0,0248	15,0174	-0,0174
16	0,6467	16,0439	-0,0439	16,0452	-0,0452
17	0,7418	17,0623	-0,0623	17,067	-0,067
18	0,8393	18,0946	-0,0946	18,0992	-0,0992
19	0,9431	19,1495	-0,1495	19,1623	-0,1623
20	1,0595	20,2319	-0,2319	20,231	-0,231
21	1,1992	21,3487	-0,3487	21,311	-0,311
22	1,3847	22,5154	-0,5154	22,5026	-0,5026
23	1,6807	23,6584	-0,6584	23,6854	-0,6854
24	2,1527	24,5061	-0,5061	24,4691	-0,4691
25		25	0	25	0

Tablo 37’de sunulan bulgular incelendiğinde MTK’ya dayalı gerçek puan eşitleme için 0,1,2,3 ve 25. Puanlar için yetenek değerleri hesaplanamamıştır. Hesaplanan c parametre değerleri kestirim toplamları 3, 0620 olarak bulunmuştur (EK 24). Bu nedenle 3 puanın altındaki yetenek değerleri kestirilmemiştir. Hem gerçek hem de gözlenen puan eşitleme yönteminden elde edilemn puanların grafiksel gösterimi Şekil 30 ve Şekil 31’de verilmiştir.



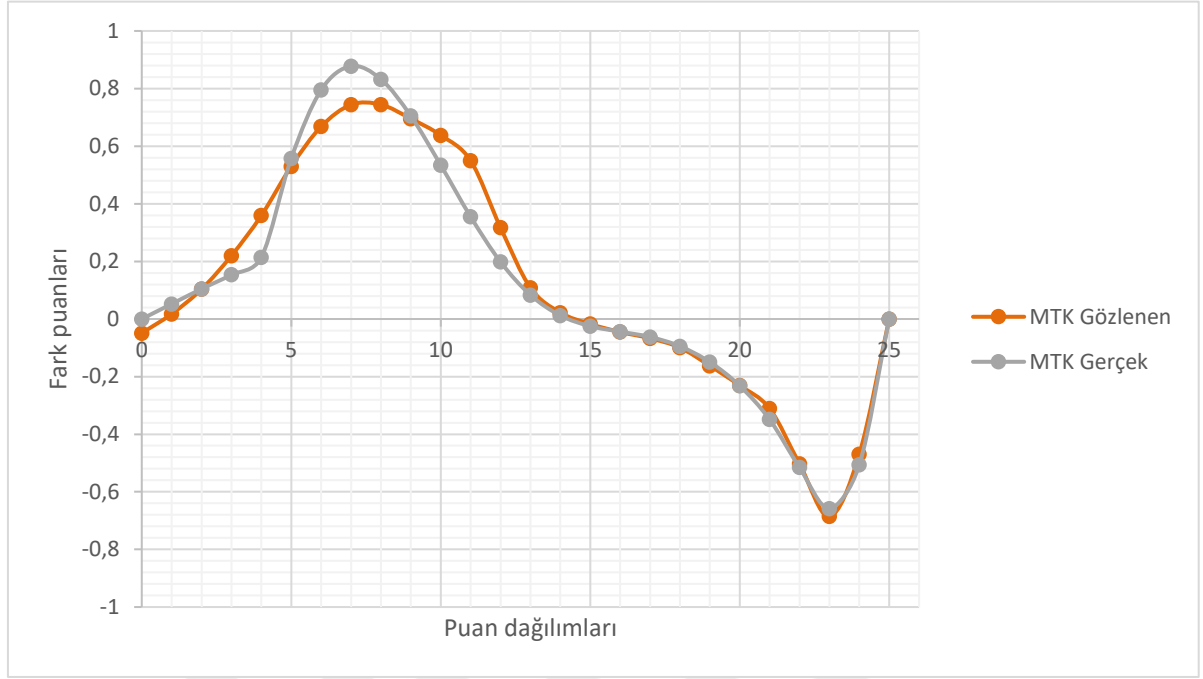
Şekil 30 MTK gerçek puan eşitleme yönteminden elde edilen eşitlenmiş puanlar

Yukarıdaki grafik incelendiğinde gerçek puan eşitleme için ham puan 15'e kadar eşitlenen puanların daha düşük olduğu, 15 dahil 25 puana kadar Y formundan daha yüksek olduğu görülmektedir. 25 ham puanın gerçek puan eşdeğeri de 25 olarak hesaplanmıştır. Sınır değerinin dışına çıkılmamıştır. Eşitlik denklemi $Y=1,0353X-0,5503$ olarak hesaplanmıştır. Gerçek puan ile ham puanlar arasında pozitif ve güçlü bir ilişki ($r=0,9984$) olduğu görülmektedir.



Şekil 31 MTK gözlenen puan eşitleme yönteminden elde edilen eşitlenmiş puanlar

Şekil 31 incelendiğinde gözlenen puan eşitleme için de gerçek puan eşitleme yöntemine benzer yorum yapılabilir. Ham puan 15'e kadar eşitlenen puanların daha düşük olduğu, 15 dahil 25 puana kadar Y formundan daha yüksek olduğu görülmektedir. Gözlenen puan eşitlemede en yüksek 25 ham puan sınır değerinin üzerinde bir değer olan 25,2795 olarak bulunmuştur. Eşitlik denklemi $Y=1,0368X-0,5679$ olarak hesaplanmıştır. Gözlenen puan ve ham puanlar arasında pozitif ve güçlü bir ilişki ($r=0,9983$) olduğu görülmektedir. Yukarıdaki Şekil 32 ve Şekil 33'te eşitlenmiş puanların dağılımları verilmiştir. MTK'daki eşitlenmiş puanlar ve ham puanlar arasındaki fark değerlerine göre dağılımlar aşağıdaki görselde sunulmuştur.



Şekil 32 MTK gözlenen ve gerçek puan eşitleme yöntemlerindeki fark puanlarının dağılımları

Yukarıdaki Şekil 32 incelendiğinde MTK gerçek puan eşitleme yöntemlerinde ham puan 0 ve 25 için eşitlenmiş puanlar bu uç değerlere eşit çıktığı için fark 0 olarak hesaplanmıştır. MTK gerçek puan için ham puan 4'e kadar doğrusal bir şekilde fark puanlarında artış varken, ham puan 4'ten sonra fark puanları hızlı bir şekilde değişim göstermiş ve artmıştır. MTK gerçek puan için ham puan 7'de MTK gerçek puan için fark değeri en yüksek değere ulaşmıştır. Ham puan 14'e kadar fark değer azalma eğilimi göstermiştir. Ham puan 14'ten 24'e kadar eşitlenmiş puanlar ham puanlar daha düşüktür. MTK gözlenen puan eşitleme yöntemlerinde kesme puan atama ile ham puan 0 ve 25 değerlerine eşitlenmiştir. MTK gerçek puan gibi MTK gözlenen puan eşitlemede de ham puan 7'ye kadar fark puanı artış gösterirken ham puan 14'e kadar azalma eğilimindedir. Ham puan 14'ten sonra eşitlenmiş puanlar ham puandan daha düşük olarak hesaplanmıştır.

MTK gerçek puan ve MTK gözlenen puan eşitleme yöntem denklemleri ile eşitlenmiş puanlar hesaplanmış ve grafiksel olarak yorumlanmaya çalışılmıştır. MTK eşitleme yöntemlerinin eşitlemenin standart hatası hesaplanmıştır. Tüm yöntemler ve bu yöntemlerin eşitleme hataları aşağıdaki Tablo 38’de sunulmuştur.

Tablo 38

MTK Eşitleme Yöntemleri ve Bu Yöntemlerin Hata Katsayıları

MTK Eşitleme Yöntemleri	Eşitleme Hata Değerleri
Gerçek Puan Eşitleme	0,0110822
Gözlenen Puan Eşitleme	0,0117915

Yukarıdaki Tablo 38’de MTK kapsamında ele alınan eşitleme yöntemlerinin içerdiği hata katsayıları verilmiştir. En az hata ile en güvenilir eşitleme sonuçlarının MTK gerçek puan eşitleme yöntemi ile elde edildiği görülmüştür. MTK gerçek puan eşitleme yöntemi, MTK gözlenen puan eşitlemine göre daha az hata ile daha uygun bir eşitleme yöntemidir.

Dördüncü Araştırma Sorusuna Yönelik Bulgular ve Yorumları:

Klasik Test Kuramında en az eşitleme hatasına sahip olan eşitleme yöntemi ile Madde Tepki Kuramında en az eşitleme hatasına sahip olan eşitleme yöntemi karşılaştırıldığında hangi kurama ait eşitleme yöntemi en az eşitleme hatası içerir?

Bu alt problem kapsamında KTK ve MTK’daki eşitleme yöntemleri, eşitleme hatalarının niceliklerine göre karşılaştırılmak istenmiştir. Tüm eşitleme yöntemleri ve karşılarında bu yöntemlerin hesaplanan eşitleme hataları Tablo 39’da gruplandırılarak verilmiştir. Kullanılan ölçme kuramları ve eşitleme yöntemlerine göre gruplanan hata değerleri aşağıdaki tabloda sunulmuştur.

Tablo 39

KTK ve MTK'ya İlişkin Eşitleme Yöntemlerinin Hata Katsayıları

Kuramlar	Eşitleme Türü	Eşitleme Yöntemleri	Sentetik Evren Ağırlıkları	Eşitleme Hata Değerleri
KTK	Doğrusal	Tucker Eşitleme	w1=1	0,1604 (Bootstrap)
				0,1404 (Delta)
			w1=0,50	0,1765 (Bootstrap)
				0,1537 (Delta)
		Levine Gözlenen	w1=1	0,1709 (Bootstrap)
				0,1515 (Delta)
			w1=0,50	0,1684 (Bootstrap)
				0,1502 (Delta)
		Levine Gerçek	-	0,3448 (Bootstrap)
				0,3403 (Delta)
		Braun-Holland	-	0,2286 (Bootstrap)
				0,1689 (Delta)
		Klasik Konjenerik	-	0,1943 (Bootstrap)
				0,1636 (Delta)
	EYE	Düzenleştirme Yapılmamış EYE	-	0,2330 (Bootstrap)
				0,0445 (Delta)
		Frekans Kestirim	w1=0,5	0,1589 (Bootstrap)
			w1=1	0,03814 (Delta)
MTK		Gerçek Puan Eşitleme		0,1995 (Bootstrap)
				0,04012 (Delta)
		Gözlenen Puan Eşitleme	-	0,0111
			-	0,0118

Tablo 39 incelendiğinde kullanılan iki ölçme kuramına ait bu araştırma için kullanılan tüm eşitleme yöntemlerinin eşitleme hata değerlerinin bulunduğu görülmektedir. Bu alt problem ifadesinden önceki diğer alt problem ifadelerinde de eşitle için en uygun yöntem belirlenirken yapılan yorumlar doğrultusunda yukarıdaki tablo incelenmiştir. Öncelikle KTK'daki yöntemler incelendiğinde doğrusal eşitleme yöntemlerinden w1=1 sentetik ağırlık kullanılan Tucker eşitleme yönteminin en az hata ile eşitlenmiş puanların elde edildiği, EYE

yöntemlerinden ise $w_1=0,5$ sentetik ağırlık kullanılan frekans kestirim yöntemi ile en az hatalı eşitlenmiş puanlar elde edilmiştir. KTK eşitleme yöntemleri hata değerlerine bakılarak karşılaştırıldığında EYE yöntemlerinden frekans kestirim yönteminin daha az hata ile eşitleme yaptığı görülmüştür. Doğrusal eşitleme yöntemlerinden Levine gerçek puan eşitleme için ise en fazla hata ile eşitlenmiş puanlar elde edilmiştir. MTK dan hesaplan gerçek ve gözlenen eşitleme yöntemleri incelendiğinde MTK gerçek puan eşitlemenin daha az hata ile eşitlenmiş puanlara ulaşıldığı saptanmıştır. Yukarıdaki tabloda yer alan hata katsayıları incelendiğinde MTK'ya ait eşitleme yöntemleri, KTK'daki tüm eşitleme yöntemlerinden daha az hata ile eşitlenmiş puanların elde edildiği görülmektedir. Her iki kuramda en az hata ile eşitleme yapılan bu yöntemler hata nicelikleri bağlamında incelendiğinde ise MTK gerçek puan yönteminde en az hata ile en uygun eşitlenmiş puanların elde edildiği görülmüştür.

BÖLÜM V

TARTIŞMA VE SONUÇ

Araştırmanın bu bölümde, alt problemlere ilişkin bulgu sonuçları özetlenmiş olup, bu sonuçlar ulaşılan literatürdeki araştırma sonuçlarıyla tartışılmıştır.

Bu çalışmada, 2012 PISA uygulamasında yer alan kitapçık 1 ve kitapçık 3 puanları üzerinde Klasik Test Kuramı ve Madde Tepki Kuramında yer alan test eşitleme yöntemleri uygulanarak en az hata ile en uygun eşitleme yöntemine karar verilmesi amaçlanmıştır. Bu amaç doğrultusunda cevaplanan alt problem cümlelerine ilişkin sonuçlar kısaca özetlenmiştir.

Birinci Alt Problemine İlişkin Sonuçlar

(Klasik Test Kuramı'nda yer alan Tucker eşitleme yöntemi, Levine gerçek puan, Levine gözlenen puan, konjenerik ve Braun-Holland doğrusal eşitleme yöntemleri kullanılarak PISA 2012'ye ait farklı kitapçıklardan elde edilen eşitlenmiş puanlar için hangi eşitleme yöntemi en az eşitleme hatası içermektedir?)

Araştırmanın ilk alt problemi için KTK'ya bağlı doğrusal eşitleme yöntemleri uygulanmıştır. Doğrusal eşitleme yöntemleri bağlamında Tucker içsel ($w_1=1$; $w_1=0,5$), Levine gerçek,

Levine gözlenen ($w_1=1$; $w_1=0,5$), klasik konjenerik Braun-Holland eşitleme yöntemleri kullanılarak eşitleme denklemleri elde edilmiştir. Elde edilen denklemler ile eşitlenmiş puanlar hesaplanmıştır. KTK'ya bağlı doğrusal eşitleme yöntemleri için eşitleme hatası değerleri incelenmiştir. Elde edilen hata değerleri niceliksel olarak karşılaştırıldığında en az hata içeren en fazla hata içeren yöntemlere doğru sıralama şu şekildedir; Tucker içsel ($w_1=1$), Levine gözlenen ($w_1=0,5$), Levine gözlenen ($w_1=1$), Tucker içsel ($w_1=0,5$), zincileme, klasik konjenerik, Braun-Holland ve Levine gerçek puan eşitleme. Eşitleme hata değerlerine bakıldığında doğrusal eşitleme yöntemleri için en az hata ile en uygun eşitleme yönteminin Tucker içsel ($w_1=1$) olduğu, en fazla hatanın ise Levine gerçek puan eşitleme yönteminin olduğu görülmüştür.

Doğrusal eşitleme sürecinde Tucker eşitleme yöntemi lehine elde edilen sonuç ulaşılabilen literatür ile kıyaslandığında, von Davier ve Kong (2005), Kane ve diğerleri (2009), Zu ve Yuan (2012), Özdemir (2017) ve İnal ve Arıkan (2017)'in yaptıkları çalışma ile benzer sonuç elde edilmiştir. Bunun aksine Klein ve Jarjoura (1985), Cope (1987), Woodruff (1989), Hanson, Zeng ve Kolen (1993) tarafından yapılan çalışmada buldukları doğrusal eşitleme yöntemi ile bu çalışma sonucu örtüşmemektedir. Lawrence ve Dorans (1988) yaptıkları çalışmada Tucker yönteminin eski ve yeni form için hesaplanan gözlenen puan parametrelerinin birbirlerine yakın olduğu durumlarda Tucker yönteminin daha iyi sonuçlar vereceğini açıklamıştır. Bu bulgu ile araştırma sonucu tutarlıdır.

Doğrusal eşitleme yöntemlerinden Levine gerçek puan eşitleme yöntemi ise en kötü eşitleme yöntemi olarak bulunmuştur. Levine gerçek puan için özellikle uç değerlerdeki eşitlenmiş puanların, ham puana göre farklılaşmasının fazla olduğu ve hesaplanan fark değerlerin diğer doğrusal yöntemlere göre farklı olduğu gözlemlenmiştir. Levine gerçek puan ortalaması ile gözlenen puan ortalamasının benzer olma varsayımından türemesine karşın bu çalışmada elde edilen fark değerlerinden dolayı kestirelen hata değerinin çok fazla olduğu görülmektedir. Teorik olarak gerçek puan, artı eksi hata değeri ile gözlenen puana eklenerek elde edilir (Spearman, 1907). Araştırmada yer alan Levine gerçek puan için eşitlenmiş puan

ile ham puan arasındaki fark değerlerinin büyük olması göze çarpmaktadır. Fark puanların dağılımları incelendiğinde değişimin fazla olduğu hususu araştırmanın bulgular kısmında açıklanmıştır. Fark değerlerinin bu kadar fazla değişkenliği hata dağılımı için yorumlandığında Levine gerçek puanın en fazla hataya sahip olan yöntem olduğu sonucuna ulaşılabilir. Alanyazında benzer sonuçlar, Chen ve Holland (2009) tarafından yapılan çalışma sonuçları ile örtüşmektedir. Levine gerçek puan için fark puanlarının doğrusal bir fonksiyon üretmeyip hata değerini arttırdığı sonucuna ulaşılmıştır. Bu sonucun aksine Hanson, Zeng ve Kolen (1993) yaptıkları çalışmalarında Levine gerçek puanın, Levine gözlenenden daha az hataya sahip olduğunu bulmuşlardır.

İkinci Alt Problemine İlişkin Sonuçlar

(Klasik Test Kuramı'nda yer alan frekans kestirim eşit yüzdelikli eşitleme yöntemleri kullanılarak PISA 2012'ye ait farklı kitapçıklardan elde edilen eşitlenmiş puanlar için hangi eşitleme yöntemi en az eşitleme hatası içermektedir?)

Araştırmanın ikinci alt problemi için KTK'ya bağlı EYE yöntemleri uygulanmıştır. EYE yönteminden önce düzgünleştirme yöntemlerinden hangisinin uygun olduğuna karar verilmiştir. Düzgünleştirme yöntemlerinde ön düzgünleştirme için C 6 polinom fonksiyonu ve beta4 binomial yöntem uygun görülürken, son düzgünleştirme için S 0,05 derecesi seçilmiştir. Hangi düzgünleştirme yönteminin daha az hataya sahip olduğu araştırılmış ve en iyi düzgünleştirme yönteminin beta4 binominal ön düzgünleştirme olduğu, C 6 derecesindeki ön düzgünleştirme yönteminin ise en fazla hatayı içerdiği görülmüştür. EYE için ön düzgünleştirmede kullanılan beta4 binomial fonksiyonun daha az hata barındırdığı bilgisi Livingstone (1993), Kahraman (2012) ve Tan (2015) tarafından yapılan çalışma sonuçları ile örtüşmektedir.

EYE olarak frekans kestirim yöntemi ($w_1=1$; $w_2=0,5$) EYE kullanılarak eşitleme denklemleri bulunmuştur. Elde edilen denklemler ile eşitlenmiş puanlar hesaplanmıştır. Farklı ağırlıklar

için frekans kestirim yönteminde hesaplanmış eşitlenmiş puanlar incelendiğinde 0,1 ve 2 ham puana karşılık hesaplanan eşitlenmiş puanlar kesme puan kullanılarak 0'a eşitlenmiştir. Kullanılan tüm EYE yöntemleri için eşitlenmiş puanlar ve ham puanlar arasında pozitif güçlü ilişki bulunmuştur. KTK'ya bağlı EYE eşitleme yöntemleri için eşitleme hatası değerleri incelenmiştir. Bulunan hata değerleri niceliksel olarak karşılaştırıldığında en az hata içeren en fazla hata içeren yöntemlere doğru sıralama şu şekildedir; frekans kestirim ($w_1=0,5$), frekans kestirim ($w_1=1$) ve düzgünleştirme yapılmamış EYE. KTK'da yer alan EYE yöntemlerinin hata değerlerine bakıldığında en az hata ile en uygun eşitleme yönteminin frekans kestirim ($w_1=0,5$) olduğu, en fazla hatanın düzgünleştirme yapılmamış EYE yöntemi olduğu görülmüştür.

Araştırmanın ikinci alt probleminden elde edilen sonuçlar Livingstone, Dorans ve Wright (1990), Livingstone (1993), Livingston ve Feryok (1987), Skaggs (2005), Hagge, Liu, He, Power, Wang ve Kolen (2011) ve Liu ve Kolen (2011)'in yaptıkları çalışmalarla örtüşmektedir. Belirtilen bu çalışmalarda da frekans kestirim yönteminin diğer EYE yöntemler arsından daha doğru sonuçlar ürettiği belirtilmiştir. Araştırma kapsamında büyük örneklem kullanıldığında frekans kestirim yöntemi ile hesaplanan eşitlenmiş puanların daha doğru sonuçlar verme eğiliminde olduğu saptanmıştır (Livingstone ve Feryok, 1987). EYE yöntemleri kendi arasında kıyaslandığında Holland, von Davier, Sinharay ve Han (2006)'ın yaptıkları çalışma ile bu araştırma sonucu örtüşmediği tespit edilmiştir. Holland ve diğerlerinin (2006) yürüttükleri çalışmalarında zincirleme EYE yönteminin, frekans kestirim yönteminden daha iyi sonuçlar ürettiği bulunmuştur.

Araştırma kapsamında doğrusal ve EYE yöntemleri araştırmanın verilerine uygulamak suretiyle KTK'da amaca uygun ve en az hata içeren eşitleme yöntemi araştırılmak istenmiştir. Doğrusal eşitleme yöntemi olarak Tucker içsel, Levine gerçek puan, Levine gözlenen, klasik konjenerik model ve Braun-Holland eşitleme yöntemleri kullanılarak eşitlenmiş puanlar elde edilmiştir. Klasik Test Kuramında EYE için frekans kestirim ve düzgünleştirme yapılmamış EYE yöntemleri kullanılmış olup eşitlenmiş puanlar

hesaplanmıştır. Yukarıda yer alan birinci ve ikinci alt probleme ilişkin sonuçlarda eşitlenmiş puan dağılımları açıklanmıştır. Sunulan hata değerleri niceliksel olarak karşılaştırıldığında en az hata içeren en fazla hata içeren yöntemlere doğru sıralama şu şekildedir; frekans kestirim ($w_1=0,5$), Tucker içsel ($w_1=1$), Levine gözlenen ($w_1=0,50$), Levine gözlenen ($w_1=1$), Tucker içsel ($w_1=0,5$), klasik konjenerik, frekans kestirim ($w_1=1$), Braun-Holland ve Levine gerçek puan eşitleme. KTK’da yer alan eşitleme hata değerlerine bakıldığında en az hata ile en uygun eşitleme yönteminin frekans kestirim ($w_1=0,5$) olduğu, en fazla hatanın ise Levine gerçek puan eşitleme yöntemi olduğu görülmüştür. Kullanılan KTK eşitleme yöntemleri kıyasalandığında EYE’nin daha az hata ile uygun olduğu sonucuna ulaşılmıştır.

Kolen ve Brennan (1995), von Davier ve Wilson (2008), Mutluer (2013), İnci (2014) ve Pektaş ve Kılınç (2016)’ın yaptıkları araştırmalarda elde ettikleri EYE’nin, doğrusal eşitleme eşitleme yönteminden daha doğru sonuçlar ürettiği ve daha az hataya sahip olduğu sonucu ile bu araştırma sonucu benzerlik göstermektedir. Bu araştırmanın sonuçları ile Wang, Lee, Brennan ve Kolen (2008), Kelecioğlu ve Gübeş (2013)’ın yapmış oldukları çalışma sonuçları örtüşmemektedir. Bu çalışmalarda doğrusal eşitleme yönteminin daha doğru sonuçlar ortaya çıkardığı anlaşılmıştır. Alanyazında Kolen ve Brennan (2004)’ın çalışmasında, büyük örneklemelerde EYE’nin daha doğru sonuçlar ürettiği; formlar arası güçlük farklılıklarının ise, çizilen eğrilerde yüzdelik karşılıklar ile dönüştürülme süreci içermesi nedeniyle daha uyumlu eşitlemeler yaptığı açıkça belirtilmiştir.

Üçüncü Alt Problemine İlişkin Sonuçlar

(Madde Tepki Kuramı’na dayalı eşitleme yöntemlerinden gerçek ve gözlenen puan eşitleme yöntemleri kullanılarak PISA 2012’ye ait farklı kitapçıklardan elde edilen eşitlenmiş puanlar için hangi eşitleme yöntemi az eşitleme hatası içermektedir?)

KTK’da eşitleme yöntemleri kullanıldıktan sonra MTK’da eşitleme yöntemleri denenmiştir. MTK’da eşitleme sürecine başlamadan önce a, b ve c parametreleri kestirilip aynı ölçeğe

dönüşüm yapılmıştır. Kalibrasyon sürecinde dört yöntem kullanılmıştır. Ortalama-ortalama, ortalama-standart sapma, Haebara ve Stocking-Lord yöntemleri ile A ve B katsayıları elde edilmiştir. Aynı ölçeğe dönüştürme sürecinde kullanılan kalibrasyon yöntemlerinin hata katsayıları hesaplanmıştır. Hesaplanan hata katsayılar niceliksel olarak kıyaslandığında en fazla hata ortalama-ortalama yönteminden, en az hatanın ise Stocking-Lord yönteminden elde edilmiştir. Kalibrasyon yöntemlerinden Stocking-Lord'un diğer ölçek dönüşüm yöntemlerinden daha uygun olduğu Stocking ve Lord (1983), Karkee ve Wright (2004), Spearman (2009), Kilmen (2010), Uysal (2014), Demirus (2015), Aksekioglu (2017) ve Yurtcu (2018)'in yaptıkları çalışmalarla da desteklenmiştir. Madde karaktersitik eğrileri baz alınarak yapılan ölçek dönüşüm süreçlerinin daha uygun olduğu belirtilmiştir. Karakteristik eğri yöntemlerinin uyumsuzluğu gideren bir yapıya sahip olduğu belirtilmiştir (Stocking ve Lord, 1983). Yetenek parametresinin farklılıklarında Stocking-Lord daha dayanıklıdır (Keller, 2007). Bunun aksine moment yöntemlerinden özellikle ortalama-ortama kalibrasyon yönteminin de uygun olduğu Gök ve Kelecioğlu (2014), Gündüz (2015) ve Tanberkan Suna (2018) tarafından yapılan çalışmalarda gözlemlenmiştir.

Uygun kalibrasyon yöntemi belirlendikten sonra MTK'ya dayalı gerçek ve gözlenen puan eşitleme yapılmıştır. MTK'da hem gerçek hem de gözlenen 0 ile 14 puana karşılık eşitlenmiş puanlar daha düşük değerlere dönüşürken, 15 ile 24 puana karşılık eşitlenmiş puanlar daha yüksek değerlere dönüşmüştür. 25 puan MTK eşitleme yöntemlerinden gerçek puan eşitlemede 25 puan olarak bulunurken, MTK gözlenen puan eşitlemede ham puandan daha yüksek olarak hesaplanmıştır. MTK eşitleme yöntemlerinden hangisinin daha az hataya sahip olduğu incelenmiştir. Eşitleme hata değerleri niceliksel olarak kıyaslandığında MTK gerçek puan eşitleme yönteminin daha robust bir çözüm ürettiği, daha az hata ile eşitleme yaptığı sonucuna varılmıştır.

Ulaşılabilen literatür tarandığında Lord ve Wingersky (1984), Han, Kolen ve Pohlmann (1997), Hagge, Liu, He, Powers, Wang ve Kolen (2011), Aksekioglu (2017), Tanberkan Suna (2018)'nin yaptıkları çalışmalarda MTK gözlenen puanın daha az eşitleme hatasına

sahip olduğunu bulmuşlardır. Gerçek puan eşitleme sürecinde, gerçek puanın gözlenen ve gerçek puanın birleşiminden oluştuğu kabul edilmektedir. Eşitleme sürecinde aynı yetenek düzeyindeki bireylerin aynı gerçek puana sahip oldukları varsayımına dayanır. Gözlenen puan eşitlemede ise, belli bir gruba odaklanılır. Bu grubun puan dağılımı, özelliklerinin denk olması sağlanarak ortak bir ölçeğe yerleştirilir (von Davier, 2008). Bu açıklamaya dayanarak Gündüz (2015), Kumlu (2019) MTK gerçek puan eşitleme yöntemi ile çalışmasına devam etmiş, MTK gerçek puanın daha az hataya sahip olduğunu çalışma sonucu olarak raporlaştırmıştır. Cho (2007) ve Keller (2007) yaptıkları çalışmalarında gerçek puana ilişkin hesaplamalarda hesaplanan MTK parametrelerinin daha uyumlu sonuçlar verdiğini belirtmiştir. Ayrıca Kolen ve Brennan (1995) MTK gerçek puan eşitleme yönteminin MTK gözlenen puan eşitleme yöntemine göre üstünlüğünü, hesaplama kolaylığı ve elde edilen dönüşümün grubun yetenek dağılımından bağımsız elde edilebilmesi, sınırlılığının ise uygulamada var olmayan gerçek puanları eşitlemesi olarak açıklamıştır.

Dördüncü Alt Problemine İlişkin Sonuçlar

(Klasik Test Kuramı'nda en az eşitleme hatasına sahip olan eşitleme yöntemi ile Madde Tepki Kuramı'ndan en az eşitleme hatasına sahip olan eşitleme yöntemi karşılaştırıldığında hangi kurama ait eşitleme yöntemi en az eşitleme hatası içermektedir?)

Bu araştırmada, KTK ve MTK kapsamındaki eşitleme yöntemlerine yer verilmiştir. KTK doğrusal eşitleme yöntemleri arasında Levine gerçek puan eşitleme yöntemi en kötü eşitleme yöntemi iken sentetik evren ağırlığı 1 olan Tucker eşitleme yöntemi en az hata ile en iyi eşitleme yöntemi olarak belirlenmiştir. KTK, EYE yöntemlerinden düzgünleştirme yapılmamış en fazla eşitleme hata değerine sahip olarak en kötü, sentetik evren ağırlığı 0,5 ile işleme alınan frekans kestirim yönteminin ise en az hata ile en iyi eşitleme yöntemi olduğu belirlenmiştir. Gerçek ve gözlenen eşitleme yöntemleri ele alınarak MTK eşitleme yöntemleri kıyaslandığında ise MTK'da en az hata ile gerçek puan eşitleme yönteminin daha uygun ve güçlü bir eşitleme yöntemi olduğu belirlenmiştir. Niceliksel olarak KTK ve

MTK'daki eşitleme yöntemlerinin hata değerleri incelendiğinde en az hataya sahip olandan en fazla hataya sahip olan eşitleme yöntemine doğru sıralama şu şekildedir; MTK gerçek puan eşitleme, MTK gözlenen puan eşitleme, frekans kestirim ($w_1=0,5$), frekans kestirim ($w_1=1$), Tucker ($w_1=1$), Levine gözlenen ($w_1=1$), Levine gözlenen ($w_1=0,50$), Tucker ($w_1=1$), klasik konjenerik, frekans kestirim ($w_1=1$), Braun-Holland, düzgünleştirme yapılmamış EYE, Levine gerçek puan.

Kuramların karşılaştırılması pek çok araştırmanın amacı olarak ele alınmıştır. Petersen, Cook ve Stocking (1983), Lord ve Wingersky (1984), Kelecioğlu (1994), Han, Kolen ve Pohlman (1997), Hagge, Liu, He, Powers, Wang ve Kolen (2011) Liu ve Kolen (2011) ve ayrıca Tanberkan Suna (2018) yaptıkları çalışmada MTK ve KTK eşitleme yöntemlerini kıyaslamışlardır. Elde ettikleri sonuçlar ile bu araştırmanın sonuçları paralellik gösterdiği MTK eşitleme yöntemlerinin daha az hata ile daha doğru sonuçlar verdiği görülmüştür.

BÖLÜM VI

ÖNERİLER

Araştırmanın bu bölümünde test eşitleme konusu üzerine çalışma yapacaklar için öneriler sıralanmıştır.

Uygulayıcılara Yönelik Öneriler

- Bu araştırma kapsamında veriler, PISA 2012 kitapçık 1 ve kitapçık 3'teki matematik testine yönelik puanlarından elde edilmiştir. PISA 2012 için seçilen bu iki kitapçık üzerinden puanlar eşitlenmeye ve en uygun eşitleme yöntemine karar verilmiştir. Yeni bir çalışma konusu olarak farklı kitapçıktan puanları, bu araştırma sonucunda çıkan MTK gerçek puan eşitleme yöntemi ile tekrar hesaplanabilir.
- PISA her üç senede bir farklı bir alana odaklanmakla birlikte araştırma sürecinde belirtilen tarihteki PISA'nın hedef öğrenme alanı dışındaki öğrenme alanından elde edilen puanlar eşitlenebilir. Bu çalışmada matematik puanları üzerinden bir araştırma yürütülmüştür. Aynı yıl verilerinden fen ve teknoloji veya okuma becerilerine yönelik puanlar da eşitlenerek yeni bir çalışma planlanabilir.
- Geniş ölçekli sınavlarda ülke başarısının belirlenmesi ve diğer ülkeler arasındaki yerin belirlenmesi amaçlandığından pek çok ülkenin katıldığı PISA 2012 için bu çalışmada kitapçık 1 ve kitapçık 3 verileri kullanılarak en başarısız ülke olarak Endonezya, en başarılı ülke olarak Şangay/Çin, ortalamanın altındaki ülkeleri temsilen Türkiye, ortalamanın üstündeki ülkeleri temsilen Finlandiya verileri kullanılmıştır. Bu çalışmada 4 ülke üzerinden araştırmanın amacı doğrultusunda bir çalışma yürütülmüştür. PISA 2012

uygulamasına katılan 65 ülkeyi temsilen yeni bir örnekleme veya evrene ilişkin bütün puanlar ile yeni bir eşitleme çalışmasının yapılması önerilmektedir.

- PISA’da rasch model üzerinden kişilerin puanları hesaplanmaktadır (OECD,2014). Kullanılan eşitleme sürecinde MTK için 3 PLM kullanılmıştır. Yapılan araştırma sonuçlarına göre 3 PLM kullanılarak öğrenci puanlarının tekrar hesaplanması ve test eşitleme süreci ile başarı sırasının yeniden hesaplanması önerilmektedir.
- Çalışma kapsamında, MTK temelli eşitleme sürecinde Stocking-Lord yöntemi ile ölçek puanlarına dönüştürme sürecinin en az hata ile eşitleme yaptığı görülmüştür. Bu nedenle uygulayıcılar için Stockin-Lord ile ölçek dönüşümü yapılarak eşitleme yapılması önerilmektedir.
- Kuramsal olarak eşitleme yöntemlerinin karşılaştırılması amacıyla yürütülen bu araştırma sonuçlardan yola çıkılarak MTK gerçek puan eşitleme yöntemi hatalardan arınık olma durumu ile farklı durumlar için puanların eşitlenebilirliğinde bu yöntemin kullanılması önerilmektedir.

-

Araştırmacılara Yönelik Öneriler

- Amaç elde edilen puanların karşılaştırılabilir olması ve birbiri yerine kullanılması, başarı sıralama sonrası bir üst aşamaya geçilmesi, görevde yükselmenin sağlanması, kurum ya da kuruluşlara atanmanın sağlanması açısından gerek ülke içi yapılan ve geniş kitleleri hedefleyen ALES, KPSS, YDS gibi sınav sonuçları için gerekse de PISA gibi ülkeler arasında mevcut başarı durumunu daha kesin bir şekilde görebilmek için eşitleme çalışmalarına yer verilmelidir.
- Araştırmada PISA 2012 kitapçık 1 ve kitapçık 3 üzerinden dört ülke puanları karşılaştırılmıştır. Bu araştırma için örneklem büyüklüğü küçük değildir. Yeni bir araştırma planlamak isteyenler için farklı örneklem büyüklükleri açısından eşitleme yöntemlerinden uygun olan belirlenebilir.
- Örnekleme ilişkin dört momente bağlı olarak normal dağılım gösterme ya da çarpık dağılım gösterme durumlarına göre KTK ve MTK’ya ait eşitleme yöntemlerinin karşılaştırılması yeni bir çalışmanın konusu olabilir.

- Ortak maddeli eşdeğer grup deseninde ortak maddeler üzerinden eşitleme yapılmaktadır. Bu araştırmada 25 maddenin 13'ü ortak maddeden oluşmaktadır. Ortak maddenin, testin toplam madde sayısına göre farklı oranlarına göre değiştirilip en uygun ortak madde oranı ve bu süreçte kullanılacak eşitleme yöntemine karar verilebilir.
- Çalışmanın verileri 0-1 matrisi üzerinden doğru sayılarına göre hesaplanan puanlardır. Çalışmada çoklu puanlama ile hesaplanan puanlara yönelik eşitleme süreci tasarlanabilir.
- Test eşitleme sürecinde tek boyutluluk varsayımı üzerine kurulu bu çalışmadan esinlenerek çok boyutlu puanlama ile MTK'ya dayalı eşitleme süreci planlanabilir.
- Bu araştırmada çoktan seçmeli test formatı kullanılmıştır. Karma test tekniği ve karma testte farklı oranlarla hazırlanmış test teknikleri ile eşitleme süreci tasarlanabilir.
- Farklı yetenekteki grupların puanları, ortak test ile eşitlemek mümkün olamayacağı için eşdeğer olmayan gruplar kovaryans deseni ile farklı yetenekteki bireylerin puanlarını eşitlemeye yönelik yeni bir çalışma planlanabilir.

KAYNAKLAR

- Angoff, W.H. (1971). Scales, norms and equivalent scores. In R.L. Thorndike (Ed.), *Educational Measurement* (2nd ed., 508-600). Washington, DC: American Council on Education.
- Angoff, W.H. (1987). Technical and practical issues in equating: A discussion of four papers. *Applied Psychological Measurement*, 11, 291-300.
- Atalay Kabasakal, K. (2014). *Değişen madde fonksiyonunun test eşitlemeye etkisi*. (Yayımlanmamış doktora tezi). Ankara: Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü.
- Baker, F. B. (1992). Equating tests under the graded response model. *Applied Psychological Measurement*, 16(1), 87-96.
- Barnard, J. (1996). *In search for equity in educational measurement: Traditional versus modern equating methods*. ASEESA Uluslararası Sempozyumda sunulmuş bildiri, Pretoria, South Africa.
- Baykul, Y. (2015). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Ankara: Pegem Akademi.
- Birnbaum, A. (1968). Some latent trait models and their use in inferring an examinee's ability. *Statistical Theories of Mental Test Scores*, 395-479.
- Bock, R. D. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37(1), 29-51.
- Braun, H. I., & Holland, P. W. (1982). *Observed-score test equating: A mathematical analysis of some ETS equating procedures*. In P. W. Holland D. B. Rubin *Test Equating*, 9-49. New York: Academic.
- Brennan, R. L., & Kolen, M. J. (1987). Some practical issues in equating. *Applied Psychological Measurement*, 11(3), 279-290.

- Brossman, B. G., & Lee, W.-C. (2013). Observed score and true score equating procedures for multidimensional item response theory. *Applied Psychological Measurement*, 37(6), 460-481.
- Büyüköztürk, Ş. (2011). *Sosyal bilimler için veri analizi el kitabı*. Ankara: Pegem A Yayıncılık.
- Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2008). *Bilimsel araştırma yöntemleri*. Ankara: Pegem A Yayıncılık.
- Büyüköztürk, Ş., Çokluk, Ö., & Köklü, N. (2010). *Sosyal bilimler için istatistik*. Ankara: Pegem A Yayıncılık.
- Choi, S. W. (2009). Firestar: Computerized adaptive testing simulation program for polytomous item response theory models. *Applied Psychological Measurement*, 33(8), 644-645.
- Cope, R. T. (1987). How well do the Angoff Design V linear equating methods compare with the Tucker and Levine methods? *Applied Psychological Measurement*, 11(2), 143-149.
- Cook, L., & Eignor, D. (1983). Practical considerations regarding the use of item response theory to equate tests. *Applications of Item Response Theory*, 175-195.
- Cook, L. L., & Eignor, D. R. (1989). Using item response theory in test score equating. *International Journal of Educational Research*, 13(2), 161-173.
- Cook, L. L., & Eignor, D. R. (1991). IRT equating methods. *Educational Measurement: Issues And Practice*, 10(3), 37-45.
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. Florida: Harcourt Brace Javonich College.
- Darroch, J. N., & Ratcliff, D. (1972). Generalized iterative scaling for log-linear models. *The Annals of Mathematical Statistics*, 1470-1480.
- DeMars, C. (2010). *Item response theory*. Oxford: Oxford University Press.
- Demirus, K. B. (2015). *Ortak maddelerin değişen madde fonksiyonu gösterip göstermemesi durumunda test eşitlemeye etkisinin farklı yöntemlerle incelenmesi*. Doktora Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara.
- De Winter, J. C. (2013). Using the Student's t-test with extremely small sample sizes. *Practical Assessment, Research & Evaluation*, 18(10).

- Doğan, N., & Tezbaşaran A. A. (2003). Klasik test kuramı ve örtük özellikler kuramının örneklemeler bağlamında karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 25(25), 58-67.
- Dorans, N. J., & 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., & Eignor, D. R. (2010). Principles and practices of test score equating. *ETS Research Report Series*, 2010(2), i-41.
- Embretson, S. E., & Reise, S. P. (2013). *Item response theory*. New Jersey: Psychology Press.
- Felan, G. D. (2002). Test Equating: Mean, Linear, Equipercentile, and Item Response Theory. *Annual Meeting of the Southwest Educational Research Association*, 1-24.
- Feldt, L., & Brennan, R. L. (1989). *Reliability: Educational measurement* (3rd ed., pp. 105–146). New York, NY: Macmillan.
- Fitzpatrick, J., & Skorupski, W. P. (2016). Equating with midtests using IRT. *Journal of Educational Measurement*, 53(2), 172-189.
- Gök, B. (2012). *Denk olmayan gruplarda ortak madde deseni kullanılarak madde tepki kuramına dayalı eşitleme yöntemlerinin karşılaştırılması*. Doktora Tezi, Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara
- Gündüz, T. (2015). *Test eşitlemede Madde Tepki Kuramına dayalı yetenek parametresine yönelik ölçek dönüştürme yöntemlerinin karşılaştırılması*. Yüksek Lisans Tezi, Gazi Üniversitesi Eğitim Bilimleri Enstitüsü, Ankara
- Haberman, S. J. (1974). Log-linear models for frequency tables with ordered classifications. *Biometrics*, 589-600.
- Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research*, 22(3), 144-149.
- Hagge, S. L., Liu, C., He, Y., Powers, S. J., Wang, W., & Kolen, M. J. (2011). A comparison of IRT and traditional equipercentile methods in mixed-format equating. *Mixed-Format Tests: Psychometric Properties With A Primary Focus On Equating*, 1, 19-50.

- Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues And Practice*, 12(3), 38-47.
- Hambleton, R. K., & Swaminathan, H. (2013). *Item response theory: Principles and applications*. New York: Springer Science & Business Media.
- Hambleton, R. K., Swaminathan, H., ve Rogers, H. J. (1991). *Fundamentals of item response theory*. Sage.
- Han, T., Kolen, M., & Pohlmann, J. (1997). A comparison among IRT true-and observed-score equatings and traditional equipercentile equating. *Applied Measurement in Education*, 10(2), 105-121.
- Hanson, B. A., Zeng, L., & Colton, D. A. (1994). *A comparison of presmoothing and posts smoothing methods in equipercentile equating* (Vol. 94, No. 4). American College Testing Program.
- Hanson, B. A., Zeng, L., & Kolen, M. J. (1993). Standard errors of Levine linear equating. *Applied Psychological Measurement*, 17(3), 225-237.
- Holland, P.W., & Dorans, N. J. (2006). *Linking and equating*. In R. L. Brennan (Ed.), *Educational measurement* (4th ed.). Westport, CT: American Council on Education and Praeger.
- Holland, P. W., von Davier, A. A., Sinharay, S., & Han, N. (2006). Testing the untestable assumptions of the chain and poststratification equating methods for the NEAT design. *ETS Research Report Series*, 2006(1), i-38.
- Hu, L. t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.
- İnal, H., & Arıkan, Ç. A. An Investigation of Group Invariance in Test Equating According to Gender. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 8(1), 128-145.
- Joreskog, K. G. (1969). Statistical analysis of sets of congeneric tests. *ETS Research Bulletin Series*, 1969(2), i-41.
- Karasar, N. (2005). *Bilimsel araştırma yöntemi*. Ankara: Nobel Yayın Dağıtım.
- Kane, M. T., Mroch, A. A., Suh, Y., & Ripkey, D. R. (2009). Linear equating for the NEAT design: Parameter substitution models and chained linear relationship models. *Measurement*, 7(4), 125-146.

- Kendall, M., & Stuart, A. (1977). *The advanced theory of statistics*. New York: Macmillan.
- Kelecioğlu, H., & Öztürk Gübeş, N. (2013). Comparing linear equating and equipercentile equating methods using random groups design. *International. Online Journal of Educational Sciences*, 5(1), 227-241.
- Kilmen, S. (2010). *Madde Tepki Kuramına dayalı test eşitleme yöntemlerinden kestirilen eşitleme hatalarının örneklem büyüklüğü ve yetenek dağılımına göre karşılaştırılması*. Yayınlanmamış doktora tezi, Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.
- Klein, L. W., & Jarjoura, D. (1985). The importance of content representation for common-item equating with nonrandom groups. *Journal of Educational Measurement*, 22(3), 197-206.
- Kline, T. (2005). *Psychological testing: A practical approach to design and evaluation*. New York: Sage.
- Kolen, M. J. (1981). Comparison of traditional and item response theory methods for equating tests. *Journal of Educational Measurement*, 18(1), 1-11.
- Kolen, M. J. (1985). Standard errors of Tucker equating. *Applied Psychological Measurement*, 9(2), 209-223.
- Kolen, M. J. (1988). Traditional equating methodology. *Educational measurement: Issues and practice*, 7(4), 29-37.
- Kolen, M., & Brennan, R. (1995). *Test equating: methods and practices*. New York: Springer-Verlag.
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices*. New York: Springer Science and Business Media.
- Kumlu, G. (2019). *Test ve alt testlerde eşitlemenin farklı koşullar açısından incelenmesi*, Doktora Tezi, Hacettepe Üniversitesi, Ankara.
- Lawrence I. L. , & Dorans, N. J. (1988). *A comparison of observed score and true score equating methods for representative samples and samples matched on an anchor test*. Princeton, NJ: Educational Testing Service.
- Lee, W.C., & Ban, J.C. (2009). A comparison of IRT linking procedures. *Applied Measurement in Education*, 23(1), 23-48.
- Levine, R. (1955). Equating the score scales of alternate forms administered to samples of different ability. *ETS Research Bulletin Series*, 1955(2), i-118.

- Liu, C. (2011). *A comparison of statistics for selecting smoothing parameters for loglinear presmoothing and cubic spline postsMOOTHING under a random groups design*. Doctoral Thesis, Educational Measurement and Statistics in the Graduate College of The University of Iowa, Iowa.
- Liu, C., & Kolen, M. J. (2011). A comparison among IRT equating methods and traditional equating methods for mixed-format tests. *Psychometric properties with a primary focus on equating, 1*, 75-94.
- Livingston, S. A. (1993). Small-sample equating with log-linear smoothing. *Journal of Educational Measurement, 30*(1), 23-39.
- Livingston, S. A. (2004). *Equating test scores (without IRT)*. Educational testing service, ETS.
- Livingston, S. A., Dorans, N. J., & Wright, N. K. (1990). What combination of sampling and equating methods works best?. *Applied Measurement in Education, 3*(1), 73-95.
- Livingston, S. A., & Feryok, N. J. (1987). Univariate vs. bivariate smoothing in frequency estimation equating. *ETS Research Report Series, 1987*(2), 1-21.
- Lord, F. M., & Novick M. R. (1968). *Statistical theories of mental test scores*. Reading, Mass: Addison-Wesley.
- Lord, F. M. (1975). A survey of equating methods based on item characteristic curve theory. *ETS Research Bulletin Series, 1975*(1), i-35.
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Erlbaum.
- Lord, F. M., & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed-score" equatings". *Applied Psychological Measurement, 8*(4), 453-461.
- Loyd, B. H., & Hoover, H. (1980). Vertical equating using the Rasch model. *Journal of Educational Measurement, 17*(3), 179-193.
- Lucke, J. F. (2005). The α and the ω of congeneric test theory: An extension of reliability and internal consistency to heterogeneous tests. *Applied Psychological Measurement, 29*(1), 65-81.
- Marco, G. L. (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement, 14*(2), 139-160.
- McDonald, R. P. (2013). *Test theory: A unified treatment*. New Jersey: Psychology Press.

- MEB (2013). *PISA 2012 Ulusal Ön Raporu*. Ankara: MEB Yenilik Ve Eğitim Teknolojileri Genel Müdürlüğü.
<http://odsgm.meb.gov.tr/test/analizler/docs/pisa/pisa2012-ulusal-on-raporu.pdf> sayfasından erişilmiştir.
- MEB (2015). PISA 2012 Araştırması Ulusal Nihai Raporu. [Çevrim-ici: <https://drive.google.com/file/d/0B2wxMX5xMcnhaGtnV2x6YWsyY2c/view?pref=2&pli=1>, Erişim tarihi: 5 Ocak 2017.]
- Mutluer, C. (2013). Yıl içinde farklı dönemlerde yapılan akademik personel ve lisansüstü eğitimi giriş sınavı (ALES) puanlarına ilişkin bir test eşitleme çalışması. Yüksek Lisans Tezi, *Abant İzzet Baysal Üniversitesi, Eğitim Bilimleri Enstitüsü, Bolu*.
- OECD (2013). PISA 2012 Results: What Makes Schools Successful? Resources, Policies and Practices (Volume IV), PISA, OECD. Publishing [Çevrim-ici: <http://dx.doi.org/10.1787/9789264201156-en>, Erişim tarihi: 8 Mart 2017.]
- OECD (2014). PISA 2012 Technical Report, OECD. Publishing [Çevrim-ici: <https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf>, Erişim tarihi: 10 Mart 2018.]
- Özdemir, B. (2017). Equating TIMSS mathematics subtests with nonlinear equating methods using neat design: circle-arc equating approaches. *International Journal of Progressive Education*, 13(2), 116-132.
- Öztürk, N. (2010). Akademik personel ve lisansüstü eğitimi giriş sınavı puanlarının eşitlenmesi üzerine bir çalışma. *Eğitim ve Bilim*, 37(165), 180-193.
- Pektaş, S., & Kılınç, M. (2016). PISA 2012 matematik testlerinden iki kitapçığın gözlenen puan eşitleme yöntemleri ile eşitlenmesi. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, 1(40). 432-444.
- Petersen, N. S., Cook, L., & Stocking, M. L. (1983). IRT versus conventional equating methods: A comparative study of scale stability. *Journal of Educational Statistics*, 8(2), 137-156.
- Peterson, J. L. (2014). *Multidimensional item response theory observed score equating methods for mixed-format tests*. Doctoral Thesis, Educational Measurement and Statistics in the Graduate College of The University of Iowa, Iowa.
- Petersen, N. S., Kolen, M. J. ve Hoover, H. D. (1989). Scaling, norming, and equating. *Educational Measurement*, 3, 221-262.

- Raykov, T. (1997). Estimation of composite reliability for congeneric measures. *Applied Psychological Measurement*, 21(2), 173-184.
- Ritchie-Scott, A. (1918). The correlation coefficient of a polychoric table. *Biometrika*, 12(1), 93-133.
- Rosenbaum, P. R., & Thayer, D. (1987). Smoothing the joint and marginal distributions of scored two-way contingency tables in test equating. *British Journal of Mathematical and Statistical Psychology*, 40(1), 43-49.
- Schumacker, R. E. ve Lomax, R. G. (1996). *A beginner's guide to structural*. New Jersey: Lawrence Erlbaum Associates.
- Skaggs, G. (2005). Accuracy of random groups equating with very small samples. *Journal of Educational Measurement*, 42(4), 309-330.
- Skaggs, G., & Lissitz, R. W. (1986). IRT test equating: Relevant issues and a review of recent research. *Review of Educational Research*, 56(4), 495-529.
- Spearman, C. (1904). The proof and measurement of association between two things. *American Journal of Psychology*, 15, 72-101.
- Spearman, C. (1907). Demonstration of formulae for true measurement of correlation. *American Journal of Psychology*, 18, 161-169.
- Spearman, C. (1913). Correlations of sums and differences. *British Journal of Psychology*, 5, 417-426.
- Speron, E. (2009). *A comparison of metric linking procedures in item response theory*. Doctoral Thesis, Graduate College of the Illinois Institute of Technology, New York.
- Stocking, M. L., & Lord, F. M. (1982). Developing a common metric in item response theory. *ETS Research Report Series*, 1982(1), i-29.
- Şencan, H. (2005). *Sosyal ve davranışsal ölçümlerde güvenirlik ve geçerlilik*. Ankara: Seçkin Yayınları.
- Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate analysis*. Boston: Allyn and Bacon.
- Tan, Ş. (2015). Küçük örneklemelerde beta4 ve polynomial loglineer öndüzgünleştirme ve kübik eğri öndüzgünleştirme metotlarının uygunluğu. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 35(1), 123-151.
- Thorndike, R. L. (1982). *Applied psychometrics*. Boston: Houghton Mifflin.

- Thurstone, L. L. (1925). A method of scaling psychological and educational tests. *Journal of Educational Psychology*, 16(7), 433.
- 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.
- Uysal, İ. (2019). *Açık uçlu maddelerde otomatik puanlamanın güvenilirliği ve test eşitleme hatalarına etkisi*, Doktora Tezi, Hacettepe Üniversitesi, Ankara.
- Yen, W. M., & Candell, G. L. (1991). Increasing score reliability with item-pattern scoring: An empirical study in five score metrics. *Applied Measurement in Education*, 4, 209–228.
- Yen, W. M. (1984). Effects of local item dependence on the fit and equating performance of the three-parameter logistic model. *Applied Psychological Measurement*, 8(2), 125-145.
- Yıldırım, H. H., Yıldırım, S., Yetişir, M. İ., & Ceylan, E. (2013). *PISA 2012 ulusal ön raporu*. Ankara: MEB.
- Zenisky, A. L., Hambleton, R. K., & Sireci, S. G. (2001). Effects of local item dependence on the validity of IRT item, test and ability statistics. *MCAT Monograph*. 3-30.
- Zhu, W. (1998). Test equating: What, why, how? *Research Quarterly for Exercise and Sport*, 69(1), 11-23.
- Zu, J., & Yuan, K. H. (2012). Standard error of linear observed-score equating for the NEAT design Wwith nonnormally distributed data. *Journal of Educational Measurement*, 49(2), 190-213.
- von Davier, A. A. (2008). New results on the linear equating methods for the nonequivalent-groups design. *Journal of Educational and Behavioral Statistics*, 33(2), 186-203.
- von Davier, A. A., & Kong, N. (2005). A unified approach to linear equating for the nonequivalent groups design. *Journal of Educational and Behavioral Statistics*, 30(3), 313-342.
- Wang, T., Lee, W.C., Brennan, R. L., & Kolen, M. J. (2008). A comparison of the frequency estimation and ed equipercntile methods under the common-item nonequivalent groups design. *Applied Psychological Measurement*, 32(8), 632-651.
- Wingersky, M. S., & Lord, F. M. (1984). An investigation of methods for reducing sampling error in certain IRT procedures. *Applied Psychological Measurement*, 8, 347–364.

- Woldbeck, T. (1998). Basic concepts in modern methods of test equating. *Annual Meeting of the Southwest Psychological Association*, 3-24.
- Woodruff, D. J. (1989). A comparison of three linear equating methods for the common-item nonequivalent-populations design. *Applied Psychological Measurement*, 13(3), 257-262.



EKLER

EK 1. Kitapçıklardaki Madde Parametrelerinin Dağılımları

Tests of Normality						
	Kolmogorov-Smirnov(a)			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
a _x	0,1428	25	0,200	0,9683	25	0,60
b _x	0,0676	25	0,200	0,9834	25	0,94
c _x	0,1886	25	0,052	0,9091	25	0,059
a _y	0,0846	25	0,200	0,9844	25	0,96
b _y	0,1149	25	0,200	0,9677	25	0,59
c _y	0,1807	25	0,055	0,8667	25	0,054
* This is a lower bound of the true significance.						

EK 2. FACTOR Programındaki Kitapçık 1'e İlişkin Özdeğerlerin Çıktısı

EXPLAINED VARIANCE BASED ON EIGENVALUES

Variable	Eigenvalue	Proportion of Variance	Cumulative Proportion of Variance
1	7.58295	0.50332	0.50332
2	0.99793	0.05712	
3	0.92543	0.03982	
4	0.87305	0.03492	
5	0.77702	0.03108	
6	0.77166	0.03087	
7	0.74095	0.02964	
8	0.67765	0.02711	
9	0.60593	0.02424	
10	0.56999	0.02280	
11	0.55163	0.02207	
12	0.51244	0.02050	
13	0.46681	0.01867	
14	0.42905	0.01716	
15	0.40920	0.01637	
16	0.38412	0.01536	
17	0.35545	0.01422	
18	0.32236	0.01289	
19	0.30091	0.01204	
20	0.27928	0.01117	
21	0.25167	0.01007	
22	0.22153	0.00886	
23	0.21628	0.00865	
24	0.15971	0.00639	
25	0.11701	0.00468	

EK 3. FACTOR Programındaki Kitapçık 3'e İlişkin Özdeğerlerin Çıktısı

EXPLAINED VARIANCE BASED ON EIGENVALUES

Variable	Eigenvalue	Proportion of Variance	Cumulative Proportion of Variance
1	7.08148	0.52326	0.52326
2	0.93965	0.05759	
3	0.87618	0.04305	
4	0.86994	0.03480	
5	0.72122	0.02885	
6	0.69481	0.02779	
7	0.62657	0.02506	
8	0.58350	0.02334	
9	0.54643	0.02186	
10	0.53623	0.02145	
11	0.49166	0.01967	
12	0.47364	0.01895	
13	0.45455	0.01818	
14	0.42821	0.01713	
15	0.39819	0.01593	
16	0.38973	0.01559	
17	0.31534	0.01261	
18	0.31035	0.01241	
19	0.30384	0.01215	
20	0.28531	0.01141	
21	0.23258	0.00930	
22	0.22099	0.00884	
23	0.21818	0.00873	
24	0.15982	0.00639	
25	0.14160	0.00566	

EK 4. Kitapçık 1’deki Madde Çiftlerinin Yerel Bağımsızlık Sonuçları

Madde Çifti				Madde Çifti			
Sayısı	Madde Çiftleri		3 PLM için Q3	Sayısı	Madde Çiftleri		3 PLM için Q3
1	Madde 1	Madde 2	-0.00530867	151	Madde 8	Madde 12	0.00211984
2	Madde 1	Madde 3	0.00071295	152	Madde 8	Madde 13	-0.00127735
3	Madde 1	Madde 4	-0.00026126	153	Madde 8	Madde 14	0.005236702
4	Madde 1	Madde 5	-0.0303494	154	Madde 8	Madde 15	-0.05138736
5	Madde 1	Madde 6	-0.0217933	155	Madde 8	Madde 16	-0.02891201
6	Madde 1	Madde 7	-0.01263352	156	Madde 8	Madde 17	-0.0334646
7	Madde 1	Madde 8	0.00080965	157	Madde 8	Madde 18	-0.05083797
8	Madde 1	Madde 9	0.1505653	158	Madde 8	Madde 19	-0.03229496
9	Madde 1	Madde 10	0.01121296	159	Madde 8	Madde 20	-0.01378846
10	Madde 1	Madde 11	0.00357951	160	Madde 8	Madde 21	0.00206215
11	Madde 1	Madde 12	0.00023884	161	Madde 8	Madde 22	0.000275446
12	Madde 1	Madde 13	0.00014977	162	Madde 8	Madde 23	0.004156138
13	Madde 1	Madde 14	0.003159513	163	Madde 8	Madde 24	-0.05033704
14	Madde 1	Madde 15	-0.02545665	164	Madde 8	Madde 25	-0.03343938
15	Madde 1	Madde 16	-0.1293175	165	Madde 9	Madde 10	0.0106992
16	Madde 1	Madde 17	-0.1440527	166	Madde 9	Madde 11	0.003096582
17	Madde 1	Madde 18	-0.02641569	167	Madde 9	Madde 12	0.001278228
18	Madde 1	Madde 19	-0.01557734	168	Madde 9	Madde 13	-0.00077614
19	Madde 1	Madde 20	-0.00530551	169	Madde 9	Madde 14	0.002697397
20	Madde 1	Madde 21	0.003280835	170	Madde 9	Madde 15	-0.04474838
21	Madde 1	Madde 22	0.001678386	171	Madde 9	Madde 16	-0.0254573
22	Madde 1	Madde 23	0.002947866	172	Madde 9	Madde 17	-0.02861658
23	Madde 1	Madde 24	-0.02474789	173	Madde 9	Madde 18	-0.04467819
24	Madde 1	Madde 25	-0.00165124	174	Madde 9	Madde 19	-0.02753784
25	Madde 2	Madde 3	-0.00259990	175	Madde 9	Madde 20	-0.01172739
26	Madde 2	Madde 4	-0.0034018	176	Madde 9	Madde 21	0.000734532
27	Madde 2	Madde 5	0.1693865	177	Madde 9	Madde 22	0.000931064
28	Madde 2	Madde 6	0.1668515	178	Madde 9	Madde 23	0.004299169
29	Madde 2	Madde 7	0.01486753	179	Madde 9	Madde 24	-0.04325892
30	Madde 2	Madde 8	-0.01074596	180	Madde 9	Madde 25	-0.02955545

31	Madde 2	Madde 9	-0.00960481	181	Madde 10	Madde 11	0.003264642
32	Madde 2	Madde 10	-0.00942217	182	Madde 10	Madde 12	0.000990632
33	Madde 2	Madde 11	-0.00650812	183	Madde 10	Madde 13	-0.00011803
34	Madde 2	Madde 12	-0.00590490	184	Madde 10	Madde 14	0.00221399
35	Madde 2	Madde 13	-0.00239792	185	Madde 10	Madde 15	-0.04451635
36	Madde 2	Madde 14	-0.00763591	186	Madde 10	Madde 16	-0.02561465
37	Madde 2	Madde 15	0.01699612	187	Madde 10	Madde 17	-0.02877364
38	Madde 2	Madde 16	0.01255973	188	Madde 10	Madde 18	-0.0442801
39	Madde 2	Madde 17	0.1467479	189	Madde 10	Madde 19	-0.02762332
40	Madde 2	Madde 18	0.1517522	190	Madde 10	Madde 20	-0.01190927
41	Madde 2	Madde 19	0.01439444	191	Madde 10	Madde 21	0.001106117
42	Madde 2	Madde 20	0.101554	192	Madde 10	Madde 22	0.000577067
43	Madde 2	Madde 21	-0.00693613	193	Madde 10	Madde 23	0.003783652
44	Madde 2	Madde 22	-0.00043934	194	Madde 10	Madde 24	-0.04380599
45	Madde 2	Madde 23	-0.00570785	195	Madde 10	Madde 25	-0.02940389
46	Madde 2	Madde 24	0.0163045	196	Madde 11	Madde 12	0.00298238
47	Madde 2	Madde 25	0.01345701	197	Madde 11	Madde 13	0.001917517
48	Madde 3	Madde 4	0.003669944	198	Madde 11	Madde 14	0.003055402
49	Madde 3	Madde 5	-0.019137	199	Madde 11	Madde 15	-0.03341348
50	Madde 3	Madde 6	-0.01288363	200	Madde 11	Madde 16	-0.01865691
51	Madde 3	Madde 7	-0.00641424	201	Madde 11	Madde 17	-0.020238
52	Madde 3	Madde 8	-0.00141002	202	Madde 11	Madde 18	-0.03363552
53	Madde 3	Madde 9	-0.00106438	203	Madde 11	Madde 19	-0.02036865
54	Madde 3	Madde 10	-0.00124136	204	Madde 11	Madde 20	-0.00802144
55	Madde 3	Madde 11	0.000766046	205	Madde 11	Madde 21	0.001666687
56	Madde 3	Madde 12	0.002523797	206	Madde 11	Madde 22	0.001571613
57	Madde 3	Madde 13	0.1750163	207	Madde 11	Madde 23	0.00498829
58	Madde 3	Madde 14	0.01290934	208	Madde 11	Madde 24	-0.03248586
59	Madde 3	Madde 15	-0.01562895	209	Madde 11	Madde 25	-0.0217505
60	Madde 3	Madde 16	-0.00552027	210	Madde 12	Madde 13	0.002229623
61	Madde 3	Madde 17	-0.00714105	211	Madde 12	Madde 14	0.1009463
62	Madde 3	Madde 18	-0.01679831	212	Madde 12	Madde 15	-0.02894246
63	Madde 3	Madde 19	-0.00932353	213	Madde 12	Madde 16	-0.01397199

64	Madde 3	Madde 20	-0.00200462	214	Madde 12	Madde 17	-0.01663733
65	Madde 3	Madde 21	0.00244649	215	Madde 12	Madde 18	-0.02991202
66	Madde 3	Madde 22	0.001067896	216	Madde 12	Madde 19	-0.01747367
67	Madde 3	Madde 23	-0.00237455	217	Madde 12	Madde 20	-0.06621877
68	Madde 3	Madde 24	-0.01522423	218	Madde 12	Madde 21	0.02088156
69	Madde 3	Madde 25	-0.00948866	219	Madde 12	Madde 22	0.1861445
70	Madde 4	Madde 5	-0.02328908	220	Madde 12	Madde 23	-0.1453884
71	Madde 4	Madde 6	-0.01699849	221	Madde 12	Madde 24	-0.02816252
72	Madde 4	Madde 7	-0.0084486	222	Madde 12	Madde 25	-0.1803151
73	Madde 4	Madde 8	-0.00041959	223	Madde 13	Madde 14	0.1856099
74	Madde 4	Madde 9	-0.00049367	224	Madde 13	Madde 15	-0.1401571
75	Madde 4	Madde 10	-0.00106905	225	Madde 13	Madde 16	-0.00636239
76	Madde 4	Madde 11	-0.00123072	226	Madde 13	Madde 17	-0.00746698
77	Madde 4	Madde 12	0.002247043	227	Madde 13	Madde 18	-0.1465767
78	Madde 4	Madde 13	0.001344016	228	Madde 13	Madde 19	-0.00854754
79	Madde 4	Madde 14	0.001190267	229	Madde 13	Madde 20	-0.02414121
80	Madde 4	Madde 15	-0.01938175	230	Madde 13	Madde 21	0.02130643
81	Madde 4	Madde 16	-0.00745602	231	Madde 13	Madde 22	0.1361862
82	Madde 4	Madde 17	-0.00887895	232	Madde 13	Madde 23	-0.04124847
83	Madde 4	Madde 18	-0.02038396	233	Madde 13	Madde 24	-0.01353742
84	Madde 4	Madde 19	-0.01180031	234	Madde 13	Madde 25	-0.008374719
85	Madde 4	Madde 20	-0.00355441	235	Madde 14	Madde 15	-0.03723895
86	Madde 4	Madde 21	0.00353378	236	Madde 14	Madde 16	-0.01998189
87	Madde 4	Madde 22	0.00448969	237	Madde 14	Madde 17	-0.02237144
88	Madde 4	Madde 23	-0.01916164	238	Madde 14	Madde 18	-0.03763809
89	Madde 4	Madde 24	-0.1931442	239	Madde 14	Madde 19	-0.02274014
90	Madde 4	Madde 25	-0.1198639	240	Madde 14	Madde 20	-0.00896016
91	Madde 5	Madde 6	0.09298697	241	Madde 14	Madde 21	0.003139582
92	Madde 5	Madde 7	0.05184469	242	Madde 14	Madde 22	0.001389966
93	Madde 5	Madde 8	-0.06020494	243	Madde 14	Madde 23	0.004517569
94	Madde 5	Madde 9	-0.05196464	244	Madde 14	Madde 24	-0.03615032
95	Madde 5	Madde 10	-0.05171003	245	Madde 14	Madde 25	-0.02369908
96	Madde 5	Madde 11	-0.03874947	246	Madde 15	Madde 16	0.05355319

97	Madde 5	Madde 12	-0.03528597	247	Madde 15	Madde 17	0.05955428
98	Madde 5	Madde 13	-0.01671408	248	Madde 15	Madde 18	0.0929539
99	Madde 5	Madde 14	-0.04317581	249	Madde 15	Madde 19	0.05961585
100	Madde 5	Madde 15	0.1067424	250	Madde 15	Madde 20	0.02273439
101	Madde 5	Madde 16	0.05089509	251	Madde 15	Madde 21	-0.03259444
102	Madde 5	Madde 17	0.06054481	252	Madde 15	Madde 22	-0.00402480
103	Madde 5	Madde 18	0.1118974	253	Madde 15	Madde 23	-0.02784052
104	Madde 5	Madde 19	0.06471856	254	Madde 15	Madde 24	0.09121744
105	Madde 5	Madde 20	0.02206005	255	Madde 15	Madde 25	0.05782029
106	Madde 5	Madde 21	-0.04009541	256	Madde 16	Madde 17	0.04819403
107	Madde 5	Madde 22	-0.00478511	257	Madde 16	Madde 18	0.04551848
108	Madde 5	Madde 23	-0.03206787	258	Madde 16	Madde 19	0.03553585
109	Madde 5	Madde 24	0.1048312	259	Madde 16	Madde 20	0.01976528
110	Madde 5	Madde 25	0.06447698	260	Madde 16	Madde 21	-0.01729386
111	Madde 6	Madde 7	0.05590755	261	Madde 16	Madde 22	-0.00196129
112	Madde 6	Madde 8	-0.04590752	262	Madde 16	Madde 23	-0.01528561
113	Madde 6	Madde 9	-0.03915116	263	Madde 16	Madde 24	0.04797799
114	Madde 6	Madde 10	-0.03924427	264	Madde 16	Madde 25	0.03449871
115	Madde 6	Madde 11	-0.02796762	265	Madde 17	Madde 18	0.05312672
116	Madde 6	Madde 12	-0.02518851	266	Madde 17	Madde 19	0.04144451
117	Madde 6	Madde 13	-0.1146321	267	Madde 17	Madde 20	0.0222735
118	Madde 6	Madde 14	-0.03180841	268	Madde 17	Madde 21	-0.01931927
119	Madde 6	Madde 15	0.0773793	269	Madde 17	Madde 22	-0.00232195
120	Madde 6	Madde 16	0.04323297	270	Madde 17	Madde 23	-0.1711922
121	Madde 6	Madde 17	0.05018445	271	Madde 17	Madde 24	0.05620852
122	Madde 6	Madde 18	0.07753623	272	Madde 17	Madde 25	0.03606492
123	Madde 6	Madde 19	0.05233162	273	Madde 18	Madde 19	0.05956545
124	Madde 6	Madde 20	0.02289039	274	Madde 18	Madde 20	0.02037517
125	Madde 6	Madde 21	-0.02926916	275	Madde 18	Madde 21	-0.03486717
126	Madde 6	Madde 22	-0.00342550	276	Madde 18	Madde 22	-0.04427933
127	Madde 6	Madde 23	-0.02325541	277	Madde 18	Madde 23	-0.02894233
128	Madde 6	Madde 24	0.07725838	278	Madde 18	Madde 24	0.09239288
129	Madde 6	Madde 25	0.04907323	279	Madde 18	Madde 25	0.05496619

130	Madde 7	Madde 8	-0.02921687	280	Madde 19	Madde 20	0.01936212
131	Madde 7	Madde 9	-0.02557035	281	Madde 19	Madde 21	-0.02068425
132	Madde 7	Madde 10	-0.02563573	282	Madde 19	Madde 22	-0.00286056
133	Madde 7	Madde 11	-0.01758148	283	Madde 19	Madde 23	-0.01661033
134	Madde 7	Madde 12	-0.01521104	284	Madde 19	Madde 24	0.05833351
135	Madde 7	Madde 13	-0.00614440	285	Madde 19	Madde 25	0.03747027
136	Madde 7	Madde 14	-0.02000925	286	Madde 20	Madde 21	-0.00811627
137	Madde 7	Madde 15	0.04958033	287	Madde 20	Madde 22	-0.00086689
138	Madde 7	Madde 16	0.03884772	288	Madde 20	Madde 23	-0.00676020
139	Madde 7	Madde 17	0.04133268	289	Madde 20	Madde 24	0.002268183
140	Madde 7	Madde 18	0.04593799	290	Madde 20	Madde 25	0.1844233
141	Madde 7	Madde 19	0.03847178	291	Madde 21	Madde 22	0.001885264
142	Madde 7	Madde 20	0.02237504	292	Madde 21	Madde 23	0.005169332
143	Madde 7	Madde 21	-0.01835264	293	Madde 21	Madde 24	-0.00328391
144	Madde 7	Madde 22	-0.00187498	294	Madde 21	Madde 25	-0.02194416
145	Madde 7	Madde 23	-0.01458958	295	Madde 22	Madde 23	0.000673269
146	Madde 7	Madde 24	0.04845413	296	Madde 22	Madde 24	-0.00417701
147	Madde 7	Madde 25	0.03397569	297	Madde 22	Madde 25	-0.00257828
148	Madde 8	Madde 9	0.01492941	298	Madde 23	Madde 24	-0.02703833
149	Madde 8	Madde 10	0.01724715	299	Madde 23	Madde 25	-0.01759331
150	Madde 8	Madde 11	0.003651173	300	Madde 24	Madde 25	0.05323576

EK 5. Kitapçık 3’ teki Madde Çiftlerinin Yerel Bağımsızlık Sonuçları

Madde Çifti				Madde Çifti 3 PLM için			
Sayısı	Madde Çiftleri		3 PLM için Q3	Sayısı	Madde Çiftleri		Q3
1	Madde 1	Madde 2	-0.02372264	151	Madde 8	Madde 12	0.042913
2	Madde 1	Madde 3	-0.0179819	152	Madde 8	Madde 13	0.03130547
3	Madde 1	Madde 4	-0.02005817	153	Madde 8	Madde 14	-0.00625892
4	Madde 1	Madde 5	-0.0274252	154	Madde 8	Madde 15	-0.00969576
5	Madde 1	Madde 6	-0.0060646	155	Madde 8	Madde 16	-0.00298666
6	Madde 1	Madde 7	-0.02771089	156	Madde 8	Madde 17	-0.0265742
7	Madde 1	Madde 8	-0.01308051	157	Madde 8	Madde 18	-0.02653372
8	Madde 1	Madde 9	-0.02872788	158	Madde 8	Madde 19	-0.1818208
9	Madde 1	Madde 10	-0.03606636	159	Madde 8	Madde 20	-0.01514755
10	Madde 1	Madde 11	-0.01495065	160	Madde 8	Madde 21	-0.00584377
11	Madde 1	Madde 12	-0.02428107	161	Madde 8	Madde 22	-0.02020968
12	Madde 1	Madde 13	-0.01502193	162	Madde 8	Madde 23	-0.01765167
13	Madde 1	Madde 14	0.000897737	163	Madde 8	Madde 24	-0.00181547
14	Madde 1	Madde 15	0.000533283	164	Madde 8	Madde 25	-0.01373274
15	Madde 1	Madde 16	0.01308459	165	Madde 9	Madde 10	0.112668
16	Madde 1	Madde 17	0.02005528	166	Madde 9	Madde 11	0.05282485
17	Madde 1	Madde 18	0.1760922	167	Madde 9	Madde 12	0.07867938
18	Madde 1	Madde 19	0.0428904	168	Madde 9	Madde 13	0.04808567
19	Madde 1	Madde 20	0.000999882	169	Madde 9	Madde 14	-0.01547474
20	Madde 1	Madde 21	0.000241045	170	Madde 9	Madde 15	-0.02195187
21	Madde 1	Madde 22	0.00390609	171	Madde 9	Madde 16	-0.0562723
22	Madde 1	Madde 23	0.00412893	172	Madde 9	Madde 17	-0.0499075
23	Madde 1	Madde 24	0.001678561	173	Madde 9	Madde 18	-0.05134458
24	Madde 1	Madde 25	0.003385633	174	Madde 9	Madde 19	-0.00366059
25	Madde 2	Madde 3	0.04891441	175	Madde 9	Madde 20	-0.0325902
26	Madde 2	Madde 4	0.05389994	176	Madde 9	Madde 21	-0.01342039
27	Madde 2	Madde 5	0.07429252	177	Madde 9	Madde 22	-0.04051137
28	Madde 2	Madde 6	0.01986391	178	Madde 9	Madde 23	-0.03751235
29	Madde 2	Madde 7	0.0736779	179	Madde 9	Madde 24	-0.00413183
30	Madde 2	Madde 8	0.04298177	180	Madde 9	Madde 25	-0.02881833

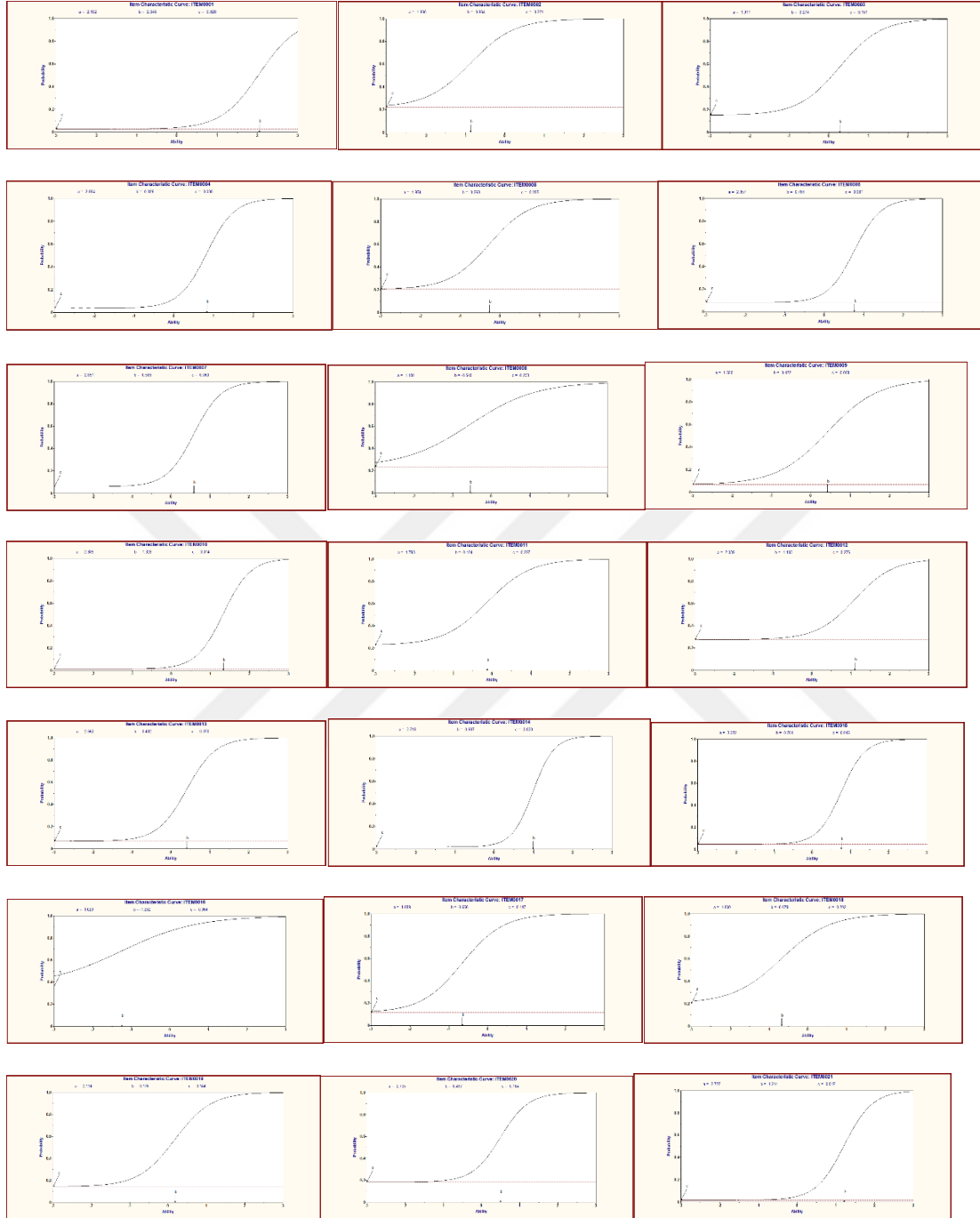
31	Madde 2	Madde 9	0.0759803	181	Madde 10	Madde 11	0.05262726
32	Madde 2	Madde 10	0.09180664	182	Madde 10	Madde 12	0.09164952
33	Madde 2	Madde 11	0.04511516	183	Madde 10	Madde 13	0.0515519
34	Madde 2	Madde 12	0.0649936	184	Madde 10	Madde 14	-0.02043032
35	Madde 2	Madde 13	0.04094088	185	Madde 10	Madde 15	-0.02762243
36	Madde 2	Madde 14	-0.01293608	186	Madde 10	Madde 16	-0.651422
37	Madde 2	Madde 15	-0.179861	187	Madde 10	Madde 17	-0.06075958
38	Madde 2	Madde 16	-0.04704553	188	Madde 10	Madde 18	-0.06124799
39	Madde 2	Madde 17	-0.0410523	189	Madde 10	Madde 19	-0.04521891
40	Madde 2	Madde 18	-0.04261208	190	Madde 10	Madde 20	-0.04065281
41	Madde 2	Madde 19	-0.0311583	191	Madde 10	Madde 21	-0.01744125
42	Madde 2	Madde 20	-0.02758391	192	Madde 10	Madde 22	-0.4954803
43	Madde 2	Madde 21	-0.01078787	193	Madde 10	Madde 23	-0.04627408
44	Madde 2	Madde 22	-0.0331285	194	Madde 10	Madde 24	-0.05209899
45	Madde 2	Madde 23	-0.03071654	195	Madde 10	Madde 25	-0.03555178
46	Madde 2	Madde 24	-0.03277881	196	Madde 11	Madde 12	0.05233794
47	Madde 2	Madde 25	-0.02406439	197	Madde 11	Madde 13	0.0363052
48	Madde 3	Madde 4	-0.04604429	198	Madde 11	Madde 14	-0.00736927
49	Madde 3	Madde 5	-0.06392749	199	Madde 11	Madde 15	-0.01105902
50	Madde 3	Madde 6	0.1874481	200	Madde 11	Madde 16	-0.03307711
51	Madde 3	Madde 7	0.06015544	201	Madde 11	Madde 17	-0.02859032
52	Madde 3	Madde 8	0.03602396	202	Madde 11	Madde 18	-0.02899164
53	Madde 3	Madde 9	0.05829383	203	Madde 11	Madde 19	-0.02007772
54	Madde 3	Madde 10	0.06354308	204	Madde 11	Madde 20	-0.01786591
55	Madde 3	Madde 11	0.03944766	205	Madde 11	Madde 21	-0.00670563
56	Madde 3	Madde 12	0.05251984	206	Madde 11	Madde 22	-0.02296231
57	Madde 3	Madde 13	0.03476248	207	Madde 11	Madde 23	-0.01988995
58	Madde 3	Madde 14	-0.01081335	208	Madde 11	Madde 24	-0.0211448
59	Madde 3	Madde 15	-0.0147573	209	Madde 11	Madde 25	-0.01527225
60	Madde 3	Madde 16	-0.03694392	210	Madde 12	Madde 13	0.04858912
61	Madde 3	Madde 17	-0.0324303	211	Madde 12	Madde 14	-0.01333394
62	Madde 3	Madde 18	-0.03274656	212	Madde 12	Madde 15	-0.01894873
63	Madde 3	Madde 19	-0.02292184	213	Madde 12	Madde 16	-0.04827596

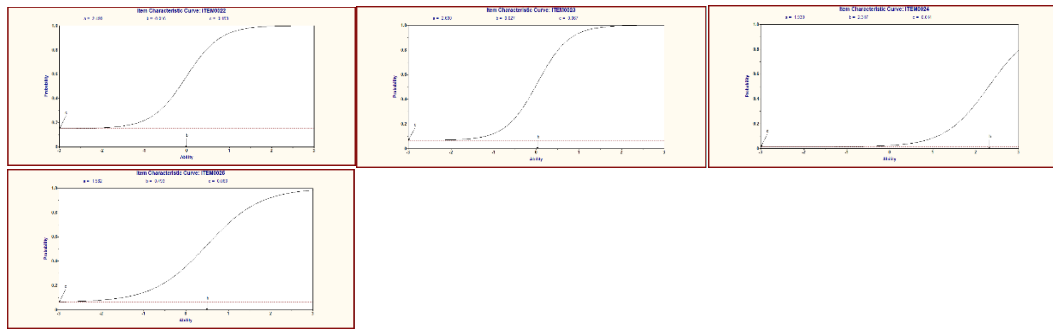
64	Madde 3	Madde 20	-0.02149483	214	Madde 12	Madde 17	-0.04263754
65	Madde 3	Madde 21	-0.08295204	215	Madde 12	Madde 18	-0.04343986
66	Madde 3	Madde 22	-0.02577135	216	Madde 12	Madde 19	-0.03100161
67	Madde 3	Madde 23	-0.02415815	217	Madde 12	Madde 20	-0.2838803
68	Madde 3	Madde 24	-0.02991408	218	Madde 12	Madde 21	-0.01158658
69	Madde 3	Madde 25	-0.176301	219	Madde 12	Madde 22	-0.03453715
70	Madde 4	Madde 5	0.06634414	220	Madde 12	Madde 23	-0.03183926
71	Madde 4	Madde 6	-0.1850145	221	Madde 12	Madde 24	-0.03552233
72	Madde 4	Madde 7	-0.06320368	222	Madde 12	Madde 25	-0.02430403
73	Madde 4	Madde 8	0.03752362	223	Madde 13	Madde 14	-0.00747337
74	Madde 4	Madde 9	0.06530315	224	Madde 13	Madde 15	-0.01106907
75	Madde 4	Madde 10	0.07271012	225	Madde 13	Madde 16	-0.03145302
76	Madde 4	Madde 11	0.0392369	226	Madde 13	Madde 17	-0.02742947
77	Madde 4	Madde 12	0.05614942	227	Madde 13	Madde 18	-0.02798109
78	Madde 4	Madde 13	0.03674935	228	Madde 13	Madde 19	-0.01958649
79	Madde 4	Madde 14	-0.1129199	229	Madde 13	Madde 20	-0.01778042
80	Madde 4	Madde 15	-0.01538963	230	Madde 13	Madde 21	-0.00694489
81	Madde 4	Madde 16	-0.03975143	231	Madde 13	Madde 22	-0.02202032
82	Madde 4	Madde 17	-0.3442421	232	Madde 13	Madde 23	-0.01995758
83	Madde 4	Madde 18	-0.03585298	233	Madde 13	Madde 24	-0.00220524
84	Madde 4	Madde 19	-0.0242419	234	Madde 13	Madde 25	-0.01543841
85	Madde 4	Madde 20	-0.02269254	235	Madde 14	Madde 15	0.003824739
86	Madde 4	Madde 21	-0.0932488	236	Madde 14	Madde 16	-0.00138651
87	Madde 4	Madde 22	-0.02803698	237	Madde 14	Madde 17	-0.00104367
88	Madde 4	Madde 23	-0.02621327	238	Madde 14	Madde 18	-0.0011875
89	Madde 4	Madde 24	-0.0366168	239	Madde 14	Madde 19	0.000049860
90	Madde 4	Madde 25	-0.1944865	240	Madde 14	Madde 20	0.002520839
91	Madde 5	Madde 6	0.02365394	241	Madde 14	Madde 21	0.001542761
92	Madde 5	Madde 7	0.09068532	242	Madde 14	Madde 22	0.001342727
93	Madde 5	Madde 8	0.05102863	243	Madde 14	Madde 23	0.002539979
94	Madde 5	Madde 9	0.09316632	244	Madde 14	Madde 24	0.009940433
95	Madde 5	Madde 10	0.1062109	245	Madde 14	Madde 25	-0.0025519
96	Madde 5	Madde 11	0.05564141	246	Madde 15	Madde 16	-0.00066561

97	Madde 5	Madde 12	0.0808251	247	Madde 15	Madde 17	-0.00011382
98	Madde 5	Madde 13	0.05208005	248	Madde 15	Madde 18	-0.00061312
99	Madde 5	Madde 14	-0.1523708	249	Madde 15	Madde 19	-0.00072659
100	Madde 5	Madde 15	-0.0215372	250	Madde 15	Madde 20	0.002379009
101	Madde 5	Madde 16	-0.05555781	251	Madde 15	Madde 21	0.0143733
102	Madde 5	Madde 17	-0.04826875	252	Madde 15	Madde 22	0.00174013
103	Madde 5	Madde 18	-0.04993934	253	Madde 15	Madde 23	0.00417704
104	Madde 5	Madde 19	-0.0352834	254	Madde 15	Madde 24	0.001436786
105	Madde 5	Madde 20	-0.0321319	255	Madde 15	Madde 25	-0.0016011
106	Madde 5	Madde 21	-0.1316717	256	Madde 16	Madde 17	-0.01545558
107	Madde 5	Madde 22	-0.03919514	257	Madde 16	Madde 18	0.1821694
108	Madde 5	Madde 23	-0.03583893	258	Madde 16	Madde 19	0.004161677
109	Madde 5	Madde 24	-0.00433572	259	Madde 16	Madde 20	0.002788041
110	Madde 5	Madde 25	-0.02731408	260	Madde 16	Madde 21	-0.00127787
111	Madde 6	Madde 7	0.02189941	261	Madde 16	Madde 22	0.05609983
112	Madde 6	Madde 8	0.02072914	262	Madde 16	Madde 23	0.002197731
113	Madde 6	Madde 9	0.02173685	263	Madde 16	Madde 24	0.000284394
114	Madde 6	Madde 10	0.1966635	264	Madde 16	Madde 25	0.004310445
115	Madde 6	Madde 11	0.02092874	265	Madde 17	Madde 18	0.1105397
116	Madde 6	Madde 12	0.1981802	266	Madde 17	Madde 19	0.003380585
117	Madde 6	Madde 13	0.1711091	267	Madde 17	Madde 20	0.001794402
118	Madde 6	Madde 14	-0.02357245	268	Madde 17	Madde 21	-0.00079295
119	Madde 6	Madde 15	-0.03862061	269	Madde 17	Madde 22	0.002812724
120	Madde 6	Madde 16	-0.1433702	270	Madde 17	Madde 23	0.000767432
121	Madde 6	Madde 17	-0.125966	271	Madde 17	Madde 24	0.00092829
122	Madde 6	Madde 18	-0.1264995	272	Madde 17	Madde 25	0.004334441
123	Madde 6	Madde 19	-0.08756455	273	Madde 18	Madde 19	0.003796692
124	Madde 6	Madde 20	-0.00712545	274	Madde 18	Madde 20	0.01684114
125	Madde 6	Madde 21	-0.00213398	275	Madde 18	Madde 21	-0.00010822
126	Madde 6	Madde 22	-0.0097336	276	Madde 18	Madde 22	0.00263755
127	Madde 6	Madde 23	-0.00874805	277	Madde 18	Madde 23	0.00140952
128	Madde 6	Madde 24	-0.00078425	278	Madde 18	Madde 24	0.000583103
129	Madde 6	Madde 25	-0.0676557	279	Madde 18	Madde 25	0.003982394

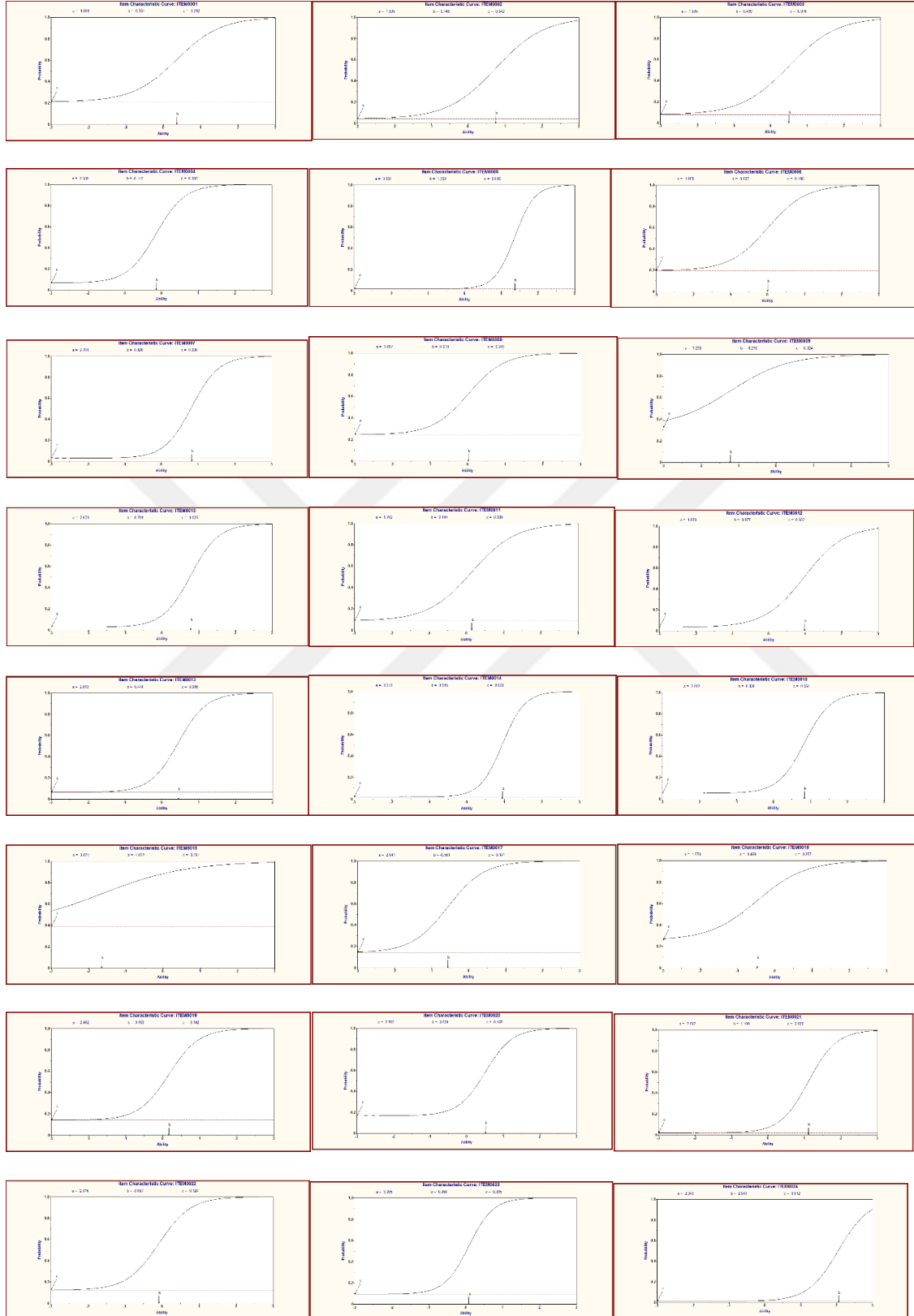
130	Madde 7	Madde 8	0.0488081	280	Madde 19	Madde 20	0.003565576
131	Madde 7	Madde 9	0.08874654	281	Madde 19	Madde 21	0.00184399
132	Madde 7	Madde 10	0.1048642	282	Madde 19	Madde 22	0.003533459
133	Madde 7	Madde 11	0.0523615	283	Madde 19	Madde 23	0.002163222
134	Madde 7	Madde 12	0.07518079	284	Madde 19	Madde 24	0.001551766
135	Madde 7	Madde 13	0.04795399	285	Madde 19	Madde 25	0.005170767
136	Madde 7	Madde 14	-0.01534349	286	Madde 20	Madde 21	0.00218087
137	Madde 7	Madde 15	-0.02122537	287	Madde 20	Madde 22	0.001638145
138	Madde 7	Madde 16	-0.05446263	288	Madde 20	Madde 23	0.002734317
139	Madde 7	Madde 17	-0.04787214	289	Madde 20	Madde 24	0.001840762
140	Madde 7	Madde 18	-0.04871233	290	Madde 20	Madde 25	-0.00119059
141	Madde 7	Madde 19	-0.03458495	291	Madde 21	Madde 22	0.001848441
142	Madde 7	Madde 20	-0.03128996	292	Madde 21	Madde 23	0.0214543
143	Madde 7	Madde 21	-0.013319	293	Madde 21	Madde 24	0.01280171
144	Madde 7	Madde 22	-0.03910452	294	Madde 21	Madde 25	-0.00060327
145	Madde 7	Madde 23	-0.0357576	295	Madde 22	Madde 23	0.003585131
146	Madde 7	Madde 24	-0.00393294	296	Madde 22	Madde 24	0.00138288
147	Madde 7	Madde 25	-0.070886	297	Madde 22	Madde 25	0.04715431
148	Madde 8	Madde 9	0.04834911	298	Madde 23	Madde 24	0.001878213
149	Madde 8	Madde 10	0.04830112	299	Madde 23	Madde 25	0.005410772
150	Madde 8	Madde 11	0.03740957	300	Madde 24	Madde 25	0.000619204

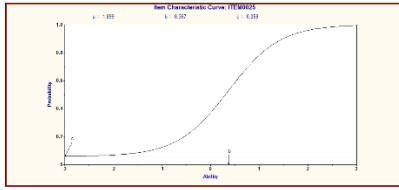
EK 6. Kitaplık 1 Maddelerine İlişkin Madde Karakteristik Eğrileri





EK 7. Kitaplık 3 Maddelerine İlişkin Madde Karakteristik Eğrileri





EK 8. MTK İçin Model Veri Uyum Sonucu

```
anova(mod1,mod2)
```

	Model	loglike	Deviance	Npars	AIC	BIC	Chisq	df	p
1	mod1	179270.8	358541.6	26	358593.6	358804.4	20094.51	24	0
2	mod2	169223.6	338447.1	50	338547.1	338952.4	NA	NA	NA

```
> anova(mod2,mod3)
```

	Model	loglike	Deviance	Npars	AIC	BIC	Chisq	df	p
1	mod2	169223.6	338447.1	50	338547.1	338952.4	82.81478	25	0
2	mod3	169182.2	338364.3	75	338514.3	339122.2	NA	NA	NA

EK 9. Tucker Doğrusal Yöntemine İlişkin Equating Error 2.0 Program Çıktısı

w₁=1 için Tucker eşitleme hatası

Input file: C:\Users\Ceren\Desktop\tC1.txt

Columns read: 2 3 4

Type of equating function: Linear

Number of replications: 500

Seed: 285947671

Scale min, max, inc: 1.000000 25.000000 1.000000

Sample sizes (new, old): 1900.000000 1921.000000

0	0.22930232	0.23762600	0.00000000	0.39432833	0.10974245	1.00000000
1	0.22176972	0.22981996	0.19595918	1.35022661	1.10033983	1.04000000
2	0.21496378	0.22274701	0.19261360	2.30612489	2.09090592	2.03000000
3	0.20895554	0.21653857	0.16023732	3.26202316	3.08143559	3.01800000
4	0.20381556	0.21119643	0.11816937	4.21792144	4.07200878	4.00600000
5	0.19961090	0.20685298	0.10947146	5.17381972	5.06253155	5.00400000
6	0.19640167	0.20351648	0.10947146	6.12971799	6.05311063	6.00400000
7	0.19423720	0.20128143	0.09998000	7.08561627	7.04362859	7.00200000
8	0.19315261	0.20016405	0.09998000	8.04151455	8.03422153	8.00200000
9	0.19316610	0.20016876	0.09998000	8.99741283	9.02481116	9.00200000
10	0.19427745	0.20131785	0.10954451	9.95331110	10.01532672	10.00000000
11	0.19646801	0.20359382	0.10954451	10.90920938	11.00590955	11.00000000
12	0.19970229	0.20693612	0.13402985	11.86510766	11.99642748	11.99400000
13	0.20393062	0.21132986	0.14119490	12.82100593	12.98700645	12.99200000
14	0.20909271	0.21666442	0.14798649	13.77690421	13.97752953	13.99000000
15	0.21512135	0.22293026	0.16690117	14.73280249	14.96810322	14.98800000
16	0.22194594	0.23000123	0.17263835	15.68870077	15.95869997	15.98600000
17	0.22949548	0.23780424	0.19369047	16.64459904	16.94923398	16.97800000
18	0.23770091	0.24632861	0.22100679	17.60049732	17.93979505	17.96600000
19	0.24649673	0.25542074	0.22925095	18.55639560	18.93033730	18.96200000

20	0.25582205	0.26510739	0.26469605	19.51229387	19.92089017	19.95600000
21	0.26562112	0.27523803	0.27147007	20.46819215	20.91144038	20.95200000
22	0.27584344	0.28585475	0.28582512	21.42409043	21.90198572	21.95200000
23	0.28644371	0.29684147	0.31607594	22.37998870	22.89258223	22.93600000
24	0.29738152	0.30815069	0.33200000	23.33588698	23.88314422	23.92400000
25	0.30862097	0.31981857	0.32260812	24.29178526	24.87367991	24.88200000
Mean	0.21670275	0.22455995	0.16042095	11.31622080	11.42765803	11.42577368



w₁=0,5 için Tucker eşitleme hatası

Input file: C:\Users\Ceren\Desktop\TC05.txt

Columns read:2 3 4

Type of equating function: Linear

Number of replications: 500

Seed: 285947671

Scale min, max, inc: 1.000000 25.000000 1.000000

Sample sizes (new, old): 1900.000000 1921.000000

0	0.22930232	0.23608967	0.00000000	0.39432833	0.18350045	1.00000000
1	0.22176972	0.22831309	0.26499811	1.35022661	1.16765780	1.07600000
2	0.21496378	0.22130521	0.22693611	2.30612489	2.15175557	2.05000000
3	0.20895554	0.21511973	0.21032356	3.26202316	3.13585285	3.04200000
4	0.20381556	0.20982812	0.17058722	4.21792144	4.11995012	4.03000000
5	0.19961090	0.20549943	0.14722771	5.17381972	5.10404740	5.01800000
6	0.19640167	0.20219552	0.10925200	6.12971799	6.08814468	6.00800000
7	0.19423720	0.19996719	0.10925200	7.08561627	7.07224195	7.00800000
8	0.19315261	0.19885061	0.10947146	8.04151455	8.05633923	8.00400000
9	0.19316610	0.19886450	0.10947146	8.99741283	9.04043650	9.00400000
10	0.19427745	0.20000863	0.09998000	9.95331110	10.02453378	10.00200000
11	0.19646801	0.20226382	0.10954451	10.90920938	11.00863106	11.00000000
12	0.19970229	0.20559351	0.13402985	11.86510766	11.99272833	11.99400000
13	0.20393062	0.20994657	0.14119490	12.82100593	12.97682561	12.99200000
14	0.20909271	0.21526095	0.14798649	13.77690421	13.96092289	13.99000000
15	0.21512135	0.22146743	0.16063624	14.73280249	14.94502016	14.98600000
16	0.22194594	0.22849335	0.18254862	15.68870077	15.92911744	15.97400000
17	0.22949548	0.23626560	0.20664946	16.64459904	16.91321471	16.96400000
18	0.23770091	0.24471308	0.23172397	17.60049732	17.89731199	17.95200000
19	0.24649673	0.25376838	0.24307201	18.55639560	18.88140927	18.94600000
20	0.25582205	0.26336880	0.25956887	19.51229387	19.86550654	19.93200000

21	0.26562112	0.27345694	0.29145154	20.46819215	20.84960382	20.91600000
22	0.27584344	0.28398082	0.32914435	21.42409043	21.83370110	21.89200000
23	0.28644371	0.29489380	0.36135301	22.37998870	22.81779837	22.86800000
24	0.29738152	0.30615438	0.38247091	23.33588698	23.80189569	23.84600000
25	0.30862097	0.31775081	0.39387308	24.29178526	24.78602495	24.80800000
Mean	0.21670275	0.22309580	0.17649499	11.31622080	11.42764884	11.42240211



EK 10. Levine Gözlenen Puan İçin Eşitleme Hata Değerleri

$w_1=1$ için Levine gözlenen puan eşitleme

Equating Error 2.0

Input file: C:\Users\Ceren\Desktop\LG1.txt

Columns read: 2 3 4

Type of equating function: Linear

Number of replications: 500

Seed: 285947671

Scale min, max, inc: 1.000000 25.000000 1.000000

Sample sizes (new, old): 1900.000000 1921.000000

0	0.22930232	0.23794702	0.00000000	0.39432833	0.02499451	1.00000000
1	0.22176972	0.23015144	0.10888526	1.35022661	1.01696567	1.01200000
2	0.21496378	0.22306946	0.14831049	2.30612489	2.00896511	1.99800000
3	0.20895554	0.21685200	0.13402985	3.26202316	3.00092862	2.99400000
4	0.20381556	0.21150216	0.11816937	4.21792144	3.99293566	3.99400000
5	0.19961090	0.20715240	0.11789826	5.17381972	4.98489227	4.99000000
6	0.19640167	0.20382565	0.11789826	6.12971799	5.97692133	5.99000000
7	0.19423720	0.20156651	0.09981984	7.08561627	6.96893956	6.99400000
8	0.19315261	0.20044552	0.09981984	8.04151455	7.96088986	7.99400000
9	0.19316610	0.20045851	0.12623787	8.99741283	8.95290728	8.99200000
10	0.19427745	0.20160927	0.12623787	9.95331110	9.94485669	9.99200000
11	0.19646801	0.20388853	0.13379088	10.90920938	10.93687336	10.99000000
12	0.19970229	0.20723567	0.15409088	11.86510766	11.92882514	11.98400000
13	0.20393062	0.21163920	0.16023732	12.82100593	12.92083996	12.98200000
14	0.20909271	0.21699386	0.16023732	13.77690421	13.91286981	13.98200000
15	0.21512135	0.22323327	0.19261360	14.73280249	14.90483006	14.97000000

16	0.22194594	0.23033415	0.20664946	15.68870077	15.89683302	15.96400000
17	0.22949548	0.23814848	0.22778060	16.64459904	16.88880088	16.95400000
18	0.23770091	0.24668516	0.23172397	17.60049732	17.88079579	17.95200000
19	0.24649673	0.25579048	0.25027185	18.55639560	18.87277189	18.94200000
20	0.25582205	0.26549157	0.27137428	19.51229387	19.86475881	19.93400000
21	0.26562112	0.27566126	0.28565714	20.46819215	20.85678967	20.92000000
22	0.27584344	0.28624478	0.32428999	21.42409043	21.84877793	21.90600000
23	0.28644371	0.29726844	0.34562407	22.37998870	22.84075347	22.88800000
24	0.29738152	0.30859676	0.36622944	23.33588698	23.83274805	23.87400000
25	0.30862097	0.32028150	0.35902089	24.29178526	24.82471759	24.84800000
Mean	0.21670275	0.22488485	0.17087296	11.31622080	11.35924545	11.40827895



w₁=0,50 için Levine gözlenen puan eşitleme

Equating Error 2.0

Input file: C:\Users\Ceren\Desktop\LG05.txt

Columns read:2 3 4

Type of equating function: Linear

Number of replications: 500

Seed: 285947671

Scale min, max, inc: 1.000000 25.000000 1.000000

Sample sizes (new, old): 1900.000000 1921.000000

0	0.22930232	0.23824511	0.00000000	0.39432833	0.01660714	1.00000000
1	0.22176972	0.23041873	0.10888526	1.35022661	1.00978545	1.01200000
2	0.21496378	0.22334737	0.14136478	2.30612489	2.00296376	1.99600000
3	0.20895554	0.21710481	0.13402985	3.26202316	2.99614207	2.99400000
4	0.20381556	0.21176436	0.11816937	4.21792144	3.98932038	3.99400000
5	0.19961090	0.20741189	0.11789826	5.17381972	4.98251824	4.99000000
6	0.19640167	0.20406641	0.11789826	6.12971799	5.97577384	5.99000000
7	0.19423720	0.20181245	0.09981984	7.08561627	6.96895531	6.99400000
8	0.19315261	0.20068556	0.09998000	8.04151455	7.96213362	7.99800000
9	0.19316610	0.20069958	0.12623787	8.99741283	8.95531193	8.99200000
10	0.19427745	0.20185427	0.12623787	9.95331110	9.94849024	9.99200000
11	0.19646801	0.20413027	0.13379088	10.90920938	10.94166855	10.99000000
12	0.19970229	0.20749068	0.15409088	11.86510766	11.93484686	11.98400000
13	0.20393062	0.21188391	0.16023732	12.82100593	12.92802517	12.98200000
14	0.20909271	0.21724733	0.16023732	13.77690421	13.92120348	13.98200000
15	0.21512135	0.22351293	0.17726816	14.73280249	14.91438271	14.97600000
16	0.22194594	0.23062268	0.20209899	15.68870077	15.90762833	15.96600000
17	0.22949548	0.23844580	0.21962696	16.64459904	16.90083840	16.95800000
18	0.23770091	0.24697124	0.23172397	17.60049732	17.89401671	17.95200000

19	0.24649673	0.25611010	0.25199206	18.55639560	18.88719502	18.95000000
20	0.25582205	0.26579911	0.26577434	19.51229387	19.88037333	19.94200000
21	0.26562112	0.27598034	0.28267296	20.46819215	20.87355164	20.93600000
22	0.27584344	0.28660133	0.30695928	21.42409043	21.86672995	21.92400000
23	0.28644371	0.29761502	0.33504627	22.37998870	22.85990826	22.91200000
24	0.29738152	0.30897939	0.36494383	23.33588698	23.85308657	23.89600000
25	0.30862097	0.32065719	0.34065232	24.29178526	24.84626488	24.86600000
Mean	0.21670275	0.22515651	0.16836836	11.31622080	11.36456062	11.41138316



EK 11. Levine Gerçek Puan Eşitlemenin Hata Değeri

Equating Error 2.0

Input file: C:\Users\Ceren\Desktop\Lge.txt

Columns read:2 3 4

Type of equating function: Linear

Number of replications: 500

Seed: 285947671

Scale min, max, inc: 1.000000 25.000000 1.000000

Sample sizes (new, old): 1900.000000 1921.000000

0	0.22930232	0.28206479	0.00000000	0.39432833	-2.07503672	1.00000000
1	0.22176972	0.27277792	0.00000000	1.35022661	-0.89922177	1.00000000
2	0.21496378	0.26442541	0.00000000	2.30612489	0.27656491	1.00000000
3	0.20895554	0.25703622	0.49638695	3.26202316	1.45241469	1.44000000
4	0.20381556	0.25069589	0.46249324	4.21792144	2.62824195	2.69000000
5	0.19961090	0.24553757	0.30659419	5.17381972	3.80401781	3.90000000
6	0.19640167	0.24159369	0.17816846	6.12971799	4.97986610	4.98400000
7	0.19423720	0.23891833	0.26390150	7.08561627	6.15570357	6.06600000
8	0.19315261	0.23758876	0.43185183	8.04151455	7.33147311	7.24800000
9	0.19316610	0.23761362	0.49909919	8.99741283	8.50731752	8.53000000
10	0.19427745	0.23897313	0.42708313	9.95331110	9.68316223	9.76000000
11	0.19646801	0.24166163	0.26058396	10.90920938	10.85893167	10.93600000
12	0.19970229	0.24564918	0.17884071	11.86510766	12.03476600	12.00400000
13	0.20393062	0.25083810	0.34585546	12.82100593	13.21053929	13.12800000
14	0.20909271	0.25720494	0.47759397	13.77690421	14.38636987	14.35200000
15	0.21512135	0.26461893	0.49254847	14.73280249	15.56221942	15.58600000
16	0.22194594	0.27299486	0.40045474	15.68870077	16.73800257	16.80600000
17	0.22949548	0.28230239	0.27319590	16.64459904	17.91382128	17.94200000

18	0.23770091	0.29239509	0.31192307	17.60049732	19.08967141	19.05200000
19	0.24649673	0.30319181	0.43852480	18.55639560	20.26546676	20.23600000
20	0.25582205	0.31468615	0.49614111	19.51229387	21.44127292	21.43800000
21	0.26562112	0.32674020	0.48438002	20.46819215	22.61712302	22.62400000
22	0.27584344	0.33928948	0.42997209	21.42409043	23.79293052	23.81800000
23	0.28644371	0.35235172	0.27738782	22.37998870	24.96872536	24.91600000
24	0.29738152	0.36580890	0.00000000	23.33588698	26.14457453	25.00000000
25	0.30862097	0.37960914	0.00000000	24.29178526	27.32039302	25.00000000
Mean	0.21670275	0.26655859	0.34475783	11.31622080	11.35959044	11.42587263



EK 12. Klasik Konjenerik Eşitleme Hatası

Equating Error 2.0

Input file: C:\Users\Ceren\Desktop\kon.txt

Columns read: 2 3 4

Type of equating function: Linear

Number of replications: 500

Seed: 285947671

Scale min, max, inc: 1.000000 25.000000 1.000000

Sample sizes (new, old): 1900.000000 1921.000000

0	0.22930232	0.23748842	0.00000000	0.39432833	0.00600585	1.00000000
1	0.22176972	0.22968689	0.10888526	1.35022661	0.99602970	1.01200000
2	0.21496378	0.22263799	0.14136478	2.30612489	1.98605354	1.99600000
3	0.20895554	0.21641526	0.12623787	3.26202316	2.97607739	2.99200000
4	0.20381556	0.21109177	0.12592061	4.21792144	3.96610124	3.98800000
5	0.19961090	0.20673701	0.12592061	5.17381972	4.95612508	4.98800000
6	0.19640167	0.20341321	0.11789826	6.12971799	5.94614893	5.99000000
7	0.19423720	0.20117146	0.12592061	7.08561627	6.93617277	6.98800000
8	0.19315261	0.20003711	0.14722771	8.04151455	7.92618641	7.98200000
9	0.19316610	0.20005209	0.14051334	8.99741283	8.91612847	8.98400000
10	0.19427745	0.20121315	0.13343163	9.95331110	9.90614431	9.98600000
11	0.19646801	0.20348192	0.14722771	10.90920938	10.89616815	10.98200000
12	0.19970229	0.20683166	0.15362291	11.86510766	11.88619200	11.98000000
13	0.20393062	0.21121094	0.17058722	12.82100593	12.87621585	12.97000000
14	0.20909271	0.21655732	0.18629010	13.77690421	13.86623969	13.96400000
15	0.21512135	0.22280118	0.22202703	14.73280249	14.85626354	14.94800000
16	0.22194594	0.22986941	0.23846174	15.68870077	15.84628738	15.94400000
17	0.22949548	0.23768847	0.26930280	16.64459904	16.83631123	16.92600000

18	0.23770091	0.24618683	0.28450659	17.60049732	17.82633507	17.91600000
19	0.24649673	0.25529666	0.30918603	18.55639560	18.81635892	18.89800000
20	0.25582205	0.26495490	0.33105891	19.51229387	19.80638276	19.88000000
21	0.26562112	0.27510379	0.35065082	20.46819215	20.79640661	20.86200000
22	0.27584344	0.28569105	0.37741754	21.42409043	21.78643046	21.83400000
23	0.28644371	0.29666960	0.38770607	22.37998870	22.77645424	22.82200000
24	0.29738152	0.30797226	0.40987803	23.33588698	23.76644268	23.80000000
25	0.30862097	0.31963339	0.41424630	24.29178526	24.75640484	24.78000000
Mean	0.21670275	0.22443760	0.19424560	11.31622080	11.31774717	11.38961263



EK 13. Braun-Holland Yöntemi için Eşitleme Hatası

Equating Error 2.0

Input file: C:\Users\Ceren\Desktop\BRAUNHOLLAND.txt

Columns read: 2 3 4

Type of equating function: Post-Smoothing

Smoothing parameter: 6.000000

Number of replications: 500

Seed: 285947671

Scale min, max, inc: 1.000000 12.000000 1.000000

Sample sizes (new, old): 1085.000000 1137.000000

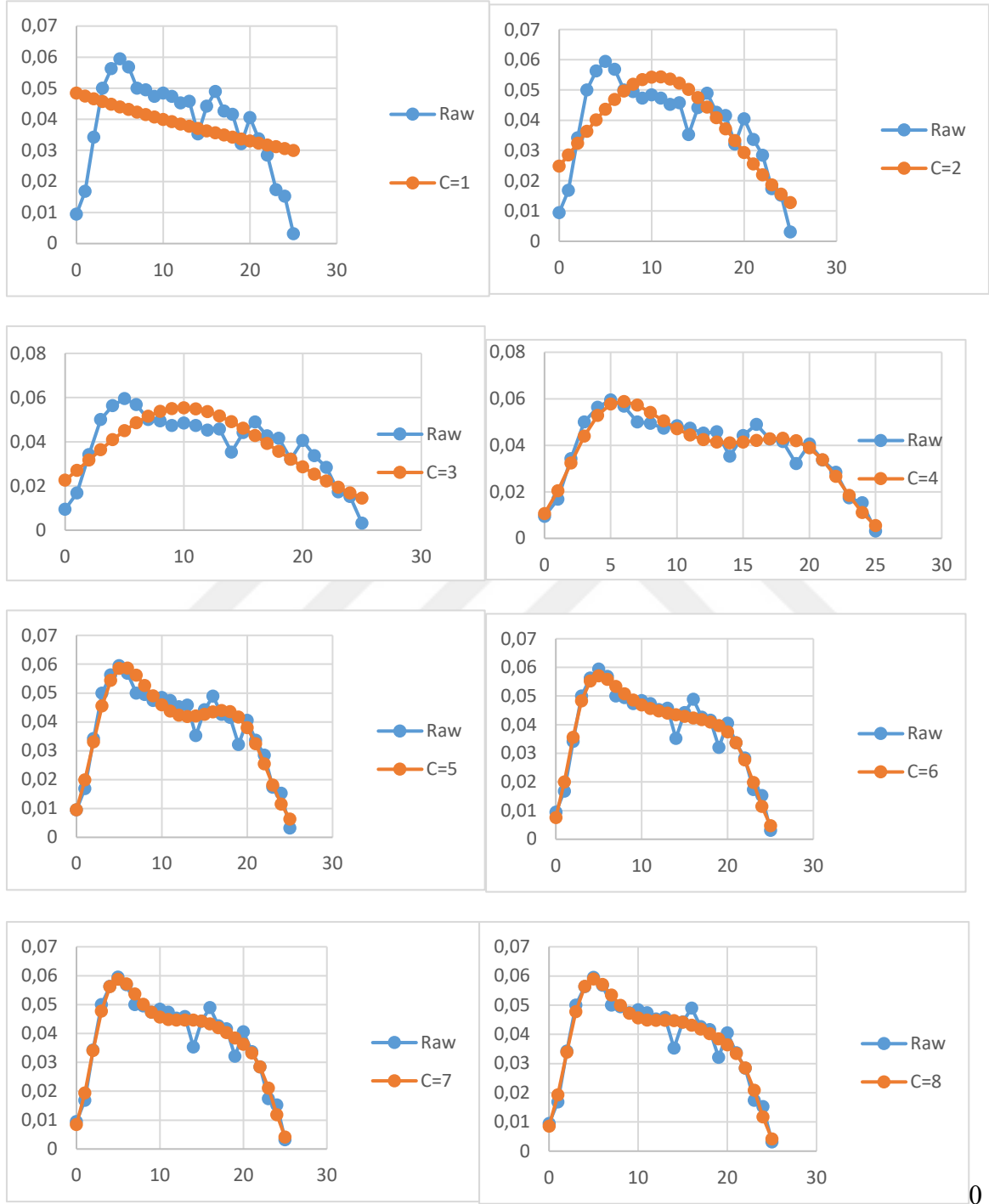
0	0.16872145	0.19401954	0.00000000	0.35723486	-0.95555134	1.00000000
1	0.17677903	0.20328356	0.00000000	1.38185706	0.22269888	1.00000000
2	0.16349599	0.18800942	0.41559115	2.36163188	1.34937244	1.22200000
3	0.15095544	0.17358864	0.49662461	3.34140671	2.47604784	2.44200000
4	0.13935797	0.16025232	0.43863424	4.32118153	3.60272324	3.74000000
5	0.12895826	0.14829336	0.25848791	5.30095635	4.72939863	4.92800000
6	0.12006797	0.13807012	0.06311894	6.28073117	5.85607403	5.99600000
7	0.11304376	0.12999276	0.00000000	7.26050599	6.98274943	7.00000000
8	0.10824953	0.12447971	0.04467662	8.24028081	8.10942483	8.00200000
9	0.10598832	0.12187904	0.12547510	9.22005563	9.23609935	9.01600000
10	0.10642170	0.12237854	0.33184936	10.19983045	10.36277304	10.12600000
11	0.10951769	0.12593771	0.49787950	11.17960527	11.48945347	11.45400000
12	0.11506156	0.13231274	0.00000000	12.15938009	12.61612595	12.00000000
Mean	0.12544617	0.14425467	0.22855692	6.95454180	6.63091102	6.62590968

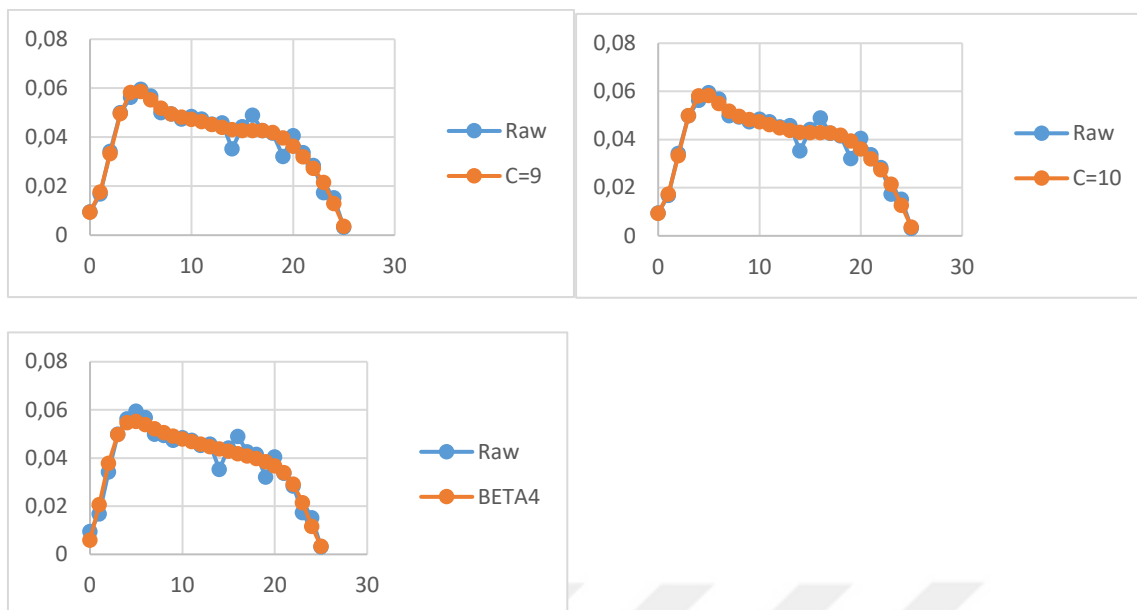


EK 14. Ön Düzgünleştirme İçin Çeşitli C Ponimomial Dereceler İçin Eşitlenmiş Puanlar

X Score	C=1	C=2	C=3	C=4	C=5	C=6	C=7	C=8	C=9	C=10	Beta4
0	0.01154	-0.07247	-0.11318	-0.13641	-0.00973	-0.06812	-0.11071	-0.10170	-0.16261	-0.16638	-0.16374
1	1.03405	0.81889	0.72963	0.75477	0.91359	0.87427	0.84167	0.84258	0.83534	0.83608	0.78336
2	2.05492	1.75283	1.64719	1.70934	1.80833	1.81406	1.81876	1.81336	1.87209	1.87715	1.73395
3	3.07413	2.72040	2.61964	2.69164	2.71287	2.74600	2.77329	2.76892	2.79836	2.79832	2.69649
4	4.09166	3.71470	3.63251	3.69219	3.64181	3.67946	3.70860	3.70908	3.68309	3.67859	3.68155
5	5.10746	4.73033	4.67489	4.70967	4.60809	4.63179	4.64469	4.64957	4.59094	4.58676	4.69699
6	6.12152	5.76297	5.73826	5.74547	5.62391	5.62346	5.61055	5.61654	5.57677	5.57810	5.74312
7	7.13380	6.80908	6.81585	6.80136	6.69668	6.67088	6.63650	6.63973	6.65938	6.66555	6.81326
8	8.14428	7.86569	7.90212	7.87804	7.82435	7.77910	7.74111	7.73925	7.80913	7.81377	7.89857
9	9.15292	8.93026	8.99251	8.97379	8.99287	8.93798	8.91843	8.91179	8.97854	8.97688	8.99163
10	10.15968	10.00053	10.08318	10.08373	10.17776	10.12430	10.13577	10.12706	10.13861	10.13205	10.08711
11	11.16455	11.07443	11.17092	11.19963	11.34985	11.30846	11.34629	11.33915	11.28509	11.27934	11.18100
12	12.16748	12.15003	12.25302	12.31100	12.48304	12.46263	12.50842	12.50573	12.42213	12.42224	12.27000
13	13.16845	13.22546	13.32718	13.40693	13.56091	13.56775	13.59966	13.60256	13.54485	13.55110	13.35114
14	14.16741	14.29882	14.39146	14.47852	14.57993	14.61318	14.61927	14.62638	14.62921	14.63639	14.42153
15	15.16434	15.36817	15.44419	15.52040	15.55101	15.60005	15.57918	15.58706	15.64581	15.64875	15.47827
16	16.15920	16.43142	16.48391	16.53269	16.49026	16.53916	16.49849	16.50332	16.58435	16.58134	16.51871
17	17.15195	17.48625	17.51027	17.51958	17.41573	17.44920	17.40459	17.40404	17.45992	17.45306	17.53915
18	18.14255	18.53352	18.52143	18.48617	18.34679	18.35349	18.32012	18.31454	18.31489	18.30896	18.53534
19	19.13097	19.56744	19.51421	19.43977	19.29817	19.27609	19.26719	19.25993	19.20364	19.20262	19.50339
20	20.11718	20.58157	20.48850	20.38674	20.27950	20.23841	20.25926	20.25518	20.18102	20.18619	20.44401
21	21.10112	21.56943	21.44577	21.33173	21.29382	21.25196	21.29241	21.29446	21.27168	21.27770	21.36068
22	22.08276	22.52258	22.38358	22.27731	22.33454	22.31062	22.34221	22.34870	22.40775	22.40534	22.26480
23	23.06207	23.44135	23.30112	23.22334	23.37941	23.38476	23.37492	23.38049	23.44181	23.43477	23.17229
24	24.03899	24.31539	24.19777	24.16494	24.37874	24.41969	24.36624	24.36488	24.32764	24.32978	24.09652
25	25.01349	25.12640	25.07318	25.08446	25.22663	25.30893	25.28041	25.26666	25.17599	25.18359	25.04417

EK 15-.Polynomial Log-Lineer (C) ve Beta 4 ile Elde Edilen Grafikler



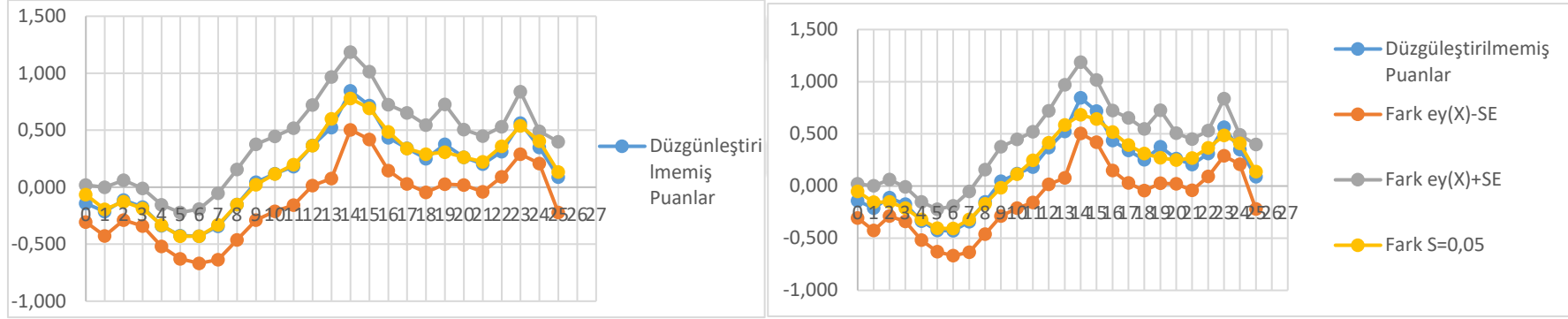


EK 16. C 6 ve Beta4 için Hata Bant Grafiği Oluşturmak için Hesaplanan İstatistikler

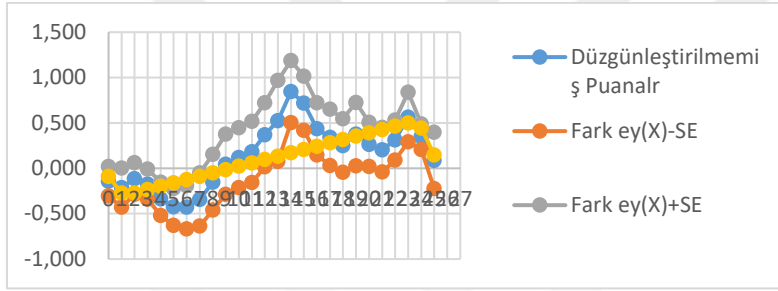
Score	SE	NoSmooth	eqY(X)-Se	eqY(X)+Se	C=6 (X)	Beta4	fark Beta4	fark C6	fark nosmooth	eqY(X)-Se	eqY(X)+Se
0	0,163	0,138	-0,025	0,301	-0,068	-0,164	-0,164	-0,068	0,138	-0,025	0,301
1	0,214	1,149	0,936	1,363	0,874	0,783	-0,217	-0,126	0,149	-0,064	0,363
2	0,174	2,081	1,907	2,255	1,814	1,734	-0,266	-0,186	0,081	-0,093	0,255
3	0,166	2,901	2,735	3,066	2,746	2,699	-0,301	-0,254	-0,099	-0,265	0,066
4	0,183	3,708	3,525	3,891	3,679	3,682	-0,318	-0,321	-0,292	-0,475	-0,109
5	0,204	4,532	4,328	4,735	4,632	4,697	-0,303	-0,368	-0,468	-0,672	-0,265
6	0,238	5,444	5,205	5,682	5,623	5,743	-0,257	-0,377	-0,556	-0,795	-0,318
7	0,292	6,648	6,356	6,939	6,671	6,813	-0,187	-0,329	-0,352	-0,644	-0,061
8	0,308	7,818	7,510	8,126	7,779	7,899	-0,101	-0,221	-0,182	-0,490	0,126
9	0,332	9,054	8,722	9,385	8,938	8,992	-0,008	-0,062	0,054	-0,278	0,385
10	0,329	10,109	9,780	10,439	10,124	10,087	0,087	0,124	0,109	-0,220	0,439
11	0,338	11,166	10,828	11,504	11,308	11,181	0,181	0,308	0,166	-0,172	0,504
12	0,353	12,369	12,017	12,722	12,463	12,270	0,270	0,463	0,369	0,017	0,722
13	0,446	13,523	13,077	13,968	13,568	13,351	0,351	0,568	0,523	0,077	0,968
14	0,341	14,845	14,504	15,186	14,613	14,422	0,422	0,613	0,845	0,504	1,186
15	0,298	15,717	15,419	16,015	15,600	15,478	0,478	0,600	0,717	0,419	1,015
16	0,288	16,425	16,137	16,713	16,539	16,519	0,519	0,539	0,425	0,137	0,713
17	0,311	17,337	17,025	17,648	17,449	17,539	0,539	0,449	0,337	0,025	0,648

18	0,296	18,177	17,881	18,473	18,353	18,535	0,535	0,353	0,177	-0,119	0,473
19	0,350	19,402	19,053	19,752	19,276	19,503	0,503	0,276	0,402	0,053	0,752
20	0,243	20,215	19,972	20,457	20,238	20,444	0,444	0,238	0,215	-0,028	0,457
21	0,245	21,149	20,904	21,394	21,252	21,361	0,361	0,252	0,149	-0,096	0,394
22	0,220	22,193	21,973	22,413	22,311	22,265	0,265	0,311	0,193	-0,027	0,413
23	0,274	23,564	23,291	23,838	23,385	23,172	0,172	0,385	0,564	0,291	0,838
24	0,141	23,769	23,628	23,911	24,420	24,097	0,097	0,420	-0,231	-0,372	-0,089
25	0,310	24,088	23,778	24,398	25,309	25,044	0,044	0,309	-0,912	-1,222	-0,602

EK 17. Farklı Kübik Spline Derecelerinin Hata Bant Değerleri Arasındaki Dağılımları







EK 18. Ön Düzgünleştirme İçin C6 Polinom Fonksiyonun Eşitleme Hatası

Equating Error 2.0

Input file: C:\Users\Ceren\Desktop\ce.txt

Columns read: 2 3 4

Type of equating function: Pre-Smoothing

Smoothing parameter: 6.000000

Number of replications: 500

Seed: 285947671

Scale min, max, inc: 1.000000 25.000000 1.000000

Sample sizes (new, old): 1900.000000 1921.000000

0	0.15212312	0.14375635	0.00000000	0.19294102	0.08032926	1.00000000
1	0.19049280	0.18449796	0.13295112	1.31170439	1.14547215	1.01800000
2	0.18454389	0.17662556	0.04467662	2.27467290	2.07550069	2.00200000
3	0.17865748	0.16822307	0.04467662	3.23780657	2.99240935	2.99800000
4	0.17333130	0.16308216	0.10888526	4.20094023	3.89809573	3.98800000
5	0.16861846	0.16228971	0.17058722	5.16407389	4.80906735	4.97000000
6	0.16457166	0.16618759	0.24115555	6.12720755	5.74841458	5.93800000
7	0.16124104	0.17342627	0.26499811	7.09034121	6.74171004	6.92400000
8	0.15867172	0.18153791	0.18629010	8.05347488	7.80488004	7.96400000
9	0.15690110	0.18745868	0.09981984	9.01660854	8.93455323	8.99400000
10	0.15595638	0.18975098	0.08908423	9.97974220	10.10260166	10.00800000
11	0.15585260	0.18705062	0.31038041	10.94287586	11.27420182	11.10800000
12	0.15659141	0.17924508	0.46249324	11.90600952	12.39699924	12.31000000
13	0.15816101	0.17063629	0.49507575	12.86914319	13.45589082	13.43000000
14	0.16053703	0.16236943	0.48989795	13.83227685	14.44745718	14.40000000
15	0.16368437	0.15652264	0.42708313	14.79541051	15.38231276	15.24000000
16	0.16755955	0.15384969	0.22992173	15.75854417	16.27604520	16.05600000
17	0.17211343	0.15618059	0.11749043	16.72167783	17.15259041	17.01400000

18	0.17729372	0.16246751	0.06324555	17.68481150	18.03169372	18.00000000
19	0.18304724	0.17362326	0.11789826	18.64794516	18.93370455	18.99000000
20	0.18932173	0.18799124	0.17124252	19.61107882	19.87326715	19.97400000
21	0.19606719	0.20256862	0.22693611	20.57421248	20.85219385	20.95000000
22	0.20323673	0.21332112	0.23846174	21.53734614	21.85620297	21.94400000
23	0.21070705	0.21761668	0.23846174	22.50045494	22.85891826	22.94400000
24	0.21527925	0.21351333	0.24576411	23.48570816	23.85620245	23.94000000
25	0.09477403	0.08662346	0.00000000	24.80312576	25.10005694	25.00000000
Mean	0.17063122	0.17469311	0.21388902	11.35302336	11.46429852	11.48442421



EK 19. Ön Düzgünleştirme Yöntemlerinden Beta4 Fonksiyona İlişkin Eşitleme Hatası

Equating Error 2.0 Çıktıları

Beta 4 için Out

Input file: C:\Users\Ceren\Desktop\beta4.txt

Columns read: 2 3 4

Type of equating function: Post-Smoothing

Smoothing parameter: 6.000000

Number of replications: 500

Seed: 285947671

Scale min, max, inc: 1.000000 25.000000 1.000000

Sample sizes (new, old): 1900.000000 1921.000000

0	0.15212312	0.14406060	0.00000000	0.19294102	0.01871515	1.00000000
1	0.19049280	0.18111373	0.04467662	1.31170439	1.07944990	1.00200000
2	0.18454389	0.17788403	0.04467662	2.27467290	1.99914596	1.99800000
3	0.17865748	0.17530120	0.10888526	3.23780657	2.93290877	2.98800000
4	0.17333130	0.17521124	0.12547510	4.20094023	3.88630569	3.98400000
5	0.16861846	0.17546391	0.13295112	5.16407389	4.86910762	4.98200000
6	0.16457166	0.17515850	0.12547510	6.12720755	5.87966119	5.98400000
7	0.16124104	0.17430786	0.09949874	7.09034121	6.91162109	6.99000000
8	0.15867172	0.17299211	0.07722694	8.05347488	7.95774639	7.99400000
9	0.15690110	0.17165503	0.07743384	9.01660854	9.01029939	9.00200000
10	0.15595638	0.17070627	0.06311894	9.97974220	10.06476399	10.00400000
11	0.15585260	0.17024418	0.06311894	10.94287586	11.11831638	11.00400000
12	0.15659141	0.17021573	0.12547510	11.90600952	12.16744050	12.01600000
13	0.15816101	0.17068540	0.18122914	12.86914319	13.20937755	13.03400000
14	0.16053703	0.17162349	0.20509510	13.83227685	14.24220759	14.04400000
15	0.16368437	0.17262489	0.24828210	14.79541051	15.26185057	15.06600000
16	0.16755955	0.17418908	0.26817159	15.75854417	16.26754492	16.07800000

17	0.17211343	0.17533568	0.24115555	16.72167783	17.25502556	17.06200000
18	0.17729372	0.17637315	0.20058913	17.68481150	18.22099245	18.04200000
19	0.18304724	0.17702485	0.15913516	18.64794516	19.16215062	19.02600000
20	0.18932173	0.17804529	0.11789826	19.61107882	20.07798004	20.01000000
21	0.19606719	0.17976384	0.08935323	20.57421248	20.97052986	20.99600000
22	0.20323673	0.18375428	0.18122914	21.53734614	21.84675850	21.96600000
23	0.21070705	0.19114443	0.31288976	22.50045494	22.71893014	22.89000000
24	0.21527925	0.19904486	0.43642181	23.48570816	23.62133205	23.74400000
25	0.09477403	0.08975100	0.04467662	24.80312576	24.85756010	24.99800000
Mean	0.17063122	0.17461592	0.14255882	11.35302336	11.46795931	11.44070105



EK 20. Son Düzgünleştirme Yöntemlerinden S 0,05 Fonksiyona İlişkin Eşitleme Hatası

S 0,05 Kübik Spline

Input file: C:\Users\Ceren\Desktop\k.txt

Columns read:2 3 4

Type of equating function: Post-Smoothing

Smoothing parameter: 6.000000

Number of replications: 500

Seed: 285947671

Scale min, max, inc: 1.000000 25.000000 1.000000

Sample sizes (new, old): 1900.000000 1921.000000

0	0.15212312	0.13645444	0.00000000	0.19294102	0.12106809	1.00000000
1	0.19049280	0.18970874	0.16497273	1.31170439	1.15782716	1.02800000
2	0.18454389	0.17437013	0.06311894	2.27467290	2.10480899	2.00400000
3	0.17865748	0.16022190	0.04467662	3.23780657	2.99873700	2.99800000
4	0.17333130	0.15838195	0.11749043	4.20094023	3.86353669	3.98600000
5	0.16861846	0.16563763	0.22202703	5.16407389	4.76065697	4.94800000
6	0.16457166	0.17525020	0.28902595	6.12720755	5.73220657	5.90800000
7	0.16124104	0.18224139	0.20948508	7.09034121	6.78393454	6.95400000
8	0.15867172	0.18295429	0.12547510	8.05347488	7.88997659	7.98400000
9	0.15690110	0.17911720	0.08944272	9.01660854	9.00028151	9.00000000
10	0.15595638	0.17633874	0.06311894	9.97974220	10.08934383	10.00400000
11	0.15585260	0.17839796	0.17058722	10.94287586	11.18160704	11.03000000
12	0.15659141	0.18309425	0.34698703	11.90600952	12.30334654	12.14000000
13	0.15816101	0.18333322	0.48538644	12.86914319	13.43122492	13.38000000
14	0.16053703	0.17317001	0.49998400	13.83227685	14.49671648	14.49600000
15	0.16368437	0.15587287	0.48587653	14.79541051	15.44501085	15.38200000
16	0.16755955	0.14662610	0.26817159	15.75854417	16.30573089	16.07800000

17	0.17211343	0.15116068	0.08908423	16.72167783	17.14933999	17.00800000
18	0.17729372	0.16307605	0.06324555	17.68481150	18.02177666	18.00000000
19	0.18304724	0.17583590	0.11789826	18.64794516	18.93306756	18.99000000
20	0.18932173	0.18535219	0.17124252	19.61107882	19.86870881	19.97400000
21	0.19606719	0.20005817	0.23846174	20.57421248	20.83423442	20.94400000
22	0.20323673	0.22271418	0.25279240	21.53734614	21.85699027	21.93600000
23	0.21070705	0.23565900	0.24543023	22.50045494	22.92488085	22.95800000
24	0.21527925	0.19998893	0.16560193	23.48570816	23.93324742	23.97600000
25	0.09477403	0.06890072	0.00000000	24.80312576	24.99287243	25.00000000
Mean	0.17063122	0.17450336	0.20480661	11.35302336	11.46610961	11.48128316



EK 21. EYE Yöntemlerden Frekans Kestirim için İstatistiksel Süreç

FREKANS	KESTİRİM YÖNTEMİ															
X	0	1	2	3	4	5	6	7	8	9	10	11	12	13	f(x)	F (X)
0	0,0052	0,0072	0,0071	0,0040	0,0022	0,0005	0,0014	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0275	0,0275
1	0,0064	0,0187	0,0171	0,0092	0,0053	0,0024	0,0023	0,0000	0,0011	0,0000	0,0000	0,0000	0,0000	0,0000	0,0626	0,0901
2	0,0088	0,0176	0,0248	0,0213	0,0209	0,0102	0,0036	0,0011	0,0000	0,0006	0,0000	0,0000	0,0000	0,0000	0,1088	0,1990
3	0,0028	0,0138	0,0218	0,0225	0,0164	0,0169	0,0082	0,0054	0,0023	0,0011	0,0000	0,0000	0,0000	0,0000	0,1112	0,3101
4	0,0008	0,0105	0,0100	0,0225	0,0164	0,0164	0,0105	0,0140	0,0074	0,0028	0,0000	0,0000	0,0000	0,0000	0,1117	0,4218
5	0,0008	0,0044	0,0065	0,0150	0,0191	0,0131	0,0196	0,0183	0,0170	0,0084	0,0000	0,0011	0,0005	0,0000	0,1246	0,5464
6	0,0000	0,0006	0,0047	0,0052	0,0102	0,0121	0,0187	0,0161	0,0176	0,0124	0,0080	0,0016	0,0005	0,0000	0,1074	0,6539
7	0,0000	0,0000	0,0005	0,0023	0,0027	0,0048	0,0091	0,0145	0,0113	0,0129	0,0080	0,0066	0,0027	0,0008	0,0761	0,7300
8	0,0000	0,0000	0,0006	0,0006	0,0004	0,0063	0,0073	0,0059	0,0119	0,0208	0,0130	0,0137	0,0049	0,0000	0,0857	0,8156
9	0,0000	0,0000	0,0000	0,0012	0,0004	0,0005	0,0023	0,0048	0,0074	0,0124	0,0150	0,0104	0,0109	0,0056	0,0709	0,8866
10	0,0004	0,0000	0,0000	0,0000	0,0000	0,0000	0,0014	0,0016	0,0040	0,0056	0,0090	0,0186	0,0125	0,0024	0,0557	0,9422
11	0,0000	0,0000	0,0000	0,0006	0,0000	0,0000	0,0005	0,0000	0,0023	0,0045	0,0070	0,0121	0,0136	0,0024	0,0432	0,9854
12	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0006	0,0020	0,0038	0,0044	0,0040	0,0146	1,0000

Y	0	1	2	3	4	5	6	7	8	9	10	11	12	13	g ₁ (y)	G ₁ (y)
0	0,0123	0,012	0,0121	0,2827	0,0043	0,0023	0,0006	0,001	0,0005	0,0015	0	0	0,0005	0	0,23	0,23
1	0,0055	0,0139	0,0181	0,0076	0,0148	0,0068	0,0018	0,0015	0,0015	0,001	0,0006	0,0005	0	0	0,07	0,3
2	0,0075	0,0184	0,02	0,0124	0,0148	0,0079	0,006	0,001	0,0015	0,0005	0,0006	0	0	0	0,09	0,39
3	0,0021	0,012	0,0116	0,0214	0,0166	0,0119	0,0096	0,0056	0,0024	0,0024	0	0	0	0	0,07	0,46
4	0,0048	0,007	0,0098	0,0205	0,0197	0,017	0,015	0,0066	0,0048	0,001	0	0,0005	0	0	0,1	0,56
5	0,0007	0,0035	0,006	0,0162	0,016	0,013	0,0192	0,0092	0,0092	0,0034	0,001	0,003	0,001	0	0,09	0,65
6	0	0,001	0,0028	0,009	0,0123	0,0136	0,0192	0,0194	0,0106	0,0083	0,007	0,0015	0	0	0,05	0,7
7	0	0,001	0,0009	0,0038	0,008	0,0096	0,013	0,0107	0,014	0,0117	0,008	0,0085	0,003	0	0,09	0,79
8	0	0	0,0009	0,0029	0,0031	0,0045	0,0084	0,0117	0,016	0,0195	0,015	0,009	0,0035	0,0014	0,05	0,84
9	0	0	0	0,0005	0,0006	0,0017	0,006	0,0082	0,0097	0,0112	0,014	0,0115	0,0086	0,001	0,06	0,9
10	0	0	0	0	0	0,0011	0,0018	0,0015	0,0044	0,0059	0,017	0,0145	0,0121	0,0024	0,05	0,95
11	0	0	0	0	0	0	0	0,0026	0,001	0,0088	0,007	0,0115	0,0091	0,0031	0,03	0,99
12	0	0	0	0	0	0	0	0	0	0,001	0,002	0,004	0,0101	0,002	0,02	1

FREKANS Kestirim Yöntemi için Koşullu Olasılık Değerleri

Vtoplam															
Y FORMU	0	1	2	3	4	5	6	7	8	9	10	11	12	13	
0	0,375	0,1739	0,147	0,081	0,039	0,03	0,01	0,01	0,0064	0,019	0	0	0,011	0	
1	0,166667	0,2029	0,22	0,131	0,134	0,08	0,02	0,02	0,0192	0,013	0,008	0,008	0	0	
2	0,229167	0,2681	0,243	0,227	0,134	0,09	0,06	0,01	0,0192	0,006	0,008	0	0	0	

3	0,0625	0,1739	0,141	0,217	0,151	0,13	0,1	0,07	0,0321	0,032	0	0	0	0
4	0,145833	0,1014	0,119	0,172	0,179	0,19	0,16	0,08	0,0641	0,013	0	0,008	0	0
5	0,020833	0,0507	0,073	0,096	0,145	0,15	0,2	0,12	0,1218	0,045	0,017	0,047	0,021	0
6	0	0,0145	0,034	0,04	0,112	0,15	0,17	0,25	0,141	0,109	0,091	0,023	0	0
7	0	0,0145	0,011	0,03	0,073	0,11	0,12	0,14	0,1859	0,154	0,107	0,132	0,063	0
8	0	0	0,011	0,005	0,028	0,05	0,09	0,15	0,2115	0,256	0,207	0,14	0,074	0,13795958
9	0	0	0	0	0,006	0,02	0,06	0,1	0,1282	0,147	0,19	0,178	0,179	0,10346968
10	0	0	0	0	0	0,01	0,02	0,02	0,0577	0,077	0,24	0,225	0,253	0,24142926
11	0	0	0	0	0	0	0	0,03	0,0128	0,115	0,099	0,178	0,189	0,31040905
12	0	0	0	0	0	0	0	0	0	0,013	0,033	0,062	0,211	0,20693937
h2(v)	0,025263	0,0726	0,093	0,104	0,094	0,08	0,08	0,08	0,0821	0,082	0,064	0,068	0,05	0,01526316

X FORMU	Vtoplam													
	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	0,206349	0,0985	0,076	0,039	0,024	0,01	0,02	0	0	0	0	0	0	0
1	0,253968	0,2576	0,184	0,088	0,057	0,03	0,03	0	0,0138	0	0	0	0	0
2	0,349206	0,2424	0,266	0,204	0,222	0,12	0,04	0,01	0	0,007	0	0	0	0
3	0,111111	0,1894	0,234	0,215	0,175	0,2	0,1	0,07	0,0276	0,014	0	0	0	0
4	0,031746	0,1439	0,108	0,215	0,175	0,2	0,12	0,17	0,0897	0,034	0,007	0	0	0
5	0,031746	0,0606	0,07	0,144	0,203	0,16	0,23	0,22	0,2069	0,103	0,014	0,016	0,011	0
6	0	0,0076	0,051	0,05	0,108	0,15	0,22	0,2	0,2138	0,151	0,122	0,024	0,011	0
7	0	0	0,006	0,022	0,028	0,06	0,11	0,18	0,1379	0,158	0,122	0,097	0,054	0,05263158
8	0	0	0,006	0,006	0,005	0,08	0,09	0,07	0,1448	0,253	0,209	0,202	0,098	0
9	0	0	0	0,011	0,005	0,01	0,03	0,06	0,0897	0,151	0,237	0,153	0,217	0,36842105
10	0,015873	0	0	0	0	0	0,02	0,02	0,0483	0,068	0,144	0,274	0,25	0,15789474
11	0	0	0	0,006	0	0	0,01	0	0,0276	0,055	0,115	0,177	0,272	0,15789474
12	0	0	0	0	0	0	0	0	0	0,007	0,029	0,056	0,087	0,26315789
h1(V)	0,032795	0,0687	0,082	0,094	0,11	0,09	0,1	0,08	0,0755	0,076	0,072	0,065	0,048	0,00989068

EK 22. Frekans Kestirim Yöntemi için Eşitleme Hata Değerleri

$w_1=0,5$ frekans kestirim

Equating Error 2.0

Input file: C:\Users\Ceren\Desktop\0,5.txt

Columns read:2 3 4

Type of equating function: Post-Smoothing

Smoothing parameter: 6.000000

Number of replications: 500

Seed: 285947671

Scale min, max, inc: 1.000000 12.000000 1.000000

Sample sizes (new, old): 1085.000000 1137.000000

0	0.16872145	0.00209889	0.00000000	0.35723486	-1.17622600	1.00000000
1	0.17677903	0.10634972	0.00000000	1.38185706	-0.93218129	1.00000000
2	0.16349599	0.26407841	0.00000000	2.36163188	0.04150998	1.00000000
3	0.15095544	0.15283651	0.47370877	3.34140671	1.42592086	1.34000000
4	0.13935797	0.13471026	0.41424630	4.32118153	2.39291146	2.22000000
5	0.12895826	0.23795457	0.46166655	5.30095635	3.60873555	3.69200000
6	0.12006797	0.10791319	0.00000000	6.28073117	5.15512826	5.00000000
7	0.11304376	0.11806016	0.00000000	7.26050599	6.06348495	6.00000000
8	0.10824953	0.11615945	0.00000000	8.24028081	7.09501053	7.00000000
9	0.10598832	0.09584428	0.00000000	9.22005563	8.10887720	8.00000000
10	0.10642170	0.07962378	0.00000000	10.19983045	8.96141141	9.00000000
11	0.10951769	0.20951227	0.09981984	11.17960527	9.90994726	9.99400000
12	0.11506156	0.22424256	0.25514702	12.15938009	11.81751459	11.93000000
Mean	0.12544617	0.15143785	0.15891377	6.95454180	5.54207707	5.65758157

w₁=1 için hata kestirim

Equating Error 2.0

Input file: C:\Users\Ceren\Desktop\1.txt

Columns read: 2 3 4

Type of equating function: Post-Smoothing

Smoothing parameter: 6.000000

Number of replications: 500

Seed: 285947671

Scale min, max, inc: 1.000000 12.000000 1.000000

Sample sizes (new, old): 1085.000000 1137.000000

0	0.16872145	0.04471793	0.00000000	0.35723486	-2.02866153	1.00000000
1	0.17677903	0.02423778	0.00000000	1.38185706	-2.14403275	1.00000000
2	0.16349599	0.39654799	0.00000000	2.36163188	-1.15845397	1.00000000
3	0.15095544	0.14299545	0.00000000	3.34140671	0.72535569	1.00000000
4	0.13935797	0.06803261	0.00000000	4.32118153	1.15797565	1.00000000
5	0.12895826	0.46600209	0.53872442	5.30095635	1.92524052	1.92400000
6	0.12006797	0.04075699	0.17058722	6.28073117	4.54718681	4.97000000
7	0.11304376	0.19528281	0.19595918	7.26050599	5.15817589	5.04000000
8	0.10824953	0.12308315	0.17600000	8.24028081	6.71800082	6.96800000
9	0.10598832	0.08919840	0.04467662	9.22005563	7.76110656	7.99800000
10	0.10642170	0.08828752	0.38584453	10.19983045	8.58045754	8.81800000
11	0.10951769	0.24310099	0.44741033	11.17960527	9.65276992	9.73200000
12	0.11506156	0.26030838	0.25848791	12.15938009	11.86708196	11.92800000
Mean	0.12544617	0.18047723	0.19951602	6.95454180	4.81085243	5.19606267

EK 23. Kalibrasyon Yöntemlerine İlişkin ST 2.0 Program Çıktısı

ST 2.0

Input file: C:\Users\Ceren\Desktop\st_for_pc_gui\st_wg_v20\st_ceren_pisa.txt

Number of common items: 13

Item Parameter Means

	a	b	c
New Form	2.3389	0.3285	0.1123
Old Form	2.3548	0.2876	0.1191

Item Parameter Standard Deviations

	a	b	c
New Form	0.7066	0.8829	0.0950
Old Form	0.6270	0.8701	0.1033

Transformation Functions

	Stocking-Lord	Haebara	Mean/Mean	Mean/Sigma
Intercept	0.009559	0.014415	-0.038712	-0.036151
Slope	0.956823	0.953890	0.993271	0.985476

EK 24. MTK Gerçek ve Gözlenen Puan Eşitleme Süreci

Results for True Score Equating

Sum of c parameters: 3.0620

Raw	Theta	True Score	Unrounded	Rounded
Score	Equivalent	Equivalent	Scale Score	Scale Score
0	-----	0.0000	0.0000	0.00
1	-----	0.9477	0.9477	1.00
2	-----	1.8955	1.8955	2.00
3	-----	2.8432	2.8432	3.00
4	-1.3124	3.7859	3.7859	4.00
5	-0.9112	4.4421	4.4421	4.00
6	-0.6470	5.2050	5.2050	5.00
7	-0.4390	6.1223	6.1223	6.00
8	-0.2673	7.1676	7.1676	7.00
9	-0.1206	8.2952	8.2952	8.00
10	0.0103	9.4658	9.4658	9.00
11	0.1310	10.6450	10.6450	11.00
12	0.2444	11.8017	11.8017	12.00
13	0.3516	12.9166	12.9166	13.00
14	0.4535	13.9877	13.9877	14.00
15	0.5513	15.0248	15.0248	15.00

16	0.6467	16.0439	16.0439	16.00
17	0.7418	17.0623	17.0623	17.00
18	0.8393	18.0946	18.0946	18.00
19	0.9431	19.1495	19.1495	19.00
20	1.0595	20.2319	20.2319	20.00
21	1.1992	21.3487	21.3487	21.00
22	1.3847	22.5154	22.5154	23.00
23	1.6807	23.6584	23.6584	24.00
24	2.1527	24.5061	24.5061	25.00
25	-----	25.0000	25.0000	25.00

Results for Observed Score Equating

Raw	Equated	Unrounded	Rounded
Score	Raw Score	Scale Score	Scale Score
0	0.0489	0.0489	0.00
1	0.9824	0.9824	1.00
2	1.8965	1.8965	2.00
3	2.7844	2.7844	3.00
4	3.6408	3.6408	4.00
5	4.4699	4.4699	4.00
6	5.3319	5.3319	5.00
7	6.2561	6.2561	6.00
8	7.2557	7.2557	7.00
9	8.3046	8.3046	8.00
10	9.3620	9.3620	9.00
11	10.4504	10.4504	10.00
12	11.6826	11.6826	12.00
13	12.8915	12.8915	13.00
14	13.9777	13.9777	14.00
15	15.0174	15.0174	15.00
16	16.0452	16.0452	16.00
17	17.0670	17.0670	17.00

18	18.0992	18.0992	18.00
19	19.1623	19.1623	19.00
20	20.2310	20.2310	20.00
21	21.3110	21.3110	21.00
22	22.5026	22.5026	23.00
23	23.6854	23.6854	24.00
24	24.4691	24.4691	24.00
25	25.2795	25.2795	25.00

Fitted observed score distributions

Raw

Score	New Form	Old Form
0	0.00232039	0.00211360
1	0.01226059	0.01313548
2	0.03091502	0.03730238
3	0.05111304	0.06506074
4	0.06464992	0.08041706
5	0.06870080	0.07764700
6	0.06532722	0.06573803
7	0.06019368	0.05713095

8	0.05821486	0.05422127
9	0.05695678	0.04991250
10	0.05027699	0.03997747
11	0.03948656	0.02967344
12	0.03386067	0.02747564
13	0.03847125	0.03501360
14	0.04614705	0.04396747
15	0.04505793	0.04375524
16	0.03324049	0.03307540
17	0.02134330	0.02159888
18	0.02112472	0.01982940
19	0.03314523	0.02897100
20	0.04304079	0.03872831
21	0.03749290	0.03680589
22	0.02978515	0.02661528
23	0.03417492	0.02604966
24	0.01653801	0.03235234
25	0.00573535	0.01300557



GAZİ GELECEKTİR..