



# HACETTEPE ÜNİVERSİTESİ

## EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

### ORTAK DEĞİŞKENLERİN ÖZELLİKLERİNE GÖRE KERNEL EŞİTLEME SONUÇLARININ KARŞILAŞTIRILMASI

Merve YILDIRIM SEHERYELİ

Doktora Tezi

Ankara, 2022

Liderlik, arařtırma, inovasyon, kaliteli eęitim ve deęiřim ile

*Daha ileriye ... En iyiye ...*



Eğitim Bilimleri Ana Bilim Dalı  
Eğitimde Ölçme ve Değerlendirme Programı

ORTAK DEĞİŞKENLERİN ÖZELLİKLERİNE GÖRE KERNEL EŞİTLEME SONUÇLARININ  
KARŞILAŞTIRILMASI

COMPARISON OF KERNEL EQUATING RESULTS ACCORDING TO PROPERTIES OF  
COVARIATES

Merve YILDIRIM SEHERYELİ

Doktora Tezi

Ankara, 2022

## Kabul ve Onay

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼ne,

Merve YILDIRIM SEHERYELİ'nin hazırladıđı “Ortak Deđiřkenlerin Özelliklerine G¼re Kernel Eřitleme Sonuçlarının Karřılařtırılması” bařlıklı bu çalıřma j¼rimiz tarafından **Eđitim Bilimleri Ana Bilim Dalı, Eđitimde Ölçme ve Deđerlendirme Bilim Dalında Doktora Tezi** olarak kabul edilmiřtir.

J¼ri Bařkanı	Prof. Dr. řener B¼Y¼K¼ZT¼RK	İmza
J¼ri Üyesi (Danıřman)	Prof. Dr. Selahattin GELBAL	İmza
J¼ri Üyesi	Prof. Dr. H¼lya KELECİOđLU	İmza
J¼ri Üyesi	Prof. Dr. řeref TAN	İmza
J¼ri Üyesi	Prof. Dr. Neře G¼LER	İmza

Bu tez Hacettepe Üniversitesi Lisans¼st¼ Eđitim, Öğretim ve Sınav Y¼netmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri ¼yeleri tarafından 29 / 07 / 2022 tarihinde uygun g¼r¼lm¼ř ve Enstit¼ Y¼netim Kurulunca ..... / ..... / ..... tarihi itibarıyla kabul edilmiřtir.

Prof. Dr. Selahattin GELBAL  
Eđitim Bilimleri Enstitüsü M¼d¼r¼

## Öz

Bu arařtırmada, “DOG-OD deseninde nasıl bir ortak deęiřken kullanmalıyız?” sorusuna yanıt aranmıřtır. Bu amaçla, Kernel eřitleme ile eřitlenmiř puan ortalamaları ve eřitleme hatalarının birer iřsel, ortak deęiřkenlerin özelliklerinin dıřsal, eřitlemedeki regresyon modelinde açıklanan varyans oranının hem iřsel hem dıřsal deęiřken olduęu yol modeli incelenmiřtir. Arařtırmada TIMSS 2019 matematik 8. sınıf Türkiye, Fransa, İsviçre ve Amerika veri setlerinde 3 ve 4. kitapçıklar (Ařama 1); 4. sınıf Kanada veri setinde 2, 3, 4, 5 ve 6. kitapçıklar (Ařama 2) kullanılmıřtır. Veriler IEA IDB Analiz Programı (v.4.0) ile düzenlenmiř analizlerin tamamı için RStudio (version 1.3.959) ara yüzü kullanılmıřtır. Kayıp veri ile başa çıkma yöntemi olarak hot-deck atama yapılmıřtır. DOG-OM deseni için KTK ve MTK'ye dayalı Kernel eřitleme, DOG-OD deseni için öncelikle ortak maddelerden alınan toplam puan kategorik hale getirilerek oluřturulan ortak deęiřkenler sonra Ařama 1'de 23, Ařama 2'de 36 kategorik ortak deęiřken kullanılmıřtır. Ařama 1'de 114, Ařama 2'de 126 eřitleme yapılmıř ve sonuçlar uç deęerler çıkarıldıktan sonra yol analizinde kullanılmıřtır. Arařtırmanın sonucunda DOG-OM deseninde KTK'deki sonuçlarda standart hatalar daha düşük, MTK'de eřitlenmiř puan ortalamaları daha yüksektir. Ortak maddelerin ortak deęiřken olarak kullanıldıęı DOG-OD desenindeki sonuçlar, özellikle ankor madde her iki ařamada da üç ya da beř kategorili olduęunda DOG-OM deseni ile benzerdir. Ařamalar birlikte düşünöldüęünde frekans daęılımları arasındaki farkın, test puanları ile ortak deęiřken arasındaki iliřkinin mutlak deęerinin; ortak deęiřkenin çarpıklıęının ve basıklıęının mutlak deęerinin, P ve Q evrenlerinde ortak deęiřken ile kurulan regresyon modelinin açıklanan varyans oranının anlamlı olarak eřitleme sonuçlarını etkiledięi; maddelerin ortak deęiřkene göre DMF içerip içermemesinin ise etkilemedięi görölmüřtür.

**Anahtar sözcükler:** test eřitleme, Kernel eřitleme, ortak deęiřken, ortak deęiřkene göre eřitleme, ortak deęiřken deseni

## Abstract

In this research, “What kind of covariate should we use in NEC design?” search for an answer to the question. For this purpose, the path model in which the mean score and equating errors equating with Kernel are internal, the covariates’ properties are external, and the rate of variance explained in the regression model in equating is both internal and external variable. In the research, TIMSS 2019 mathematics 8<sup>th</sup>-grade in Turkey, France, Switzerland and America datasets (Phase-1); 4<sup>th</sup>-grade in Canadian dataset (Phase-2). The RStudio interface was used for all analyzes whose data were arranged with the IEA IDB Analysis Program. Hot-deck imputation has been made for coping missing data. 114 equalizations were made in Phase 1 and 126 in Phase 2, and the results were used in the path analysis after removing the outliers. As a result of the research, standard errors were lower in the results of NEAT design in CTT, and the equating mean scores in IRT were higher. The results in NEC design, in which the common items are used as a covariate, are similar to NEAT design, especially when anchor item has three or five categories. When phases are considered together, the difference between the frequency distributions is the absolute value of the relationship between the test scores and the covariate; The absolute value of the skewness and kurtosis of the covariate and the explained variance proportion of the regression model established with the covariate in the P and Q populations significantly affect the equating results.

**Keywords:** test equating, Kernel equating, covariate, equating with covariate, covariate design

## Teşekkür

Tanıştığımız ilk günden itibaren tez sürecimde, çalışma hayatımda, zorlandığım her konuda omzumdaki elinin bana güç verdiği değerli danışmanım Prof. Dr. Selahattin Gelbal'a en içten teşekkürlerimi sunarım.

Ben yorulduğumda, daha fazlasını yapamayacağımı düşündüğümde bile bana inanan, hem önerileri hem de bakış açıları ile ufkumu açan, sorduğum her soruyu sabırla dinleyen ve yanıtlayan kıymetli öğretmenlerim ve TIK üyelerim Prof. Dr. Hülya Kelecioğlu ve Prof. Dr. Şeref Tan'a,

Her yeni günde mutluyken, hayat tarafından sınanırken, sabretmem gerekirken, bildiklerimi çevreme ve öğrencilerime anlatırken, büyüklerimle ve küçüklerimle sohbet ederken nasıl davranmam ve ne yapmam gerektiğini izleyerek, görerek öğrenme fırsatını bana tanıyan canım hocam Prof. Dr. Şener Büyüköztürk'e,

Aldığım eğitimler, katıldığım seminerler, yaptığım görevlerin her birinde beni koşulsuz destekleyen, öğrenmem için yeni fırsatlar veren, sabrına, sakinliğine ve çözümlerine hayranlıkla tanık olduğum hocam Doç. Dr. Ufuk Akbaş'a,

İyi niyeti, güler yüzü, desteği, sıcacık kalbi ile her zaman içimi ısıtan, ilgiyle dinleyip ilgiyle cevaplayan öğretmenim Prof. Dr. Neşe Güler'e,

Ben her hocam dediğimde, yardıma ihtiyaç duyduğumu fark etmediğimde bile bana elini uzatan, en kritik zamanlarda yaptığı yönlendirmeleri ile beni güvende hissettiren öğretmenim, akıl hocam Doç. Dr. Kübra Atalay Kabasakal'a,

Doktora eğitimim boyunca kendisinden aldığım derslerde, yaptığımız sohbetlerde, analizlerde yaşadığım her sorunda bana öğrettikleri ile yoluma devam edebildiğim kıymetli hocam Prof. Dr. Burcu Atar'a,

Hem tez önerisi hem de tez sürecimde sadece e-postalarımın yanıt vermekle kalmayıp bana en güzel yardımı değerli vaktini bolca ayırarak yapan hocam Doç. Dr. Okan Bulut'a,

Tez konumu ve sonrasında gideceğim yolu belirlememde beni yönlendiren hocalarım Arş. Gör. Halime Yıldırım Hoş, Dr. Öğr. Üyesi Çiğdem Akın Arıkan ve Dr. Öğr. Üyesi Meltem Yurtçu'ya teşekkür ederim.

Yol arkadaşım, en çok anlatanım en çok dinleyenim Arş. Gör. Sebahat Gören'e; yaşadığım her duygu değişiminde, iyi ve kötü anlarımda benden vazgeçmeyen Arş. Gör. Büşra Nur Durmaz, Arş. Gör. Mihriban Özcan ve Arş. Gör. Büşra Bilir'e en kalpten teşekkürlerimi iletiyorum.

Canım ailem; annem, babam, kardeşim ve eşim. Bana olan desteğiniz ve anlayışınız, varlığınızın bana verdiği azim beni daha da güçlendirecek. Her konuda yanımda olduğunuzu benimle yürüdüğünüzü bilmek, benim için en güvenli limandan bile daha değerli. Bana bu şansı sağladığınız için teşekkür ederim.



## İçindekiler

Kabul ve Onay .....	ii
Öz.....	iii
Abstract .....	iv
Teşekkür.....	v
İçindekiler .....	vii
Tablolar Dizini.....	x
Şekiller Dizini .....	xii
Simgeler ve Kısaltmalar Dizini .....	xiii
Bölüm 1 Giriş .....	1
Problem Durumu.....	1
Araştırmanın Amacı ve Önemi .....	8
Araştırma Problemi .....	9
Sayıtlılar.....	10
Sınırlılıklar.....	10
Tanımlar .....	11
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar .....	12
Araştırmanın Kuramsal Temeli.....	12
İlgili Araştırmalar .....	40
Bölüm 3 Yöntem .....	59
Araştırmanın Deseni .....	59
Araştırmanın Evreni ve Örnekleme .....	59
Veri Toplama Süreci .....	62
Verilerin Analizi .....	68
Bölüm 4 Bulgular ve Yorum .....	83
Aşama 1 .....	83

Aşama 2 .....	95
Bölüm 5 Sonuç, Tartışma ve Öneriler .....	107
Sonuç ve Tartışma.....	107
Öneriler.....	112
Kaynaklar.....	114
EK-A: TR veri setine ilişkin koşul incelemeleri ve DOG-OD desene ilişkin eşitleme sonuçları .....	cxxviii
EK-B: FC veri setine ilişkin koşul incelemeleri ve DOG-OD desene ilişkin eşitleme sonuçları .....	cxxix
EK-C: SW veri setine ilişkin koşul incelemeleri ve DOG-OD desene ilişkin eşitleme sonuçları .....	cxxx
EK-Ç: US veri setine ilişkin koşul incelemeleri ve DOG-OD desene ilişkin eşitleme sonuçları .....	cxxxı
EK-D: TR veri setine ilişkin DMF analizi sonuçları.....	cxxxii
EK-E: FC veri setine ilişkin DMF analizi sonuçları .....	cxxxiii
EK-F: SW veri setine ilişkin DMF analizi sonuçları .....	cxxxiv
EK-G: US veri setine ilişkin DMF analizi sonuçları .....	cxxxv
EK-H: 2 ile 3. Kitapçıktan elde edilen veri setine ilişkin koşul incelemeleri ve DOG-OD desene ilişkin eşitleme sonuçları.....	cxxxvi
EK-I: 4 ile 5. Kitapçıktan elde edilen veri setine ilişkin koşul incelemeleri ve DOG-OD desene ilişkin eşitleme sonuçları .....	cxxxviii
EK-İ: 5 ile 6. Kitapçıktan elde edilen veri setine ilişkin koşul incelemeleri ve DOG-OD desene ilişkin eşitleme sonuçları .....	cxl
EK-J: Aşama 1 için P ve Q evrenlerinde yol şemaları .....	cxlii
EK-K: Aşama 2 için P ve Q evrenlerinde yol şemaları.....	cxliii
EK-L: Aşama 1 için ham puanlar ile eşitlenmiş puanlar arasındaki fark puan dağılımları .....	cxliv

EK-M: Aşama 2 için ham puanlar ile eşitlenmiş puanlar arasındaki fark puan dağılımları .....	cxlix
EK-N: Araştırma Etik Komisyon İzin Muafiyeti Formu.....	clvii
EK-O: Etik Beyanı .....	clviii
EK-Ö: Yüksek Lisans/Doktora Tez Çalışması Orijinallik Raporu .....	clix
EK-P: Thesis/Dissertation Originality Report.....	clx
EK-R: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı .....	clxi

## Tablolar Dizini

<b>Tablo 1</b> Eşitleme desenleri .....	16
<b>Tablo 2</b> Eşitleme desenlerine göre eşitleme fonksiyonları .....	22
<b>Tablo 3</b> Veri setinde yer alan kişi sayıları .....	61
<b>Tablo 4</b> Birinci aşamada kullanılan ortak değişkenler .....	63
<b>Tablo 5</b> TR, FC, SW ve US veri setlerinde testten alınan toplam puanlar ile P ve Q evrenlerindeki değişkenler arasındaki korelasyon katsayıları .....	64
<b>Tablo 6</b> Madde gruplarının kitapçıklara dağılımı ve madde sayıları .....	65
<b>Tablo 7</b> İkinci aşamada kullanılan ortak değişkenler.....	66
<b>Tablo 8</b> Eşitlenen kitapçıklara göre testten alınan toplam puanlar ile P ve Q evrenlerindeki değişkenler arasındaki korelasyon katsayıları .....	67
<b>Tablo 9</b> P ve Q evrenlerindeki değişkenlerle yapılan yol analizine ilişkin model veri uyumu sonuçları .....	74
<b>Tablo 10</b> Veri setlerine ilişkin tek boyutluluk ve KR-20 değerleri.....	75
<b>Tablo 11</b> Veri setlerine ilişkin yerel bağımsızlık sonuçları .....	75
<b>Tablo 12</b> P ve Q evrenleri için uyum indeksleri .....	76
<b>Tablo 13</b> P ve Q evrenlerinde 2PL model için M2 istatistikleri.....	76
<b>Tablo 14</b> MTK'ye dayalı kestirilen madde güçlük ve ayırt edicilik parametreleri ile güvenilirlik katsayıları .....	77
<b>Tablo 15</b> KTK'ye dayalı elde edilen madde güçlük ve ayırt edicilikleri ile güvenilirlik katsayıları .....	78
<b>Tablo 16</b> Kitapçıklara göre veri setlerine ilişkin tek boyutluluk ve KR-20 değerleri .....	79
<b>Tablo 17</b> Veri setlerine ilişkin yerel bağımsızlık sonuçları .....	80
<b>Tablo 18</b> MTK uyum indeksleri .....	80
<b>Tablo 19</b> MTK'ye dayalı 2PL model için M2 istatistikleri .....	81
<b>Tablo 20</b> MTK'ye dayalı kestirilen madde güçlük ve ayırt edicilik parametreleri ile güvenilirlik katsayıları .....	81

<b>Tablo 21</b> KTK'ye dayalı elde edilen madde güçlük ve ayırt edicilikleri ile güvenilirlik katsayıları .....	82
<b>Tablo 22</b> DOG-OM deseninde KTK ve MTK Kernel eşitleme sonuçları .....	83
<b>Tablo 23</b> Ankor maddelerle oluşturulmuş ortak değişkenler için DOG-OD deseninde KTK Kernel eşitleme sonuçları.....	85
<b>Tablo 24</b> P evreninden elde edilen değişkenler için yol analizi regresyon katsayıları (Standartlaştırılmış) .....	91
<b>Tablo 25</b> Q evreninden elde edilen değişkenler için yol analizi regresyon katsayıları (Standartlaştırılmış) .....	93
<b>Tablo 26</b> DOG-OM deseninde KTK ve MTK Kernel eşitleme sonuçları .....	95
<b>Tablo 27</b> Ankor maddelerle oluşturulmuş ortak değişkenler için DOG-OD deseninde KTK Kernel eşitleme sonuçları.....	97
<b>Tablo 28</b> P evreninden elde edilen değişkenler için yol analizi regresyon katsayıları (Standartlaştırılmış) .....	102
<b>Tablo 29</b> Q evreninden elde edilen değişkenler için yol analizi regresyon katsayıları (Standartlaştırılmış) .....	104

## Şekiller Dizini

<b>Şekil 1</b> Madde Karakteristik Eğrisi ( $a=1, b=0, c=0$ ) .....	35
<b>Şekil 2</b> Yol analizinde kurulan model .....	73
<b>Şekil 3</b> KTK'ye göre eşitlenmiş puan dağılımları ve SEE'ler .....	84
<b>Şekil 4</b> MTK'ye göre eşitlenmiş puan dağılımları ve SEE'ler .....	84
<b>Şekil 5</b> DOG-OM ile ankor maddelerden elde edilen ortak değişkenlere göre DOG-OD desenlerin eşitleme fonksiyonları arasındaki hata bantları ve eşitleme hataları-TR verisi.	86
<b>Şekil 6</b> DOG-OM ile ankor maddelerden elde edilen ortak değişkenlere göre DOG-OD desenlerin eşitleme fonksiyonları arasındaki hata bantları ve eşitleme hataları-FC verisi.	87
<b>Şekil 7</b> DOG-OM ile ankor maddelerden elde edilen ortak değişkenlere göre DOG-OD desenlerin eşitleme fonksiyonları arasındaki hata bantları ve eşitleme hataları-SW verisi	88
<b>Şekil 8</b> DOG-OM ile ankor maddelerden elde edilen ortak değişkenlere göre DOG-OD desenlerin eşitleme fonksiyonları arasındaki hata bantları ve eşitleme hataları-US verisi	89
<b>Şekil 9</b> KTK'ye göre eşitlenmiş puan dağılımları ve SEE'ler .....	96
<b>Şekil 10</b> MTK'ye göre eşitlenmiş puan dağılımları ve SEE'ler .....	96
<b>Şekil 11</b> DOG-OM ile ankor maddelerden elde edilen ortak değişkenlere göre DOG-OD desenlerin eşitleme fonksiyonları arasındaki hata bantları ve eşitleme hataları-2 ile 3. kitapçık eşitleme verisi .....	99
<b>Şekil 12</b> DOG-OM ile ankor maddelerden elde edilen ortak değişkenlere göre DOG-OD desenlerin eşitleme fonksiyonları arasındaki hata bantları ve eşitleme hataları-4 ile 5. kitapçık eşitleme verisi .....	100
<b>Şekil 13</b> DOG-OM ile ankor maddelerden elde edilen ortak değişkenlere göre DOG-OD desenlerin eşitleme fonksiyonları arasındaki hata bantları ve eşitleme hataları-5 ile 6. kitapçık eşitleme verisi .....	101

## Simgeler ve Kısaltmalar Dizini

**AIC:** Akaike Bilgi Kriteri (Akaike Information Criterion)

**BIC:** Bayesian Bilgi Kriteri (Bayesian Information Criterion)

**EG:** Denk Grup

**DOG-OD:** Denk Olmayan Gruplarda Ortak Değişken

**DOG-OM:** Denk Olmayan Gruplarda Ortak Madde

**IEA:** Uluslararası Eğitim Başarısının Değerlendirilmesi Derneği (International Association for the Evaluation of Educational Achievement)

**RMSE:** Ortalama Hata Kareleri Kökü (Root Mean Squared Errors)

## Bölüm 1

### Giriş

Bu bölümde araştırmaya temel olan problem durumu, araştırmmanın amacı ve önemi, araştırmmanın problemi belirtilmiş, araştırmmanın sayıtları ve sınırlılıkları ile birlikte bazı tanımlar verilmiştir.

#### Problem Durumu

Bir meslek sahibi olmak ailelerin çocuklarından beklentisi olduğu gibi toplumun da kendi varlığını ve kurallarını sürdürmek için bireylerden beklentisidir. İyi bir meslek sahibi olmak için iyi bir üniversite, iyi bir lise, iyi bir ortaokul gibi eğitim sisteminin kademelerini tamamlamak gerekmektedir. Dünyanın birçok ülkesinde olduğu gibi ülkemizde de bu kademeler arasındaki geçişlerde sınavlar uygulanmaktadır. Bu sınavların sonuçları sayesinde hem okullara öğrenci seçimi yapılmakta hem de sistemin ve sistemde yer alan tüm elemanların kendisini geliştirmesi ve güncellemesi için geri bildirim mekanizmaları oluşturulmaktadır. Bu geri bildirimler; ülkelerin eğitim politikalarının, felsefelerinin gözden geçirilmesi, iller ve okullar arasındaki öğretim faaliyetlerinin incelenmesi, sınıf içi öğretim ve değerlendirme tekniklerinin güncellenmesi, öğrenciye ve veliye öğrencinin düzeyi hakkında bilgi verilmesi gibi birden fazla amaçla kullanılabilir (Erden, 1998).

Burs hakkı kazanma, görevde yükselme, meslek için belge alma, ordu, yükseköğretim, üniversite ya da ortaöğretim kurumlarına kabul gibi önemli kararlarda sınavlar kullanılmaktadır. Bu sınavların sonuçlarına göre öğrenciler birer birey olarak toplumda yer edinme yarışına girmiş olmaktadır. Adaylar, yalnız tek bir sınav hakkı ile “*ya çok şey kaybedebilir ya da çok şey kazanabilirler*” (Büyüköztürk, 2016). Yüksek performans beklentisi nedeni ile kendilerini bıkkın ve baskı altında hisseden öğrencilerin kaygı düzeyleri de artmaktadır. Yine bu sınav sonuçları öğrencilere, velilere, okullara birer ödül-ceza mekanizması olarak yansıtıldığında da istenmeyen sonuçlar yaşanmaktadır. Çıktıları bu



denli önemli olan sınavlar, yüksek-riskli sınavlar olarak adlandırılmaktadır (Kumandaş & Kutlu, 2015). Amerika'da SAT (Scholastic Assessment Test) ve ACT (American Collage Testing), İsveç'te SweSAT (Swedish Scholastic Assessment Test) gibi ulusal sınavlar geniş ölçekli yüksek-riskli sınavlardır. Ülkemizde ise ÖSYM (Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı) tarafından uygulanan YKS (Yükseköğretim Kurumları Sınavı), KPSS (Kamu Personel Seçme Sınavı), ALES (Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı), YDS (Yabancı Dil Bilgisi Seviye Tespit Sınavı), YÖKDİL (Yükseköğretim Kurumları Yabancı Dil Sınavı) ve MEB tarafından uygulanan liselere geçiş sistemi (LGS) sınavı bu sınavlar arasındadır. Sonuç değerlendirmeye yönelik bu sınavlar seçme işlemini öğrencileri puanlarına göre sıralayarak gerçekleştirmektedir. Bu durumda doğru ya da yanlış işaretledikleri ve hatta boş bıraktıkları bir soru bile öğrencilerin yerleşeceği okulu etkilemektedir. Bu sınavlara ek olarak Ekonomik İşbirliği ve Kalkınma Örgütü (OECD) tarafından uygulanan PISA (The Programme for International Student Assessment) ve Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu (IEA) tarafından uygulanan TIMSS (Trends in International Mathematics and Science Study) gibi öğrenci gelişimini izlemeye yönelik uluslararası sınavlar da geniş ölçekli testlerdir. Bu araştırmada da dört yılda bir uygulanan TIMSS 2019 matematik uygulamasına ilişkin veriler kullanılmıştır. TIMSS uygulamasında, dört ve sekizinci sınıf öğrencilerinin matematik ve fen bilimleri alanlarındaki performanslarının belirlenmesi için bilişsel testlerin yanı sıra katılan öğrencilere, velilerine, öğretmenlerine ve okul müdürleri için de anketler kullanılmaktadır. Bu sayede öğrencilerin gelişimlerinin yanı sıra ülkelerin eğitim sistemleri izlenebilmekte ve eğitim politikalarını yönlendirebilecekleri kanıtlar da elde edilebilmektedir (MEB, 2020; Mullis ve ark., 2020). Ülkemizde ise öğrencilerin üst düzey düşünme becerilerinin ölçülmesine yönelik MEB tarafından yapılan Akademik Becerilerin İzlenmesi ve Değerlendirilmesi (ABİDE) projesi ve okul öğrenmelerini izlemeye yönelik yapılan Öğrenci Başarısının Belirlenmesi Sınavı (ÖBBS) bu uluslararası sınavlar ile benzer amaçlar doğrultusunda uygulanmıştır. Geniş ölçekli testler, eğitim programlarının değerlendirilmesinde, öğretimin aksayan yanlarının belirlenmesinde ve giderilmesinde, öğrenci-öğretmen-okullar hakkında verilecek kararlarda

sıklıkla kullanılmaktadır (Çakan, 2003). Bu geniş ölçekli testlerin sonuçlarına göre alınacak önemli kararların isabetli olabilmesi ise geçerli, güvenilir ve adil ölçme süreçleri ile mümkün olmaktadır.

Ölçme süreçleri dolaylı ya da doğrudan olabilmektedir. Fiziksel bir büyüklükle temsil edilebilen, doğrudan gözlenebilen, iyi tanımlanmış birimleri ve bir ölçeği olan değişkenlerin ölçülmesinde geçerlik, güvenilirlik ve özellikle adalet sorunlarının olmadığı söylenebilir (Turgut & Baykul, 1992). Eğitimde ya da psikolojide ise yapılan tüm ölçmeler dolaylıdır (Morgan, 2013). Ölçeklerin araştırılan özelliği mümkün olan en az hata ile tam, doğru ve adil ölçebilmesi, karar vericilerin bireyler hakkında verdikleri yargıların doğruluğunu pozitif yönde etkiler. Ölçekler aracılığı ile bireylerde var olan özelliklerin, niteliklerin sayısallaştırılması, bireylerin ölçülen özelliğe sahip oluş derecelerinin (performanslarının) belirlenmesi sağlanmaya çalışılır (Crocker & Algina, 2006). Bireylerin performanslarındaki farklılıklar, bireylerin kişisel özelliklerinden, kullanılan testin özelliklerinden ya da bireylerin bu testin bağlamına karşı duyarlılıklarından kaynaklanabilmektedir. Bir ölçme aracının amaçlarından biri de bu farklılıkların yalnız bireysel farklılıklardan kaynaklanmasını sağlamak (madde etkisi), ölçümlerin amaçlanan kullanımı için geçerli yorumlar yapılabilmesine olanak sağlamaktır. Ölçülmek istenen yapıyla ilgisi olmayan özellikler nedeniyle bireylerin avantaj ya da dezavantajının olmadığı durumlarda ise bir geçerlik konusu olan “test adaleti” sağlanmış olur (American Educational Research Association, 2014). Fakat bir testin hem testi uygulayanlar hem de teste giren tüm gruplar için adil olduğunu söylemek pratikte çok mümkün değildir (Lord, 1980). Bu nedenle, değişen madde fonksiyonu, yanlılık, eşitleme ve bağlama çalışmalarının her biri test adaletini sağlamak amacıyla incelenen ölçme konularıdır. Bu araştırmada da eşitleme konusu ele alınmıştır.

American Educational Research Association (2014) tarafından yayınlanan “Eğitim ve Psikolojideki Testler için Standartlar” kitabında test eşitlemenin amacı, alternatif formlardan elde edilen test puanları arasındaki güçlük farkını düzelterek puanların birbiri yerine kullanılmasını sağlamak olarak belirtilmiştir. Alternatif formlar, bir testin aynı içerik ve

istatistiksel özelliklere göre oluşturulmuş, aynı koşullar altında uygulanmış farklı formlarıdır. Bu formlardan alınan puanların aynı ölçek üzerine yerleştirilerek formların birbiri yerine kullanılabilmesinin sağlanması süreci test eşitleme olarak tanımlanmıştır. Örneğin; yükseköğretime öğrenci kabul sürecini düşünelim. Farklı kabul koşulları olmakla birlikte genel olarak ALES, Yabancı dil puanı ve lisans not ortalaması ölçütleri ile aday öğrenciler başvuru yapabilmektedir. ALES ve dil puanları (e-YDS, YDS, YÖKDİL) beş yıl süre ile geçerlidir ("Lisansüstü Eğitim Öğretim Yönetmeliği," 2016, 04 20). Yılda iki kez ALES, dört kez e-YDS ve iki kez YÖKDİL yapılır (ÖSYM Sınav Takvimi, 2020). ALES için 10 farklı zamandaki 10 sınava, dil puanı için 30 farklı zamandaki 30 sınava giren farklı öğrencilerin puanları adaylık için geçerlidir. Bu tür tekrarlı yapılan sınavlarda adayların tartıştığı konulardan biri aynı özelliği ölçmek için hazırlanan ve aynı amaçla kullanılan sınavlardan birinin diğerinden daha kolay ya da zor olduğu yönündedir (Kolen & Brennan, 2014; Wu ve ark., 2016). O zaman, bu sınavlardan alınan puanlar karşılaştırılabilir mi? Biraz daha özelleştirirsek; diğer tüm ölçütlerden eşit puan alan iki adaydan biri zor sınavdan 89, diğeri kolay sınavdan 90 puan almış olsun. 90 puan alan adayın yükseköğretime kabul edilirken 89 puan alan adayın kabul edilmemesi doğru bir karar mıdır? Benzer bir durum A ve B gibi farklı test formları olan sınavlar için de geçerlidir. Aynı sorular farklı sırayla testte yer alsaydı bile bazı test formları öğrencilere hep daha kolay gelir: "Ben B grubu olsaydım daha yüksek puan alırdım." Böyle bir durumda zor testten ve kolay testten aynı puanı alan iki bireyin ölçülen özelliğe aynı derecede sahip olduğu söylenebilir mi? Bir test geliştiricisinin geliştirdiği iki test formundan aynı puanı alan iki kişinin ölçülen özelliğe sahip olma düzeylerinin aynı olduğunu söyleyebilmesi için testlerin psikometrik kanıtlarını yeterli düzeyde ortaya koyması gerekir (Kolen & Brennan, 2014).

Bu tartışmaların önüne geçmek için aynı test formu tekrarlı olarak kullanılabilir ya da birbirine çok benzeyen mümkünse paralel formlar uygulanabilir. Uygulamada birbirine tamamen paralel formlar oluşturmak neredeyse mümkün değildir. Üstelik aynı formların kullanıldığı farklı oturumlarda soruların adaylar arasında paylaşılması, aynı formun tekrarlı

uygulanmasında ise adayların hatırlama ya da aktarma etkileri geçerlik ve güvenilirlik sorunları olarak ortaya çıkmaktadır (Haladyna & Downing, 2004). Bu nedenle sınavda yan yana oturanlara, farklı günlerde ya da birden fazla kez sınava girenlere genellikle farklı test formlarının verilmesi istenir (Petersen ve ark., 1989). Sınavlardaki soruların halka arz edilmesi yine testin gelecekteki formunda başka maddelerin kullanılmasını gerektirir. Konu kapsamının eskimesi, bilginin güncellenmesi ve gelişmesi gibi nedenlerle soruların değiştirilmesi ya da yeni sorular eklenmesi istenebilir. Böyle durumlarda farklı uygulamalarda aynı soruların kullanılması daha da güçleşmektedir. Bu durumlarda aynı özelliği ölçmek için hazırlanan ve sonuçları aynı amaç için kullanılacak farklı test formlarından alınan puanların birbirinin yerine kullanılması gerekmektedir (González & Wiberg, 2017; Hambleton & Swaminathan, 1985). İşte bu puanları karşılaştırabilmek ve test formlarını birbiri yerine kullanılabilir duruma getirmek için izlenmesi gereken prosedürler bulunmaktadır. Bu süreç “Test Eşitleme” olarak ifade edilmektedir (von Davier, 2013). Yanıtladığı test formundan bağımsız olarak adayların ölçülen özelliğe sahip olma derecesinin olabildiğince en doğru şekilde belirlenmesi için testlerin eşitlenmesi gerekir (Tan, 2015).

Test uygulayıcılar tarafından farklı formlarının eşitlenmesi planlanan testlerde ilk önce eşitleme desenine karar vermek gerekmektedir. Çünkü kullanılacak maddeler, evren ve örneklem seçimi gibi konular da veri toplama sürecinde eşitleme desenleri ile birlikte belirlenmektedir. Ortak evrene dayalı desenler (Andersson & Wiberg, 2014) kullanılabilirliği gibi ortak madde ya da teste dayalı desenler (Akın Arıkan, 2017; Andersson & Wiberg, 2017; Kim ve ark., 2006; Kumlu, 2019; Moses ve ark., 2011; Moses ve ark., 2010; Özgürlük, 2019; Tanberkan Suna & Tan, 2017; Uysal, 2019; Wiberg & von Davier, 2017; Yıldırım Seheryeli ve ark., 2021) de literatürde sıklıkla kullanılmaktadır. Son yıllarda ise bu iki yöntemin dışında bir bakış açısı ile ortak değişkene dayalı yöntemler (Akın Arıkan, 2020; Altıntaş & Wallin, 2021; Bränberg & Wiberg, 2011; Sansivieri, 2017; Wiberg & Bränberg, 2015; Yurtçu, 2018) de incelenmeye başlanmıştır.

Ortak evren, ortak madde ya da ortak deęişkene dayalı yöntemlerden birinde verilerin toplanmasından önce her bir yöntemin sınırlılıkları göz önünde bulundurularak desen seçip eşitleme süreci planlanmalıdır. Ortak evrene dayalı desenler; tek grup, eş/denk/rasgele grup ve dengelenmiş grup desenleridir. Tek grup deseninde tek bir örnekleme her iki form da uygulandığı için katılımcıların yeteneklerindeki farklılık kontrol altına alınabilir. Fakat katılımcılar ilk testten etkilenecek diğer testi yanıtlamaktadır. Sıra etkisi nedeniyle testlerin güçlükleri arasındaki fark doğru belirlenemeyebilir (González & Wiberg, 2017). Rastgele gruplar deseninde seçkisiz olarak belirlenen ve ölçülen özellik bakımından birbiri ile aynı ya da denk olan gruplara formların uygulanması gerekmektedir. Bu nedenle geniş örneklemlere ihtiyaç duyulur. Bunun yanında test formlarının tamamen aynı özelliği ölçtüğüne ilişkin güçlü kanıtlar gereklidir (Kolen & Brennan, 2014). Dengelenmiş grup deseninde ise evrenden seçilen tek örneklem seçkisiz olarak ikiye ayrılır. İlk gruba önce verilen test ikinci gruba ikinci sırada verilerek her iki testi de tüm katılımcıların yanıtlaması sağlanır. Bu desen küçük örneklemlerde de iyi çalışsa da kullanılabilirliği düşüktür (American Educational Research Association, 2014). Ortak maddeye dayalı yöntemlerde ise gruplar arasında denklik aranmaz, ortak maddelerin kullanılması grupların ölçülen özellik açısından farklılıklarını yansıtmalarını sağlar. Her iki formda da grupların yanıtladığı ortak maddeler ya da testler yer alır. Burada ortak maddelerin testin tamamı ile aynı özelliği ölçmesi ve güçlük açısından da testi temsil etmesi beklenmektedir. Hatta her iki testte de ortak maddelerin aynı sıra ile aynı konumda bulunması gerekir (American Educational Research Association, 2014; González & Wiberg, 2017; Kolen & Brennan, 2014). Geniş ölçekli testlerde her yıl sınava girenler farklılaştığı için ortak evrene dayalı desenlerde grupların denk olmasını sağlamak neredeyse imkânsızdır. Denk olmayan gruplar için ortak maddeye dayalı desenlerin kullanılmasını zorlaştırıcı en büyük neden ise test güvenliğidir. Test formlarının gruplara tekrarlı verilmesi ya da aynı maddelerin kullanılması maddelerin katılımcılar arasında yayılmasına neden olacaktır. Bu nedenle her iki desen yerine grupların ölçülen özellik açısından yetenek farklılıklarını ayarlamak için test puanları üzerinde etkili

olduğu bilinen bir kovaryant değişkenin kullanılması önerilmiştir (Bränberg, 2010; Livingston & Lewis, 2009).

Desen seçiminin önemli olduğu kadar eşitleme yöntemi de eşitleme hatasının daha düşük kestirilmesi için önemlidir. Klasik Test Kuramı'na (KTK) dayalı klasik eşitleme yöntemleri özdeşlik, ortalama, doğrusal ve eşit yüzdelikli eşitleme olarak sınıflandırılırken Madde Tepki Kuramı'na (MTK) dayalı yöntemler gözlenen ve gerçek puan eşitleme olarak sınıflandırılmıştır (Kolen & Brennan, 2014). KTK'ye göre gerçek puan, bireyin ölçülen özellik açısından aldığı her testte değişmediği varsayılan ve hatalardan arınık olan puandır. Gözlenen puan ise bireyin testteki her bir maddeye verdiği yanıtların toplamı ile hesaplanan puandır. Gözlenen puan, gerçek puan ile tesadüfi hataların toplamıdır. Bu hata örneklem için kestirilir ve gruptaki herkes için aynı değerdedir (Baykul, 2010). MTK'de toplam puan yerine bireyin testteki maddelere verdiği yanıt örüntüleri ile ilgilenilir. Bir maddenin doğru yanıtlanma ihtimali, maddenin güçlük ve ayırt ediciliği ile yetenek parametresinin bir fonksiyonudur. Hata ise her bir madde ve yetenek düzeyi için ayrı kestirilebilir. İlgili maddenin güçlüğü yetenek düzeyine yaklaştıkça ve ayırt ediciliği arttıkça hata azalır (Hambleton ve ark., 1991). Yine MTK'ye dayalı yöntemlerin kullanılabilmesi için varsayımların da karşılanıp karşılanmadığı incelenmelidir. Bu varsayımlar karşılandığında KTK'de elde edilen istatistiklerin testin ve grubun özelliklerine bağlı olması sınırlılığı MTK'de kullanılan yöntemle bağlı olarak giderilebilmektedir. Çünkü MTK'de madde parametreleri gruptan, yetenek parametreleri ise maddelerin özelliklerinden bağımsız olarak kestirilebilmektedir (Baykul, 2010; Crocker & Algina, 2006). KTK'de bireyin yeteneği değişmemesine rağmen kolay bir testten yüksek puan alanların gerçek puanı da yüksektir (Hambleton & Jones, 1993). Burada MTK'ye dayalı eşitleme yöntemlerinden önce birlikte ya da ayrı (moment ya da karakteristik eğri yöntemleri ile) kalibrasyon yapılarak ölçek dönüştürme işlemi gerçekleştirilmektedir (De Ayala, 2009). Bu yöntemler kullanılırken KTK'ye dayalı yöntemlerde en az 400, MTK'ye dayalı yöntemlerde Rasch model için en az 400, 3PL model için en az 1500 kişinin tek bir formu almış olması gerekmektedir (Zhu,

1998). Bu yöntemlerin dışında hem KTK hem de MTK'ye dayalı olarak kullanılabilen Kernel eşitleme yöntemi de bulunmaktadır. Kernel eşitleme, kesikli puan dağılımlarını sürekli hâle dönüştüren bir eşit yüzdelikli gözlenen puan eşitleme yöntemidir. Bu yöntemde ön ya da son düzgünleştirme kullanarak veriler daha tutarlı hâle gelir, eşitleme sonucu daha küçük standart hatalar elde edilir ve bu hatalar örneklem değişkenliğinden daha az etkilenir (González & Wiberg, 2017).

### **Araştırmanın Amacı ve Önemi**

Bu araştırmada TIMSS 2019 matematik verileri kullanılarak denk olmayan gruplarda ortak değişken deseni ve ortak madde deseninde yapılan gözlenen puan Kernel eşitleme sonuçları ortak değişkenlerin özelliklerine göre karşılaştırılmıştır. Ortak değişkenin özellikleri olarak gözlenen test puanları ile ortak değişken arasındaki ilişkilerin, ortak değişkene göre maddelerin DMF içerip içermeme durumlarının, ortak değişkenin P ve Q evrenlerindeki frekans dağılımları arasındaki farkın, ortak değişkenlerin basıklık ve çarpıklıklarının, ortak değişkenle yapılan eşitlemede kullanılan regresyon modelinin açıkladığı varyans oranı ele alınmıştır.

Ülkemizde ortak evrene, ortak maddeye dayalı birçok çalışma (Akın Arıkan, 2020; Ertoprak, 2017; Karagül, 2020; Kumlu, 2019; Özgürlük, 2019; Tanberkan Suna & Tan, 2017; Uysal, 2019) yapılmış olsa da Kernel eşitleme (Akın Arıkan, 2017, 2020; Altıntaş & Wallin, 2021; Bağcı ve ark., 2016 Ekim; Özsoy, 2021) ve ortak değişkene göre yapılan eşitleme (Akın Arıkan, 2020; Altıntaş & Wallin, 2021; Yurtçu, 2018) çalışmalarına çok az rastlanmaktadır.

Bu çalışma ile henüz hiçbir ulusal sınavımızda kullanılmayan test eşitlemenin önündeki test güvenliği engellerinin ortak değişken kullanılarak giderilebileceği düşünülmektedir. Gruplar arasındaki yetenek farklılıklarını ortak değişken sayesinde giderme ve Kernel eşitleme ile daha küçük standart hatalar elde edilerek eşitleme yapma, ölçme ve değerlendirme uygulamalarının geniş ölçekli testlerde de kullanılması avantaj

sağlayacaktır. Bu çalışmada “Nasıl bir ortak değişken kullanılmalı?” sorusuna yanıt aranacağından bu çalışmanın ortak madde bulunmayan testlerin eşitlenmesinde kullanılacak değişkenleri seçmede ön bilgi teşkil edeceği düşünülmektedir. Dolayısıyla bu çalışma geniş ölçekli sınavlarda test adaletinin daha doğru sağlanarak kişiler hakkında verilen kararların geçerliğinin artırılmasına sağlayacağı katkı açısından önemlidir.

### **Araştırma Problemi**

Denk olmayan gruplarda ortak değişken (DOG-OD) deseninde ortak değişkenin özellikleri KTK'ye dayalı gözlenen puan Kernel eşitleme sonucu elde edilen eşitleme hatasını, eşitlenmiş ve ham puanlar arasındaki fark puanların dağılıma ilişkin hataları, beklenen hata ortalamalarını ve eşitleme doğruluğunu nasıl etkilemektedir? Bu problem incelenmeden önce ise denk olmayan gruplarda ortak madde (DOG-OM) deseninde KTK ve MTK'ye dayalı eşitleme sonuçları incelenmiştir. Bu sonuçlar ile denk olmayan gruplarda ortak değişken (DOG-OD) desenindeki sonuçların karşılaştırılabilmesi için ortak maddeler de kategorik hâle dönüştürülerek ortak değişken olarak kullanılmıştır. Bu dönüştürme işleminin, gerçekte kategorik olan ortak değişkenler ile sürekli değişken olan ortak maddelerle yapılan eşitleme sonuçlarını incelemeyi kolaylaştıracağı düşünülmüştür.

**Alt problemler.** Bu çalışma kapsamında aşağıdaki sorulara yanıt aranacaktır.

1. DOG-OM desende KTK ve MTK'ye dayalı gözlenen puan Kernel eşitleme yapıldığında eşitlenmiş puan ortalamaları ( $eqY_x$ ), eşitleme hataları (SEE), yüzde görelî hatalar (PRE), ham puanlar ile eşitlenmiş puanlar arasındaki fark puanlar ile elde edilen ortalama hatalar (RMSE) ve eşitleme doğruluğu (RMSD) ne düzeydedir?
2. KTK'ye dayalı DOG-OM desen ile ortak maddelerin birer ortak değişken olarak modele dahil edildiği DOG-OD desende yapılan Kernel eşitleme sonuçları ( $eqY_x$ , SEE, PRE, RMSE, RMSD) ve eşitleme fonksiyonları arasındaki farklara ilişkin hatalar (SEED) nasıl dağılmaktadır?
3. DOG-OD desende KTK'ye dayalı gözlenen puan Kernel eşitleme yapıldığında;



- a) ortak deęişkenin P ve Q evrenlerindeki frekans daęılımları arasındaki fark (D);
- b) gözlenen test puanları ile ortak deęişken arasındaki ilişkinin mutlak deęeri ( $absPR$  ve  $absQR$ );
- c) ortak deęişkenin çarpıklık ve basıklığının mutlak deęeri ( $absPÇarpıklık$ ,  $absPBasıklık$ ,  $absQÇarpıklık$ ,  $absQBasıklık$ );
- d) testteki maddelerin P ve Q evrenlerinde ortak deęişkene göre DMF gösterip göstermemesi (DMF)
- e) P ve Q evrenlerinde ortak deęişken ile kurulan modelin açıklanan varyans oranı ( $PRkare, QRkare$ );

eşitleme sonuçlarını ve hatalarını ( $eqYx$ , SEE, PRE, RMSE, RMSD) nasıl etkilemektedir?

### Sayıtlılar

- Ankor maddeleri P ve Q evrenlerindeki bireylerin çözme sırasının dięer bir deyişle sıra etkisinin eşitleme sonuçlarını etkilemedięi varsayılmıştır.
- Veri setlerinde madde takımı (testlet) şeklindeki maddelerin %50'sini doğru yanıtlayanlar 1, dięerleri 0 olarak kodlanmıştır. Dolayısı ile madde takımı etkisi göz ardı edilmiştir.
- Veri setlerinde kısmi puanlanan maddelerde %50'sini doğru yanıtlayanlar 1, dięerleri 0 olarak kodlanmıştır.

### Sınırlılıklar

- Araştırma kullanılan gerçek veri setinin ilgili kitapçıkları (Aşama 1 için 8. Sınıf matematik Türkiye, Fransa, İsviçre ve Amerika veri setlerinin 3 ve 4. kitapçıkları; aşama 2 için 4. Sınıf matematik Kanada veri setinin 2, 3, 4, 5 ve 6. kitapçıklar) ve veri setindeki deęişkenler ile sınırlıdır.
- DOG-OD deseninde Kernel eşitlemede hiyerarşik veri yapısı için bir çözüm bulunmamaktadır. Bu nedenle TIMSS 2019 uygulamasındaki öğrenci, sınıf ve okul ağırlıkları eşitlemeye dahil edilmemiştir.

- Araştırma sonuçlarında tüm parametreleri aynı anda kontrol edebilmek için simülasyon çalışması yapılmak istenmiştir. Değişkenlerin niteliksel özellikleri dikkate alındığında ortak değişkene göre denk olmayan gruplar için eşitleme verisi üretmek uygun görülmemiştir.
- DOG-OD deseninde Kernel eşitlemede ortak değişkenler yalnız dış ankor olarak kullanılabilir. Bu nedenle DOG-OM deseninde de ortak maddeler dış ankor olarak ele alınmıştır.
- Araştırmanın ikinci aşamasında kullanılan veri setinde B ve C düzeyinde DMF içeren maddeler bulunmamaktadır. Bu nedenle DMF koşulu aşama 2'de incelenememiştir.

### **Tanımlar**

- Ankor test: Denk olmayan gruplar deseninde her iki popülasyonun da ortak yanıtladığı maddelerden oluşan test.
- Ortak değişken: Denk olmayan gruplarda ortak değişken deseninde eşitleme yaparken her iki popülasyonun da geçmiş bilgilerine ilişkin değişken (Background variable).

## Bölüm 2

### Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Bu bölümde araştırmanın kuramsal temellerinden bahsedilmiş, ulusal ve uluslararası literatürde yer alan çalışmalara değinilmiştir.

#### Araştırmanın Kuramsal Temeli

##### ***Bağlama (Linking)***

Kişiler hakkında doğru ve adil kararlar verebilmek için kullanılan ölçme aracının biriminin değerlendirmeye alınan her bir kişi için aynı olması gerekmektedir. Kişiler aynı testi aldığıında, kişilerin puanlarını karşılaştırmada herhangi bir problem yoktur. Fakat test güvenliği, testi farklı zamanda uygulama vb. nedenlerle herkesin aynı testi alamadığı durumlarda puanların doğrudan karşılaştırılabilir olmasını sağlamak için “linking (bağlama)” prosedürlerinin kullanılması gerekmektedir (Kolen, 2006). Holland ve Dorans (2006) göre puanlar arasında dönüşüm yapmayı sağlayan “bağlama”nın üç farklı yolu vardır: Yordama, ölçekleme ve eşitleme. X ve Y birer değişken olmak üzere; Y puanları, X puanlarının doğrusal bir fonksiyonu olsun. Yordama, aynı evrendeki kişilerin X puanlarını kullanarak Y puanlarını tahmin etme sürecidir. X puanları bir testten alınan puanlar olabileceği gibi demografik özellikler de olabilmektedir (Holland, 2007). Ölçekleme, X ve Y puanlarının karşılaştırılabilir olması için ortak bir ölçeğe dönüştürme sürecidir. Bu nedenle ölçekleme, bağlamanın dolaylı bir yoludur (Holland & Dorans, 2006). Eşitleme ise, aynı yapıyı ölçen ve benzer güçlük ve güvenilirlikteki farklı test formlarından alınan puanların birbiri yerine kullanılabilmesini sağlama sürecidir (Dorans ve ark., 2010). Test formlarının içeriği, güçlüğü ve testi alan bireylerin yetenek dağılımları benzerlik gösteriyorsa yatay eşitleme, kasıtlı olarak zorluk bakımından farklı tasarlanmasına rağmen aynı genel bilgi ve beceri alanını ölçmeyi amaçlayan testlerde dikey ölçekleme kullanılmaktadır (Zhu, 1998). Yatay eşitleme, test güvenliği nedeniyle aynı genel içeriği ölçen farklı formlar kullanıldığında; dikey ölçekleme ise ya öğretmen yapımı testlerde ya da boylamsal izleme çalışmalarında aynı

genel içeriğin sınıflar/yıllar arasında test edilmesini içerir (Hambleton & Swaminathan, 1985). Bu çalışma kapsamında test eşitleme süreci ile yatay eşitleme kastedilmektedir.

### ***Test Eşitleme (Equating)***

Farklı formları kullanılan testler, içerik olarak aynı olsa da zorluk açısından farklı olabilmektedir. Eşitleme, zorluk ve içerik açısından benzer oluşturulmuş test formlarından alınan puanlar arasındaki, testlerin zorluk düzeylerinden kaynaklanan farklılıkları ayarlamak için kullanılan bir süreçtir (Kolen & Brennan, 2014). Eşitleme, ölçekleme yapılacak test formlarının katı gereklilikleri olan bir versiyonu olarak düşünülebilir. Bu gereklilikler test formlarının aynı yapıyı ölçmesi, neredeyse aynı güçlük ve güvenilirlik düzeyinde olmalarıdır. İdealde iki formun birbirine tamamen paralel olması beklenmektedir, dolayısı ile bu durumda eşitleme gerekli değildir. Uygulamada ise paralel formlar oluşturmak neredeyse imkânsızdır (Dorans ve ark., 2010). Bu nedenle ham puanların genellikle alternatif formlar arasında tutarlı bir anlamı bulunmamaktadır. Lineer ya da lineer olmayan dönüşümler kullanılarak anlamlı yorumlar yapılabilir. Bir eşitleme sürecinin adımları aşağıdaki gibi verilmiştir (Kolen & Brennan, 2014).

- “1. Eşitlemenin amacına karar verme.
2. Alternatif formlar oluşturma.
3. Veri toplama için bir tasarım seçme. Eşitleme, test formlarının istatistiksel olarak nasıl farklılaştığına dair bilgi sağlamak için verilerin toplanmasını gerektirir.
4. Veri toplama tasarımını uygulama. Test yönetilir ve veriler tasarımın belirlediği şekilde toplanır.
5. Bir veya daha fazla operasyonel denklem tanımı seçme. Eşitleme, formlar arasındaki ne tür ilişkilerin tahmin edileceği konusunda bir seçim yapılmasını gerektirir.
6. Bir veya daha fazla istatistiksel tahmin yöntemi seçme.
7. Eşitlemenin sonuçlarını değerlendirme.” (s.7.)

**2.1.2.1. Eşitleme Koşulları.** Bir test eşitleme çalışması yapabilmek için testlerin eşitleme koşullarını sağlaması gerekmektedir. Bu koşullardan bazıları bireylerin puanlarına odaklanırken bazıları da puan dağılımlarına odaklanmaktadır. İdealde bir bireyin aldığı puanın yanıtladığı test formundan bağımsız olması beklenir ve bu beklenti puan dağılımları için de geçerlidir. Literatürde eşitleme koşullarının farklı sınıflandırmaları (Angoff, 1971; Dorans & Holland, 2000; Hambleton & Swaminathan, 1985; Kolen, 1988; Kolen & Brennan, 2014; Lord, 1980; Petersen ve ark., 1989) bulunsa da hepsinde genel anlamda bahsedilen koşullar aşağıdaki gibidir:

*Aynı yapı/Benzer içerik.* Test formlarının eşitlenebilmesi için aynı psikolojik özelliği (yapıyı) ölçmesi gereklidir. İçerik farklı olduğunda yapı da farklılaşacağından eşitleme yapılamaz (Dorans & Holland, 2000).

*Simetri.* Test eşitleme yapılırken formlardan biri referans diğeri temel test olarak isimlendirilir. Simetriklik, test formlarından hangisi referans ya da temel test olarak kullanılırsa kullanılsın sonuçların değişmemesidir (Lord, 1980).

*Eşitlik.* Eğer ölçmenin hatası puanlara dahil olmasaydı kişinin alacağı puan, gerçek puanını ifade ederdi. Lord'un (1980) eşitlik özelliği her gerçek puan (T) için eşitleme fonksiyonunun koşullu frekans dağılımlarının Y'nin koşullu frekans dağılımları ile aynı olmasıdır. Bu dağılımların varyansı sifıra eşit olduğu sürece ölçümlerin güvenilir olduğu söylenebilir. Eğer testlerden elde edilen ölçümler güvenilir değilse eşitleme yapılamaz. Sınava girenlerin popülasyonu için Form Y'deki puanların kümülatif dağılımı G, Form X'teki puanları Form Y ölçeğine dönüştüren fonksiyon  $eq(x)$ , aynı popülasyon için  $eq(x)$ 'in kümülatif dağılımı  $G^*$  olsun. Eşitlik özelliği aşağıdaki gibi tanımlanır:

$$\forall T \text{ için } G^*[eq(x)|T] = G(Y|T)$$

Bu eşitlik eğer X ile Y aynı formlar ise sağlanır. Pratikte ise bu mümkün değildir. Bu nedenle daha az katı olan birinci sıra eşitlik (BSE) tanımlanmıştır. Burada belirli bir gerçek puana sahip olan adaylar Form X'te, Form Y'den aldıkları ile aynı ortalama dönüştürülmüş

puana sahiptir. Daha az katı olan ikinci sıra eşitlik özelliği (İSE) ise eşitlemeden sonra koşullu standart ölçme hatalarının alternatif formlar için benzer olduğu ölçüde sağlanmaktadır (Kolen & Brennan, 2014; Zhu, 1998).

Gözlenen puan eşitlemede ise ortalama eşitleme özelliğinde iki formdaki dönüştürülmüş puanlar aynı ortalamaya, lineer eşitlemede iki formdaki dönüştürülmüş puanlar aynı ortalama ve standart sapmaya sahip olmalıdır. Eşit yüzdelli eşitlemede lineer ve ortalama eşitleme de geçerli olmalıdır. Lineer eşitleme özelliği korunduğunda ise ortalama eşitleme özelliği de geçerli olmalıdır (Lord, 1980).

*Grup değişmezliği.* Eşitleme sonuçlarının uygulama yapılan gruptan bağımsız olmasıdır. Eşitlemeden elde edilen dönüşümün benzer koşullarda da uygulanabilmesi istenir (Hambleton ve ark., 1991; Zhu, 1998).

**2.1.2.2. Eşitleme Desenleri.** Test formlarının eşitlenmesi planlandığında hem pratikte hem de istatistiksel olarak avantajları değişen, veri toplama desenlerinden birini seçmek gerekmektedir. Bu eşitleme desenleri random gruplar, tek grup, dengelenmiş tek grup, denk olmayan gruplarda ortak madde ve denk olmayan gruplarda ortak değişken deseni olarak isimlendirilmektedir. González ve Wiberg (2017) eşitleme desenlerini veri toplama desenleri olarak isimlendirmiş ve aşağıdaki gibi özetlemiştir.

Tablo 1 incelendiğinde her popülasyonda iki grup ya da iki farklı popülasyondan veriler toplanmış gibi görünse de herhangi bir sayıda popülasyondaki herhangi bir sayıda test katılımcısı grubundan veri elde etmek mümkündür. Benzer şekilde, teoride test formu, ankor test (ortak maddelerden oluşan test) ve ortak değişken sayıları da tabloda belirtilenden fazla olabilir. Aşağıda bu desenler ayrıntılı açıklanmıştır.

**Tablo 1** Eşitleme desenleri

Desen	Grup	Hedef Popülasyon				P popülasyonu				Q popülasyonu			
		X	Y	A	C	X	Y	A	C	X	Y	A	C
Tek Grup	G	✓	✓										
Random (Denk) Grup	G1	✓											
	G2		✓										
Dengelenmiş tek grup	G1	✓	✓										
	G2	✓	✓										
Denk olmayan gruplarda ortak madde	G1					✓		✓					
	G2									✓		✓	
Denk olmayan gruplarda ortak değişken	G1					✓		✓					
	G2									✓			✓

C: Ortak değişken; A: Ortak test; X, Y: Test formları

*Tek grup deseni (TG)*. Tek bir popülasyondaki tek bir gruba eşitlenmesi planlanan test formlarının ikisi de uygulanır. Aynı grup her iki testi de yanıtladığından testlerin zorlukları katılımcıların yeteneklerinden bağımsızdır. Bu nedenle de en düşük eşitleme hatasına sahip desendir (Livingston, 2014). Küçük örneklerde kullanışlıdır. Bu desenin ilk dezavantajı formlardan birinin önce uygulanması nedeniyle daha sonra uygulanan formda katılımcıların yorgun ya da formdaki sorulara aşına olması nedeniyle diğer testin daha zor/kolay olarak algılanması söz konusudur. Buna ek olarak katılımcıların her iki formu da yanıtlaması gerektiği için uygulama ya da toplam test süresi uzundur (American Educational Research Association, 2014; González & Wiberg, 2017).

*Random gruplar deseni (RG)*. Aynı popülasyondan bağımsız iki grup random olarak seçilir. Her grup yalnız tek bir test formunu yanıtlar. Gruplar random oluşturulduğu için her iki grubun ölçülen özellik açısından denk olduğu varsayılır. Bu varsayım sayesinde testler arasındaki güçlük farkı, katılımcıların performansları arasındaki fark ile açıklanabilir. Bu desende, gruplardan biri testi farklı zamanda aldığı anda katılımcıların olgunlaşma ihtimali grupların denk olması varsayımını tehlikeye düşürür (González & Wiberg, 2017). Bu

varsayımın incelenerek eşitleme sürecine devam edilmesi puanların geçerliğini tehlikeye atmamak açısından önemlidir.

*Dengelenmiş tek grup deseni (TG-D)*. Tek bir popülasyondan seçilen örneklem bağımsız iki grup oluşturacak şekilde bölünür. Eşitlenmesi planlanan test formlarının ikisi de uygulanır fakat sırası farklıdır. İlk formu birinci grup önce yanıtlarken ikinci formu diğer grup önce yanıtlar (Kolen & Brennan, 2014). Küçük örneklerde kullanışlıdır. Tek grup deseninden farklı olarak sıra etkisinin de dikkate alınabilmesi mümkündür fakat yine katılımcıların her iki formu da yanıtlaması gerektiği için uygulama ya da toplam test süresi uzundur (González & Wiberg, 2017).

*Denk olmayan gruplarda ortak madde deseni (DOG-OM)*. Bu desen, aynı popülasyonda yer almayan kişilere ya da farklı test formlarının aynı kişiye uygulanamadığı durumlarda kullanılmaktadır. Farklı popülasyonlardaki bağımsız iki gruba test formlarından yalnız biri ve ek olarak her iki gruba da ortak bir test uygulanır. Bu maddelere/teste ankor denir. Ortak testin eşitlenecek olan test ile benzer ortalama ve madde gücüne sahip, içerik olarak da eşitlenecek testi temsil eden bir yapıda olması, eşitlenecek test ile arasındaki korelasyonun yüksek düzeyde olması önerilmektedir (Dorans ve ark., 2010; Wiberg & von Davier, 2017). Bu testten alınan puanlar katılımcıların puanlarına dâhil edildiğinde ortak teste iç ankor, dâhil edilmeyerek sadece eşitleme için kullanıldığında dış ankor adı verilir. Bu desenin ilk dezavantajı ortak maddeler katılımcılara daha önce uygulandığında yanıtları hatırlanabilir ya da katılımcılar maddelerin kullanılış amacını bilerek bu maddeleri önemsemeden yanıtlayabilir. Buna ek olarak standartlaştırılmış ya da geniş ölçekli testlerin tasarımında ortak maddelerin yer almaması da bir sorundur. Ortak testte maddelerin sırasının farklı olması durumunda sıra etkisi tekrar gündeme gelecektir (González & Wiberg, 2017; Kolen & Brennan, 2014). Diğer bir sorun da gruplar arasındaki farklılık ne kadar fazla ise kullanılan istatistiksel yöntemlerin grup ile form farklılıklarını ayırması o kadar zorlaşır (Dorans ve ark., 2010; Wiberg & von Davier, 2017).



*Denk olmayan gruplarda ortak deęişken deseni (DOG-OD).* Eşitleme sürecinde ortak bir test olmadığına ya da ortak testin özellikleri eşitleme yapmaya uygun olmadığına katılımcılara ilişkin ortak bir deęişkenin (vekil ya da kovaryant deęişken) kullanılması ortak madde desenine önemli bir alternatiftir. Bu desenele ortak madde yerine ortak deęişken kullanılarak eşitlemedeki grup farklılıkları açıklanmaya çalışılır. Gruplar arasındaki sistematik farklılıkları en iyi açıklayan ve test ile yüksek ilişkili ortak deęişkenlerin seçilmesi önerilmektedir. Böylelikle grupların yetenek farkı kontrol altına alınabilir. Ayrıca ortak deęişkenin kategorilerindeki birey sayısının ya da dağılımlarının yaklaşık olarak aynı olması da önemlidir çünkü grup başarısı yerine katılımcıların ortak deęişken ile ölçülen özelliğe sahip olması ön planda olacaktır (Wiberg & Bränberg, 2015). Ortak deęişkenler doğru seçildiğine eşitleme hatası azalarak eşitlemenin doğruluęu artmaktadır. Eşitlenecek gruplara göre yanlılık içeren bir deęişken seçildiğinde ise deęerlendirmenin yanlı olması sebebiyle test adaleti sağlanamayacak ve puanların geçerlięi düşecektir (Bränberg & Wiberg, 2011; Wiberg & Bränberg, 2015).

**2.1.2.3. Eşitleme Hatası.** Her ölçme işleminde olduğu gibi eşitleme süreci de rastgele ve sistematik hatalar içermektedir.

*Rastgele hata.* Eşitlemeyi tasarlama ve yürütmedeki temel amaç eşitlemedeki tahminlerin hatalarını azaltmaktır. Eşitleme sürecinde test formları katılımcılara rastgele atandığından eşitlenmiş puanların kestiriminde de örneklemeden kaynaklanan hataların ortaya çıkması kaçınılmazdır. Bu hata örnekleme hatası, tesadüfi hata ya da eşitlemenin standart hatası (SEE: Standard error of equating) olarak isimlendirilmektedir. Kavramsal olarak rastgele hata tekrarlı eşitleme süreçlerinde eşitlenmiş puanların standart sapmasıdır. Örnekleme büyüklüęü artırılarak ya da bu hatayı elimine eden desenler seçilerek eşitlemenin standart hatası azaltılabilir hatta bu ikisi birlikte yapılarak rastgele hatanın önemsiz hâle gelmesi sağlanabilir (González & Wiberg, 2017; Zhu, 1998).

*Sistematik hata.* Eşitleme yönteminin seçiminden, eşitleme varsayımlarının ve koşulların ihlalinden kaynaklanan hatadır. Gruplar büyük ölçüde farklılık gösterdiğinde,

ortak maddeler içerik ve istatistiksel özellikler açısından test formunu temsil etmediğinde ya da bu maddeler bir uygulamadan diğerine farklı çalıştığında sistematik hatalar meydana gelir (Kolen & Brennan, 2014; Zhu, 1998).

Zaman ilerledikçe eşitleme ve ölçekleme sürecine çok sayıda form dâhil edildiğinde hem rastgele hem de sistematik hatalar birikmektedir. Rastgele hatalar örneklem büyüklüğü ile kontrol edilebilirken sistematik hataları kontrol etmek oldukça güçtür. Eşitleme süreçleri bu hatalar en aza indirilecek şekilde planlansa da bazı durumlarda çok yüksek hatalarla kestirimler yapılabilir. Böyle durumlarda eşitleme her zaman savunulabilen bir süreç olmamaktadır (Kolen & Brennan, 2014).

Geleneksel (KTK'ye dayalı) eşitleme yöntemlerinde doğrusal eşitleme için genellikle form başına en az 400 ve eşit yüzdellik eşitleme için en az 1500 kişiden veri toplanması önerilmektedir. MTK'ye dayalı yöntemlerde ise genellikle Rasch model için 400 ve 3PL model için en az 1500 örneklem büyüklüğü gereklidir (Zhu, 1998). Örneklem küçük olduğunda ise log-lineer düzgünleştirme yöntemlerinin kullanılması önerilmektedir (Livingston, 1993). Kolen ve Brennan (2014) örneklem büyüklüğünü belirlemek için eşitleme hata varyansı ve puan dağılımlarına dayanan formüller ile tek grup ve random gruplar için farklı formüller önermiştir. Genel olarak doğrusal eşitlemede 1000 örneklem büyüklüğü ile elde edilen ortalama kesinlik, eşit yüzdellikli eşitlemede 1500'ün üzerinde örneklem büyüklüğüne ihtiyaç duymaktadır. Bu durum doğrusal eşitlemenin daha iyi olduğu anlamına gelmemektedir. İlişki doğrusal olmadığından yanlış modelin kullanımından kaynaklanacak sistematik hata nedeniyle, eşit yüzdellikli eşitlemedeki standart hata daha yüksek olsa bile popülasyonun daha doğru bir tahminini sağlayabilir.

**2.1.2.4. Eşitleme Dönüşümlerinde Tahmin Ediciler.** Olasılık modeli gibi veri üretme (kestirim) mekanizmaları  $\theta$  parametresi ile tamamen karakterize edildiğinde bu modele parametrik model denmektedir. Burada  $\theta$ 'nın elemanı olduğu parametre uzayı sonlu boyutludur. Bu boyut hem sonlu hem de sonsuz parametrelerin karışımı olduğunda model yarı-parametrik olarak isimlendirilmektedir. Başka bir durumda da olasılık dağılımını

karakterize eden sonlu dağılımlar olmadığında yalnız parametrelerin kendileri ile ilgilenilir. O zaman da model parametrik olmayan model olarak isimlendirilir (González & Wiberg, 2017). Bu modeller farklı tahmin ediciler kullanıldığında farklı süreklileştirme (continuization) yöntemleri üretir. Aşağıda bu tahmin edicilerden bazıları verilmiştir.

*Eşit yüzdellikli eşitleme fonksiyonu.* Gözlenen puanların  $x_1, x_2, \dots, x_{n_x} \sim F_X$  ve  $y_1, y_2, \dots, y_{n_y} \sim F_Y$  dönüşümleri için parametrik olmayan ampirik bir dağılım fonksiyonudur.  $X$  ve  $Y$ 'nin kümülatif dağılım fonksiyonları  $F_X$  ve  $F_Y$ 'dir.

$$\varphi(x) = F_Y^{-1}(F_X(x)) \text{ eşitliği... (2.1) ile tanımlanır.}$$

Bu tahmin ediciyi kullanırken dağılımların kesikli olması problemi devam ettiğinden süreklileştirme için doğrusal interpolasyon kullanılmaktadır (González & Wiberg, 2017).

Eşit yüzdellikli eşitlemede gözlenen en yüksek puan ve en düşük puan için eşitleme ilişkisi elde edilir. Bu puanların dışında ham puanı olan birinin puanı eşitlenmek istediğinde bir dönüşüm yapmak mümkün olmamaktadır, bu sınırlılık düzgülleştirme yöntemleri kullanılarak kısmen giderilebilmektedir (Livingston, 2014).

*Doğrusal eşitleme fonksiyonu.* Ortalama, doğrusal, Tucker ve Levine eşitleme dahil olmak üzere doğrusal bir eşitleme fonksiyonu kullanır. Bu tahmin edicide  $X$  ve  $Y$ 'nin ortalama ve standart sapmaları ile parametrik konum ölçeği puan dağılımları kullanılmaktadır.  $H$  bir dağılım fonksiyonu olmak üzere,

$$F_X(x) = H\left(\frac{x - \mu_X}{\sigma_X}\right) \text{ ve } F_Y(y) = H\left(\frac{y - \mu_Y}{\sigma_Y}\right) \text{ olarak tanımlanırsa}$$

$$\varphi(x; \theta) = F_Y^{-1}(F_X(x; \mu_X, \sigma_X); \mu_Y, \sigma_Y) = \mu_Y + \frac{\sigma_Y}{\sigma_X}(x - \mu_X) \text{ eşitliği ... (2.2) elde edilir.}$$

Burada  $\theta$  parametresi tamamen parametrik olan bir eşitleme tahmin edicisinde verileri üreten puan dağılımlarını ifade etmektedir (González & Wiberg, 2017).

Doğrusal eşitlemede, eşitlenmiş puanlar sınava giren gruba bağlı olduğu gibi aynı zamanda düşük ve yüksek ham puanların eşitlenmiş puanları, ölçeğin sınır puanlarının dışında olabilmesi bir sınırlılıktır. 50 soruluk bir testten 52 almanın oldukça zor açıklanacağı gerçeği

bulunmaktadır (Livingston, 2014). Böyle bir durumda ya bu puanların sınırların dışında kalması kabul edilebilir ya da en yüksek ve en düşük puanlar eşitleme sürecine dahil edilmeyebilir (Kolen & Brennan, 2014).

*Kernel eşitleme fonksiyonu.* Bu fonksiyon hem sonlu parametrelerin hem de dağılım fonksiyonlarının  $\varphi$  tahmininde yer aldığı yarı-parametrik bir tahmin edicidir.  $r_j = Pr (X = x_j)$  ve  $s_k = Pr (Y = y_k)$  olmak üzere,

$$\varphi(x; \theta) = F_{h_Y}^{-1}(F_{h_X}(x; r), s) \text{ eşitliği ... (2.3) elde edilir.}$$

Burada  $r$  ve  $s$ ,  $r_j$  ve  $s_k$  elemanlarından oluşan birer vektör,  $h_X$  ve  $h_Y$  düzgünleştirme miktarını kontrol eden bant genişliği parametreleridir (González & Wiberg, 2017).

Bir TG-D deseninde, eşit yüzdellikli ve doğrusal olmak üzere iki eşitleme fonksiyonu kullanılmaktadır. Kernel eşitleme bant genişlikleri optimal düzeyde olduğunda eşit yüzdellikli eşitleme fonksiyonuna, geniş bant genişliği kullanıldığında ise doğrusal eşitlemeye yaklaşır (von Davier ve ark., 2006). DOG-OM deseninde ise ankor maddeler tarafından sağlanan bilgiyi kullanmanın üç yolu vardır. Ankor puanlar eşitlenecek puanları tahmin etmek için kullanıldığında son tabakalama eşitleme (PSE); simetrik bağlantı ilişkilerinden yararlanan zincir eşitleme (CE); eşitlenecek formlar ile ankordan elde edilen gerçek puanlar arasındaki tahmini ilişkilere dayanan Levine gözlenen puan doğrusal eşitleme (Kolen & Brennan, 2014). Bu yöntemlerin Kernel eşitlemedeki karşılıkları aşağıdaki Tablo 2’de özetlenmiştir.

Tablo 2’de görüldüğü gibi Levine gözlenen puan doğrusal eşitlemeye yaklaşan bir Kernel eşitleme karşılığı yoktur. Bu araştırmada R programındaki kequate paketi (Andersson ve ark., 2022) kullanılarak Kernel eşitleme için optimal bant genişliği ile son tabakalama eşitleme kullanılmıştır.

**Tablo 2** Eşitleme desenlerine göre eşitleme fonksiyonları

Desen	Eşitleme yöntemi	Kernel Eşitleme Türü
TG-D	Klasik eşit yüzdelikli	KE optimal bant genişliği ile
	Klasik doğrusal	KE geniş bant genişliği ile
DOG-OM	Zincir eşit yüzdelikli	KE zincir, optimal bant genişliği ile
	Frekans kestirimi eşit yüzdelikli	KE son tabakalama, optimal bant genişliği ile
	Zincir doğrusal	KE zincir, geniş bant genişliği ile
	Tucker	Doğrudan karşılığı olmasa da KE son tabakalama geniş bant genişliği ile)
	Braun-Holland doğrusal	KE son tabakalama geniş bant genişliği ile
	Levine gözlenen puan doğrusal	-

**2.1.2.5. Eşitleme Yöntemleri.** González ve Wiberg (2017) eşitleme yöntemlerini Geleneksel, Madde Tepki Kuramı'na dayalı ve Kernel eşitleme yöntemleri şeklinde sınıflandırmıştır.

**Geleneksel eşitleme yöntemleri.** Geleneksel eşitleme yöntemleri, Klasik Test Kuramı'nı temel alan yöntemlerdir. Eşit yüzdelikli, doğrusal ve ortalama eşitleme yöntemleri eşitleme desenleri için açıklanmıştır.

TG, RG, TG-D desenleri için eşitleme. Eşit yüzdelikli eşitlemede Y formundaki y puanları ile aynı yüzdellik dereceye sahip X formundaki x puanları tanımlanarak eşitleme işlemi yapılır. X ve Y'nin kümülatif dağılım fonksiyonlarının sürekli rastgele değişkenler olduğu varsayılarak 2.1 eşitliği tanımlanır. Tüm eşitleme dönüşümleri eşitlik 2.1'in özel durumlarıdır. Doğrusal eşitleme test formları arasındaki zorluk farklılıklarının tamamen ilk iki moment (ortalama ve standart sapma) ile tanımlanabileceği varsayımına dayanır (Eşitlik 2.2). Ortalama eşitleme yöntemi ise X ve Y puan dağılımları için konum aileleri kullanıldığında ya da her iki konum-ölçek dağılımlarının varyanslarının eşit olduğu varsayıldığında kullanılır. X ve Y formlarının yalnız ortalamaları açısından farklılaştığı varsayımına dayanır. Eşitlik 2.2'de varyanslar eşit kabul edilirse

$$\varphi(x; \theta) = X + \mu_Y - \mu_X \text{ eşitliği ...}(2.4) \text{ elde edilir.}$$

Eşitleme desenlerinden ortak evrene dayalı TG, RG ve TG-D desenlerinde çıkarımlar yapmak için özel varsayımlara ihtiyaç yoktur. Tüm veriler tek bir popülasyondan elde edildiği için her şey doğrudan gözlemlenebilir ve tahmin edilebilir. Fakat DOG desenlerinde örneklemeler farklı popülasyonlara aittir. Bu desende hedef popülasyon T, P ve Q popülasyonlarının ağırlıklandırılması ile elde edilen sentetik bir popülasyon oluşturularak tanımlanır.  $w_P$  ve  $w_Q$  popülasyonların ağırlıkları,  $w_P, w_Q \geq 0$  ve  $w_P + w_Q = 1$  olmak üzere;

$$T = w_P P + w_Q Q \text{ eşitliği ... (2.5) tanımlanır.}$$

$f_{XP}$  ve  $f_{XQ}$ , P ve Q evrenlerinden elde edilen ve T üzerinde tanımlanan X puan dağılımları,  $f_{YP}$  ve  $f_{YQ}$ , P ve Q evrenlerinden elde edilen ve T üzerinde tanımlanan Y puan dağılımları olmak üzere;

$$f_{XT}(x) = w_P f_{XP}(x) + w_Q f_{XQ}(x) \text{ ve}$$

$$f_{YT}(y) = w_P f_{YP}(y) + w_Q f_{YQ}(y) \text{ eşitlikleri ... (2.6) elde edilir.}$$

Burada X formunun P, Y formunun da Q evrenlerindeki örneklemelere uygulanması nedeniyle 2.6 eşitliğini doğrudan elde etmek mümkün değildir. Bu nedenle hem P hem de Q popülasyonunda bir A ankor testi kullanılması gerekmektedir.

DOG-OM deseni için eşitleme. Bir DOG tasarımında kullanılacak iki yaklaşım vardır: Zincirleme eşitleme (ZE) ve frekans kestirimi eşitleme (FKE). FKE'de eşitleme dönüşümünün oluşturulduğu T evrenindeki X ve Y puan dağılımlarını elde etmek için sentetik evren kullanılır. ZE yaklaşımında X ve Y formları A formuna eşitlenir ve sentetik bir evren oluşturulmaz. Zincirleme eşitleme için  $\varphi_A(x)$ , X puanlarını A'ya ve  $\varphi_Y(a)$ , A puanlarını Y'ye dönüştüren fonksiyonu göstermek üzere, X puanlarını Y'ye dönüştüren eşitleme fonksiyonu

$$\varphi(x) = \varphi_Y(x) = \varphi_Y(\varphi_A(x)) \text{ ... (2.7) şeklindedir.}$$

DOG-OM deseni için doğrusal eşitlemede öncelikle T evrenindeki X ve Y puan dağılımları için ortalama ve standart sapma değerlerine ihtiyaç vardır. Sentetik evrene göre parametreler aşağıdaki gibidir:

$$\mu_{XT} = w_P \mu_{XP} + w_Q \mu_{XQ} \dots (2.8)$$

$$\mu_{YT} = w_P \mu_{YP} + w_Q \mu_{YQ} \dots (2.9)$$

$$\sigma^2_{XT} = w_P \sigma^2_{XP} + w_Q \sigma^2_{XQ} + w_P w_Q [\mu_{XP} - \mu_{XQ}]^2 \dots (2.10)$$

$$\sigma^2_{YT} = w_P \sigma^2_{YP} + w_Q \sigma^2_{YQ} + w_P w_Q [\mu_{YP} - \mu_{YQ}]^2 \dots (2.11)$$

Bu eşitliklerin hiçbirinde X ve Y formlarını aynı evrendeki kişiler almadığından parametreler doğrudan tahmin edilemez. Aşağıdaki yöntemler farklı varsayımlarla bu parametreleri tahmin etmek için kullanılırlar.

Tucker eşitlemede iki varsayım bulunmaktadır. Birincisi, X'in A üzerindeki koşullu olasılığının beklenen değeri ve Y'nin A üzerindeki koşullu olasılığının beklenen değeri her iki evrede de aynıdır. İkincisi bu koşullu olasılıkların varyansları da her iki evrende aynıdır (Gulliksen, 2013). Doğrusal regresyon modellerinde olduğu gibi beklenen değerlerin doğrusal, varyansların da sabit olduğu varsayılır. Bu nedenle de eğim, kesişim ve varyanslar her iki regresyonda da eşit olmaya zorlanabilir. P ve Q evrenlerinde  $\delta$  ve  $\gamma$  eğim ve kesişim katsayıları olmak üzere;

$$\delta_P = \mu_{XP} - \gamma_P \mu_{AP}; \quad \delta_Q = \mu_{XQ} - \gamma_Q \mu_{AQ}; \quad \gamma_P = \frac{\sigma_{XP,AP}}{\sigma_{AP}^2} \quad \text{ve} \quad \gamma_Q = \frac{\sigma_{XQ,AQ}}{\sigma_{AQ}^2} \quad \dots (2.12) \text{ eşitliği}$$

tanımlanır. Sentetik evrende eğim ve kesişim katsayılarının eşit olduğu varsayımı denkleme eklendiğine ise aşağıdaki eşitlikler elde edilir.

$$\mu_{XT} = \mu_{XP} - w_Q \gamma_P [\mu_{AP} - \mu_{AQ}] \dots (2.13)$$

$$\mu_{YT} = \mu_{YQ} - w_P \gamma_Q [\mu_{AP} - \mu_{AQ}] \dots (2.14)$$

$$\sigma^2_{XT} = \sigma^2_{XP} - w_Q \gamma_P^2 [\sigma^2_{AP} - \sigma^2_{AQ}] + w_P w_Q \gamma_P^2 [\mu_{AP} - \mu_{AQ}]^2 \dots (2.15)$$

$$\sigma^2_{YT} = \sigma^2_{YQ} - w_P \gamma_Q^2 [\sigma^2_{AP} - \sigma^2_{AQ}] + w_P w_Q \gamma_Q^2 [\mu_{AP} - \mu_{AQ}]^2 \dots (2.16)$$

Nominal ağırlıklı eşitleme, Tucker eşitlemenin basitleştirilmiş ve daha küçük örneklem için kullanılan hâlidir (Albano, 2016). Ortak test formunun X ve Y formları ile P ve Q evrenlerinde mükemmel ilişkili olduğu ve test formlarının istatistiksel özelliklerinin benzer olduğu varsayımına dayanır. Eşitlik 2.12 yerine  $J_X$ ,  $J_Y$  ve  $J_A$  madde sayılarını göstermek üzere,

$$\gamma_P = \frac{J_X}{J_A} \text{ ve } \gamma_Q = \frac{J_Y}{J_A} \text{ eşitliği kullanılır.}$$

Levine gözlenen puan eşitleme varsayımları Klasik Test Kuramına dayanmaktadır. Burada gözlemlenen X, Y ve A (iç ankor) formlarından elde edilen toplam puanların her birinin gerçek puan ve hata puanlarının toplamından oluştuğu kabul edilir. Her ne kadar uygulamada gözlenen puanlar kullanılsa da varsayımlar gerçek puan türünden ifade edilmektedir. (i) X ve Y formlarındaki gerçek puanlar ile A formundaki gerçek puanlar arasındaki ilişki birdir. (ii) X ve Y formlarındaki gerçek puanların A üzerindeki regresyon katsayıları eşittir. (iii) Evrenlerdeki ölçme hata varyansı X, Y ve A formları için aynıdır. Bu varsayımları kullanarak eğimlerin gerçek puanların standart sapmaları cinsinden tanımlamak mümkündür.

$$\gamma_P = \frac{\sigma_{XP}^2}{\sigma_{XP,AP}} \text{ ve } \gamma_Q = \frac{\sigma_{YQ}^2}{\sigma_{YQ,AQ}} \dots(2.17)$$

Burada öncelikle eşitlik 2.17, eşitlik 2.13, 2.14, 2.15 ve 2.16'da yerine yazılır. Daha sonra elde edilen ortalama ve varyanslar eşitlik 2.2 genel doğrusal eşitleme denkleminde yerine konur.

Levine gerçek puan eşitleme yine Levine gözlenen puan eşitleme ile aynı varsayımlara sahiptir. Gözlenen puanlar yerine gerçek puanlar eşitlenmektedir. Eşitlik 2.17'deki eğim parametreleri tekrar kullanılarak aşağıdaki eşitliğe ulaşılır (Kolen & Brennan, 2014).

$$\varphi(\tau_X; \mu_{XP}, \mu_{YQ}, \mu_{AP}, \mu_{AQ}, \gamma_P, \gamma_Q) = \frac{\gamma_Q}{\gamma_P} [\tau_X - \mu_{XP}] + \mu_{YQ} + \gamma_Q [\mu_{AP} - \mu_{AQ}] \dots(2.18)$$



Zincirleme doğrusal eşitlemede P ve Q evrenlerindeki test puanları ve ankor test puanlarının ortalamaları ve standart sapmaları hesaplanarak önce X puanlarının A puanlarına daha sonra A puanlarının Y puanlarına bağlanmasını sağlayan doğrusal dönüşümler kullanılır (Eşitlik 2.7). Son aşamada ise aşağıdaki eşitlik elde edilir.

$$\varphi(x) = \mu_{YQ} + \frac{\sigma_{YQ}}{\sigma_{AQ}} [\mu_{AP} - \mu_{AQ}] - \frac{\sigma_{YQ}/\sigma_{AQ}}{\sigma_{XP}/\sigma_{AP}} \mu_{XP} + \frac{\sigma_{YQ}/\sigma_{AQ}}{\sigma_{XP}/\sigma_{AP}} x \dots (2.19)$$

DOG-OM deseni için eşit yüzdelikli eşitlemede frekans kestirimi ve zincirleme eşit yüzdelikli eşitleme kullanılmaktadır.

Frekans kestirimi eşitlemede hedef evren T'de X ve Y puanlarını tahmin etmek için A ankor testi kullanılır. Elde edilen puan dağılımlarından yüzdeler sıralar bulunur ve eşitleme işlemi gerçekleştirilir (Braun & Holland, 1982).  $f_{XP}(x|a)$ , P evreninde X puanlarının A üzerinde koşullu frekans dağılımını gösterebilir. Benzer şekilde  $f_{YP}(y|a)$ ,  $f_{XQ}(x|a)$ ,  $f_{YQ}(y|a)$  da koşullu frekans dağılımlarıdır.  $f_{AP}(a)$  ve  $f_{AQ}(a)$  da ankor test puanlarının marjinal dağılımları olsun. Eşitlik 2.5'te verilen T hedef popülasyonu için (i)  $f_{XP}(x|a) = f_{XQ}(x|a)$  ve (ii)  $f_{YP}(y|a) = f_{YQ}(y|a)$  varsayımları kullanıldığında aşağıdaki eşitlikler elde edilir.

$$f_{XT}(x) = w_P f_{XP}(x) + w_Q \sum_a f_{XP}(x|a) f_{AQ}(a) \dots (2.20)$$

$$f_{YT}(y) = w_Q f_{YQ}(y) + w_P \sum_a f_{YQ}(y|a) f_{AP}(a) \dots (2.21)$$

Zincirleme eşit yüzdelikli eşitlemede, zincirleme eşitlemedeki gibi önce X formu A'ya daha sonra A formu Y'ye bağlanır. Zincirleme eşit yüzdelikli eşitleme dönüşümü (i)  $\varphi_{AP}(x) = \varphi_{AQ}(x) = \varphi_A(x)$  ve (ii)  $\varphi_{YP}(a) = \varphi_{YQ}(a) = \varphi_Y(a)$  varsayımları altında aşağıdaki gibi tanımlanır.

$$\varphi_Y(x) = F_{YQ}^{-1} \left( F_{AQ} \left( F_{AP}^{-1} (F_{XP}(x)) \right) \right) = \varphi_{YQ}(\varphi_{AP}(x)) \dots (2.22)$$

Braun-Holland eşitleme yöntemi doğrusal ve eşit yüzdelikli eşitlemenin birlikte kullanıldığı, frekans tahmin yönteminin doğrusal bir uyarlamasıdır (Braun & Holland, 1982). Eşitlik 2.20'deki  $f_{XT}(x)$  ve  $f_{YT}(y)$ , sentetik hedef evrenin ortalama ve standart sapma kestirimine dayanmaktadır. Eşitlik 2.2'de aşağıda tanımlanan ortalama ve standart

sapmalar yerine yazıldığında Braun-Holland doğrusal eşitleme dönüşümü elde edilmektedir.

$$\mu_{XT} = \sum_x x f_{XT}(x) \text{ ve } \mu_{YT} = \sum_y y f_{YT}(y) \dots(2.23)$$

$$\sigma_{XT}^2 = \sum_x [x - \mu_{XT}]^2 f_{XT}(x) \text{ ve } \sigma_{YT}^2 = \sum_y [y - \mu_{YT}]^2 f_{YT}(y) \dots(2.24)$$

Geleneksel test eşitleme yöntemlerinin karşılaştırılması.

Her yöntem için farklılaşan hata hesaplamalarında, sistematik hatalar genel olarak yanlılık olarak incelenirken, rastgele hatalar için bootstrap standart hatalar incelenir.

*Eşitlemenin bootstrap standart hatası.* Bir eşitleme prosedürünün sınava girenlerin popülasyonundan örneklemeler kullanılarak çok kez tekrarlandığını varsayalım. Bu durumda her bir puanın karşılığı olabilecek çok sayıda eşitlenmiş puan elde edilebilir. Her bir puan için elde edilen standart hatalar eşitlenmiş puanların tekrarları üzerinden standart sapmasına eşittir (Kolen & Brennan, 2014). Fakat pratikte genellikle tek bir veri setine sahip olduğu düşünülürse bu ideal duruma ulaşmak mümkün değildir. Efron ve Tibshirani (1993) göre bu sorun eldeki verilerden çok sayıda rastgele örneklemeler çekilerek bir istatistiğin standart hatasının kestirilebileceği bootstrap yöntemidir. Bu yöntemde  $l$  kez,  $X$  ve  $Y$  formlarını alan kişilere ait veri seti arasından belirlenen bir büyüklükte bir örneklem çekilir ve eşitleme yapılır. Bu yöntemde eşitleme hatası aşağıdaki eşitlik ile hesaplanır.

$$SE(x_i) = \sqrt{\frac{1}{L} \sum_{l=1}^L [\hat{\phi}_l(x_i) - \bar{\phi}_l(x_i)]^2} \dots (2.25)$$

Bu eşitlikte  $\bar{\phi}_l(x_i) = \frac{1}{L} \sum_{l=1}^L \hat{\phi}_l(x_i)$  ve  $\hat{\phi}_l(x_i)$  fonksiyonu repkilyasyonun  $l$ . adımındaki eşitlenmiş fonksiyonları göstermektedir. Parametrik bootstrap yöntemi için izlenmesi gereken adımlar aşağıdaki gibidir (González & Wiberg, 2017).

1.  $X$  ve  $Y$ 'nin ampirik dağılımlarına uyan bir log-lineer model kullanılır.
2.  $X$  ve  $Y$  test puanlarından belirlenen büyüklükte rastgele örneklemeler çekilir. Elde edilen dağılımlar  $X$  ve  $Y$ 'nin parametrik bootstrap örneklem dağılımlarıdır.

3. Elde edilen dağılımlar kullanılarak eşitleme dönüşümleri kestirilir.

4. 2 ve 3. adımlar  $l$  kez tekrarlanır.

5. Eşitlik 2.25 kullanılarak her bir puanın standart hatası kestirilir.

**Yanlılık ve RMSE.** Standart hatada olduğu gibi yanlılıkta da her bir puan için bir yanlılık (BIAS) değeri aşağıdaki eşitlik ile hesaplanır.  $L$  tekrar ya da örneklem sayısını göstermek üzere,

$$bias(x_i) = \frac{1}{L} \sum_{l=1}^L [\hat{\varphi}_l(x_i) - \varphi(x_i)] \dots (2.26)$$

Eşitlik 2.25 ve 2.26 kullanılarak da RMSE değeri hesaplanır.

$$RMSE(x_i) = \sqrt{bias(x_i)^2 + SE(x_i)^2} \dots (2.27)$$

**Kernel Eşitleme.** Kernel eşitleme yöntemi 1980 yıllarının başında Holland ve Thayer tarafından sunulmuş (Holland ve ark., 1989; Holland & Thayer, 1989), test eşitlemede log-linear modellerde ön-düzdüğüleştirme yaparak standart eşitleme hataları üretmesi konusunda üstünlük göstermiştir. Tek grup, denk grup, son-düzdüğüleştirme ile denk olmayan gruplarda ortak madde desenleri de ilk sunulan desenler olmuştur. Dengelenmiş gruplar, zincir eşitleme ile denk olmayan gruplarda ortak madde desenleri daha sonra (2000'li yılların başında); denk olmayan gruplarda ortak değişken deseni ise özellikle 2010 yılından sonra çalışılmıştır. Kernel eşitlemede ayırık puanların sürekli yaklaşımları elde edilir. Bu tür yaklaşımlar; gerçekte ayırık puanlanan rastgele değişkenin, Kernel'i karakterize eden sürekli bir rastgele değişkenin, süreklilik için düzdüğüleştirme derecesini kontrol eden bir parametrenin fonksiyonu olan sürekli bir rastgele değişken üretir. Puanların dönüşümünde 2.3. eşitliğinin genişletilmiş hâli kullanılır.

$$\hat{\varphi}(x; r, s) = F_{h_Y}^{-1}(F_{h_X}(x; \hat{r}), \hat{s}) = \widehat{F}_{h_Y}^{-1}(\widehat{F}_{h_X}(x)) \dots (2.28)$$

Burada  $\hat{r}$  ve  $\hat{s}$ ,  $r_j$  ve  $s_k$  elemanlarından oluşan birer vektör,  $h_X$  ve  $h_Y$  düzdüğüleştirme miktarını kontrol eden bant genişliği parametreleridir (González & Wiberg, 2017). Kernel eşitleme yöntemi beş adımdan oluşan bir süreç olarak tanımlanmıştır.

1. Puan olasılıklarının ön-düzgünleştirilmesi. Hem daha kararlı eşitleme fonksiyonları üretmek hem de varyansı azaltmak için gözlenen puanlara parametrik bir dönüşüm uygulanır. Her bireyin toplam puanını hesaplamak ve bu puan olasılıklarını düzgünleştirmek için log-lineer modeller kullanılır (Holland ve ark., 1989; Holland & Thayer, 1989, 2000). Bunun dışında her bir maddeye verilen yanıt örüntülerini kullanarak bir MTK modeli kullanıldığında da düzgünleştirilmiş puan olasılıkları elde edilir. Bu durumda ön-düzgünleştirme yapmadan gözlenen puanlar kullanılabilir.

Ön-düzgünleştirmede Log-lineer model uyumlarının değerlendirilmesi. Model uyumlarının değerlendirilmesinde olabilirlik oranı testi ve AIC bilgi kriteri kullanılabilir (Moses & Holland, 2009) gibi gözlemlenen frekanslar Poisson dağılımına sahip olduğunda yaklaşık olarak standart normal dağılıma sahip olan Freeman-Tukey artıkları (FTres) (Bishop ve ark., 1975) kullanılmaktadır.  $n_i$ ,  $i$ . gözlenen frekans ve  $\hat{n}_i$ ,  $i$ . uyum frekansı olmak üzere, Freeman-Tukey artıkları aşağıdaki gibi hesaplanır.

$$FT_i = \sqrt{n_i} + \sqrt{n_i + 1} + \sqrt{4\hat{n}_i + 1} \dots (2.29)$$

İki değişkenli frekans dağılımlarının çok fazla sıfır verisi içermesi ihtimalinden dolayı iki değişkenli durumlarda olabilirlik oranı testi veya AIC bilgi kriteri kullanılarak karşılaştırmaların yapılması önerilmektedir. Yine ortalama, varyans, çarpıklık, basıklık gibi koşullu parametreler de incelenebilir. Eğer koşullu parametreler, tahmin edilen ve gözlenen dağılımlar arasında çok farklılık oluşturuyorsa model değişikliği gerekebilir (González & Wiberg, 2017). Yine kurulan modelin uyumunu değerlendirmek için log-lineer modelde polinomial regresyon (Poisson dağılım ile) analizinde hesaplanması önerilen McFadden  $R^2$  açıklanan varyans oranı da kullanılabilir (Jackman ve ark., 2015). Model, test edilen modeli, null ise sadece kesişimlerin (intercept only) olduğu modeli göstermek üzere aşağıdaki gibi hesaplanmaktadır (Mittlböck & Schemper, 1996).

$$\text{McFadden } R^2 = 1 - (\log \text{likelihood}_{\text{model}} / \log \text{likelihood}_{\text{null}})$$

2. Puan olasılıklarının hesaplanması. Parametrik model oluşturulduktan sonra parametre tahminleri her desen için gerekli olan puan olasılıklarını hesaplamada kullanılır. Kullanılan ön-düzgünleştirme yönteminde göre bu adım da farklılaşır. Denk gruplar ve zincir eşitleme ile denk olmayan gruplarda ortak madde desenlerinde log-lineer yöntemler ya da MTK modellerinin kullanımında bu adım aynı; tek grup ve son-düzgünleştirme ile denk olmayan gruplarda ortak madde deseninde farklıdır. Bu farklılık log-lineer yöntemlerde puan olasılıklarını hesaplamak için tahmini iki değişkenli dağılımların; MTK modellerinde ise birlikte (eş zamanlı) kalibrasyon ya da eşitleme katsayılarının kullanılmasından kaynaklanmaktadır. Burada son-düzgünleştirme ile denk olmayan gruplarda ortak madde deseninde diğerlerinden farklı olarak puan olasılıkları, P ve Q evrenlerinin bir karması olan sentetik bir evren için tanımlanır.

3. Kesikli test puanı dağılımının sürekli dağılım olarak incelenmesi. Puan olasılıkları hesaplandığında elde edilen dağılım kesikli bir dağılımdır. Eşitleme yapılabilmesi için bu dağılımların sürekli dağılıma dönüştürülmesi gerekir. Gauss, lojistik veya tek biçimli (uniform) Kernel yöntemleri ile yapılan dönüşümlerde test puan dağılımlarının ortalaması ve varyansı da korunmaktadır (von Davier ve ark., 2004). Bunlara ek olarak Epanechnikov ve adaptif Kernel yöntemleri (González & Wiberg, 2017) de kullanılabilir. Ayrık puanlanan rastgele bir değişkenin süreklileştirilmesinde hem sürekli bir rastgele değişken hem de bant genişliği parametreleri kullanılmaktadır.  $X(h_X)$  süreklileştirilmiş bir rastgele değişken,  $V$  Kernel'i karakterize eden sürekli bir rastgele değişken ve  $h_X$  bant genişliği parametresi olsun.

$\mu_X = \sum_j x_j r_j$ ;  $\sigma_X^2 = \sum_j (x_j - \mu_X)^2 r_j$ ;  $a_X = \sqrt{\sigma_X^2 / (\sigma_X^2 + \sigma_V^2 h_X^2)}$  olmak üzere aşağıdaki eşitlik elde edilir.

$$X(h_X) = a_X(X + h_X V) + (1 - a_X)\mu_X \dots(2.30)$$

Bu durumda  $K$  seçilen bir Kernel fonksiyonu olmak üzere  $F_X$  kernel düzgünleştirme fonksiyonu aşağıdaki gibi tanımlanır.

$$F_{(h_X)} = \sum_j r_j K \left( \frac{x - a_X x_j - (1 - a_X) \mu_X}{a_X h_X} \right) \dots (2.31)$$

Burada özellikle dikkat edilmesi gereken noktalardan biri de bant genişliği parametresinin seçilmesidir. Bant genişliği, tahmin edilen yoğunluk fonksiyonları düzgün olacak şekilde fakat orijinal ayırık puan dağılımlarının özellikleri korunacak şekilde belirlenmelidir. Bu belirleme işlemi için von Davier ve ark. (2004) aşağıdaki ceza (penalty) fonksiyonu minimum yapacak olan değeri önermişlerdir.

$$PEN(h_X) = \sum_j (\hat{r}_j - \hat{f}_{h_X}(x_j))^2 + k \sum_j \rho_j \dots (2.32)$$

Bu eşitlikte  $\hat{f}_{h_X}(x)$  kestirilen yoğunluk fonksiyonu;  $k$  sabit;  $\rho_j$  ise  $\hat{f}_{h_X}(x)$  fonksiyonunun türevi olan  $\hat{f}_{h_X}'(x_j - v) < 0$  ve  $\hat{f}_{h_X}'(x_j + v) > 0$  olduğunda bire, diğer durumlarda sıfıra eşit olan bir gösterge değişkendir.  $v$  değeri, tipik olarak 0.25 olarak seçilen,  $x$  puanı etrafındaki aralığın genişliğini gösteren bir parametredir (Lee & Davier, 2009). Bu yöntemin dışında son yıllarda çapraz geçerlik (Liang & von Davier, 2014), çift düzgünleştirme (Häggström & Wiberg, 2014) ve kural tabanlı yöntem (Andersson & von Davier, 2014) gibi alternatifler de önerilmiştir.

4. Eşitleme.  $X$  testindeki her bir puana belirli bir tasarıma göre eşit olan bir değer belirlenir. Bu dönüşüm Eşitlik 2.31'deki eşitleme fonksiyonu ile sağlanır.

5. Eşitleme hatalarının hesaplanması. Pratikte eşitleme fonksiyonları kesin olarak bilinemez ve bu nedenle tahmin edilir. Bu tahminlerin varyansı da örneklem değişkenliğinden etkilenmektedir. Kernel eşitlemede eşitlemenin standart hatası (SEE), yüzde görelî hata (PRE), eşitleme farklarının standart hatası (SEED) gibi doğruluk ölçüleri kullanılarak yapılan dönüşümler değerlendirilebilmektedir (González & Wiberg, 2017; von Davier ve ark., 2004).

SEE, tahmini eşitleme dönüşümündeki belirsizliğin bir ölçüsüdür.  $X$  formunun  $Y$ 'ye eşitlemesinde SEE aşağıdaki gibi hesaplanır.

$$SEE_Y(x) = \sigma_Y(x) = \sqrt{Var(\hat{\phi}(x))} = \sqrt{Var(\phi(x; \hat{r}, \hat{s}))} \dots (2.33)$$

SEED, farklı Kernel eşitleme dönüşümlerini karşılaştırmak için kullanılır ve iki SEE vektörü arasındaki farkın Öklid Normu olarak tanımlanmaktadır. İki eşitleme fonksiyonunun örneklem büyüklüğüne bağlı olarak nasıl değiştiğini belirlemek için  $\pm 2$  SEED hata bandı belirlenebilmektedir. Eşitleme fonksiyonları arasındaki fark bu bant genişliği içinde ise endişe edilecek kadar büyük olarak değerlendirilmezler; örneğin, farklılıklar örnekleme hatasından kaynaklanabilir (Liu & Low, 2007).  $\widehat{\varphi}_1$  ve  $\widehat{\varphi}_2$  farklı dönüşümleri göstermek üzere SEED aşağıdaki gibi hesaplanır.

$$SEED_Y(x) = \sqrt{Var(\widehat{\varphi}_1(x) - \widehat{\varphi}_2(x))} \dots (2.34)$$

PRE, eşitlenen değerlerin dağılımlarının momentleri ile eşitlenmiş puan dağılımlarının momentleri arasındaki farkı ölçer. Fark ne kadar küçükse etkili bir puan eşitlemeyi gösterir (Jiang ve ark., 2012).

p. momentteki PRE(p);  $\mu_p(\varphi(X)) = \sum_j (\varphi(x_j))^p r_j$  ve  $\mu_p(Y) = \sum_k (y_k)^p s_k$  olmak üzere aşağıdaki gibi tanımlanır.

$$PRE(p) = 100 \frac{\mu_p(\varphi(X)) - \mu_p(Y)}{\mu_p(Y)} \dots (2.35)$$

Bunlara ek olarak eşitlenmiş puanlar ile ham puanlar arasındaki fark puanlarının beklenen ortalamaları RMSE değeri ile gösterilmiştir. Bu değer 0,80 ile 2,00 arasında olması kabul edilebilir değerlerdir (von Davier ve ark., 2006).  $\bar{d}$ , fark puanlarının ortalaması ve  $sd$  bu fark puanları dağılımının standart sapması olmak üzere;

$$RMSE = \sqrt{(\bar{d})^2 + sd^2} \dots (2.36) \text{ eşitliği elde edilir.}$$

Eşitlemenin doğruluğunu tanımlamak için kullanılan eşitleme hatası ise RMSD (Root Mean Square Differences) ile belirtilmiştir.  $eqX$  eşitlenmiş puanları,  $X$  ham puanları ve  $f_i$  hedef evrendeki puanların frekansını göstermek üzere eşitlik aşağıdaki gibidir.

$$RMSD = \sqrt{\frac{\sum_i f_i (eqX - X)^2}{\sum_i f_i}} \dots (2.37)$$

***Madde Tepki Kuramı'na dayalı Eşitleme Yöntemleri.*** Klasik Test Kuramı'nda öğrencinin bir testten aldığı toplam puan, diğer bir deyişle gözlenen puan, ölçülen özelliğin öğrencide var olma derecesini ifade eder. Fakat zor bir sınavdan düşük puan alan ya da kolay bir sınavdan yüksek puan alan öğrencinin yeteneği değişmemesine rağmen, gözlenen puanı değişir (Hambleton & Jones, 1993). Üstelik iki farklı öğrenci farklı maddeleri doğru yanıtlamış olsa bile aynı toplam puanı alırsa bu durum onların yeteneklerinin de aynı olduğu anlamına gelmez. Bu nedenle MTK öğrencilerin testteki maddeleri doğru ya da yanlış yaptığını gösteren cevap örüntüleri ile ilgilenir. Öğrencilerin her bir maddeyi doğru yanıtlama ihtimali, madde parametreleri ve gizil yetenek parametresi ile hesaplanır. MTK'de bireylerin yetenekleri uygulanan testin, teste ilişkin parametreler de uygulanan grubun özelliklerinden bağımsızdır (Hambleton ve ark., 1991). Maddelerin iki kategorili ya da çok kategorili puanlanma, ölçülen yapının tek boyutlu ya da çok boyutlu olma, toplanan verilerin hiyerarşik düzeylere göre elde edilmesi, testte yer alan maddelerin farklı puanlama türlerini barındırması gibi durumlarda farklı MTK modellerine ait kestirimler kullanılmaktadır. Bu modellerden uygun olan kullanılmadığında parametre kestirimleri hatalı ya da yanlış olabilmekte ve parametre değişmezliği sağlanma ihtimali azalmaktadır (Baker, 2016; Zhu, 1998). MTK'deki modellerden hangisinin veriye uygun olacağına karar vermeden önce ise ölçümlerin tek boyutluluk ve yerel bağımsızlık varsayımlarını incelemek gerekmektedir (Baker, 2016).

MTK varsayımları. Tek boyutluluk, testteki maddelerin birlikte tek bir yeteneği ölçmesi, tek bir örtük özellik için maddelerin istatistiksel olarak bağımlı olmasıdır. İki kategorili puanlanan maddeler için tekrakorik, çok kategorili puanlanan maddeler için polikorik korelasyon matrisi kullanılarak faktör analizi, Stout'un temel tek boyutluluk testi (DIMTEST, POLYDIMTEST), artıkların analizi gibi boyutluluk testleri kullanılabilir. Yerel bağımsızlık ise testte yer alan bir maddenin aynı yetenek düzeyindeki bireyler tarafından doğru yanıtlanma ihtimalinin testteki diğer maddeye bağlı olmaması durumudur. Başka bir deyişle aynı yetenek düzeyindeki bir cevaplayıcının cevap örüntüsünün oluşma



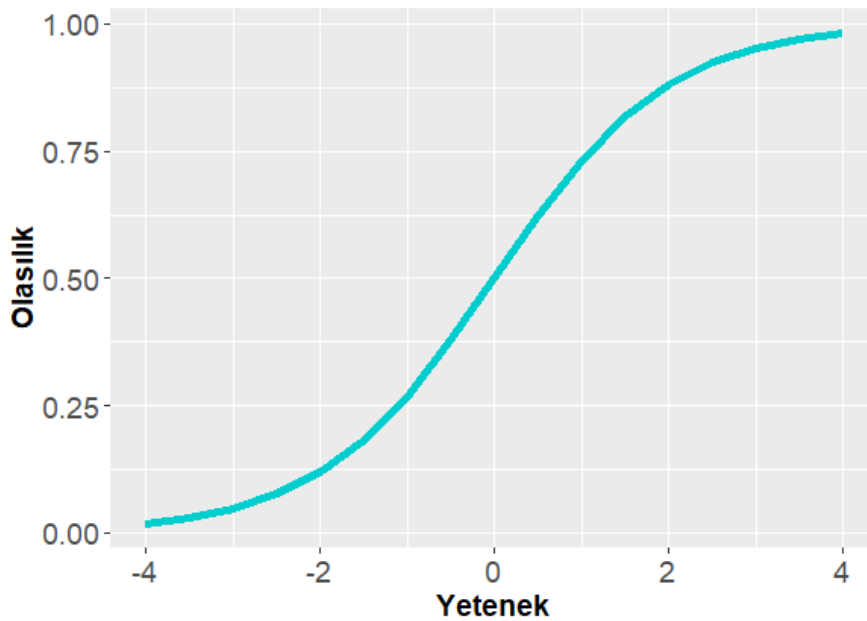
olasılığının maddelerin ayrı ayrı yanıtlanma olasılıklarının çarpımına eşit olmasıdır. Tüm modellerde maddelere verilen yanıtları etkileyen tek faktörün bireyin yeteneği olması gerekmektedir. Bu nedenle testteki bir maddenin diğer bir maddeye yanıt oluşturmayacak şekilde düzenlenmesi beklenmektedir (Hambleton ve ark., 1991). Yardım, kopya çekmek, soruların ipucu vermesi, aşamalı soruların testte yer alması, zaman, yorgunluk, madde formatı, hız, puanlama yanlışlıkları yerel bağımsızlığı etkilemektedir. Yerel bağımlılık görmezden gelindiğinde madde güçlüğü yanlı tahmin edilebilir, madde ayırıcılıkları, test bilgi fonksiyonu, güvenilirlik ve yetenek parametreleri olduğundan yüksek kestirilebilmekte, test eşitlemede hatalı sonuçlar görülebilmektedir. Yerel bağımsızlığı test etmek için madde çiftleri arasındaki korelasyonlar, hatalar arasındaki korelasyonlar ya da bu değişkenler ile hesaplanan Yen'in Q3 istatistiği incelenebilmektedir (DeMars, 2016). Boyutluluk ve yerel bağımsızlık incelendikten sonra verilerin hangi modele uyum sağladığına AIC, BIC, Ki-kare, -2LogLikelihood değerleri ve M2 istatistikleri incelenerek karar verilebilmektedir. Bu değerler birbirine yakın sonuçlar verdiğinde ise parametre sayısı daha az olan modeli tercih etmek daha uygun görülmektedir (Embretson & Reise, 2000). Aşağıda kısaca iki kategorili puanlanan maddeler için MTK modellerinden bahsedilmiştir.

MTK Modelleri. Belirli bir yetenek ( $\theta$ -theta) düzeyine sahip olan öğrencilerin ilgili maddeyi doğru yanıtlama olasılığı  $P(\theta)$  olsun. Dolayısı ile bir maddenin doğru yanıtlanma olasılığı düşük yetenek düzeyindeki öğrenciler için sıfıra, yüksek yetenek düzeyindeki öğrenciler için bire yaklaşıyor. Çünkü yetenek düzeyi yüksek olan öğrencinin maddeyi doğru yanıtlama olasılığının daha yüksek olması beklenir. Maddenin doğru yanıtlanma olasılığını, yetenek parametresinin bir fonksiyonu olarak veren bu fonksiyon monoton artan bir özelliktedir (Baker, 2016). Bu fonksiyon ile elde edilen grafik Madde Karakteristik Eğrisi olarak isimlendirilmektedir. Bu eğriye şeklini veren parametreler; madde ayırt ediciliği ( $a$  parametresi), madde güçlüğü ( $b$  ya da konum parametresi) ve alt asimptottur ( $c$  parametresi).

$$P(\theta) = c_i + (1 - c_i) \frac{1}{1 + \exp(-a_i(\theta - b_i))} \dots (2.36)$$

Bu parametrelerin bir ya da birkaçının sıfıra eşit olma durumlarına göre MTK modelleri farklılaşmaktadır. a, b ve c parametrelerin her biri sıfırdan farklı ise model 3PL;  $c=0$  ise 2PL;  $c=0$  ve a tüm maddeler için sabit ise model 1PL olarak isimlendirilir. Özel olarak 1PL modelde  $a=1$  olduğunda modelin adı Rasch olarak değişmektedir. Burada verilen modeller lojistik fonksiyon üzerine temellenmiştir (De Ayala, 2009; Paek & Cole, 2019). Maddelerin puanlanma kategorileri farklılaştığında normal ogive modeller de kullanılabilir. Fakat bu ayırım yalnız birim farklılığından kaynaklandığından model değişimi gerektirmemektedir (De Ayala, 2009). Aşağıda 2PL modelde  $a=1$  ve  $b=0$  için çizilen bir madde karakteristik eğrisi gösterilmiştir.

**Şekil 1** Madde Karakteristik Eğrisi ( $a=1$ ,  $b=0$ ,  $c=0$ )



a parametresi, eğrinin orta noktasındaki eğimidir. Eğri orta noktada dikleştikçe maddenin ayırt ediciliği artar. b parametresi, %50 olasılıkla maddeyi doğru yanıtlayabilecek kişilerin yetenek düzeyidir. Maddenin hangi yetenek düzeyinde en işlevsel olduğunu gösterir. b parametresi arttıkça eğri x-ekseninde sağa doğru kayar. Bu nedenle konum parametresi olarak da isimlendirilir. c parametresi (pseudo guessing) ise gruptaki en düşük yetenek düzeyindeki bireylerin bile maddeyi doğru yanıtlama ihtimalini gösterir. c parametresi sıfırdan farklı olduğunda a ve b parametrelerinin yorumları c birimlik ötelemeye

istinaden yapılır. b parametresi için %50 ihtimalle maddenin doğru yanıtlanma olasılığı  $\frac{1+c_i}{2}$  yetenek düzeyini; a parametresi için ise b noktasındaki eğimi  $\frac{a_i(1+c_i)}{4}$  formülü ile ifade edilir (Baker, 2016; Embretson & Reise, 2000). Öğrencilerin yetenek parametreleri MTK puanları olarak kullanılmadan önce negatif değerlerden kaçınmak için ( $\theta \in [-3,+3]$ ) uygun bir ortalama ve standart sapma ile puanlar yeniden hesaplanabilir (González & Wiberg, 2017; Paek & Cole, 2019).

Hangi MTK modelinin veri ile daha uyumlu olduğuna karar verdikten sonra ise parametre kestirimleri yapılır. Parametre kestirimleri eşitlenmek istenen iki test formundan elde edilen ölçümlerin birlikte ya da ayrı kalibrasyon yapıma durumuna göre farklılaşmaktadır. Birlikte kalibrasyon, eşitlenecek olan X ve Y formlarında yer alan maddelere ilişkin parametreleri aynı anda kestirilmesidir. Ayrı kalibrasyon ise eşitlenecek olan X ve Y formlarındaki maddeler ayrı ayrı ölçeklenerek ayrı parametrelerin elde edilmesidir. Daha sonra ortak maddelerin/testlerin yardımı ile kalibrasyon yapılmaktadır (Kolen & Brennan, 2014). MTK ölçeklerinin dönüşümü, benimsenen eşitleme desenine bağlı olarak farklı olacaktır. Örneğin, hem TG hem de TG-D tasarımlarında yetenekler ile ilgilenilmediğinde, tahmin, yetenek dağılımının ortalaması ve varyansı için aynı kısıtlamalara sahip olduğu sürece, ölçeklerin ek dönüşümüne gerek yoktur. Fakat DOG deseninde gruplar denk olmadığı için bu ölçek dönüşümleri gereklidir.

Parametre Dönüşümleri. Farklı test formlarına ilişkin parametreler ayrı kestirildiğinde parametrelerin aynı ölçeğe dönüştürülmesi gerekir. Bu dönüşümü gerçekleştirmek için aşağıdaki lineer denklem kullanılır. A ve B eşitleme sabitleri olmak üzere;

$$\theta_{Y_i} = A\theta_{X_i} + B \dots(2.37)$$

Bu eşitlik benzer şekilde parametreler için de yazılabilir.

$$a_{Y_i} = a_{X_j}/A \dots(2.38)$$

$$b_{Y_i} = Ab_{X_j} + B \dots(2.39)$$

$$c_{Y_i} = c_{X_j} \dots(2.40)$$

Bahsedilen dönüşümleri gerçekleştirmek için ihtiyaç duyulan eşitleme katsayılarının kestirimi için aşağıda moment ve karakteristik eğri yöntemlerinden bahsedilmiştir.

*Eşitleme katsayılarının kestiriminde moment yöntemi.* Bu yöntemde A ve B sabitlerini kestirmek için ortalama ve standart sapma gibi farklı momentler kullanılmaktadır. Bu momentler X ve Y formlarında ortak olarak bulunan maddelerden hesaplanmaktadır.  $\mu_{a_x}$  ve  $\mu_{a_y}$ , sadece ortak maddeler üzerinden alınan madde ayırt edicilik parametrelerinin ortalamaları;  $\sigma_{a_x}$  ve  $\sigma_{a_y}$ , benzer şekilde bu parametrelerin standart sapması olsun. González ve Wiberg (2017) katsayıları aşağıdaki gibi tanımlamıştır.

Ortalama-ortalama yönteminde A katsayısı:

$$A = \frac{\mu_{a_x}}{\mu_{a_y}} \dots(2.41)$$

Ortalama-sigma yönteminde A katsayısı:

$$A = \frac{\sigma_{a_x}}{\sigma_{a_y}} \dots(2.42)$$

Ortalama-geometrik ortalama yönteminde  $n_a$  ortak madde sayısını göstermek üzere A katsayısı:

$$A = \prod_{j=1}^{n_a} \left( \frac{a_{X_j}}{a_{Y_j}} \right)^{1/n_a} \dots(2.43) \text{ eşitliği ile bulunur.}$$

Bu üç durumda da B katsayısı aşağıdaki formülle hesaplanır.

$$B = \mu_{b_y} - A\mu_{b_x} \dots(2.44)$$

*Eşitleme katsayılarının kestiriminde karakteristik eğri yöntemi.* Karakteristik eğri yönteminde, madde karakteristik eğrilerine bağlı iteratif bir süreç söz konusudur. A ve B katsayıları optimal düzeye gelene kadar iterasyon devam eder. Haebara ve Stocking-Lord yöntemlerinin her ikisinde de X ve Y formlarındaki ortak maddeleri kullanılmaktadır. Haebara yönteminde öğrencilerin bir yetenek düzeyindeki madde karakteristik eğrileri

arasındaki fark, her bir madde için madde karakteristik eğrileri farkının kareleri toplanarak hesaplanmaktadır. Stocking-Lord yönteminde ise bu fark, her bir madde için madde karakteristik eğrilerinin toplamı arasındaki farkın karesi ile hesaplanmaktadır (Kolen & Brennan, 2014).

*Puanların eşitlemesi.* Yukarıda verilen yöntemlerden biri seçilerek yapılacak kalibrasyon işleminden sonra puan eşitlemede gözlenen ve gerçek puan eşitleme türleri seçilmektedir. Gerçek puan eşitleme, koşullu puan dağılımlarının ortalamasına dayalıdır. Öncelikle bir formdaki gerçek puanlar kestirilir, ters fonksiyon özelliği ile ilgili puana karşılık gelen yetenek düzeyi bulunur (Newton-Raphson metodu ile), en son diğer formdaki gerçek puana karşılık gelen yetenek düzeyi belirlenir. Gözlenen puan eşitleme, marjinal puan dağılımlarına dayalıdır ve koşullu puan olasılıklarını tanımlamak için MTK modellerini kullanır. Puanların dağılımları, özellikleri denk olacak şekilde ortak bir ölçeğe eşit yüzdelikli eşitleme yöntemi kullanılarak getirilir (De Ayala, 2009).

Eşitlemenin son adımında eşitlemenin standart hatası hesaplanır ve elde edilen puanların sahip olması gereken özellikler (eşitleme koşulları) kontrol edilir. Bu koşullar simetri, aynı yapı ve benzer içerik, eşitlik, gözlenen puanların eşitliği ve grup değişmezliği özellikleridir (Kolen & Brennan, 2014).

**Yerel Eşitleme.** Tüm eşitleme yöntemleri, test katılımcılarının tüm popülasyonları için tek eşitlik dönüşümleri kullanmıştır. van der Linden (2010) çalışmasında her yetenek seviyesi için ayrı ayrı eşitleme yapılması gerektiğini göstermiştir. Yerel eşitleme her bir yetenek seviyesi için gerçekten gerekli olan farklı dönüşümler arasında bir uzlaşma ortaya koymaktadır. Eğer eşitlenen her iki test için tüm puan dağılımları mevcutsa yerel eşitleme ile daha adil bir eşitleme yapılmış olur. Bu nedenle tüm katılımcılar için tek bir eşitleme dönüşümü yerine yetenek parametresinin ilgili düzeyleri için eşitleme dönüşümleri ailesi elde edilebilir. Yıldırım Seheryeli ve ark. (2021) bireylerin yetenek dağılımlarına en yakın yetenek düzeyinde yerel eşitleme hatalarının en az olduğunu belirlemişlerdir. Yetenek düzeyindeki birey sayıları az olan gruplarda ise puanlar daha yüksek hata ile kestirilmiştir.

Xin ve Zhang (2015) ilgili yetenek düzeyinin temsil edilip edilememesinin bu yöntemin bir sınırlılığı olduğunu belirtmiştir. van der Linden (2010) eşitlik ilkesi ve grup değişmezliği özelliği sağlanmadığında yerel eşitleme kullanılmasını önermiştir.

Yerel eşitlemede de diğer eşitleme yöntemlerinde olduğu gibi hem TG, DTG, DOG-OM desenlerinde doğrusal ve eşit yüzdelikli eşitleme, Kernel yerel eşitleme, MTK gözlenen puan yerel eşitleme, her iki test formu da MTK için uygunsu MTK gözlenen puan Kernel yerel eşitleme yöntemleri de kullanılabilir (González & Wiberg, 2017).

**2.1.2.6. Eşitlemede Güncel Gelişmeler.** Kernel eşitleme yöntemi olarak alan yazında genellikle Gauss kernel kullanılsa da uniform ve lojistik kernel de alternatif olarak yer almaktadır. Bunlara ek olarak 2015 yılından sonra düzgünleştirme adımı bant genişliği ve yoğunluk fonksiyonlarının farklılaştığı Epanechnikov kernel ve adaptive kernel de önerilmeye başlanmıştır (Cid & von Davier, 2015; González & Davier, 2016). Yine 2015 yılından sonra bant genişliği seçiminde standart hata ve yanlışlık hatasını minimize etmek için kural temelli ve çift düzgünleştirme türleri (Andersson & von Davier, 2014; Häggström & Wiberg, 2014; Liang & von Davier, 2014) önerilmiştir. MTK modellerinde iki kategorili puanlanan maddeler için uygun olan model ile birlikte Gauss Kernel kullanılabilmesinin yanında ikiden fazla kategori kullanıldığında hem aşamalı tepki modeli hem de kısmi puan modeli ile de DG ve DOG-OM desenlerinde eşitleme yapılabilmektedir (Andersson & Wiberg, 2014). İstatistiksel yöntem olarak sıklıkla kullanılan Bayes Yöntemleri, hipotez testlerinde p anlamlılık değerindeki parametre sayısı ve örneklem büyüklüğünden kaynaklanan olası belirsizlikleri modele bir önsel bilgi ekleyerek düzeltmeye çalışır. Burada parametre sayısı sınırlıdır. Parametre sayısındaki bu sınırı ortadan kaldırmak için ise Parametrik Olmayan Bayes Yöntemleri kullanılabilir. Yine bu iki yöntem de eşitlemede önerilen yöntemler arasındadır (Ghosh & Ramamoorthi, 2003; González ve ark., 2015).

## **İlgili Araştırmalar**

Literatür incelendiğinde test eşitleme çalışmalarının genellikle temel alınan kuram, parametre dönüşüm yöntemi ve puan eşitleme türüne göre farklılaştığı görülmektedir. En az farklılaşmaya ise test eşitlemede veri toplama deseninde rastlanmıştır, denk olmayan gruplarda ortak madde deseninde yapılan eşitlemeler diğer desenlerden oldukça fazladır. TG ve DG desenlerinde yapılan çalışmalar ise tek başına çalışılan az sayıda çalışma (Öztürk & Anıl, 2012) olsa da genellikle DOG-OM desen ile incelendiğinden ayrı bir başlık olarak ele alınmamıştır.

### ***Ortak madde deseninde yapılan çalışmalar***

Mutluer (2021) çalışmasında denk olmayan gruplarda ortak madde deseni kullanarak KTK ve MTK'ye göre eşitleme sonuçlarının karşılaştırılmasını amaçlamıştır. PISA 2012 matematik alt testi 1 ve 3. kitapçıklar kullanılmıştır. Tucker eşitleme, Levine gerçek puan, Levine gözlenen puan, konjenerik ve Braun-Holland doğrusal eşitleme ile MTK'de Stocking Lord ve Haebara parametre dönüşümleri ile gerçek ve gözlenen puan eşitleme yöntemlerini karşılaştırmıştır. Araştırmanın sonucunda KTK'ye dayalı yöntemler arasında en az hata Tucker eşitleme en fazla hata Levine gerçek puan eşitleme yönteminden elde edilmiştir. Eşit yüzdellikli eşitleme yöntemleri öncesinde yapılan düzgünleştirmeler sonucunda hata bant değeri ve hata miktarlarında en iyi sonuç beta4 binomial fonksiyonda, en fazla hata ise C6 polinom derecesinde elde edilmiştir. Doğrusal ve eşit yüzdellikli eşitlemede ise eşit yüzdellikli eşitleme daha az hata ile eşitleme yapmıştır. MTK'ye dayalı yöntemlerde öncelikle kalibrasyonlar yapılmış Stocking-Lord yönteminin daha az hata ile kestirim yaptığı görülmüştür. Yine MTK'ye dayalı gerçek ve gözlenen puan eşitlemede ise gerçek puan eşitleme sonucunda daha az hatalar elde edilmiştir. Bu yöntemlerin, örneklemelerin ya da MTK modellerinin farklı kombinasyonlarının yapıldığı çalışmalara ek olarak DOG-OM desen yerine DOG-OD deseninin kullanılabileceği çalışmaların yapılması önerilmiştir.

Güneş (2021) çalışmasında DMF içeren ortak maddeler testten çıkarıldığında eşitlenmiş puanların ve eşitleme hatalarının değişiminin doğrusal olan ve olmayan eşitleme yöntemleri kullanılarak incelenmesini amaçlamıştır. Doğrusal eşitleme için Tucker, Levine, zincir eşitleme, doğrusal olmayan eşitleme için rekans kestirim, zincir eşit yüzdelli ve Daire-Yay Levine yöntemleri kullanılmıştır. DMF değişkenleri olarak İngilizce ve İspanyolca dilleri ile bu dilleri alanların cinsiyet değişkeni olarak belirlenmiştir. DMF belirlemek için Mantel-Haenzel ve SIBTEST kullanılmıştır. Araştırma sonucunda B ve C düzeyinde DMF içeren maddelerin çıkarılmasında eşitleme hatalarında gözlenen değişim tutarlı bir örüntü sergilememiştir. Bu sonucun kitapçıkların güçlükleri arasındaki farkın, DMF değişkeninde gruplar arasındaki yetenek farkının ya da eşitleme varsayımlarının tam olarak karşılanmamış olmasının neden olabileceği belirtilmiştir.

Diao ve Keller (2020) çalışmasında DOG-OM desende tasarlanan bir sınava tekrarlı olarak giren öğrencilerin gruptan çıkarılması, kayma (drift) gösteren maddelerin analizden çıkarılması ve tekrarlı olarak sınava giren öğrenciler çıkarıldığında Rasch gerçek puan eşitleme yönteminin kullanılmasının eşitleme performansına etkisini incelemektir. Araştırmanın sonucunda ankor maddelerden biri madde kaymasına maruz kaldıysa örneklemdaki tekrarlı kişilerin gruptan çıkarılmasının daha uygun olduğu; kişi ya da madde çıkarmanın mümkün olmadığı durumlarda daire-yay eşitleme yönteminin kullanılması gerektiği belirtilmiştir. Rasch gerçek puan eşitleme yöntemi ise diğer çözümlerden daha iyi performans göstermiştir.

Karagül (2020) çalışmasında çok kategorili puanlanan maddelerden oluşan testler ve küçük örneklerde KTK'ye göre hangi yöntemin test eşitlemede kullanılabileceğinin belirlenmesini amaçlamıştır. Simülasyon çalışması olarak tasarlanan araştırmada örneklem büyüklüğü, madde sayısı, kategori sayısı, eşik parametreleri arası fark değiştirilerek üretilen veri setlerinde 100 tekrar yapılmıştır. KTK'ye göre ortalama, Tucker, Braun-Holland ve düşük ağırlıklar doğrusal test eşitleme yöntemleri eşitlemenin standart hatası, yanlılık, ortalama hata karekökü açısından incelemiştir. Araştırmanın sonucunda örneklem



küçüldükçe ve eşik parametreleri farkı arttıkça hataların arttığı görülmüştür. 15 kişiden fazla örneklemlerde tüm koşullarda en düşük hata ortalama yöntemi ile, en yüksek hata düşük ağırlıklar doğrusal eşitleme yöntemi ile elde edilmiştir. Tucker ve Braun-Holland doğrusal test eşitleme arasındaki farklar manidar bulunmamıştır. Örneklem sayısı 15 olduğunda ortalama, 30 olduğunda ortalama ve Tucker, 60 ve üzeri olduğunda ortalama, Tucker ve Braun-Holland yöntemlerinin birlikte kullanılarak sonuçların karşılaştırılması ve desteklenmesi önerilmiştir.

Choi (2019) araştırmasında çok boyutlu madde tepki kuramı yöntemlerine göre denk olmayan gruplarda ortak madde deseninde eşitleme prosedürlerini incelemiştir. Araştırmanın sonucunda çok boyutlu MTK yöntemleri tek boyutlu MTK'den daha iyi sonuçlar ortaya koymuştur. Eş zamanlı kalibrasyon tüm yöntemlerde daha iyi performans göstermiştir. Eşitlenecek formların içerik ve boyutlarının aynı olması gerektiği vurgulanmıştır. Gruplar arasındaki yetenek farkı arttıkça çok boyutlu MTK eşitleme yöntemlerinde de hataların daha yüksek kestirildiği görülmüştür.

Wang (2019) çalışmasında çeşitli koşullar altında mevcut eşitleme yöntemleriyle karşılaştırarak, beta gözlemlenen puan ve gerçek puan yöntemlerini daha iyi anlamayı amaçlamıştır. Denk gruplar ve denk olmayan gruplarda ortak madde desenlerinde beta gerçek puan yönteminin üç olası biçimi (B2B-T, B4B-T ve B4CB-T) ve gözlemlenen puan yönteminin üç olası biçimi (B2B-O, B4B-O ve B4CB-O) incelenmiştir. Gerçek veriler ve simülasyon verileri kullanılarak; yumuşatma, ön düzeltme ve düzeltme sonrası olmayan eşit yüzdelli yöntemler ve MTK gerçek puan ve gözlemlenen puan yöntemleri ve altı form beta gerçek puan ve gözlemlenen puan yöntemleri dahil on beş eşitleme yöntemi kullanılmıştır. Yumuşatma, ön düzeltme ve düzeltme sonrası olmayan frekans tahmin ve eşit yüzdelli yöntemler incelenmiştir. Örneklem büyüklüğü ve test uzunluğunun etkileri ile grup yetenek farkının etkileri incelenmiştir. Araştırmanın sonucunda genel olarak, sözde ve simüle edilmiş veri sonuçları, incelenen yöntemlerden en iyi performansı MTK yöntemlerinin ve ardından zincir eşitleme yöntemlerinin izlediğini göstermiştir. Beta

yöntemlerinden elde edilen sonuçlar, geleneksel ve IRT yöntemlerinden farklı sonuçlar vermiştir. Beta gerçek puan yöntemleri, geleneksel yöntemlere kıyasla grup farklılığına daha az duyarlıydı. Ortak maddelerin uzunluğu, beta gerçek puan yöntemlerinin sonuçlarının kararlılığında önemli bir rol oynamıştır.

Kumlu (2019) çalışmasında çok boyutlu bir testin test geneli ve alt test düzeylerinde tek boyutlu ve çok boyutlu MTK'ye dayalı eşitlenmesi ile elde edilen RMSE, BIAS ve SE değerlerini incelemiştir. Simülasyon çalışması olarak tasarlanan bu araştırmada eşitleme hatasını etkileyen koşullardan alt test arası ilişki düzeyi, örneklemin büyüklüğü 1000-3000, alt testler arası ortak madde oranı %10, %20 ve %30 ve alt test madde sayısı, alt testler arası güçlük farkı ve model türü koşul olarak ele alınmıştır. Ortalama-ortalama ve Stocking-Lord parametre dönüşüm yöntemlerini kullanmış, MTK gerçek puan eşitleme yöntemine göre eşitleme yapılmıştır. Araştırmanın sonucunda her bir alt testin parametreleri tek boyutlu 3PL model ile kestirildikten sonra alt test düzeyinde eşitleme yapılması ile en az hata elde edilmiştir.

Özgürlük (2019) çalışmasında karma testlerin eşitlenmesinde örneklem büyüklüğü, eşitleme yöntemi, ölçek dönüştürme yöntemi ve cinsiyetin eşitleme özelliklerinden dağılımın benzerliği ve eşitlik ilkesine etkisinin incelenmesi amaçlamıştır. Gerçek (ABİDE) veri setinin ve eşitlemede üç formun kullanıldığı bu araştırmada maddelerin formatı (çok kategorili puanlanan, iki kategorili puanlanan), testi alan alt gruplar (kız, erkek), örneklem büyüklüğü (1000, 3000 ve 5000) ve ölçek dönüştürme yöntemleri (ortalama-ortalama, ortalama-sigma, Haebara ve Stocking-Lord) koşul olarak ele alınmıştır. MTK gerçek ve gözlenen puan eşitleme yöntemlerine göre eşitleme yapılmıştır. Araştırmanın sonucunda karakteristik eğri yöntemlerinin moment yöntemlerinden daha kararlı sonuçlar verdiği görülmüştür. Eşitleme öncesindeki puanların birikimli dağılım fonksiyonlarının benzer ya da farklı olmasının ADÖ (aynı dağılım özelliği) korunumunu etkilediği, madde formatının ve eşitleme türünün eşitlenen formlara göre farklılaştığı sonucuna ulaşılmıştır.

Uysal (2019) çalışmasında karma formatlı testlerin gerçek puanlayıcı ve farklı otomatik puanlama yöntemleri ile puanlanmasının test eşitlemeye etkisi incelenmiştir. Gerçek (ABİDE) veri setinin ve eşitlemede iki formun kullanıldığı bu araştırmada KTK'ye dayalı Tucker doğrusal, zincir doğrusal, zincir eşit yüzdelikli ve frekans eşit yüzdelikli eşitleme ve MTK'ye dayalı ortalama-ortalama, ortalama-sigma, haebara ve Stocking Lord parametre dönüşüm yöntemleri ile gözlenen puan eşitleme yapılmıştır. Diğer çalışmalardan farklı olarak puanlarda ön düzgünleştirme yapılmıştır. Eşitleme sonucunda SEE, BIAS ve RMSE değerleri karşılaştırılmıştır. Otomatik puanlamada %10, %20 ve %33 test verisi, yöntem olarak destek vektör makinesi, lojistik regresyon, çok terimli sade bayes, kısa uzun süreli bellek ve iki yönlü kısa uzun süreli bellek kullanılmıştır. Uyumlular için uyum yüzdeleri, karesel ağırlıklı Kappa, Gwet'in AC1 katsayıları incelenmiştir. Araştırmanın sonucunda gerçek puanlayıcı ile en iyi uyum iki yönlü kısa uzun süreli bellek yöntemi ile elde edilmiştir. Test eşitlemede kestirilen hataların tüm otomatik puanlama yöntemlerinde gerçek puanlayıcı ile yakın sonuçlar verdiği görülmüştür. Gerçek puanlama ve otomatik puanlamada da MTK'ye dayalı moment yöntemleri, karakteristik eğri yöntemlerinden ve KTK'ye dayalı yöntemlerden daha düşük hata değerleri üretmiştir.

Kim ve Lu (2018) çalışmalarında teste girenlerin sözde eşdeğer gruplarını (PEG) oluşturmak için ortak değişken verilerini kullanarak test puanlarını eşitlemenin etkinliğini değerlendirmeyi amaçlamışlardır. Her biri 100 madde içeren, 20 (güçlü ilişki) veya 10 (zayıf ilişki) ortak madde ve 30 000'den fazla test katılımcısı tarafından alınan 4 test formu kullanılmıştır. Tek grup tasarımı ile doğrudan eşitleme yapılarak ölçüt olarak alınmıştır. Gruplar yetenek açısından önemli ölçüde farklılık gösterdiğinden, zincirleme eşit yüzdelikli eşitleme yöntemi hem güçlü hem de zayıf ankor madde koşullarında son tabakalama eşitleme yönteminden daha doğru sonuçlar üretmiştir. 10 ortak madde kullanılan son tabakalama eşitleme en az doğru sonuçları vermiştir. PEG'nin etkililiğini arttırmak için, yetenekteki grup farklılıklarını ayarlayabilen ilgili arka plan değişkenleri toplandığında bu eşitleme yöntemi kullanılabilir.

Atar ve Yeşiltaş (2017) çalışmalarında DOG-OM desende elde edilmiş çok boyutlu bir veri yapısında Stocking-Lord, ortalama-ortalama ve ortalama-sigma yöntemlerinin karşılaştırılmasını amaçlamışlardır. Simülasyon koşullarında örneklem büyüklüğü, yetenek dağılımları, boyutlar arasındaki korelasyonlar ve ortak madde yüzdeleri birer parametre olarak alınmıştır. Araştırmanın sonucunda Stocking-Lord yönteminin daha küçük RMSE ve yanlılık değerlerine sahip olduğu görülmüştür. Yetenek dağılımı parametresinin yöntemlerin hiçbirinde etkili olmadığı, örneklem büyüklüğü veya ortak madde yüzdesi arttıkça hatanın azaldığı belirtilmiştir.

Wiberg ve von Davier (2017) çalışmalarında ankor testi sonuçlarında zaman içinde olağandışı bir değişiklik olup olmadığını ve test sonuçlarının farklı gruplar için değişiklik gösterip göstermediğini incelemek için değişkenlerle bağlantılı olarak ankor testleri ve maddelerin incelenmesini amaçlamışlardır. Araştırma sonucunda yaşın, eğitim geçmişinin, cinsiyetin, sınava hangi dönem girildiğinin test puan ortalamaları üzerinde etkili olduğu bulunmuştur. Bu değişkenlerin alt gruplarda da eşitleme fonksiyonlarında farklılaşmaya sebep olup olmadığını incelenmesi önerilmiştir. Özellikle alınan puan uzun yıllar geçerli olduğunda test puanlarının tahmin edicilerini doğru tahmin edebilmenin önemli olduğunu vurgulamışlardır.

Tanberkan Suna ve Tan (2017) çalışmasında Fast and Valid Aptitude Test (FVAT) genel yetenek testinin eski ve yeni iki formunu kullanmış, eşitleme fonksiyonlarını grup değişmezliği açısından incelemiştir. Cinsiyet ve yaş alt gruplarına göre elde edilen eşitleme fonksiyonlarına ait RMSD değerlerinin farklılıklarını incelemiştir. KTK'ye dayalı lineer eşitleme (Tucker, Levine, Broun-Holland), eşit yüzdelikli eşitleme (düzgünleştirilmiş, düzgünleştirilmemiş) yöntemlerine, MTK'ye dayalı ortalama-ortalama, ortalama-standart sapma, Haebara ve Stocking-Lord parametre dönüşüm yöntemlerini kullanarak gerçek ve gözlenen puan eşitleme yöntemlerini koşul olarak ele almıştır. Araştırmanın sonucunda grup değişmezliğinin kısmen ihlal edildiği duruma KTK'ye göre Tucker ve Braun-Holland, MTK'ye göre ortalama-ortalama parametre dönüşüm yönteminin kabul edilebilir hata ile

eşitleme yaptığı görülmüştür. Grup değişmezliğinin büyük ölçüde ihlal edildiği durumda hiçbir yöntem kabul edilebilir hata sınırının altında değerler üretememiştir.

Ertoprak (2017) çalışmasında kayıp verileri mekanizmaları, oranı, testteki konumları, başa çıkma yöntemlerine göre oluşturulan koşullarda eşitleme ve yanlılık hatalarının incelenmesini amaçlamıştır. Simülasyon verilerinde üç farklı kayıp veri yeri (eşitlenecek test-ET, her iki test-HİT, ortak test-OT), üç farklı kayıp veri mekanizması (tamamen rastgele kayıp-TRK, rastgele kayıp-RK, rastgele olmayan kayıp-ROK) ve üç farklı kayıp veri oranına (%10, %20, %40) sahip kayıp verili ve 50 tekrarlı veri setleri üretmiştir. Daha sonra kayıp verili maddelerin bulunduğu veri setleri üzerinde dört farklı başa çıkma yönteminin (uygulanmamış gibi davranma-UGD, yanlış yanıtlanmış gibi davranma-YYGD, lojistik regresyona dayalı çoklu değer atama-LRÇDA, diskriminant fonksiyonuna dayalı çoklu değer atama-DFÇDA) kullanılması ile kayıp veri sorununu çözerek asıl veri setlerine ulaşmıştır. MTK'ye dayalı Stocking-Lord parametre dönüşüm yöntemini kullanmıştır. Araştırmanın sonucunda kayıp verilerin konumunda "Her iki test", kayıp veri mekanizmasında "Tamamen rastgele kayıp", kayıp veri oranında %10, başa çıkma yöntemlerinde ise Diskriminant Fonksiyonuna Dayalı Çoklu Değer Atamanın eşitleme hatalarının en az olduğu görülmüştür.

Cağlak (2016) çalışmasında denk olmayan gruplar deseninde küçük örneklerde eşitleme performansını test etmek için özdeşlik, nominal ağırlık ortalaması, sentetik eşitleme fonksiyonu ve daire-yay eşitleme yöntemlerini karşılaştırmıştır. Örneklem büyüklüklerini 10, 20, 50, 100 olarak belirlemiş, log-lineer dönüşümlerle ön düzgünleştirme yaparak zincir eşitleme prosedürünü kullanmıştır. Araştırmanın sonucunda genel olarak özdeşlik eşitlemenin en iyi, bazı durumlarda nominal ağırlık ortalamasının özdeşlik ve Tucker Lineer eşitlemeden daha iyi sonuçlar verdiği görülmüştür. Eşitlenen testlerin puan dağılımları benzer olduğunda daha düşük hatalar elde edildiği belirtilmiştir.

Uysal (2014) çalışmasında ikili ve çok kategorili puanlanan maddelerin yer aldığı karma testlerde ortak madde oranının %10, %20 ve %30; yetenek dağılımlarının normal ya

da çarpık olduğu durumlarda farklı eşitleme yöntemlerinin sonuçlarının simulasyon verisi kullanılarak incelenmesini amaçlamıştır. Ortalama-ortalama, ortalama-standart sapma, sağlam ortalama-standart sapma, Haebara ve Stocking Lord kalibrasyon yöntemleri kullanılmıştır. Araştırmanın sonucunda en fazla hata sağlam ortalama-standart sapma, en az hata ise sağa çarpık-sağa çarpık dağılıma sahip populasyonların eşitlendiği ve ortak madde oranının %30 olduğu durumda Stocking Lord yönteminde elde edilmiştir. Ek olarak yetenek dağılımları eşitlenen iki grupta da benzer olduğunda daha az hatalı kestirime ulaşılmıştır.

Moses ve ark. (2010) çalışmalarında DOG-OM desende yapılan eşitlemeye birden fazla ankorun dahil edilmesi için, son tabakalaşma, imputasyon ve eğilim skoru eşleştirmesi dahil olmak üzere üç yöntemin değerlendirilmesini amaçlamaktadır. Araştırmanın sonucunda son tabakalaşma yönteminin varsayımlarını ve eşitlenmesini uygulamak için yaklaşımların nasıl kullanılabileceğini açıklanmıştır. Üç yaklaşım, iki ankor kullanmanın daha garantili olduğunu göstermiştir. Tabakalaşma sonrası, atama ve eğilim puanı eşleştirme yaklaşımlarının, iki ortak testi birleştirmek ve eşitleme ve ölçekleme işlevlerini hesaplamak için benzer şekillerde kullanılabileceğini göstermiştir. Son tabakalaşma ve atama yöntemleri, bu çalışmanın her iki örneği için temelde aynı olan sonuçlar üretmiştir.

Livingston ve ark. (1990) çalışmalarında test formlarını eşitlemek için beş yöntem ve eşitleme için iki örnekleme yönteminin kombinasyonlarını eşitleme doğruluğu açısından karşılaştırmayı amaçlamışlardır. Örnekleme yöntemlerinden ilki eski formu alan bireyler arasından rastgele örneklem alınması (temsili örnekleme), ikincisi ise ankor maddeleri kullanarak bir katmanlaştırıcı dönüşüm ile eski ve yeni formdaki ankor puan dağılımlarının benzer olmasının sağlanması ile oluşturulan örneklemdir (eşleşen örnekleme). Eşitleme yöntemi olarak ise Tucker, Levine eşit derecede güvenilir, zincirleme eş yüzdelikli, frekans tahmini ve MTK 3PL yöntemlerini kullanmışlardır. Araştırmanın sonucunda, eşitlemelerdeki tutarsızlıkların çoğu genel olarak yanlışlığa atfedilebileceğini belirtmiştir. Eşleşen örneklemdaki tüm eşitleme yöntemlerinin sonuçları, temsili örneklemdaki Tucker ve frekans

tahmin yöntemlerine bezer sonuçlar vermiştir. Temsili örnekleme, zincirleme eş yüzdelikli yöntem daha küçük bir sapma göstermiştir. MTK (3PL) ve Levine yöntemleri ise birbirleriyle benzer sonuçlar vermiş ve yanlılıkların tutarlı olmadığı görülmüştür.

### ***Ortak madde deseninde ve Kernel eşitleme yöntemi ile yapılan çalışmalar***

Koçar ve Sayın (2018 Ekim) bildirimlerinde DOG-OM deseninde, 250-1000 örneklem büyüklüklerinde ve 0-1 yetenek düzeylerinde simülasyon verilerini kullanarak Kernel tekniklerinden (gauss, lojistik, standardize edilmiş gauss ve tek biçimli) elde eşitleme hatalarını karşılaştırmışlardır. Araştırmanın sonucunda aynı yetenek düzeyi ve farklı örneklem büyüklüklerinde tek biçimli teknik dışındaki tüm tekniklerin birbirine çok yakın hata değerleri ürettiği, tek biçimli teknikte daha düşük hata değerlerine ulaşıldığı görülmüştür. Örneklem büyüklüğü arttıkça da yine hata değerlerinin azaldığı bulunmuştur.

Andersson ve Wiberg (2017) çalışmalarında denk olmayan gruplarda ortak madde deseninde MTK gözlenen puan Kernel eşitleme yönteminin kullanılmasını anlatmışlardır. Hem simülasyon hem de gerçek verilerde 2PL ve 3PL model kullanılmıştır. Hem zincir eşitleme hem de son tabakalama ile eşitleme yapılmıştır. Araştırmanın sonucunda Kernel eşitlemenin 2PL model için minimum 1000, 3PL model için minimum 3000 örneklem büyüklüğüne ihtiyaç duyduğu, dağılımın normal olmadığı yakınsama problemlerinin yaşanabildiği, bu durumu düzeltmek için bazı parametrelerin sifira eşitlenebileceği belirtilmiştir MTK yöntemlerinin düşük puan aralıklarında log-lineer yöntemlerden daha küçük SEE ler ürettiği, yüksek puan aralıklarında ise SEE lerin benzer olduğu bulunmuştur. Model yanlış belirlendiğinde ise yanlılık hatasının artma eğiliminde olduğu belirtilmiştir.

Akın Arıkan (2017) çalışmasında Kernel eşitleme yöntemleri ile MTK gerçek puan eşitleme yöntemlerini örneklem büyüklüğü, yetenek dağılımı, ortak madde tipi, ortak madde oranı ve ortak maddelerin güçlük dağılımına göre karşılaştırmıştır. Simülasyon çalışmasında örneklem sayısının 500-1500-5000 olduğu 3PLM'ye göre 100 tekrar ile veriler üretilmiştir. Kernel son tabakalama doğrusal, Kernel zincirleme doğrusal, Kernel son tabakalama eşit yüzdelikli, Kernel zincirleme eşit yüzdelikli ve MTK gerçek puan eşitleme

(parametre dönüşüm yöntemi olarak Haebara) yöntemlerini kullanarak eşitlemenin hatası ve eşitlemenin standart hatasını incelemiştir. Araştırmanın sonucunda örneklem büyüklüğünün arttıkça tüm yöntemlerde eşitlemenin standart hatasının ve toplam hatanın azaldığı, fakat örneklem büyüklüğünün standart hatayı daha fazla etkilediği görülmüştür. Yetenek dağılımının benzer olduğu durumlarda son tabakalama eşitlemenin, farklı olduğu durumlarda zincirleme eşitleme yöntemlerinin daha az hata verdiği bulunmuştur. Ortak madde oranı arttıkça özellikle son tabakalama eşitleme ve MTK gerçek puan eşitlemede hatalar daha da az elde edilmiştir. Tüm eşitlemelerde dış ortak test daha düşük hata vermiştir. Ortak maddelerin güçlük dağılımında ise iç ortak testte yetenek dağılımı benzerken mini (aynı kapsam, istatistiksel açıdan benzer test) ve midi (aynı kapsam, madde güçlükleri orta güçlükte) testten elde edilen sonuçlar da benzer, yetenek dağılımı farklı iken midi test daha düşük hata vermiştir. Dış ortak testte ise midi test sonuçları daha karardır. Tüm koşullar incelendiğinde ise Kernel eşitleme yöntemlerinin orta puan ölçeğinde düşük hata uçlarda frekans azlığı nedeniyle yüksek hata verdiği sonucuna ulaşılmıştır. Örneklem büyüklüğü 5000 olduğunda ise tüm yöntemler benzer etki göstermiştir.

Bağcı ve ark. (2016 Ekim) bildirimlerinde küçük örneklemelerde (50, 100, 150, 200) farklı eşitleme yöntemlerinin doğruluğunun incelenmesini amaçlamışlardır. Eşitlenmiş puanların ortalama standart kata kestirimlerini Tucker-lineer, Tucker-ortalama, Tucker-circle, Levine-lineer, Levine-ortalama, Levine-circle, Braun-Holland lineer, Braun-Holland ortalama, Braun-Holland circle, nominal weights-ortalama, nominal weights-circle, eşit yüzdellikli-zincirleme, lineer-zincirleme, ortalama-zincirleme, zincirleme-circle, frekans kestirim ve Kernel eşitleme yöntemlerinde karşılaştırmışlardır. Araştırmanın sonucunda örneklem büyüklüğü arttıkça tüm yöntemlerde hatanın azaldığı, özellikle 50 örneklemeden 100 örnekleme yaşanan hata düşüşün 100 örneklemeden 200 örnekleme yaşanan düşüşe göre oldukça fazla olduğu görülmüştür. 100-150 ve 200 örneklemlerden elde edilen hata değerleri ise birbirine yakındır. Frekans kestirim yöntemi diğer bütün yöntemlerden daha fazla hata üretirken Kernel yöntemi franks kestirimi, Braun-Holland lineer ve zincirleme eşit



yüzdellikli eşitleme yöntemlerine göre daha az, circle-ark yöntemi ise en az hata üretmiştir. Ortalama yaklaşımları ve Nominal Weights yöntemi de DOG-OM desende az hata ile kesitim yaptığı görülmüş, kullanılması önerilmiştir.

Kim (2014) çalışmasında DOG-OM tasarımı altında çeşitli düzgünleştirme yöntemlerini incelemeyi amaçlamıştır. Örneklem büyüklüğü, grup yetenek farkı, ortak madde oranı, form zorluğu, test uzunluğu, ortak madde zorluğu gibi koşullar altında simülasyon verilerini kullanmışlardır. Log-lineer ön yumuşatma ve kübik spline post yumuşatma ile frekans tahmini, değiştirilmiş frekans tahmini, zincirleme eşit yüzdellikli eşitleme ve kernel eşitleme yöntemleri kullanılmıştır. Araştırmanın sonucunda, en az bir düzgünleştirme yöntemi düzgünleştirilmemiş yöntemden daha küçük hata üretmiştir. Frekans kestirimi yöntemleri kullanıldığında, sistematik ve toplam hatalar açısından polinom log-lineer ön-düzleştirmeden daha iyi performans gösterme eğiliminde olmuştur. Kübik spline son düzgünleştirme, kullanılan eşitleme yönteminden bağımsız olarak en az miktarda rastgele hata üretmiştir. Kernel eşitleme, zincir eşitleme ile karşılaştırıldığında çoğu test koşulu altında daha doğru sonuçlar vermiştir. Log-lineer ön düzgünleştirme, kübik spline son düzgünleştirmeden daha fazla test koşulu altında daha küçük hata üretmiştir.

Moses ve ark. (2011) çalışmalarında DOG-OM desende yapılan eşitlemede birden fazla ankor testin kullanılması ile model doğruluğunun artırılması için son tabakalaşma ve eksik veri atama yöntemlerinin incelenmesini amaçlamıştır. Kernel'in genişletilmiş bir hâlini kullanmışlardır. Araştırmanın sonucunda her iki yaklaşımın da hemen hemen eşdeğer eşitleme sonuçları ürettiğini, ancak son tabakalaşma yaklaşımının standart hatalar açısından bazı esneklik ve doğruluk avantajlarına sahip olduğunu göstermektedir. Çok farklı sınava giren gruplar tarafından alınan test formlarını içeren eşitleme durumlarının incelenmesi önerilmiştir.

von Davier ve ark. (2006) çalışmalarında Kernel eşitlemenin EG ve DOG-OM desenlerde diğer desenler kadar eşitleme için uygun sonuçlar üretip üretmediği ve varsa bu farkın ne kadar olduğunu incelemiştir. Araştırmanın sonucunda, Kernel eşitleme

sonuçlarının diğer yöntemler ile benzer sonuçlar verdiğini, Kernel eşitleme yöntemlerinin eşitleme kriterine daha yakın olduğu belirtmişlerdir. Yine dış ankor ve iç ankor için de büyük farklılıklar görülmemiştir. Bunların dışında Kernel eşitlemede doğruluğu değerlendirmek için diğer yöntemlerde bulunmayan doğruluk ölçüleri de yer almaktadır.

### ***Ortak değişken deseninde yapılan çalışmalar***

Albano ve Wiberg (2019) çalışmalarında DOG-OM ve DOG-OD desende, değişen örneklem büyüklükleri, popülasyonların yetenek farklılıkları ve ankor testin uzunlukları ve ankor türü gibi koşullar altında bağlama ve eşitleme doğruluğunu incelemeyi amaçlamışlardır. Araştırmada gerçek veriler kullanılarak yeniden örnekleme ile elde edilen veri setleri kullanılmıştır. Tüm popülasyondan rastgele örneklenen denekler araştırmanın birinci kısmını, farklı iki alt gruptan rastgele örneklenen denekler araştırmanın ikinci kısmını oluşturmaktadır. EG deseninde özdeşlik, doğrusal, eşit yüzdelikli ve ön düzgünleştirilmiş eşit yüzdelikli eşitlemenin tümü gerçekleştirilmiştir. DOG-OM, DOG-OD ve DOG-OMC tasarımları altında, ön düzgünleştirme ile eşitleyen Tucker lineer ve frekans tahmini eşit yüzdelikli eşitleme kullanılmıştır. DOG-OM ve DOG-OD tasarımlarında zincirleme eşit yüzdelikli eşitleme de yapılmıştır. Sonuçların değerlendirilmesi için replikasyonlara dayalı ortalama bir RMSE değeri hesaplanmıştır. Araştırma sonucunda, ortak değişken olarak okuma puanları ve matematik puanları alınmış benzer sonuçlar verdiği için yalnız okuma puanları raporlanmıştır. Birinci çalışmada, EG tasarımı altında ortak değişken ya da madde olmadığında eşit yüzdelikli eşitleme, DOG-OD desen EG'de ortak değişken olmamasından daha doğru sonuçlar vermiştir. DOG-OMC tasarımında ise en küçük hatalar frekans tahmininden gelmiştir. Eşitleme doğruluğundaki en büyük değişim örneklem büyüklüğünden kaynaklanmıştır. İkinci çalışmada ise EG tasarımı altında özdeşlik eşitlemenin her zaman doğrusal veya eşit yüzdelikli eşitlemeye tercih edildiğini göstermiştir. Yine örneklem büyüklüğünün etkisi ilk çalışma ile benzer; ortak madde sayısının fazla olması ilk çalışmadan daha etkili bulunmuştur. Ortak değişkenlere duyulan ihtiyacın eşdeğer olmayan gruplar tasarımlarında daha fazla olduğu belirtilmiştir. DOG-OMC tasarımında frekans

tahmini yöntemi yine en az hata üretme eğilimindedir. Ortak değişkenler ve toplam puanlar arasındaki ilişkilere ne ölçüde bağlı olduğunun; popülasyon değişmezliği kaybı arasındaki dengenin; küçük örneklem büyüklüklerinde ortak değişkenin etkisinin incelenmesini önermişlerdir.

Yurtçu (2018) çalışmasında ortak madde ve ortak değişken kullanılarak parametrik olmayan Bayes (PoB) yöntemi ile eşitlenmiş puan dağılımlarının ve eşitleme hatalarının incelenmesini amaçlamıştır. Parametrik olmayan Bayes yöntemini, MTK'ye dayalı parametre dönüşüm yöntemlerinde ortalama-ortalama, ortalama-standart sapma, Haebara ve Stocking-Lord yöntemlerini kullanmıştır. PISA 2012 verilerinin Kanada ve İtalya örneklemelerinde 5 ve 6. kitapçıkları eşitleme formu olarak, cinsiyet, matematik öz yeterliği değişkenleri ile birlikte ortak maddeleri ortak değişken olarak kullanmıştır. Eşitleme sonuçlarını karşılaştırmak için eşitlenmiş puanların dağılımı ile hedef testin dağılımı arasındaki Hellinger uzaklıklarını incelemiştir. Araştırmanın sonucunda en yüksek hata değerlerinin Stocking-Lord ile elde edilmiş olmasına rağmen, hedef testin dağılımına olan uzaklıklar incelendiğinde en yakın sonuçların bu yöntemle elde edildiği görülmüştür. Cinsiyet ele alındığında dağılımlar benze olsa da 6. kitapçığın 5. kitapçıktan daha kolay olduğu, cinsiyetin matematik performansı üzerinde önemli bir etkisinin olmadığı görülmüştür. Matematik öz yeterliği ortak değişkeninin düşük-orta-yüksek olarak sınıflandığı durumda yalnız düşük düzeyin diğer düzeylerin dağılımında farklılaştığı, her iki kitapçıkta da dağılımların benzer olduğu görülmüştür. Bu değişkenin kitapçıklar arasındaki farkı ortaya çıkaramadığı belirtilmiştir. İki değişkenin birlikte ele alındığı durumda ise daha ayrıntılı bilgiler elde edildiği, hem bireyler hem de kitapçıklar arasındaki farkın ortaya konabildiği, uzaklıklar ve dağılımlar incelendiğinde ise eşitlenmiş puan dağılımlarının ve hedef test puan dağılımlarının birbirine yaklaştığı sonucuna ulaşılmıştır. Ortak maddelerin ortak değişken olarak ele alındığı durumda ise ortak maddelerden düşük ve yüksek puan alan bireylerde puan dağılımları benzerlik göstermiştir. İki kitapçık için de bu durum ortak maddelerin direk test puanlarına etki ederek dağılımları değiştirecek nitelikte olduğunu göstermiştir. Ortak

maddelerin kullanılmasının da dağılımları yaklaştırdığı fakat iki değişken kullanılması kadar olmadığı görülmüştür. PoB modelleri dağılımlar üzerinden eşitleme yaptığından gruptaki kişi sayılarının az ya da aynı olması gibi sınırlılıklarının bulunmadığı belirtilmiştir. Fakat sınıflama yapıldığında gruptaki birey sayısının az olmasının güven aralıklarının geniş olmasına yol açtığı bu nedenle de ortak değişkenlerin bilgi vermeyen önseller olduğunda güven aralıklarının geniş olduğu sonucuna ulaşılmıştır.

González ve ark. (2015) araştırmalarında ortak bir değişkenin kullanıldığı durumda Bayes parametrik olmayan modeli kullanarak eşitleme yapmanın yolunu önermişlerdir. Bu modelde puan dağılımları grafiğinin şeklinin ortak değişkenlerin bir fonksiyonu olarak değişebilmesi modelin güçlü yanı olarak belirtilmiştir. Araştırmanın sonucunda simülasyon verilerinde önerilen modelin gerçek eşitleme fonksiyonlarını tahmin edebildiği, eşit yüzdelli eşitleme ve Kernel eşitlemeden ya daha iyi ya da benzer düzeyde eşitleme sonuçları verdiği görülmüştür. Gerçek veri kullanıldığında ise önerilen modelin diğer alternatif parametrik olmayan modellerinden daha iyi sonuçlar verdiği belirtilmiştir. Ayrıca bu model ile geleneksel eşitleme yöntemlerinden farklı olarak alt popülasyonlardaki değişmezlik de eşitleme fonksiyonları ortak olarak incelenebilmektedir. Bu incelemenin yazarlar tarafından devam ettiğini belirtmişlerdir. Bu model ile bazı puan değerlerinde bulunan atipik dağılımları ya da yoğunluğun en düşük ve en yüksek puanlarda yığılması gibi durumlarda da eşitleme tahminlerini gerçekleştirebilmektedir. Bu iyi özelliklerin bağımlı Bayes parametrik olmayan bir yöntem kullanıldığı durumda ve denk gruplar deseninde geçerli olduğunu vurgulamışlardır.

Bränberg ve Wiberg (2011) çalışmalarında denk gruplar ve denk olmayan gruplarda ortak madde deseninde ortak değişken kullanımının test eşitlemeye etkisinin eşitleme hatası, yanlılık ve varyans açısından incelenmesi amaçlamışlardır. Ortak değişken olarak cinsiyet, yaş ve sınıf düzeyi değişkenlerini kullanmışlardır. Araştırmalarında Swedish Scholastic Aptitude Test (SweSAT) verilerini ve bu verilerin parametrelerinden 20 000 tekrarlı, örneklem sayısı koşulu 200-500-2000 olan bir simülasyon çalışması

oluşturmuşlardır. Araştırmanın sonucunda ankor testin bir ortak değişken olarak kullanıldığı durumda gruplar arasındaki farklılıkları düzeltmede en iyi olduğu görülmüştür. Fakat ortak maddelerin kullanılmadığı durumlarda (tüm ankor testlerin tek boyutlu olması, ankor testlerin formlarla aynı yapıyı pratikte ölçememesi, ankor testin gruplar arasındaki farklılıkları tam olarak açıklayamaması) bir ya da daha fazla ortak değişkenin kullanılmasını önermişlerdir.

Liou ve ark. (2001) araştırmalarında ortak maddelerden elde edilen puanlar ile test puanları arasındaki korelasyon çok küçük olduğunda ortak değişkenlerin de modele nasıl dahil edileceğinin gösterildiği bir model önermişlerdir. Eşdeğer olmayan gruplarda parametreler EM algoritması ile incelenmiş ve genel lineer model kullanılmıştır. Araştırmanın sonucunda daha düşük korelasyon olmasına rağmen değişkenlerin ortak maddeler gibi puanlarının dağılımları tahmin etmek için işe yaradığını, örneklem yanlılığının beklendiği kadar büyük olmadığı belirtilmiştir.

Wright ve Dorans (1993) çalışmalarında popülasyonlar arasındaki sistematik farklılıkları açıklayan bir değişkeni bildiğimizde ve o değişkeni bir ortak test tasarımında veya eşitleme yapılabilecek bir değişken olarak kullandığımızda eşitleme sonuçlarının iyileşip iyileşmeyeceğini incelemeyi amaçlamışlardır. Bu nedenle hem gerçek veriler hem de simülasyon verileri ile değişkenler birer ortak madde gibi ele alınmıştır. Ortak değişken olarak sayısal test için sözel, sözel test için sayısal test puanları kullanılmıştır. Araştırmanın sonucunda Tucker ve frekans tahmini sonuçları, ortak değişken üzerinde eşitleme sonuçlarını iyileştirmiştir, zincirleme eşit yüzdelik ve Levine sonuçları benzer bulunmuştur. Bir ankor olarak ortak değişken ile elde edilen sonuçlarda ise hem Tucker hem de frekans tahmin yöntemleri iyi sonuçlar verirken zincirleme eşit yüzdelikli ve Levine eşitleme için varsayımlar sağlanamamıştır.

### ***Ortak değişken deseninde ve Kernel eşitleme yöntemi ile yapılan çalışmalar***

Altıntaş ve Wallin (2021) çalışmalarında Ankara Üniversitesi Yabancı Uyruklu öğrenci sınavının ortak maddesi olmayan ve aynı yapıyı ölçtüğü varsayılan 2017 ve 2018

yıllarındaki uygulamalarının DOG-OD desende Kernel eşitleme yöntemi ile eşitlenmesini araştırmışlardır. Ortak değişken olarak veri setinde yer alan cinsiyet ve yaş değişkenleri kullanılmıştır. Araştırmanın sonucunda test formlarının farklı puan dağılımlarına sahip olduğunu ve ilişkinin doğrusal olmadığı görülmüştür. Bu nedenle de bu farklılıkları ortadan kaldıracak şekilde eşitleme süreci tasarlanmıştır. Eşitleme sonucunda test puanlarında uçlara doğru gidildikçe eşitleme hatasının arttığı, bu durumun uçlardaki puanları alan kişi sayısının az olmasından kaynaklandığı belirtilmiştir.

Özsoy (2021) çalışmasında DOG-OM ve DOG-OD desende Kernel eşitleme yöntemlerinin kestirilen hata değerlerinin TIMSS 2019 Fen testi bağlamında karşılaştırılmasını amaçlamıştır. DOG-OM desende Kernel sontabakalama eşit yüzdelikli, Kernel zincirleme eşit yüzdelikli, Kernel sontabakalama doğrusal ve Kernel zincirleme doğrusal eşitleme yöntemleri kullanılmıştır. DOG-OD desende bilgisayar/tablet ile cinsiyet değişkenleri ortak değişken olarak ayrı ve birlikte ele alınmış ve Kernel eşit yüzdelikli ve Kernel doğrusal eşitleme yöntemleri karşılaştırılmıştır. Araştırmanın sonucunda DOG-OM desende Kernel sontabakalama eşit yüzdelikli eşitleme, DOG-OD desende ise bilgisayar/tablet sahibi olup olmama değişkeni ve Kernel doğrusal eşitleme kullanıldığında daha düşük kestirim hataları elde edilmiştir. Ayrıca DOG-OD desende ortak değişkenler ayrı ayrı kullanıldığında yine daha düşük hatalar kestirilmiştir. DOG-OM ve DOG-OD desen karşılaştırıldığında ise DOG-OD desenin daha yüksek hatalar ürettiği görülmüştür. Ortak değişkenlerin ortak madde yerine kullanılamayacağı, ortak madde olmadığı durumda kullanılmasının uygun olduğu sonucuna ulaşılmıştır.

Akın Arıkan (2020) çalışmasında Kernel eşitlemede PSE ve CE eşitleme yöntemlerinde DOG-OM ve DOG-OD desenlerin kullanımından elde edilen sonuçları karşılaştırmıştır. Araştırmada 2016 ABIDE Matematik alt testine ilişkin veriler kullanılmıştır. Ortak değişken olarak test puanları ile ilişkisiz olan cinsiyet ve test puanları ile düşük ilişkili sosyoekonomik düzey ve bu iki değişken birlikte ele alınmıştır. Sosyoekonomik düzey değişkeni kümeleme analizi ile üç gruba ayrılmıştır. Eşitleme sonuçlarını karşılaştırmak için

test puanları ve eşitlenmiş puanlar arasındaki fark puanları, hatalar için RMSD ve SEE değerleri incelenmiştir. Araştırmanın sonucunda eşitlenmiş puanlar ile ham puanlar arasındaki farkın en fazla DOG-OM CE yönteminde, en az DOG-OD desende olduğu görülmüştür. Fark puanları dağılımında DOG-OD desenin DOG-OM PSE ile benzer dağılım gösterdiği belirtilmiştir. Standart hatalarda ise 0-4 puan aralığında yöntemlerin benzer sonuçlar verdiği ve puanlar dörtten büyüdükçe hataların da farklılaştığı belirtilmiştir. DOG-OD desende ortak değişken olarak sosyoekonomik düzey ve cinsiyetin birlikte kullanıldığında kestirim hatalarının azaldığı, sosyoekonomik düzeyin kullanıldığı modelde cinsiyete göre daha az hatalar elde edildiği görülmüştür. Ortak maddelerin yer almadığı durumlarda DOG-OD desen kullanılabilirliği, ABIDE verilerinin diğer alt testlerinin incelenebileceği, Parametrik olmayan Bayes ile Kernel eşitleme sonuçlarının karşılaştırılabileceği önerilmiştir.

Wallin ve Wiberg (2019) çalışmalarında gerçek ve simülasyon verileri kullanarak DOG-OD desende Kernel eşitleme yöntemi ile eşitleme sonuçlarını incelemeyi amaçlamıştır. Fakat ortak değişkenlerin kategorilerinde gözlem olmama ihtimalini düşünerek doğrudan ortak değişkenleri kullanmak yerine eğilim (propensity) puanlarını kullanmışlardır. Gerçek verilerin kullanıldığı çalışmada eğilim puanlarının doğrudan ortak değişkenlerin kullanılmasından daha küçük bir standart hata ürettiği görülmüştür. DOG-OD deseni yine EG ve DOG-OM deseninden uç puanlarda daha düşük hatalarla kestirim yapmıştır. Tüm koşullarda PSE yaklaşımları CE'den daha iyi sonuçlar vermiştir. Fakat bu yorumların tüm veri setleri için geçerli olabilmesi için daha çok çalışmaya ihtiyaç olduğunu ve hem ortak değişken hem de eğilim puanları kullanmanın altında yatan varsayımların test edilmesini önermişlerdir.

Sansivieri (2017) çalışmasında geleneksel eşitleme yöntemleri, Kernel eşitleme ve MTK'ye dayalı test eşitlemeyi inceleyerek ortak madde içermeyen testlerde DOG-OD desene göre yeni bir IRT modeli önermeyi amaçlamışlardır. Hem gerçek hem de gerçek ve gözlenen puan eşitlemede ortak değişken kullanımı hem simülasyon hem de gerçek veri

kullanılarak incelenmiştir. Ortak değişken olarak sınava girenlerin cinsiyeti ve uyruğu alınmıştır. Araştırmanın sonucunda simülasyon çalışmasında MTK gözlenen puan eşitleme yönteminin desteklendiği, DOG-OD tasarımında önerilen yöntemin EG tasarımındaki geleneksel yöntemlerden daha küçük hatalar ürettiği görülmüştür. MTK gerçek puan eşitlemede ise DOG-OD deseninde büyüklüğü, DIF varlığı ve madde sayısı koşulları değiştirildiğinde eşitleme hatalarında herhangi bir iyileşme görülmemiştir. Bu durumun gerçek veri setindeki ortak değişkene göre ciddi başarı farkı olmasından kaynaklanabileceğini belirtmişlerdir. Araştırmada DOG-OD desende önerilen MTK gözlenen puan eşitleme yönteminin kullanılabileceği ayrıca kullanılan ortak değişkenlerin ulaşılabilirlik nedeniyle seçildiği başka araştırmalarda farklı değişkenlerin denenmesi önerilmiştir.

Longford (2015) çalışmasında iki farklı sınava girenlerin, bir dizi ortak değişkenden elde edilen eğilim puanları üzerinde eşleştirilmesini amaçlamıştır. Bu iki puan alt kümesi daha sonra eşitlenmiş ve ilişkili test performansları kümeleri eşdeğer olarak ele alınmıştır. Araştırmanın sonucunda ortak değişkenler baz alınarak oluşturulan eğilim puanı Kernel son-düzenleştirme yapılarak eşitleme için kullanılmıştır. Burada ortak değişkenler ortak madde gibi düşünülmüştür. Bu yöntemin bir sınırlılığı olarak, ortak değişkenlerin yöntemden tamamen bağımsız olması varsayımı test edilebilir değildir, bunu gidermek için çok sayıda değişken kümesinden bir eğilim puanı elde etmek daha uygun görülmüştür.

Wiberg ve Bränberg (2015) çalışmalarında ortak kişi ya da ortak madde kullanılmayan bir durumda test puanları ile ilişkili başka bir değişkenin test eşitlemede kullanılabilmesi durumunu incelemişlerdir. EG, DOG-OM ve DOG-OD desenlerinin kullanımlarını karşılaştırmış, gözlenen puan Kernel eşitleme yöntemlerini incelemişlerdir. Ortak değişken olarak öğrencilerin sınıf düzeylerini ve sözel test puanlarını kullanmışlardır. Araştırmada Swedish Scholastic Aptitude Test (SweSAT) verilerini kullanarak DOG-OD ve ankor test kullanılan EG deseni karşılaştırmışlardır. Eşitlenmiş puanlar ile satır toplam puanları arasındaki farkların dağılımı ve SEE ler incelenmiştir. Araştırmanın sonucunda



ortak test ile ortak deęişkenlerin birlikte kullanıldığı durumda en küçük standart hatalar elde edildięi görülmüştür. DOG-OM CE ve DOG-OM PSE'nin DOG-OD e yakın sonuçlar verdięi belirtilmiştir. Ortak deęişkenlerin dağılımındaki farklılıklardan kaynaklanan test puan dağılımlarının farklılıęını ayarlayabilmek için ortak deęişken kullanmanın uygulamada çok yararlı olduęundan bahsetmişlerdir.

### **Özet**

Literatür incelendięinde denk olmayan gruplarda ortak maddeye dayalı desenlerin dięer eşitleme desenlerine göre daha fazla kullanıldığı, Kernel eşitleme yöntemlerinin ve ortak deęişkene dayalı desenlere ise daha az rastlandığı görülmektedir. Ortak maddeye dayalı desenlerde kullanılan ortak madde sayısının, örneklem büyüklüęünün, varsayımların ihlalinin eşitleme doğruluęunda önemli bir parametre olduęu görülmektedir. Ortak deęişkene dayalı desenlerde ise deęişkenlerin eşitleme yapabilmek için kullanılabileceęi ve istatistiksel olarak nasıl modele dâhil edileceęi üzerinde durulmuştur. Kernel eşitleme çalışmalarında daha çok eşitleme yöntemleri karşılaştırılmıştır. Simülasyon çalışmaları DOG-OM desende parametrelerden veri üretilerek yapılsa da DOG-OD desende daha çok gerçek veriler ve gerçek verilerden örneklem seçimi ile oluşturulan veri setleri kullanılmıştır. Özellikle veri setinde var olan deęişkenler analizlere dâhil edilmiştir. Bu deęişkenlerin yanlılık varsayımı ve test puanları ile ilişkili olması dışında (ki bu durumun da ihlal edildięi örnekler de yer almaktadır, cinsiyet, okul kademesi gibi) hangi özelliklerine göre seçilmesi gerektięi ile ilgili herhangi bir çalışmaya rastlanmamıştır.

## Bölüm 3

### Yöntem

Bu bölümde araştırmanın deseni, evreni, örnekleme, veri toplama süreci, verilerin analizinden bahsedilmiştir.

#### Araştırmanın Deseni

Bu çalışmada denk olmayan gruplarda ortak değişken deseninde gözlenen puan Kernel eşitleme sonucu elde edilen eşitleme hatası, eşitlenmiş ve ham puanlar arasındaki fark puanların yüzde görelî hataları ve fark puanlarının beklenen ortalamaları ortak değişkenin özelliklerine göre nasıl farklılık gösterdiği incelenmiştir. Çalışma bu yönü ile betimsel bir araştırma niteliğindedir. Betimsel araştırmalar ilişkiyi ya da farkı ortaya koyabilmek için sonuçların sadece olduğu gibi ortaya koyulmasına yönelik yapılan çalışmalardır (Fraenkel ve ark., 2012). Yine bu çalışmada DOG-OD deseninde Kernel eşitleme sonuçlarını ve hatalarına ortak değişken özelliklerinin etkisi incelenmiştir. Bu yönü ile de araştırmanın yordayıcı korelasyonel çalışma olduğu söylenebilir. Nicel araştırma türlerinden biri olan yordayıcı korelasyonel çalışmalar ise birden fazla bağımsız değişkenlerden yola çıkarak yine birden fazla bağımlı değişkeni yordamayı amaçlamaktadır (Büyüköztürk ve ark., 2020; Fraenkel ve ark., 2012).

#### Araştırmanın Evreni ve Örnekleme

Araştırmada ilgili probleme yanıt bulabilmek için denk olmayan gruplarda ortak madde ve ortak değişken desenlerinde eşitleme sonuçlarını karşılaştırmak hedeflendiğinden, bu koşulları sağlayacak olan veri setlerine ulaşılmaya çalışılmıştır. TIMSS 2019 izleme çalışmasının matematik bilişsel testi verileri ve öğrencilere uygulanan anketlerde öğrencilerin geçmiş özelliklerine ait değişkenlerin (background variables) verileri kullanılmıştır. Verilere IEA'nın veri tabanından IEA IDB Analiz Programı (v.4.0) kullanılarak ulaşılmıştır (<https://timss2019.org/international-database/>). TIMSS uygulamalarının

tamamında tabakalı örnekleme yöntemi kullanılmaktadır. Önce araştırmaya katılacak okullar, sonra ise o okullardan sınıflar seçilmektedir (Martin ve ark., 2020).

Örneklem yeterince büyük olduğunda, formlar arasındaki eşitleme, popülasyon düzeyindeki eşitlemeyi temsil edebilir; örneklem küçük olduğunda ise eşitleme, popülasyondan önemli ölçüde farklı olabilir (Kim ve ark., 2006). Bu nedenle kullanılan ilk veri setinde, küçük örnekleme sahip farklı ülkelerin veri setlerinin ortak değişkenlerin niteliklerinin farklı olabileceği ihtimali ile birleştirilememesi ve popülasyonlar arasındaki farklılıkların eşitleme sonuçlarındaki etkisinin kontrol altına alınamaması gibi nedenlerle araştırma ikinci bir veri seti üzerinden tekrarlanmıştır. İlk veri seti ile ilgili yapılan çalışmalar Aşama 1, ikinci veri seti ile ilgili tekrarlanan çalışmalar Aşama 2 olarak isimlendirilmiştir.

### **Aşama 1**

Birinci aşamada, TIMSS 2019 izleme çalışmasının 8. sınıf matematik bilişsel testi verileri ve öğrencilerin anketlere verdikleri yanıtlara ilişkin veriler (background variables) kullanılmıştır. Toplam 14 kitapçığın yer aldığı veri setinde ortak madde sayısı en fazla olan 3 ve 4. kitapçıklara ait veriler tercih edilmiştir. Türkiye verisine ek olarak Türkiye ile aynı maddeleri alan ülkeler arasından TIMSS 2019 ülkelere ait başarı ortalamaları sonuçlarına göre üç ülke daha seçilmiştir. Başarı ortalaması Türkiye'den ( $\bar{X}=496$ ) anlamlı farklılık göstermeyen İsveç ( $\bar{X}=503$ ), anlamlı olarak yüksek olan Amerika Birleşik Devletleri ( $\bar{X}=515$ ) ve anlamlı olarak düşük olan Fransa ( $\bar{X}=483$ ) örneklemden kişi sayıları dikkate alınarak seçilmiştir. Ülkelerin başarı sıralamaları ise büyükten küçüğe Amerika Birleşik Devletleri (12.), İsveç (16.), Türkiye (20.) ve Fransa (22.) şeklindedir. Veri setinde 3 ve 4. kitapçıklardaki toplam ortak madde sayısı 12, ortak olmayan madde sayısı 14'tür. İlgili kitapçığı alan kişi sayıları Tablo 3'te verilmiştir.

**Tablo 3** Veri setinde yer alan kişi sayıları

Ülkeler	Kitapçık numaraları	
	3	4
Türkiye (TR)	289	290
Fransa (FC)	277	281
İsveç (SW)	282	287
ABD (US)	626	606

Tablo 3 incelendiğinde Türkiye, Fransa ve İsviçre veri setlerinde kişi sayısının 290'dan küçük, ABD veri setinde ise 600'den büyük olduğu görülmektedir. En fazla katılımcının ABD'den olduğu, diğer ülkelerin ise katılımcı sayılarının hem ülkeler bazında hem de 3 ve 4 numaralı kitapçıklarda birbirine yakın olduğu söylenebilir.

## Aşama 2

İkinci aşamada kullanılacak veri seti seçilirken hem kitapçıklardaki kişi sayısı yüksek hem de TIMSS ölçek orta noktası olan 500 ortalama başarı puanının üzerinde yer alan bir ülkeye ait veri seti olmasına özen gösterilmiştir. Bu nedenle TIMSS 2019 izleme çalışmasının Kanada'ya ait 4. sınıf matematik bilişsel testi verileri ve yine öğrencilerin anketlere verdikleri yanıtlara ilişkin veriler (background variables) kullanılmıştır. 14 kitapçıktan P evrenindeki madde sayısı aynı ve ortak madde sayısı görece yüksek olan beş kitapçık (2, 3, 4, 5 ve 6. kitapçıklar), üç eşitleme (2 ile 3; 4 ile 5; 5 ile 6) yapabilmek için seçilmiştir. Kanada'nın başarı ortalaması ( $\bar{X}=512$ ), başarı sıralaması 32'dir. Sınava giren toplam kişi sayısı ise 13 653'tür. 2. kitapçığı alan 942, 3. kitapçığı alan 975, 4 kitapçığı alan 957, 5. kitapçığı alan 964 ve 6. kitapçığı alan 952 kişi analizlere dahil edilmiştir. Kitapçıklarda yalnız 3 ve 4. kitapçıklar kullanıldığında ortak değişkene göre yapılan eşitleme sayısı yol analizi için küçük örneklem oluşturmuştur. Bu nedenle eşitleme sayısı artırılarak yol analizi verisinde örneklem sayısı yüksek tutulmaya çalışılmıştır. Merkezi limit teoremine göre örneklem büyüklüğü geniş olduğunda, veriler normal dağılıma yaklaşmaktadır (Johnson & Wichern, 2002).

## Veri Toplama Süreci

Bu araştırma ikincil veriler üzerinden yürütüldüğünden veri toplama süreci IEA'nın veri tabanından ulaşılan veri setlerinin düzenlenmesi üzerinedir. Veri setleri oluşturulduktan sonra veriler analize uygun hale getirilebilmek için incelenmiş ve düzenlenmiştir. R (for Windows 4.0.1) programlama yazılımının RStudio (version 1.3.959) ara yüzünde temel paket ile birlikte dplyr (Wickham ve ark., 2022), data.table (Lim & Qin, 2016), psych (Revelle & Revelle, 2015) paketleri kullanılmıştır.

### Aşama 1

Birinci aşamada ham veri setinden, öncelikle bilişsel test verileri daha sonra ortak değişkenlere ilişkin öğrenci anket verileri seçilmiştir.

**Bilişsel test verileri.** Türkiye, Fransa, İsviçre ve ABD ülkelerinde 3. kitapçığı yanıtlayan kişiler P evreni, 4. kitapçığı yanıtlayan kişiler ise Q evreni olarak ikiye bölünmüş ankor maddelerin formlardaki konumu aynı olacak şekilde maddeler sıralanmıştır. P evreninde ankor maddeler dışında kalan maddeler X, Q evreninde ankor maddeler dışında kalan maddeler Y ve ankor maddeler de A formu olarak isimlendirilmiştir. X formundaki 14, A formundaki 12 ve 23. maddeler kısmî puanlanmaktadır. Bu maddelere verilen yanlış yanıtlar 0, kısmî doğrular ve tam doğru yanıtlar ise (%50 ve üzerinde doğru yanıtlayanlar) 1 ile kodlanarak maddeler kullanılmıştır. Hem X ve Y hem de A formlarında yerel bağımsızlığı etkileyeceği düşünülen madde takımı şeklindeki (testlet) maddeler de benzer şekilde %50 ve üzerinde doğru yanıtlayanlar 1, diğerleri 0 ile kodlanarak analize dahil edilmiştir. Son durumda X ve Y formlarında 14, A formunda 12 madde kullanılmıştır.

**Ortak değişkenlerin seçimi.** Ankor teste göre yapılan eşitleme sonuçlarını ortak değişkene göre yapılan eşitleme sonuçları ile karşılaştırabilmek için ortak maddeler ortak değişken olarak ele alınmış, farklı kategori sayılarına (3-5-8-13) bölünmüştür. DOG-OD deseninde yapılan eşitlemede ortak değişkenler birer dış ankor olarak eşitlemeye dahil edilebildiğinden ortak maddeler de dış ankor olarak eşitlemeye dahil edilmiştir. TIMSS 2019

uygulanmasında öğrencilere ilişkin özelliklerin yer aldığı veri setindeki toplam 60 kategorik değişkenin ham puanlar ile korelasyonları incelenmiştir. Bu değişkenlerin kategori sayıları 2 ile 6 arasında değişmektedir. Kategori sayısı 2, 5 ve 6 olan (sırayıyla cinsiyet, öğretmenlerin ne kadar sıklıkla matematik ödevi verdiği ve matematik ödevlerine kaç dakika harcadığı) yalnız birer değişken; 3 kategorili 10 ve 4 kategorili 47 değişken bulunmaktadır. Araştırmaya dahil edilen ortak değişkenler seçilirken korelasyonun düşük, orta ve yüksek düzeyde farklılaşmasına ve özellikle test puanları ile mantıksal açıdan da ilişkili olmasına dikkat edilmiştir. Ortak maddelerden dönüştürülen ortak değişkenler dışında, araştırma kapsamında koşullar baz alınarak seçilen 23 ortak değişken Tablo 4'te verilmiştir.

**Tablo 4** Birinci aşamada kullanılan ortak değişkenler

Ortak Değişken	Tanımı	Kategori Sayısı
ankor3	Ankor testten alınan toplam puan 3 kategoriye bölünmüş	3
ankor5	Ankor testten alınan toplam puan 5 kategoriye bölünmüş	5
ankor8	Ankor testten alınan toplam puan 8 kategoriye bölünmüş	8
ankor13	Ankor testten alınan toplam puan 13 kategoriye bölünmüş	13
BSBG13C	Kendisini okula ait hissetme	4
BSBM15	Kendi ödevini yapabilme	4
BSBM16C	Matematiğin sıkıcı olduğunu düşünme	4
BSBM16I	Matematiği favori ders olarak görme	4
BSBM17B	Matematik öğretmeninden kolay öğrendiğini düşünme	4
BSBM17F	Öğretmenin dersleri bağlantılı anlattığını düşünme	4
BSBM19A	Matematikte genellikle iyi olduğunu düşünme	4
BSBM19B	Matematiğin daha zor olduğunu düşünme	4
BSBM19C	Matematikte kendisini güçlü hissetmeme	4
BSBM19D	Matematiği kolay öğrenme	4
BSBM19F	Matematikte problemlerin üstesinden gelmede iyi olma	4
BSBM19H	Matematiğin kendisi için zor olduğunu düşünme	4
BSBM19I	Matematiğin kafa karıştırıcı olduğunu düşünme	4
BSBM20H	Ebeveynlerinin matematiğin önemli olduğunu düşünmeleri	4
BSBM26AA	Matematik öğretmenin ev ödevi verme sıklığı	5
BSBM26BA	Matematik ev ödevlerinde harcanan dakika	6
BSBM27AA	Son 12 ayda ekstra matematik dersi alma durumu	3
BSBM27BA	Ayda kaç ekstra matematik dersi aldığı	4
BSDGHER	Eğitim kaynakları indeksi	3
BSDGSB	Öğrenci zorbalığı indeksi	3
BSDGICM	Matematik derslerinde öğretimin niteliği indeksi	3
BSDGDML	Matematik derslerinde düzensiz davranış indeksi	3
BSDGSCM	Matematikte kendine güvenme indeksi	3

Tablo 4 incelendiğinde ankor maddelerden üretilen ortak değişkenler dışındaki değişkenlerin kategori sayılarının 3 ile 6 arasında olduğu görülmektedir. Bu değişkenlere

ilişkin korelasyonlar TR, FC, SW ve US veri setleri için ayrı olacak şekilde Tablo 5'te verilmiştir.

**Tablo 5** TR, FC, SW ve US veri setlerinde testten alınan toplam puanlar ile P ve Q evrenlerindeki değişkenler arasındaki korelasyon katsayıları

Ortak değişken	TR		FC		SW		US	
	P evreni	Q evreni	P evreni	Q evreni	P evreni	Q evreni	P evreni	Q evreni
ankor3	0,82	0,93	0,65	0,68	0,65	0,77	0,83	0,92
ankor5	0,82	0,89	0,65	0,68	0,62	0,73	0,83	0,90
ankor13	0,81	0,89	0,63	0,68	0,63	0,73	0,80	0,85
ankor8	0,78	0,88	0,63	0,67	0,64	0,76	0,82	0,89
BSBM19B	0,42	0,51	0,40	0,50	0,44	0,52	0,38	0,43
BSBM19C	0,42	0,56	0,46	0,45	0,41	0,50	0,37	0,44
BSBM27BA	0,38	0,29	-0,13	-0,31	-0,28	-0,40	-0,14	-0,24
BSBM19H	0,29	0,52	0,36	0,42	0,44	0,52	0,35	0,41
BSBM19I	0,26	0,48	0,44	0,40	0,36	0,37	0,27	0,31
BSBM16C	0,20	0,43	0,12	0,19	0,38	0,29	0,23	0,23
BSBM26BA	0,11	0,15	-0,25	-0,11	-0,17	-0,18	0,11	0,12
BSDGDML	0,07	-0,11	0,08	-0,06	-0,10	0,05	-0,27	-0,25
BSBM26AA	0,03	0,10	0,10	0,02	0,18	0,10	-0,18	-0,17
BSDGICM	-0,06	-0,32	0,05	-0,15	-0,14	-0,17	-0,17	-0,14
BSBM17F	-0,07	-0,33	-0,05	-0,18	-0,18	-0,24	-0,23	-0,18
BSDGSB	-0,08	-0,12	-0,10	-0,15	-0,21	-0,06	-0,25	-0,23
BSBG13C	-0,09	0,08	-0,03	-0,14	-0,18	-0,11	-0,25	-0,20
BSBM17B	-0,09	-0,22	-0,07	-0,21	-0,15	-0,21	-0,15	-0,15
BSBM16I	-0,23	-0,38	-0,38	-0,35	-0,42	-0,44	-0,26	-0,33
BSBM20H	-0,24	-0,33	-0,04	-0,15	-0,22	0,03	-0,21	-0,17
BSBM19F	-0,31	-0,39	-0,39	-0,50	-0,46	-0,53	-0,33	-0,37
BSBM19D	-0,35	-0,38	-0,39	-0,43	-0,41	-0,45	-0,30	-0,37
BSBM27AA	-0,35	-0,30	0,09	0,32	0,24	0,40	0,20	0,23
BSDGHER	-0,36	-0,46	-0,49	-0,40	-0,40	-0,47	-0,37	-0,39
BSBM15	-0,38	-0,46	-0,25	-0,20	-0,21	-0,35	-0,26	-0,25
BSBM19A	-0,39	-0,46	-0,47	-0,51	-0,40	-0,47	-0,35	-0,38
BSDGSCM	-0,51	-0,62	-0,51	-0,56	-0,44	-0,60	-0,38	-0,48

Not: Kırmızı olan değerlerin mutlak değeri 0,30'un üzerindedir.

Tablo 5 incelendiğinde ankor maddelerden alınan toplam puan 3, 5, 8 ve 13 kategorili ortak değişken olarak kullanıldığı durumda tüm veri setlerinde diğer değişkenlere göre oldukça yüksek korelasyon göstermiştir. Bu değerler TR veri seti için P evreninde -0,51 ile 0,82 ve Q evreninde -0,62 ile 0,93; FC veri seti için P evreninde -0,51 ile 0,65 ve Q evreninde -0,56 ile 0,68; SW veri seti için P evreninde -0,46 ile 0,65 ve Q evreninde -0,60 ile 0,77; US veri seti için P evreninde -0,38 ile 0,83 ve Q evreninde -0,48 ile 0,92 arasındadır.

Bu değerlerin mutlak değerleri incelendiğinde ise ankor maddeler dışında 0,30 ve üzerinde korelasyon gösteren ortak değişkenler genel olarak BSBM19B, BSBM19C, BSBM19H, BSBM19I, BSBM16I, BSBM19F, BSBM19D, BSDGHER, BSBM19A ve BSDGSCM değişkenleridir. Ortak değişkenlerin özelliklerine ilişkin ayrıntılı bilgiler EK-A, EK-B, EK-C ve EK-Ç'de verilmiştir.

## Aşama 2

TIMSS 2019 uygulamasında her kitapçık numara olarak kendisinden sonraki diğer kitapçık ile ortak maddeye sahiptir. Aşama 2'de Kanada verisi için iki kitapçıktan fazla sayıda kitapçık eşitlemeye dahil edilmiştir.

**Bilişsel test verileri.** Kitapçıklarda bulunan bu madde grupları birer form olarak düşünülmüştür. İkinci ve üçüncü kitapçıkta M2, dördüncü ve beşinci kitapçıkta M4, beşinci ve altıncı kitapçıkta M5, ortak maddeleri göstermektedir. Her eşitleme sürecinde kitapçık numarası küçük olan grup P ve büyük olan grup Q evreni olarak isimlendirilmiştir. Kısmî puanlanan maddelere verilen yanlış yanıtlar 0, kısmî doğrular ve tam doğru yanıtlar ise 1 olarak kodlanarak maddeler kullanılmıştır. Tüm formlarda yerel bağımsızlığı etkileyeceği düşünülen madde takımı şeklindeki (testlet) maddeler ise ilgili grupta maddeleri %50 ve üzerinde doğru yanıtlayanlar 1, diğerleri 0 ile kodlanarak analize dahil edilmiştir. Son durumda her bir formdaki ve kitapçıktaki madde sayıları Tablo 6'da verilmiştir.

**Tablo 6** Madde gruplarının kitapçıklara dağılımı ve madde sayıları

Madde Grubu	Kitapçıklar				
	K02	K03	K04	K05	K06
M02	12				
M03	11	11			
M04		12	12		
M05			12	12	
M06				11	11
Toplam	23	23	24	23	23

Tablo 6 incelendiğinde her eşitleme (K02 ile K03, K04 ile K05, K05 ile K06) için P evreninde 12 madde olduğu görülmektedir. Kitapçıklardaki madde sayıları ise 23 ve 24'tür.



**Ortak deęişkenlerin seçimi.** Dış ankor maddeler ortak deęişken olarak ele alınmış, farklı kategori sayılarına (3-5-8-13) bölünmüştür. Veri setindeki toplam 94 kategorik deęişkenin ham puanlar ile korelasyonları incelenmiş; düşük, orta ve yüksek düzeyde farklılaşmasına dikkat edilmiştir. Bu deęişkenlerin kategori sayıları 2 ile 6 arasında deęişmektedir. 2 kategorili 17, 3 kategorili 12, 4 kategorili 63 ve 5 kategorili 2 deęişken bulunmaktadır. Ortak maddelerden dönüştürülen ortak deęişkenler dışında, araştırma kapsamında koşullar baz alınarak seçilen 36 ortak deęişken Tablo 7’de verilmiştir.

**Tablo 7** İkinci aşamada kullanılan ortak deęişkenler

Ortak Deęişken	Tanımı	Kategori Sayısı
ankor3	Ankor testten alınan toplam puan 3 kategoriye bölünmüş	3
ankor5	Ankor testten alınan toplam puan 5 kategoriye bölünmüş	5
ankor8	Ankor testten alınan toplam puan 8 kategoriye bölünmüş	8
ankor13	Ankor testten alınan toplam puan 13 kategoriye bölünmüş	13
ASBE01BA	Yazmada zorlanma (PC)	2
ASBE01BB	Numara takımını kullanmada zorluk (PC)	2
ASBE01BC	Nesneleri sürüklemeye zorluk (PC)	2
ASBE01BF	Tablet ya da bilgisayarla problem yaşama	2
ASBE02B	Okuldaki matematik görevlerinde tablet ya da bilgisayar kullanma sıklıkları	4
ASBE02C	Okuldaki fen bilimleri görevlerinde tablet ya da bilgisayar kullanma sıklıkları	4
ASBE02D	Test ya da kuizde tablet ya da bilgisayar kullanma sıklıkları	4
ASBG04	Evdeki kitap miktarı	5
ASBG05C	Kendi odası olup olmaması	2
ASBG05D	Evde internet olup olmaması	2
ASBG08	Okulda ne sıklıkla devamsızlık yaptığı	5
ASBG10C	Kendisini okula ait hissetme	4
ASBG10D	Öğretmenlerin adaletli olduğunu düşünme	4
ASBG10E	Bu okula gitmekle duyduğu gurur	4
ASBG11J	Fotoğraflarını çevrim içi paylaşma sıklığı	4
ASBM02A	Matematik öğrenirken eğlenme	4
ASBM02C	Matematiğin sıkıcı olduğunu düşünme	4
ASBM02E	Matematikten hoşlanma	4
ASBM02G	Matematik problemlerinden hoşlanma	4
ASBM02I	Matematiği favori ders olarak görme	4
ASBM03C	Matematikte yanıtların açık olduğunu düşünme	4
ASBM03F	Matematik öğretmeninin tekrar açıklama yaptığını düşünme	4
ASBM04A	Matematikte öğrencilerin dinlemediğini düşünme	4
ASBM04D	Matematikte sessiz olmak için uzun süre beklendiğini düşünme	4
ASBM05A	Matematikte genellikle iyi olduğunu düşünme	4
ASBM05B	Matematiğin daha zor olduğunu düşünme	4
ASBM05C	Sadece matematikte iyi olmadığını düşünme	4
ASBM05D	Matematiği kolay öğrenme	4

ASBM05E	Matematiğin kendisini gergin hissettirdiğini düşünme	4
ASBM05F	Matematikte problemlerin üstesinden gelmede iyi olma	4
ASBM05H	Matematiğin kendisi için zor olduğunu düşünme	4
ASBM05I	Matematiğin kafa karıştırıcı olduğunu düşünme	4
ASDGHRL	Eğitim kaynakları indeksi	3
ASDGSCM	Matematikte kendine güvenme indeksi	3
ASDGSLM	Öğrencilerin matematiği sevme indeksi	3
ASDGSLs	Öğrencilerin fen bilimlerini sevme indeksi	3

Tablo 7 incelendiğinde ankor maddelerden üretilen ortak değişkenler dışındaki değişkenlerin kategori sayılarının 2 ile 6 arasında olduğu görülmektedir. Bu değişkenlere ilişkin korelasyonlar eşitlenen kitapçıklar için ayrı olacak şekilde Tablo 8'de verilmiştir.

**Tablo 8** Eşitlenen kitapçıklara göre testten alınan toplam puanlar ile P ve Q evrenlerindeki değişkenler arasındaki korelasyon katsayıları

Ortak değişken	2 ile 3.kitapçık		4 ile 5.kitapçık		5 ile 6.kitapçık	
	P evreni	Q evreni	P evreni	Q evreni	P evreni	Q evreni
ankor3	0,71	0,71	0,66	0,72	0,75	0,72
ankor5	0,71	0,70	0,68	0,73	0,72	0,69
ankor8	0,71	0,69	0,67	0,72	0,73	0,71
ankor13	0,69	0,69	0,68	0,71	0,72	0,69
ASBM05B	0,36	0,39	0,37	0,36	0,38	0,41
ASBG11J	0,35	0,19	0,22	0,21	0,19	0,15
ASBM05H	0,35	0,39	0,38	0,39	0,36	0,43
ASBM05C	0,33	0,35	0,33	0,33	0,33	0,31
ASBM05I	0,29	0,31	0,35	0,38	0,38	0,37
ASBG04	0,26	0,16	0,22	0,23	0,20	0,20
ASBE02D	0,22	0,15	0,26	0,20	0,18	0,20
ASBM05E	0,22	0,29	0,28	0,29	0,30	0,28
ASBG08	0,19	0,15	0,20	0,24	0,21	0,13
ASBE01BB	0,18	0,21	0,09	0,16	0,11	0,17
ASBE01BA	0,16	0,24	0,19	0,27	0,18	0,22
ASBM04D	0,16	0,08	0,11	0,23	0,21	0,10
ASBE02B	0,15	0,19	0,28	0,12	0,12	0,21
ASBM02C	0,15	0,21	0,18	0,25	0,21	0,20
ASBE02C	0,14	0,21	0,27	0,15	0,16	0,21
ASBE01BF	0,13	0,03	0,24	0,11	0,03	0,26
ASBM04A	0,07	0,00	-0,01	0,10	0,07	0,06
ASBE01BC	0,06	0,21	0,13	0,19	0,21	0,13
ASBG10E	0,01	0,01	0,01	0,04	0,07	0,01
ASBM03C	0,00	-0,07	-0,04	-0,08	-0,01	-0,09
ASDGSLs	-0,02	0,00	0,02	0,01	0,03	0,00
ASBG10C	-0,03	-0,09	-0,10	-0,05	-0,01	-0,09
ASBG10D	-0,07	-0,08	0,00	-0,06	-0,01	-0,07
ASBM03F	-0,10	-0,06	-0,12	-0,04	-0,01	-0,01
ASBG05C	-0,12	-0,07	-0,01	-0,05	0,00	-0,03
ASBM02E	-0,17	-0,25	-0,19	-0,25	-0,26	-0,23
ASBM02G	-0,17	-0,23	-0,13	-0,19	-0,20	-0,21
ASBM02A	-0,18	-0,25	-0,18	-0,26	-0,22	-0,22
ASDGSLM	-0,18	-0,28	-0,17	-0,26	-0,28	-0,27

ASBG05D	-0,20	-0,23	-0,25	-0,23	-0,22	-0,27
ASBM02I	-0,21	-0,30	-0,21	-0,24	-0,26	-0,27
ASBM05D	-0,23	-0,29	-0,29	-0,35	-0,27	-0,31
ASBM05A	-0,27	-0,33	-0,32	-0,36	-0,33	-0,35
ASBM05F	-0,27	-0,29	-0,29	-0,28	-0,26	-0,27
ASDGHRL	-0,27	-0,08	-0,27	-0,16	-0,18	-0,22
ASDGSCM	-0,35	-0,41	-0,44	-0,43	-0,42	-0,44

Not: Kırmızı olan değerlerin mutlak değeri 0,30'un üzerindedir.

Tablo 8 incelendiğinde ankor maddelerden alınan toplam puan 3, 5, 8 ve 13 kategorili ortak değişken olarak kullanıldığı durumda tüm veri setlerinde diğer değişkenlere göre oldukça yüksek korelasyon göstermiştir. Bu değerler 2 ve 3. kitapçıklar için P evreninde -0,35 ile 0,71 ve Q evreninde -0,41 ile 0,71; 4 ve 5. kitapçıklar için P evreninde -0,44 ile 0,68 ve Q evreninde -0,43 ile 0,73; 5 ve 6. kitapçıklar için P evreninde -0,42 ile 0,75 ve Q evreninde -0,44 ile 0,72 arasındadır. Bu değerlerin mutlak değerleri incelendiğinde ise ankor maddeler dışında 0,30 ve üzerinde korelasyon gösteren ortak değişkenler genel olarak ASBM05B, ASBM05H, ASBM05C, ASBM05I, ASBM05A ve ASDGSCM değişkenleridir. Ortak değişkenlere ilişkin ayrıntılı bilgiler EK-H, EK-I ve EK-I'de verilmiştir.

### Verilerin Analizi

IEA IDB (International Database Analyzer) Analiz Programı (v.4.0) ile SPSS veri dosyası olarak elde edilen veri setlerinde analizlerin tümü için R (for Windows 4.0.1) programlama yazılımının RStudio (version 1.3.959) ara yüzü kullanılmıştır. Kitapçıklardaki kayıp veri mekanizmaları incelendiğinde Tamamen Rastgele Kayıp olduğu görülmüştür. Hem maddelere verilen yanıtlarda hem de geçmiş bilgilerde (background variables) veri silme yöntemi kullanıldığında kişi sayısının oldukça azaldığı görülmüştür. Bu nedenle kayıp veri içeren satırların analizden çıkarılması yerine çok değişkenli kayıp veri atama yöntemlerinden biri olan kesirli Hot-deck atama kullanılmıştır (Kim & Fuller, 2004). Bu yöntemde eksik değer, kendisi ile benzer yanıt örüntüsüne sahip olan en yakın aday hücreler arasından bir dizi atanmış değer ile değiştirilir. Atanmış değerler, hücre içi veri homojenliğini korumak için hücrelerle aynı atama hücreleri arasından rastgele seçilir. Orijinal veri yapısını korumak için her bir atanmış değere kesirli ağırlıklar verilir (Hücre

olasılıkları ağırlıklandırma yöntemiyle EM (Expectation Maximization) kullanılarak tahmin edilir.) ve çıktı olarak yalnız bir veri seti elde edilir (Im ve ark., 2018). Bu araştırmada atama yönteminin Im vd (2018) tarafından genişletilmiş bir hâli kullanılmıştır: İlk avantaj bu yeni halinde atama hücrelerinin yeniden yapılmasını gerektirmeden hücreler, atama sürecinin bir yan ürünü olarak maddeler arasındaki korelasyonları korumak için oluşturulur. İkinci olarak da keyfi kayıp örüntülere sahip çok değişkenli eksik veri setlerine de uygulanabilmektedir. Kayıp veri atama işleminden sonra ise tüm veri seti ülkelere ve evrenlere göre ayrılmıştır.

Veri setlerinde öncelikle tek boyutluluk, yerel bağımsızlık, model-veri uyumu varsayımları mirt paketi (Chalmers ve ark., 2015) ve CTT paketi (Willse & Willse, 2018) kullanılarak test edilmiş daha sonra KTK ve MTK'ye dayalı madde güçlük ve ayırt edicilikleri ile güvenilirlik katsayıları incelenmiştir. DOG-OM deseninde dış ankor kullanılarak KTK ve MTK'ye dayalı olarak son tabakalama (PSE) Kernel eşitleme yöntemi ise kequate paketi (Andersson ve ark., 2022) ile analiz edilmiştir.

Ham puanlar ile aralarındaki ilişkiye göre seçilen ortak değişkenlere ankor maddelerden dönüştürülen ortak değişkenler de dahil edilmiştir. Ankor maddelerden elde edilen toplam puan kategorik hâle dönüştürülürken z puanlarından yararlanılmıştır. Gauss eğrisinin altında kalan bölge üç kategori için %33, beş kategori için %20, sekiz kategori için %12,5 ve 13 kategori için %7,69'luk alanlara ayrılmıştır. ankor3, ankor5, ankor8 ve ankor13 olarak isimlendirilmiştir.

P ve Q evrenlerinden elde edilen yanıtlarda kategorilerdeki kişi sayılarının dağılımları arasındaki fark için iki örneklem Kolmogorov-Smirnov fark testi kullanılmıştır. Tüm değişkenlerde P ve Q evrenlerinde frekans dağılımlarından hesaplanan D değerlerinin maksimumu D kritik değeri ile karşılaştırılmış, istatistiksel olarak fark olmadığı görülmüştür. Bu nedenle D değerlerinin toplamı (D) alınarak evrenler arasındaki farkın bir ölçüsü olarak kabul edilmiştir. Aşama 1 için bu değer TR veri setinde 0,02 (BSBM20H ve BSDGDML) ile 0,68 (ankor13) arasında, FC veri setinde 0,01 (BSBM15) ile 0,38 (ankor13) arasında, SW

veri setinde 0,03 (BSBM19A ve BSDGSB) ile 0,89 (BSBM26AA) arasında, US veri setinde 0,01 (BSDGDML) ile 0,44 (ankor13) arasındadır. Aşama 2 için ise 2 ve 3. kitapçıklara ait veri setinde 0,00 (ASBG05D, ASBE01BA ve ASBG05C) ile 0,11 (ankor13) arasında, 4 ve 5. kitapçıklara ait veri setinde 0,00 (ASBG11J ve ASBG05D) ile 0,34 (ankor13) arasında, 5 ve 6. kitapçıklara ait veri setinde 0,00 (ASBG05D) ile 0,09 (ankor13) arasındadır.

Ortak değişkenler (kategorik) ile gözlenen puanlar (sürekli) arasındaki ilişkiler polyserial korelasyon katsayısı (Revelle & Revelle, 2015) kullanılarak hesaplanmıştır. Katsayılar P ve Q evrenlerinde negatif ve pozitif, orta ve düşük korelasyon gösteren değişkenler ile P ve Q evrenlerinde korelasyonları farklılaşan değişkenler incelenmiştir. Değişkenler bu ilişkilere göre seçildiğinden veri toplama sürecinde ortak değişkenlerin seçimi başlığında anlatılmıştır.

Seçilen ortak değişkenlerin dağılımları için çarpıklık ve basıklık katsayıları incelenmiştir. Aşama 1 için bu değer TR veri setinde P evreninde çarpıklık değerleri -0,67 ile 1,97 ve basıklık değerleri -1,79 ile 3,5 arasında, Q evreninde çarpıklık değerleri -0,53 ile 2,12 ve basıklık değerleri -1,71 ile 4,12 arasındadır. Her iki evrende de BSBM26BA, BSDGDML, BSBM26AA, BSBM19A, BSDGHER, BSBM19D ve ankor3 değişkenleri normal dağılımdan sapma göstermemiştir. FC veri setinde P evreninde çarpıklık değerleri -2,24 ile 2,18 ve basıklık değerleri -1,50 ile 3,7 arasında, Q evreninde çarpıklık değerleri -2,65 ile 3,22 ve basıklık değerleri -1,71 ile 9,59 arasındadır. Her iki evrende de BSBM20H, BSDGDML, BSBM17F, BSBM17B, ankor13, BSBM19A, BSDGICM, BSBM16C, BSBM19F ve BSBM19D değişkenleri normal dağılımdan sapma göstermemiştir. SW veri setinde P evreninde çarpıklık değerleri -1,48 ile 1,97 ve basıklık değerleri -1,32 ile 3,09 arasında, Q evreninde çarpıklık değerleri -1,63 ile 2,16 ve basıklık değerleri -1,43 ile 4,13 arasındadır. Her iki evrende de BSDGHER, BSBG13C, BSBM17B, BSBM17F, BSBM19A, BSBM26BA, ankor13, BSBM26AA, BSBM19D, BSBM19F ve BSBM19I değişkenleri normal dağılımdan sapma göstermemiştir. US veri setinde P evreninde çarpıklık değerleri -1,77 ile 2,32 ve basıklık değerleri -1,51 ile 5,28 arasında, Q evreninde çarpıklık değerleri -1,92 ile 2,56 ve

basıklık değerleri -1,65 ile 6,10 arasındadır. Her iki evrende de BSDGHER, BSBM26BA, BSBM26AA, BSBG13C, BSDGDML, BSDGICM, BSBM19D, BSBM19F değişkenleri normal dağılımdan sapma göstermemiştir. Aşama 2 için 2 ve 3. kitapçıklara ait veri setinde P evreninde çarpıklık değerleri -4,57 ile 5,42 ve basıklık değerleri -1,65 ile 27,45 arasında, Q evreninde çarpıklık değerleri -4,84 ile 5,87 ve basıklık değerleri -1,76 ile 32,45 arasındadır. Her iki evrende de ASBM05D, ASBM02E, ASBM05C, ankor13, ASBM05F, ASBM02G, ASBG04, ASBM04A, ASBE02D, ASDGSLS, ASBM05H ve ASBM05E değişkenleri normal dağılımdan sapma göstermemiştir. 4 ve 5. kitapçıklara ait veri setinde P evreninde çarpıklık değerleri -4,76 ile 4,71 ve basıklık değerleri -1,99 ile 23,83 arasında, Q evreninde çarpıklık değerleri -4,80 ile 4,59 ve basıklık değerleri -2,00 ile 24,53 arasındadır. Her iki evrende de ankor13, ASBG04, ASBM05C, ASBE02D, ASBM04A, ASBM05H, ASBM02G, ASBM02C, ASBM02E, ASBM05F, ASDGSLS ve ASBM05D değişkenleri normal dağılımdan sapma göstermemiştir. 5 ve 6. kitapçıklara ait veri setinde P evreninde çarpıklık değerleri -4,80 ile 4,59 ve basıklık değerleri -2,00 ile 24,53 arasında, Q evreninde çarpıklık değerleri -4,87 ile 4,43 ve basıklık değerleri -1,99 ile 23,61 arasındadır. Her iki evrende de ankor13, ASBE02D, ASBM05F, ASBG04, ASBM05C, ASBM02E, ASBM02A, ASBM05H, ASBM02G, ASBM02C, ASDGSLS, ASBM04A ve ASBM05D değişkenleri normal dağılımdan sapma göstermemiştir.

Ortak değişkenlere göre maddelerin DMF içerip içermeme durumları incelenmiştir. Tek biçimli DMF belirlemek için çok kategorili puanlanan verilerde kullanılabilen ve KTK'ye dayalı yöntemlerden biri olan genelleştirilmiş Lojistik Regresyon kullanılmıştır. Etki büyüklüğü için difR paketinde (Magis ve ark., 2020) yer alan kriterler baz alınmıştır. A, ihmal edilebilir; B, orta düzeyde ve C, geniş etki büyüklüğünü ifade etmek üzere

*Zumbo & Thomas (ZT):*  $0 \rightarrow 'A' \rightarrow 0,13 \rightarrow 'B' \rightarrow 0,26 \rightarrow 'C' \rightarrow 1$  ve

*Jodoin & Gierl (JG):*  $0 \rightarrow 'A' \rightarrow 0,035 \rightarrow 'B' \rightarrow 0,07 \rightarrow 'C' \rightarrow 1$  şeklindedir.

Aşama 1 için elde edilen DMF sonuçları ülkelere göre EK-D, EK-E, EK-F, EK-G'de verilmiştir. Sonuçlar incelendiğinde genel olarak evrenlerde var olan DMF düzeylerinin A ve

B düzeyinde olduğu, yalnız US veri setindeki bir maddede C düzeyinde DMF olduğu görülmüştür. A düzeyi DMF kabul edilebilir düzeyde olduğundan yalnız B ve C düzeyindekiler için 1, A düzeyindekiler için 0 kodlanarak DMF değişkeni elde edilmiştir. Eşitleme sürecinde DMF içeren maddeler analizden çıkarılmamıştır. Aşama 2'de ise 14 kitapçığın tamamı incelenmiş B ve C düzeyinde DMF'ye rastlanmamıştır. Bu nedenle sonuçlar değerlendirilirken DMF parametresi kullanılamamıştır.

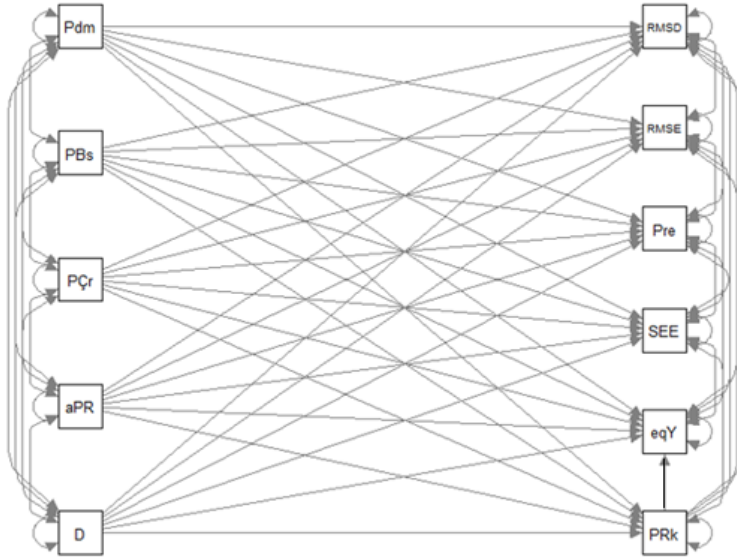
Ortak değişkenler belirlendikten sonra DOG-OD deseninde KTK'ye dayalı Kernel eşitleme yapılmıştır. Bu yöntemde frekansları yordamak için kurulan polinomial regresyon modelinde (generalized linear model) modelin uygunluğu McFadden'in pseudo  $R^2$  (Jackman ve ark., 2015) ile ifade edilmektedir. Tüm değişkenlerde kurulan eşitleme modelinde McFadden  $R^2$  değerlerinin 0,30 ve üzerinde olduğu görülmüştür.

Aşama 1'de dört ülke veri setinde DOG-OM deseninde KTK ve MTK'ye dayalı Kernel eşitleme (MTK: 2x1, KTK: 4x1); 27 ortak değişken (ankor maddelerden ortak değişken yapılanlar dahil) ile yapılan DOG-OD deseninde KTK'ye dayalı Kernel eşitleme (4x27) olmak üzere toplam 114 eşitleme gerçekleştirilmiştir. Aşama 2'de veri setinde (5 kitapçık 3 eşitleme) DOG-OM deseninde KTK ve MTK'ye dayalı Kernel eşitleme (3x2); 40 ortak değişken (ankor maddelerden ortak değişken yapılanlar dahil) ile yapılan DOG-OD deseninde KTK'ye dayalı Kernel eşitleme (3x40) olmak üzere toplam 126 eşitleme gerçekleştirilmiştir.

Eşitleme sonuçlarının değerlendirilmesi için eşitlenmiş puanların ortalamaları ( $E(Y_x)$ ), eşitleme hatalarının ortalamaları (SEE), eşitlenmiş puanlar ile ham puanlar arasındaki farkların dağılımı ile bu dağılımdan elde edilen ortalama hatalar (RMSE) ile eşitleme doğruluğu (RMSD), eşitlenen değerlerin dağılımlarının momentleri ile eşitlenmiş puan dağılımlarının 10. momentleri arasındaki farklar (10. PRE) incelenmiştir. Ortak değişkenin hangi özelliklerinin eşitleme sonuçlarını etkilediğini belirlemek için eşitleme sonuçları ve ortak değişken ile kurulan modeldeki açıklanan varyans oranı ( $R^2$ ) birer içsel değişken olarak düşünülmüştür. Koşullar ve eşitleme sonuçları Aşama 1 için EK-A,

EK-B, EK-C ve EK-Ç'de; Aşama 2 için EK-H, EK-I ve EK-İ'de verilmiştir. Üçüncü araştırma sorusunda yer alan parametreler (ortak değişkenin özellikleri (D, korelasyon, Dmf, çarpıklık, basıklık), açıklanan varyans dışında) ise dışsal değişken olarak ele alınarak yol analizi yapılmıştır. Aşağıda P ve Q evrenleri için ayrı ayrı kurulan modelde P evrenine ilişkin yol modeli örnek olarak ifade edilmiştir.

**Şekil 2** Yol analizinde kurulan model



Şekil 2 incelendiğinde olası tüm yol katsayılarının modele dahil edildiği görülmektedir. Fakat modelin tanımlanabilmesi için bazı yol katsayıları sınırlandırılarak (constrained parameter) yazılım tarafından belirli sınırlar içinde kestirilmesi sağlanmıştır. Böylece hangi içsel değişkenlerin dışsal değişkenleri etkilediğinin, diğer bir deyişle değişkenlerin anlamlı olup olmadığı ve birlikte eşitleme sonuçlarında açıkladıkları varyans oranının belirlenmesi amaçlanmıştır. Model uyumları değerlendirilirken  $\chi^2 / sd$  değeri 3'ten küçük; RMSEA ve SRMR değerlerinin 0,08'den küçük; CFI ve TLI değerlerinin 90'dan büyük olması model uyumunun iyi olduğunun göstergesi olarak kabul edilmiştir (Hu & Bentler, 1999; Kline, 2011). P ve Q evrenlerindeki değişkenler ve her iki aşama için yapılan yol analizine ilişkin model veri uyumu sonuçları Tablo 9'da verilmiştir.



**Tablo 9** P ve Q evrenlerindeki değişkenlerle yapılan yol analizine ilişkin model veri uyumu sonuçları

		$\chi^2$	sd	p	$\chi^2/sd$	CFI	TLI	RMSEA	SRMR
Aşama 1	P	30,870	19	0,042	1,625	0,971	0,932	0,078	0,066
	Q	20,380	19	0,372	1,073	0,987	0,982	0,027	0,060
Aşama 2	P	17,859	17	0,398	1,051	0,989	0,989	0,021	0,076
	Q	20,013	17	0,274	1,177	0,988	0,985	0,039	0,077

Tablo 9 incelendiğinde  $\chi^2/sd$  değerlerinin 2'den küçük, RMSEA ve SRMR değerlerinin 0,08'den küçük, CFI ve TLI değerlerinin 0,93'ten büyük olduğu görülmektedir. Dolayısı ile kurulan yol modellerinin analiz için uygun olduğu söylenebilir.

Değişkenlerdeki uç değerler ( $|z|>3$ ) analizden çıkarıldıktan sonra bağımsız değişkenlerin dağılımlarının normal dağılımdan sapma göstermesi ve örneklem büyüklüğünün küçük olması (<200) (Yıldırım ve ark., 2018) nedeniyle lavaan paketinde (Rosseel, 2012) yol analizinde robust WLS yöntemlerinden biri olan DWLS (Diagonally Weighted Least Square) yöntemi tercih edilmiştir. Yol katsayılarının anlamlılıkları ve açıklanan varyans oranları incelenmiştir. Grafikler için semPlot paketi (Epskamp ve ark., 2017) kullanılmıştır.

### **Aşama 1**

İlk veri setine ilişkin tek boyutluluk, yerel bağımsızlık, model-veri uyumu, madde güçlük ve ayırt edicilikleri ile güvenilirlik katsayıları bu kısımda verilmiştir.

**Tek boyutluluk ve Yerel bağımsızlık.** Araştırma kapsamında ortak değişkenlerin niteliği ülkelere göre değişebileceğinden her bir ülke için 12 maddenin (tüm ankorlar) dâhil edildiği veri setinde tek boyutluluk incelenmiştir. P ve Q evrenlerinde toplam 26 madde kullanılmıştır. Tetrakorik korelasyon matrisi kullanılarak elde edilen özdeğerler, açıklanan varyans oranları ve güvenilirlikleri için KR-20 katsayıları aşağıdaki Tablo 10'da verilmiştir.

**Tablo 10** Veri setlerine ilişkin tek boyutluluk ve KR-20 değerleri

Ülke	Evren	Özdeğerler			Özdeğerlerin oranı		Açıklanan varyans	KR-20
		1.	2.	3.	1/2	2/3		
TR	P	10,095	1,617	1,382	6,24	1,17	0,37	0,87
	Q	12,796	1,638	1,218	7,81	1,34	0,47	0,91
FC	P	6,659	2,019	1,741	3,30	1,16	0,23	0,76
	Q	8,585	1,731	1,650	4,96	1,05	0,31	0,83
SW	P	7,667	1,748	1,637	4,39	1,07	0,27	0,79
	Q	9,805	1,841	1,539	5,33	1,20	0,36	0,84
US	P	10,705	1,665	1,129	6,43	1,47	0,39	0,88
	Q	12,404	1,209	1,095	10,26	1,10	0,46	0,90

Tablo 10 incelendiğinde birinci özdeğerin ikinci özdeğere oranı ile ikinci özdeğerin üçüncü özdeğere oranı arasında büyük bir düşüş olduğu görülmektedir (Lord, 1980). Bu ilk oranın tüm ülkelerde ve evrenlerde 3'ten büyük olması (Büyüköztürk, 2019) tek boyutlu bir yapının göstergesi olarak kabul edilebilmektedir. Bu ülkelerde verilerin tek boyutluluk varsayımını karşıladığı söylenebilir. Açıklanan varyans oranları incelendiğinde değerlerin 0,23'ün üzerinde ve iç tutarlılık anlamındaki güvenilirlik için hesaplanan KR-20 katsayılarının ise 0,76 ve üzerinde olduğu görülmektedir.

Maddelerin yerel bağımsızlık varsayımını karşılayıp karşılamadığını incelemek için hatalar arasındaki ilişkiler ve Yen'in Q3 istatistiği incelenmiştir. Hatalar arasındaki ilişkilerin sifıra yakın ve ikili artık korelasyonların mutlak değerlerinin 0,20'den küçük olması yerel bağımsızlık varsayımının karşılandığını göstermektedir (DeMars, 2016). Q3 istatistiğinin mutlak değerinin 0,20'den büyük olan hücrelerin oranı da yine Tablo 11'de gösterilmiştir.

**Tablo 11** Veri setlerine ilişkin yerel bağımsızlık sonuçları

Ülke	Evren	Q3  > 0,20 oranları*	Minimum	Maksimum
TR	P	0,000	-0,20	0,12
	Q	0,018	-0,27	0,21
FC	P	0,012	-0,23	0,23
	Q	0,003	-0,25	0,17
SW	P	0,009	-0,24	0,13
	Q	0,009	-0,25	0,20
US	P	0,000	-0,15	0,13
	Q	0,000	-0,20	0,13

\*Q3 istatistiği matrisinde yerel bağımsızlığı sağlamayan hücre oranı

Tablo 11 incelendiğinde Yen'in Q3 istatistiklerinin tüm artık korelasyonlar arasında %0 ile %2 arasında beklenen değerlerde olmadığı, dolayısı ile yaklaşık %98 sağlandığı görülmektedir. Bu durumda yerel bağımsızlık varsayımının da karşılandığı söylenebilir.

**Model-Veri Uyumu.** Verilerin model uyumunda MTK'ye göre 1PL ve 2PL model karşılaştırılmıştır. Örneklem büyüklüğü 300'ün altında olan ülkeler için 4PL modelde örneklem büyüklüğü analizler için yeterli olmamıştır. Bununla birlikte 3PL modelde iterasyon sayıları 500'den 2000'e çıkarılmış yine modelin zayıf ya da örneklem sayısının yetersiz olması nedeniyle kestirimlerin tamamlanamadığı görülmüştür. Verilerin hangi model ile uyumlu olduğuna karar vermek için AIC, BIC, ki-kare ve logLikelihood değerleri Tablo 12'de verilmiştir.

**Tablo 12** P ve Q evrenleri için uyum indeksleri

Evren	Ülke	Model	AIC	BIC	logLik	X2	df	p
P	TR	1PL	8139,764	8238,758	-4042,882			
		2PL	8042,041	8232,696	-3969,021	147,723	25	0,00*
	FC	1PL	8135,186	8233,035	-4040,593			
		2PL	8002,961	8191,409	-3949,480	182,226	25	0,00*
	SW	1PL	7990,454	8088,786	-3968,227			
		2PL	7826,661	8016,040	-3861,330	213,794	25	0,00*
	US	1PL	17510,160	17630,020	-8728,080			
		2PL	17316,170	17547,010	-8606,084	243,992	25	0,00*
Q	TR	1PL	7591,667	7690,754	-3768,834			
		2PL	7479,608	7670,441	-3687,804	162,060	25	0,00*
	FC	1PL	7578,466	7676,702	-3762,233			
		2PL	7518,409	7707,603	-3707,204	110,058	25	0,00*
	SW	1PL	7809,169	7907,975	-3877,584			
		2PL	7539,975	7730,268	-3717,988	319,194	25	0,00*
	US	1PL	16280,210	16399,200	-8113,106			
		2PL	16102,630	16331,790	-7999,317	227,576	25	0,00*

\*p<0.05

Tablo 12 incelendiğinde P ve Q evrenlerinde genel olarak 2PL modelin veriye daha uygun olduğu görülmüştür. Bu model için M2 istatistikleri aşağıdaki Tablo 13'te verilmiştir.

Tablo 13 incelendiğinde RMSEA değerlerinin 0,5'ten küçük ve SRMSR değerlerinin 0,05 civarında ve 0,05'ten küçük; TLI ve CFI değerlerinin 0,95'ten büyük olduğu, dolayısı ile SW veri setinde P evreni dışında model uyumunun oldukça iyi olduğu söylenebilir (Kline, 2005).

**Tablo 13** P ve Q evrenlerinde 2PL model için M2 istatistikleri

Evren	Ülke	M2	df	p	RMSEA	SRMSR	TLI	CFI
P	TR	307,284	299	0,358	0,010	0,049	0,998	0,998
	FC	365,018	299	0,005	0,028	0,056	0,954	0,958
	SW	292,339	299	0,598	0,000	0,048	1,004	1,000
	US	411,988	299	0,000	0,025	0,037	0,989	0,990
Q	TR	412,116	299	0,000	0,036	0,054	0,984	0,985
	FC	362,801	299	0,007	0,028	0,053	0,972	0,974
	SW	355,068	299	0,014	0,026	0,051	0,984	0,985
	US	378,076	299	0,001	0,021	0,035	0,994	0,995

**Madde Güçlük ve Ayırt Edicilikleri ile Güvenirlik Katsayıları.** Ankor maddelerin testten elde edilen puana dahil edilmediği (dış ankor) durumda hem KTK'ye hem de MTK'ye dayalı elde edilen madde güçlük ve ayırt edicilik değerleri ile güvenirlik katsayıları Tablo 14 ve Tablo 15'te verilmiştir.

**Tablo 14** MTK'ye dayalı kestirilen madde güçlük ve ayırt edicilik parametreleri ile güvenirlik katsayıları

		b parametreleri				a parametreleri				marginal_rxx	ampirik_rxx
		Min	Maks	Ort.	S.S	Min	Maks	Ort.	S.S		
P	TR	-0,834	2,360	<b>0,660</b>	0,881	0,569	1,857	1,234	0,378	0,77	0,78
	FC	-0,945	<b>14,720</b>	2,580	4,543	0,123	1,509	0,716	0,445	<b>0,63</b>	<b>0,64</b>
	SW	<b>-17,918</b>	<b>23,399</b>	0,248	8,215	-0,094	2,072	0,828	0,561	<b>0,68</b>	<b>0,68</b>
	US	-1,099	5,920	0,502	1,757	0,244	1,893	1,273	0,485	0,79	0,79
Q	TR	-1,069	3,642	0,367	1,200	0,795	2,982	1,577	0,590	0,81	0,82
	FC	-1,201	<b>8,992</b>	1,170	2,565	0,223	3,059	1,225	0,733	0,74	0,74
	SW	-1,619	<b>44,350</b>	3,606	11,785	0,004	3,639	1,485	1,083	0,80	0,81
	US	-0,839	2,154	0,185	0,834	0,879	3,143	1,666	0,686	0,83	0,84

Tablo 14 incelendiğinde ortalama güçlüklerin P evrenlerinde 0,248 ile 2,580, Q evrenlerinde 0,185 ile 3,606 arasında; ortalama ayırt ediciliklerin ise P evrenlerinde 0,716 ile 1,273 arasında, Q evrenlerinde ise 1,225 ile 1,666 arasında değiştiği görülmektedir. Minimum ve maksimum değerlerde ise ayırt edicilik parametreleri P evrenlerinde -0,094 ile 2,072 arasında, Q evrenlerinde 0,004 ile 3,639 arasında; güçlük (konum) parametreleri P evrenlerinde -17,918 ile 23,399 arasında ve Q evrenlerinde -1,619 ile 44,350 arasındadır. Baker (2016) güçlük parametrelerinin teorik olarak  $-\infty$  ile  $+\infty$  arasında değer alabileceğini belirtmiş olsa da madde parametrelerinin çok yüksek ya da çok düşük olmasının varsayımların karşılanmadığı durumlarda yanlış kestirimler nedeniyle ortaya çıkabileceğini

belirtmiştir. Bu veri setlerinde ise özellikle kişi sayısının MTK için düşük olmasının bu duruma yol açabileceği düşünülmektedir. Bu nedenle MTK'ye dayalı eşitleme yöntemlerinin yalnız TR ve US veri setlerinde kullanılmasına karar verilmiştir. Yine testlerin güçlük parametrelerinin her ülke veri seti için P ve Q evrenlerinde birbirinden farklı olduğu görülmektedir. En az farklılaşma TR veri setinde iken en fazla farklılaşmanın SW veri setindedir. P evrenlerinde uygulanan testler SW dışında Q evrenlerinde uygulanan testlerden daha zor olduğu söylenebilir. Güvenirlik katsayılarında ise madde parametrelerinin kestirim hataları ile hesaplanan marjinal güvenirlik katsayılarının ve yetenek parametresinin kestirim hataları ile hesaplanan ampirik güvenirlik katsayılarının P evrenlerinde FC ve SW veri setlerinde görece düşük olduğu söylenebilir. Diğer veri setlerinde ise katsayıların 0,70'in üzerinde olması ölçümlerin güvenirliliğinin iyi olduğunun göstergesidir. Tablo 15'te madde güçlük ve ayırt edicilikleri ile güvenirlik katsayılarının KTK'ye göre sonuçları verilmiştir.

**Tablo 15** KTK'ye dayalı elde edilen madde güçlük ve ayırt edicilikleri ile güvenirlik katsayıları

		Güçlük (p)				Ayırt Edicilik (r)				KR-20	omega
		Min	Maks	Ort.	S.S	Min	Maks	Ort.	S.S		
P	TR	0,149	0,716	0,381	0,150	0,204	0,503	0,390	0,090	0,78	0,79
	FC	0,094	0,690	0,380	0,165	0,026	0,412	0,209	0,128	<b>0,52</b>	<b>0,59</b>
	SW	0,128	0,848	0,490	0,208	-0,008	0,436	0,243	0,127	<b>0,59</b>	<b>0,63</b>
	US	0,193	0,756	0,486	0,176	0,082	0,525	0,392	0,133	0,78	0,79
Q	TR	0,038	0,759	0,431	0,216	0,187	0,632	0,441	0,109	0,82	0,84
	FC	0,093	0,822	0,390	0,252	0,086	0,469	0,308	0,098	<b>0,69</b>	0,73
	SW	0,171	0,840	0,415	0,221	0,003	0,584	0,352	0,164	0,74	0,77
	US	0,091	0,713	0,467	0,188	0,301	0,681	0,469	0,109	0,83	0,85

Tablo 15 incelendiğinde ortalama güçlüklerin P evrenlerinde 0,380 ile 0,490, Q evrenlerinde 0,390 ile 0,467 arasında; ortalama ayırt ediciliklerin ise P evrenlerinde 0,209 ile 0,392 arasında, Q evrenlerinde ise 0,308 ile 0,469 arasında değiştiği görülmektedir. Minimum ve maksimum değerlerde ise ayırt edicilik parametreleri P evrenlerinde -0,008 ile 0,525 arasında, Q evrenlerinde 0,003 ile 0,681 arasında; güçlük parametreleri P evrenlerinde 0,094 ile 0,848 arasında ve Q evrenlerinde 0,038 ile 0,840 arasındadır. KTK'ye göre ülkeler bazında P ve Q evrenlerinde uygulanan testlerin ortalama madde güçlüklerinin

birbirine yakın olduğu görülmektedir. Hem madde güçlüğü ile hesaplanan KR-20 hem de faktör yükleri ile hesaplanan Omega güvenilirlik katsayıları, MTK'ye benzer olarak SW ve FC ülkelerinde P evrenlerinden elde edilen ölçümlerde görece düşüktür. Diğer veri setlerinde ise katsayıların 0,70 civarında ve üzerinde olması ölçümlerin güvenilirliğinin iyi olduğunun göstergesidir.

## Aşama 2

İkinci veri setine ilişkin tek boyutluluk, yerel bağımsızlık, model-veri uyumu, madde güçlük ve ayırt edicilikleri ile güvenilirlik katsayıları bu kısımda verilmiştir.

**Tek boyutluluk ve Yerel bağımsızlık.** Araştırma kapsamında her bir kitapçık için tüm ankor maddelerin dâhil edildiği veri setlerinde tek boyutluluk incelenmiştir. Tetrakorik korelasyon matrisi kullanılarak elde edilen özdeğerler, açıklanan varyans oranları ve güvenilirlikleri için KR-20 katsayıları aşağıdaki Tablo 16'da verilmiştir.

**Tablo 16** Kitapçıklara göre veri setlerine ilişkin tek boyutluluk ve KR-20 değerleri

Kitapçık	Özdeğerler			Özdeğerlerin oranı		Açıklanan varyans	KR-20
	1.	2.	3.	1/2	2/3		
K02	7,167	1,272	1,166	5,63	1,09	0,28	0,81
K03	7,364	1,431	1,160	5,15	1,23	0,29	0,81
K04	7,691	1,184	1,157	6,50	1,02	0,29	0,82
K05	7,428	1,284	1,233	5,79	1,04	0,29	0,82
K06	7,632	1,253	1,187	6,09	1,06	0,30	0,82

Tablo 16 incelendiğinde birinci özdeğerin ikinci özdeğere oranı ile ikinci özdeğerin üçüncü özdeğere oranı arasında büyük bir düşüş olduğu görülmektedir (Lord, 1980). Bu ilk oranın tüm kitapçıklarda 3'ten büyük olması (Büyüköztürk, 2019) tek boyutlu bir yapının göstergesi olarak kabul edilebilmektedir. Bu durumda verilerin tek boyutluluk varsayımını karşıladığı söylenebilir. Açıklanan varyans oranları incelendiğinde değerlerin 0,28'in üzerinde ve iç tutarlılık anlamındaki güvenilirlik için hesaplanan KR-20 katsayılarının ise 0,81 ve üzerinde olduğu görülmektedir.

Maddelerin yerel bağımsızlık varsayımını karşılayıp karşılamadığını incelemek için hatalar arasındaki ilişkiler ve Yen'in Q3 istatistiği incelenmiştir. Hatalar arasındaki ilişkilerin

sıfıra yakın ve ikili artık korelasyonların mutlak değerlerinin 0,20'den küçük olması yerel bağımsızlık varsayımının karşılandığını göstermektedir (DeMars, 2016). Q3 istatistiğinin mutlak değerinin 0,20'den büyük olan hücrelerin oranı da yine Tablo 17'de gösterilmiştir.

**Tablo 17** Veri setlerine ilişkin yerel bağımsızlık sonuçları

Evren	Q3  > 0,20 oranları*	Minimum	Maksimum
K02	0,000	-0,15	0,08
K03	0,000	-0,15	0,18
K04	0,000	-0,17	0,08
K05	0,000	-0,18	0,15
K06	0,004	-0,18	0,23

\*Q3 istatistiği matrisinde yerel bağımsızlığı sağlamayan hücre oranı

Tablo 17 incelendiğinde Yen'in Q3 istatistiklerinin tüm artık korelasyonlar arasında yaklaşık %0 oranında beklenen değerlerde olmadığı, dolayısı ile tamamında yerel bağımsızlık varsayımının da karşılandığı söylenebilir.

**Model-Veri Uyumu.** Verilerin model uyumunda MTK'ye göre 1PL ve 2PL model karşılaştırılmıştır. Bununla birlikte 3PL ve 4PL modelde iterasyon sayıları 500'den 2000'e çıkarılmış yine modelin zayıf ya da örneklem sayısının yetersiz olması nedeniyle kestirimlerin tamamlanamadığı görülmüştür. Verilerin hangi model ile uyumlu olduğuna karar vermek için AIC, BIC, ki-kare ve logLikelihood değerleri Tablo 18'de verilmiştir.

**Tablo 18** MTK uyum indeksleri

Kitapçık	Model	AIC	BIC	logLik	X2	df	p
K02	1PL	25079,50	25195,86	-12515,75			
	2PL	24886,39	25109,40	-12397,19	237,12	22	0,00*
K03	1PL	25885,34	26002,52	-12918,67			
	2PL	25593,47	25818,06	-12750,74	335,87	22	0,00*
K04	1PL	26701,53	26823,12	-13325,76			
	2PL	26490,73	26724,20	-13197,37	256,79	23	0,00*
K05	1PL	25597,12	25714,03	-12774,56			
	2PL	25366,47	25590,54	-12637,24	274,65	22	0,00*
K06	1PL	25113,13	25229,74	-12532,57			
	2PL	24846,15	25069,64	-12377,07	310,99	22	0,00*

\*p<0,05.

Tablo 18 incelendiğinde evrenlerde 2PL modelin veriye daha uygun olduğu görülmüştür. Bu model için M2 istatistikleri aşağıdaki Tablo 19'da verilmiştir.

**Tablo 19** MTK'ye dayalı 2PL model için M2 istatistikleri

Kitapçık	M2	df	p	RMSEA	SRMSR	TLI	CFI
K02	279,954	230	0,014	0,015	0,029	0,992	0,992
K03	379,455	230	0,000	0,026	0,034	0,977	0,979
K04	351,085	252	0,000	0,020	0,032	0,986	0,987
K05	389,639	230	0,000	0,027	0,033	0,976	0,979
K06	407,264	230	0,000	0,028	0,034	0,974	0,977

Tablo 19 incelendiğinde RMSEA ve SRMSR değerlerinin 0,05'ten küçük; TLI ve CFI değerlerinin 0,97'den büyük olduğu, dolayısı ile model uyumunun mükemmel düzeyde olduğu söylenebilir (Kline, 2005).

**Madde Güçlük ve Ayırt Edicilikleri ile Güvenirlik Katsayıları.** Ankor maddelerin testten elde edilen puana dahil edilmediği (dış ankor) durumda hem KTK'ye hem de MTK'ye dayalı elde edilen madde güçlük ve ayırt edicilik değerleri ile güvenirlik katsayıları Tablo 10 ve Tablo 21'de verilmiştir.

**Tablo 20** MTK'ye dayalı kestirilen madde güçlük ve ayırt edicilik parametreleri ile güvenirlik katsayıları

Evren		b parametreleri				a parametreleri				Marginal rxx	Ampirik rxx
		Min	Maks	Ort.	S.S	Min	Maks	Ort.	S.S		
P	M02	-2,171	2,069	-0,198	1,106	0,636	1,631	1,064	0,363	0,71	0,71
	M04	-0,681	1,803	0,371	0,734	0,699	2,156	1,169	0,466	0,73	0,73
	M05	-1,122	1,789	0,231	0,952	0,345	1,747	0,982	0,336	0,68	0,69
Q	M03	-0,801	2,319	0,503	0,911	0,594	2,268	1,145	0,577	0,71	0,72
	M05	-1,968	0,947	-0,238	0,972	0,547	2,202	1,248	0,466	0,73	0,74
	M06	-0,701	1,621	0,534	0,821	0,378	1,904	1,074	0,443	0,70	0,70

Tablo 20 incelendiğinde ortalama güçlüklerin -0,238 ile 0,534 arasında; ortalama ayırt ediciliklerin 0,982 ile 1,248 arasında değiştiği görülmektedir. Minimum ve maksimum değerlerde ise ayırt edicilik parametreleri 0,345 ile 2,268 arasında; güçlük (konum) parametreleri -2,171 ile 2,319 arasındadır. Baker (2016) güçlük parametrelerinin teorik olarak  $-\infty$  ile  $+\infty$ ; pratikte ise -3 ile +3 arasında değer alabileceğini belirtmiştir. Testlerin güçlük parametrelerinin P ve Q evrenlerinde M02 ve M05'te görece küçük; P ve Q evrenlerinde M04 ve M06 madde gruplarının görece büyük olduğu söylenebilir. Bu durum formların güçlük düzeylerinde farklılaşma olabileceğinin göstergesidir. Güvenirlik katsayılarında ise madde parametrelerinin kestirim hataları ile hesaplanan marjinal



güvenirlilik katsayılarının ve yetenek parametresinin kestirim hataları ile hesaplanan ampirik güvenirlik katsayılarının 0,68 ile 0,74 arasında değiştiği görülmektedir. Tablo 21’de madde güçlük ve ayırt edicilikleri ile güvenirlik katsayılarının KTK’ye göre sonuçları verilmiştir.

**Tablo 21** KTK’ye dayalı elde edilen madde güçlük ve ayırt edicilikleri ile güvenirlik katsayıları

		Güçlük (p)				Ayırt Edicilik (r)					
		Min	Maks	Ort.	S.S	Min	Maks	Ort.	S.S	KR-20	omega
P	M02	0,214	0,787	0,540	0,183	0,211	0,439	0,328	0,082	0,69	0,73
	M04	0,135	0,643	0,423	0,156	0,270	0,475	0,352	0,068	0,72	0,74
	M05	0,174	0,665	0,444	0,157	0,133	0,458	0,307	0,082	0,67	0,70
Q	M03	0,097	0,652	0,407	0,154	0,217	0,473	0,331	0,090	0,69	0,72
	M05	0,227	0,854	0,526	0,188	0,208	0,470	0,359	0,085	0,71	0,74
	M06	0,202	0,607	0,396	0,140	0,152	0,435	0,326	0,084	0,69	0,71

Tablo 21 incelendiğinde ortalama güçlüklerin 0,396 ile 0,540 arasında; ortalama ayırt ediciliklerin ise 0,307 ile 0,359 arasında değiştiği görülmektedir. Minimum ve maksimum değerlerde ise ayırt edicilik parametreleri 0,133 ile 0,475 arasında; güçlük parametreleri 0,097 ile 0,854 arasındadır. KTK’ye göre uygulanan madde gruplarında P evreninde M04 ve Q evreninde M06 görece daha zor; P evreninde M02 ve Q evreninde M05 görece daha kolay olduğu söylenebilir. Bu durum formların güçlük düzeylerinde farklılaşma olabileceğinin göstergesidir. Hem madde güçlüğü ile hesaplanan KR-20 hem de faktör yükleri ile hesaplanan Omega güvenirlik katsayıları 0,67 ile 0,74 arasındadır.

## Bölüm 4

### Bulgular ve Yorum

Bu bölümde her bir araştırma sorusuna yanıt oluşturacak şekilde bulgular ve bulgulara dayalı yapılan yorumlar yer almaktadır. Veri setlerine göre iki aşamaya ayrılan araştırmada bulgular da benzer şekilde sunulmuştur.

#### Aşama 1

##### *Birinci araştırma sorusuna yönelik bulgular*

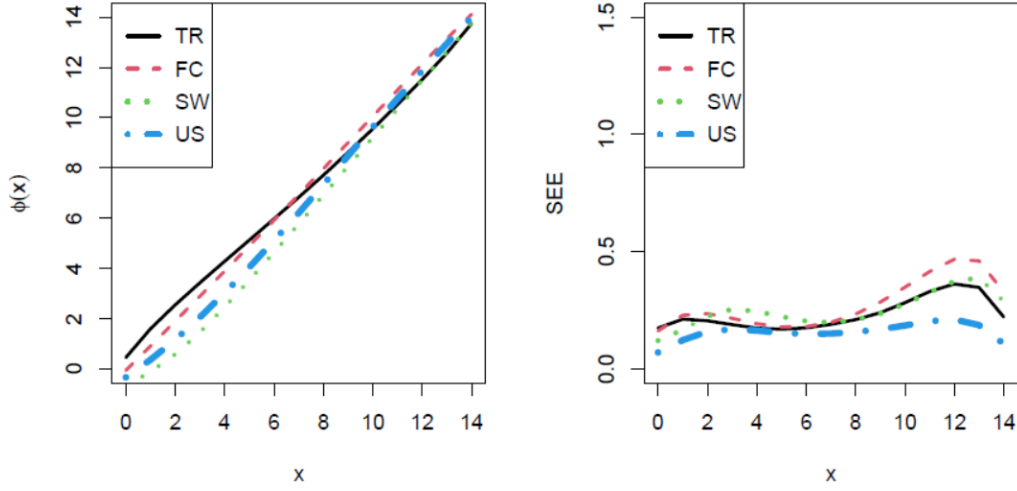
“DOG-OM desende KTK ve MTK’ye dayalı gözlenen puan Kernel eşitleme yapıldığında eşitlenmiş puan ortalamaları (eqYx), eşitleme hataları (SEE), yüzde görelî hatalar (PRE), ham puanlar ile eşitlenmiş puanlar arasındaki fark puanlar ile elde edilen ortalama hatalar ve eşitleme doğruluğu (RMSE ve RMSD) ne düzeydedir?” sorusuna yanıt bulmak için DOG-OM deseninde KTK ve MTK Kernel eşitleme yapılmıştır. Sonuçları Tablo 22’de verilmiştir.

**Tablo 22** DOG-OM deseninde KTK ve MTK Kernel eşitleme sonuçları

	KTK					MTK			
	eqYx	SEE	10.PRE	RMSE	RMSD	eqYx	SEE	RMSE	RMSD
TR	6,964	0,232	-2,287	0,407	0,138	7,622	0,605	0,703	0,564
FC	6,987	0,276	-0,194	0,096	0,008				
SW	6,002	0,249	-1,089	1,088	1,687				
US	6,442	0,157	-0,313	0,655	0,504	6,904	0,450	0,501	0,237

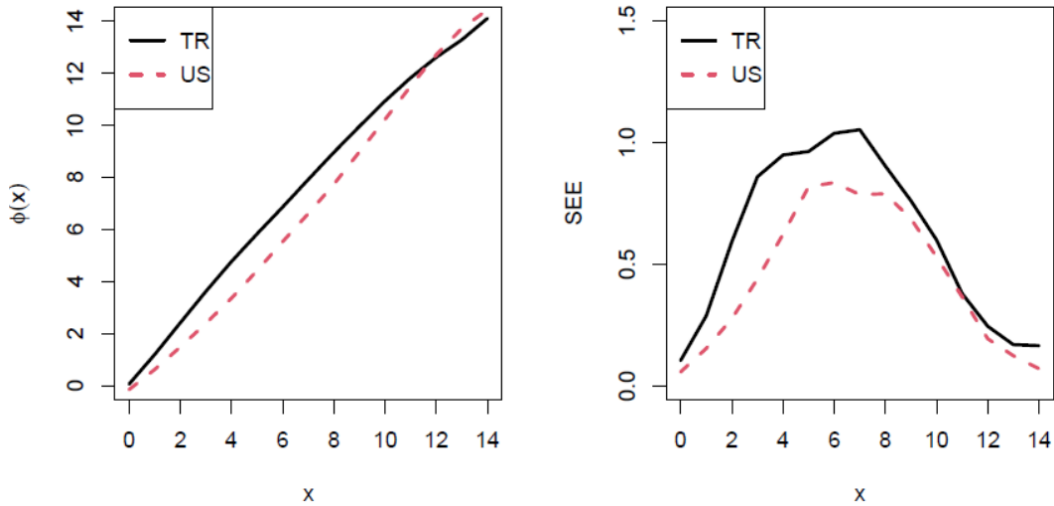
Tablo 22 incelendiğinde eşitlenmiş puan ortalamalarının KTK’de 6,002 ile 6,987 arasında değiştiği, MTK’de ise TR ve US veri setlerinde KTK’ye göre daha yüksek ortalama ve daha yüksek SEE ile kestirildiği görülmüştür. KTK’ye göre 10. momentteki PRE’lerde ise %0,2 ile %2,3 arasında yüzde görelî hata bulunmaktadır. RMSE değerleri incelendiğinde değerlerin 0,096 ile 1,088 arasında değiştiği görülmektedir. MTK’ye göre RMSE ve RMSD değerleri TR verisinde daha yüksek, US verisinde daha düşüktür. Genel olarak FC veri setinde eşitlenmiş puanlar ile ham puanların birbirine daha yakın olduğu söylenebilir. Eşitlenmiş puanlar ve SEE hataların dağılımı aşağıda gösterilmiştir.

**Şekil 3** KTK'ye göre eşitlenmiş puan dağılımları ve SEE'ler



Şekil 3 incelendiğinde testten alınan toplam puan azaldıkça ülkelerdeki eşitlenmiş puanların da birbirinden uzaklaştığı söylenebilir. SEE'lerin dağılımları birbirine benzer olmakla birlikte 10 ham puanın üzerindeki puanlarda üzerinde görece daha yüksek hata ile kestirim yapıldığı görülmektedir. TR ve US veri setlerinde MTK 2PL modele göre yapılan Kernel eşitleme sonuçları Şekil 4'te verilmiştir.

**Şekil 4** MTK'ye göre eşitlenmiş puan dağılımları ve SEE'ler



Şekil 4'e göre TR ve US veri setlerinde yaklaşık 12 puana kadar eşitlenmiş TR puanları daha yüksek iken sonrasında eşitlenmiş US puanları daha yüksektir. SEE'ler incelendiğinde ise TR veri setinin US'den daha yüksek hata ile eşitleme yaptığı görülmektedir. İki ülke veri setine ilişkin hataların orta grupta daha yüksek olduğu

söylenbilir. Bu sonuçlar KTK'ye göre incelendiğinde ise KTK'de daha düşük eşitlenmiş puanlar ve daha düşük standart hatalar (SEE) ile kestirim yapıldığı görülmektedir.

### **İkinci araştırma sorusuna yönelik bulgular**

“KTK'ye dayalı ortak maddelerin birer ortak değişken olarak modele dahil edildiği DOG-OD desende yapılan Kernel eşitleme sonuçları (eqYx, SEE, PRE, RMSE, RMSD) nasıldır? DOG-OM ile bu DOG-OD desende eşitleme fonksiyonları arasındaki farklara ilişkin hatalar (SEED) nasıl dağılmaktadır?”

DOG-OM deseninde yapılan Kernel eşitleme sonuçları ile DOG-OD deseninde elde edilen Kernel eşitleme sonuçlarının karşılaştırılabilmesi için öncelikle ankor maddeler ortak değişken olarak kullanılmıştır. Ankor maddelerden alınan toplam puanın 3, 5, 8 ve 13 kategori yapılarak oluşturulan ortak değişkenlere göre Kernel eşitleme sonuçları Tablo 23'te verilmiştir.

**Tablo 23** Ankor maddelerle oluşturulmuş ortak değişkenler için DOG-OD desende KTK Kernel eşitleme sonuçları

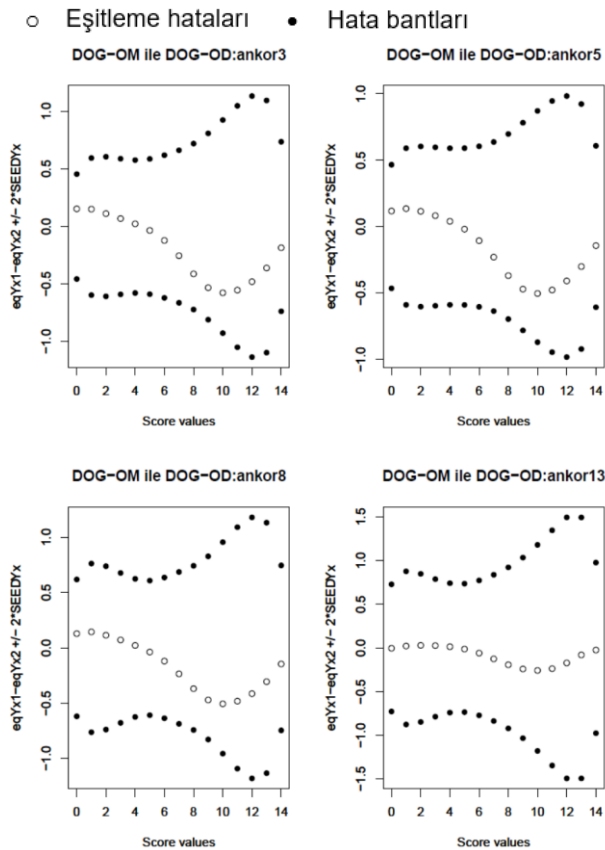
Ülke	Ortak Değişken	eqYx	SEE	10.PRE	RMSE	RMSD
TR	ankor3	7,164	0,251	-0,086	0,233	0,060
TR	ankor5	7,134	0,239	-0,322	0,234	0,056
TR	ankor8	7,138	0,239	-0,303	0,234	0,058
TR	ankor13	7,052	0,247	-1,496	0,324	0,090
FC	ankor3	7,032	0,271	-0,033	0,159	0,012
FC	ankor5	7,059	0,268	-0,311	0,155	0,008
FC	ankor8	7,030	0,269	-0,293	0,156	0,012
FC	ankor13	6,983	10,219	0,304	0,073	0,005
SW	ankor3	5,999	0,251	-1,298	1,088	1,685
SW	ankor5	5,988	0,250	-1,018	1,101	1,728
SW	ankor8	5,974	0,247	-1,142	1,111	1,750
SW	ankor13	6,077	9,595	-0,688	1,052	1,612
US	ankor3	6,472	0,167	-0,218	0,633	0,469
US	ankor5	6,494	0,162	-0,273	0,617	0,441
US	ankor8	6,480	0,160	-0,253	0,637	0,471
US	ankor13	6,454	0,162	-0,266	0,643	0,482

Tablo 23 incelendiğinde eşitlenmiş puanların 5,974 ile 7,164 arasında SEE hataların ise 0,160 ile 10,219 arasında değiştiği görülmektedir. Bu yüksek hatanın 13 kategoriye bölünerek oluşturulan ortak değişkende kategori başına düşen kişi sayılarının az olmasından kaynaklandığı söylenebilir. Her bir ülkede benzer şekilde kategori sayıları

arttıkça hataların arttığı söylenebilir. 10. momentte eşitlenmiş puanlar ile ham puanların dağılımlarının farkının (10. PRE) SW veri seti dışında en az 3 kategorili daha sonra 5 kategorili ortak değişkende olduğu görülmektedir. 10. PRE, RMSE ve RMSD değerlerinin doğru orantılı olarak birbirine çok benzer sonuçlar verdiği söylenebilir.

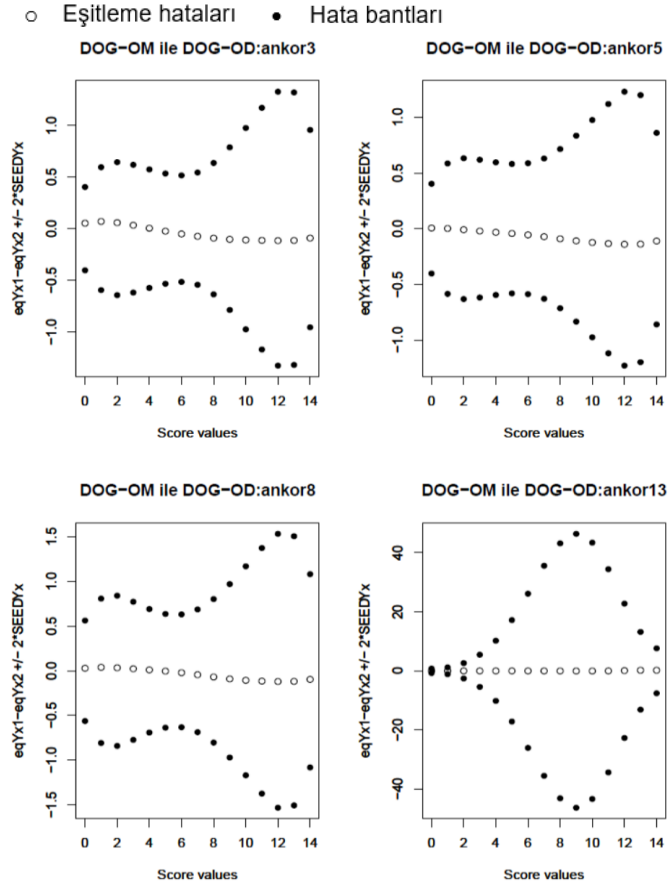
Tablo 22 ve Tablo 23 birlikte değerlendirildiğinde DOG-OM ile DOG-OD desende eşitleme sonuçları karşılaştırıldığında eşitlenmiş puanların ortalamalarının DOG-OD deseninde SW veri seti dışında daha yüksek, SEE'lerin DOG-OD deseninde FC veri seti dışında görece daha düşük, 10. PRE'lerin mutlak değerinin DOG-OD deseninde TR ve US veri setleri dışında daha düşük, RMSE ve RMSD değerlerinin DOG-OD deseninde FC ve SW veri setleri dışında daha düşük olduğu görülmektedir. Elde edilen bu eşitleme fonksiyonları ile DOG-OM deseninde elde edilen eşitleme sonuçlarını karşılaştırmak için SEED'ler incelenmiştir. TR verisi için dağılımların grafiği Şekil 5'te verilmiştir.

**Şekil 5** DOG-OM ile ankor maddelerden elde edilen ortak değişkenlere göre DOG-OD desenlerin eşitleme fonksiyonları arasındaki hata bantları ve eşitleme hataları-TR verisi



Şekil 5 incelendiğinde hata bantlarının genişliklerinin 3 kategoride  $\pm 1,5$ ; 5 kategoride  $\pm 1,0$ ; 8 kategoride  $\pm 1,5$  ve 13 kategoride  $\pm 1,5$  olduğu görülmektedir. Eşitleme hatalarının, eşitleme fonksiyonları arasındaki hata bantları dahilinde olduğu görülmektedir. Bu durumda, TR verisinde DOG-OM ve DOG-OD desenlerinde eşitleme fonksiyonları birbirinden büyük ölçüde farklılaşmamakla birlikte ankor maddenin 5 kategorili hâli en uygundur. Şekil 6'da FC verisine ilişkin sonuçlar verilmiştir.

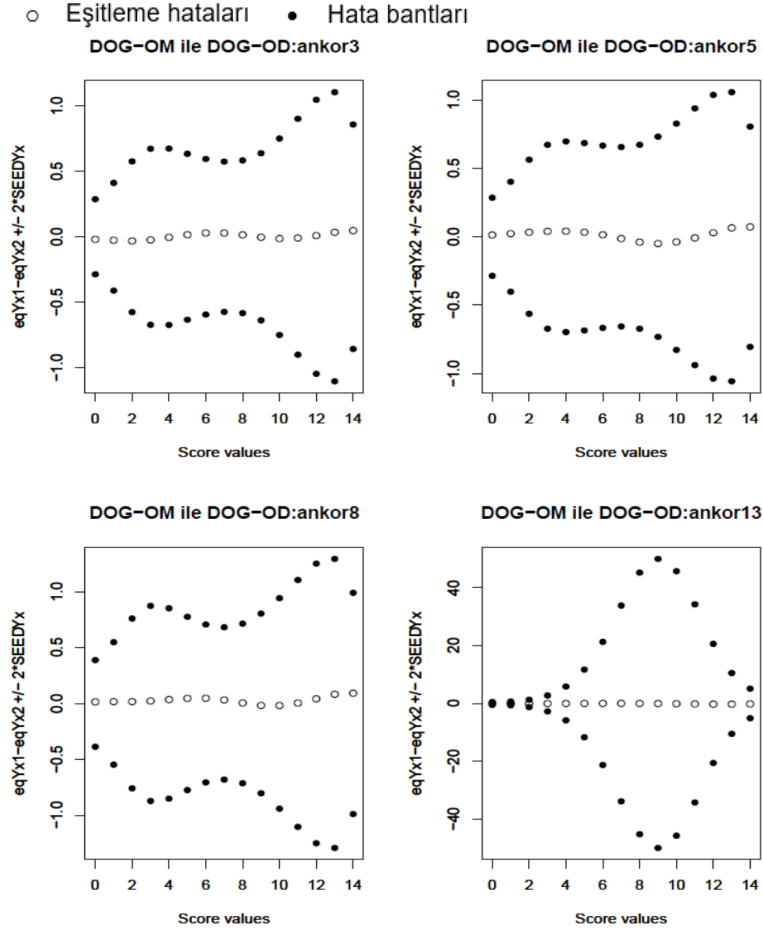
**Şekil 6** DOG-OM ile ankor maddelerden elde edilen ortak değişkenlere göre DOG-OD desenlerin eşitleme fonksiyonları arasındaki hata bantları ve eşitleme hataları-FC verisi



Şekil 6 incelendiğinde hata bantlarının genişliklerinin 3 ve 5 birbirine çok yakın, 8 kategoride  $\pm 1,5$  değerinin biraz dışında fakat 13 kategoride çok yüksek olduğu görülmektedir. Eşitleme hatalarının, eşitleme fonksiyonları arasındaki hata bantları dahilinde olduğu görülmektedir. Bu durumda, FC verisinde DOG-OM ve DOG-OD desenlerinde eşitleme fonksiyonları birbirinden büyük ölçüde farklılaşmamakla birlikte ankor

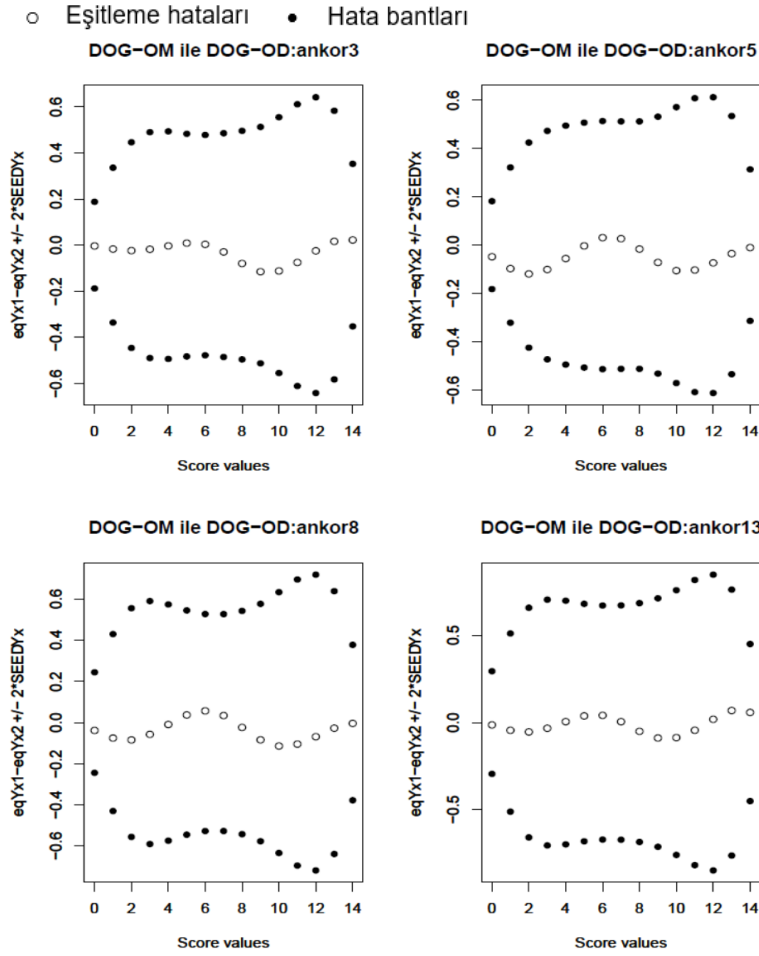
maddenin 3 ya da 5 kategorili hâli en uygundur. Şekil 7'de SW verisine ilişkin sonuçlar verilmiştir.

**Şekil 7** DOG-OM ile ankor maddelerden elde edilen ortak değişkenlere göre DOG-OD desenlerin eşitleme fonksiyonları arasındaki hata bantları ve eşitleme hataları-SW verisi



Şekil 7 incelendiğinde hata bantlarının genişliklerinin FC veri setine benzer olmakla birlikte 3 ve 5 birbirine çok yakın, 8 kategoride  $\pm 1,5$  değerinin biraz dışında fakat 13 kategoride çok yüksek olduğu görülmektedir. Eşitleme hatalarının, eşitleme fonksiyonları arasındaki hata bantları dahilinde olduğu görülmektedir. Bu durumda, SW verisinde DOG-OM ve DOG-OD desenlerinde eşitleme fonksiyonları birbirinden büyük ölçüde farklılaşmamakla birlikte ankor maddenin 3 ya da 5 kategorili hâli en uygundur. Şekil 8'de US verisine ilişkin sonuçlar verilmiştir.

**Şekil 8** DOG-OM ile ankor maddelerden elde edilen ortak değişkenlere göre DOG-OD desenlerin eşitleme fonksiyonları arasındaki hata bantları ve eşitleme hataları-US verisi



Şekil 8 incelendiğinde hata bantlarının genişliklerinin 3 ve 5 kategoride  $\pm 0,6$ ; 8 kategoride  $\pm 0,8$  ve 13 kategoride  $\pm 1,0$  olduğu görülmektedir. Eşitleme hatalarının, eşitleme fonksiyonları arasındaki hata bantları dahilinde olduğu görülmektedir. Bu durumda, US verisinde DOG-OM ve DOG-OD desenlerinde eşitleme fonksiyonları birbirinden büyük ölçüde farklılaşmamakla birlikte ankor maddenin 3 ya da 5 kategorili hâli en uygundur.

Tüm sonuçlar birlikte değerlendirildiğinde ise 3 ve 5 kategori ile oluşturulan ortak değişkenler tüm ülke veri setlerinde DOG-OM desende KTK'ye göre yapılan Kernel eşitleme sonuçları baz alındığında daha doğru eşitleme sonuçları vermiştir.

### **Üçüncü araştırma sorusuna yönelik bulgular**

“DOG-OD desende KTK'ye dayalı gözlenen puan Kernel eşitleme yapıldığında;



- a) ortak deęişkenin P ve Q evrenlerindeki frekans daęılımları arasındaki fark (D);
- b) gözlenen test puanları ile ortak deęişken arasındaki ilişkinin mutlak deęeri (absPR ve absQR);
- c) ortak deęişkenin çarpıklık ve basıklığının mutlak deęeri (absPÇarpıklık, absPBasıklık, absQÇarpıklık ve absQBasıklık);
- d) testteki maddelerin P ve Q evrenlerinde ortak deęişkene göre DMF gösterip göstermemesi (DMF)
- e) P ve Q evrenlerinde ortak deęişken ile kurulan modelin açıklanan varyans oranı (PRkare ve QRkare);

eşitleme sonuçlarını ve hatalarını (eqYx, SEE, PRE, RMSE, RMSD) nasıl etkiler?” sorusunu yanıtlamak için ortak deęişkenlerle yapılan 103 eşitleme sonucunun ve hataları ile e'deki parametrenin içsel, a, b, c ve d'deki parametrelerin dışsal deęişken olduęu yol analizi P ve Q evrenlerinden elde edilen deęişkenler için ayrı ayrı yapılmıştır. Bu deęişkenlerin eşitleme sonuçlarına olan etkisi için P evreninden elde edilen deęişkenlere ait sonuçlar Tablo 24'te verilmiştir.

Tablo 24 incelendiğinde ortak deęişken ile kurulan eşitleme modelin P evrenindeki açıklanan varyans oranını sadece absÇarpıklık deęişkeninin etkiledięi görülmektedir ( $\beta_{\text{absPÇarpıklık}} = 0,554$ ;  $p < 0,05$ ). Modeldeki dięer deęişkenler kontrol altına alındığında, ortak deęişkenin P evrenindeki çarpıklığının mutlak deęeri arttıkça ortak deęişken ile kurulan eşitleme modelin açıklanan varyans oranı da artmaktadır. PRkare deęişkenindeki varyansın %25'i bu model ile açıklanabilmiştir.

**Tablo 24** P evreninden elde edilen değişkenler için yol analizi regresyon katsayıları (Standartlaştırılmış)

Yordanan	Yordayıcı	Regresyon katsayıları	Hata	z	p	Açıklanan varyans
PRkare	D	-0,045	0,095	-0,471	0,638	0,245
	absPR	0,233	0,122	1,901	0,057	
	absPÇarpıklık	0,554	0,115	4,813	0,000*	
	absPBasıklık	-0,072	0,108	-0,665	0,506	
	Pdmf	0,041	0,109	0,372	0,710	
eqYx	D	0,084	0,088	0,951	0,342	0,101
	absPR	-0,175	0,042	-4,145	0,000*	
	absPÇarpıklık	0,079	0,051	1,567	0,117	
	absPBasıklık	0,004	0,049	0,078	0,938	
	Pdmf	0,012	0,093	0,128	0,898	
	PRkare	-0,280	0,088	-3,183	0,001*	
SEE	D	0,692	0,123	5,647	0,000*	0,381
	absPR	-0,175	0,069	-2,522	0,012*	
	absPÇarpıklık	0,079	0,050	1,593	0,111	
	absPBasıklık	0,004	0,050	0,077	0,938	
	Pdmf	0,074	0,055	1,350	0,177	
	PRkare	0,043	0,060	0,710	0,478	
Pre	D	-0,039	0,045	-0,864	0,388	0,051
RMSE	absPR	-0,175	0,042	-4,199	0,000*	0,051
RMSD	absPÇarpıklık	0,079	0,051	1,560	0,119	0,051
	absPBasıklık	0,004	0,049	0,078	0,938	
	Pdmf	0,074	0,059	1,246	0,213	
	PRkare	-0,106	0,049	-2,166	0,030*	

\*p<0,05.

Tablo 24'e göre DOG-OD deseninde Kernel eşitleme sonucu elde edilen eşitlenmiş puan ortalamalarını (eqYx) absPR ve PRkare değişkenleri etkilemektedir ( $\beta_{absPR} = -0,175$ ;  $\beta_{PRkare} = -0,280$ ;  $p < 0,05$ ). Modeldeki diğer değişkenler kontrol altına alındığında, P evreninde gözlenen test puanları ile ortak değişken arasındaki ilişkinin mutlak değeri arttıkça eşitlenmiş puan ortalamaları azalmaktadır. Benzer şekilde, ortak değişken ile kurulan eşitleme modelinin açıklanan varyans oranı arttığında eşitlenmiş puan ortalamaları yine azalmaktadır. PRkare değişkenini yordayan diğer değişkenler ise dolaylı olarak eqYx değişkenindeki varyansa katkı yapmaktadır. Dolayısı ile absPÇarpıklık değişkeninin doğrudan etkisi anlamlı olmamasına rağmen dolaylı etkisi anlamlıdır. eqYx değişkenindeki varyansın %10'u bu model ile açıklanabilmiştir.

Tablo 24'e göre DOG-OD deseninde Kernel eşitleme sonucu elde edilen standart hata ortalamalarını (SEE) D ve absPR değişkenleri etkilemektedir ( $\beta_D = 0,692$ ;  $\beta_{absPR} = -0,175$ ;  $p < 0,05$ ). Modeldeki diğer değişkenler kontrol altına alındığında, ortak değişkenin P ve Q evrenlerindeki frekans dağılımları arasındaki fark arttığında eşitleme hatalarının

ortalamları artmaktadır. Benzer şekilde diğer değişkenler kontrol altına alındığında, P evreninde gözlenen test puanları ile ortak değişken arasındaki ilişkinin mutlak değeri arttıkça eşitleme hatalarının ortalamları azalmaktadır. SEE değişkenindeki varyansın %38'i bu model ile açıklanabilmektedir.

Tablo 24'e göre DOG-OD deseninde Kernel eşitlemedeki yüzde görel hatanın 10. moment ortalamalarını (PRE), fark puanları ile elde edilen beklenen hata ortalamalarını (RMSE) ve yine fark puanları ile hesaplanan eşitleme doğruluğuna ilişkin hatayı (RMSD); absPR ve PRkare değişkenleri etkilemektedir ( $\beta_{absPR} = -0,175$ ;  $\beta_{PRkare} = -0,106$ ;  $p < 0,05$ ). Modeldeki değişkenler kontrol altına alındığında, gözlenen test puanları ile ortak değişken arasındaki ilişkinin mutlak değeri arttıkça yüzde görel hatanın 10. moment ortalamaları; beklenen hata ortalamaları azalmakta; eşitleme doğruluğu ise artmaktadır. Benzer şekilde diğer değişkenler kontrol altına alındığında, ortak değişken ile kurulan eşitleme modelinin açıklanan varyans oranı arttığında yüzde görel hatanın 10. moment ortalamaları; beklenen hata ortalamaları azalmakta; eşitleme doğruluğu ise artmaktadır. PRkare değişkenini yordayan diğer değişkenler ise dolaylı olarak PRE, RMSE ve RMSD değişkenlerindeki varyansa da katkı yapmaktadır. Dolayısı ile absPBasıklık değişkeninin dolaylı etkisi de anlamlıdır. PRE, RMSE ve RMSD değişkenlerinin her birindeki varyansın %5'i bu model ile açıklanabilmektedir. Parametre olarak belirlenen değişkenlerin eşitleme sonuçlarına olan etkisi için Q evreninden elde edilen değişkenlere ait sonuçlar Tablo 25'te verilmiştir.

Tablo 25 incelendiğinde ortak değişken ile kurulan eşitleme modelin Q evrenindeki açıklanan varyans oranını D ve absQÇarpıklık değişkenlerinin etkilediği görülmektedir ( $\beta_D = 0,182$ ;  $\beta_{absPÇarpıklık} = 0,583$ ;  $p < 0,05$ ). Modeldeki diğer değişkenler kontrol altına alındığında, ortak değişkenin P ve Q evrenlerindeki frekans dağılımları arasındaki fark arttıkça ortak değişken ile kurulan eşitleme modelin açıklanan varyans oranı da artmaktadır. Benzer şekilde diğer değişkenler kontrol altına alındığında, ortak değişkenin Q evrenindeki çarpıklığın mutlak değeri arttıkça ortak değişken ile kurulan eşitleme modelin açıklanan

varyans oranı da artmaktadır. QRkare değişkenindeki varyansın %31'i bu model ile açıklanabilmiştir.

**Tablo 25** Q evreninden elde edilen değişkenler için yol analizi regresyon katsayıları (Standartlaştırılmış)

Yordanan	Yordayıcı	Regresyon katsayıları	Hata	z	p	Açıklanan varyans
QRkare	D	0,182	0,083	2,204	0,028*	0,307
	absQR	0,137	0,105	1,308	0,191	
	absQÇarpıklık	0,583	0,113	5,174	0,000*	
	absQBasıklık	-0,031	0,108	-0,285	0,775	
	Qdmf	0,095	0,049	1,925	0,054	
eqYx	D	0,097	0,094	1,028	0,304	0,038
	absQR	-0,109	0,039	-2,813	0,005*	
	absQÇarpıklık	0,045	0,052	0,869	0,385	
	absQBasıklık	0,041	0,053	0,781	0,435	
	Qdmf	-0,066	0,110	-0,598	0,550	
	QRkare	-0,160	0,100	-1,606	0,108	
SEE	D	0,682	0,111	6,159	0,000*	0,384
	absQR	-0,109	0,054	-2,017	0,044*	
	absQÇarpıklık	0,045	0,053	0,859	0,390	
	absQBasıklık	0,041	0,057	0,717	0,473	
	Qdmf	0,046	0,055	0,833	0,405	
	QRkare	-0,080	0,055	-1,447	0,148	
Pre	D	-0,050	0,052	-0,948	0,343	0,042
RMSE	absQR	-0,109	0,038	-2,844	0,004*	0,042
RMSD	absQÇarpıklık	0,045	0,052	0,868	0,385	0,042
	absQBasıklık	0,041	0,053	0,781	0,435	
	Qdmf	0,046	0,060	0,765	0,444	
	QRkare	-0,139	0,047	-2,951	0,003*	

\*p<0,05.

Tablo 25'e göre DOG-OD deseninde Kernel eşitleme sonucu elde edilen eşitlenmiş puan ortalamalarını (eqYx) yalnız absQR değişkeni etkilemektedir ( $\beta_{absQR} = -0,109$ ;  $p < 0,05$ ). Modeldeki diğer değişkenler kontrol altına alındığında, Q evreninde gözlenen test puanları ile ortak değişken arasındaki ilişkinin mutlak değeri arttığında eşitlenmiş puan ortalamaları azalmaktadır. eqYx değişkenindeki varyansın %4'ü bu model ile açıklanabilmiştir.

Tablo 25'e göre DOG-OD deseninde Kernel eşitleme sonucu elde edilen standart hata ortalamalarını (SEE) D ve absQR değişkenleri etkilemektedir ( $\beta_D = 0,682$ ;  $\beta_{absQR} = -0,109$ ;  $p < 0,05$ ). Modeldeki diğer değişkenler kontrol altına alındığında, ortak değişkenin P ve Q evrenlerindeki frekans dağılımları arasındaki fark arttığında eşitleme hatalarının ortalamaları artmaktadır. Benzer şekilde diğer değişkenler kontrol altına alındığında, Q

evreninde gözlenen test puanları ile ortak değişken arasındaki ilişkinin mutlak değeri arttığında eşitleme hata ortalamaları azalmaktadır. SEE değişkenindeki varyansın %38'i bu model ile açıklanabilmiştir.

Tablo 25'e göre DOG-OD deseninde Kernel eşitlemedeki yüzde görelî hatanın 10. moment ortalamalarını (PRE), fark puanları ile elde edilen beklenen hata ortalamalarını (RMSE) ve yine fark puanları ile hesaplanan eşitleme doğruluğuna ilişkin hatayı (RMSD); absQR ve QRkare değişkenleri etkilemektedir ( $\beta_{absQR} = -0,109$ ;  $\beta_{QRkare} = -0,139$ ;  $p < 0,05$ ). Modeldeki değişkenler kontrol altına alındığında, Q evreninde gözlenen test puanları ile ortak değişken arasındaki ilişkinin mutlak değeri arttıkça yüzde görelî hatanın 10. moment ortalamaları; beklenen hata ortalamaları azalmakta; eşitleme doğruluğu ise artmaktadır. Benzer şekilde diğer değişkenler kontrol altına alındığında, ortak değişken ile kurulan eşitleme modelinin açıklanan varyans oranı arttığında yüzde görelî hatanın 10. moment ortalamaları; beklenen hata ortalamaları azalmakta; eşitleme doğruluğu ise artmaktadır. PRkare değişkenini yordayan diğer değişkenler ise dolaylı olarak PRE, RMSE ve RMSD değişkenlerindeki varyansa da katkı yapmaktadır. Dolayısı ile D ve absQÇarpıklık değişkenlerinin dolaylı etkisi de anlamlıdır. PRE, RMSE ve RMSD değişkenlerinin her birindeki varyansın %4'ü bu model ile açıklanabilmiştir.

Sonuç olarak EK-J'de yol şemaları verilen eşitleme sonuçlarını ve eşitleme hatalarını ortak değişkenin P ve Q evrenlerindeki frekans dağılımları arasındaki farkın (D), gözlenen test puanları ile ortak değişken arasındaki ilişkinin mutlak değerinin (absPR ve absQR); ortak değişkenin çarpıklığının (PÇarpıklık, QÇarpıklık), P ve Q evrenlerinde ortak değişken ile kurulan modelin açıklanan varyans oranının (PRkare, QRkare) anlamlı olarak etkilediği görülmüştür. D değeri azaldığında, korelasyonun mutlak değeri arttığında, çarpıklığın mutlak değeri arttığında, eşitleme modelinde açıklanan varyans oranı arttığında eşitleme sonuçları iyileşmektedir. Testteki maddelerin P ve Q evrenlerinde ortak değişkene göre DMF gösterip göstermemesinin ve ortak değişkenlerin basıklıklarının mutlak değerinin ise bu sonuçlar üzerinde doğrudan ya da dolaylı anlamlı bir etkisine rastlanmamış olmasına

rağmen Pdmf ile PÇarpıklık, P ve Q Basıklık ile Çarpıklık değişkenlerinin kovaryansı anlamlıdır. Bu değerler baz alındığında veri setindeki ankor3 (ankor maddelerden üretilen 3 kategorili ortak değişken), ankor5 (ankor maddelerden üretilen 5 kategorili ortak değişken), BSBM19A (Matematikte genellikle iyi olduğunu düşünme), BSBM19B (Matematiğin daha zor olduğunu düşünme), BSBM19C (Matematikte kendisini güçlü hissetmeme), BSDGHER (Eğitim kaynakları indeksi) ve BSDGSCM (Matematikte kendine güvenme indeksi) değişkenlerinin görece bu özelliklere sahip olduğu görülmektedir.

## Aşama 2

### *Birinci araştırma sorusuna yönelik bulgular*

“DOG-OM desende KTK ve MTK’ye dayalı gözlenen puan Kernel eşitleme yapıldığında eşitlenmiş puan ortalamaları (eqYx), eşitleme hataları (SEE), yüzde görelî hatalar (PRE), ham puanlar ile eşitlenmiş puanlar arasındaki fark puanlar ile elde edilen ortalama hatalar ve eşitleme doğruluğu (RMSE ve RMSD) ne düzeydedir?” sorusuna yanıt bulmak için DOG-OM deseninde KTK ve MTK Kernel eşitleme yapılmıştır. 2 ile 3, 4 ile 5 ve 5 ile 6. kitapçıkların eşitleme sonuçları Tablo 26’da verilmiştir.

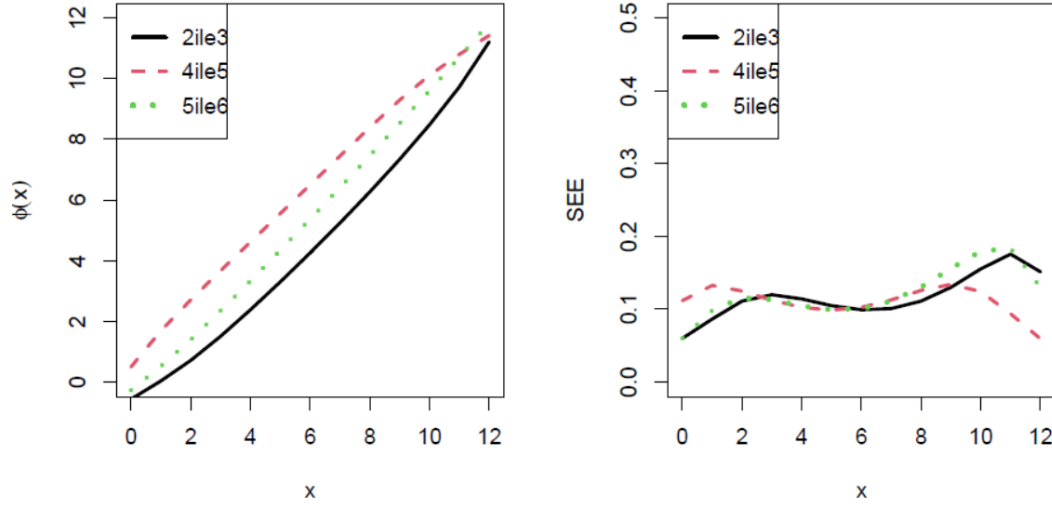
**Tablo 26** DOG-OM deseninde KTK ve MTK Kernel eşitleme sonuçları

	KTK					MTK			
	eqYx	SEE	10.PRE	RMSE	RMSD	eqYx	SEE	RMSE	RMSD
2 ile 3	4,606	0,117	-10,368	1,449	2,408	4,849	0,236	1,275	1,991
4 ile 5	6,361	0,110	2,878	0,524	0,270	6,417	0,239	0,650	0,483
5 ile 6	5,505	0,122	-1,798	0,521	0,350	5,607	0,241	0,455	0,290

Tablo 26 incelendiğinde eşitlenmiş puan ortalamalarının KTK’de 4,606 ile 6,361 arasında değiştiği eşitleme hatalarının ise birbirine yakın olduğu görülmektedir. MTK’de eşitlenmiş puan ortalamaları KTK’ye göre daha yüksek ortalama ve daha yüksek SEE ile kestirilmiştir. 10. momentteki PRE’lerde ise %0,2 ile %10 arasında yüzde görelî hata bulunmaktadır. RMSE değerleri incelendiğinde değerlerin 0,455 ile 1,275 arasında değiştiği görülmektedir. MTK’ye göre RMSE ve RMSD değerleri 2 ile 3.kitapçık ve 5 ile 6. kitapçık

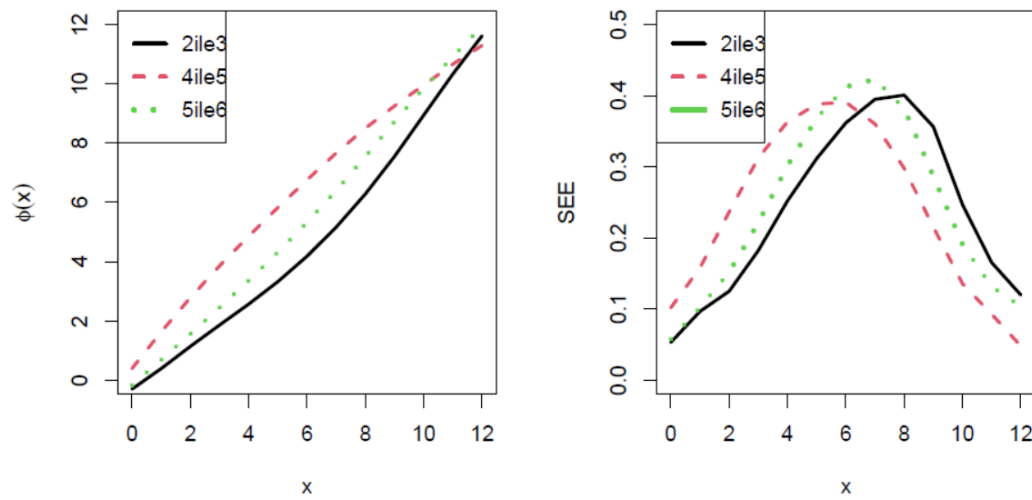
verilerinde daha düşük, 4 ile 5. kitapçık verilerinde daha yüksektir. Eşitlenmiş puanlar ve SEE hataların dağılımı aşağıda gösterilmiştir.

**Şekil 9** KTK'ye göre eşitlenmiş puan dağılımları ve SEE'ler



Şekil 9 incelendiğinde testten alınan toplam puan azaldıkça ve arttıkça eşitlenmiş puanların da birbirine yaklaştığı söylenebilir. SEE'lerin dağılımları birbirine benzer olmakla birlikte 4 ile 5. kitapçıkların eşitlenmesinde 2 ham puanın altında daha yüksek, 10 ham puan üzerinde görece daha düşük hata ile kestirim yapıldığı görülmektedir. MTK 2PL modele göre yapılan Kernel eşitleme sonuçları Şekil 10'da verilmiştir.

**Şekil 10** MTK'ye göre eşitlenmiş puan dağılımları ve SEE'ler



Şekil 10'a göre 4 ile 5. kitapçıkların eşitlenmiş puanları yaklaşık 11 puana kadar daha yüksektir. SEE'ler incelendiğinde ise en yüksek hataların 5 ile 6. kitapçığın eşitlenmesi sonucu elde edildiği görülmektedir. Veri setlerine ilişkin hataların orta grupta daha yüksek olduğu söylenebilir. Düşük puan alan bireylerde 4 ile 5, yüksek puan alan bireylerde 2 ile 3. kitapçığın eşitlenmesinde daha yüksek hatalar görülmektedir. Bu sonuçlar KTK'ye göre incelendiğinde ise KTK Kernel eşitlemenin daha düşük eşitlenmiş puanlar elde edilirken, daha az standart hata (SEE) ile kestirildiği görülmektedir.

### ***İkinci araştırma sorusuna yönelik bulgular***

“KTK'ye dayalı ortak maddelerin birer ortak değişken olarak modele dahil edildiği DOG-OD desende yapılan Kernel eşitleme sonuçları (eqYx, SEE, PRE, RMSE, RMSD) nasıldır? DOG-OM ile bu DOG-OD desende eşitleme fonksiyonları arasındaki farklara ilişkin hatalar (SEED) nasıl dağılmaktadır?”

DOG-OM deseninde yapılan Kernel eşitleme sonuçları ile DOG-OD deseninde elde edilen Kernel eşitleme sonuçlarının karşılaştırılabilmesi için öncelikle ankor maddeler ortak değişken olarak kullanılmıştır. Ankor maddelerden alınan toplam puanın 3, 5, 8 ve 13 kategori yapılarak oluşturulan ortak değişkenlere göre Kernel eşitleme sonuçları Tablo 27'de verilmiştir.

**Tablo 27** Ankor maddelerle oluşturulmuş ortak değişkenler için DOG-OD deseninde KTK Kernel eşitleme sonuçları

Kitapçık	Ortak Değişken	eqYx	SEE	10.PRE	RMSE	RMSD
2 ile 3	ankor3	4,526	0,119	-11,786	1,519	2,605
2 ile 3	ankor5	4,566	0,118	-10,976	1,481	2,504
2 ile 3	ankor8	4,573	0,117	-11,016	1,474	2,467
2 ile 3	ankor13	4,632	0,120	-9,780	1,430	2,361
4 ile 5	ankor3	6,342	0,116	3,200	0,530	0,278
4 ile 5	ankor5	6,310	0,114	3,100	0,517	0,262
4 ile 5	ankor8	6,305	0,113	3,105	0,511	0,255
4 ile 5	ankor13	6,339	0,113	2,957	0,495	0,253
5 ile 6	ankor3	5,454	0,125	-2,606	0,566	0,387
5 ile 6	ankor5	5,478	0,124	-2,211	0,540	0,357
5 ile 6	ankor8	5,463	0,123	-2,397	0,554	0,372
5 ile 6	ankor13	5,504	0,124	-1,772	0,522	0,352



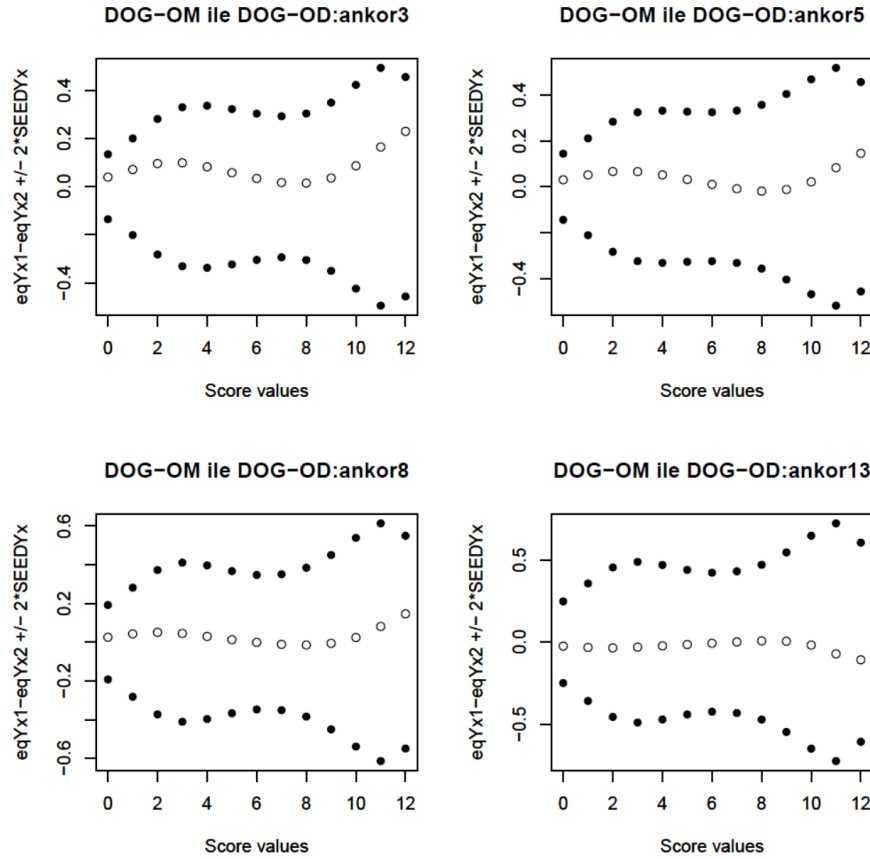
Tablo 27 incelendiğinde eşitlenmiş puanların 4,526 ile 6,342 arasında SEE hataların ise 0,113 ile 0,125 arasında değiştiği görülmektedir. SEE'ler birbirine yakın olmakla birlikte en yüksek SEE değeri 5 ile 6. kitapçıkların eşitlenmesinde kullanılan üç kategorili ortak değişkende elde edilmiştir. 10. momentte eşitlenmiş puanlar ile ham puanların dağılımlarının farkı (10. PRE) en az, 5 ile 6. kitapçıkların eşitlenmesinde kullanılan 13 kategorili ankor maddededir. 10. PRE, RMSE ve RMSD değerlerinin doğru orantılı olarak birbirine çok benzer sonuçlar verdiği söylenebilir.

Tablo 26 ve Tablo 27 birlikte değerlendirildiğinde DOG-OM ile DOG-OD desende eşitleme sonuçları karşılaştırıldığında eşitlenmiş puanların ortalamalarının DOG-OD deseninde 2 ile 3. kitapçıkta 13 kategorili ankor değişkeni dışında daha düşük, SEE'lerin DOG-OD deseninde daha yüksek, 10. PRE'lerin mutlak değerinin DOG-OD deseninde daha yüksek, RMSE ve RMSD değerlerinin DOG-OD deseninde 2 ile 3. kitapçıkların eşitlenmesinde daha yüksek olduğu görülmektedir. Elde edilen bu eşitleme fonksiyonları ile DOG-OM deseninde elde edilen eşitleme sonuçlarını karşılaştırmak için SEED'ler incelenmiştir. 2 ile 3. kitapçıklara ait veri seti için dağılımların grafiği Şekil 11'de verilmiştir.

**Şekil 11** DOG-OM ile ankor maddelerden elde edilen ortak değişkenlere göre DOG-OD desenlerin eşitleme fonksiyonları arasındaki hata bantları ve eşitleme hataları-2 ile 3.

kitapçık eşitleme verisi

- Eşitleme hataları    ● Hata bantları

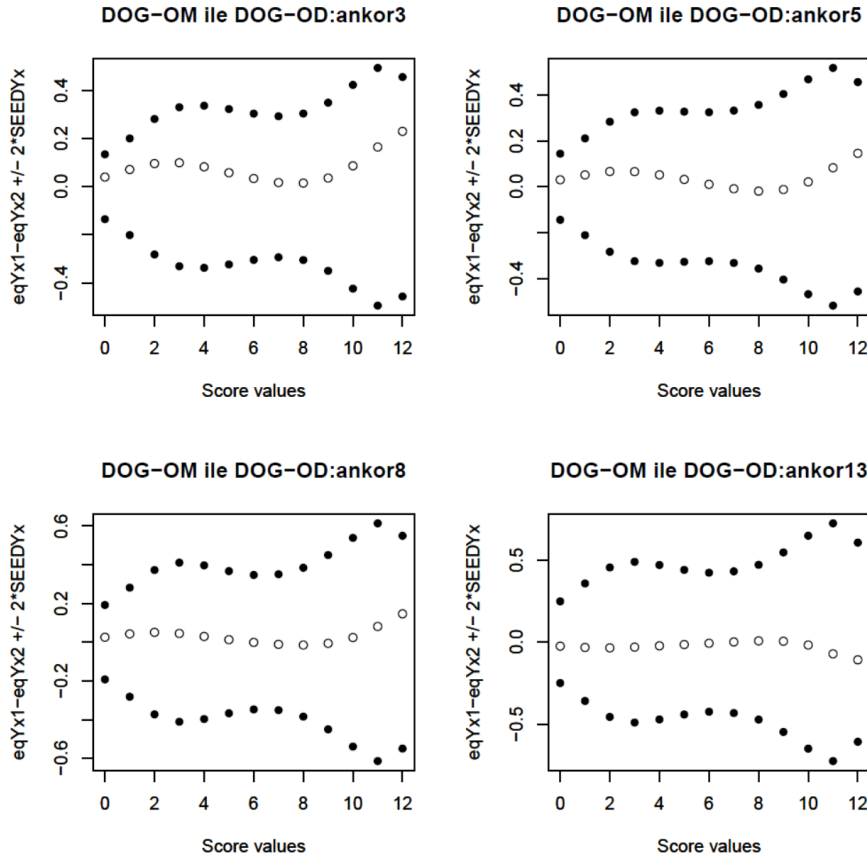


Şekil 11 incelendiğinde hata bantlarının genişliklerinin 3 ve 5 kategoride  $\pm 0,5$ ; 8 ve 13 kategoride  $\pm 0,6$  olduğu görülmektedir. Eşitleme hatalarının, eşitleme fonksiyonları arasındaki hata bantları dahilinde olduğu görülmektedir. Bu durumda, 2 ile 3. kitapçık eşitleme verisinde DOG-OM ve DOG-OD desenlerinde eşitleme fonksiyonları birbirinden büyük ölçüde farklılaşmamakla birlikte ankor maddenin 3 ya da 5 kategorili hâli en uygundur. 4 ile 5. kitapçık eşitleme verisi için dağılımların grafiği Şekil 12’de verilmiştir.

**Şekil 12** DOG-OM ile ankor maddelerden elde edilen ortak değişkenlere göre DOG-OD desenlerin eşitleme fonksiyonları arasındaki hata bantları ve eşitleme hataları-4 ile 5.

kitapçık eşitleme verisi

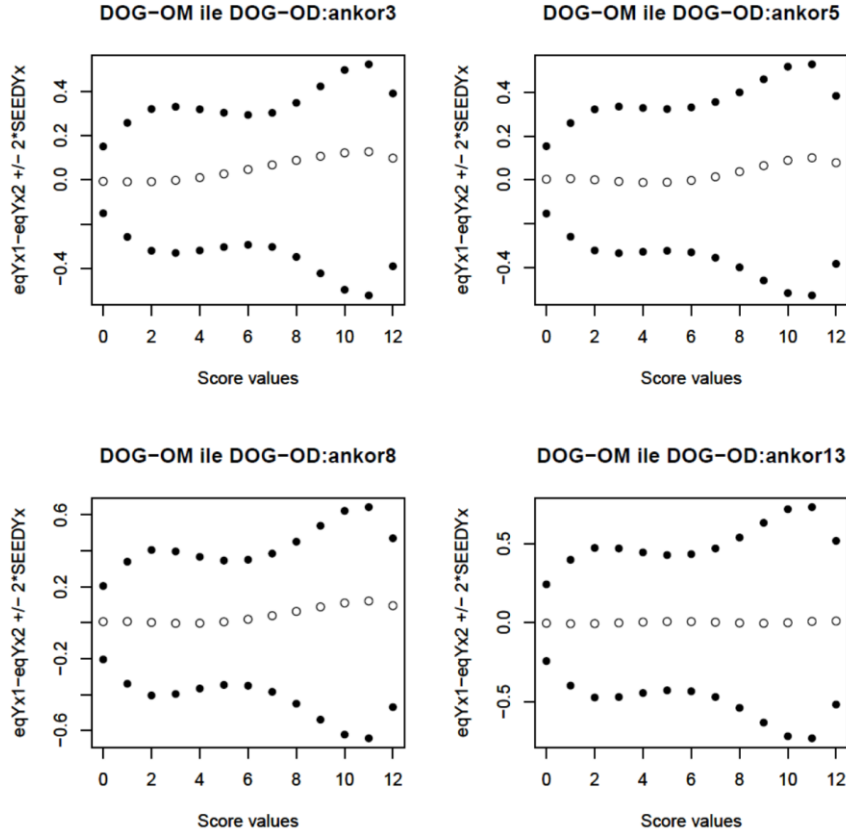
- Eşitleme hataları    ● Hata bantları



Şekil 12 incelendiğinde hata bantlarının genişliklerinin 3 ve 5 kategoride  $\pm 0,5$ ; 8 ve 13 kategoride  $\pm 0,6$  olduğu görülmektedir. Eşitleme hatalarının, eşitleme fonksiyonları arasındaki hata bantları dahilinde olduğu görülmektedir. Bu durumda, 4 ile 5. kitapçık eşitleme verisinde DOG-OM ve DOG-OD desenlerinde eşitleme fonksiyonları birbirinden büyük ölçüde farklılaşmamakla birlikte ankor maddenin 3 ya da 5 kategorili hâli en uygundur. 5 ile 6. kitapçık eşitleme verisi için dağılımların grafiği Şekil 13'te verilmiştir.

**Şekil 13** DOG-OM ile ankor maddelerden elde edilen ortak değişkenlere göre DOG-OD desenlerin eşitleme fonksiyonları arasındaki hata bantları ve eşitleme hataları-5 ile 6. kitapçık eşitleme verisi

- Eşitleme hataları • Hata bantları



Şekil 13 incelendiğinde hata bantlarının genişliklerinin 3 ve 5 kategoride  $\pm 0,5$ ; 8 ve 13 kategoride  $\pm 0,6$  olduğu görülmektedir. Eşitleme hatalarının, eşitleme fonksiyonları arasındaki hata bantları dahilinde olduğu görülmektedir. Bu durumda, 5 ile 6. kitapçık eşitleme verisinde DOG-OM ve DOG-OD desenlerinde eşitleme fonksiyonları birbirinden büyük ölçüde farklılaşmamakla birlikte ankor maddenin 3 ya da 5 kategorili hâli en uygundur.

Tüm sonuçlar birlikte değerlendirildiğinde ise 3 ve 5 kategori ile oluşturulan ortak değişkenler tüm veri setlerinde DOG-OM deseninde KTK'ye göre yapılan Kernel eşitleme sonuçları baz alındığında daha doğru eşitleme sonuçları vermiştir.

### **Üçüncü Araştırma sorusuna yönelik bulgular**

“DOG-OD desende KTK’ye dayalı gözlenen puan Kernel eşitleme yapıldığında;

- ortak değişkenin P ve Q evrenlerindeki frekans dağılımları arasındaki fark (D);
- gözlenen test puanları ile ortak değişken arasındaki ilişkinin mutlak değeri (absPR ve absQR);
- ortak değişkenin çarpıklık ve basıklığının mutlak değeri (absPÇarpıklık, absPBasıklık, absQÇarpıklık ve absQBasıklık);
- P ve Q evrenlerinde ortak değişken ile kurulan modelin açıklanan varyans oranı (PRkare ve QRkare);

eşitleme sonuçlarını ve hatalarını (eqYx, SEE, PRE, RMSE, RMSD) nasıl etkiler?” sorusunu yanıtlamak için ortak değişkenlerle yapılan 117 eşitleme sonucunun ve hataları ile d’deki parametrenin içsel, a, b ve c’deki parametrelerin dışsal değişken olduğu yol analizi P ve Q evrenlerinden elde edilen değişkenler için ayrı ayrı yapılmıştır. Bu değişkenlerin eşitleme sonuçlarına olan etkisi için P evreninden elde edilen değişkenlere ait sonuçlar Tablo 28’de verilmiştir.

**Tablo 28** P evreninden elde edilen değişkenler için yol analizi regresyon katsayıları (Standartlaştırılmış)

Yordanan	Yordayıcı	Regresyon katsayıları	Hata	z	p	Açıklanan varyans
PRkare	D	-0,135	0,066	-2,038	0,042*	0,600
	absPR	0,213	0,084	2,540	0,011*	
	absPÇarpıklık	0,726	0,053	13,591	0,000*	
	absPBasıklık	0,144	0,061	2,343	0,019*	
eqYx	D	0,047	0,104	0,453	0,651	0,012
	absPR	-0,032	0,024	-1,353	0,176	
	absPÇarpıklık	0,128	0,035	3,639	0,000*	
	absPBasıklık	0,003	0,010	0,363	0,717	
	PRkare	-0,092	0,108	-0,851	0,395	
SEE	D	-0,049	0,030	-1,655	0,098	0,042
	absPR	-0,032	0,025	-1,291	0,197	
	absPÇarpıklık	0,128	0,028	4,507	0,000*	
	absPBasıklık	0,003	0,010	0,364	0,716	
	PRkare	0,050	0,083	0,596	0,551	
Pre	D	-0,049	0,030	-1,631	0,103	0,022
RMSE	absPR	-0,032	0,024	-1,350	0,177	0,022
RMSD	absPÇarpıklık	0,128	0,035	3,653	0,000*	0,022
	absPBasıklık	0,003	0,010	0,363	0,717	
	PRkare	-0,198	0,058	-3,399	0,001*	

\*p<0,05.

Tablo 28 incelendiğinde ortak değişken ile kurulan eşitleme modelin P evrenindeki açıklanan varyans oranını D, absPR, absPÇarpıklık ve absPBasıklık değişkeninin etkilediği görülmektedir ( $\beta_D = -0,135$ ;  $\beta_{absPR} = 0,213$ ;  $\beta_{absP\text{Çarpıklık}} = 0,726$ ;  $\beta_{absPBasıklık} = 0,144$ ;  $p < 0,05$ ). Modeldeki diğer değişkenler kontrol altına alındığında, ortak değişkenin P ve Q evrenlerindeki frekans dağılımları arasındaki fark arttıkça ortak değişken ile kurulan eşitleme modelinin açıklanan varyans oranı azalmaktadır. Benzer şekilde P evreninde gözlenen test puanları ile ortak değişken arasındaki ilişkinin mutlak değeri, ortak değişkenin P evrenindeki çarpıklığının ve basıklığının mutlak değeri arttıkça ortak değişken ile kurulan eşitleme modelin açıklanan varyans oranı da artmaktadır. PRkare değişkenindeki varyansın %60'ı bu model ile açıklanabilmiştir.

Tablo 28'e göre DOG-OD deseninde Kernel eşitleme sonucu elde edilen eşitlenmiş puan ortalamalarını (eqYx) yalnız absPÇarpıklık değişkeni etkilemektedir ( $\beta_{absP\text{Çarpıklık}} = 0,128$ ;  $p < 0,05$ ). Modeldeki diğer değişkenler kontrol altına alındığında, P evreninde ortak değişkenin çarpıklığının mutlak değeri arttıkça eşitlenmiş puan ortalamaları artmaktadır. eqYx değişkenindeki varyansın %1'i bu model ile açıklanabilmiştir.

Tablo 28'e göre DOG-OD deseninde Kernel eşitleme sonucu elde edilen standart hata ortalamalarını (SEE) yalnız absPÇarpıklık değişkeni etkilemektedir ( $\beta_{absP\text{Çarpıklık}} = 0,128$ ;  $p < 0,05$ ). Modeldeki diğer değişkenler kontrol altına alındığında, P evreninde ortak değişkenin çarpıklığının mutlak değeri arttıkça eşitleme hataları artmaktadır. SEE değişkenindeki varyansın %4'ü bu model ile açıklanabilmiştir.

Tablo 28'e göre DOG-OD deseninde Kernel eşitlemedeki yüzde görel hatanın 10. moment ortalamalarını (PRE), fark puanları ile elde edilen beklenen hata ortalamalarını (RMSE) ve yine fark puanları ile hesaplanan eşitleme doğruluğuna ilişkin hatayı (RMSD); absPÇarpıklık ve PRkare değişkenleri etkilemektedir ( $\beta_{absP\text{Çarpıklık}} = 0,128$ ;  $\beta_{PRkare} = -0,198$ ;  $p < 0,05$ ). Modeldeki değişkenler kontrol altına alındığında, P evreninde ortak değişkenin çarpıklığının mutlak değeri arttıkça yüzde görel hatanın 10. moment ortalamaları; beklenen hata ortalamaları azalmakta; eşitleme doğruluğu ise artmaktadır. Benzer şekilde diğer

değişkenler kontrol altına alındığında, ortak değişken ile kurulan eşitleme modelinin açıklanan varyans oranı arttığında yüzde görelî hatanın 10. moment ortalamaları; beklenen hata ortalamaları azalmakta; eşitleme doğruluğu ise artmaktadır. PRkare değişkenini yordayan diğer değişkenler ise dolaylı olarak PRE, RMSE ve RMSD değişkenlerindeki varyansa da katkı yapmaktadır. Dolayısı ile D, absPR, absPÇarpıklık ve absPBasıklık değişkenlerinin dolaylı etkisi de anlamlıdır. PRE, RMSE ve RMSD değişkenlerinin her birindeki varyansın %2'si bu model ile açıklanabilmiştir. Parametre olarak belirlenen değişkenlerin eşitleme sonuçlarına olan etkisi için Q evreninden elde edilen değişkenlere ait sonuçlar Tablo 29'da verilmiştir.

**Tablo 29** Q evreninden elde edilen değişkenler için yol analizi regresyon katsayıları (Standartlaştırılmış)

Yordanan	Yordayıcı	Regresyon katsayıları	Hata	z	p	Açıklanan varyans
QRkare	D	-0,124	0,064	-1,947	0,052	0,599
	absQR	0,137	0,086	1,590	0,112	
	absQÇarpıklık	0,726	0,057	12,782	0,000*	
	absQBasıklık	0,107	0,060	1,802	0,072	
eqYx	D	0,032	0,103	0,309	0,757	0,012
	absQR	-0,048	0,022	-2,219	0,026*	
	absQÇarpıklık	0,112	0,032	3,492	0,000*	
	absQBasıklık	-0,006	0,008	-0,792	0,428	
	QRkare	-0,134	0,105	-1,280	0,201	
SEE	D	-0,036	0,030	-1,208	0,227	0,035
	absQR	-0,048	0,023	-2,077	0,038*	
	absQÇarpıklık	0,112	0,026	4,261	0,000*	
	absQBasıklık	-0,006	0,008	-0,785	0,432	
	QRkare	0,050	0,084	0,595	0,552	
Pre	D	-0,036	0,030	-1,195	0,232	0,014
RMSE	absQR	-0,048	0,022	-2,212	0,027*	0,014
RMSD	absQÇarpıklık	0,112	0,032	3,507	0,000*	0,014
	absQBasıklık	-0,006	0,008	-0,792	0,428	
	QRkare	-0,144	0,057	-2,541	0,011*	

\*p<0,05.

Tablo 29 incelendiğinde ortak değişken ile kurulan eşitleme modelin Q evrenindeki açıklanan varyans oranını absQÇarpıklık değişkeninin etkilediği görülmektedir ( $\beta_{\text{absPÇarpıklık}} = 0,726$ ;  $p < 0,05$ ). Modeldeki diğer değişkenler kontrol altına alındığında, ortak değişkenin Q evrenindeki çarpıklığın mutlak değeri arttıkça ortak değişken ile kurulan eşitleme modelin

açıklanan varyans oranı da artmaktadır. QRkare değişkenindeki varyansın %60'ı bu model ile açıklanabilmiştir.

Tablo 29'a göre DOG-OD deseninde Kernel eşitleme sonucu elde edilen eşitlenmiş puan ortalamalarını (eqYx) absQR ve absQÇarpıklık değişkenleri etkilemektedir ( $\beta_{absQR} = -0,048$ ;  $\beta_{absPÇarpıklık} = 0,112$ ;  $p < 0,05$ ). Modeldeki diğer değişkenler kontrol altına alındığında, Q evreninde gözlenen test puanları ile ortak değişken arasındaki ilişkinin mutlak değeri arttığında eşitlenmiş puan ortalamaları azalmaktadır. Benzer şekilde diğer değişkenler kontrol altına alındığında, ortak değişkenin Q evrenindeki çarpıklığın mutlak değeri arttıkça eşitlenmiş puan ortalamaları artmaktadır. eqYx değişkenindeki varyansın %1'i bu model ile açıklanabilmiştir.

Tablo 29'a göre DOG-OD deseninde Kernel eşitleme sonucu elde edilen standart hata ortalamalarını (SEE) absQR ve absQÇarpıklık değişkenleri etkilemektedir ( $\beta_{absQR} = -0,048$ ;  $\beta_{absPÇarpıklık} = 0,112$ ;  $p < 0,05$ ). Modeldeki diğer değişkenler kontrol altına alındığında, Q evreninde gözlenen test puanları ile ortak değişken arasındaki ilişkinin mutlak değeri arttığında eşitleme hataları azalmaktadır. Benzer şekilde diğer değişkenler kontrol altına alındığında, ortak değişkenin Q evrenindeki çarpıklığın mutlak değeri arttıkça eşitleme hataları artmaktadır. SEE değişkenindeki varyansın %4'ü bu model ile açıklanabilmiştir.

Tablo 29'a göre DOG-OD deseninde Kernel eşitlemedeki yüzde görelî hatanın 10. moment ortalamalarını (PRE), fark puanları ile elde edilen beklenen hata ortalamalarını (RMSE) ve yine fark puanları ile hesaplanan eşitleme doğruluğuna ilişkin hatayı (RMSD); absQR, absQÇarpıklık ve QRkare değişkenleri etkilemektedir ( $\beta_{absQR} = -0,109$ ;  $\beta_{absPÇarpıklık} = 0,112$ ;  $\beta_{QRkare} = -0,139$ ;  $p < 0,05$ ). Modeldeki değişkenler kontrol altına alındığında, Q evreninde gözlenen test puanları ile ortak değişken arasındaki ilişkinin mutlak değeri arttıkça yüzde görelî hatanın 10. moment ortalamaları; beklenen hata ortalamaları azalmakta; eşitleme doğruluğu ise artmaktadır. Benzer şekilde diğer değişkenler kontrol altına alındığında, ortak değişken ile kurulan eşitleme modelinin açıklanan varyans oranı arttığında ve Q evreninde ortak değişkenin çarpıklığı azaldığında, yüzde görelî hatanın 10.



moment ortalamaları; beklenen hata ortalamaları azalmakta; eşitleme doğruluğu ise artmaktadır. PRkare değişkenini yordayan diğer değişkenler ise dolaylı olarak PRE, RMSE ve RMSD değişkenlerindeki varyansa da katkı yapmaktadır. Dolayısı ile absPÇarpıklık değişkeninin dolaylı etkisi de anlamlıdır. PRE, RMSE ve RMSD değişkenlerinin her birindeki varyansın %1'i bu model ile açıklanabilmiştir.

Sonuç olarak EK-K'de yol şemaları verilen eşitleme sonuçlarını ve eşitleme hatalarını ortak değişkenin P ve Q evrenlerindeki frekans dağılımları arasındaki farkın (D), gözlenen test puanları ile ortak değişken arasındaki ilişkinin mutlak değerinin (absPR ve absQR); ortak değişkenin çarpıklık ve basırlığının mutlak değerinin (absPÇarpıklık, absPBasıklık, absQÇarpıklık), P ve Q evrenlerinde ortak değişken ile kurulan modelin açıklanan varyans oranının (PRkare, QRkare) anlamlı olarak etkilediği görülmüştür. D değeri azaldığında, korelasyonun mutlak değeri arttığında, çarpıklık ve basırlığın mutlak değeri arttığında, eşitleme modelinde açıklanan varyans oranı arttığında eşitleme sonuçları iyileşmektedir. Bu değerler baz alındığında veri setindeki ankor3 (ankor maddelerden üretilen 3 kategorili ortak değişken), ankor5 (ankor maddelerden üretilen 5 kategorili ortak değişken), ASBM05A (Matematikte genellikle iyi olduğunu düşünme), ASBM05B (Matematiğin daha zor olduğunu düşünme), ASBM05C (Sadece matematikte iyi olmadığını düşünme), ASBM05H (Matematiğin kendisi için zor olduğunu düşünme), ASBM05I (Matematiğin kafa karıştırıcı olduğunu düşünme) ve ASDGSCM (Matematikte kendine güvenme indeksi) değişkenlerinin görece bu özelliklere sahip olduğu görülmektedir.

## Bölüm 5

### Sonuç, Tartışma ve Öneriler

Bu araştırmada ortak maddelerin kullanımının uygun olmadığı ya da eşitleme deseninde yer almadığı durumlarda test eşitleme için ortak bir değişkenin kullanımı ve ortak değişkenin bazı özelliklerinin; eşitlenmiş puanlar, eşitleme hataları, fark puanlardan elde edilen hatalar ile eşitleme doğruluğunu etkileyip etkilemediği incelenmiştir. Bu bölümde ise bulgulara dayalı olarak sonuçlar tartışılmış ve hem araştırmacılara hem de uygulayıcılara önerilerde bulunulmuştur.

#### **Sonuç ve Tartışma**

Bulgulara dayalı olarak sonuç ve tartışma da alt problemlerin sunulduğu sırası ile açıklanmıştır. Aşama 1 ve 2'de elde edilen bulgular birlikte değerlendirilmiştir.

#### ***Birinci alt problem***

DOG-OM desende KTK ve MTK'ye dayalı olarak yapılan Kernel eşitlemede eşitlenmiş puanlar ve eşitleme hataları incelenmiştir. Eşitleme sonucunda hem eşitlenmiş puan ortalamaları hem de SEE'lerin KTK'ye göre MTK'de daha yüksek kestirildiği görülmüştür. Bu bulgu Mutluer'in (2021) KTK'ye dayalı eşit yüzdelikli eşitlemede düzleştirme yapıldığında diğer yöntemlere göre daha az hata ile kestirim yapıldığını gördüğü çalışması ile benzerdir. Yine hem Choi (2019) tek boyutlu ve çok boyutlu MTK eşitleme yöntemlerinde hem de Uysal (2019) karma formatlı testlerde MTK'ye dayalı moment yöntemlerinde gruplar arasındaki yetenek farkı arttıkça hataların daha yüksek kestirildiğini belirtmiştir. Fakat özellikle örneklem büyüklüğünün fazla olduğu çalışmalarda genellikle MTK'ye dayalı karakteristik eğri ya da moment yöntemlerinin daha düşük hatalarla kestirim yaptığı görülmüştür. (Atar ve Yeşiltaş, 2017; Kumlu, 2019; Livingston ve ark., 1990; Özgürlük, 2019; Tanberkan Suna ve Tan, 2017; Wang, 2019). Andersson ve Wiberg (2017) Kernel eşitlemenin 2PL model için minimum 1000, 3PL model için minimum 3000 örneklem büyüklüğüne sahip olması gerektiğini belirtmişlerdir. Araştırmada kullanılan

örneklerde ise kişi sayısı 1000'den küçüktür. Örneklem büyüklüğü küçüldüğünde ve araştırmada kullanılan yöntemlerin sınırlılıkları olduğunda bu araştırmadan farklı olarak ise özdeşlik yönteminin (Albano ve Wiberg, 2019; Çağlak, 2016) ve daire-yay yönteminin (Diao ve Keller, 2020) en iyi sonuçlar verdiği çalışmalara da rastlanmaktadır.

10. momentteki PRE'ler ise aşama 1'de %2,3'ten aşama 2'de %10'dan küçüktür. RMSE değerleri her iki aşamada da 1,3'ten küçüktür. Bu değerler (González & Wiberg, 2017) tarafından belirtilen sınırlar içinde olmakla birlikte Kernel eşitlemenin kullanılabileceğinin göstergeleridir. Ek olarak en az bir düzgünleştirme yönteminin ve son tabakalaşma yöntemlerinin kullanıldığı durumlarda hataların azalma eğiliminde olması (Kim, 2014; Moses ve ark., 2010; 2011; von Davier ve ark., 2006) da bu çalışmaların bulgularını desteklemektedir.

Aşama 1'de MTK'ye göre KTK'den farklı olarak kişi sayısı arttıkça SEE, RMSE ve RMSD azalmıştır. Aşama 2'de ise eşitlenen kitapçıklardaki kişi sayıları birbirine yakın olduğundan SEE'ler de birbirine yakındır. Koğar ve Sayın (2018 Ekim) araştırmalarında örneklem büyüklüğü arttıkça da yine hata değerlerinin azaldığı, belirli bir büyüklükten sonra ise bu artışın daha az olduğunu belirtmişlerdir. Bu bulgu çalışmanın Aşama 2 kısmında SEE'lerin daha yakın olmasının nedeni olarak görülebilir.

Hata dağılımlarında KTK'ye göre yapılan eşitlemede yüksek puanlarda hataların da yüksek olduğu; MTK'de ise orta puanlarda hataların yüksek olduğu görülmüştür. Benzer şekilde Kernel eşitleme yöntemlerinin orta puan ölçeğinde düşük hata uçlarda frekans azlığı nedeniyle yüksek hata vermesi de Akın Arıkan (2017) ve Andersson ve Wiberg (2017) tarafından yapılan çalışmalar ile desteklenmektedir.

### ***İkinci alt problem***

Ankor maddelerden alınan toplam puanın 3, 5, 8 ve 13 kategorili ortak değişken olarak kullanıldığı DOG-OD desende KTK'ye dayalı olarak yapılan Kernel eşitlemede eşitlenmiş puanlar ve eşitleme hataları incelenmiştir. Kategori sayısı arttıkça P ve Q

evrenlerinde frekans dağılımları arasındaki fark ve SEE'ler artmaktadır. En yüksek SEE, Aşama 1'de FC veri setinde 13 kategorili ortak değişkende elde edilmiştir. Aşama 2'de ise hatalar ile kategori sayıları arasında bir örüntüye rastlanmamıştır. 5 ile 6. kitapçıkların eşitlenmesinde kullanılan üç kategorili ortak değişkende en düşük SEE elde edilmiştir. Aşama 1'de genel olarak 3 ve 5 kategorili ortak değişkende, Aşama 2'de 5 kategorili ortak değişkende SEE, 10. PRE, RMSE ve RMSD'ler daha düşüktür. Fakat bu durum eşitleme sonuçlarının iyi olmadığı şeklinde yorumlanmamaktadır. Eşit yüzdelli eşitlemedeki standart hata daha yüksek olsa bile popülasyonun daha doğru bir tahminini sağlayabilmektedir (Brennan, 2014). Kernel eşitlemede kullanılan bant genişlikleri optimum düzeyde olduğunda eşitleme sonuçları eşit yüzdelli eşitlemeye yaklaşmaktadır. Bu bulgular ile birlikte SEED'ler incelendiğinde ise her iki aşamada da 3 ve 5 kategorili ortak değişkenin hata bantlarının daha düşük olduğu, eşitleme fonksiyonları arasındaki hataların dağılımının bant genişliği içinde olması nedeniyle benzer olduğu görülmüştür. Dolayısı ile özellikle ankor maddeler 5 kategorili ortak değişken olarak kullanıldığında sonuçların benzer olduğu söylenebilmektedir. Bu bulgu, Özsoy'un (2021) DOG-OM ve DOG-OD desenlerini karşılaştırdığı çalışmada, DOG-OD desenin daha yüksek hatalar ile kestirim yapması nedeniyle ortak değişkenin ortak madde yerine kullanılmayacağı bulgusunu desteklememektedir. Bu farklılığın nedeninin, ortak değişken olarak gerçek kovaryantlar (bilgisayar/tablet sahibi olup olmama ve cinsiyet) kullanıldığında ortak değişken özelliklerinin ankor madde özelliklerinden farklı olmasının eşitleme sonuçlarını etkilemesi olduğu söylenebilir. Ankor maddeleri birer ortak değişken olarak ele alan çalışmalarda da (Longford, 2015; Wright ve Dorans, 1993; Yurtçu, 2018) gerçek veriler ile simülasyon verileri; Kernel eşitleme, Tucker ve frekans tahmini yöntemleri ve Bayes yöntemleri kullanılarak incelenmiştir.

### ***Üçüncü alt problem***

Ortak değişkene ait hangi özellikler Kernel eşitleme ile elde edilen eşitlenmiş puanları, SEE, 10. PRE, RMSE ve RMSD'leri etkilemekte olduğunu belirlemek için bu değişkenlerin

içsel değişken olduğu yol modeli test edilmiştir. Dışsal değişken olarak ortak değişkenin P ve Q evrenlerindeki frekans dağılımları arasındaki fark (D); gözlenen test puanları ile ortak değişken arasındaki ilişkinin mutlak değeri (absPR ve absQR); ortak değişkenin çarpıklık ve basıklığının mutlak değeri (absPÇarpıklık, absPBasıklık, absQÇarpıklık ve absQBasıklık); testteki maddelerin P ve Q evrenlerinde ortak değişkene göre DMF gösterip göstermemesi (DMF) alınmıştır. Bu değişkenlerin öncelikle P ve Q evrenlerinde ortak değişken ile kurulan eşitleme modelinin açıklanan varyans oranını (PRkare ve QRkare) etkileyip etkilemediği daha sonra ayrı ayrı ve birlikte eşitleme sonuçlarını yordamaları incelenmiştir.

Aşama 1'de D ve Çarpıklık değişkenleri Aşama 2'de D, absR, absÇarpıklık ve absBasıklık değişkenleri eşitleme modelinin açıklanan varyans oranını daha sonra gözlenen test puanları ile ortak değişken arasındaki ilişkinin mutlak değeri ile birlikte eşitleme sonuçlarını etkilediği görülmüştür. Akın Arıkan (2020) ortak değişken olarak test puanları ile ilişkisiz olan cinsiyet ve test puanları ile düşük ilişkili sosyoekonomik düzey ve bu iki değişken birlikte dahil ettiği DOG-OD deseninde sosyoekonomik düzeyin kullanıldığı modelde cinsiyete göre daha az hatalar elde edildiğini belirtmiştir. Wiberg ve Bränberg (2015) de ortak değişken olarak öğrencilerin sınıf düzeylerini ve sözel test puanlarının birlikte ankor maddelerin kullanıldığı durumlarda daha küçük SEE'ler elde ettiklerini belirtmişlerdir. Ortak değişkenin özellikleri ile ilgili herhangi bir bilgi belirtmemişlerdir. Ancak eşitleme sonuçlarının görece iyi olmasının nedeninin ankor maddelerin varlığının kurulan modelde açıklanan varyans oranını arttırmasından kaynaklanabileceği söylenebilir. Çünkü benzer bir birleştirme işlemi Wallin ve Wiberg (2019) ortak değişkenlerin kategorilerinde gözlem olmama ihtimalini düşünerek doğrudan ortak değişkenleri kullanmak yerine eğilim (propensity) puanlarını kullanmanın da eşitleme sonuçlarını iyileştirdiğini belirtmişlerdir.

Liou ve ark. (2001) ortak maddelerden elde edilen puanlar ile test puanları arasındaki korelasyon çok küçük olduğunda da ortak değişkenlerin modele dahil edilebileceğini

belirtmiştir. Literatürde de cinsiyet, sosyoekonomik düzey, etnik köken gibi test puanları ile ilişkili olmadığı belirtilen değişkenler veri setinde yer aldığı için modele dahil edildiği belirtilmiş ve eşitleme yapılabileceği belirtilmiştir (Bränberg ve Wiberg, 2011; Akın Arıkan, 2020; Altıntaş ve Wallin, 2021; Yurtçu, 2018). Fakat bu çalışmada ortak değişkenler test puanları ile düşük ilişki gösterdiğinde, iyi bir yordayıcı olmadığına eşitleme hataları (SEE, 10. PRE, RMSE, RMSD) da daha yüksek kestirilmektedir. Dolayısı ile düşük ilişki gösteren ortak değişkenlerin seçilmesi önerilmemektedir. Özellikle alınan puanlar uzun yıllar geçerli olduğunda eşitleme sonuçlarının tahmin edicilerinin de iyi belirlenmesinin test adaletini korumada önemli olduğu göz ardı edilmemelidir.

P ve Q evrenlerinde testlerin DMF içerip içermeme durumlarına ait yol katsayıları anlamlı bulunamamıştır. Güneş (2021) de Kernel ve DOG-OD deseninden farklı olarak DOG-OM deseninde yaptığı çalışmasında DMF içeren ortak maddeler testten çıkarıldığında eşitleme hatalarında gözlenen değişimin tutarlı bir örüntü sergilemediğini belirtmiştir. Bu sonucun kitapçıkların güçlükleri arasındaki farkın, DMF değişkeninde gruplar arasındaki yetenek farkının ya da eşitleme varsayımlarının tam olarak karşılanmamış olmasının neden olabileceğini belirtilmiştir. Benzer bir bulgu Sansivieri (2017) tarafından ortak değişken olarak sınava girenlerin cinsiyeti ve uyruğu alındığında DMF varlığının ya da yokluğunun MTK'ye dayalı eşitleme hataları üzerinde bir iyileşmeye yol açmadığı şeklinde belirtilmiştir.

Tüm özellikler birlikte değerlendirildiğinde ise veri setlerinde var olan değişkenlerden Aşama 1 ve Aşama2'de "ankor maddelerden üretilen 3 kategorili ortak değişken", "ankor maddelerden üretilen 5 kategorili ortak değişken", "Matematikte genellikle iyi olduğunu düşünme", "Matematiğin daha zor olduğunu düşünme" ve "Matematikte kendine güvenme indeksi" değişkenlerinin kullanılmasının eşitleme sonuçlarını iyileştirdiği görülmüştür. Aşama 1'de farklı olarak "Matematikte kendisini güçlü hissetmeme" ve "Eğitim kaynakları indeksi", Aşama 2'de farklı olarak "Sadece matematikte iyi olmadığını düşünme", "Matematiğin kendisi için zor olduğunu düşünme" ve "Matematiğin kafa karıştırıcı olduğunu düşünme" değişkenleri yer almaktadır.

## **Öneriler**

Bahsedilen sonuçlar ve araştırmanın sınırlılıkları temel alındığında araştırmacılara ve uygulayıcılara yönelik öneriler aşağıdaki gibidir.

### ***Uygulayıcılara yönelik***

- Ortak maddelerin kullanılmadığı ya da test tasarımında yer almadığı durumlarda ortak değişkenler kullanılabilir ve eşitleme yöntemi olarak Kernel eşitleme yapılabilir.
- Ortak değişken seçilirken kategorilerdeki frekansların sıfırdan farklı olması ile birlikte, P ve Q evrenlerindeki frekans dağılımlarının yakın olmasına dikkat edilmelidir.
- Özellikle eşitleme için kurulan regresyon modelinin açıklanan varyansını ve eşitleme sonuçlarını etkilemesi nedeniyle ortak değişkenin çarpıklık ve basıklığının normal dağılımdan büyük sapmalar göstermeyecek şekilde seçilmesi önerilmektedir.
- P ve Q evrenlerinde ortak değişken ile ham puanlar arasındaki ilişkiler yüksek olan değişkenler kullanılabilir.
- Her iki aşamada da ortak olan “ankor maddelerden üretilen 3 kategorili ortak değişken”, “ankor maddelerden üretilen 5 kategorili ortak değişken”, “Matematikte genellikle iyi olduğunu düşünme”, “Matematiğin daha zor olduğunu düşünme” ve “Matematikte kendine güvenme indeksi” değişkenlerinin Kernel eşitlemede kullanılması önerilmektedir.
- Son durumda kullanılması önerilen değişkenlerin kategori sayıları 3, 4 ve 5 olduğundan değişkenler seçilirken kategori sayısının en fazla 5 olması önerilmektedir.

### ***Araştırmacılara yönelik***

- Ortak değişkenin bu araştırmada kullanılmayan farklı durumlar (birden fazla ortak değişkenin bir arada yer alması, hem ankor test hem de ortak değişkenin bir arada kullanılması) modele dahil edilerek yeni bir araştırma yapılabilir.

- Çalışmada TIMSS 2019 matematik 2, 3, 4, 5, 6. kitapçıklar kullanılmıştır. Benzer araştırma gerçek ve simülasyon veri setlerinin bir arada kullanıldığı ve daha fazla kitapçık kullanılarak üçten fazla eşitleme için tekrarlanabilir.
- Araştırmadaki maddeler 1-0 puanlanan maddeler olacak şekilde düzenlenmiştir. Kategorik puanlanan maddeler ya da karma testler için eşitleme yöntemleri incelenebilir. Madde takımı etkisinin dahil edildiği DOG deseninde yapılacak eşitleme sonuçları karşılaştırılabilir.
- Bu çalışmadaki testler tek boyutlu olarak ele alınmıştır. Benzer bir çalışma çok boyutlu testler için test eşitleme tasarımının da araştırmacı tarafından kontrol edilebileceği şekilde planlanarak tekrarlanabilir.
- TIMSS 2019 uygulamasındaki öğrenci, sınıf ve okul ağırlıkları eşitlemeye dahil edilerek örneklemden kaynaklanan hataların azaltılmasına yönelik çalışmalar tasarlanabilir.
- Yalnız Aşama 1'de DMF etkisi modele dahil edilebilmiştir ve DMF içeren madde sayısı ve düzeyine müdahale edilmemiştir. DMF içeren ve içermeyen maddelerin oranları değiştirilerek DOG-OD deseninde Kernel eşitleme sonuçları incelenebilir.
- Çalışma DOG-OD deseninde KTK'ye dayalı Kernel eşitleme ile yapılmıştır. Benzer çalışma Bayesian yöntemler ya da MTK'ye dayalı yöntemler kullanılarak tekrarlanabilir.



## Kaynaklar

- Akın Arıkan, Ç. (2017). *Kernel eşitleme ve madde tepki kuramına dayalı eşitleme yöntemlerinin karşılaştırılması* [Doktora Tezi, Hacettepe Üniversitesi]. Ankara.
- Akın Arıkan, Ç. (2020). The Impact of Covariate Variables on Kernel Equating under the Non-equivalent Groups. *Journal of Measurement and Evaluation in Education and Psychology*, 11(4), 362-373. <https://doi.org/10.21031/epod.706835>
- Albano, A. D. (2016). equate: An R package for observed-score linking and equating. *Journal of Statistical Software*, 74, 1-36. <http://138.232.16.156/article/view/v074i08>
- Albano, A. D., & Wiberg, M. (2019). Linking With External Covariates: Examining Accuracy by Anchor Type, Test Length, Ability Difference, and Sample Size. *Applied Psychological Measurement*, 43(8), 597-610. <https://doi.org/10.1177/0146621618824855>
- Altıntaş, Ö., & Wallin, G. (2021). Equality of admission tests using kernel equating under the non-equivalent groups with covariates design. *International Journal of Assessment Tools in Education*, 8(4), 729-743. <https://doi.org/https://doi.org/10.21449/ijate.976660>
- American Educational Research Association, A. P. A., & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Andersson, B., Branberg, K., Wiberg, M., & Andersson, M. B. (2022). Package 'kequate'. <https://cran.microsoft.com/snapshot/2017-04-11/web/packages/kequate/kequate.pdf>
- Andersson, B., & von Davier, A. A. (2014). Improving the bandwidth selection in kernel equating. *Journal of Educational Measurement*, 51(3), 223-238. <https://doi.org/https://doi.org/10.1111/jedm.12044>

- Andersson, B., & Wiberg, M. (2014). IRT observed-score kernel equating with the R package kequate. In: Citeseer.
- Andersson, B., & Wiberg, M. (2017). Item Response Theory Observed-Score Kernel Equating. *Psychometrika*, 82(1), 48-66. <https://doi.org/10.1007/s11336-016-9528-7>
- Angoff, W. H. (1971). Scales, norms and equivalent scores. In R. L. Thorndike (Ed.), *Educational measurement* (pp. 508–600). American Council on Education.
- Atar, B., & Yeşiltaş, G. (2017). Çok Boyutlu Eşitleme Yöntemlerinin Eşdeğer Olmayan Gruplarda Ortak Madde Deseni için Performanslarının İncelenmesi. *Journal of Measurement and Evaluation in Education and Psychology*, 8(4), 421-434. <https://doi.org/https://doi.org/10.21031/epod.335284>
- Bağcı, A. G. V., Erdemir, A. G. A., & Çavdar, A. G. D. (2016 Ekim). *Küçük Örneklemlerde Ortak Maddeli Eşdeğer Olmayan Gruplar Deseni Altında Farklı Eşitleme Yaklaşımlarının Karşılaştırılması* 5. Ulusal Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi, Antalya. <https://www.epodder.org/wp-content/uploads/2020/07/EPOD-2016.pdf>
- Baker, F. B. (Ed.). (2016). *Madde tepki kuramının temelleri* (M. İlhan & N. Güler, Trans.). Pegem Akademi.
- Baykul, Y. (2010). *Eğitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması*. Pegem Akademi.
- Bishop, Y., Fienberg, S., & Holland, P. (1975). *Discrete Multivariate Analysis: Theory and Practice*. MIT Press.
- Bränberg, K. (2010). *Observed score equating with covariates* Department of Statistics, Umeå university]. Sweden.
- Bränberg, K., & Wiberg, M. (2011). Observed score linear equating with covariates. *Journal of Educational Measurement*, 48(4), 419-440. <https://doi.org/https://doi.org/10.1111/j.1745-3984.2011.00153.x>

Braun, H. I., & Holland, P. (1982). Observed-score test equating: A mathematical analysis of some ETS equating procedures. In P. W. Holland & D. B. Rubin (Eds.), *Test equating* (pp. 9–49). Academic Press.

Büyüköztürk, Ş. (2016). Sınavlar üzerine düşünceler. *Kalem Eğitim ve İnsan Bilimleri Dergisi*, 6(2), 345-356.  
[http://kalemacademy.org/Cms\\_Data/Contents/KalemAcademyDB/Folders/SayiMakaleleri/~contents/MXT2L44R7L8HB5U6/2016-kalemueibd-11-kissayisi-makale-001-11.pdf](http://kalemacademy.org/Cms_Data/Contents/KalemAcademyDB/Folders/SayiMakaleleri/~contents/MXT2L44R7L8HB5U6/2016-kalemueibd-11-kissayisi-makale-001-11.pdf)

Büyüköztürk, Ş. (2019). *Veri analizi el kitabı*. Pegem Akademi.

Büyüköztürk, Ş., Kılıç-Çakmak, E., Akgün, Ö., Karadeniz, Ş., & Demirel, F. (2020). *Bilimsel araştırma yöntemleri*. Pegem Akademi.

Cağlak, S. (2016). Comparison of several small sample equating methods under the NEAT design. *Turkish Journal of Education*, 5(3), 96-118.  
<https://doi.org/10.19128/turje.16916>

Chalmers, P., Pritikin, J., Robitzsch, A., Zoltak, M., KwonHyun, K., Falk, C., & Meade, A. (2015). Package 'mirt'. *Zugriff am*, 21(01), 2016.  
<http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.359.6801&rep=rep1&type=pdf>

Choi, J. (2019). *Comparison of MIRT observed score equating methods under the common-item nonequivalent groups design* [University of Iowa].  
<https://iro.uiowa.edu/esploro/outputs/doctoral/Comparison-of-MIRT-observed-score-equating/9983776729702771>

Cid, J. A., & von Davier, A. A. (2015). Examining potential boundary bias effects in kernel smoothing on equating: An introduction for the adaptive and Epanechnikov kernels. *Applied Psychological Measurement*, 39(3), 208-222.  
<https://doi.org/https://doi.org/10.1177/0146621614555901>

- Crocker, L., & Algina, J. (2006). *Introduction to classical and modern test theory*. Cengage Learning.
- Çakan, M. (2003). Geniş ölçekli başarı testlerinin eğitimindeki yeri ve önemi. *Eğitim ve Bilim*, 28(128). <http://eb.ted.org.tr/index.php/EB/article/view/5130/1212>
- De Ayala, R. (2009). *Methodology in the social sciences*. Guilford Press.
- DeMars, C. (2016). *Madde tepki kuramı* (E. Özberk & H. Kelecioğlu, Eds.).
- Diao, H., & Keller, L. (2020). Investigating Repeater Effects on Small Sample Equating: Include or Exclude? *Applied Measurement in Education*, 33(1), 54-66. <https://doi.org/https://doi.org/10.1080/08957347.2019.1674302>
- 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. <https://doi.org/https://doi.org/10.1111/j.1745-3984.2000.tb01088.x>
- 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. <https://doi.org/https://doi.org/10.1002/j.2333-8504.2010.tb02236.x>
- Efron, B., & Tibshirani, R. J. (1993). An introduction to the bootstrap. *Monographs on statistics and applied probability*, 57, 1-436. [https://cindy.informatik.uni-bremen.de/cosy/teaching/CM\\_2011/Eval3/pe\\_efron\\_93.pdf](https://cindy.informatik.uni-bremen.de/cosy/teaching/CM_2011/Eval3/pe_efron_93.pdf)
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Lawrence Erlbaum Associates.
- Epskamp, S., Epskamp, M. S., & MplusAutomation, S. (2017). Package 'semPlot'. Recuperado de <https://cran.r-project.org/web/packages/semPlot/semPlot.pdf>
- Erden, M. (1998). *Eğitimde program değerlendirme*. Anı Yayıncılık.

- Ertoprak, D. G. (2017). *Kayıp verinin test eşitlemeye etkisinin incelenmesi* Hacettepe Üniversitesi].
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). *How to design and evaluate research in education*. McGraw-Hill.
- Ghosh, J., & Ramamoorthi, R. (2003). *Bayesian nonparametrics*. Springer.
- González, J., Barrientos, A. F., & Quintana, F. A. (2015). Bayesian nonparametric estimation of test equating functions with covariates. *Computational Statistics & Data Analysis*, 89, 222-244.  
<https://doi.org/https://doi.org/10.1016/j.csda.2015.03.012>
- González, J., & Davier, A. A. v. (2016). An illustration of the Epanechnikov and adaptive continuation methods in kernel equating. The Annual Meeting of the Psychometric Society,
- González, J., & Wiberg, M. (2017). *Applying test equating methods*.
- Gulliksen, H. (2013). *Theory of mental tests*. Routledge.
- Güneş, F. (2021). *Değişen Madde Fonksiyonu Gösteren Ortak Maddelerin Test Eşitlemeye Etkisinin İncelenmesi* [Yüksek Lisans Tezi, Hacettepe Üniversitesi, Ankara].
- Häggström, J., & Wiberg, M. (2014). Optimal bandwidth selection in observed-score kernel equating. *Journal of Educational Measurement*, 51(2), 201-211.  
<https://doi.org/https://doi.org/10.1111/jedm.12042>
- Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement: Issues and Practice*, 23(1), 17-27.  
<https://doi.org/https://doi.org/10.1111/j.1745-3992.2004.tb00149.x>
- 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.

[https://www.academia.edu/download/47596077/ITEMS\\_Module\\_16.pdf](https://www.academia.edu/download/47596077/ITEMS_Module_16.pdf)

Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications*. Kluwer-Nijhoff Publishing.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.

Holland, P. W. (2007). A framework and history for score linking. In N. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 5-30). Springer.

Holland, P. W., & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan (Ed.), *Educational measurement* (pp. 187-220). American Council on Education and Praeger Publishers.

Holland, P. W., King, B. F., & Thayer, D. T. (1989). The standard error of equating for the kernel method of equating score distributions. . *ETS Research Report Series*, 1989(1), 83-89. <https://doi.org/https://doi.org/10.1002/j.2330-8516.1989.tb00332.x>

Holland, P. W., & Thayer, D. T. (1989). The kernel method of equating score distributions. *ETS Research Report Series*, 1989(1), i-45. <https://doi.org/https://doi.org/10.1002/j.2330-8516.1989.tb00333.x>

Holland, P. W., & Thayer, D. T. (2000). Univariate and bivariate loglinear models for discrete test score distributions. *Journal of Educational and Behavioral Statistics*, 25(2), 133-183. <https://doi.org/https://doi.org/10.3102/10769986025002133>

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. <https://doi.org/https://doi.org/10.1080/10705519909540118>

- Im, J., Cho, I. H., & Kim, J.-K. (2018). FHDI: An R Package for Fractional Hot Deck Imputation. *R J.*, 10(1), 140. [https://www.researchgate.net/profile/In-Ho-Cho-2/publication/328074285\\_FHDI\\_An\\_R\\_package\\_for\\_fractional\\_hot\\_deck\\_imputation/links/605ce5d7299bf173676ba434/FHDI-An-R-package-for-fractional-hot-deck-imputation.pdf](https://www.researchgate.net/profile/In-Ho-Cho-2/publication/328074285_FHDI_An_R_package_for_fractional_hot_deck_imputation/links/605ce5d7299bf173676ba434/FHDI-An-R-package-for-fractional-hot-deck-imputation.pdf)
- Jackman, S., Tahk, A., Zeileis, A., Maimone, C., Fearon, J., Meers, Z., Jackman, M. S., & Imports, M. (2015). Package 'pscl'. *Political Science Computational Laboratory*, 18(04.2017). <https://cran.microsoft.com/snapshot/2015-03-01/web/packages/pscl/pscl.pdf>
- Jiang, Y., von Davier, A. A., & Chen, H. (2012). Evaluating equating results: Percent relative error for chained kernel equating. *Journal of Educational Measurement*, 49(1), 39-58. <https://doi.org/https://doi.org/10.1111/j.1745-3984.2011.00159.x>
- Johnson, R. A., & Wichern, D. W. (2002). *Applied multivariate statistical analysis* (Vol. 5). Prentice hall Upper Saddle River, NJ.
- Karagül, A. E. (2020). *Küçük örneklemlerde çok kategorili puanlanan maddelerden oluşan testlerde klasik test eşitleme yöntemlerinin karşılaştırılması* [Ankara Üniversitesi].
- Kim, H. Y. (2014). *A comparison of smoothing methods for the common item nonequivalent groups design* [The University of Iowa].
- Kim, J. K., & Fuller, W. (2004). Fractional hot deck imputation. *Biometrika*, 91(3), 559-578. <https://doi.org/https://doi.org/10.1093/biomet/91.3.559>
- Kim, S., & Lu, R. (2018). The Pseudo-Equivalent Groups Approach as an Alternative to Common-Item Equating. *ETS Research Report Series*, 2018(1), 1-13. <https://doi.org/https://doi.org/10.1002/ets2.12195>
- Kim, S., von Davier, A. A., & Haberman, S. (2006). An alternative to equating with small samples in the non-equivalent groups anchor test design. *ETS Research Report*

- Series, 2006(2), i-40. <https://doi.org/https://doi.org/10.1002/j.2333-8504.2006.tb02033.x>
- Kline, R. B. (2011). *Principles and practice of structural equation modeling*. Guilford publications.
- Kline, T. (2005). *Psychological testing: A practical approach to design and evaluation*. Sage.
- Koğar, H., & Sayın, A. (2018 Ekim). *Denk olmayan gruplarda Kernel eşitleme yönteminden elde edilen sonuçların karşılaştırılması* 6. Uluslararası Eğitimde ve Psikolojide Ölçme ve Değerlendirme Kongresi, Kosova. <https://avesis.gazi.edu.tr/yayin/e6c10b76-a60e-4e77-903a-74bd7f488323/denk-olmayan-gruplarda-kernel-esitleme-yonteminden-elde-edilen-sonuclarin-karsilastirilmasi>
- Kolen, M. J. (1988). Traditional equating methodology. *Educational Measurement: Issues and Practice*, 7(4), 29-37. <https://doi.org/https://doi.org/10.1111/j.1745-3992.1988.tb00843.x>
- Kolen, M. J. (2006). Scaling and norming. In R. L. Brennan (Ed.), *Educational measurement* (Vol. 4, pp. 156-186). American Council on Education and Praeger Publishers.
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking*. Springer.
- Kumandaş, H., & Kutlu, Ö. (2015). High stake tests. *Journal of Educational Science Research*, 5(2), 63-75. <https://doi.org/http://dx.doi.org/10.12973/jesr.2015.52.4>
- Kumlu, G. (2019). *Test ve alt testlerde eşitlemenin farklı koşullar açısından incelenmesi* Hacettepe Üniversitesi].
- Lee, Y.-H., & Davier, A. A. v. (2009). Equating through alternative kernels. In A. A. v. Davier (Ed.), *Statistical models for test equating, scaling, and linking* (pp. 159-173). Springer.



- Liang, T., & von Davier, A. A. (2014). Cross-validation: An alternative bandwidth-selection method in kernel equating. *Applied Psychological Measurement*, 38(4), 281-295. <https://doi.org/https://doi.org/10.1177/0146621613518094>
- Lim, B., & Qin, H. (2016). The data.table Package for Data Cleaning and Preparation in R. [http://homepage.divms.uiowa.edu/~kcowles/STAT5400\\_2016/LimQin.pdf](http://homepage.divms.uiowa.edu/~kcowles/STAT5400_2016/LimQin.pdf)
- Liou, M., Cheng, P. E., & Li, M.-Y. (2001). Estimating comparable scores using surrogate variables. *Applied Psychological Measurement*, 25(2), 197-207. <https://doi.org/https://doi.org/10.1177/01466210122032000>
- Lisansüstü Eğitim Öğretim Yönetmeliği. (2016, 04 20). *Resmî Gazete*. <https://www.mevzuat.gov.tr/mevzuat?MevzuatNo=21510&MevzuatTur=7&MevzuatTertip=5>
- Liu, J., & Low, A. C. (2007). An exploration of kernel equating using SAT® data: Equating to a similar population and to a distant population. *ETS Research Report Series*, 2007(1), i-22. <https://doi.org/https://doi.org/10.1002/j.2333-8504.2007.tb02059.x>
- Livingston, S. A. (1993). Small-sample equating with log-linear smoothing. *Journal of Educational Measurement*, 30(1), 23-39. <https://doi.org/https://doi.org/10.1111/j.1745-3984.1993.tb00420.x>
- Livingston, S. A. (2014). Equating test scores (without IRT). *ETS Research Report Series*. <https://eric.ed.gov/?id=ED560972>
- 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. [https://doi.org/https://doi.org/10.1207/s15324818ame0301\\_6](https://doi.org/https://doi.org/10.1207/s15324818ame0301_6)
- Livingston, S. A., & Lewis, C. (2009). Small-sample equating with prior information. *ETS Research Report Series*, 2009(1), i-14. <https://doi.org/10.1002/j.2333-8504.2009.tb02182.x>

- Longford, N. T. (2015). Equating without an anchor for nonequivalent groups of examinees. *Journal of Educational and Behavioral Statistics*, 40(3), 227-253. <https://doi.org/10.3102/1076998615574773>
- Lord, F. M. (1980). *Applications of Item response theory to practical testing problems*. Erlbaum.
- Magis, D., Beland, S., & Raiche, G. (2020). Package 'difR'. In.
- Martin, M., von Davier, M., & Mullis, I. (2020). *Methods and Procedures: TIMSS 2019 technical report*.
- MEB. (2020). *TIMSS 2019 Türkiye ön raporu*. (Eğitim Analiz ve Değerlendirme Raporları Serisi, Issue).
- Mittlböck, M., & Schemper, M. (1996). Explained variation for logistic regression. *Statistics in medicine*, 15(19), 1987-1997. [https://doi.org/https://doi.org/10.1002/\(SICI\)1097-0258\(19961015\)15:19<1987::AID-SIM318>3.0.CO;2-9](https://doi.org/https://doi.org/10.1002/(SICI)1097-0258(19961015)15:19<1987::AID-SIM318>3.0.CO;2-9)
- Morgan, C. T. (2013). *Psikolojiye giriş* (S. Karakaş & R. Eski, Trans.). Eğitim Yayınevi.
- Moses, T., Deng, W., & Zhang, Y.-L. (2011). Two approaches for using multiple anchors in NEAT equating: A description and demonstration. *Applied Psychological Measurement*, 35(5), 362-379. <https://doi.org/https://doi.org/10.1177/0146621611405510>
- Moses, T., Deng, W., & Zhang, Y. L. (2010). The use of two anchors in nonequivalent groups with anchor test (NEAT) equating. *ETS Research Report Series*, 2010(2), i-33. <https://doi.org/https://doi.org/10.1002/j.2333-8504.2010.tb02230.x>
- Moses, T., & Holland, P. W. (2009). Selection strategies for univariate loglinear smoothing models and their effect on equating function accuracy. *Journal of Educational Measurement*, 46(2), 159-176. <https://doi.org/https://doi.org/10.1111/j.1745-3984.2009.00075.x>

- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*.
- Mutluer, C. (2021). *Klasik test kuramına ve madde tepki kuramına dayalı test eşitleme yöntemlerinin karşılaştırması: Uluslararası öğrenci değerlendirme programı (PISA) 2012 matematik testi örneği* Gazi Üniversitesi, Ankara.].  
<https://tez.yok.gov.tr/UlusalTezMerkezi/>
- ÖSYM Sınav Takvimi. (2020). Ölçme, Seçme ve Yerleştirme Merkezi Başkanlığı.  
<https://www.osym.gov.tr/TR,8797/takvim.html>
- Özgürlük, B. (2019). *Örnekleme Büyüklüğünün ve Madde Formatının Sekizinci Sınıf Abide Testlerinin Eşitlenmesine Etkisi* Hacettepe Üniversitesi].
- Özsoy, Ş. N. (2021). *Kernel eşitleme yöntemlerinin karşılaştırılması: TIMSS 2019 fen testi örneği* Abant İzzet Baysal Üniversitesi].
- Öztürk, N., & Anıl, D. (2012). Akademik personel ve lisansüstü eğitimi giriş sınavı puanlarının eşitlenmesi üzerine bir çalışma. *Eğitim ve Bilim*, 37(165).  
<http://eb.ted.org.tr/index.php/EB/article/view/1141>
- Paek, I., & Cole, K. (2019). *Using R for item response theory model applications*. Routledge.  
<https://doi.org/https://doi.org/10.4324/9781351008167>
- Petersen, N. S., Kolen, M. J., & Hoover, H. D. (1989). Scaling, norming, and equating.
- Revelle, W., & Revelle, M. W. (2015). Package 'psych'. *The Comprehensive R Archive Network*, 337, 338. <https://cran.rstudio.org/web/packages/psych/psych.pdf>
- Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48, 1-36. <https://doi.org/10.18637/jss.v048.i02>
- Sansivieri, V. (2017). *Item response theory equating with the non-equivalent groups with covariates design* Bologna University, Italy. ].

- Tan, Ş. (2015). Küçük örneklerde beta4 ve polynomial loglineer öndüzdüğüleştirme ve kübik eğri sondüzdüğüleştirme metotlarının uygunluğu. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 35(1), 123-151. <https://dergipark.org.tr/en/download/article-file/77508>
- Tanberkan Suna, H., & Tan, Ş. (2017). Grup Değişmezliği Özelliğinin Farklı Eşitleme Yöntemlerinde İncelenmesi. *Karaelmas Eğitim Bilimleri Dergisi*, 5(2), 316-327. <https://dergipark.org.tr/en/pub/kebd/issue/67222/1049187>
- Turgut, M. F., & Baykul, Y. (1992). *Ölçekleme teknikleri*.
- Uysal, İ. (2014). *Madde Tepki Kuramı'na dayalı test eşitleme yöntemlerinin karma modeller üzerinde karşılaştırılması* Abant İzzet Baysal Üniversitesi, Bolu.]. <https://tez.yok.gov.tr/UlusalTezMerkezi/>
- Uysal, İ. (2019). *Açık uçlu maddelerde otomatik puanlamanın güvenilirliği ve test eşitleme hatalarına etkisi* Hacettepe Üniversitesi].
- van der Linden, W. J. (2010). Local observed-score equating. In A. A. v. Davier (Ed.), *Statistical models for equating, scaling, and linking* (pp. 201-223). Springer.
- von Davier, A. A. (2013). Observed-score equating: An overview. *Psychometrika*, 78(4), 605-623. <https://link.springer.com/article/10.1007/s11336-013-9319-3>
- von Davier, A. A., Holland, P. W., Livingston, S. A., Casabianca, J., Grant, M. C., & Martin, K. (2006). An evaluation of the kernel equating method: A special study with pseudotests constructed from real test data. *ETS Research Report Series*, 2006(1), i-31. <https://doi.org/https://doi.org/10.1002/j.2333-8504.2006.tb02008.x>
- von Davier, A. A., Holland, P. W., & Thayer, D. T. (2004). *The kernel method of test equating*. Springer.
- Wallin, G., & Wiberg, M. (2019). Kernel equating using propensity scores for nonequivalent groups. *Journal of Educational and Behavioral Statistics*, 44(4), 390-414. <https://doi.org/https://doi.org/10.3102/1076998619838226>

Wang, S. (2019). *Beta observed score and true score equating methods* The University of Iowa].

[https://search.proquest.com/openview/a6d133387faa382bce91229d25154746/1?pq-origsite=gscholar&cbl=18750&diss=y&casa\\_token=YX1DAylvn2QAAAAA:ydf3\\_TAwCkHA-ITvcGQFmS7NA9dQaYd\\_hTUTF0\\_-rbDxH2AjP7g1ayYf7dRT67muzQWbaUs6Qpw](https://search.proquest.com/openview/a6d133387faa382bce91229d25154746/1?pq-origsite=gscholar&cbl=18750&diss=y&casa_token=YX1DAylvn2QAAAAA:ydf3_TAwCkHA-ITvcGQFmS7NA9dQaYd_hTUTF0_-rbDxH2AjP7g1ayYf7dRT67muzQWbaUs6Qpw)

Wiberg, M., & Bränberg, K. (2015). Kernel equating under the non-equivalent groups with covariates design. *Applied Psychological Measurement*, 39(5), 349-361.

<https://doi.org/https://doi.org/10.1177/0146621614567939>

Wiberg, M., & von Davier, A. A. (2017). Examining the impact of covariates on anchor tests to ascertain quality over time in a college admissions test. *International Journal of Testing*, 17(2), 105-126.

<https://doi.org/https://doi.org/10.1080/15305058.2016.1277357>

Wickham, H., François, R., Henry, L., & Müller, K. (2022). dplyr: A Grammar of Data Manipulation. *The Comprehensive R Archive Network*. <https://cran.r-project.org/web/packages/dplyr/dplyr.pdf>

Willse, J. T., & Willse, M. J. T. (2018). Package 'CTT'. *The Comprehensive R Archive Network*. <https://mirror.rcg.sfu.ca/mirror/CRAN/web/packages/CTT/CTT.pdf>

Wright, N. K., & Dorans, N. J. (1993). Using the selection variable for matching or equating. *ETS Research Report Series*, 1993(1), i-22.

<https://doi.org/https://doi.org/10.1002/j.2333-8504.1993.tb01515.x>

Wu, M., Tam, H. P., & Jen, T.-H. (2016). *Educational measurement for applied researchers*. Springer.

- Xin, T., & Zhang, J. (2015). Local equating of cognitively diagnostic modeled observed scores. *Applied Psychological Measurement*, 39(1), 44-61. <https://doi.org/https://doi.org/10.1177/0146621614542427>
- Yıldırım, H., Saraç, M. U., & Büyüköztürk, Ş. (2018). Farklı Örneklem Büyüklüğü ve Dağılımı Koşullarında WLS ve Robust WLS Yöntemlerinin Karşılaştırılması. <http://openaccess.hku.edu.tr/xmlui/handle/20.500.11782/1305>
- Yıldırım Seheryeli, M., Yahşi Sari, H., & Kelecioğlu, H. (2021). Comparison of Kernel Equating and Kernel Local Equating in Item Response Theory Observed Score Equating. *Journal of Measurement and Evaluation in Education and Psychology*, 12(4), 348-357. <https://doi.org/10.21031/epod.900843>
- Yurtçu, M. (2018). *Parametrik Olmayan Bayes Yöntemiyle Ortak Değişkenlere Göre Yapılan Test Eşitlemelerinin Karşılaştırılması* Hacettepe Üniversitesi].
- Zhu, W. (1998). Test equating: What, why, how? *Research Quarterly for Exercise and Sport*, 69(1), 11-23. <https://doi.org/10.1080/02701367.1998.10607662>

**EK-A: TR veri setine ilişkin koşul incelemeleri ve DOG-OD desene ilişkin eşitleme sonuçları**

Kovaryantlar	kategori sayısı	$n_p = 289$						$n_q = 290$				DOG-OD desende eşitleme					
		D-top	r	Çarpıklık	Basıklık	$R^2$	DMF	r	Çarpıklık	Basıklık	$R^2$	DMF	Puan Ort.	SEE	10.PRE	RMSE	RMSD
BSBM20H	4	0,02	-0,24	1,97	3,5	0,705		-0,33	2,04	3,67	0,667		7,67	0,32	1,56	0,71	0,563
BSDGDML	3	0,02	0,07	0,03	-0,26	0,534		-0,11	0,07	-0,38	0,469		7,65	0,33	1,54	0,69	0,535
BSDGHER	3	0,03	-0,36	-0,09	-0,48	0,576		-0,46	-0,18	-0,57	0,555		7,68	0,31	1,48	0,71	0,561
BSBM17F	4	0,05	-0,07	1,52	1,86	0,613		-0,33	1,67	2,46	0,612		7,62	0,32	1,51	0,66	0,494
BSBM19H	4	0,06	0,29	0,29	-1,33	0,34	9*	0,52	0,29	-1,43	0,384		7,64	0,31	1,48	0,67	0,508
BSBM19I	4	0,06	0,26	0,39	-1,2	0,382	9*	0,48	0,29	-1,3	0,365		7,62	0,32	1,47	0,65	0,473
BSDGICM	3	0,06	-0,06	1,3	0,7	0,617		-0,32	1,64	1,76	0,637		7,59	0,33	1,46	0,63	0,44
BSBG13C	4	0,06	-0,09	1,61	1,68	0,624		0,08	1,44	1,36	0,559		7,65	0,33	1,55	0,69	0,534
BSBM26AA	5	0,07	0,03	0,31	-0,67	0,428		0,1	0,35	-0,42	0,414		7,66	0,33	1,56	0,70	0,544
BSDGSB	3	0,08	-0,08	1,75	2,15	0,661		-0,12	1,29	0,69	0,62		7,66	0,33	1,57	0,70	0,553
BSDGSCM	3	0,09	-0,51	-0,67	-0,71	0,561		-0,62	-0,53	-1,06	0,521		7,44	0,30	0,81	0,45	0,229
BSBM19A	4	0,09	-0,39	0,62	-0,17	0,542		-0,46	0,63	-0,47	0,464	1	7,52	0,31	1,12	0,55	0,344
BSBM15	4	0,09	-0,38	0,35	-1,08	0,498	9*	-0,46	0,49	-0,90	0,497	11*	7,56	0,31	1,40	0,59	0,386
BSBM27BA	4	0,09	0,38	0,55	-1,16	0,436		0,29	0,42	-1,21	0,305		7,61	0,32	1,49	0,64	0,466
BSBM16C	4	0,1	0,2	-0,05	-1,44	0,398		0,43	-0,25	-1,38	0,362		7,56	0,32	1,38	0,59	0,391
BSBM19D	4	0,11	-0,35	0,57	-0,44	0,481		-0,38	0,54	-0,6	0,421		7,52	0,32	1,22	0,54	0,333
BSBM16I	4	0,11	-0,23	0,37	-1,34	0,396		-0,38	0,4	-1,29	0,336	1	7,51	0,32	1,25	0,54	0,317
BSBM19B	4	0,12	0,42	0,02	-1,23	0,418		0,51	-0,15	-1,35	0,391		7,48	0,31	1,11	0,51	0,291
BSBM17B	4	0,12	-0,09	1,81	2,94	0,657		-0,22	2,12	4,12	0,679		7,59	0,33	1,43	0,63	0,441
BSBM27AA	3	0,13	-0,35	-0,09	-1,79	0,472		-0,3	0,17	-1,71	0,352		7,57	0,32	1,38	0,61	0,412
ankor3	3	0,14	0,82	0,68	-0,27	0,671		0,93	0,45	-0,96	0,686		7,16	0,25	-0,09	0,23	0,060
BSBM19C	4	0,15	0,42	0,02	-1,34	0,412		0,56	-0,18	-1,44	0,441		7,44	0,31	1,01	0,46	0,234
BSBM19F	4	0,18	-0,31	-0,04	-1,15	0,398		-0,39	0,11	-1,12	0,32		7,46	0,32	1,07	0,48	0,26
ankor5	5	0,22	0,82	0,19	-1,27	0,637	10*	0,89	0,03	-1,56	0,672		7,13	0,24	-0,32	0,23	0,056
BSBM26BA	6	0,23	0,11	0,09	-0,15	0,47		0,15	0,45	0,1	0,489		7,77	0,32	1,72	0,81	0,748
ankor8	8	0,3	0,78	0,29	-1,18	0,567		0,88	0,07	-1,55	0,62		7,14	0,24	-0,30	0,23	0,058
ankor13	13	0,68	0,81	0,41	-1,26	0,614		0,89	0,17	-1,68	0,604		7,05	0,25	-1,50	0,32	0,090

Tabloda D-top sütununa göre sıralanmıştır;

\*B düzeyindeki DMF'yi göstermektedir.

**EK-B: FC veri setine ilişkin koşul incelemeleri ve DOG-OD desene ilişkin eşitleme sonuçları**

Kovaryantlar	kategori sayısı	$n_P = 277$						$n_Q = 281$					DOG-OD desende eşitleme				
		D-top	r	Çarpıklık	Basıklık	$R^2$	DMF	r	Çarpıklık	Basıklık	$R^2$	DMF	Puan Ort.	SEE	10.PRE	RMSE	RMSD
BSBM15	4	0,01	-0,25	1,25	0,71	0,717		-0,2	1,2	0,51	0,672	1	7,24	0,30	0,47	0,34	0,054
BSBM26AA	5	0,02	0,1	1,07	0,73	0,691		0,02	1,03	0,59	0,646		7,25	0,30	0,59	0,34	0,059
BSDGDML	3	0,02	0,08	0,05	-0,31	0,689		-0,06	0,08	-0,38	0,648		7,23	0,30	0,52	0,33	0,053
BSDGHER	3	0,03	-0,49	-0,15	0,88	0,78		-0,4	-0,27	1,4	0,751		7,24	0,30	0,74	0,35	0,06
BSBM19F	4	0,04	-0,39	-0,19	-0,93	0,591		-0,5	-0,24	-0,91	0,605		7,28	0,29	0,61	0,36	0,071
BSBM19D	4	0,05	-0,39	0,05	-0,85	0,619	6*	-0,43	0,14	-0,93	0,581		7,24	0,29	0,59	0,33	0,058
BSBM27AA	3	0,05	0,09	-2,24	3,7	0,789		0,32	-2,65	5,62	0,808		7,18	0,30	0,58	0,30	0,041
BSBG13C	4	0,05	-0,03	0,98	1,48	0,737		-0,14	1,16	1,71	0,715		7,23	0,30	0,49	0,33	0,051
BSBM19C	4	0,06	0,46	0,06	-1,27	0,573	1	0,45	0	-1,36	0,539		7,18	0,29	0,18	0,29	0,033
BSDGICM	3	0,06	0,05	0,14	-0,56	0,655		-0,15	0,25	-0,77	0,617		7,24	0,30	0,54	0,33	0,055
BSBM19B	4	0,07	0,4	-0,3	-0,91	0,607	1	0,5	-0,29	-1,07	0,582	4*	7,20	0,29	0,39	0,28	0,038
ankor3	3	0,08	0,65	0,27	-1,5	0,698	5, 6*	0,68	0,11	-1,71	0,69		7,03	0,27	-0,03	0,16	0,012
BSBM19A	4	0,08	-0,47	0,25	-0,78	0,65	12*	-0,51	0,38	-0,74	0,621		7,09	0,28	0,70	0,22	0,026
BSDGSCM	3	0,08	-0,51	-0,55	-0,82	0,673		-0,56	-0,46	-1,1	0,658		7,05	0,29	0,88	0,15	0,016
BSBM20H	4	0,09	-0,04	0,73	0,62	0,735		-0,15	0,92	0,54	0,662		7,26	0,30	0,57	0,35	0,064
BSBM17F	4	0,09	-0,05	0,2	-0,86	0,646		-0,18	0,45	-0,39	0,599		7,24	0,30	0,52	0,34	0,056
BSBM17B	4	0,1	-0,07	0,68	0,08	0,657		-0,21	0,61	-0,43	0,599		7,25	0,30	0,47	0,34	0,058
BSBM16C	4	0,1	0,12	-0,2	-0,62	0,601		0,19	-0,22	-0,85	0,528		7,19	0,30	0,57	0,29	0,04
BSDGSB	3	0,1	-0,1	1,57	1,52	0,758		-0,15	1,62	1,09	0,801	5*	7,16	0,31	0,58	0,31	0,044
BSBM16I	4	0,11	-0,38	-0,28	-1,19	0,578		-0,35	-0,14	-1,32	0,526		7,16	0,29	0,32	0,28	0,031
BSBM26BA	6	0,12	-0,25	1,08	1,00	0,693		-0,11	1,28	1,44	0,693		7,21	0,30	0,40	0,32	0,044
BSBM19H	4	0,13	0,36	-0,35	-1,08	0,578		0,42	-0,47	-1,16	0,564		7,10	0,29	0,17	0,23	0,022
BSBM19I	4	0,15	0,44	-0,17	-0,94	0,605		0,4	-0,33	-1,04	0,559		7,10	0,29	0,00	0,25	0,024
BSBM27BA	4	0,15	-0,13	2,18	3,63	0,783		-0,31	3,22	9,59	0,832		7,08	0,30	0,72	0,30	0,048
ankor5	5	0,17	0,65	0,24	-1,17	0,66	10*	0,68	0,09	-1,45	0,641		7,06	0,27	-0,31	0,16	0,008
ankor8	8	0,27	0,63	0,38	-1,04	0,638		0,67	0,15	-1,39	0,636	5*	7,03	0,27	-0,29	0,16	0,012
ankor13	13	0,38	0,63	0,51	-0,45	0,646		0,68	0,36	-0,69	0,6		6,98	10,22	0,30	0,07	0,005

Tabloda D-top sütununa göre sıralanmıştır;

\*B düzeyindeki DMF'yi göstermektedir.



**EK-C: SW veri setine ilişkin koşul incelemeleri ve DOG-OD desene ilişkin eşitleme sonuçları**

Kovaryantlar	kategori sayısı	n <sub>P</sub> = 282						n <sub>Q</sub> = 287				DOG-OD desende eşitleme					
		D-top	r	Çarpıklık	Basıklık	R <sup>2</sup>	DMF	r	Çarpıklık	Basıklık	R <sup>2</sup>	DMF	Puan Ort.	SEE	10.PRE	RMSE	RMSD
BSBM19A	4	0,03	-0,4	0,46	-0,36	0,614		-0,47	0,41	-0,51	0,555		6,15	0,27	-0,81	0,99	1,444
BSDGSB	3	0,03	-0,21	1,97	3,09	0,768		-0,06	1,81	2,44	0,707		6,17	0,29	-0,78	0,98	1,433
BSBM27AA	3	0,04	0,24	-1,48	1,02	0,700		0,4	-1,63	1,41	0,686		6,06	0,28	-0,85	1,07	1,716
BSBM17B	4	0,04	-0,15	0,57	-0,43	0,571		-0,21	0,6	-0,26	0,522	12*	6,14	0,29	-0,73	1,01	1,506
BSBG13C	4	0,04	-0,18	0,8	-0,21	0,594		-0,11	0,71	-0,24	0,53		6,14	0,29	-0,72	1,01	1,502
ankor3	3	0,05	0,65	0,2	-1,24	0,677		0,77	0,12	-1,41	0,55	1*	6,00	0,25	-1,30	1,09	1,685
BSBM27BA	4	0,05	-0,28	1,90	3,02	0,73	1	-0,4	2,16	4,13	0,726		6,07	0,28	-0,84	1,07	1,708
BSBM16C	4	0,05	0,38	0,08	-1,14	0,542		0,29	0,05	-1,09	0,425	1	6,11	0,28	-0,87	1,04	1,59
BSDGHER	3	0,05	-0,4	-0,81	-0,18	0,76		-0,47	-0,49	0,41	0,706		6,23	0,28	-0,45	0,92	1,241
BSDGICM	3	0,05	-0,14	0,27	-0,93	0,584		-0,17	0,27	-1,14	0,458		6,13	0,29	-0,74	1,01	1,525
BSDGDML	3	0,05	-0,1	0,12	1,03	0,696		0,05	0,02	0,31	0,616		6,15	0,29	-0,68	0,99	1,464
BSBM19I	4	0,06	0,36	-0,24	-1,00	0,538		0,37	-0,26	-0,98	0,451	1	6,19	0,28	-0,67	0,96	1,373
BSBM17F	4	0,06	-0,18	0,68	-0,09	0,606		-0,24	0,49	-0,45	0,546		6,14	0,28	-0,72	1,01	1,519
BSBM19F	4	0,07	-0,46	-0,15	-0,85	0,581	7*	-0,53	-0,12	-0,9	0,539		6,07	0,27	-1,19	1,07	1,675
BSBM19H	4	0,07	0,44	-0,66	-0,95	0,589		0,52	-0,51	-1,12	0,535		6,20	0,27	-0,52	0,94	1,311
BSBM19D	4	0,07	-0,41	0,28	-0,73	0,589	10*	-0,45	0,17	-0,84	0,513		6,15	0,27	-1,02	1,00	1,474
ankor5	5	0,08	0,62	0,01	-1,09	0,614	1	0,73	-0,07	-1,19	0,589	1	5,99	0,25	-1,02	1,10	1,728
BSBM20H	4	0,08	-0,22	1,1	0,75	0,681		0,03	0,83	0,33	0,643		6,13	0,29	-0,79	1,04	1,623
BSBM19C	4	0,09	0,41	0,12	-1,17	0,521		0,5	0,04	-1,32	0,474		6,09	0,27	-1,07	1,05	1,609
BSBM16I	4	0,09	-0,42	-0,64	-0,89	0,58		-0,44	-0,49	-1,16	0,491		6,06	0,27	-1,13	1,07	1,675
BSBM15	4	0,09	-0,21	0,49	-1,13	0,612		-0,35	0,34	-1,22	0,551		6,21	0,28	-0,50	0,94	1,324
BSBM26BA	6	0,09	-0,17	0,37	-0,63	0,549		-0,18	0,32	-0,61	0,488		6,14	0,29	-0,69	1,00	1,49
BSDGSCM	3	0,13	-0,44	-0,29	-0,85	0,627		-0,6	-0,42	-1,18	0,577		6,10	0,27	-0,97	0,99	1,401
ankor8	8	0,15	0,64	0,06	-1,32	0,611	1	0,76	-0,03	-1,43	0,599	1	5,97	0,25	-1,14	1,11	1,750
BSBM19B	4	0,15	0,44	-0,37	-0,67	0,585		0,52	-0,29	-1,09	0,505		6,13	0,27	-1,20	0,98	1,415
ankor13	13	0,21	0,63	0,47	-0,38	0,582	1*	0,73	0,29	-0,68	0,562	14*	6,08	9,60	-0,69	1,05	1,612
BSBM26AA	5	0,89	0,18	-0,57	0,02	0,632		0,1	-0,28	-0,81	0,611		6,10	55,89	-0,87	1,05	1,64

Tabloda D-top sütununa göre sıralanmıştır;

\*B düzeyindeki DMF'yi göstermektedir.

**EK-Ç: US veri setine ilişkin koşullu incelemeleri ve DOG-OD desene ilişkin eşitleme sonuçları**

Kovaryantlar	kategori sayısı	n <sub>P</sub> = 626						n <sub>Q</sub> = 606				DOG-OD desende eşitleme					
		D-top	r	Çarpıklık	Basıklık	R <sup>2</sup>	DMF	r	Çarpıklık	Basıklık	R	DMF	Puan Ort.	SEE	10.PRE	RMSE	RMSD
BSDGDML	3	0,01	-0,27	0,04	-0,32	0,65		-0,25	0,05	-0,38	0,536		6,81	0,22	0,79	0,44	0,214
BSBM19D	4	0,02	-0,3	0,45	-0,74	0,52		-0,37	0,41	-0,8	0,422		6,83	0,21	0,79	0,43	0,208
BSBM19H	4	0,03	0,35	-0,33	-1,28	0,463		0,41	-0,3	-1,34	0,356		6,82	0,21	0,78	0,43	0,209
BSBM19A	4	0,03	-0,35	0,93	0,32	0,652		-0,38	0,95	0,23	0,586		6,79	0,21	0,70	0,45	0,227
BSDGHER	3	0,03	-0,37	-0,14	0,48	0,744		-0,39	-0,25	0,5	0,686		6,75	0,21	0,67	0,48	0,257
BSBM19I	4	0,03	0,27	0,04	-1,14	0,425		0,31	0,08	-1,12	0,307		6,82	0,22	0,81	0,43	0,211
BSBM17F	4	0,03	-0,23	1,19	0,68	0,683		-0,18	1,14	0,4	0,561		6,82	0,22	0,79	0,43	0,209
BSBM20H	4	0,03	-0,21	1,75	2,79	0,785	1	-0,17	1,66	2,57	0,739		6,79	0,22	0,74	0,45	0,231
BSDGICM	3	0,03	-0,17	0,87	-0,51	0,596		-0,14	0,82	-0,65	0,472		6,82	0,22	0,79	0,44	0,214
BSBM17B	4	0,03	-0,15	1,04	0,06	0,611		-0,15	0,98	-0,02	0,513		6,82	0,22	0,80	0,44	0,213
BSDGSCM	3	0,04	-0,38	-0,3	-1,04	0,561		-0,48	-0,38	-1,1	0,488		6,87	0,21	0,87	0,41	0,183
BSBM16C	4	0,05	0,23	0,21	-1,08	0,418		0,23	0,19	-1,02	0,303		6,76	0,22	0,69	0,47	0,25
BSBG13C	4	0,05	-0,25	0,72	-0,41	0,546		-0,2	0,76	-0,36	0,469		6,77	0,22	0,69	0,46	0,245
BSDGSB	3	0,05	-0,25	1,2	0,4	0,721		-0,23	1,47	1,09	0,668		6,76	0,22	0,67	0,46	0,245
BSBM27AA	3	0,05	0,2	-1,77	1,85	0,741		0,23	-1,92	2,21	0,709		6,75	0,22	0,67	0,48	0,259
BSBM27BA	4	0,05	-0,14	2,32	5,28	0,789		-0,24	2,56	6,1	0,77		6,74	0,22	0,66	0,48	0,266
BSBM19C	4	0,06	0,37	-0,13	-1,36	0,448		0,44	-0,05	-1,4	0,348		6,86	0,21	0,88	0,41	0,189
BSBM19B	4	0,06	0,38	-0,33	-1,08	0,477		0,43	-0,19	-1,16	0,373		6,88	0,21	0,91	0,41	0,187
BSBM15	4	0,08	-0,26	1,02	0,08	0,693		-0,25	1,13	0,25	0,666		6,73	0,22	0,62	0,49	0,278
BSBM26AA	5	0,08	-0,18	0,89	0,02	0,566		-0,17	0,87	-0,23	0,473		6,77	0,22	0,69	0,46	0,241
BSBM26BA	6	0,08	0,11	0,72	0,52	0,612		0,12	0,73	0,35	0,557		6,82	0,22	0,80	0,44	0,213
ankor3	3	0,11	0,83	0,24	-1,51	0,766	1	0,92	0,03	-1,65	0,762		6,47	0,17	-0,22	0,63	0,469
BSBM19F	4	0,11	-0,33	0,27	-0,95	0,48		-0,37	0,19	-0,97	0,38		6,91	0,21	0,97	0,40	0,172
BSBM16I	4	0,14	-0,26	0,05	-1,5	0,399		-0,33	-0,13	-1,51	0,303		6,90	0,22	0,98	0,41	0,18
ankor5	5	0,18	0,83	0,04	-1,42	0,726	10*	0,9	-0,13	-1,49	0,746	1	6,49	0,16	-0,27	0,62	0,441
ankor8	8	0,27	0,82	0,03	-1,41	0,66	10**	0,89	-0,12	-1,49	0,689		6,48	0,16	-0,25	0,64	0,471
ankor13	13	0,44	0,8	0,17	-0,88	0,644	10*	0,85	0,06	-1,03	0,651		6,45	0,16	-0,27	0,64	0,482

Tabloda D-top sütununa göre sıralanmıştır;

\*B düzeyindeki, \*\*C düzeyindeki DMF'yi göstermektedir.



## EK-E: FC veri setine ilişkin DMF analizi sonuçları

FC	P														Q													
	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12	m13	m14	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12	m13	m14
BSBM19B	1																	4										
BSBM19C	1																											
BSBM27BA							7					13	14												11			
BSBM19H																								10				
BSBM19I			3																									
BSBM16C																												
BSBM26BA		2					7																					
BSDGDML												13	14															
BSBM26AA							7																					14
BSDGICM										10		13																
BSBM17F										10																		
BSDGSB																		5							11			
BSBG13C																												
BSBM17B																												
BSBM16I	1				5	6						12																
BSBM20H								9																				
BSBM19F										10	11																	
BSBM19D	1				6						11																	
BSBM27AA												14					3											
BSDGHER																							9					
BSBM15															1													
BSBM19A											12																	
BSDGSCM		2																							12			
ankor3					5	6						13																14
ankor5						6					10		13	14			4											
ankor8				4		6						12		14			4	5						10		12		
ankor13				4																				10				

Gri: Hiçbir maddede DMF yok; Beyaz: A düzeyinde; Sarı: B düzeyinde; Turuncu: C düzeyinde DMF'yi göstermektedir.



## EK-G: US veri setine ilişkin DMF analizi sonuçları

US	P														Q													
	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12	m13	m14	m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12	m13	m14
US																												
BSBM19B																												14
BSBM19C			3	4					9														10					14
BSBM27BA		2			5																							
BSBM19H			3				7		9	10																		14
BSBM19I			3						9																			14
BSBM16C							7										3											
BSBM26BA																												14
BSDGDML			3									12																14
BSBM26AA				4															5		7					13		
BSDGICM																												
BSBM17F																		4					10					
BSDGSB																									13		14	
BSBG13C		2																										14
BSBM17B																2												
BSBM16I			3								12								5									
BSBM20H	1											13														13		
BSBM19F			3								12	13																
BSBM19D			3		5				9			13																
BSBM27AA		2			5																							
BSDGHER																												10
BSBM15											12																	10
BSBM19A			3			6																			11		13	
BSDGSCM			3																							12		
ankor3	1	2							9	10							3	4	5		7		9					
ankor5									9	10		13				1		4	5	6	7		9					14
ankor8									9	10		13						4	5	6	7	8						14
ankor13		2								10								4										

Gri: Hiçbir maddede DMF yok; Beyaz: A düzeyinde; Sarı: B düzeyinde; Turuncu: C düzeyinde DMF'yi göstermektedir.

**EK-H: 2 ile 3. Kitapçıkta elde edilen veri setine ilişkin koşul incelemeleri ve DOG-OD desene ilişkin eşitleme sonuçları**

Kovaryantlar	kategori sayısı	D-top	n <sub>P</sub> = 942				n <sub>O</sub> = 975				Puan Ort.	DOG-OD desende eşitleme			
			r	Çarpıklık	Basıklık	R <sup>2</sup>	r	Çarpıklık	Basıklık	R <sup>2</sup>		SEE	10.PRE	RMSE	RMSD
ASBG05C	2	0,00	-0,12	1,30	-0,31	0,79	-0,07	1,30	-0,32	0,79	4,59	0,14	-11,05	1,46	2,45
ASBE01BA	2	0,00	0,16	-1,95	1,80	0,84	0,24	-1,96	1,84	0,84	4,59	0,14	-11,09	1,47	2,45
ASBG05D	2	0,00	-0,20	5,42	27,45	0,90	-0,23	5,87	32,45	0,91	4,58	0,14	-11,11	1,47	2,47
ASBM05A	4	0,01	-0,27	1,15	0,97	0,81	-0,33	1,16	0,94	0,81	4,58	0,14	-11,13	1,47	2,46
ASBG10D	4	0,01	-0,07	1,83	2,98	0,85	-0,08	1,85	3,17	0,85	4,58	0,14	-11,09	1,47	2,46
ASBG11J	4	0,01	0,35	-4,57	20,53	0,93	0,19	-4,84	23,68	0,92	4,59	0,14	-10,98	1,46	2,46
ASBM05I	4	0,01	0,29	-0,46	-1,12	0,66	0,31	-0,48	-1,08	0,65	4,58	0,14	-11,17	1,47	2,48
ASBG07	2	0,01	0,14	2,08	2,34	0,85	-0,01	2,28	3,19	0,86	4,59	0,14	-11,01	1,46	2,44
ASBG06A	4	0,02	-0,05	1,28	0,92	0,85	0,00	1,42	1,59	0,84	4,59	0,14	-11,00	1,46	2,44
ASBE01BB	2	0,02	0,18	-2,16	2,68	0,85	0,21	-2,40	3,75	0,86	4,57	0,14	-11,22	1,48	2,50
ASBM02G	4	0,02	-0,17	0,68	-0,71	0,67	-0,23	0,66	-0,76	0,68	4,60	0,14	-10,95	1,46	2,43
ASDGSL	3	0,02	-0,02	0,61	-0,86	0,70	0,00	0,57	-0,91	0,70	4,59	0,14	-11,06	1,46	2,45
ASBE01BF	2	0,02	0,13	-3,40	9,54	0,89	0,03	-4,32	16,67	0,90	4,58	0,14	-11,15	1,47	2,48
ASBM05F	4	0,02	-0,27	0,48	-0,79	0,67	-0,29	0,54	-0,70	0,68	4,58	0,14	-11,17	1,48	2,49
ASBG06B	4	0,02	-0,01	1,12	0,59	0,83	-0,03	1,19	0,73	0,83	4,60	0,14	-10,91	1,45	2,41
ASBM05C	4	0,03	0,33	-0,90	-0,41	0,74	0,35	-0,84	-0,56	0,73	4,60	0,13	-10,93	1,45	2,38
ankor3	3	0,03	0,71	0,22	-1,58	0,80	0,71	0,17	-1,59	0,80	4,53	0,12	-11,79	1,52	2,60
ASBG05I	2	0,03	0,05	0,11	-1,99	0,67	0,07	0,00	-2,00	0,68	4,58	0,14	-11,14	1,47	2,47
ASBM03F	4	0,03	-0,10	2,10	3,98	0,86	-0,06	1,98	3,46	0,85	4,59	0,14	-10,97	1,46	2,44
ASBM04A	4	0,03	0,07	-0,23	-0,79	0,70	0,00	-0,21	-0,85	0,70	4,59	0,14	-11,01	1,46	2,44
ASBE02B	4	0,03	0,15	0,33	-1,44	0,62	0,19	0,32	-1,43	0,63	4,58	0,14	-11,12	1,47	2,46
ASDGSLM	3	0,03	-0,18	0,26	-1,31	0,63	-0,28	0,20	-1,33	0,66	4,61	0,14	-10,81	1,45	2,39
ASBG10E	4	0,03	0,01	1,51	1,64	0,82	0,01	1,55	1,72	0,82	4,59	0,14	-11,06	1,46	2,44
ASBM05D	4	0,03	-0,23	0,88	-0,13	0,71	-0,29	0,78	-0,39	0,72	4,60	0,14	-10,91	1,45	2,39
ASBE02D	4	0,03	0,22	-0,57	-0,87	0,65	0,15	-0,56	-0,89	0,67	4,60	0,14	-10,94	1,46	2,43
ankor5	5	0,04	0,71	0,19	-1,26	0,77	0,70	0,14	-1,25	0,77	4,57	0,12	-10,98	1,48	2,50
ASBM02E	4	0,04	-0,17	0,96	-0,19	0,72	-0,25	0,90	-0,42	0,73	4,60	0,14	-10,90	1,45	2,40
ASBG10C	4	0,04	-0,03	1,26	0,70	0,78	-0,09	1,19	0,41	0,77	4,60	0,14	-10,97	1,46	2,42
ASBM03C	4	0,04	0,00	1,65	2,48	0,84	-0,07	1,38	1,40	0,83	4,59	0,14	-11,08	1,47	2,46
ASBM02A	4	0,04	-0,18	1,02	0,17	0,74	-0,25	0,99	0,05	0,76	4,61	0,14	-10,80	1,44	2,39
ASBM05E	4	0,05	0,22	-0,65	-0,95	0,69	0,29	-0,59	-1,00	0,69	4,60	0,14	-10,88	1,45	2,39
ASDGSCM	3	0,05	-0,35	0,18	-1,26	0,67	-0,41	0,11	-1,14	0,70	4,63	0,13	-10,36	1,42	2,33
ASBM02I	4	0,05	-0,21	0,54	-1,21	0,65	-0,30	0,46	-1,33	0,66	4,61	0,14	-10,89	1,44	2,38
ASDGHRL	3	0,05	-0,27	-0,45	-1,65	0,86	-0,08	-0,22	-1,76	0,85	4,55	0,14	-11,57	1,50	2,56
ASBM02C	4	0,05	0,15	-0,67	-0,89	0,67	0,21	-0,59	-1,04	0,67	4,61	0,14	-10,80	1,44	2,38
ASBM04D	4	0,06	0,16	0,07	-1,21	0,65	0,08	-0,02	-1,18	0,66	4,57	0,14	-11,27	1,48	2,51

ASBE01BC	2	0,07	0,06	-1,11	-0,77	0,77	0,21	-1,56	0,44	0,82	4,55	0,14	-11,52	1,50	2,58
ASBG04	5	0,07	0,26	0,07	-0,64	0,65	0,16	0,02	-0,76	0,63	4,57	0,14	-11,27	1,49	2,53
ASBG08	5	0,07	0,19	-1,25	0,45	0,77	0,15	-1,39	0,81	0,79	4,58	0,14	-11,13	1,48	2,50
ASBM05B	4	0,08	0,36	-0,49	-1,07	0,67	0,39	-0,36	-1,14	0,66	4,65	0,13	-10,26	1,40	2,26
ankor8	8	0,08	0,71	0,21	-1,32	0,77	0,69	0,14	-1,33	0,76	4,57	0,12	-11,02	1,47	2,47
ASBM05H	4	0,09	0,35	-0,75	-0,83	0,72	0,39	-0,67	-0,98	0,72	4,64	0,13	-10,42	1,41	2,29
ASBE02C	4	0,09	0,14	-0,14	-1,35	0,58	0,21	0,00	-1,37	0,60	4,62	0,14	-10,64	1,43	2,35
ankor13	13	0,11	0,69	0,25	-0,63	0,75	0,69	0,14	-0,64	0,75	4,63	0,12	-9,78	1,43	2,36

Tabloda D-top sütununa göre sıralanmıştır.



**EK-I: 4 ile 5. Kitapçıkta elde edilen veri setine ilişkin koşul incelemeleri ve DOG-OD desene ilişkin eşitleme sonuçları**

Kovaryantlar	kategori sayısı	D-top	r	n <sub>P</sub> = 942			r	n <sub>Q</sub> = 975			Puan Ort.	DOG-OD desende eşitleme			
				Çarpıklık	Basıklık	R <sup>2</sup>		Çarpıklık	Basıklık	R <sup>2</sup>		SEE	10.PRE	RMSE	RMSD
ASBG05D	2	0,00	-0,25	4,71	20,18	0,91	-0,23	4,59	19,10	0,90	6,51	0,13	3,59	0,69	0,54
ASBG11J	4	0,00	0,22	-4,76	23,83	0,92	0,21	-4,80	24,53	0,92	6,51	0,13	3,59	0,69	0,53
ASBE01BA	2	0,01	0,19	-1,66	0,74	0,82	0,27	-1,71	0,94	0,82	6,50	0,13	3,59	0,68	0,53
ASBE01BC	2	0,01	0,13	-1,50	0,25	0,82	0,19	-1,55	0,41	0,80	6,50	0,13	3,58	0,68	0,53
ASDGHRL	3	0,01	-0,27	-0,56	-1,62	0,87	-0,16	-0,60	-1,57	0,87	6,51	0,13	3,61	0,69	0,55
ASBM03C	4	0,01	-0,04	1,52	2,11	0,83	-0,08	1,56	2,24	0,83	6,51	0,13	3,59	0,69	0,54
ASBE01BF	2	0,01	0,24	-3,84	12,75	0,89	0,11	-3,41	9,65	0,88	6,52	0,13	3,60	0,70	0,55
ASBG07	2	0,02	0,04	2,46	4,06	0,86	-0,04	2,22	2,91	0,84	6,51	0,13	3,59	0,69	0,54
ASBM05D	4	0,02	-0,29	0,84	-0,27	0,72	-0,35	0,81	-0,32	0,73	6,49	0,13	3,62	0,68	0,52
ASBE02B	4	0,02	0,28	0,31	-1,46	0,63	0,12	0,30	-1,48	0,57	6,50	0,13	3,58	0,68	0,53
ASDGSCM	3	0,02	-0,44	0,17	-1,06	0,73	-0,43	0,14	-1,13	0,69	6,53	0,13	3,62	0,71	0,57
ASDGSLM	3	0,02	-0,17	0,23	-1,34	0,60	-0,26	0,19	-1,32	0,60	6,52	0,13	3,62	0,70	0,56
ASBG05C	2	0,02	-0,01	1,56	0,43	0,81	-0,05	1,39	-0,08	0,79	6,51	0,13	3,59	0,69	0,54
ASBM04D	4	0,02	0,11	0,09	-1,14	0,62	0,23	0,10	-1,17	0,62	6,51	0,13	3,59	0,69	0,54
ASBG06B	4	0,02	-0,07	1,22	0,93	0,82	-0,03	1,28	1,31	0,81	6,50	0,13	3,57	0,68	0,52
ASBG05I	2	0,03	0,10	-0,09	-1,99	0,65	-0,05	0,01	-2,00	0,60	6,51	0,13	3,59	0,69	0,54
ASDGSLs	3	0,03	0,02	0,64	-0,90	0,68	0,01	0,58	-0,92	0,66	6,50	0,13	3,58	0,69	0,53
ASBM05A	4	0,03	-0,32	1,17	1,12	0,82	-0,36	1,14	0,85	0,81	6,49	0,13	3,58	0,67	0,51
ASBM05E	4	0,03	0,28	-0,62	-0,95	0,68	0,29	-0,58	-1,02	0,66	6,52	0,13	3,64	0,70	0,57
ASBM05F	4	0,03	-0,29	0,50	-0,82	0,65	-0,28	0,54	-0,76	0,64	6,49	0,13	3,57	0,67	0,51
ASBG06A	4	0,03	-0,13	1,48	1,78	0,84	0,01	1,33	1,33	0,84	6,49	0,13	3,55	0,67	0,51
ASBM03F	4	0,04	-0,12	2,11	4,46	0,87	-0,04	2,01	3,54	0,85	6,53	0,13	3,61	0,71	0,57
ASBG08	5	0,04	0,20	-1,45	1,09	0,79	0,24	-1,45	1,07	0,79	6,50	0,13	3,60	0,68	0,52
ASBM02E	4	0,04	-0,19	0,96	-0,27	0,71	-0,25	0,88	-0,37	0,70	6,52	0,13	3,63	0,70	0,56
ASBM02C	4	0,04	0,18	-0,61	-0,93	0,67	0,25	-0,67	-0,88	0,65	6,51	0,13	3,60	0,69	0,54
ASBM02G	4	0,04	-0,13	0,59	-0,91	0,63	-0,19	0,59	-0,82	0,64	6,51	0,13	3,60	0,69	0,54
ASBM05B	4	0,05	0,37	-0,48	-0,98	0,67	0,36	-0,43	-1,10	0,64	6,51	0,13	3,59	0,69	0,54
ASBE01BB	2	0,05	0,09	-1,69	0,87	0,82	0,16	-2,22	2,91	0,84	6,48	0,13	3,54	0,66	0,50
ASBM05H	4	0,05	0,38	-0,75	-0,85	0,72	0,39	-0,66	-0,96	0,69	6,54	0,13	3,68	0,72	0,59
ASBM05I	4	0,06	0,35	-0,51	-0,99	0,67	0,38	-0,40	-1,15	0,64	6,55	0,13	3,69	0,73	0,61
ASBM04A	4	0,06	-0,01	-0,24	-0,79	0,68	0,10	-0,22	-0,87	0,68	6,51	0,13	3,59	0,69	0,54
ASBM02A	4	0,06	-0,18	1,01	0,08	0,74	-0,26	0,88	-0,11	0,73	6,52	0,13	3,64	0,70	0,56
ASBM02I	4	0,07	-0,21	0,51	-1,23	0,63	-0,24	0,38	-1,37	0,62	6,53	0,13	3,63	0,71	0,57
ASBE02D	4	0,07	0,26	-0,57	-0,82	0,67	0,20	-0,53	-0,89	0,62	6,53	0,13	3,65	0,71	0,58
ASBG10D	4	0,08	0,00	1,63	2,08	0,83	-0,06	1,86	3,15	0,85	6,50	0,13	3,58	0,68	0,53
ASBG10E	4	0,09	0,01	1,43	1,07	0,79	0,04	1,59	1,76	0,82	6,51	0,13	3,59	0,69	0,54

ASBG10C	4	0,09	-0,10	1,11	0,10	0,75	-0,05	1,24	0,58	0,77	6,49	0,13	3,56	0,67	0,51
ASBM05C	4	0,09	0,33	-1,00	-0,15	0,75	0,33	-0,86	-0,49	0,71	6,57	0,13	3,74	0,75	0,64
ankor3	3	0,10	0,66	0,07	-1,37	0,77	0,72	-0,10	-1,32	0,79	6,34	0,12	3,20	0,53	0,28
ASBE02C	4	0,12	0,27	-0,10	-1,28	0,61	0,15	-0,04	-1,35	0,52	6,55	0,13	3,68	0,73	0,61
ASBG04	5	0,15	0,22	-0,01	-0,67	0,63	0,23	0,03	-0,67	0,60	6,57	0,13	3,72	0,75	0,64
ankor5	5	0,18	0,68	0,35	-1,06	0,75	0,73	0,18	-1,14	0,77	6,31	0,11	3,10	0,52	0,26
ankor8	8	0,30	0,67	0,34	-1,29	0,74	0,72	0,15	-1,38	0,76	6,31	0,11	3,10	0,51	0,25
ankor13	13	0,34	0,68	0,52	-0,33	0,71	0,71	0,26	-0,54	0,75	6,34	0,11	2,96	0,49	0,25

Tabloda D-top sütununa göre sıralanmıştır.

**EK-İ: 5 ile 6. Kitapçıkta elde edilen veri setine ilişkin koşul incelemeleri ve DOG-OD desene ilişkin eşitleme sonuçları**

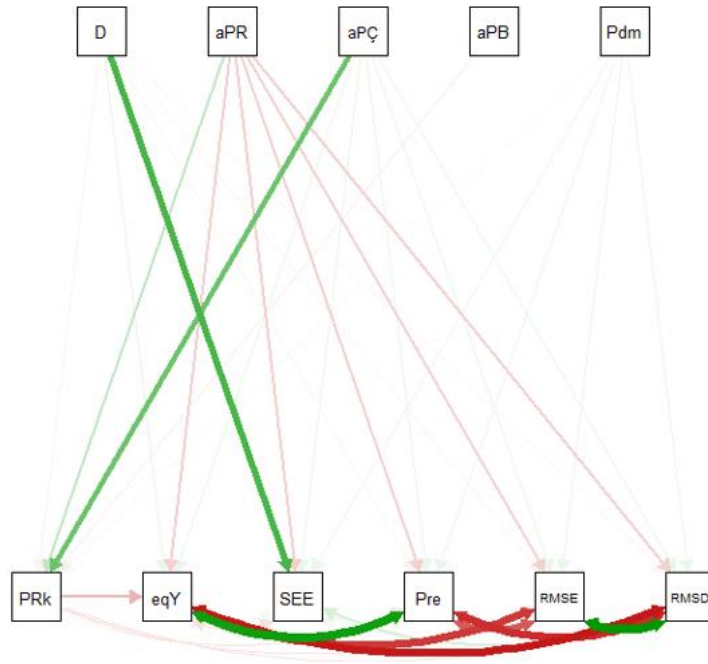
Kovaryantlar	kategori sayısı	D-top	n <sub>P</sub> = 942				n <sub>Q</sub> = 975				Puan Ort.	DOG-OD desende eşitleme			
			r	Çarpıklık	Basıklık	R <sup>2</sup>	r	Çarpıklık	Basıklık	R <sup>2</sup>		SEE	10.PRE	RMSE	RMSD
ASBG05D	2	0,00	-0,22	4,59	19,10	0,91	-0,27	4,43	17,67	0,90	5,51	0,15	-1,74	0,51	0,34
ASDGHRL	3	0,01	-0,18	-0,60	-1,57	0,89	-0,22	-0,62	-1,54	0,87	5,51	0,15	-1,72	0,51	0,34
ASBG05C	2	0,01	0,00	1,39	-0,08	0,82	-0,03	1,43	0,04	0,80	5,51	0,15	-1,75	0,52	0,35
ASBG07	2	0,01	-0,01	2,22	2,91	0,87	-0,12	2,11	2,47	0,84	5,51	0,15	-1,73	0,52	0,34
ASBM05B	4	0,01	0,38	-0,43	-1,10	0,71	0,41	-0,42	-1,12	0,68	5,51	0,14	-1,75	0,51	0,34
ASBM05D	4	0,02	-0,27	0,81	-0,32	0,76	-0,31	0,77	-0,38	0,74	5,51	0,14	-1,71	0,51	0,34
ASBE01BA	2	0,02	0,18	-1,71	0,94	0,85	0,22	-1,87	1,48	0,83	5,49	0,15	-1,82	0,53	0,36
ASDGSLM	3	0,02	-0,28	0,19	-1,32	0,70	-0,27	0,21	-1,35	0,65	5,49	0,14	-1,86	0,53	0,36
ASBG06A	4	0,02	0,05	1,33	1,33	0,85	-0,08	1,36	1,49	0,84	5,50	0,15	-1,77	0,52	0,35
ASBE01BF	2	0,02	0,03	-3,41	9,65	0,90	0,26	-4,10	14,85	0,89	5,49	0,15	-1,80	0,53	0,36
ASBE01BB	2	0,02	0,11	-2,22	2,91	0,87	0,17	-1,96	1,82	0,84	5,52	0,15	-1,68	0,50	0,33
ASBM04A	4	0,02	0,07	-0,22	-0,87	0,71	0,06	-0,25	-0,85	0,72	5,50	0,15	-1,77	0,52	0,35
ASBE01BC	2	0,02	0,21	-1,55	0,41	0,83	0,13	-1,74	1,04	0,82	5,49	0,15	-1,83	0,53	0,37
ASBG06B	4	0,02	0,02	1,28	1,31	0,83	-0,04	1,30	1,21	0,82	5,52	0,15	-1,68	0,51	0,33
ASBG11J	4	0,02	0,19	-4,80	24,53	0,93	0,15	-4,87	23,61	0,92	5,52	0,15	-1,72	0,51	0,33
ASDGSCM	3	0,03	-0,42	0,14	-1,13	0,74	-0,44	0,15	-1,22	0,72	5,50	0,14	-1,89	0,53	0,35
ASBM05I	4	0,03	0,38	-0,40	-1,15	0,69	0,37	-0,45	-1,10	0,67	5,50	0,14	-1,82	0,53	0,36
ASBM05E	4	0,03	0,30	-0,58	-1,02	0,71	0,28	-0,65	-0,89	0,70	5,49	0,14	-1,85	0,54	0,37
ASBG05I	2	0,03	0,02	0,01	-2,00	0,70	0,01	0,13	-1,99	0,68	5,51	0,15	-1,74	0,52	0,34
ankor5	5	0,03	0,72	-0,14	-1,13	0,80	0,69	-0,14	-1,18	0,77	5,48	0,12	-2,21	0,54	0,36
ASBG10D	4	0,03	-0,01	1,86	3,15	0,85	-0,07	1,70	2,53	0,83	5,51	0,15	-1,71	0,51	0,34
ankor3	3	0,04	0,75	0,11	-1,26	0,83	0,72	0,09	-1,36	0,81	5,45	0,12	-2,61	0,57	0,39
ASDGSLs	3	0,04	0,03	0,58	-0,92	0,72	0,00	0,66	-0,82	0,71	5,51	0,15	-1,74	0,52	0,34
ASBM02C	4	0,04	0,21	-0,67	-0,88	0,71	0,20	-0,62	-0,97	0,68	5,51	0,15	-1,72	0,51	0,34
ASBG10E	4	0,04	0,07	1,59	1,76	0,83	0,01	1,50	1,42	0,80	5,50	0,15	-1,79	0,52	0,35
ASBM02G	4	0,04	-0,20	0,59	-0,82	0,69	-0,21	0,65	-0,80	0,67	5,50	0,15	-1,80	0,52	0,35
ASBM04D	4	0,04	0,21	0,10	-1,17	0,67	0,10	0,07	-1,12	0,65	5,51	0,15	-1,69	0,51	0,34
ASBM05H	4	0,04	0,36	-0,66	-0,96	0,73	0,43	-0,73	-0,89	0,72	5,47	0,14	-1,97	0,55	0,38
ASBG10C	4	0,05	-0,01	1,24	0,58	0,78	-0,09	1,15	0,27	0,76	5,52	0,15	-1,70	0,51	0,34
ASBM02A	4	0,05	-0,22	0,88	-0,11	0,76	-0,22	1,00	0,09	0,75	5,49	0,15	-1,84	0,54	0,37
ASBM02E	4	0,05	-0,26	0,88	-0,37	0,75	-0,23	0,91	-0,44	0,73	5,49	0,14	-1,85	0,53	0,36
ASBG08	5	0,05	0,21	-1,45	1,07	0,82	0,13	-1,50	1,25	0,79	5,49	0,15	-1,87	0,54	0,37
ASBM05C	4	0,05	0,33	-0,86	-0,49	0,75	0,31	-0,85	-0,62	0,73	5,50	0,14	-1,87	0,53	0,36
ASBM03C	4	0,05	-0,01	1,56	2,24	0,85	-0,09	1,65	2,46	0,84	5,50	0,15	-1,77	0,52	0,35
ankor8	8	0,06	0,73	-0,11	-1,34	0,79	0,71	-0,10	-1,40	0,77	5,46	0,12	-2,40	0,55	0,37
ASBG04	5	0,06	0,20	0,03	-0,67	0,66	0,20	-0,01	-0,67	0,63	5,48	0,15	-1,91	0,54	0,38

ASBM05A	4	0,06	-0,33	1,14	0,85	0,84	-0,35	1,02	0,51	0,80	5,56	0,14	-1,45	0,46	0,27
ASBE02B	4	0,06	0,12	0,30	-1,48	0,64	0,21	0,23	-1,45	0,61	5,50	0,15	-1,79	0,53	0,36
ASBM03F	4	0,06	-0,01	2,01	3,54	0,86	-0,01	2,34	5,42	0,87	5,51	0,15	-1,75	0,52	0,35
ASBM02I	4	0,07	-0,26	0,38	-1,37	0,68	-0,27	0,49	-1,31	0,66	5,48	0,14	-1,91	0,54	0,37
ASBM05F	4	0,07	-0,26	0,54	-0,76	0,70	-0,27	0,42	-0,92	0,66	5,55	0,14	-1,50	0,47	0,29
ASBE02D	4	0,07	0,18	-0,53	-0,89	0,68	0,20	-0,64	-0,73	0,68	5,48	0,15	-1,93	0,55	0,39
ASBE02C	4	0,09	0,16	-0,04	-1,35	0,60	0,21	-0,09	-1,29	0,60	5,48	0,15	-1,90	0,54	0,38
ankor13	13	0,09	0,72	0,11	-0,72	0,76	0,69	0,03	-0,81	0,74	5,50	0,12	-1,77	0,52	0,35

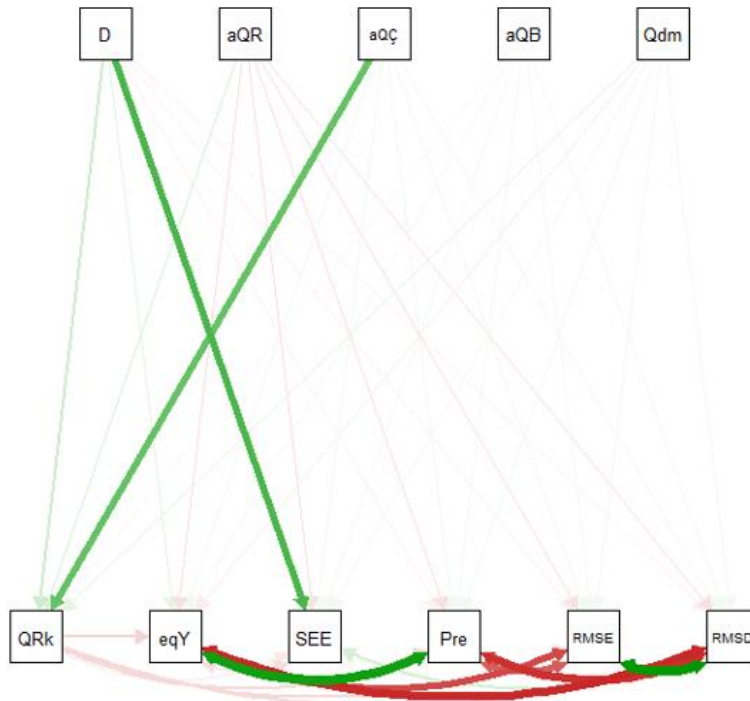
Tabloda D-top sütununa göre sıralanmıştır.

## EK-J: Aşama 1 için P ve Q evrenlerinde yol şemaları

P evreni için yol şeması

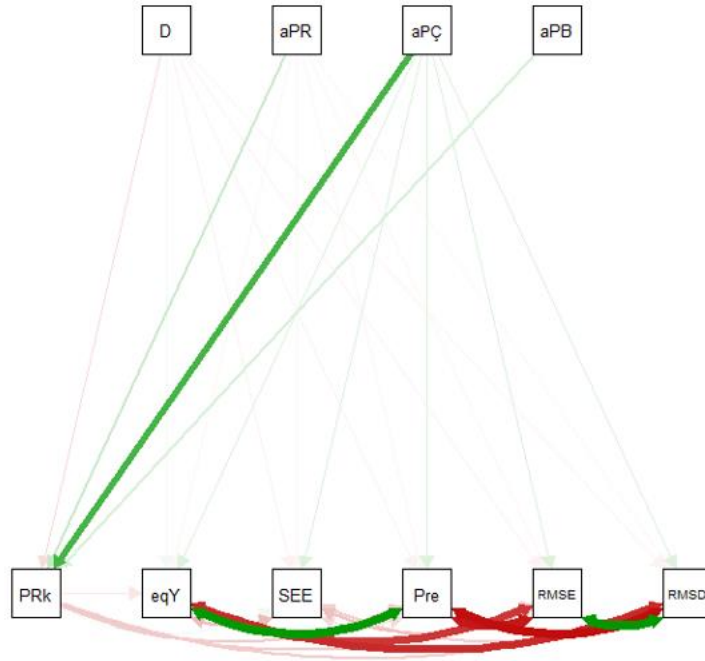


Q evreni için yol şeması

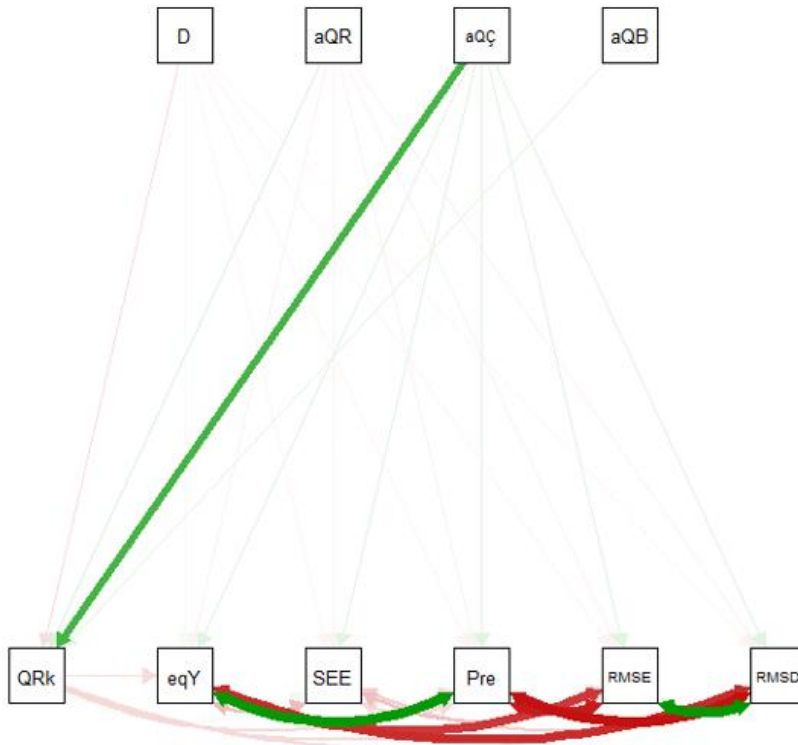


### EK-K: Aşama 2 için P ve Q evrenlerinde yol şemaları

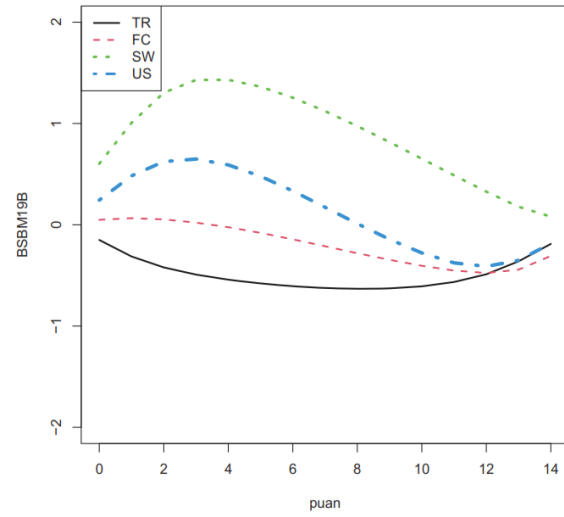
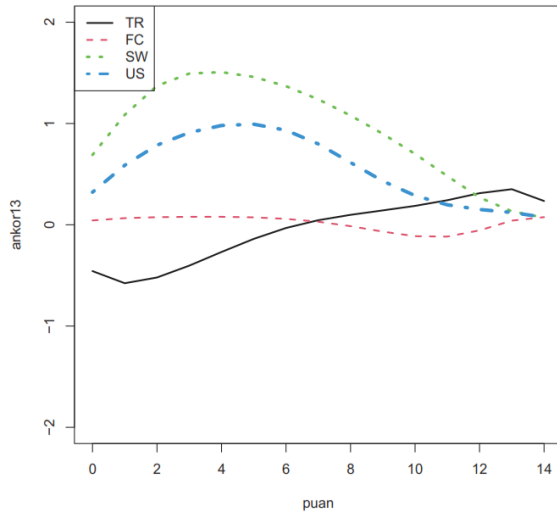
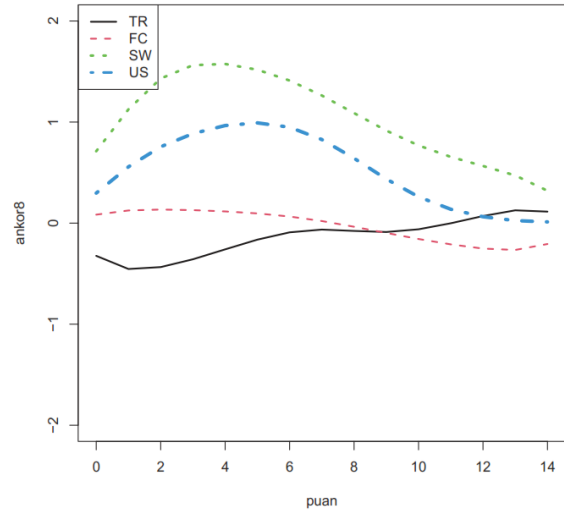
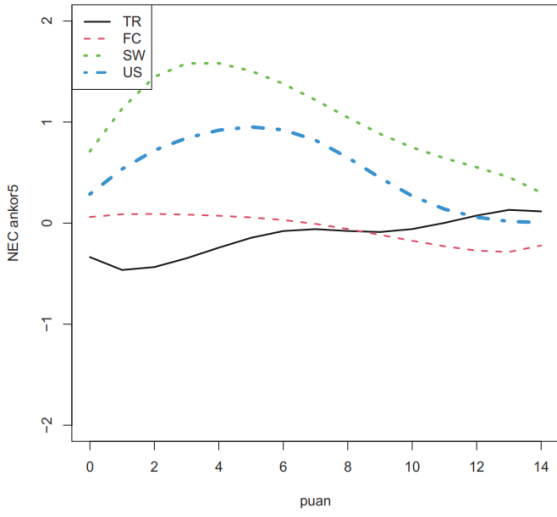
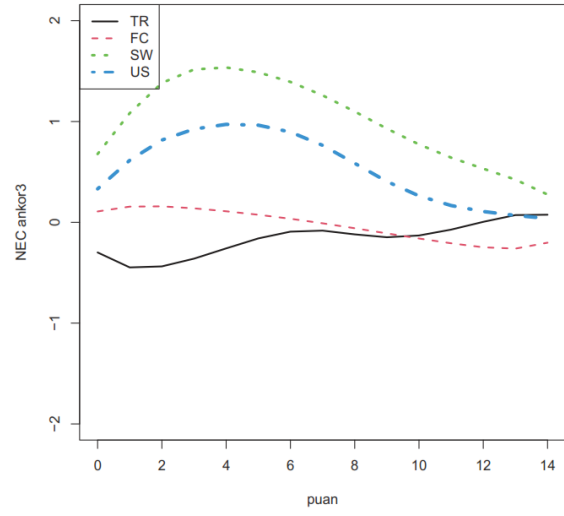
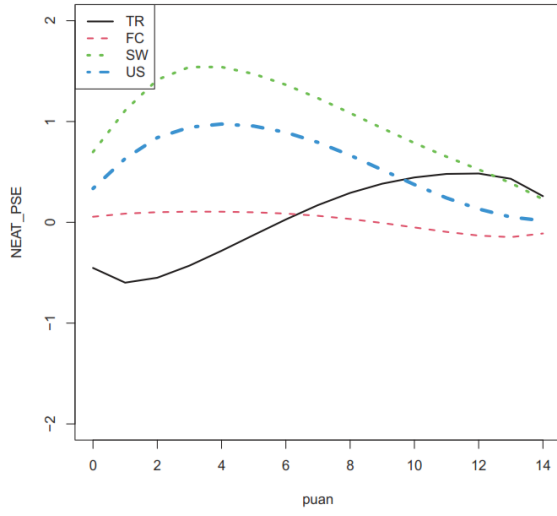
P evreni için yol şeması

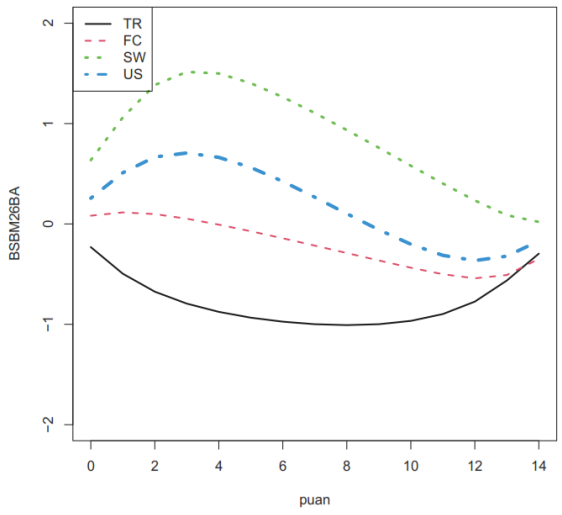
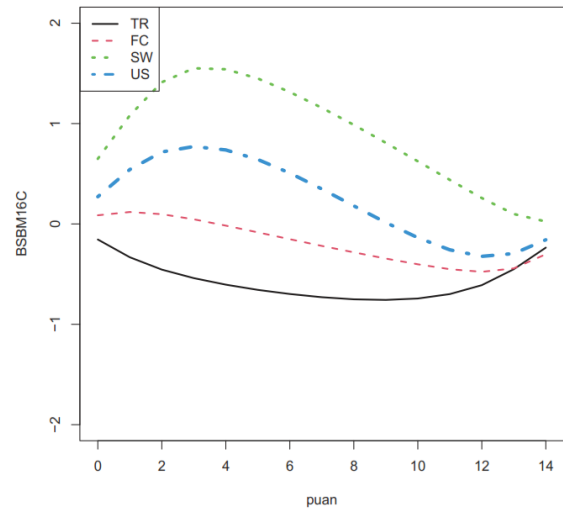
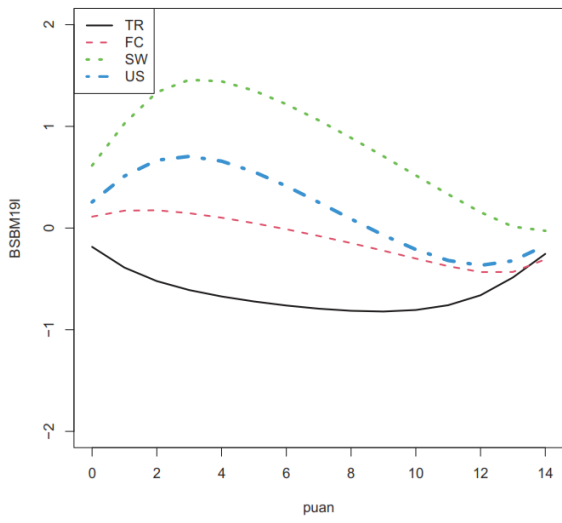
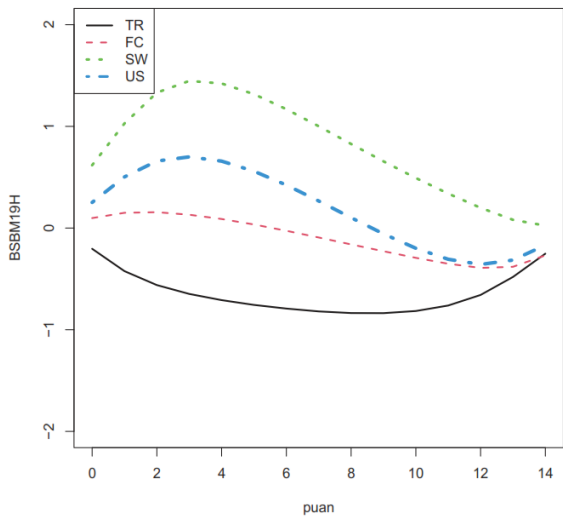
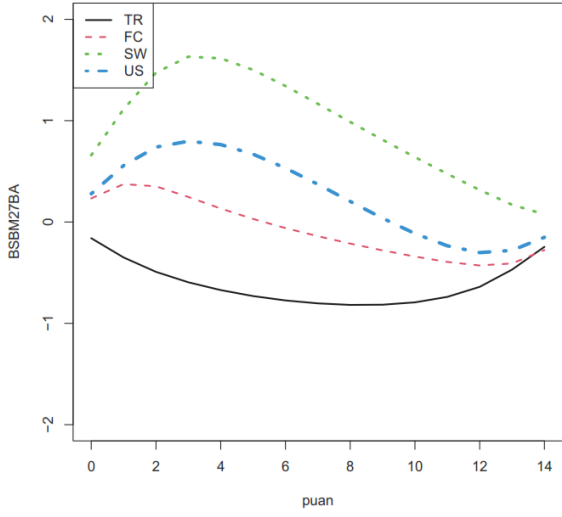
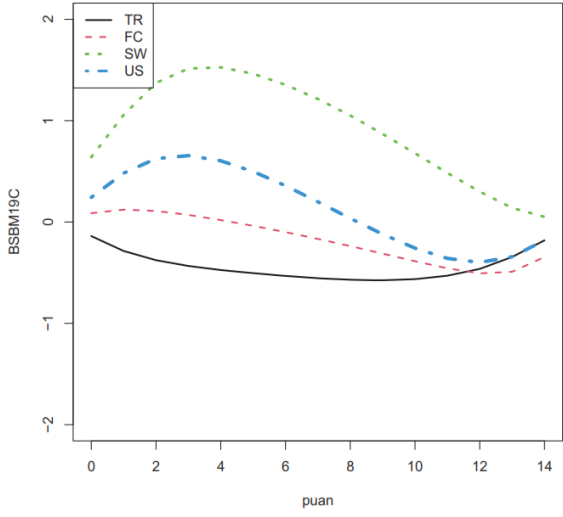


Q evreni için yol şeması

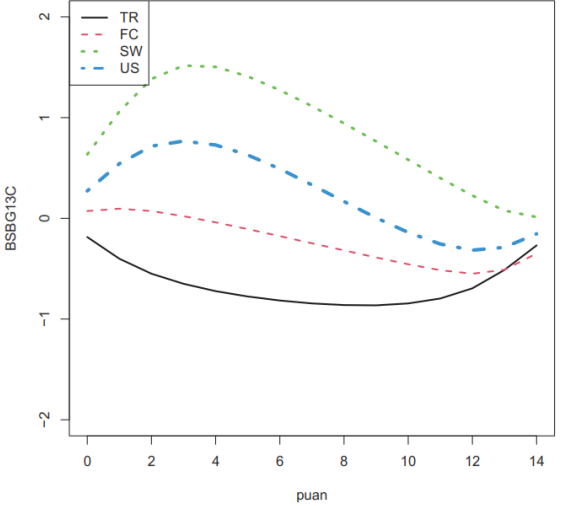
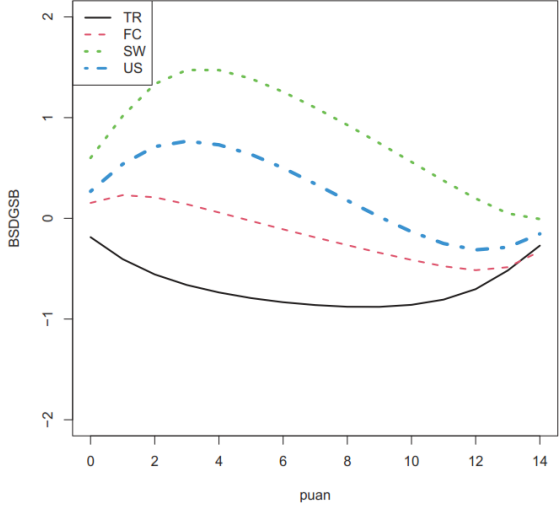
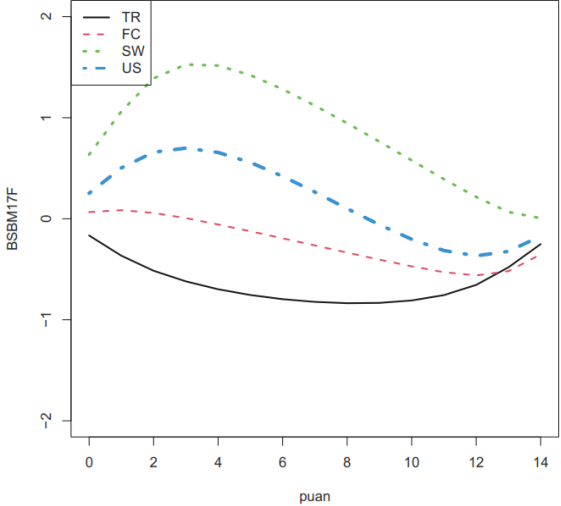
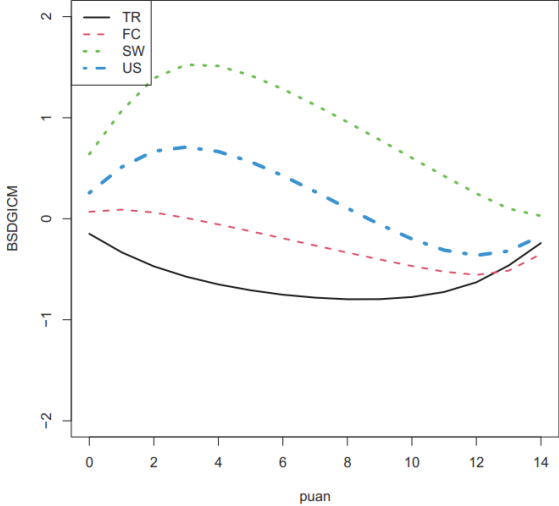
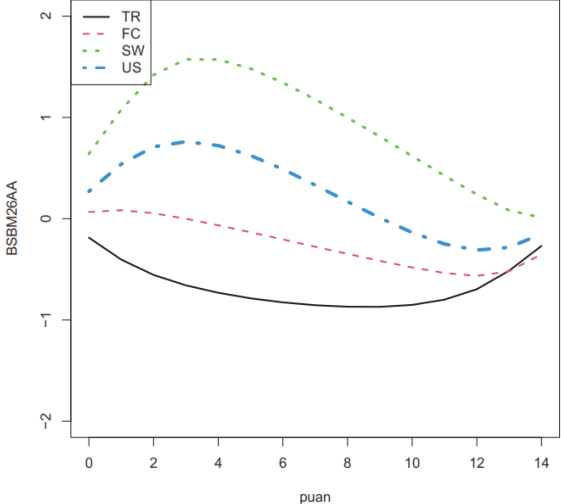
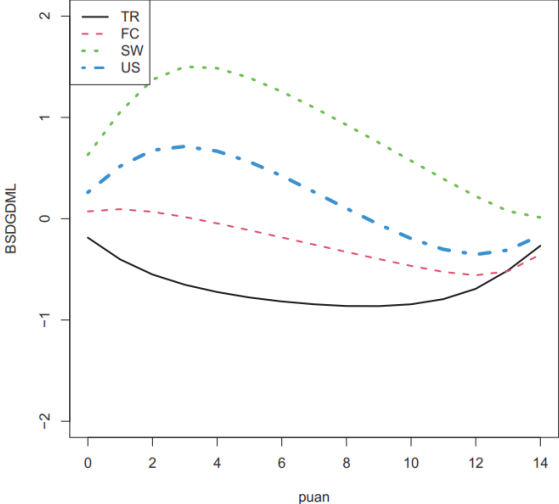


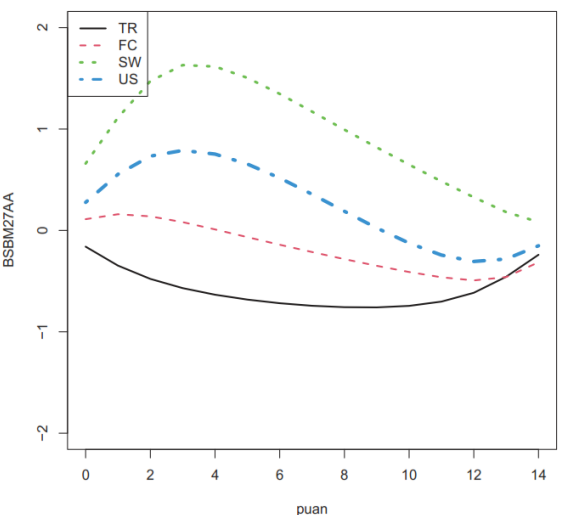
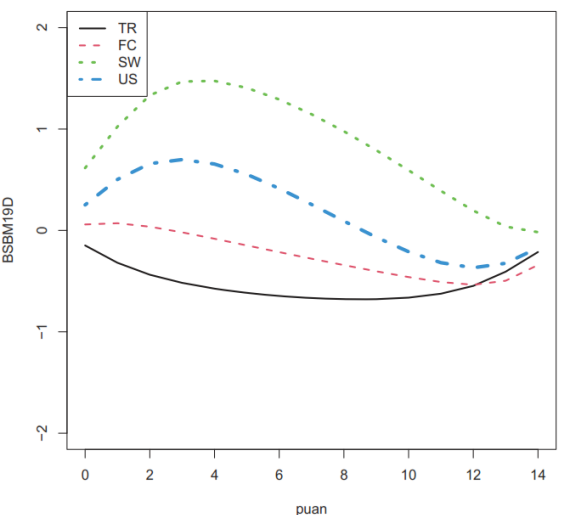
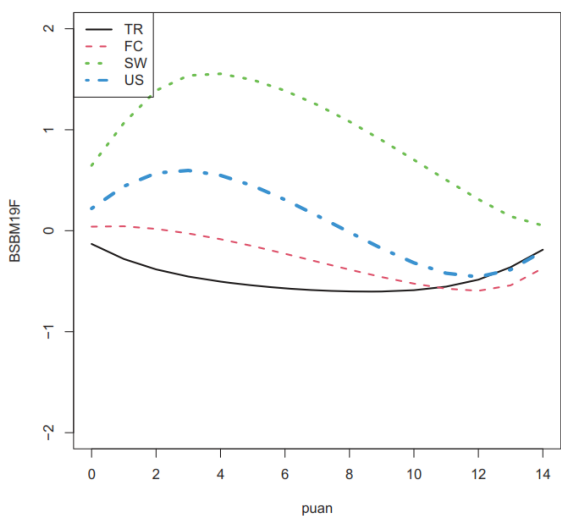
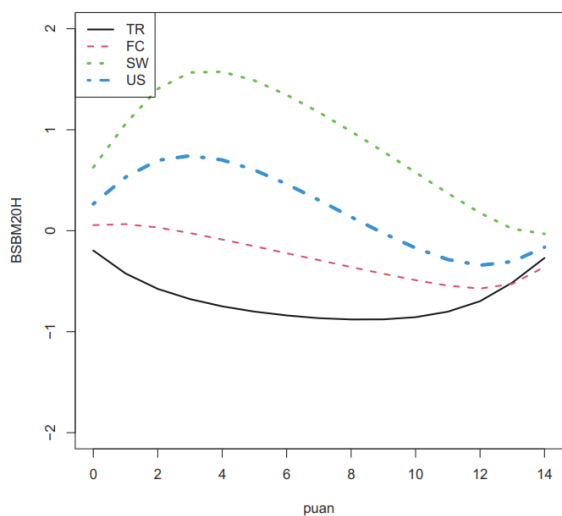
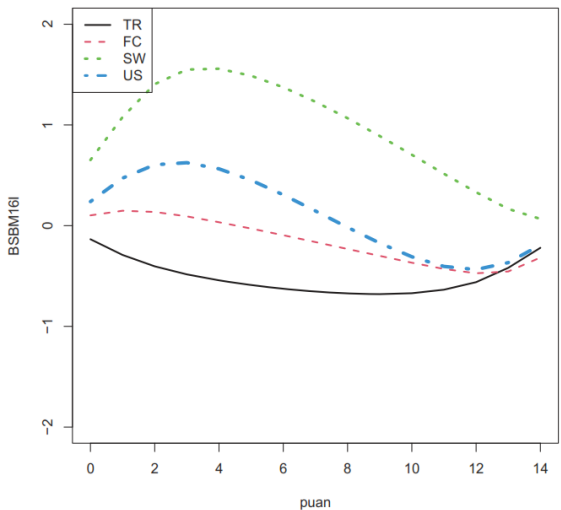
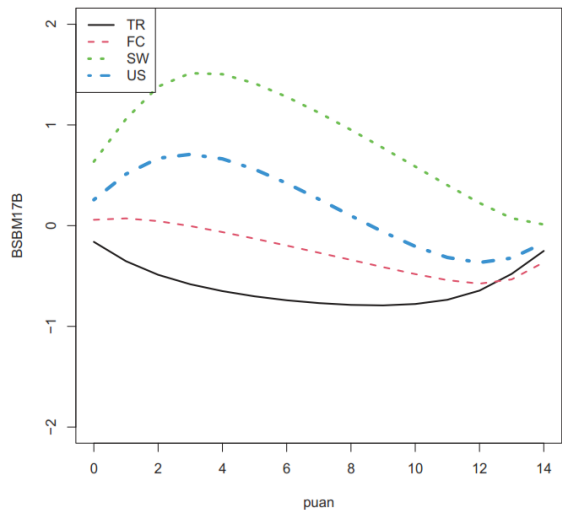
## EK-L: Aşama 1 için ham puanlar ile eşitlenmiş puanlar arasındaki fark puan dağılımları

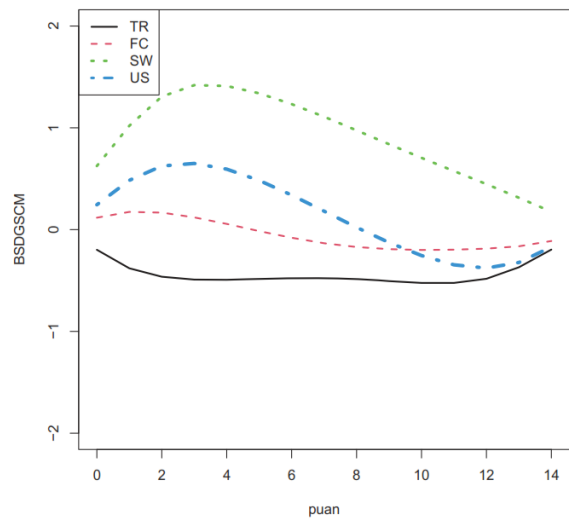
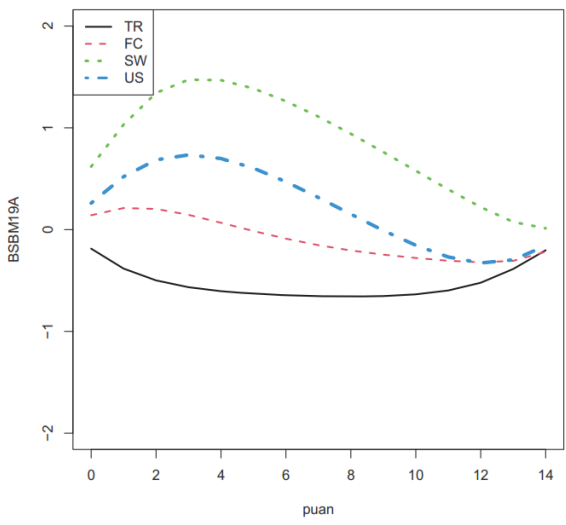
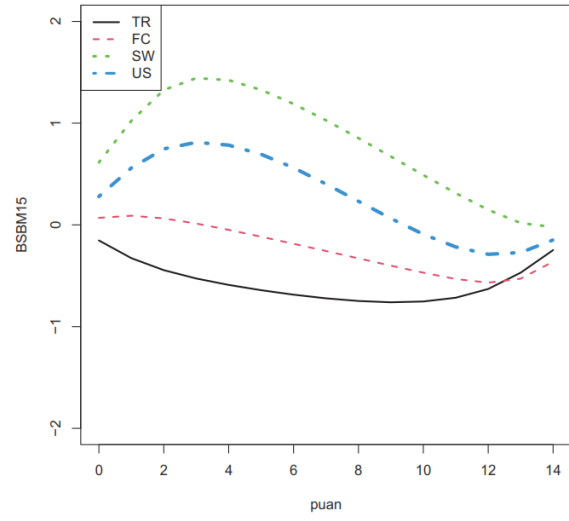
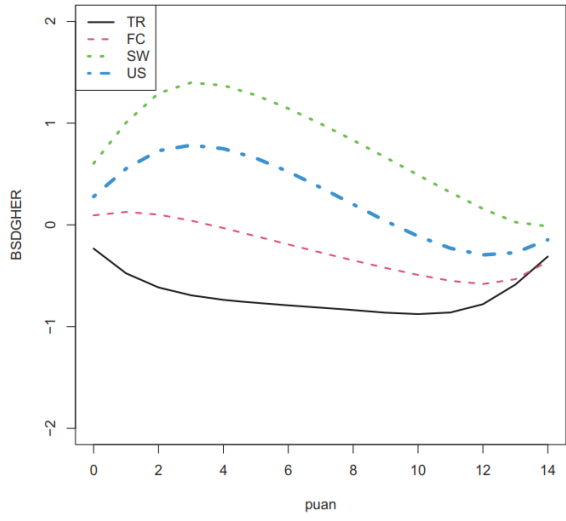




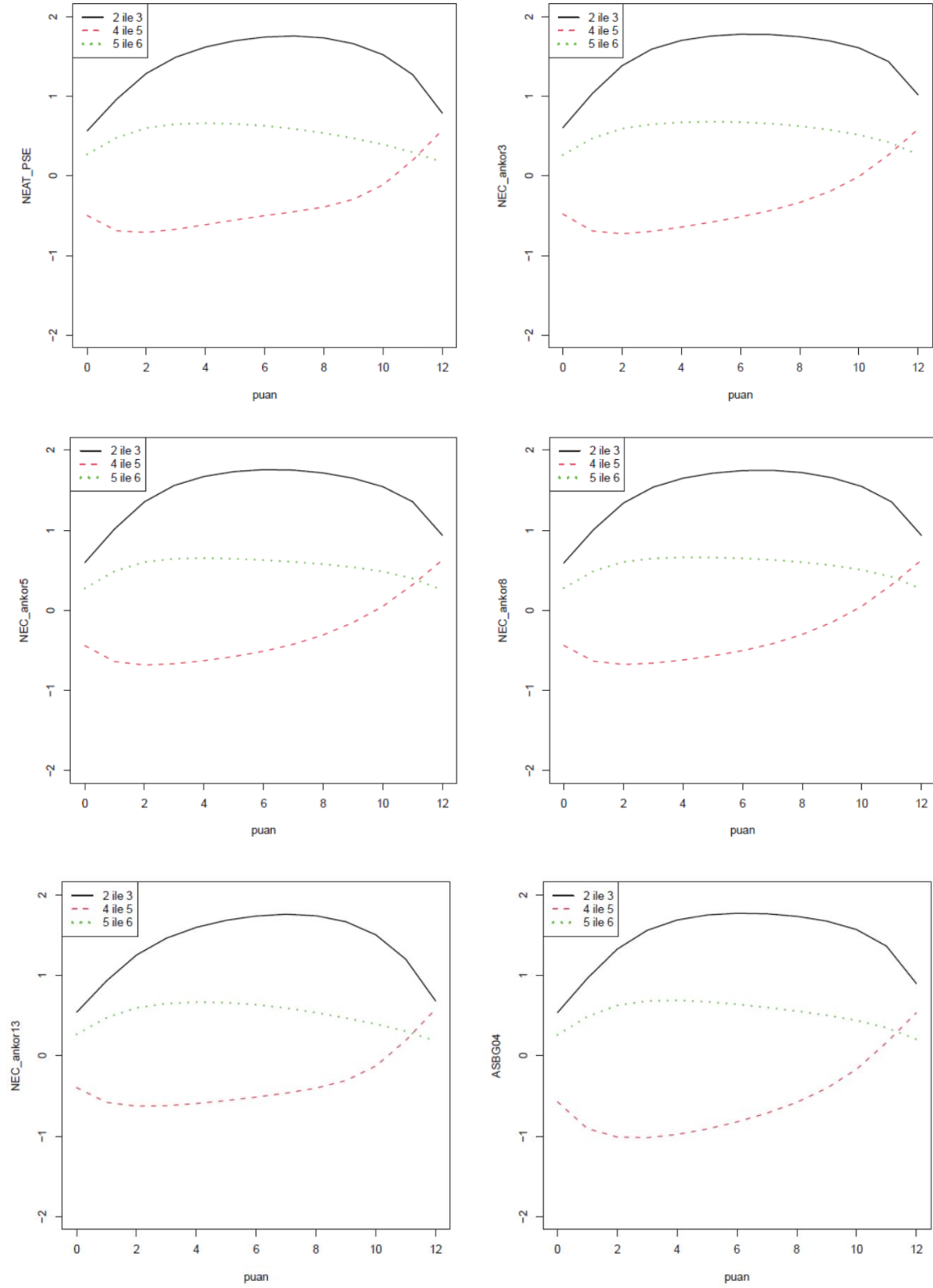


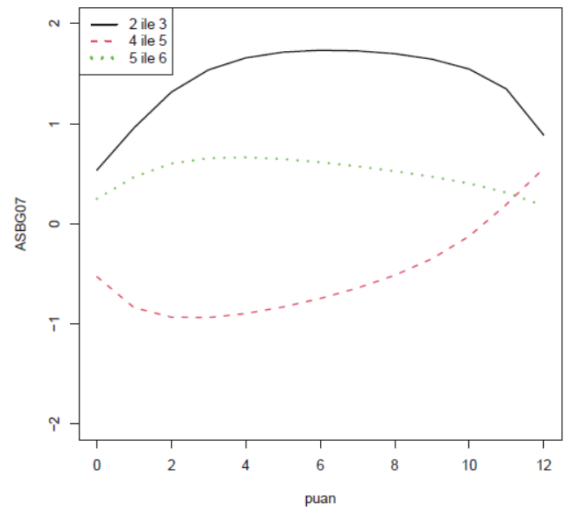
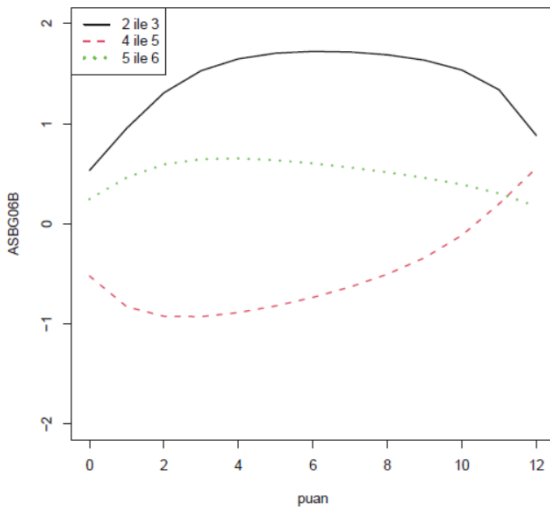
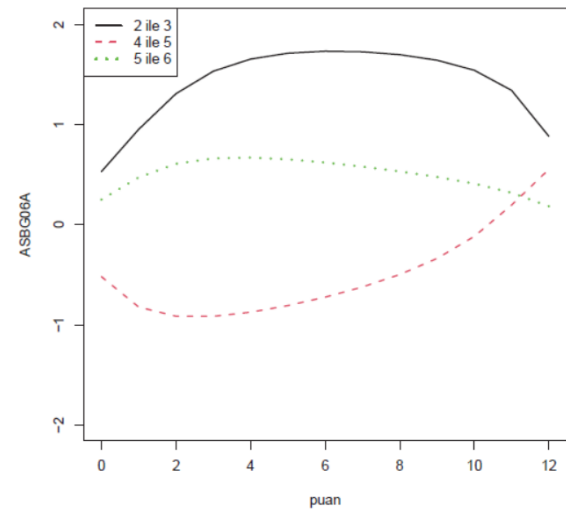
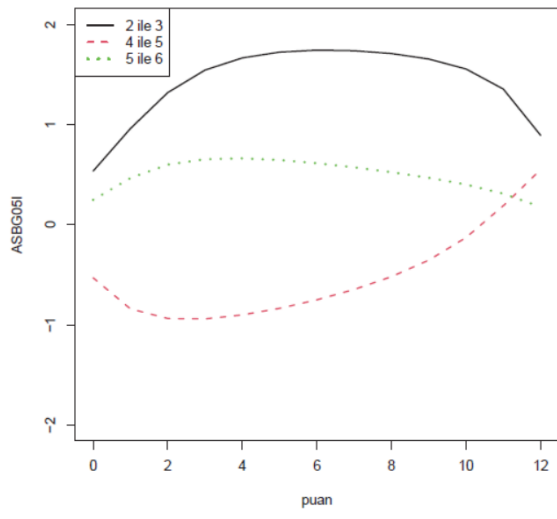
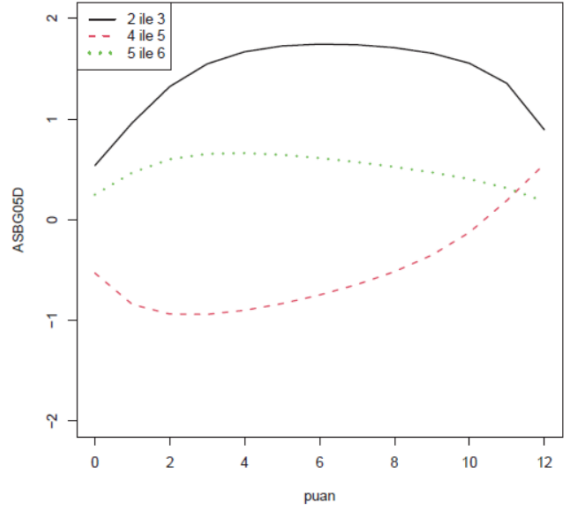
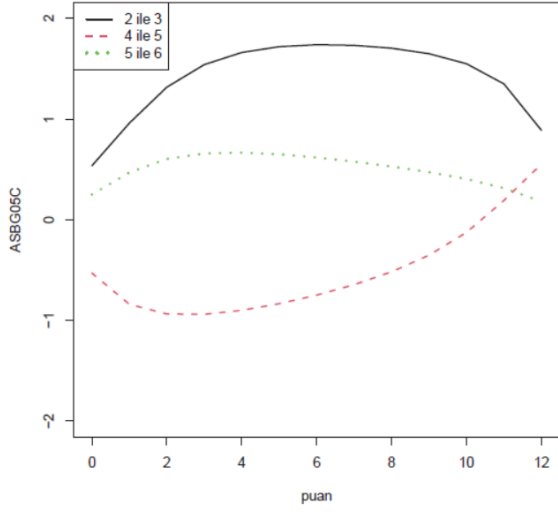


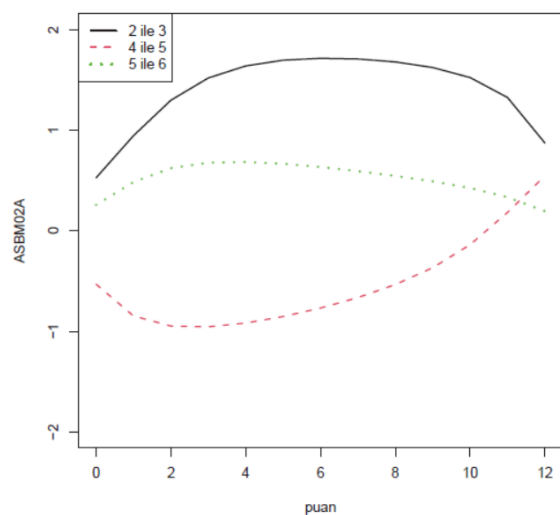
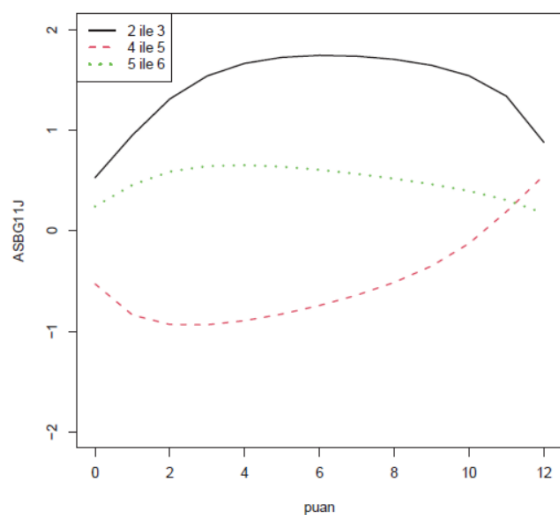
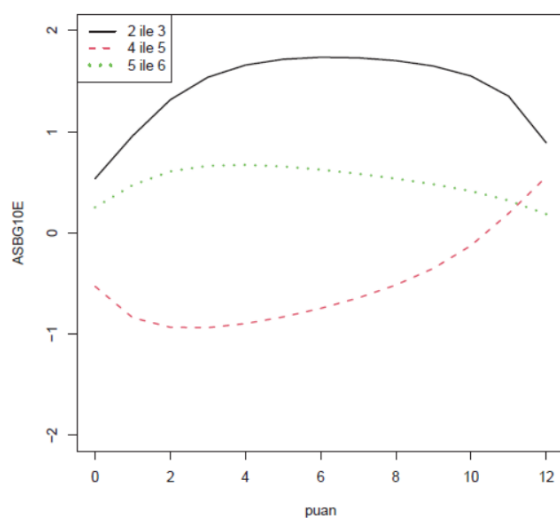
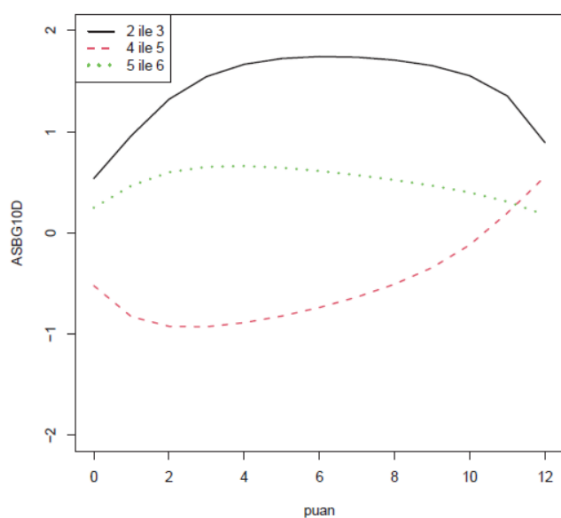
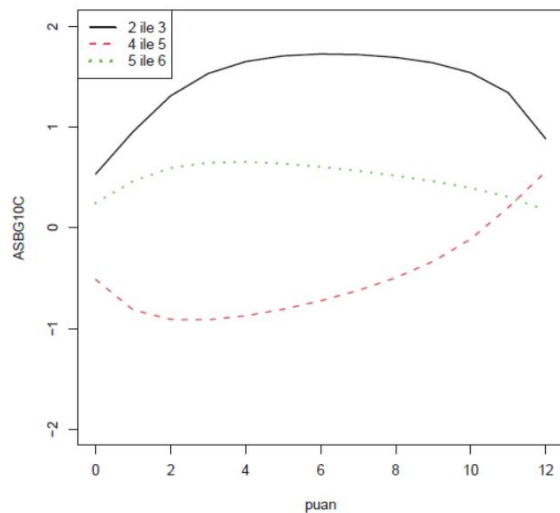
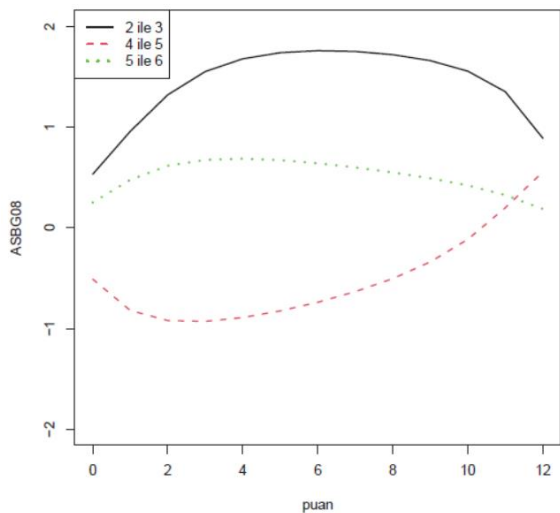


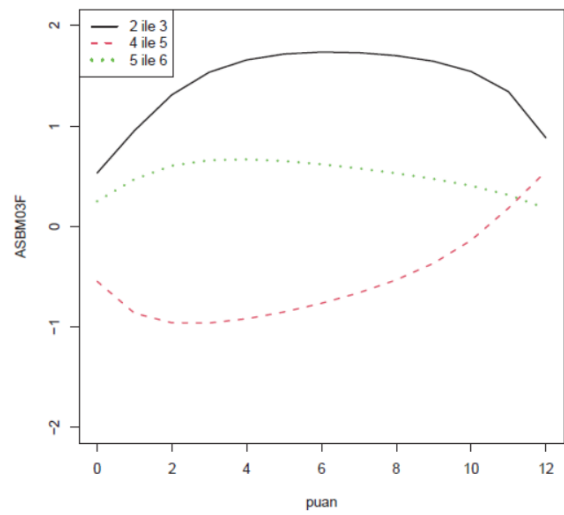
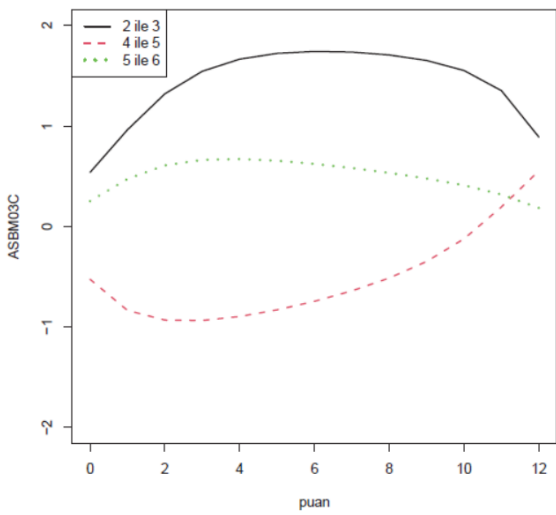
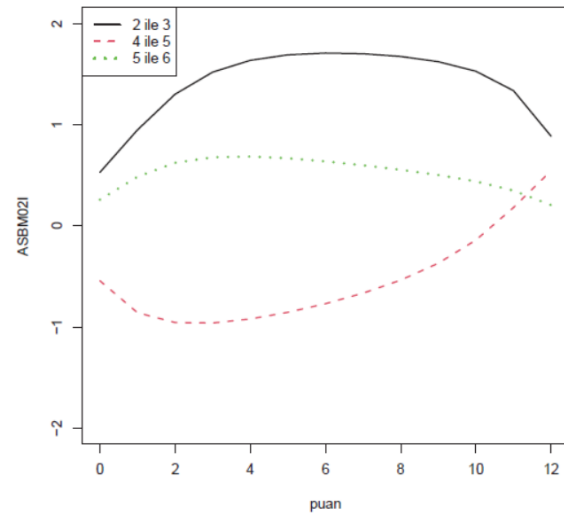
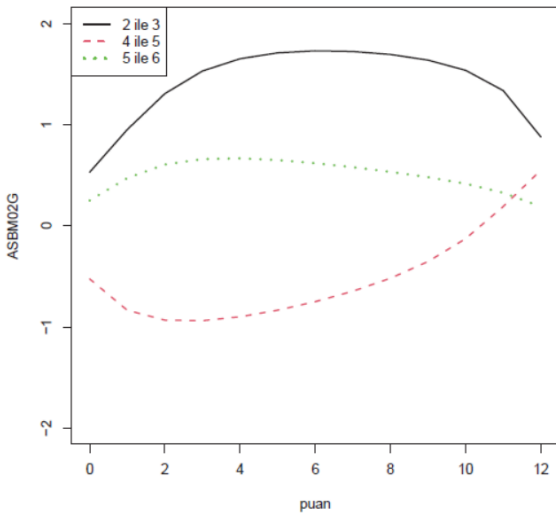
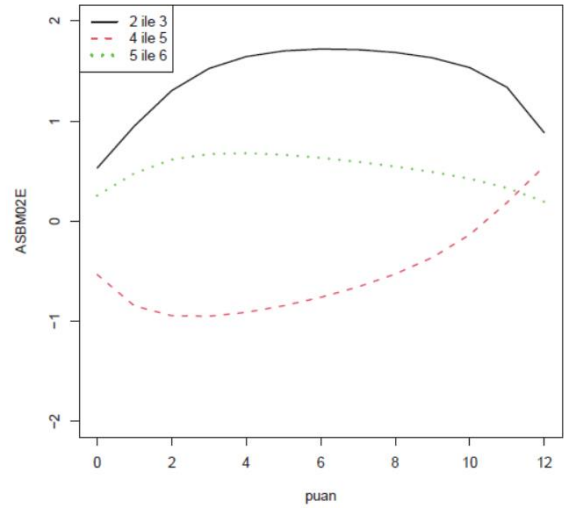
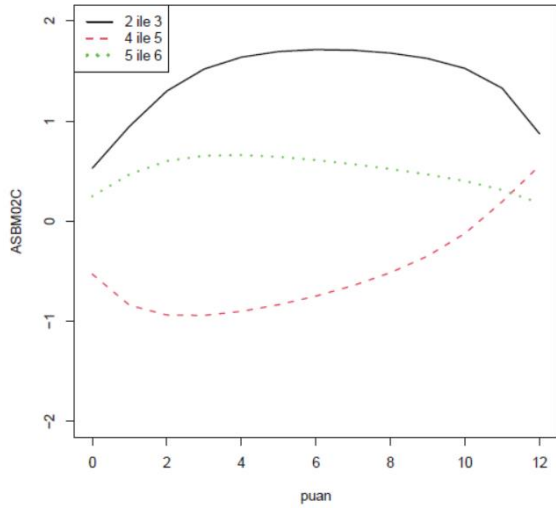


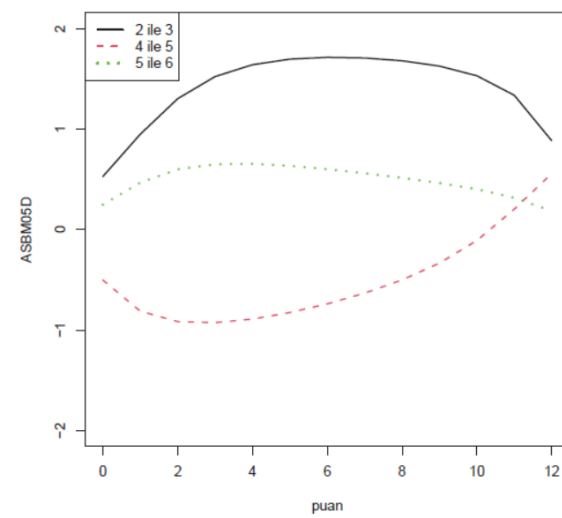
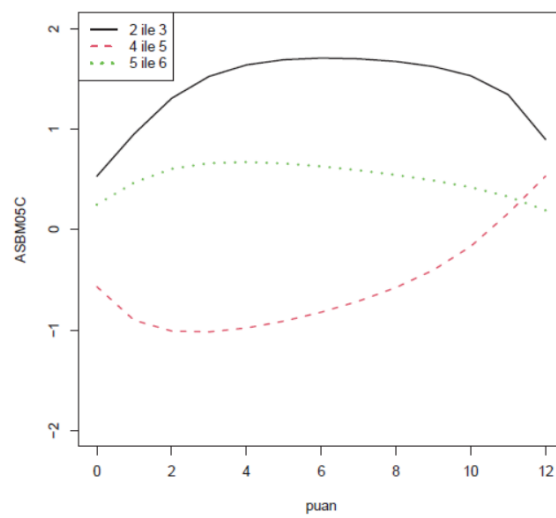
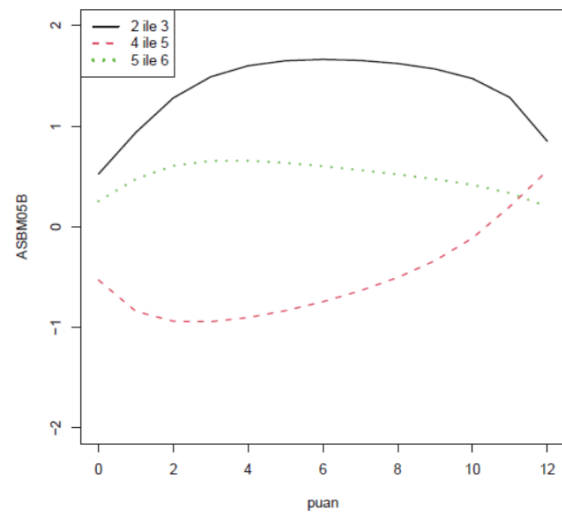
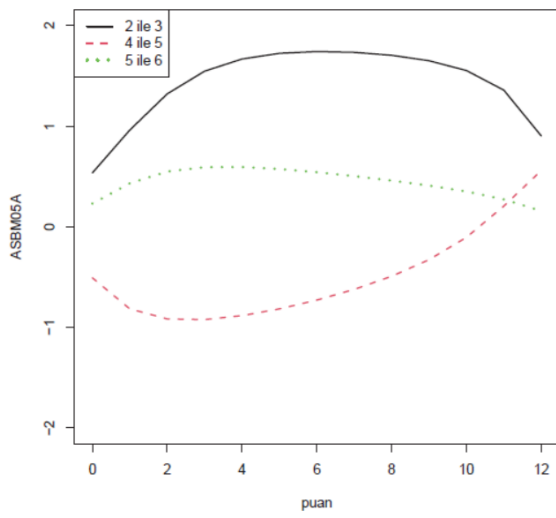
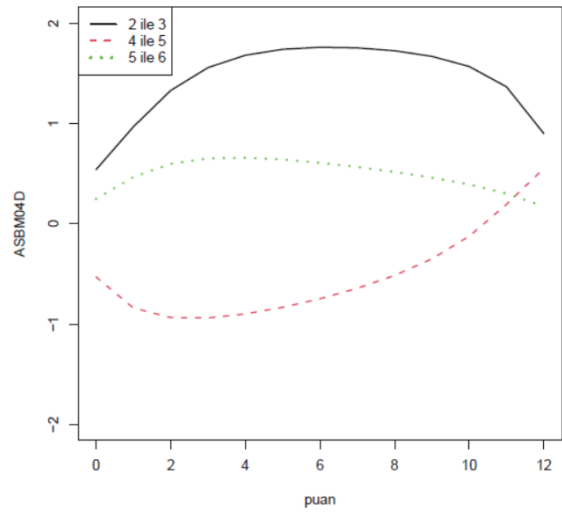
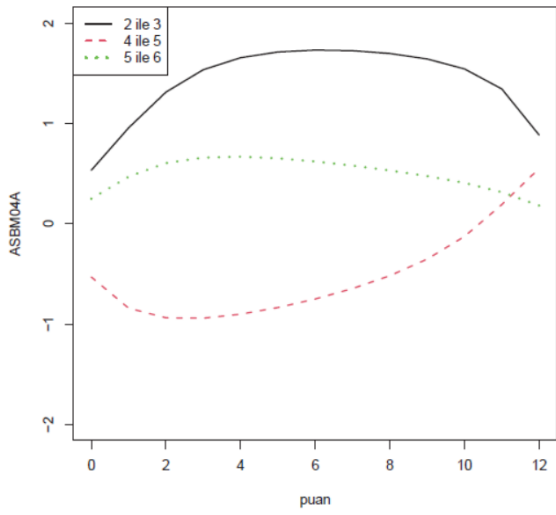
## EK-M: Aşama 2 için ham puanlar ile eşitlenmiş puanlar arasındaki fark puan dağılımları



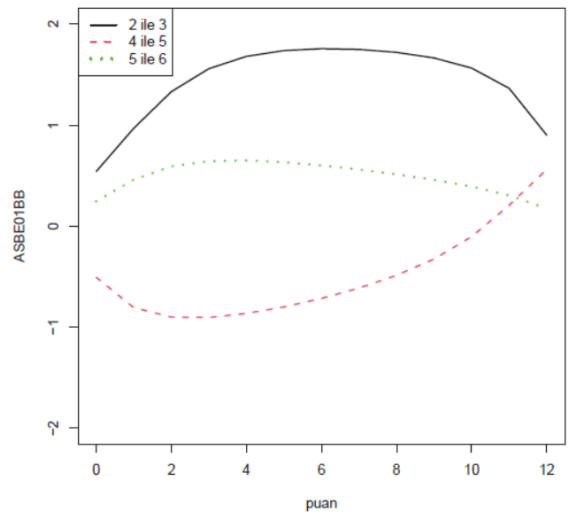
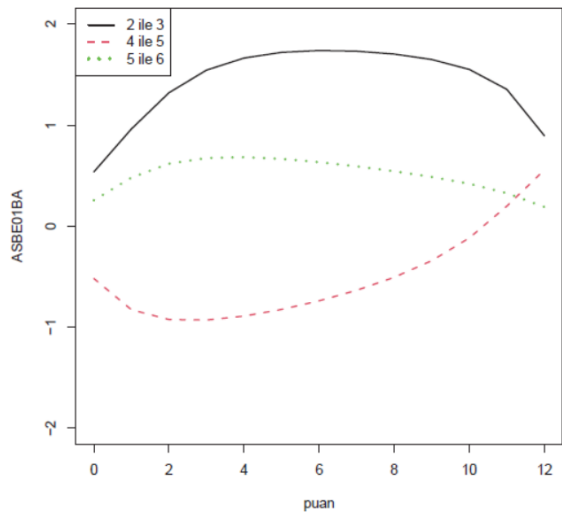
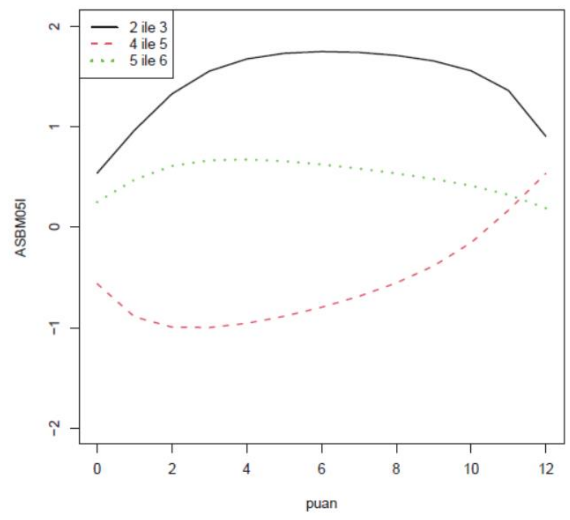
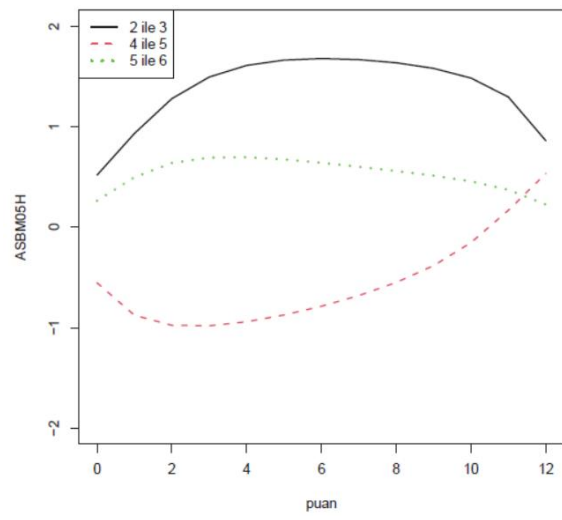
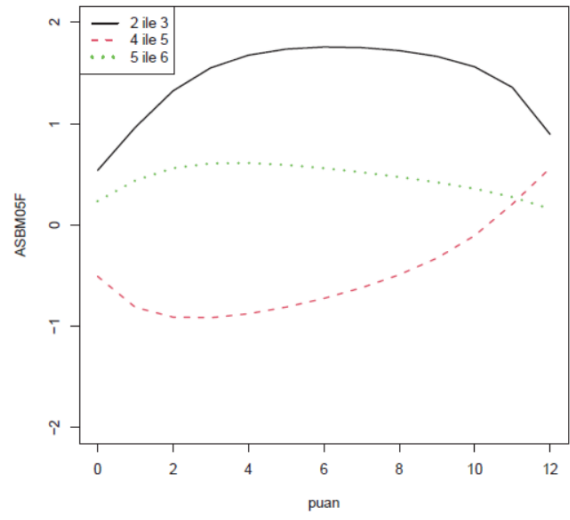
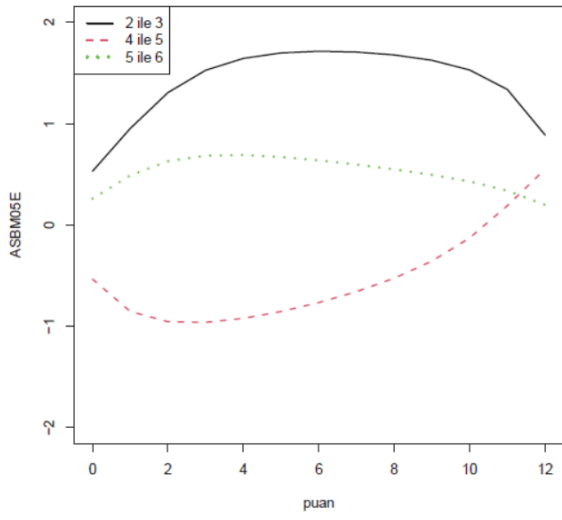


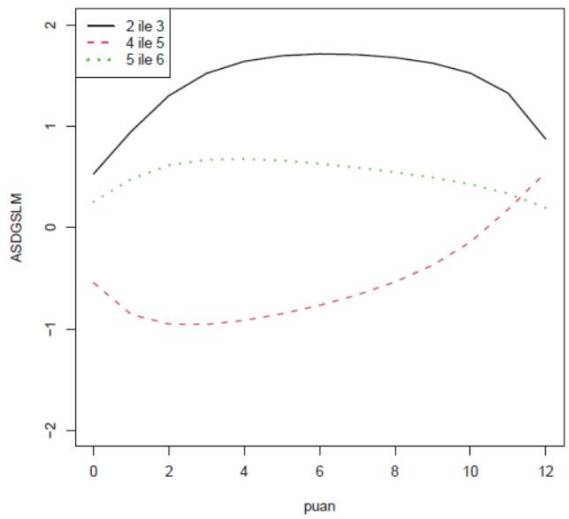
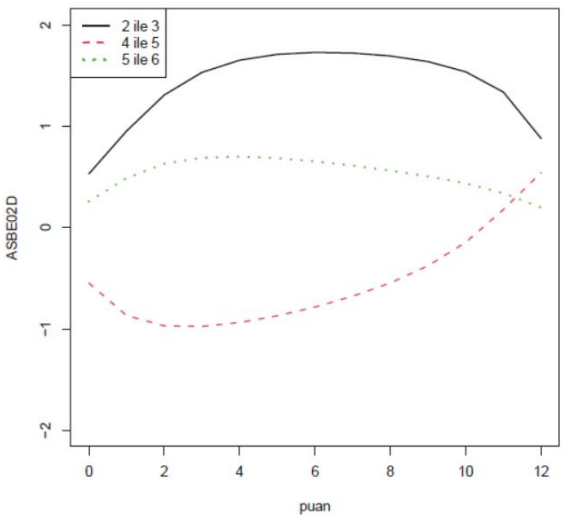
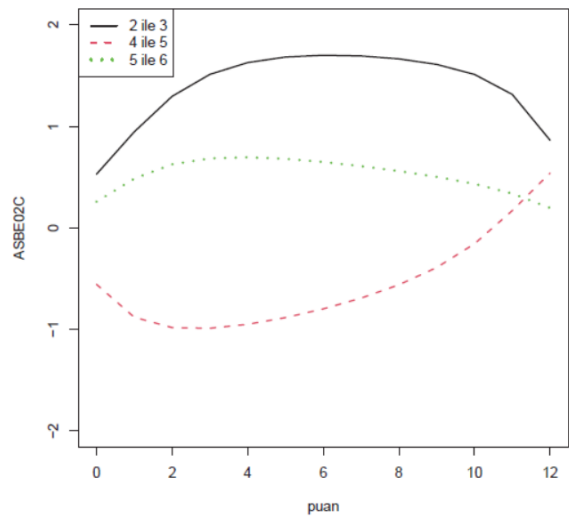
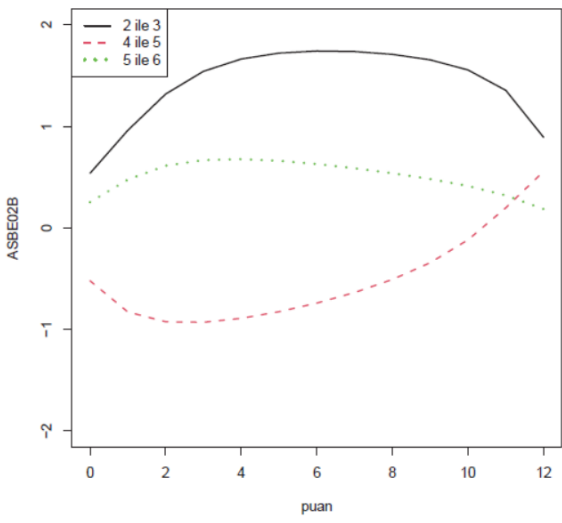
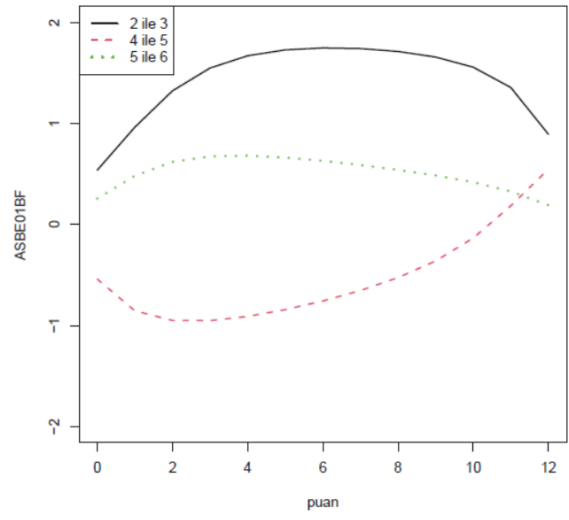
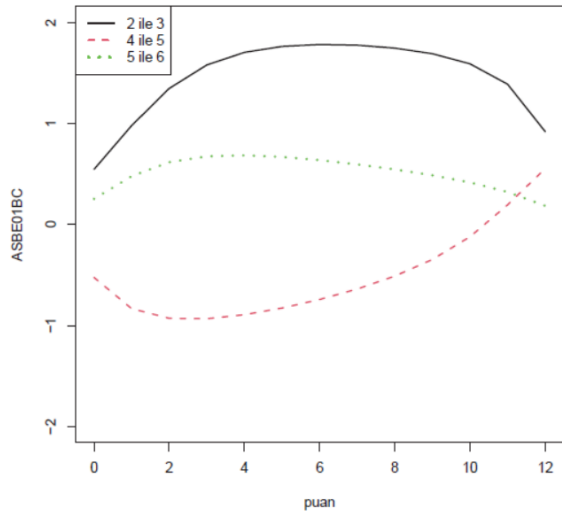


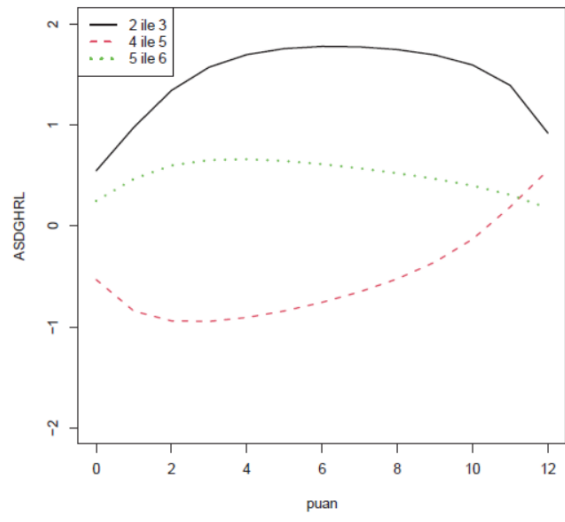
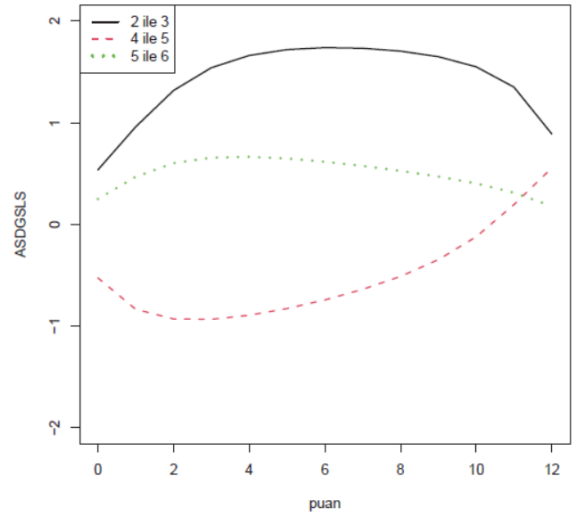
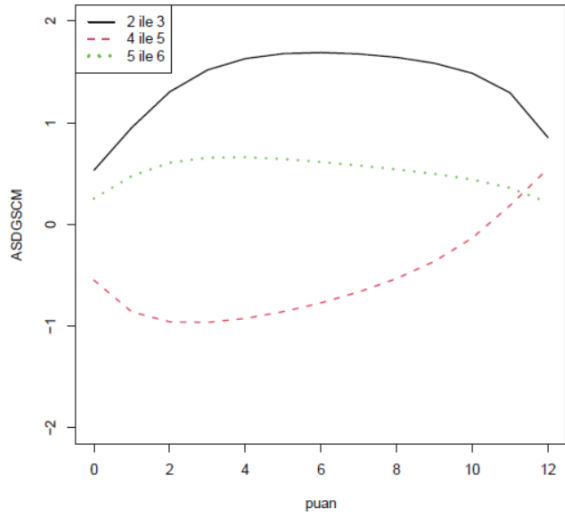













## EK-N: Araştırma Etik Komisyon İzin Muafiyeti Formu

	<b>Hacettepe Üniversitesi</b> <b>Eğitim Bilimleri Enstitüsü</b> <b>Tez Çalışması/Araştırma Etik Komisyon İzin Muafiyeti Formu</b>	<b>F46</b>
26/ 05 / 2022		
Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Eğitim Bilimleri Ana Bilim Dalı Başkanlığına		
<b>Tez/Araştırma Başlığı</b>	<b>ORTAK DEĞİŞKENLERİN ÖZELLİKLERİNE GÖRE KERNEL EŞİTLEME SONUÇLARININ KARŞILAŞTIRILMASI</b>	
Yukarıda başlığı/konusu verilen tez/araştırma çalışmam,		
1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır. 2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir. 3. Beden bütünlüğüne veya ruh sağlığına müdahale içermemektedir. 4. Anket, ölçek (test), mülakat, odak grup çalışması, gözlem, deney, görüşme gibi teknikler kullanılarak katılımcılardan veri toplanmasını gerektiren nitel ya da nicel yaklaşımlarla yürütülen araştırmalar niteliğinde değildir. 5. Diğer kişi ve kurumlardan temin edilen veri kullanımını (kitap, belge vs.) gerektirmektedir. Ancak bu kullanım, diğer kişi ve kurumların izin verdiği ölçüde Kişisel Bilgilerin Korunması Kanuna riayet edilerek gerçekleştirilecektir.		
Çalışmada kullanacağım veriler: ( X ) Kamusal erişime açık (buraya yazınız): TIMSS- Trends in International Mathematics and Science Study: Uluslararası Matematik ve Fen Eğilimleri Araştırması-2019 ( ) Özel izin ve onaya tabi (buraya yazınız): ..... ( ) Üretilmiş veri (buraya yazınız): ..... ( ) Diğer (buraya yazınız): .....		
Yükseköğretim Kurumları Etik Kurulları ve Komisyonlarının Yönergelerini inceledim ve bunlara göre çalışmamın yürütülebilmesi için herhangi bir Etik Komisyondan/Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.		
Gereğini saygılarımla arz ederim.		
<b>Merve YILDIRIM SEHERYELİ</b> <small>(Araştırmacı Adı Soyadı, İmzası)</small>		
<b>Araştırmacı Bilgileri</b>		
<b>Adı Soyadı</b>	<b>Merve YILDIRIM SEHERYELİ</b>	
<b>Öğrenci ise No</b>	N18149765	
<b>Ana Bilim Dalı</b>	Eğitim Bilimleri	
<b>Programı</b>	Eğitimde Ölçme ve Değerlendirme	
<b>Statüsü</b>	<input type="checkbox"/> Yüksek Lisans <input checked="" type="checkbox"/> Doktora <input type="checkbox"/> Bütünleşik Dr. <input type="checkbox"/> Diğer	
<b>Danışman Görüşü ve Onayı*</b>		
<b>Prof. Dr. Selahattin GELBAL</b> <small>(İmza)</small> <small>(Danışmanın Ünvanı, Adı ve Soyadı)</small>		
*Tez ve tezden üretilen yayınlarda gerekli		
<small>Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü. Beytepe Yerleşkesi, 06800, Çankaya / ANKARA  Telefon: 0(312) 297 85 72 Belgegeçer: 0(312) 297 85 66 e-Ağ: <a href="http://ebe.hacettepe.edu.tr/">http://ebe.hacettepe.edu.tr/</a> e-Posta: <a href="mailto:ebe@hacettepe.edu.tr">ebe@hacettepe.edu.tr</a></small>		

**EK-O: Etik Beyanı**

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

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

beyan ederim.

...../...../.....

(İmza)

Merve YILDIRIM SEHERYELİ

**EK-Ö: Yüksek Lisans/Doktora Tez Çalışması Orijinallik Raporu**

...../...../.....

HACETTEPE ÜNİVERSİTESİ

Eğitim Bilimleri Enstitüsü

Eğitim Bilimleri Ana Bilim Dalı Başkanlığına,

Tez Başlığı: ORTAK DEĞİŞKENLERİN ÖZELLİKLERİNE GÖRE KERNEL EŞİTLEME SONUÇLARININ KARŞILAŞTIRILMASI

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

Rapor Tarihi	Sayfa Sayısı	Karakter Sayısı	Savunma Tarihi	Benzerlik Oranı	Gönderim Numarası
18/07/2022	106	170802	29/07/2022	%3	18720166053

Uygulanan filtreler:

1. Kaynaklar hariç
2. Alıntılar dâhil
3. 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

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

**Ad Soyadı:** Merve YILDIRIM SEHERYELİ

**Öğrenci No.:** N18149765

**Ana Bilim Dalı:** Eğitim Bilimleri

**Programı:** Eğitimde Ölçme ve Değerlendirme

**Statüsü:**  Y.Lisans  Doktora  Bütünleşik Dr.

İmza

**DANIŞMAN ONAYI**

UYGUNDUR.

Prof. Dr. Selahattin GELBAL

## EK-P: Thesis/Dissertation Originality Report

...../...../.....

HACETTEPE UNIVERSITY  
Graduate School of Educational Sciences  
To The Department of Educational Sciences

Thesis Title: COMPARISON OF KERNEL EQUATING RESULTS ACCORDING TO PROPERTIES OF COVARIATES

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

Time Submitted	Page Count	Character Count	Date of Thesis Defense	Similarity Index	Submission ID
18/07/2022	106	170802	29/07/2022	%3	18720166053

Filtering options applied:

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

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

I respectfully submit this for approval.

**Name Lastname:** Merve YILDIRIM SEHERYELİ  
**Student No.:** N18149765  
**Department:** Educational Sciences  
**Program:** Educational Mesurement and Evaluation  
**Status:**  Masters  Ph.D.  Integrated Ph.D.

Signature

**ADVISOR APPROVAL**

APPROVED

Prof. Dr. Selahattin GELBAL

## EK-R: Yayınlama ve Fikrî Mülkiyet Hakları Beyanı

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

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

Yükseköğretim Kurulu tarafından yayınlanan "**Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge**" kapsamında tezim aşağıda belirtilen koşullar haricince YÖK Ulusal Tez Merkezi / H.Ü. Kütüphaneleri Açık Erişim Sisteminde erişime açılır.

- Enstitü/ Fakülte yönetim kurulu kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 2 yıl ertelenmiştir. (1)
- Enstitü/Fakülte yönetim kurulunun gerekçeli kararı ile tezimin erişime açılması mezuniyet tarihinden itibaren 6 ay ertelenmiştir. (2)
- Tezimle ilgili gizlilik kararı verilmiştir. (3)

..... /..... /.....

(imza)

Merve YILDIRIM SEHERYELİ

---

"Lisansüstü Tezlerin Elektronik Ortamda Toplanması, Düzenlenmesi ve Erişime Açılmasına İlişkin Yönerge"

- (1) Madde 6. 1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu iki yıl süre ile tezinerişime açılmasının ertelenmesine karar verebilir.
- (2) Madde 6.2. Yeni teknik, materyal ve metodların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internette paylaşılması durumunda 3 şahıslara veya kurumlara haksız kazanç; imkânı oluşturabilecek bilgi ve bulguları içeren tezler hakkında tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulunun gerekçeli kararı ile altı ayı aşmamak üzere tezin erişime açılması engellenebilir.
- (3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir\*. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.

Madde 7.2. Gizlilik kararı verilen tezler gizlilik süresince enstitü veya fakülte tarafından gizlilik kuralları çerçevesinde muhafaza edilir, gizlilik kararının kaldırılması halinde Tez Otomasyon Sistemine yüklenir

\*Tez danışmanının önerisi ve enstitü anabilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir.



