



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

KAYIP VERİYLE BAŞA ÇIKMA YÖNTEMLERİNİN TEST EŞİTLEMeye ETKİSİNİN
İNCELENMESİ

Gülden ÖZDEMİR

Doktora Tezi

Ankara, 2023



Liderlik, araştırma, inovasyon, kaliteli eğitim ve deęişim ile

Daha ileriye ... En İyiyeye ...



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

KAYIP VERİYLE BAŞA ÇIKMA YÖNTEMLERİNİN TEST EŞİTLEMeye ETKİSİNİN
İNCELENMESİ

INVESTIGATION OF THE IMPACT OF TECHNIQUES OF HANDLING MISSING DATA ON
THE TEST EQUATING

Gülden ÖZDEMİR

Doktora Tezi

Ankara, 2023

Kabul ve Onay

Eğitim Bilimleri Enstitüsü Müdürlüğüne,

Gülden ÖZDEMİR'in hazırladığı "Kayıp Veriyle Başa Çıkma Yöntemlerinin Test Eşitlemeye Etkisinin İncelenmesi" başlıklı bu çalışma jürimiz tarafından **Eğitim Bilimleri Ana Bilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalında Doktora Tezi** olarak kabul edilmiştir.

Jüri Başkanı	Prof. Dr. Selahattin GELBAL	İmza
Jüri Üyesi (Danışman)	Prof. Dr. Burcu ATAR	İmza
Jüri Üyesi	Doç. Dr. Kübra ATALAY KABASAKAL	İmza
Jüri Üyesi	Doç. Dr. Celal Deha DOĞAN	İmza
Jüri Üyesi	Doç. Dr. Serkan ARIKAN	İmza

Enstitü Yönetim Kurulunun
...../...../..... Tarihli ve
sayılı kararı.

Bu tez Hacettepe Üniversitesi Lisansüstü Eğitim, Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca yukarıdaki jüri üyeleri tarafından / / tarihinde uygun görülmüş ve Enstitü Yönetim Kurulunca / / tarihi itibarıyla kabul edilmiştir.

Prof. Dr. İsmail Hakkı MİRİCİ
Eğitim Bilimleri Enstitüsü Müdürü

Öz

Bu araştırmada; kayıp veriyle başa çıkma yöntemlerinin, farklı koşullar altında, test eşitleme yöntemlerinin performansı üzerindeki etkilerini incelemek amaçlanmıştır. Araştırmada; örneklem büyüklüğü (750, 1500), kayıp veri oranı (%10, %20, %30) ve kayıp verilerin bulunduğu test formu (her iki test, eşitlenecek test) koşul olarak alınmıştır. Test eşitleme yöntemlerinin performansları; hata kareleri ortalamasının karekökü (RMSE) ve eşitleme yanlılığına (BIAS) göre değerlendirilmiştir. Araştırmada kullanılan tam veri setleri, TIMSS 2019'a elektronik uygulama (eTIMSS) ile katılan tüm ülkelerdeki, 6 ve 7 numaralı kitapçıkları alan ve iki kategorili (1-0) puanlanan maddeleri yanıtlayan öğrenci verilerinden elde edilmiştir. Tamamen rastgele kayıp veri mekanizması altında veri silinerek oluşturulan kayıp verili test formlarına sıfır atama (SA), hot deck atama (HDA) ve çoklu atama (PMM ve LOGREG) yöntemleri ile veri ataması yapılmıştır. Tamamlanmış test formları MTK'ye dayalı test eşitleme yöntemleri ile eşitlenmiştir. Araştırmadaki tüm analizler R programlama dili aracılığıyla gerçekleştirilmiştir. Referans değere en yakın ortalama RMSE ve BIAS değerleri üreten yöntemlerin PMM ve LOGREG olduğu, HDA yönteminin ise bu yöntemlere benzer performans gösterdiği belirlenmiştir. En hatalı ve en yanlı sonuçlar üreterek referans değere en uzak kestirimler yapan yöntemin ise SA olduğu görülmüştür. Araştırmada ele alınan örneklem büyüklüğü koşulunun tüm kayıp veriyle başa çıkma yöntemleri üzerinde, kayıp veri oranı ve kayıp verilerin bulunduğu test formu koşulunun ise sadece SA yöntemi ile elde edilen ortalama RMSE ve BIAS değerleri üzerinde anlamlı bir etkisinin olduğu tespit edilmiştir. PMM, LOGREG ve HDA yöntemlerinin araştırmada ele alınan tüm koşullarda, SA yönteminin ise en iyi performans gösterdiği büyük örneklemelerde ve düşük kayıp veri oranlarında tercih edilmesi önerilmiştir.

Anahtar sözcükler: kayıp veri, kayıp veriyle başa çıkma yöntemleri, test eşitleme, eşitleme hatası, bootstrap

Abstract

In this study; It is aimed to examine the effects of missing data handling methods on the performance of test equating methods under different conditions. In the research; sample size (750, 1500), missing data rate (10%, 20%, 30%) and test form with missing data (both tests, test to be equalized) were taken as conditions. Performances of test equating methods; root mean square error (RMSE) and equating bias (BIAS). The full datasets used in the research were obtained from the data of students in all countries participating in TIMSS 2019 with the electronic application (eTIMSS), who took booklets 6 and 7 and answered the items scored in two categories (1-0). Zero imputation (ZI), hot deck imputation (HDI) and multiple imputation (PMM and LOGREG) methods were used to assign data to test forms with missing data, which were created by deleting data under the missing completely at random data mechanism. Completed test forms were equated with test equating methods based on IRT. All analyzes in the research were carried out using the R programming language. It was determined that the methods that produced the mean RMSE and BIAS values closest to the reference value were PMM and LOGREG, while the HDI method showed similar performance to these methods. It has been observed that the method that produces the most erroneous and biased results and makes the furthest estimations from the reference value is ZA. It was determined that the sample size condition in the research had a significant effect on the methods of handling with all missing data, while the missing data rate and the test form condition with missing data had a significant effect on the average RMSE and BIAS values obtained only by the ZI method. It has been suggested that PMM, LOGREG and HDA methods should be preferred in all conditions discussed in the study, while SA method should be preferred in large samples and low loss data rates where it performs best.

Keywords: missing value, missing data handling methods, test equating, equating error, bootstrap

Teşekkür

Doktora eğitimim ve tez sürecimde; bilgi birikimi ve tecrübesi ile yoluma ışık olan, yaratıcı ve çözüm odaklı yaklaşımıyla çıkmaz sokaklarıma yeni yollar açan, sabrı, sakinliği ve güler yüzüyle hep güvende hissettirerek yolculuğuma eşlik eden değerli tez danışmanım Prof. Dr. Burcu ATAR'a teşekkürlerin en büyüğünü borç bilirim.

Gerek akademik donanımı gerekse yönetim tecrübesi ve duruşu ile örnek olan kıymetli hocam Prof. Dr. Selahattin GELBAL'a çok teşekkür ederim. Tezimin gelişimine farklı görüş ve önerileri ile katkı sağlayan değerli hocalarım Doç. Dr. Kübra ATALAY KABASAKAL, Doç. Dr. Celal Deha DOĞAN ve Doç. Dr. Serkan ARIKAN'a çok teşekkür ederim. Özellikle analiz sürecinde ne zaman ihtiyacım olsa desteğini büyük bir incelikte gösteren Dr. Sinan YAVUZ'a çok teşekkür ederim. Tez yazım sürecimde manevi desteğini hep yanımda hissettiğim sevgili hocam Dr. Berna GÜCÜM'e ve arkadaşlarım Elif KARAKUŞ ile Hale ÜÇKARDEŞ'e çok teşekkür ederim. Eğitimlerim boyunca akademik hayatımın gelişimine yön veren tüm hocalarıma emekleri için çok teşekkür ederim.

Derdimle dertlenip mutluluğumla uçan kıymetli dostlarım Muhlise ÇEVİKKALP ve Adil Sakin KARGIN'a çok teşekkür ederim.

Canım anneme ve babama, varlıkları ile hayatıma kattıkları anlam ve yaşam yolculuğumdaki emekleri için sonsuz teşekkür ederim.

İçindekiler

Kabul ve Onay.....	ii
Öz.....	iii
Abstract.....	iv
Teşekkür.....	v
Tablolar Dizini.....	viii
Şekiller Dizini.....	ix
Simgeler ve Kısaltmalar Dizini.....	x
Bölüm 1 Giriş.....	1
Problem Durumu.....	1
Araştırmanın Amacı ve Önemi.....	5
Araştırma Problemi.....	7
Sayıtlılar.....	8
Sınırlılıklar.....	8
Tanımlar.....	9
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	10
Madde Tepki Kuramı.....	10
Test Eşitleme.....	15
Kayıp Veri.....	25
İlgili Araştırmalar.....	35
Bölüm 3 Yöntem.....	41
Araştırmanın Türü.....	41
Araştırmanın Evreni ve Örneklemi.....	41
Veri Toplama Aracı.....	43
Eşitleme Deseni.....	45
Araştırmada Ele Alınan Koşullar.....	46
Veri Setlerinin Oluşturulması.....	48

Verilerin Analizi	49
Bölüm 4 Bulgular, Yorumlar ve Tartışma.....	66
Birinci Alt Probleme İlişkin Bulgular, Yorumlar ve Tartışma	66
İkinci Alt Probleme İlişkin Bulgular, Yorumlar ve Tartışma	75
Üçüncü Alt Probleme İlişkin Bulgular, Yorumlar ve Tartışma	79
Bölüm 5 Sonuç ve Öneriler.....	86
Sonuçlar	86
Öneriler	89
Kaynaklar	93
EK-A: Yerel Bağımsızlık Testi Sonuçları	cxiv
EK-B: Kitapçık 6'da Yer Alan Maddelere Ait Madde Karakteristik Eğrileri	cxxi
EK-C: Kitapçık 7'de Yer Alan Maddelere Ait Madde Karakteristik Eğrileri	cxxiii
EK-Ç: Analiz Sonucu Elde Edilen RMSE ve BIAS Değerleri	cxxvi
EK-D: Araştırma Etik Komisyon İzin Muafiyeti Formu	cxxix
EK-E: Etik Beyanı	cxxxi
EK-F: Doktora Tez Çalışması Orijinallik Raporu.....	cxxxii
EK-G: Dissertation Originality Report	cxxxiii
EK-H: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı.....	cxxxiii

Tablolar Dizini

Tablo 1 <i>Denk Olmayan Gruplarda Ortak Madde Deseni</i>	17
Tablo 2 <i>Ülkelere ve Test Formlarını Alan Öğrenci Sayılarına İlişkin Bilgiler</i>	42
Tablo 3 <i>Kitapçıklardaki Maddelere Ait Özellikler</i>	44
Tablo 4 <i>Araştırmada Kullanılan Eşitleme Deseni</i>	45
Tablo 5 <i>Araştırmada Ele Alınan Koşullar</i>	46
Tablo 6 <i>Betimsel İstatistikler</i>	50
Tablo 7 <i>Kitapçıklara İlişkin Güvenirlik Katsayıları</i>	52
Tablo 8 <i>KMO ve Bartlett Test Sonuçları</i>	53
Tablo 9 <i>Faktör Analizi Sonuçları</i>	53
Tablo 10 <i>Paralel Analiz Sonuçları</i>	56
Tablo 11 <i>Model Veri Uyumu</i>	57
Tablo 12 <i>Madde Parametre Kestirimleri</i>	59
Tablo 13 <i>Madde Parametrelerinin Ortalamaları</i>	60
Tablo 14 <i>Araştırma Kapsamında Oluşturulan Veri Setleri</i>	63
Tablo 15 <i>Ortalama RMSE ve BIAS Değerleri İçin ANOVA Sonuçları</i>	76

Şekiller Dizini

Şekil 1 <i>Tipik Bir Madde Karakteristik Eğrisi</i>	11
Şekil 2 <i>MTK'ye Dayalı Ölçek Dönüştürme Yöntemleri</i>	19
Şekil 3 <i>Kayıp Veri Mekanizmaları</i>	26
Şekil 4 <i>Çoklu Atama Aşamalarının Grafikselleştirilmesi</i>	32
Şekil 5 <i>Kitapçık 6 İçin Puan Dağılımları</i>	51
Şekil 6 <i>Kitapçık 7 İçin Puan Dağılımları</i>	51
Şekil 7 <i>Yamaç Birikinti Grafiği (Kitapçık 6)</i>	54
Şekil 8 <i>Yamaç Birikinti Grafiği (Kitapçık 7)</i>	55
Şekil 9 <i>750 Örneklem Büyüklüğü İçin RMSE Grafiği</i>	67
Şekil 10 <i>1500 Örneklem Büyüklüğü İçin RMSE Grafiği</i>	67
Şekil 11 <i>750 Örneklem Büyüklüğü İçin BIAS Grafiği</i>	71
Şekil 12 <i>1500 Örneklem Büyüklüğü İçin BIAS Grafiği</i>	72
Şekil 13 <i>Eşitlemeler ile Elde Edilen Ortalama RMSE Değerleri</i>	80
Şekil 14 <i>Eşitlemeler ile Elde Edilen Ortalama BIAS Değerleri</i>	81
Şekil 15 <i>Yöntemlere Göre Elde Edilen Ortalama RMSE ve BIAS Değerleri</i>	83

Simgeler ve Kısaltmalar Dizini

ALES: Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı

BIAS: Yanlılık

ET: Eşitlenecek Test Formu (Kitapçık 6)

HIT: Her İki Test Formu (Kitapçık 6 ve Kitapçık 7)

HDA: Hot Deck Atama (Sıcak Deste Atama)

İOKBS: İlköğretim ve Ortaöğretim Kurumları Bursluluk Sınavı

KPSS: Kamu Personel Seçme Sınavı

LOGREG: Logistic Regression (Lojistik Regresyon ile Çoklu Atama Yöntemi)

MTK: Madde Tepki Kuramı

PMM: Predictive Mean Matching (Tahmine Dayalı Ortalama Eşleştirme ile Çoklu Atama Yöntemi)

RMSE: Root Mean Square Error (Hata Kareleri Ortalamasının Karekökü)

SA: Sıfır Atama

TIMSS: Trend in International Mathematics and Science Study (Uluslararası Matematik ve Fen Eğitimleri Araştırması)

TRK: Tamamen Rastgele Kayıp

YDS: Yabancı Dil Bilgisi Seviye Tespit Sınavı

YKS: Yükseköğretim Kurumları Sınavı

YÖKDİL: Yükseköğretim Kurumları Yabancı Dil Sınavı

Bölüm 1

Giriş

Bu bölümde; problem durumu, araştırmanın amacı ve önemi, problem cümlesi, sayıtlılar, sınırlılıklar ve tanımlar yer almaktadır.

Problem Durumu

Ülkemizde öğrencilerin belirli özelliklere ilişkin düzeylerinin ölçülmesi ve ölçme sonuçlarına göre geribildirim verilmesi genellikle sınavlar aracılığı ile yapılmaktadır. Özellikle bir üst öğrenime öğrenci seçme, bir işe personel yerleştirme gibi bireyler hakkında kritik kararların alınmasında ulusal düzeyde gerçekleştirilen merkezi sınavlar (Sınavla Öğrenci Alacak Ortaöğretim Kurumlarına İlişkin Merkezî Sınav, Yükseköğretim Kurumları Sınavı (YKS), İlköğretim ve Ortaöğretim Kurumları Bursluluk Sınavı (İOKBS), Kamu Personel Seçme Sınavı (KPSS) vb.) önemli rol oynamaktadır. Bu sınavlardan bazıları yılda bir kez, bazıları ise yılda birkaç kez uygulanmaktadır.

Büyük ölçekli sınavlarda aynı sınavı birkaç kez alan adayların daha önce görmüş olabilecekleri test maddeleri ile karşılaşmaları, maddeler ile ilgili daha önceden bilgi sahibi oldukları için avantaj sağlayacaktır. Sınava giren adayların hiçbirine avantaj sağlamamak dolayısıyla adaletli bir sınav yapmak için aynı testin birçok formu geliştirilmektedir. Başka bir ifade ile aynı amaç için belirli zaman dilimlerinde uygulanan sınavlarda, soruların (maddelerin) güvenliğini sağlamak amacıyla test formları geliştirilmektedir. Ancak aynı yapıyı ölçmek için geliştirilen ve farklı maddelerden oluşan test formlarından elde edilen puanların güvenirlik ve maddelerin güçlük gibi özellikleri farklı olabilir. Bu durum sınava giren ve farklı test formlarını alan adaylar arasında eşitsizliğe yol açabilir. Adayların bir kısmının, diğer adaylardan daha zor bir test formu alarak dezavantajlı hale geçmelerini önlemek ve bu formlardan elde edilen puanların karşılaştırılabilir olmasını sağlamak için test formlarının eşitlenmesi gerekmektedir (Cook & Eignor, 1991). Eşitleme sonucu test

formları birbirinin yerine kullanılabilir ancak bu alternatif formların her biri için geçerlik kanıtları ayrı ayrı sunulmalıdır.

Öğrenci özelliklerinin ölçülmesi sürecinde kullanılan ölçme araçlarının niteliği, bu sınav sonuçlarına dayalı olarak öğrenciler hakkında verilecek kararların doğru ve amaca uygun olması ile yakından ilişkilidir. Ölçme araçlarında bulunması istenen niteliklerden geçerlik ve güvenirliğin yanı sıra adalet kavramı da önem taşımaktadır. Çünkü adalet, temel bir geçerlik meselesidir (AERA, APA & NCME, 2014). Ülkemizde bazı üniversitelerin lisansüstü eğitim programlarına kabul alma şartları arasında yer alan, Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı (ALES) ile Yabancı Dil Bilgisi Seviye Tespit Sınavı (YDS) ya da Yükseköğretim Kurumları Yabancı Dil Sınavı (YÖKDİL) yılda birkaç kez uygulanan sınavlardır. Örneğin, aynı lisansüstü eğitim programına başvuran adaylar farklı dönemlerde girdikleri ALES puanlarını kullanabilmekte ya da YDS puanı yerine YÖKDİL puanı ile başvuru yapabilmektedirler. Ancak bu sınavlar eşitlenmemesine rağmen eşit kabul edilmektedir. Bu durum adalet kavramı dolayısıyla test puanlarının geçerliği üzerinde soru işaretleri oluşturmaktadır.

Test geçerliğini tehlikeye düşüren ve sınavı alan adaylarla ilgili doğru kararlar verilmesi hususunda engel teşkil eden bir diğer faktör ise kayıp verilerdir (Hohensinn & Kubinger, 2011). Sınavlarda yer alan her maddeye ya da duruma tam ve eksiksiz cevap almak çoğu zaman mümkün olmaz. Testi alan bireylerin bir maddeyi yanıtlamama durumları üç madde altında özetlenebilir (De Ayala ve diğerleri, 2001);

1. Eşdeğer formlardan oluşan testlerde bireyler ortak maddelerin dışındaki maddelerin tamamını görememektedirler. Maddenin yanıtlanmama durumu bireyin kararı değildir. Testin yapısı ve tasarımına bağlı olarak, test geliştirme aşamasında planlandığı için bu maddeler “uygulanmamış madde” olarak kodlanmaktadır ve parametre kestirimlerinde herhangi bir soruna yol açmaz.
2. Hız testlerinde bazı bireyler öngörülen sürede tüm maddelere erişememektedirler. Maddenin yanıtlanmama durumu bireyin kararı değil süre

yetersizliđidir. Test geliřtirme ařamasında planlanmış ya da öngörölmüş olmayan bu maddeler “eriřilememiř madde” olarak adlandırılmaktadır.

3. Yeterli süreye sahip test uygulamalarında bazı bireyler bazı maddeleri bilinçli olarak yanıtlamamaktadırlar. Maddenin yanıtlanmama durumu bireyin kararıdır. Yanıtı bilmeme, bilgisine güvenmeme gibi birçok nedenden dolayı kasıtlı olarak boş bırakılan bu maddeler “atlanmış madde” olarak adlandırılmaktadır.

Yanıtlamama ya da boş bırakma durumları, amaçlanan gözlemlerde kayıp verilerin oluşması sorununa yol açmaktadır. Kayıp veriler ile veri seti arasındaki ilişki incelenerek kayıp veri mekanizmaları belirlenebilir. Little ve Rubin (1987) tarafından tanımlanan üç farklı kayıp veri mekanizması vardır. Bunlar; tamamen rastgele kayıp (TRK), rastgele kayıp (RK) ve rastgele olmayan kayıp (ROK) veri mekanizmalarıdır. Kayıp verinin olma olasılığı; kayıp verinin bulunduğu deđişken veya arařtırmada ölçölen ya da ölçölmeyen diđer deđişkenler ile ilişkili deđilse TRK’dır. Bir test maddesinin gözden kaçırıldığı için yanıtlanmaması bu duruma örnek olarak verilebilir. Kayıp verinin olma olasılığı; kayıp verinin bulunduğu deđişken dışında arařtırmada ölçölen ya da ölçölmeyen diđer deđişkenler ile ilişkili ise RK’dır. Erkek katılımcıların kadın katılımcılar lehine yanlılık içeren bir maddeyi anlayamadıkları için yanıtlamamaları ya da süre yetmediđi için okunamayan maddelerin yanıtlanmaması bu duruma örnek olarak verilebilir. Kayıp verinin olma olasılığı; kayıp verinin bulunduğu deđişken ile bizzat ilişkiliyse ROK’tur. Bir test maddesinin, cevabı bilinmediđi için ya da açık ve anlaşılır olmadığı için yanıtlanmaması bu duruma örnek olarak verilebilir.

Kayıp veriler veri setinin daralmasına ve dolayısıyla yapılacak kestirimlerin gücünün azalmasına neden olmaktadır. Ayrıca maddeyi yanıtlayanlar ile yanıtlamayanlar arasında yanlılık çıkma olasılığı vardır. Standart analiz yöntemleri, tam veri setlerine göre hazırlandığı için kayıp veriye sahip veri setleri üzerinde kullanılamamaktadır (Rubin, 1987). Veri eksikliđinden dolayı istatistiksel yöntemlerin kullanımını zorlařtıran dolayısıyla istatistiksel analizin gücünün azalmasına ve zaman, çaba, maliyet sonucu elde edilen verinin verimsiz

kullanılmasına yol açan kayıp veriler (Peng ve diğerleri, 2007) ile farklı başa çıkma yöntemleri geliştirilmiştir. Bu yöntemlerden bazılarının diğer yöntemlere göre daha iyi olduğu söylenebilir ancak hiçbir yöntem “en iyi” olarak tanımlanamaz (Allison, 2002). Bu nedenle araştırmacılar kayıp veri mekanizması ve oranı, örneklem büyüklüğü gibi etkenleri göz önünde bulundurarak kayıp veri ile başa çıkılması için uygun bir yöntem belirlemelidirler (Cheema, 2012). Büyük bir veri setinde verilerin %5’ten fazlası TRK ya da RK ise, küçük ve orta büyüklükteki bir veri setinde çok miktarda ROK veri varsa bu kayıp veriler ihmal edilebilir değildir (Finch, 2008). İhmal edilemeyen bu kayıp veriler için veri setine en uygun kayıp veriyle başa çıkma yöntemine/yöntemlerine karar verilmelidir. Alan yazında yapılan çalışmalar, kayıp veriyle başa çıkma yöntemlerinin farklı koşullar altındaki performanslarının farklılaştığını göstermiştir. Benzer şekilde test eşitleme yöntemlerinin performanslarının karşılaştırıldığı çalışmalarda da ele alınan koşullara göre sonuçların değişiklik gösterdiği raporlanmıştır.

Kayıp veriler, kayıp verilerle başa çıkma yöntemlerinin kullanıldığı test formlarında eşitleme sonuçlarını da etkilemektedir (Ngudgratoke, 2009; Kim, 2015; Shin, 2009; Shin, 2016; Ertoprak, 2017). Bu nedenle eşitleme çalışması öncesinde farklı test formlarındaki muhtemel kayıp verilere ilişkin detaylar incelenerek kayıp veri sorunu uygun başa çıkma yöntemleri ile çözülmeli ve bu yöntemlerin test eşitlemeye yönelik kestirimlerdeki performansları incelenmelidir. Ancak eşitleme çalışmalarının çoğunda kayıp veriye sahip maddeler cevaplayıcılara ya hiç uygulanmamış ya da cevaplayıcılar tarafından yanlış cevaplanmış gibi davranılmaktadır. Bu durumun göz ardı edilmesi, farklı test formlarının yer aldığı bir sınava katılan bireylerin puanları hakkında onların hayatlarını etkileyecek hatalı kararlar verilmesine yol açabilir. Bu nedenle, farklı kayıp veri ile başa çıkma yöntemlerinin test eşitleme üzerindeki etkisini, gerçek bir veri seti üzerinde ve farklı koşullar altında derinlemesine belirlemek için bu çalışmanın yapılmasına ihtiyaç duyulmuştur.

Araştırmanın Amacı ve Önemi

Aynı amaç ve benzer yapıları ölçmek için hazırlanan gerek sınıf içi gerekse ulusal ve standartlaştırılmış uluslararası uygulamalara sıklıkla rastlanılmaktadır. Bu uygulamalar ile bireyler ve ülkeler kendi başarılarını değerlendirebilmekte, özellikle TIMSS, PISA, PIRLS gibi sınavlarda uluslararası karşılaştırmalar yapılabilmekte ve bu uygulamaların sonuçları ülkelerin eğitim politikalarının revize edilmesinde önemli rol oynamaktadır. Bu nedenle söz konusu karşılaştırmaların adilliği, doğruluğu ve kesinliği hem bireyler hem de politika geliştiricileri açısından önem arz etmektedir.

Söz konusu uygulamaların içeriğine bakıldığında; aynı düzeydeki bireylere, aynı zamanda, farklı test formlarının uygulandığı (TIMSS, PISA, PIRLS, ABİDE) ya da farklı düzeydeki bireylere, farklı zamanlarda, farklı test formlarının uygulandığı (ALES, YDS, YÖKDİL) görülmektedir. Farklı test formlarından elde edilen puanların anlamlı bir şekilde karşılaştırılması ya da aynı amaç ve benzer yapıları ölçmek için geliştirilen test formlarının birbiri yerine kullanılabilmesi için test eşitleme yapılmalıdır. Ancak sınavlarda yer alan maddelere çoğu zaman yanıt verilmemesinden dolayı oluşan kayıp veriler nedeniyle tam veri setleriyle çalışan standart analiz yöntemleri kullanılamamaktadır. Bu durumda madde ve yetenek parametrelerinin kestirimi, dolayısıyla test eşitleme süreci etkilenmektedir. Bu nedenle test eşitlemenin hatasız ya da en az hata ile yapılabilmesi için kayıp veriye sahip veri setinin en uygun kayıp veriyle başa çıkma yöntemi kullanılarak çözümlenmesi gereklidir.

Alanyazında test eşitleme ile ilgili yapılan çalışmalar incelendiğinde; örneklem büyüklüğü, yetenek dağılımı, madde türü, boyut sayısı, şans başarısı, değişen madde fonksiyonu (DMF), grup değişmezliği, eşitleme deseni gibi faktörlerin test eşitleme yöntemlerinin performanslarını etkilediği görülmüştür. Söz konusu çalışmalarda; test eşitleme yöntemlerinin performanslarını etkileyen etmenlerin belirlenmesi, farklı koşullar altında test eşitleme yöntemlerinin performanslarının karşılaştırılması ve en az hata veren yöntemin belirlenmesi amaçlanmıştır (Aksekioğlu, 2017; Aytekin Kazanç, 2019; Branberg

& Wiberg, 2011; Gündüz, 2015; Karagül, 2020; Kilmen, 2010; Kim, 2015; Lee & Ban, 2010; Liu & Kolen, 2011; Mutluer, 2021; Powers, 2010; Tanberkan Suna, 2018), Kayıp veri ile ilgili yapılan çalışmalar incelendiğinde ise kayıp veriyle başa çıkma yöntemlerinin iç tutarlılık, madde ve model veri uyumu, standart hata, yapısal eşitlik modelleri, ölçme değişmezliği ve DMF gibi faktörler üzerinde etkisinin olduğu belirlenmiştir. Örneklem büyüklüğü, kayıp veri oranı, test uzunluğu, puanlama türü, kayıp veri mekanizması gibi koşulların ele alındığı çalışmalarda, kayıp veriyle başa çıkma yöntemlerinin performanslarının karşılaştırılması ve bu yöntemlerin performanslarını etkileyen etmenlerin belirlenmesi amaçlanmıştır (Bayhan, 2018; Çüm & Gelbal, 2015; Hohensinn & Kubinger, 2011, Işıkoğlu, 2017; Jia & Wu, 2019; Karaman, 2022; Koçak, 2016; Liu & Sriutaisuk, 2019; Tamcı, 2018).

Bu araştırmada ise en genel hali ile kayıp veriyle başa çıkma yöntemlerinin, test eşitleme üzerindeki etkisini incelemek amaçlanmıştır. Çünkü alanyazında yer alan çalışmalar incelendiğinde kayıp veriyle başa çıkma yöntemleri (Ada, 2015; Ak, 2020; Akbaş, 2014; Başman, 2014; Bayhan, 2018; Bayram, 2020; Cheema, 2012; Demir, 2013; Işıkoğlu, 2017; Karaman, 2022; Koçak, 2016; Öztemür, 2014; Şahin Kürşad, 2014; Tamcı, 2018; Weaver & Maxwell, 2014; Zhu, 2014) ve test eşitleme yöntemleri (Aksekioglu, 2017; Aşiret, 2014; Atalay Kabasakal, 2014; Aytekin Kazanç, 2019; Bastari, 2000; Bozdağ, 2007; Chen, 2001; Demirus, 2015; Gök, 2012; Gündüz, 2015; He, 2013; İnci, 2014; Karagül, 2020; Kilmen, 2010; Kim & Lee, 2006; Meng, 2012; Mutluer, 2013; Mutluer, 2021; Speron, 2009; Tian, 2011; Uysal, 2014; Uysal, 2019; Yurtçu, 2018) ile ilgili ayrı ayrı ve çok sayıda araştırma olduğu görülmektedir. Ancak bu kavramları birlikte ele alan araştırmalar sınırlı sayıdadır (Ertoprak, 2017; Kim, 2015; Ngudgratoke, 2009; Shin, 2009; Shin, 2016). Söz konusu araştırmalar incelendiğinde; yurt dışında yapılan çalışmalarda (Kim, 2015; Ngudgratoke, 2009; Shin, 2009; Shin, 2016) test uzunluğu, sınava giren gruplar arasındaki yetenek farkı, yaş düzeyleri, madde parametreleri, model, grup farklılığı ve kayıp veri ile başa çıkma yöntemlerine ilişkin koşullara yer verildiği ancak örneklem büyüklüğü, kayıp verilerin bulunduğu test formu ve kayıp veri oranına ilişkin koşulların ele alınmadığı görülmüştür.

Türkiye’de yapılan çalışmanın (Ertoprak, 2017) ise sabit bir örneklem büyüklüğü, Madde Tepki Kuramı’na (MTK) dayalı karakteristik eğri dönüştürme yöntemlerinden Stocking-Lord ve sadece simülatif şekilde üretilen veri seti ile sınırlı olduğu görülmüştür. Ancak simülasyon kusursuz bir araştırma yöntemi değildir. Çalışma sonuçlarının uygulandığı gerçek veriler, varsayılan bir MTK modelinden simüle edilmiş verilere yakın olmayabilir. Harris ve Crouse (1993)’a göre, gerçek veriler simüle edilmiş verilere yakın olduğunda, sonuçlar simülasyon çalışmalarının bulguları ile benzerlik göstermektedir.

Bu çalışmada; gerçek veri seti üzerinde, kayıp veriyle başa çıkma yöntemlerinin (sıfır atama (SA), hot deck atama (HDA), çoklu atama (PMM ve LOGREG)), farklı koşullar altında, MTK’ye dayalı test eşitleme yöntemlerinin (karakteristik eğri dönüştürme yöntemleri) performansları üzerindeki etkileri incelenmiştir. Test eşitleme yöntemlerinin performansları, hata kareleri ortalamasının karekökü (RMSE) ve eşitleme yanlılığına (BIAS) göre değerlendirilmiştir. Araştırmada; örneklem büyüklüğü (750, 1500), kayıp veri oranı (%10, %20, %30) ve kayıp verilerin bulunduğu test formu (her iki test, eşitlenecek test) koşul olarak alınmıştır. Bu amaç doğrultusunda araştırmadan elde edilen sonuçların kayıp veriyle başa çıkma yöntemleri ve test eşitleme kavramlarının birlikte ele alındığı çalışmalara farklı koşullar ile katkı sağlayacağı düşünülmektedir.

Araştırma Problemi

Tam veriye sahip test formları ve bu formlardan farklı örneklem büyüklüğü, kayıp veri oranı ve kayıp verilerin bulunduğu test formuna göre elde edilen formlar kayıp veriyle başa çıkma yöntemlerine göre tamamlanıp Haebara ve Stocking-Lord yöntemlerine göre eşitlendiğinde RMSE ve BIAS değerleri nasıl değişmektedir?

Alt Problemler

1. Tam veriye sahip test formları ve kayıp veriyle başa çıkma yöntemleri (SA, HDA, PMM, LOGREG) uygulanarak tamamlanan test formları, Haebara ve Stocking-Lord yöntemlerine göre eşitlendiğinde, araştırmada ele alınan koşullara göre;

- a. RMSE değerleri,
 - b. BIAS değerleri
- nasıl değişmektedir?
2. Referans değer ve her bir kayıp veriyle başa çıkma yöntemi (SA, HDA, PMM, LOGREG) için araştırmada ele alınan koşulların (örneklem büyüklüğü, kayıp veri oranı ve kayıp verilerin bulunduğu test formu) ortalama RMSE ve BIAS değerleri üzerinde anlamlı bir etkisi var mıdır?
 3. Kayıp veriyle başa çıkma yöntemleri (HDA, SA, PMM, LOGREG) uygulanarak tamamlanan test formları, Haebara ve Stocking-Lord yöntemlerine göre eşitlendiğinde referans değere en yakın ve en uzak;
 - a. RMSE ve BIAS değeri üreten kayıp veriyle başa çıkma yöntemi hangisidir?
 - b. ortalama RMSE ve BIAS değeri üreten kayıp veriyle başa çıkma yöntemi hangisidir?
 - c. ortalama RMSE ve BIAS değeri üreten karakteristik eğri dönüştürme yöntemi hangisidir?

Sayıltılar

- Öğrenciler e-TIMSS 2019 fen bilimleri testinde maddeleri cevaplarken gerçek performanslarını göstermişlerdir.
- e-TIMSS 2019 fen bilimleri testine katılan bütün öğrenciler kitapçıkları benzer koşullar altında cevaplandırmışlardır.

Sınırlılıklar

Bu araştırma;

- e-TIMSS 2019 fen bilimleri testinde; en çok sayıda, iki kategorili puanlanan madde, ortak madde ve yanıtlayıcının bulunduğu 7. ve 8. kitapçıklarda çoklu bağlantı probleminin tespit edilmesi ve bu probleme yol açan maddeler çıkarıldığında geriye kalan madde sayısının araştırma için yeterli olmamasından dolayı söz konusu kriterleri ikinci sırada sağlayan 6. ve 7. kitapçıklardan elde edilen veriler,

- Gerçek veri seti kullanıldığı için sınırlı bir evren içerisinde tekrar tekrar seçilecek örneklem sayısı göz önüne alındığında 50 tekrar ile sınırlıdır.

Tanımlar

LOGREG: Lojistik regresyon ile çoklu atama yönteminin kısaltması. R kütüphanesinde yer alan “mice” paketinde bu yöntem “Imputation by logistic regression” olarak ifade edilmiş ve kısaltması “*logreg*” olarak belirtilmiştir. Bu araştırmada LOGREG olarak kullanılmıştır.

PMM: Tahmine dayalı ortalama eşleştirme ile çoklu atama yönteminin kısaltması. R kütüphanesinde yer alan “mice” paketinde bu yöntem “Imputation by predictive mean matching” olarak ifade edilmiş ve kısaltması “*pmm*” olarak belirtilmiştir. Bu araştırmada PMM olarak kullanılmıştır.

HDA: Hot deck atama (sıcak deste atama) yönteminin kısaltması. Her kayıp değer yerine aynı araştırmadan elde edilen “benzer” bir cevap kümesinden gözlemlenen bir yanıtın atanması.

Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Araştırmanın kuramsal temeli; Madde Tepki Kuramı, test eşitleme ve kayıp veri olmak üzere üç konu başlığı çerçevesinde oluşturulmuştur.

Madde Tepki Kuramı

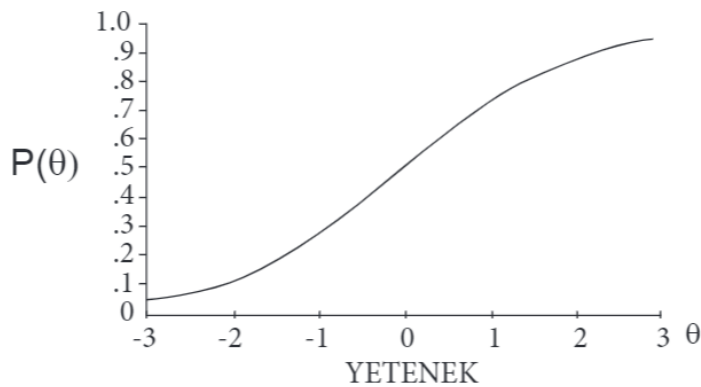
Psikoloji ve eğitim alanlarında, bireylerin yetenek, özellik, tutum gibi örtük değişkenlere sahip olma derecesini belirlemek amacıyla ölçme çalışmaları gerçekleştirilmektedir. Bu çalışmalar Klasik Test Kuramı (KTK) ve Madde Tepki Kuramı (MTK) gibi uygulamalara dayanmaktadır. Bir yeteneğin ölçülmesi için bireylerin uygun yanıt verebileceği belirli sayıda ve birbirinden bağımsız maddelerden oluşan bir test geliştirilir. Test iki kategorili ve/veya çok kategorili maddelerden oluşabilir. İki kategorili puanlanan maddelerden oluşan bir testi alan bireylerin her bir madde için verdikleri doğru cevaplar 1 (bir), yanlış cevaplar 0 (sıfır) olacak şekilde puanlanır. Çok kategorili puanlanan maddelerden oluşan bir testi alan bireylerin maddeye ilişkin puanları kısmi puanları içerebilir.

KTK'de her bir test maddesinden alınan puanların toplanması ile bireylerin ham puanları belirlenirken MTK'de ham puanların yerine yetenekler kestirilir. Her bir yetenek düzeyindeki (θ) bireylerin maddeleri doğru cevaplama olasılığı olan $P(\theta)$ ile ilgilenilir (Baker, 2001). KTK test geliştirme çalışmalarında sıklıkla kullanılmasına rağmen teorik olarak daha doğrulanabilir ölçme ilkeleri sunması ve ölçme problemlerini çözmek için pratik çözümler getirmesi nedeniyle özellikle standartlaştırılmış testlerin geliştirilmesinde MTK'nin kullanımı giderek daha fazla yaygınlaşmıştır (Embretson & Reise, 2000). MTK, bireyin sahip olduğu örtük özellikler ile söz konusu özellikleri ölçmek için geliştirilmiş ölçme aracının maddelerine verdikleri yanıtlar arasındaki ilişkiyi matematiksel bir fonksiyon ile anlatan bir model ortaya koymaktadır (Crocker & Algina, 1986).

Kuramın önemli özelliklerinden biri yetenek parametrelerinin (θ) ölçme aracından, test ve madde parametrelerinin ise testi alan gruptan bağımsız olmasıdır. Gruptan bağımsız olarak kestirilen madde parametrelerinin değişmezlik özelliği sayesinde ölçme sonuçları gruplar arasında karşılaştırılabilir. Tüm yetenek düzeyleri için standart hataların ayrı ayrı kestirildiği MTK'de (Embretson & Reise, 2000) ölçmenin standart hatası ve güvenilirlik bilgi fonksiyonu ile hesaplanır. Madde bilgisi arttıkça güvenilirlik de artar, standart hata ise azalır. Madde düzeyinde tanımlanabilen madde bilgi fonksiyonlarının toplanması ile test bilgi fonksiyonu elde edilir (Demars, 2016). Tüm kuramlarda olduğu gibi MTK'nin de sağlanması gereken varsayımları vardır. Bu varsayımlar temel olarak; tek boyutluluk, yerel bağımsızlık ve değişmezlik olarak sırlanabilir (Embretson & Reise, 2000). Değişmezlikten kasıt aynı maddelerin farklı bireylere uygulanması ile kestirilen madde parametrelerinin yani madde karakteristik eğrisinin (MKE) değişmezliğidir (Hambleton ve diğerleri, 1991). Baker (2001)'a göre; MKE, MTK'nin temel yapı taşıdır. Bir testte yer alan her bir madde için ayrı MKE'ler bulunur. Şekil 1'de gösterilen ve S şeklinde olan bu eğri, yetenek düzeyi ile maddeyi doğru cevaplama olasılığı arasındaki ilişkinin bir ifadesidir.

Şekil 1

Tipik Bir Madde Karakteristik Eğrisi



İki kategorili (1-0) puanlanan maddeler için dört farklı MKE modeli mevcuttur. Bu modeller, yetenek düzeyleri ile madde yanıtları (1-0) arasındaki ilişkiye dair matematiksel

eşitlikte kullanılan madde parametre sayısına göre; bir parametrelili lojistik model, iki parametrelili lojistik model, üç parametrelili lojistik model ve dört parametrelili lojistik model olarak adlandırılır

Bir Parametrelili Lojistik Model (1PLM)

Bir parametrelili lojistik model en basit tepki modelidir ve MKE'nin şekli sadece güçlük (b) parametresi ile belirlenir. MKE'nin büküm noktası olan b parametresi, bir bireyin i maddesini doğru cevaplama olasılığının %50 olduğu yetenek düzeyini ifade eder (Hambleton & Swaminathan, 1985). 1PLM, Rasch model olarak da bilinir ancak Rasch modelde ayırt edicilik (a) parametresi sabit olarak 1 değerini alırken 1PLM'de a parametresi madde ayırt ediciliklerin ortalama değerini alır (De Ayala, 2009). Modelde sadece b parametresi kestirilmekte olup a parametresi tüm maddeler için aynı kabul edilmektedir. Şans (c) parametresi ise sıfır kabul edilmekte yani ihmal edilmektedir (Hambleton & Swaminathan, 1985). 1PLM için belirli bir yetenek düzeyindeki bireyin maddeyi doğru cevaplama olasılığı Eşitlik 1 ile gösterilmektedir (Crocker & Algina, 1986; Hambleton ve diğerleri, 1991):

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

(1)

θ = Bireyin yetenek düzeyi

$P_i(\theta)$ = θ yetenek düzeyindeki bireyin i maddesini doğru cevaplama olasılığı

b_i = i maddesinin güçlüğü

İki Parametrelili Lojistik Model

İki parametrelili lojistik modelde MKE'nin şekli a ve b parametreleri ile belirlenir. Dolayısıyla 2PLM'de belirli bir yetenek düzeyindeki bireyin maddeyi doğru cevaplama olasılığı hesaplanırken a parametresi de kullanılır (Crocker & Algina, 1986). MKE'nin büküm noktasının eğimi olan a parametresi, bir maddenin farklı yetenek düzeylerine sahip bireyleri

birbirinden ayırt edebilme derecesini gösterir (Hambleton & Swaminathan, 1985). c parametresi ise yine sıfır kabul edilir. Modelde bir bireyin i maddesine doğru cevap verme olasılığı Eşitlik 2’de gösterilmektedir (Crocker & Algina, 1986):

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

(2)

θ = Bireyin yetenek düzeyi

$P_i(\theta)$ = θ yetenek düzeyindeki bireyin i maddesini doğru cevaplama olasılığı

a_i = i maddesinin ayırt ediciliği

b_i = i maddesinin güçlüğü

Üç Parametrelili Lojistik Model (3PLM)

Üç parametrelili lojistik modelde a ve b parametrelerine c parametresi de eklenerek MKE’nin şekli belirlenir. Bireyin bir maddeyi şansla doğru cevaplama olasılığı olan ve şans parametresi (guess parameter) olarak da adlandırılan c parametresi MKE’de sıfır olmayan düşük asimptota karşılık gelir. MKE’nin büküm noktası olan b parametresi ise 1PLM ve 2PLM’den farklı olarak, bir bireyin i maddesini doğru cevaplama olasılığının $(1+c)/2$ olduğu yetenek düzeyini gösterir (Hambleton & Swaminathan, 1985). 3PLM’de bir bireyin i maddesine doğru cevap verme olasılığı Eşitlik 3 ile hesaplanmaktadır (Crocker & Algina, 1986; Embretson & Reise, 2000; Hambleton ve diğerleri, 1991):

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

(3)

θ = Bireyin yetenek düzeyi

$P_i(\theta)$ = θ yetenek düzeyindeki bireyin i maddesini doğru cevaplama olasılığı

a_i = i maddesinin ayırt ediciliği

b_i = i maddesinin güçlüğü

c_i = i maddesinin şansla doğru cevaplanma olasılığı

Dört Parametrelili Lojistik Model (4PLM)

3PLM, düşük yetenek düzeyindeki bireylerin bir maddeyi şansla doğru cevaplama olasılığını dikkate almasına rağmen yüksek yetenek düzeyindeki bireylerin bir maddeyi dikkatsizlikle yanlış cevaplama olasılığını sıfır (0) kabul etmektedir. Bu durumun yüksek yetenek düzeyindeki bireyler için dezavantaja yol açtığını belirten Barton ve Lord (1981) tarafından 3PLM'ye d parametresi eklenerek dört parametrelili lojistik model geliştirilmiştir. Kayma parametresi (slipping parameter) olarak da adlandırılan d parametresi 0'dan yüksek 1'den düşük değer aldığı için yüksek yetenek düzeyindeki bireylerin dikkatsizlik sonucu verecekleri yanlış cevaplardan dolayı kaybedecekleri puan miktarı daha az olacaktır. 4PLM'de bir bireyin i maddesine doğru cevap verme olasılığı Eşitlik 4 ile hesaplanmaktadır (Barton & Lord, 1981):

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

(4)

θ = Bireyin yetenek düzeyi

$P_i(\theta)$ = θ yetenek düzeyindeki bireyin i maddesini doğru cevaplama olasılığı

a_i = i maddesinin ayırt ediciliği

b_i = i maddesinin güçlüğü

c_i = i maddesinin şansla doğru cevaplanma olasılığı

d_i = i maddesinin dikkatsizlikle yanlış cevaplanma olasılığı

Test Eşitleme

Testlerden elde edilen puanlar, bireyler hakkında genellikle önemli kararlar verirken (bir üst eğitim kurumuna kabul edilme, işe yerleştirme vb.) bir bilgi olarak kullanılır. Bilgi ne kadar doğru ise karar da o kadar doğru olur. Karar vermek, testlerin birden çok durumda uygulanmasını gerektirir. Ancak bu uygulamalar sonucunda, sınava girmemiş kişiler test soruları hakkında daha önce sınava girmiş kişiler tarafından bilgilendirilebilir ya da birden çok kez girilebilecek sınavlarda bir kişiye farklı dönemlerde gireceği iki sınavda da aynı test soruları verilebilir (Kolen, 1988). Bu durumun yol açacağı test güvenliği sorunu için içerik ve istatistiksel özellikler açısından mümkün olduğunca benzer olan farklı test soruları koleksiyonu yani test formları geliştirilir (Schmeiser & Welch, 2006). Ancak aynı yapıyı ölçmek için tasarlanmış bu formlar benzer olsa da tamamen eşit olmayabilir, genellikle güçlük açısından biraz farklılık gösterebilir. İçerik ve güçlük bakımından benzer olacak şekilde oluşturulmuş formlar “test eşitleme” adı verilen istatistiksel yöntem ile ortak bir ölçeğe yerleştirilir ve test eşitleme sonucunda farklı formlardan alınan puanlar birbirinin yerine kullanılabilir. Başarılı bir ölçekleme ve test eşitlemeden sonra bir test formundan aynı puanı alan adayların başka bir test formundan aynı puanı alan adaylar ile aynı başarı düzeyinde olduğu düşünülebilir. Yani elde edilen eşit puanlar, test formlarının kime ya da ne zaman uygulandığından bağımsız olarak aynı anlama sahiptir (Kolen & Brennan, 2014).

Benzer güçlük düzeyindeki test formlarının benzer yetenek düzeyindeki bireylere uygulanması sonucu gerçekleştirilen süreç “yatay eşitleme”, farklı güçlük düzeyindeki test formları farklı yetenek düzeyindeki bireylere uygulanması sonucu gerçekleştirilen süreç ise “dikey ölçekleme” olarak adlandırılır (Hambleton & Swaminathan, 1985; Kolen, 1988). Dikey ölçeklemede, örneğin ilkökul düzeyindeki öğrencilerin gelişimsel puanlarının yıldan yıla takip edilebilmesi için, farklı sınıf seviyelerindeki öğrencilere mevcut sınıf seviyeleriyle eşleşen içeriği test eden sorular sorulur. Ancak testlerin içeriği farklı olduğu için, farklı eğitim seviyelerine yönelik bu testlerin puanları birbiri yerine kullanılamaz (Kolen & Brennan,

2014). Bu araştırmanın amacı doğrultusunda, kullanılan test formları için yatay eşitleme uygulanmıştır.

Test eşitleme sürecinin yürütülebilmesi için eşitleme koşullarının sağlanması gerekmektedir. Bu koşullar; Klasik Test Kuramı'na (KTK) ve Madde Tepki Kuramı'na (MTK) dayalı eşitleme yöntemlerine göre farklılık göstermektedir (Hambleton & Swaminathan, 1985). KTK'ye dayalı yapılan eşitleme yöntemlerinde; aynı yapıyı ölçme, simetri, eşitlik, grup değişmezliği (Lord, 1980) ve eşit güvenirlik (Dorans & Holland, 2000) gibi eşitleme koşullarının sağlanması gerekmektedir. MTK'ye dayalı yapılan eşitleme yöntemlerinde ise model veri uyumu koşulunun sağlanması gerekmektedir (Kolen, 1981). Bu koşul sağlandıktan sonra sırası ile aşağıdaki basamaklar takip edilir (Hambleton & Swaminathan, 1985);

- Eşitleme desenine karar verilir
- Madde ve yetenek parametreleri kestirilir ardından eşitleme yöntemi seçilir
- Madde ve yetenek parametreleri ölçek dönüştürme yöntemi kullanılarak ortak bir ölçeğe yerleştirilir (tek grup ya da denk grup deseninin kullanıldığı durumlarda parametreler aynı ölçek üzerinde olacağı için bu basamak gerekli değildir)
- Test puanlarının raporlanacağı ölçeğe (gerçek puanlar cinsinden ve/veya gözlenen puanlar cinsinden) karar verilir.

Eşitleme Desenleri

Eşitleme çalışmalarında veri toplama desenleri; tek grup deseni, denk grup deseni, tesadüfi grup deseni, denk olmayan gruplarda ortak madde deseni (Kolen & Brennan, 2014), denk olmayan gruplarda ortak değişken deseni, denk olmayan gruplarda ortak madde ve ortak değişken deseni (Wiberg & Branberg, 2015) olarak sınıflandırılabilir. Bu araştırmada denk olmayan gruplarda ortak madde deseni kullanılmıştır. Bu desen genellikle başarı testlerinden oluşan büyük ölçekli ulusal ya da uluslararası sınavlarda ve test güvenliği endişesinden dolayı aynı formun tekrar uygulanamadığı durumlarda kullanılır.

Desende, her bir cevaplayıcıya sadece bir test formu uygulanır ve bu test formları ortak maddeler içerir. Farklı test formlarını alan ve denk olmayan grupların, grup farklılıklarını form farklılıklarından ayırmak amaçlanır (Kolen & Brennan, 2014).

Tablo 1’de denk olmayan gruplarda ortak madde deseni verilmiştir.

Tablo 1

Denk Olmayan Gruplarda Ortak Madde Deseni

Evren	Örneklem	X Formu	Çapa Test	Y Formu
A	A ₁	✓	✓	
B	B ₁		✓	✓

✓ Örneklemdeki grupların ilgili sütundaki testleri aldıklarını belirtir.

Denk olmayan gruplarda ortak madde ya da çapa test deseni olarak adlandırılan bu desende, Tablo 1’de belirtildiği gibi, birbirine denk olmayan iki farklı gruptan A₁, X formunu; B₁, Y formunu ve ayrıca her iki grup ortak maddelerden oluşan çapa testi almaktadır. Çapa testin rolü gruplar arasındaki yetenek farklılıklarını ölçmek, eşitlemedeki yanlılığı ortadan kaldırmak ve eşitleme fonksiyonunun tahmininde kesinliği artırmaktır. İyi bir test eşitleme için çapa testin, X ve Y formları ile aynı yapıyı ölçen ve X ve Y formlarını çok iyi temsil eden bir mini test olması gerekmektedir (Angoff, 1984; Dorans ve diğerleri, 2010). Başka bir deyişle, ortak madde seti eski ve yeni test formlarında benzer şekilde davranmalı ve içerik ile istatistik özellikler açısından toplam test formlarını orantılı olarak temsil etmelidir. Ayrıca ortak maddeler farklı test formlarında, (madde numarası olarak) aynı yerde bulunmalı ve (ifade değişikliği, yeniden düzenleme vb. olmadan) tamamen aynı olmalıdır (Kolen & Brennan, 2014).

Uluslararası sınavların çoğu (TIMSS, PISA, PIRLS vb.) ilgilenilen evreni temsil eden bir örneklemde tanımlanmış bir dizi beceriyi ölçer. Örneğin, TIMSS uygulaması dördüncü ve sekizinci sınıf öğrencilerinin matematik ve fen becerilerine odaklanır. Bilgisayara tabanlı uygulamada (eTIMSS) sekizinci sınıf öğrencilerinin fen değerlendirmesi 212 maddeden oluşmaktadır ve bu madde sayısı, sınav süresi açısından bakıldığında sınava giren tüm

öğrencilerin her maddede test edilmesini imkânsız hale getirmektedir. Bunun yerine 212 madde 14 fen bloğuna dağıtılarak sınava giren her öğrencinin kitapçığında bu bloklardan iki tanesi yer alır. Bu kitapçık tasarımı, her blok (ve dolayısıyla her bir madde) iki farklı kitapçıkta da görüldüğü için kitapçıklar arasında bağlantı kurulmasını sağlar. Söz konusu tasarımdan dolayı öğrencilere tüm maddelerin yalnızca bir kısmı uygulanır (TIMSS 2019 kitapçık tasarımına ilişkin detaylı bilgi için bakınız Mullis ve diğerleri, 2020).

Gonzalez ve Rutkowski'ye (2010) göre kitapçık tasarımları genel olarak tamamlanmış ve tamamlanmamış matrix tasarımlar olarak ele alınır. Tamamlanmış matrix kitapçık tasarımlarında, tüm formlar tüm maddeleri içerir ve bu nedenle sınava giren tüm kişilere değerlendirmedeki tüm maddeler uygulanır. Tamamlanmamış matrix kitapçık tasarımlarında ise herhangi bir form tüm maddelerin yalnızca bir alt kümesine sahiptir ve sınava giren herhangi bir kişiye toplam madde sayısının yalnızca bir alt kümesi uygulanır.

Test Eşitleme Yöntemleri

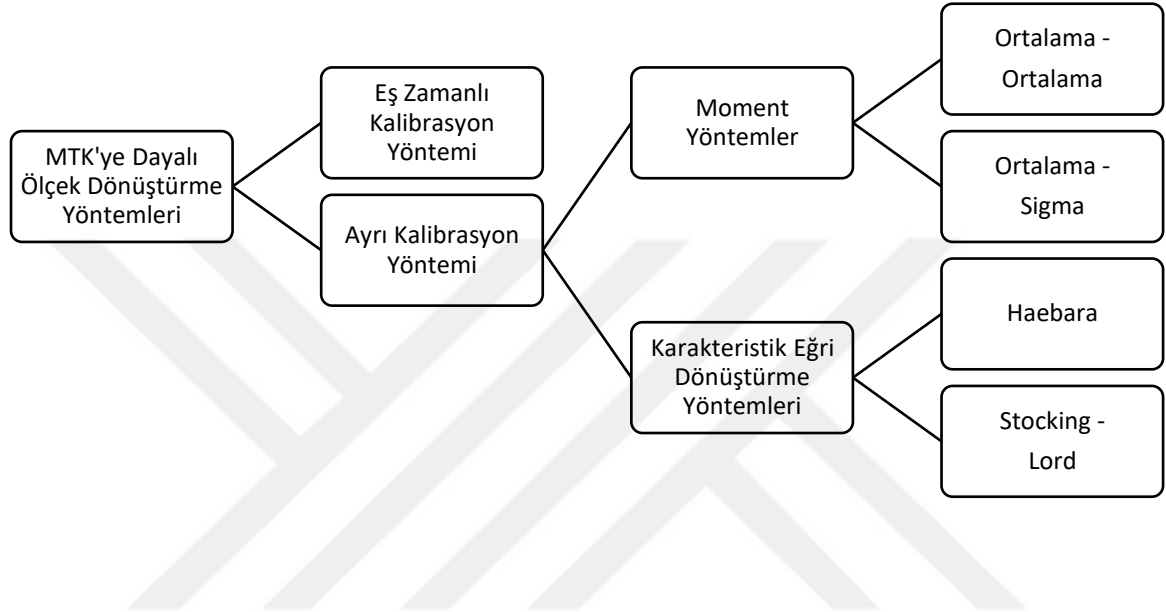
Uygun eşitleme deseni seçilip veri toplandıktan sonra madde ve yetenek parametreleri kestirilir. Ardından eşitlenecek test formları için eşitleme yöntemine ya da yöntemlerine karar verilir. Test eşitleme yöntemleri, KTK'ye ve MTK'ye dayalı yöntemler olmak üzere iki başlık altında incelenebilir. KTK'ye dayalı test eşitlemede test formları arasındaki farklılıkları ayarlamak için; ortalama eşitleme, doğrusal eşitleme ve eşit yüzdelikli eşitleme yöntemleri kullanılır (KTK'ye dayalı test eşitleme yöntemleri ile ilgili ayrıntılı bilgi için bakınız Hambleton & Jones, 2012; Hanson ve diğerleri 1994; Kolen, 1988; Kolen & Brennan, 2014).

MTK'ye dayalı test eşitlemede ise test formları için kestirilen madde ve yetenek parametreleri ölçek dönüştürme yöntemi kullanılarak ortak bir ölçeğe yerleştirilir (Hambleton & Swaminathan, 1985). Örneğin, denk olmayan gruplarda ortak madde deseni seçildiğinde test formlarını farklı gruplar alacağı için söz konusu formlarından elde edilen parametreler aynı ölçekte değildir. Bu durumda iki test formunu aynı ölçeğe yerleştirmek için ortak

maddeler kullanılarak doğrusal bir dönüştürme yapılmalıdır. MTK'ye dayalı ölçek dönüştürme yöntemleri Şekil 2'de gösterilmiştir.

Şekil 2

MTK'ye Dayalı Ölçek Dönüştürme Yöntemleri



Şekil 2'ye göre; MTK'ye dayalı ölçek dönüştürme yöntemleri; eş zamanlı kalibrasyon (concurrent calibration) ve ayrı kalibrasyon (separate calibration) yöntemleri olarak ele alınmaktadır (Kolen & Brennan, 2014). Bu iki yöntem, ortak maddeler için madde parametre kestirimlerinin ele alınma şekillerinde farklılık gösterir.

Eş Zamanlı Kalibrasyon Yöntemi

Eş zamanlı kalibrasyon yönteminde her iki test formundaki parametre kestirimleri tek bir analiz ile aynı anda ve tek aşamada gerçekleştirilir (Hanson & Beguin, 2002). Bu yöntem, madde ve yetenek parametrelerinin aynı anda, tek bir bilgisayar çalışmasında tahmin edilmesini içerir (Lord, 1980). Söz konusu yöntemin, temel gruptan ortak maddelerin parametre tahminlerinin sabit olarak ayarlandığı ve geriye kalan madde parametrelerinin hedef gruptan alınan veriler kullanılarak tahmin edildiği bir çeşidi de mümkündür. Ancak her iki durumda da ortak maddeler için yalnızca bir parametre tahmini olacaktır (Kolen & Brennan, 2014).

Ayrı Kalibrasyon Yöntemi

İki test formu arasındaki karşılaştırmayı sağlamak için ayrı kalibrasyonla test eşitlemede aşağıdaki adımlar izlenir:

1. Temel ve hedef formlardaki madde parametreleri kestirilir.
2. Farklı ölçekte olan parametrelerin karşılaştırılabilmesi için ortak maddelerin a ve b parametrelerine dayalı A (eğim) ve B (kesişim) eşitleme katsayıları hesaplanır.
3. Bir formdaki θ değeri, A ve B katsayıları kullanılarak diğer formdaki θ değerine dönüştürülür.

Ayrı kalibrasyon yöntemleri, moment yöntemler ve karakteristik eğri dönüştürme yöntemleri olmak üzere ikiye ayrılmaktadır.

Moment Yöntemler. Moment yöntemler, ortalama-ortalama (mean-mean) ve ortalama-sigma (mean-sigma) yöntemleri olarak iki kategoride ele alınır. Ortalama-ortalama yönteminde; A eşitleme katsayısı, ortak maddelerden elde edilen madde ayırt edicilik (a) parametresinin ortalaması ile B eşitleme katsayısı ise ortak maddelerden elde edilen madde güçlük (b) parametresinin ortalaması ile kestirilir (Lloyd & Hoover, 1980). Ortalama-sigma yönteminde; A ve B eşitleme katsayıları, ortak maddelerden elde edilen güçlük (b) parametresinin ortalama ve standart sapması ile kestirilir (Marco, 1977).

Bu çalışmada karakteristik eğri dönüştürme yöntemleri kullanıldığı için aşağıda bu yöntemler hakkında detaylı bilgi verilmiştir.

Karakteristik Eğri Dönüştürme Yöntemleri. Çok farklı parametre kestirimlerine (örneğin, farklı b parametrelerine) sahip bir madde çok benzer madde karakteristik eğrileri ürettiğinde moment yöntemler bu parametre kestirimleri arasındaki farktan aşırı derecede etkilenebilir. Tüm madde parametrelerinin aynı anda dikkate alınmamasından kaynaklanan bu soruna çözüm olarak Haebara (1980) ile Stocking ve Lord (1983) tarafından yöntemler geliştirilmiştir (Kolen & Brennan, 2014). Karakteristik eğri dönüştürme yöntemleri olarak adlandırılan bu yöntemler; madde ayırt edicilik (a), güçlük (b) ve şans (c) parametrelerini

aynı anda dikkate alarak test veya madde karakteristik eğrileri arasındaki farkı açıklayan fonksiyonları kullanır.

Haebara (1980) yöntemi tanımlanan bir kayıp fonksiyonu en aza indirmek üzerine kurulmuştur. Bu yöntemde kullanılan fonksiyona göre; belirli bir yetenek düzeyindeki cevaplayıcılar için madde karakteristik eğrileri arasındaki fark, her bir maddeye ait madde karakteristik eğrileri arasındaki farkın karelerinin toplamıdır. Fonksiyonun matematiksel ifadesi en sade şekli ile Kolen ve Brennan (2014) tarafından Eşitlik 5'te gösterilmiştir:

$$H_{crit} = \sum_i \sum_{j:V} [p_{ij}(\theta_{Ti}; a_{Tj}, b_{Tj}, c_{Tj}) - p_{ij}(\theta_{Ti}; \frac{a_{Hj}}{A}, Ab_{Hj} + B, c_{Hj})]^2 \quad (5)$$

Eşitlik 5' te;

H_{crit} : Haebara kriterini (criterion),

p_{ij} : i cevaplayıcısının j maddesini doğru cevaplama olasılığını,

θ_{Ti} : i cevaplayıcısının temel formdaki yetenek düzeyini,

a_{Tj}, b_{Tj}, c_{Tj} : temel formdaki j ortak madde için sırasıyla a, b ve c parametrelerini,

$\frac{a_{Hj}}{A}, Ab_{Hj} + B, c_{Hj}$: hedef formdaki j ortak madde için sırasıyla a, b ve c parametrelerini,

A : Dönüştürme denkleminde eğim katsayısını,

B : Dönüştürme denkleminde kesişim katsayısını

ifade eder.

Haebara yönteminde kestirim süreci, madde karakteristik eğrileri arasındaki farkların karelerinin toplamını en aza indiren dönüşüm katsayıları A ve B'yi bularak ilerler. Böylece toplam cevaplayıcılar üzerinden Haebara kriteri hesaplanır.

Stocking-Lord (1983) yöntemi gerçek puan kavramı üzerine kurulmuştur. İdeal koşullar altında yeni form, eski formun ölçeğine dönüştürüldükten sonra kestirilen gerçek

puan, eski formunki ile aynı olmalıdır. Ancak MTK modellerinin uyumsuzluğu gibi nedenlerden dolayı gerçek puanlar arasında çoğunlukla fark vardır. Bu nedenle eşitlenen iki formun kestirilen gerçek puanları arasındaki farkın en aza indirilmesi gerekir. Bu amaç için Stocking ve Lord (1983) tarafından Haebara (1980) yöntemine benzer bir yöntem geliştirilmiştir. Yöntemler arasında kayıp fonksiyonun hesaplanması sürecinde farklılık vardır. Stocking-Lord yönteminde kullanılan fonksiyona göre; belirli bir yetenek düzeyindeki cevaplayıcılar için madde karakteristik eğrileri arasındaki fark, her bir maddeye ait madde karakteristik eğrileri arasındaki farkın toplamının karesidir. Fonksiyonun matematiksel ifadesi en sade şekli ile Kolen ve Brennan (2014) tarafından Eşitlik 6'da gösterilmiştir:

$$SL_{crit} = \sum_i \left[\sum_{j:V} p_{ij}(\theta_{Ti}; a_{Tj}, b_{Tj}, c_{Tj}) - \sum_{j:V} p_{ij}\left(\theta_{Ti}; \frac{a_{Hj}}{A}, Ab_{Hj} + B, c_{Hj}\right) \right]^2 \quad (6)$$

Eşitlik 6'da;

SL_{crit} : Stocking-Lord kriterini (criterion),

p_{ij} : i cevaplayıcısının j maddesini doğru cevaplama olasılığını,

θ_{Ti} : i cevaplayıcısının temel formdaki yetenek düzeyini,

a_{Tj}, b_{Tj}, c_{Tj} : temel formdaki j ortak madde için sırasıyla a, b ve c parametrelerini,

$\frac{a_{Hj}}{A}, Ab_{Hj} + B, c_{Hj}$: hedef formdaki j ortak madde için sırasıyla a, b ve c parametrelerini,

A : Dönüştürme denkleminde eğim katsayısını,

B : Dönüştürme denkleminde kesişim katsayısını

ifade eder.

Stocking-Lord yönteminde kestirim sürecinde, tüm θ_i düzeyinde test karakteristik eğrileri arasındaki farkın toplamının karesinin minimize edildiği dönüşüm katsayıları (A ve B) elde edilir. Böylece A ve B'yi minimum yapan Stocking-Lord kriteri bulunur. Bu kriter toplam cevaplayıcılar üzerinden hesaplanmıştır.

Ayrı kalibrasyon yöntemlerinden karakteristik eğri dönüştürme yöntemleri (Haebara ve Stocking-Lord), moment yöntemlere (ortalama-ortalama, ortalama-sigma) göre daha küçük standart hatalar ve daha kararlı sonuçlar üretmektedir (Hanson & Beguin, 2002; Kolen & Brennan, 2014, Lee & Ban, 2010; Ogasawara, 2001). Ayrı kalibrasyon yöntemlerinde eşitleme iki aşamada gerçekleşir. Önce temel ve hedef test formlarındaki madde parametreleri kestirilir sonra kestirilen madde parametreleri aynı ölçeğe yerleştirilir. Ancak eş zamanlı kalibrasyon yönteminde temel ve hedef test formlarındaki parametre kestirimi ve ölçeğe yerleştirme süreci aynı anda ve tek aşamada gerçekleşir. Dolayısıyla MTK modelinin varsayımları sağlandığı zaman eşzamanlı kalibrasyon yöntemi, karakteristik eğri dönüştürme yöntemlerine göre daha iyi performans göstermektedir (Chu, 2002; Hanson & Beguin, 2002). Ancak MTK modelinin varsayımları sağlanamadığında ayrı kalibrasyon yöntemi daha doğru sonuçlar üretmektedir. Özellikle tek boyutluluk varsayımı sağlanamadığında ayrı kalibrasyon yöntemi tercih edilebilir. Avantajlarına rağmen eşzamanlı kalibrasyon yöntemi, birden fazla sınıf düzeyinde testler aynı anda yapıldığından çok boyutluluk ve yakınsama sorunlarıyla karşı karşıyadır (Kolen & Brennan, 2014; Patz & Hanson, 2002; Tong & Kolen, 2010). Test eşitleme yöntemlerinin varsayımları azaldıkça performansları artmaktadır. Tek aşamada gerçekleşen eşitleme sürecindeki hatanın iki aşamada gerçekleşen sürece göre daha az olması, ölçek dönüştürme aşamasındaki hata ile ilgilidir (Chu, 2002).

Test Eşitleme Hatası

Test eşitleme hatası, bireye uygulanan ve uygulanmayan testler için kestirilen yetenek düzeyleri arasındaki fark ile ifade edilir. Dolayısıyla eşitleme çalışmalarının hatadan arınık olması için farklı testlerden elde edilen yetenek düzeylerinin eşit olması gerekmektedir (Cook & Eignor, 1991). Bireyin aldığı test formunda kestirilen yetenek düzeyi ile almadığı test formu için dönüştürülmüş yetenek düzeyinin eşit olmaması test eşitleme hatası ile açıklanmaktadır. Test eşitlemede rastgele hata ve sistematik hata olmak üzere iki genel hata kaynağı bulunmaktadır (Kolen, 1988; Kolen & Brennan, 2014). Rastgele

eşitleme hatası; evrenden örneklem alındığında ve örneklemde yer alan katılımcıların sınav puanları, eşitleme ilişkilerini tahmin etmek için kullanıldığında ortaya çıkan hatadır. Özellikle küçük örneklem kullanıldığında tahmini eşitleme ilişkisi, evrendeki eşitleme ilişkisinden farklı olur. Söz konusu eşitleme ilişkilerinin tahmininde örneklem büyüklüğü arttıkça rastgele hata miktarı ihmal edilebilir hale gelir. Sistemik hata (BIAS) ise farklı nedenlerden meydana gelebilir. Sistemik hataya yol açan durumlar şunlardır: eşitleme ilişkilerini tahmin etmede kullanılan yöntemin eşitleme ilişkisini tahmin etmede yanlılık göstermesi, eşitleme yönteminde kullanılan istatistiksel varsayımların ihlal edilmesi, seçilen test eşitleme deseninin uygun olamayan bir şekilde uygulanması (örneğin; denk olmayan gruplarda ortak madde deseni uygulanırken, ortak maddelerin Form X'in başında Form Y'nin sonlarına doğru yer alması yani ortak maddelerin farklı yerleşimleri nedeniyle bu maddeler iki form üzerinde çok farklı davranabilir) ve eşitleme yapmak için kullanılan sınava giren katılımcıların eşitlenmiş formu alan katılımcılardan önemli ölçüde farklı olması. Sistemik hatayı kontrol edebilmek için test geliştirme sürecinin dikkatli yapılması, tercih edilen eşitleme deseninin doğru şekilde uygulanması ve uygun istatistiksel yöntemlerinin kullanılması gereklidir. Rastgele ve sistemik hatayı birbirinden ayıran en önemli fark örneklem büyüklüğü arttıkça rastgele hatanın azalması ancak sistemik hatanın örneklem büyüklüğünden doğrudan etkilenmemesidir (Kolen & Brennan, 2014). Rastgele hata ile eşitleme yanlılığı kullanılarak test eşitlemede oluşan tüm hataya ilişkin hata kareleri ortalamasının karekökü (root mean squared errors-RMSE) hesaplanabilmektedir (Gonzales & Wiberg, 2017; Kolen & Brennan, 2014). RMSE, test eşitlemedeki genel hatayı temsil eder ve Eşitlik 7'de gösterildiği gibi SE (standart hata) ile BIAS (eşitleme yanlılığı) bileşenlerinden oluşur.

$$SE^2 + BIAS^2 = RMSE^2$$

(7)

Kayıp Veri

Araştırmalarda kullanılan ölçme araçları ile elde edilen verilere eksiksiz ve tam ulaşmak her zaman mümkün değildir ve bu durum kayıp veri problemine yol açmaktadır. Elde edilmesi planlanan ve elde edilen veri seti arasındaki farkı ifade eden kayıp veriler (Enders, 2010), tam veri setlerine yönelik geliştirilen istatistik programlarının kullanımı için önemli bir sorun oluşturmaktadır. İstatistiksel bir model kullanılarak yapılacak kestirimlerde yanlışlıkların ortaya çıkması, bilgi eksikliğinin oluşması sonucu istatistiksel analizlerin gücünün azalması, sıklıkla tercih edilen istatistiksel yöntemlerin kayıp veriye sahip veri setlerinde kullanımının uygun olmaması ve daha önemli alanlarda değerlendirilebilecek kaynakların boşa harcanmış olması kayıp verilere bağlı olarak ortaya çıkabilecek sorunlardır (Peng ve diğerleri, 2007).

Kayıp veriler ile veri seti arasındaki ilişkinin doğası, kayıp veri mekanizmalarının ortaya çıkmasını sağlamıştır (Enders, 2010).

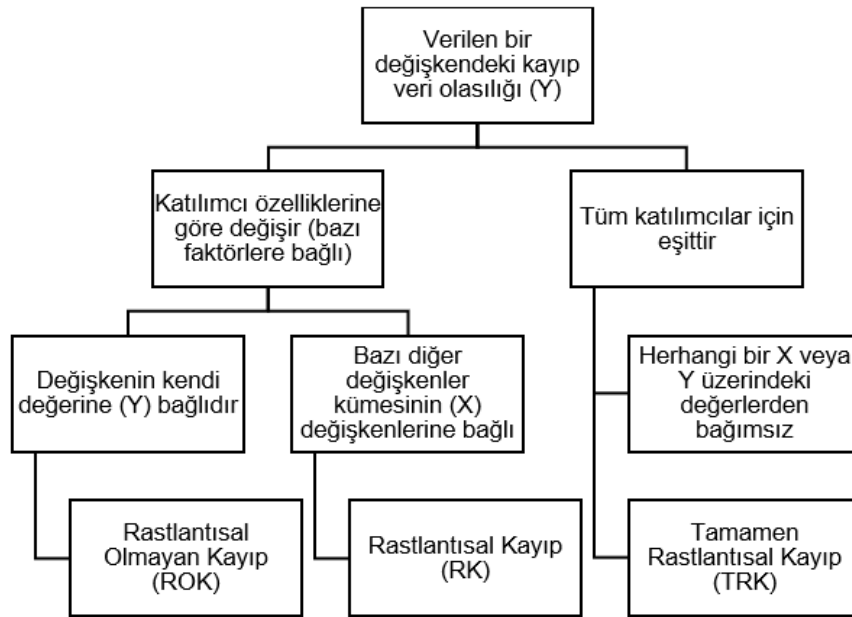
Kayıp Veri Mekanizması

Little ve Rubin (2002) tarafından geliştirilen ve araştırmacılar tarafından sıklıkla kullanılan kayıp veri mekanizmaları; rastgele olmayan kayıp (ROK, missing not at random), rastgele kayıp (RK, missing at random) ve tamamen rastgele kayıp (TRK, missing completely at random) olmak üzere üç başlık altında toplanmıştır. ROK veri mekanizmasında; Y kayıp veri içeren bir değişken olmak üzere, bir Y değişkeninde kayıp veri görülme olasılığı, veri setindeki diğer değişkenler kontrol altına alındıktan sonra dahi Y değişkeninin kendisi ile ilişkili olduğunu göstermektedir. RK veri mekanizmasında; söz konusu Y değişkeninde kayıp veri görülme olasılığı, veri setindeki diğer değişkenler kontrol altına alındıktan sonra Y değişkeninin kendisi ile ilişkisiz olduğunu göstermektedir. TRK veri mekanizması ise söz konusu Y değişkeninde kayıp veri görülme olasılığı, veri setindeki diğer değişkenler ya da Y değişkeninin kendisi ile ilişkisiz olduğunu göstermektedir (Enders, 2010). Örneğin; tüm durumlar için gözlenen X (eğitim yılı) ve kayıp veri içeren Y (suç işleme)

olmak üzere, X ve Y'den oluşan bir veri setinde, Y değişkeninin kayıp veri içermesi ne X değişkenine ne de Y değişkenine bağlıdır. Yani suç işleme değişkeninin kayıp veri içermesi eğitim yılı ve/veya suç işleme ile ilişkili değildir (Allison, 2003). Kayıp veri mekanizmaları Şekil 3'te şematize edilmiştir (Myers, 2011).

Şekil 3

Kayıp Veri Mekanizmaları



Rubin'e (1976) göre; TRK ya da RK veri mekanizmalarında kayıp olma olasılığı, kayıp olma durumunun varsayımsal değerlerine bağlı olmadığı için ihmal edilebilir (ignorable) kayıp veri mekanizmalarıdır. Ancak ROK veri mekanizmasında modeldeki diğer değişkenler dikkate alındıktan sonra bile kayıp verilerin varsayımsal değerlerine doğrudan bağlı olduğundan ihmal edilemez (non-ignorable) kayıp veri mekanizmasıdır. Schafer ve Graham (2002) ise; RK veri mekanizmasını ihmal edilebilir cevapsızlık (ignorable nonresponse), ROK veri mekanizmasını ihmal edilemez cevapsızlık (nonignorable nonresponse) olarak tanımlar. Kayıp verilerin RK ya da ROK veri mekanizmasına sahip olduğunu kanıtlamak için kayıp verilere ait bilgilere ulaşarak söz konusu bilgilerin tamamlanması gereklidir. Yani kayıp verilere ait gerçek değerler bilinmeden RK ve ROK

veri mekanizmaları ayırt edilemez. Kayıp veri dağılımlarının RK ya da ROK veri mekanizmasına sahip olduğuna dair istatistiksel kanıtlar sunulamıyor olsa da TRK veri mekanizması Little'ın TRK testi (Little's MCAR test) ile deneysel olarak test edilebilmektedir (Enders, 2010). Little'ın TRK testi, çok değişkenli nicel veriler için farklı kayıp veri örüntülerinin ortalamaları arasında hiçbir fark olmadığını sıfır hipotezi altında ki-kare (χ^2) istatistiği ile test eder. Sıfır hipotezinin reddedilmesi, verilerin TRK olmadığını belirtmek için yeterli kanıt sağlar (Li, 2013). Kayıp verilerin ihmal edilebilir olup olmadığını belirlemenin başka bir yolu Finch (2008) tarafından sunulmuştur. Finch (2008)'e göre sadece kayıp veri mekanizması değil kayıp veri mekanizması ile kayıp verinin veri seti içerisindeki oranı birlikte değerlendirilmelidir. TRK ya da RK veri mekanizmasına sahip büyük bir veri setinde kayıp veri oranı %5 ve daha az ise kayıp veriler ihmal edilebilir. ROK veri mekanizmasına sahip küçük veya orta büyüklükte bir veri setinde çok sayıda kayıp veri olması durumunda kayıp veriler ihmal edilemez.

Sonuç olarak kayıp veri mekanizması, kayıp veri ile grup üyeliği arasındaki ilişkiye göre belirlenir (Kim, 2015). ROK'ın aksine RK ve TRK veri mekanizmaları daha küçük yanlılıklar ile nispeten daha iyi sonuçlar vermektedir (Ertoprak, 2017; Kim, 2015). Kayıp veri mekanizmasının belirlenmesinden sonraki aşama, ihmal edilemeyen kayıp veriler için veri setine en uygun kayıp veriyle başa çıkma yöntemine/yöntemlerine karar verilmesidir.

Kayıp Veriyle Başa Çıkma Yöntemleri

Kayıp veriyle başa çıkma yöntemleri genel olarak veri silme ve veri atama yöntemleri olmak üzere iki grupta incelenebilir. Liste bazında silme (listwise deletion) ve çiftler bazında silme (pairwise deletion) yöntemleri, geleneksel yöntem olarak da adlandırılan veri silme yöntemleridir. Silme yöntemleri verilerin TRK veri mekanizmasına sahip olmasını gerektirir ancak TRK veri mekanizması geçerli olduğunda bile verilerin atılması istatistiksel gücün azalmasına neden olur (Enders, 2010; Rubin, 1987). Bir veri setinde yer alan kayıp verilere sahip maddeler görmezden gelinerek cevaplayıcılara hiç uygulanmamış gibi davranılması ya da bu maddelerin cevaplayıcılar tarafından yanlış yanıtlanmış gibi davranılması da

analizlerin gücünün ve doğruluğunun azalmasına yol açar. Ayrıca kestirim sonuçlarında yanlışlıklara sebep olur (Allison, 2002; Cheema, 2014). MULTILOG, PARSCALE, STATA, R programlama dili ("itm" paketinde yer alan "t_{pm}" fonksiyonu) gibi istatistik programlarında kayıp veriler ele alınırken varsayılan olarak uygulanmamış gibi davranma yöntemi kullanılır. Böylece veri setindeki kayıp veriler görmezden gelinir ve sadece gözlenen veriler kullanılarak parametre kestirimleri gerçekleştirilir. BILOG-MG programında ise kayıp veriler ele alınırken varsayılan olarak yanlış yanıtlanmış gibi davranma yöntemi kullanılır. Veri setinde ikili puanlanan (1-0) maddelerdeki kayıp veriler yerine 0 (sıfır) atama yapılır ve böylece bu veriler yanlış yanıtlanmış gibi değerlendirilerek parametreler kestirilir. Söz konusu istatistik programlarında varsayılan olarak kullanılan bu yöntemlerin neden olduğu olumsuz sonuçlardan dolayı kayıp verilerin yerine veri setindeki bilgileri kullanarak farklı değerler oluşturan atama yöntemlerinin kullanımı artmıştır. Çok sayıda veri atama yöntemi geliştirilmiş olup bunlar arasında; ortalama atama (mean imputation), rastgele atama (random imputation), çoklu atama (multiple imputation), sıfır atama (zero imputation), beklenti maksimizasyonu (expectation maximization-EM), sıcak/soğuk deste atama (hot/cold deck imputation), en yakın komşu (nearest neighbor), son gözlemi ileri taşıma (last observed value carried forward) ve regresyon atama (regression imputation) yer almaktadır.

Atama yöntemleri, silme yöntemlerinden farklı olarak eksiksiz bir veri setinin oluşturulmasına izin vermesi ve bilgi kaybı olmaması açısından avantajlıdır. Silme yöntemlerinde olduğu gibi geleneksel atama yöntemlerinin çoğu da kayıp veri mekanizmasının TRK olmasını gerektirir (Enders, 2010; Rubin, 1987). Silme yöntemleri, ortalama atama ve regresyon temelli tekli atama yöntemi, geleneksel yöntemler; beklenti maksimizasyonu, tam bilgi en çok olabilirlik ve çoklu atama yöntemleri, modern yöntemler arasındadır (Graham, 2009). Geleneksel olarak kullanılan yöntemler katı varsayımlara dayandıkları için genellikle yanlış tahminler üretirler. Kayıp veriyle başa çıkma yöntemleri yıllar içinde büyük ölçüde iyileştirilmiş olsa da kullanım kolaylıkları nedeniyle geleneksel yöntemler günümüzde de tercih edilmektedir (Enders, 2010; Little & Rubin, 2002; Schafer

& Graham, 2002). Ancak kayıp veriyle başa çıkmada modern yöntemlerin tek başına ya da geleneksel yöntemler ile birlikte kullanıldığı çok sayıda araştırma da mevcuttur (Bayram, 2020; Chung & Cai, 2019; Çüm & Gelbal, 2015; Ertoprak, 2017; Işıkoğlu, 2017; Karaman, 2022; Kim, 2015; Lim & Cheung, 2022; Li & Lomax, 2016; Shin, 2016; Tamcı, 2018).

Bu araştırmada kullanılan kayıp veriyle başa çıkma yöntemlerinden sıfır atama (SA), hot deck atama (HDA) ve çoklu atama (ÇA) yöntemlerine ilişkin bilgilere yer verilmiştir. Geleneksel yöntemler (SA ve HDA), TRK veri mekanizmasına ve iki kategorili (1-0) puanlanan verilere uygun oldukları için tercih edilmişlerdir. Modern yöntemler (ÇA) ise mevcut kayıp veri literatüründe en yaygın olarak önerilen yöntemlerden biri olduğu için kullanımına karar verilmiştir.

Sıfır Atama (SA-Zero Imputation). İki kategorili (1-0) verilerde sıklıkla kullanılan ve uygulaması kolay ancak kullanışlılığı sınırlı bir veri atama yöntemidir. Bu yöntemde veri setinde yer alan hiçbir bilgi kullanılmadan kayıp verilerin yerine sıfır (0) atanır. SA yönteminde kayıp veriye sahip tüm maddelere cevaplayıcılar tarafından yanlış cevap verilmiş gibi davranılır (McKnight ve diğerleri, 2007; Finch, 2008). Kayıp verilerin yerine sıfır konulması, çok spesifik koşullarda, örneğin cevaplayıcının doğru cevabı bilmediği için kayıp değer mantıklı olarak varsayılabileceği kayıp başarı puanlarıyla uğraşırken mantıksal olarak anlamlıdır. Bununla birlikte bu yöntem, kayıp verinin ortaya çıkmasından başka nedenler (kaygı, yorgunluk vb.) sorumlu olduğunda yanlış parametre kestirimleri üretir (McKnight ve diğerleri, 2007).

Hot Deck Atama (HDA-Sıcak Deste Atama). “Deck” yani “deste” kavramı, veri setlerinin delikli kart destelerine kaydedilmesinden gelmektedir. “Hot” yani “sıcak” kavramı bu kartların sıcak zimbasına atıfta bulunur. Deste atamada her kayıp değer yerine “benzer” bir cevap kümesinden gözlemlenen bir yanıt atanır. Soğuk deste atama ve hot deck atama olmak üzere iki tür deste atama yöntemi vardır. Soğuk deste atama yönteminde kayıp verilere değer atama için kullanılan deste önceden yapılmış benzer bir araştırmanın veri kümesidir. Hot deck atama yönteminde ise kayıp verilere değer atama için kullanılan deste

aynı araştırmadan elde edilen veri kümesidir (Kowarik & Templ, 2016; Little & Rubin, 2002). Hot deck atama yöntemi sosyal bilimler alanında yapılan çalışmalarda yaygın olarak kullanılmasına rağmen diğer kayıp veri atama yöntemlerine kıyasla teorik olarak daha az gelişmiştir (Myers, 2011). Hot deck atama sürecinde önce bir veri dosyasının satırları (yani yanıt verenleri), deste adı verilen bir dizi değişken içinde sıralanır. Sonra her bir kayıp veri, benzer bir birimden gözlemlenen bir cevapla değiştirilir. Başka bir ifade ile HDA yöntemi, yanıt vermeyen katılımcıya (alıcı) ait bir ya da daha fazla değişkenin kayıp verilerinin, her iki durumda da gözlemlenen özellikler açısından yanıt vermeyenlere benzer bir yanıt verenden (donör) gözlemlenen değerlerle değiştirilmesini içerir (Andridge & Little, 2010). Donörden bilgi seçilmesi rastgele, düzeltilmiş olarak ya da ikili ilişkiler üzerinden yapılabilir (Toka, 2012). Rastgele HDA yönteminde donör, bir dizi potansiyel donörden oluşan donör havuzundan rastgele seçilir. Başka bir ifade ile katılımcılar destelere ayrıldıktan sonra belirli bir deste içindeki tüm cevaplayıcılar rastgele sıralanır ve belirli bir değişkende eksik olan herhangi bir cevaplayıcıya daha sonra bu rastgele izin verilen veri dosyasında eksik verisi olmayan kendisine en yakın cevaplayıcının değeri atanır (Andridge & Little, 2010; Myers, 2011). Diğer versiyonlarda sadece bir donör tanımlanır ve değerler bu durumdan, genellikle bazı ölçütlere dayalı olarak “en yakın komşu” olarak hesaplanır. Donör seçiminde rastgelelik olmadığı için bu yöntemlere deterministik HDA yöntemleri denir. Donörlerle alıcıları eşleştirmek için ölçüt seçimi ve bu ölçüte dahil edilen değişkenler aracılığı ile örtük varsayımlar yapan bu HDA yönteminin başka varsayımları olmadığı için araştırmacılar tarafından tercih edilmektedir. Değerler donör havuzunda yer alan gözlemlenen yanıtlardan geldiği için yalnızca gerçek değerlerin atanması yöntemin bir diğer avantajıdır. Ayrıca bu yöntemde atama sınıflarını tanımlayan değişkenler ile yanıt verme eğilimi ve atanacak değişken arasındaki ilişkinin ölçüsüne göre yanıt vermeme yanlılığında bir azalma olur (Andridge & Little, 2010). Bununla birlikte, küçük veri setlerinde kayıp veri yerine atanabilecek kombinasyona sahip destenin bulunmama ihtimalinin yüksek olması, kayıp verinin yerine atama yapma tercihinin araştırmacıya bırakmasından dolayı olası yanlış tercihlerin yanlış sonuçlar doğurabilmesi ve araştırmacının gözlemlenmiş olabilecek

değerleri bilmemesi gerçeğinden dolayı belirsizliği hesaba katmaması HDA yönteminin sınırlılıklarındandır (Myers, 2011; Siddique & Belin, 2008, Toka, 2012). Söz konusu sınırlılıklara rağmen model tabanlı tekniklere (çoklu atama, beklenti maksimizasyonu, en çok olabilirlik) kıyasla HDA yönteminin kullanım kolaylığı geniş kapsamlı işlemleri kolaylaştırma potansiyeline sahiptir ve birçok kayıp veri problemi için istatistiksel olarak geçerli bir yaklaşımdır (Myers, 2011).

Çoklu Atama (Multiple Imputation). Rubin tarafından 1977 yılında ilk tohumları atılan, bir yıl sonra açıkça önerilen ve on yıl sonra temel referans olarak yayınlanan çoklu atama yöntemi, rastgele kayıp veya tamamen rastgele kayıp veri mekanizması altında, kayıp verilerle karşılaşıldığında geçerli istatistiksel çıkarımlar elde etme sorununa yönelik uygulamalı bir yaklaşım sunmaktadır (Little & Rubin, 2002). Basit atama yöntemlerinde her bir kayıp verinin yerine, atanan değer gerçekmiş gibi, tek bir değer oluşturulur. Çoklu atama yöntemlerinde ise her bir kayıp verinin yerine, atanan değer belirsizliği hesaba katılarak, bir dizi olası değer üretilir (Rubin, 1976). Başka bir ifade ile çoklu atama yönteminde veri setinde yer alan kayıp değerlerin yerine birden fazla atama yapılarak, çok sayıda tam veri seti elde edilir (Little & Rubin, 2002). Kayıp veri problemi çözülmüş tam veri setleri uygun yöntemler kullanılarak analiz edilir ve sonuçlar Rubin'in formülü kullanılarak birleştirilir. Bu formülün matematiksel ifadesi Eşitlik 8'de gösterilmiştir (Rubin, 1987).

$$\bar{\theta} = \frac{1}{m} \sum_{i=1}^m \hat{\theta}_i$$

(8)

Eşitlik 8'de;

m= atama sayısını,

$\hat{\theta}_i$ = parametre kestirimlerini ($\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \dots, \hat{\theta}_m$),

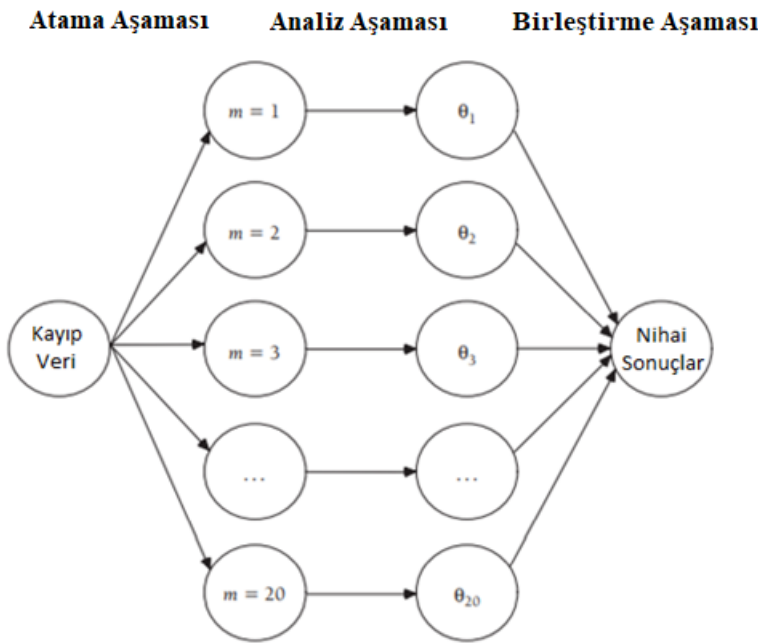
$\bar{\theta}$ = i=1'den m'e kadar elde edilen parametre kestirimlerinin ortalamasını

ifade eder.

Çoklu atama; atama (imputation), analiz (analysis) ve birleştirme (pooling) olmak üzere üç aşamadan oluşur (Şekil 4). Atama aşamasında veri setinin özelliğine göre (kategorik, sürekli) hangi modelin kullanılacağına karar verilir. Belirtilen model kullanılarak çok sayıda ($m=20$) veri seti oluşturulur. Analiz aşamasında m sayıdaki tam veri setinin her biri için standart prosedür kullanılarak ayrı ayrı m kez analiz gerçekleştirilir. Birleştirme aşamasında ise m sayıdaki veri setinin parametre tahminleri ve standart hataları birleştirilir. Parametre tahminleri, aritmetik ortalama değerleri kullanılarak bir araya toplanır. Standart hatalar ise atamalar içindeki ve arasındaki varyanslar kullanılarak birleştirilir. Bu aşamada her şey tek bir sonuç kümesinde birleştirilerek nihai sonuca ulaşılır (Enders, 2010).

Şekil 4

Çoklu Atama Aşamalarının Grafikselleştirilmesi



Çoklu atama yöntemlerinin hepsinde bu üç aşama yer alır ancak yöntemlerdeki farklılık veri atama aşamasında kullanılacak modele göre hesaplama algoritmalarının farklı olmasından kaynaklanır. Örneğin; atama aşamasında sürekli, ikili ve çok kategorili veriler ile bu verilerin karışımını içeren çok değişkenli kayıp veriler için çoklu atamalar oluşturan algoritma, zincirleme denklemlerle çok değişkenli atama (multivariate imputation by chained equations-mice) olarak tanımlanır. "mice" yöntemi, her bir kayıp değişkenin ayrı bir model

tarafından atandığı tam koşullu belirleme yaklaşımına dayanmaktadır. Bu yöntem; R kütüphanesinde yer alan “mice” paketi, SOLAS, WinMICE, ice yazılımı ve SPSS’te kayıp veri analizi ek modülünün çoklu atama prosedürü bölümünün aracılığıyla kullanılabilir (van Buuren & Groothuis-Oudshoorn, 2011).

Bu araştırmada R kütüphanesindeki “mice” paketinde, “*defaultMethod*” (varsayılan yöntem) argümanında yer alan; tahmine dayalı ortalama eşleştirme ile çoklu atama ve lojistik regresyon ile çoklu atama olmak üzere iki farklı çoklu atama yöntemi kullanılmıştır.

Tahmine Dayalı Ortalama Eşleştirme ile Çoklu Atama (Imputation by Predictive Mean Matching-PMM). Çok uzun zamandır var olan PMM yöntemi (Rubin, 1986; Little, 1988), günümüzde çoklu atama yaklaşımı uygulayan birçok yazılım paketinde (R, Stata, SPSS ve SAS) yer almasıyla birlikte yaygın olarak kullanılabilir olmuş ve kullanımı pratik hale gelmiştir (Allison, 2015). Kayıp veri içeren değişkenin ölçek türü; sayısal, ikili, sıralı ya da sınıflamalı verilerden herhangi biri olduğunda kullanılan çoklu atama algoritmasıdır (van Buuren, 2018). İlk kez Rubin (1986) tarafından istatistiksel eşleştirme bağlamında önerilen bu yöntem basit bir çoklu atama prosedürü içerir. Bu prosedüre göre, belirli bir yanıt vermeyen kişiye en yakın k yanıtlayıcıyı bulunur ve m sayıda tahmin değeri sağlamak için bu kümeden örneklem alınır. k’nın seçimi, tahmine dayalı dağılımı etkili bir şekilde simüle edecek kadar büyük ve eşleştirmelerin kalitesini koruyacak kadar küçük olmayı gerektirir. m’nin seçiminde ise bu değerin iki olması bile daha iyi tahminler sunacaktır (Little, 1988). PMM yönteminde kayıp verilere, doğrusal regresyon modeli kullanılarak ortak değişkenlere bağlı en yakın komşu değerlerden oluşturulan donör havuzundan rastgele seçilen tek bir donörün gözlemlenen değeri atanır. Çoklu atama için bu işlem m kez tekrarlanır. Sonrasında analiz ve birleştirme aşamalarına devam edilir (Kleinke, 2017). Kayıp hücrelerin dağılımı ile aday donörlerin gözlenen verilerinin aynı olduğu varsayımına dayanır (van Buuren, 2018). Tüm mesafelerin hesaplanmasını içeren donör havuzunun büyüklüğü 3 ile 10 arasında olduğu zaman genel olarak çoğu durumda en iyi sonuçları sağlar (Morris ve

diğerleri, 2015). Donör havuzunun büyüklüğünün 5 olması genel olarak birçok durum için iyi bir varsayılan değerdir. PMM, doğrusal regresyon ve normal dağılıma dayalı standart yöntemlere göre gerçek değerlere çok daha fazla benzeyen emsal değerler üretir. Çünkü bu yöntemde atanan değerler, gerçek verileri olan bireylerden "ödünç alınan" gerçek değerlerdir (Allison, 2015). Dolayısıyla gözlemlenen veri aralığının dışında anlamsız bir verinin atanması (örneğin, negatif bir boy uzunluğu) problemi ile karşılaşılmayacaktır. Ayrıca PMM yöntemi, atama modelinin yanlış tanımlanmasına karşı dirençlidir ve performansı teorik olarak diğer üstün yöntemler kadar iyidir. Ancak küçük örneklemelerde ve gözlenen verilerden çok daha fazla kayıp veri olması durumunda aynı donör değerinin birçok kez tekrarlanma riskini artıracakı unutulmamalıdır (van Buuren, 2018).

Lojistik Regresyon ile Çoklu Atama (Imputation by Logistic Regression-LOGREG). Lojistik regresyon; iki kategorili bir bağımlı değişkeninin, kategorik ya da sürekli olan bir veya daha fazla bağımsız değişken ile tahmin edilmesidir (Agresti, 2007). Kayıp veri içeren değişkenin ölçek türü ikili veriler olduğunda kullanılan çoklu atama algoritmasıdır. Bu algoritmaya göre lojistik regresyon kullanılarak Bayesçi atama ile kayıp kategorik veriler yerine atama yapılır. Parametreler tekrarlı olarak yeniden ağırlıklandırılmış en küçük kareler ile tahmin edilir. Mükemmel tahmin veri artırma yöntemiyle gerçekleştirilir (van Buuren, 2018). Başka bir ifade ile 1 (doğru yanıt) ve 0 (yanlış yanıt) olarak puanlanan iki kategorili bağımlı değişkende kayıp veriler olması durumunda, model parametreleri ile veri setindeki gözlenen değerlere ait parametre dağılımları kullanılarak lojistik regresyonun olasılık kurallarına göre kayıp verilerin yerine 1 ya da 0 değeri atanır (Wu ve diğerleri, 2015). Tüm çoklu atama yöntemlerinde olduğu gibi bu yöntemde de atama aşaması m kez tekrarlanır. Ardından analiz ve birleştirme aşamaları gerçekleştirilir. Hosmer ve Lemeshow (2000)'a göre, yanıt değişkeninin her iki kategorisinde yeterli yanıt sağlamak için lojistik regresyonda örneklem büyüklüğünün 400'den büyük olması önerilmektedir. Austin ve van Buuren (2022) tarafından yapılan simülasyon çalışmasının sonucuna göre, örneklem büyüklüğü küçük (≤ 1000) ve kayıp veri oranı %90 ya da daha fazla olduğunda, çoklu atama kullanılırken

lojistik regresyon katsayılarının standart sapması ve ortalama tahmin edilen standart hatalar çok büyük olma eğilimi göstermektedir. Lojistik regresyon ile çoklu atama yöntemi; R kütüphanesinde yer alan “mice” paketi ile SAS, PROC MI ve IVEware gibi programlarda yer almaktadır (Wu ve diğerleri, 2015).

İlgili Araştırmalar

Bu bölümde kayıp veriyle başa çıkma yöntemleri ve test eşitleme yöntemleri ile ilgili Türkiye’de ve dünyada yapılmış çalışmalara yer verilmiştir.

Ngudgratoke (2009) tarafından yapılan çalışmada teorik olarak daha güçlü bir yöntem olan tabakalaştırma sonrası eşitleme (post-stratification equating method-PSE) ile eşitleme yanlılıklarını azaltmak amaçlanmış, çapa testi puanı ve çapa testi gerçek puanı gibi tek bir değişken yerine sınava girenler hakkındaki tamamlayıcı bilgiler (çapa testi puanı, alt puanlar ve sınava girenlerin demografik değişkenleri) kullanılmıştır. Hem MT21 adlı çalışmadan alınan gerçek verilerin hem de beş boyutlu olarak simüle edilmiş verilerin kullanıldığı bu çalışmada; test uzunluğu (10’u ortak 40 madde ve 15’i ortak 60 madde), sınava giren gruplar arasındaki yetenek farkları (fark yok, büyük ölçüde fark var) ve kayıp veri ile başa çıkma yöntemleri (eğilim puanı tabakalaştırma-propensity score stratification ve çoklu atama-multiple imputation) gibi farklı koşullar altında, 100 tekrar yapılarak oluşturulan veri setleri ile test eşitleme gerçekleştirilmiştir. Eşitleme deseni olarak denk olmayan gruplarda ortak madde (common item non-equivalent group-CINEG) deseni kullanılmıştır. Gözlemlenen puan eşitleme yöntemlerinden, zincir eşitleme (chain equating-CE) ve tabakalaştırma sonrası eşitleme (PSE) yöntemleri kullanılmıştır. Gerçek verilerden elde edilen demografik değişkenler ve sınava girenlerin madde yanıtları eşleştirilerek simüle edilmiş test puanlarıyla ilişkilendirilmiştir. Bu nedenle her simülasyon veri seti aynı demografik değerlere sahiptir ve örneklem büyüklükleri (Test 1 için 1.361, Test 2 için 1.266) değişmemiştir. Çalışma sonuçları; gruplar yetenekler açısından büyük ölçüde farklılık gösterdiğinde, eşitleme yanlılığını azaltmak için daha fazla tamamlayıcı bilgi kullanmanın

makul olduğunu göstermiştir. Alt puanların veya alt puanların diğer tamamlayıcı bilgilerle bir eğilim puanı biçiminde birleştirilmesinin, uzun testler için eşitleme yanlılığını azaltma potansiyeline sahip olduğu belirlenmiştir. Bununla birlikte demografik değişkenlerin çoklu atama yöntemi için de eşitleme yanlılığını azaltma potansiyeline sahip olduğu gözlenmiştir. Eşitlemenin standart hataları açısından çoklu atama yönteminin, eğilim puanı yönteminden daha küçük değerlere sahip olduğu tespit edilmiştir. Bu araştırmada test uzunluğu, yetenek farkları ve kayıp veri ile başa çıkma yöntemlerinden eğilim puanı tabakalaştırma ile çoklu atama üzerinde durulmuştur. Farklı örneklem büyüklükleri, kayıp verilerin bulunduğu test formu ve kayıp veri oranına ilişkin değişkenler dikkate alınmamıştır. Ayrıca karakteristik eğri dönüştürme yöntemleri ile gerçek puan eşitlemeye yer verilmemiştir.

Shin (2009) “Rasch model tabanlı eşitlemede kayıp verilerle (ihmal edilen yanıtlar) nasıl başa çıkılır?” isimli çalışmasında, ele alınan kayıp veriyle başa çıkma yöntemlerinin (uygulanmamış gibi davranma ve sıfır atama) performanslarını belirlemek için CINEG deseni altında gerçek puan eşitlemede ortak maddeler arasında parametre kaymasına sahip olanların (drift items) doğru tespit edilme düzeylerini karşılaştırmayı amaçlamıştır. Bu amaç için standardize edilmiş İngiliz Dili Sanatları (English Language Arts-ELA) testinin 2005 ve 2006 yıllarındaki uygulamaları ile elde edilen madde parametre tahminlerinden 20 ortak madde ve 30 ortak olmayan madde olmak üzere toplam 50 çoktan seçmeli madde kullanılarak Monte Carlo simülasyon çalışması yürütülmüştür. 20 ortak madde içerisinden seçilen 5 maddenin madde güçlük parametrelerine 1.2 logit eklenerek kayma maddeleri (drift items) oluşturulmuştur. 200, 500, 1000 ve 3000 örneklem büyüklükleri için ortak maddelere verilen yanıtlar üzerinde yüksek, orta ve düşük beceri gruplarına yönelik sırasıyla %7, %10 ve %20 kayıp veriler yaratılmıştır. Sonrasında bu kayıp verilerin yerine sıfır atanarak ya da uygulanmamış gibi davranılarak 48 koşul üretilmiş ve her koşul için 100 tekrar yapılmıştır. Ortak madde tarama istatistiği kullanılarak bir çift ortalama isabet frekansı ve ortalama yanlış frekansı elde edilmiştir. Sonuç olarak kayıp verilere sıfır atanması yerine uygulanmamış gibi davranıldığında kayma maddelerinin daha doğru tespit edildiği

raporlanmıştır. Bu araştırmada farklı madde güçlük parametresi, farklı örneklem büyüklüğü ve farklı kayıp veri oranı altında parametre kaymasına sahip olan maddelere odaklanılmıştır. Kayıp veri ile başa çıkma yöntemlerinden sıfır atama ve uygulanmamış gibi davranma yöntemleri ele alınmıştır. Kayıp verilerin sadece ortak maddelerde meydana geldiği durumlar göz önüne alınmış olup kayıp verilerin test formları içinde meydana geldiği diğer yerlerdeki durumlar dikkate alınmamıştır. Ayrıca hot deck atama ve çoklu atama yöntemlerine yer verilmemiştir.

Kim (2015) çalışmasında, çoklu atama kullanan yeni bir MTK eşzamanlı kalibrasyon yöntemini tanıtmayı ve farklı koşullar altında diğer eşitleme yöntemiyle (beklenti maksimizasyonu) karşılaştırmayı amaçlanmıştır. Araştırmada hem simüle edilmiş hem de gerçek veri setleri kullanılmıştır. İki kategorili puanlanan maddelerden oluşan bu verilerde karşılaştırılabilirlik için CINEG deseni altında eşitleme yapılmıştır. Simüle edilmiş veri setleri; farklı yetenek parametresi dağılımına sahip 500'er kişilik temel ve hedef grup ile farklı madde parametrelerine (a, b ve c), farklı modellere (2PLM, 3PLM), farklı madde sayısına (90 ve 100), farklı ortak madde sayısına (10 ve 20), farklı madde parametre dağılımlarına (LogN, N ve beta) ve rastlantısal kayıp veri mekanizması altında farklı kayıp veri oranına (%40 ve %44.4) göre 100 tekrar yapılarak oluşturulmuştur. Gerçek veri seti için ise 2000 yılında ABD'den alınan Uluslararası Öğrenci Değerlendirme Programı'ndaki (PISA) Kitapçık 1 ve Kitapçık 5'te yer alan okuma testleri kullanılmıştır. 40 maddeden oluşan temel form 434, 39 maddeden oluşan hedef form ise 416 öğrenci tarafından alınmıştır. Formlardaki ortak madde sayısı 16'dır. Gözlenen ve gerçek puan eşitleme yapılmıştır ve performanslar hata kareleri ortalaması (MSE) ile eşitleme yanlılığına (BIAS) göre değerlendirilmiştir. Sonuçta hem çoklu atamanın hem de beklenti maksimizasyonunun, 2PLM için tüm koşullarda benzer sonuçlar gösterdiği görülmüştür. Bununla birlikte 3PLM için ortalama MSE değerlerine göre çoklu atamanın, beklenti maksimizasyonuna göre daha düşük sapmaya sahip olduğu tespit edilmiştir. "Planlanan kayıp verilerle eşitleme: Çoklu atama ile eş zamanlı kalibrasyon" isimli bu çalışmada beklenti maksimizasyonu ile çoklu atama

yöntemleri kullanılarak tamamlanmış kayıp veriler eş zamanlı kalibrasyon ile ortak bir ölçeğe yerleştirilmiştir. Ancak sıfır atama ve hot deck atama ile karakteristik eğri dönüştürme yöntemleri ele alınmamıştır. Ayrıca örneklem büyüklüğü ve kayıp verilerin bulunduğu test formuna ilişkin koşullara yer verilmemiştir.

Shin (2016) tarafından yapılan çalışmada farklı kayıp veriyle başa çıkma yöntemlerinin MTK parametreleri ve dikey ölçeklemeyi etkileyip etkilemediğini araştırmak amaçlanmıştır. Bu amaçla kullanılan ve farklı yaş düzeylerine uygulanan Bilişsel Yetenekler Testi Formu 7 (CogAT 7)'deki kayıp veri oranı %0.93 ile %9.17 arasında, örneklem büyüklüğü ise 5824 ile 9986 arasında değişmektedir. Sekiz farklı kayıp veriyle başa çıkma yönteminin (Liste bazında silme (LW), yanlış olarak puanlama (IN), sunulmadı olarak puanlama (NP), atlanan öğeleri yanlış ve ulaşılamayan öğeleri sunulmadı olarak puanlama (INNR), kısmi bir puan atama (BN), veri büyütme algoritması ile stokastik regresyon kullanan çoklu atama (MISR), zincirleme denklemlerle çoklu atama (MICE) ve hata ile iki yönlü atama altında çoklu atama (MITW)) ele alındığı çalışmada üç yetenek kestirim yöntemi (EAP, MLE, QD) ile hem eşzamanlı hem de ayrı kalibrasyon yöntemi (Stocking-Lord) kullanılarak eşitleme yapılmıştır. Eşitleme 3PL model ile CINEG deseni altında gerçekleştirilmiştir. Dikey ölçek, dereceler arası büyüme, dereceler arası değişkenlik ve etki büyüklüğü dâhil olmak üzere üç özelliğe dayalı olarak değerlendirilmiştir. Kayıp veriyle başa çıkma yöntemlerinin madde parametre tahminleri üzerindeki etkisi ise; madde ayırt ediciliği, madde güçlüğü ve şans parametresi için özet istatistikler karşılaştırılarak da incelenmiştir. Analiz sonuçlarında genel olarak; IN yönteminin daha yüksek madde ayırt ediciliği ve madde güçlüğü parametre tahminleri ürettiği, ancak diğer yöntemlere kıyasla daha düşük şans parametre tahminleri verdiği görülmüştür. Ayrıca IN yönteminin daha yüksek ortalama teta tahminleri ve daha büyük büyüme üretirken, MITW daha küçük teta tahminleri ve büyüme sağlamıştır. MICE ve MISR yöntemlerinin, INNR ve NP yöntemlerine benzer şekilde performans gösterme eğiliminde oldukları tespit edilmiştir. Kayıp veriyle başa çıkma yönteminin seçimi; eşzamanlı kalibrasyondan ziyade ayrı kalibrasyon ile ve EAP ya da

QD'den ziyade MLE ile sonuçlar üzerinde daha büyük bir etkiye sahip olduğu belirlenmiştir. Araştırmada sekiz farklı kayıp veriyle başa çıkma yönteminin MTK parametreleri ve dikey ölçekleme üzerindeki etkilerine odaklanılmış olup yatay ölçekleme ele alınmamıştır. Ayrıca hot deck atama ve çoklu atama (PMM ve LOGREG) yöntemleri ile kayıp verilerin bulunduğu test formuna ilişkin koşul kullanılmamıştır.

Ertoprak (2017)'in "Kayıp Verinin Test Eşitlemeye Etkisi" isimli çalışmasında, simülatif olarak 3PL modele uygun olarak üretilen ve ikili puanlanan maddelerden oluşan iki test formundan, farklı kayıp veri mekanizmaları (TRK, RK, ROK,) altında, kayıp verilerin bulunduğu farklı yerlerde (eşitlenecek test-ET, ortak test-OT, her iki test-HİT), farklı oranlarda (%10, %20, %30) kayıp verili veri setleri oluşturulmuştur. Bu veri setlerindeki veri sorunu, kayıp veriyle farklı başa çıkma yöntemleri (uygulanmamış gibi davranma-UGD, yanlış yanıtlanmış gibi davranma-YYGD, lojistik regresyona dayalı çoklu değer atama-LRÇDA, diskriminant fonksiyonuna dayalı çoklu değer atama-DFÇDA) ile çözülmüş ve ortak madde deseni kapsamında, MTK'ye dayalı Stocking-Lord yöntemi kullanılarak eşitlenmiştir. Araştırma sonuçları RMSE ve BIAS değerlerine göre karşılaştırılmış olup HİT kayıp veri yeri, TRK kayıp veri mekanizması, %10 kayıp veri oranı, özellikle yetenek parametresi için DFÇDA yönteminde en az hatalı ve en az yanlı değerlere ulaşıldığı görülmüştür. Ayrıca kayıp veri ile başa çıkma yöntemi olarak sıklıkla kullanılan UGD ve YYGD'nin, test eşitlemede yanlı ve hatalı kestirimler yaptıkları belirlenmiştir. Bu araştırmada simülasyon verisi üretilmiş ve farklı koşullar altında kayıp veriye sahip veri setleri oluşturulmuştur. Ancak kayıp veriyle başa çıkma yöntemlerinden hot deck atama ve çoklu atama yöntemlerinden PMM'e yer verilmemiştir. Ayrıca sabit örneklem büyüklüğü baz alınmış olup farklı örneklem büyüklükleri kullanılmamıştır.

Sonuç olarak alanyazında yapılan çalışmalar değerlendirildiğinde hem kayıp veri hem de test eşitleme kavramlarını birlikte ele alan çalışma sayısının çok sınırlı olduğu görülmüştür. Yurt dışında yapılan çalışmalarda ya gerçek veri seti ya da gerçek veri seti ile birlikte simülatif olarak üretilmiş veri setinin kullanıldığı ancak Türkiye'de yapılan çalışmada

sadece simülasyon verisinin kullanıldığı görülmüştür. Söz konusu çalışmalarda kayıp veriyle başa çıkma yöntemlerinden; yanlış yanıtlanmış gibi davranma, uygulanmamış gibi davranma ve çoklu atama yöntemlerinin sıklıkla kullanıldığı belirlenmiştir. Ancak hot deck atama ve çoklu atama yöntemlerinden PMM'in test eşitleme üzerindeki etkisine ilişkin herhangi bir çalışmaya rastlanmamıştır. Ayrıca örneklem büyüklüğü, kayıp veri oranı ve kayıp verilerin bulunduğu test formuna ilişkin koşulların birlikte ele alındığı çalışmaya ulaşılamamıştır.



Bölüm 3

Yöntem

Bu bölümde; araştırmanın türü, araştırmanın evreni ve örnekleme, veri toplama aracı, eşitleme deseni, araştırmada ele alınan koşullar, veri setlerinin oluşturulması ve verilerin analizi ile ilgili kısımlara yer verilmiştir.

Araştırmanın Türü

Araştırmada kayıp veriyle başa çıkma yöntemlerinin, farklı koşullar altında, Madde Tepki Kuramı'na (MTK) dayalı test eşitleme yöntemlerinin performansı üzerindeki etkileri incelenmiştir. Bu nedenle araştırmada, farklı koşulların eşitleme yöntemlerinin performansı ile ilişkisinin ortaya çıkarılmasını sağlamak için ilişkisel araştırma türü kullanılmıştır. İlişkiler ve bağlantıların incelendiği ilişkisel araştırmalarda araştırmacının tahminlerde bulunmasına olanak sağlanır (Büyüköztürk ve diğerleri, 2011).

Araştırmanın Evreni ve Örnekleme

Araştırmada Uluslararası Matematik ve Fen Eğitimi Araştırması (TIMSS – Trend in International Mathematics and Science Study) 2019 uygulamasına katılan ve örgün eğitim sürecinin sekizinci sınıf düzeyinde bulunan öğrenci verileri kullanılmıştır. TIMSS 2019 uygulamasına sekizinci sınıf düzeyinde 250.000 öğrenci katılmıştır. TIMSS 2019'da 22 ülke bilgisayar tabanlı (eTIMSS), 14 ülke ise kalem kağıt tabanlı (paperTIMSS) uygulama yapmıştır (Martin ve diğerleri, 2020).

Uygulamaya katılan ülkelerin öğrenci sayıları incelenmiş ve eşitleme sürecinde örneklem büyüklüğünün de etkisi göz önüne alınarak TIMSS 2019 fen alanında, eTIMSS uygulaması yapan bütün ülkeler araştırmanın örnekleme olarak belirlenmiştir. MTK'ye dayalı eşitleme yöntemlerinin kullanılacağı test formlarını belirlemede ise kriter olarak; en çok sayıda, iki kategorili puanlanan madde, ortak madde ve yanıtlayıcının bulunma durumları göz önüne alınmıştır. Yapılan inceleme sonucunda 6 ve 7 numaralı kitapçıkların belirlenen kriterlere uygun olduğu tespit edilmiştir.

Ülkelere ve test formlarını alan öğrenci sayılarına ilişkin detaylı bilgilere Tablo 2’de yer verilmiştir.

Tablo 2

Ülkelere ve Test Formlarını Alan Öğrenci Sayılarına İlişkin Bilgiler

Ülke	Fen Başarı Sırası*	Kayıp Veri İçeren Test Formları		Tam Veriye Sahip Test Formları	
		Kitapçık 6	Kitapçık 7	Kitapçık 6	Kitapçık 7
Singapur	1	340	345	297	301
Tayvan	2	347	352	244	275
Güney Kore	4	271	277	194	207
Rusya	5	279	277	143	151
Finlandiya	6	334	347	213	246
Litvanya	7	279	267	157	133
Macaristan	8	322	328	188	200
Amerika Birleşik Devletleri	11	633	623	505	493
İsveç	12	272	276	120	129
Portekiz	13	234	234	127	114
İngiltere	14	240	238	130	138
Türkiye	15	292	291	166	188
İsrail	16	269	265	170	157
Hong Kong	17	236	231	135	135
İtalya	18	265	253	135	141
Norveç	20	315	327	149	136
Fransa	21	280	279	125	143
Katar	25	276	277	186	182
Birleşik Arap Emirlikleri	26	1590	1585	980	1015
Şili	28	290	288	122	125
Malezya	29	511	500	426	420
Gürcistan	33	234	235	66	80
Toplam		8109	8095	4978	5109

* TIMSS 2019 fen başarı dağılımında eTIMSS uygulaması yapan tüm ülkeler seçilmiştir. Diğer ülkeler kağıt-kalem tabanlı uygulama (paperTIMSS) yaptıkları için sıralamaya dahil edilmemiştir.

Tablo 2’de yer alan verilere göre; fen testinde Kitapçık 6’yı alan 8109, Kitapçık 7’yi alan 8095 öğrenci araştırmanın evrenini oluşturmaktadır. Söz konusu test formları (Kitapçık 6 ve Kitapçık 7) kayıp veri içermektedir. Ancak araştırmanın amacına göre tam veriye sahip test formlarına ihtiyaç duyulmaktadır. Dolayısıyla araştırmada kullanılacak test formlarının tam veriye sahip olması örnekleme için bir ölçüttür. Bu ölçüte göre kayıp veri içeren test formlarında yer alan kayıp veriler liste bazında silme yöntemi kullanılarak silinmiştir. Örneklemin belirlenen ölçütü karşılayan kişi, olay, durum ya da nesnelerden oluşturulmasına ölçüt örnekleme denir (Büyüköztürk ve diğerleri, 2011). Sonuç olarak

Kitapçık 6'yı alan ve tam veriye sahip 4978, Kitapçık 7'yi alan ve tam veriye sahip 5109 öğrenci araştırmannın örneklemini oluşturmaktadır. Tam veriye sahip test formlarını alan toplam öğrenci sayılarının birbirine yakın olduğu görölmektedir.

Örneklem, referans grup olarak belirlenmiş olup bu gruptan araştırmannın amacına uygun farklı koşullara sahip veri setleri oluşturulmuştur. Veri setlerinin oluşturulmasına ilişkin ayrıntılı bilgiler “Veri Setlerinin Oluşturulması” başlığı altında yer almaktadır.

Veri Toplama Aracı

Bu araştırmada Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu (IEA- International Association for the Evaluation of Educational Assessment) tarafından gerçekleştirilen TIMSS 2019 uygulamasına ait veriler kullanılmıştır. Verilere IEA' nın resmî internet sayfasında yer alan (<https://timss2019.org/international-database/>) veri tabanından ulaşılmıştır.

TIMSS

Dördüncü ve sekizinci sınıflarda dört yıl aralıklarla yürütölen ve katılımcı ölkelere öğrencilerinin matematik ve fen başarıları hakkında eğitim politikası ve uygulamalarının iyileştirilmesine yönelik kanıta dayalı bilgiler sağlayan bir tarama araştırmasıdır. IEA tarafından 1995 yılından bu yana dört yılda bir uygulanan ve TIMSS değerlendirme serisinin yedincisi olan TIMSS 2019 uygulamasına katılan ölkelerin yarısı ilk kez bilgisayar tabanlı uygulamaya (eTIMSS) geçiş yaparken diğör yarısı kağıt-kalem (paperTIMSS) tabanlı uygulamaya devam etmiştir. TIMSS 2019'da, her sınıf düzeyinde 14'ü matematik ve 14'ü fen maddelerinden oluşın toplam 28 blok bulunmaktadır. Bu blokların çeşitli kombinasyonları bir araya getirilerek 14 öğrenci kitapçığı oluşturulur. Her öğrenci kitapçığı, iki blok matematik ve iki blok fen olmak üzere toplam dört bloktan oluşur. Kitapçıklar arasındaki bağlantıyı etkinleştirmek için her blok iki kitapçıkta da yer alır (Mullis & Martin, 2017; Mullis ve diğörleri, 2020).

Araştırma, TIMSS 2019 uygulaması kapsamında örgün eğitim sürecinin sekizinci sınıf düzeyinde bulunan öğrencilere uygulanan fen test verileri üzerinden yürütülmüştür. Veri setinde eTIMSS uygulaması yapan 22 ülkeye ait 14 kitapçık incelenmiştir. Ardından hangi maddelerin hangi kitapçıkta yer aldığını belirlemek için frekans analizi yapılmıştır. Son olarak uygulanmamış ve kısmî puanlanan (partial credit) maddeler çıkarıldığında en fazla; iki kategorili puanlanan madde, ortak madde ve yanıtlayıcı sayısına 6 ve 7 numaralı kitapçıkların sahip olduğu görülmüştür. Bu nedenle bu araştırmada 6 ve 7 numaralı kitapçıklardan seçilen maddeleri yanıtlayan öğrenci verileri kullanılmıştır. Kitapçık 6'nın "ET19DCS06" ve "ET19DCS07", kitapçık 7'nin ise "ET19DCS07" ve "ET19DCS08", kodlu fen bloklarından oluştuğu görülmüştür. Kitapçıklardan seçilen maddelere ait özellikler Tablo 3'te verilmiştir.

Tablo 3

Kitapçıklardaki Maddelere Ait Özellikler

Madde Türü	Kitapçık Numarası	Madde Adı	Madde Etiketi
Ortak Maddeler	Kitapçık 6 ve Kitapçık 7	SE52090A	Influenza: Cause
		SE52090B	Influenza: How is it spread
		SE52262	Biceps and triceps
		SE52267	Acquired characteristics
		SE52051	Is element metal or nonmetal
		SE52026	Example of chemical change
		SE52130	Gas inside ping-pong ball
		SE52028	Electromagnet attracting needles
		SE52217	Water pressure on fish
		SE52038	Energy for water cycle
Ortak Olmayan Maddeler	Kitapçık 6	SE52099	Trees protect soil from erosion
		SE62090	Water cycle in forest ecosystem
		SE62284	Hair color of young rabbits
		SE62032	Hot metal ball on balance
		SE62043	Electromagnet and paper clips
		SE62158	Graphs of musical notes
		SE62159	Box pulled by three forces
		SE62005	Syringes with solid and air
		SE62075	Hugo's chemical reaction
		SE62004	Block pounded into flat sheet
	Kitapçık 7	SE62175	Power plant geographic factor
		SE62173A	Temperature and geography
		SE62173B	Climate and geography
		SE72070	Symbiosis between crocodiles and birds
		SE72462	Largest transfer of energy In energy pyramid
		SE72443	Human DNA
		SE72145	John's mixture separation steps
		SE72100	Number of hydrogen atoms in Diagram A
		SE72137	How carbon dioxide extinguishes a fire

SE72298	Information Kerry needs to calculate her average speed
SE72260	Components julia must use to build a circuit
SE72265	Statements about magnets
SE72347	Area that would most benefit from desalination plant
SE72351	Why oil, gas, and coal are nonrenewable resources
SE72367	Main cause of changing tides

Tablo 3 incelendiğinde, her iki kitapçıkta da 11'i ortak ve 12'si ortak olmayan madde olmak üzere toplam 23 madde bulunduğu görülmektedir. Söz konusu maddeler sadece çoktan seçmeli (multiple choice) ve yanıtı doğru ya da yanlış olan yapılandırılmış yanıt (constructed response) oluşmaktadır. Elde edilen tam veri seti, bu maddelere verilen doğru yanıtlar için "1", yanlış yanıtlar için "0" değeri kodlanarak iki kategorili puanlanan 1-0 madde biçimine dönüştürülmüştür.

Eşitleme Deseni

Araştırmada kullanılan tam veri seti; 11'i ortak, 12'si ortak olmayan madde içeren iki farklı test formu (Kitapçık 6 ve Kitapçık 7) ve bu test formlarını alan iki farklı yanıtlayıcı grubundan oluşmuştur. Dolayısıyla Kitapçık 6 ve Kitapçık 7'den elde edilen puanlar, her iki kitapçıkta da ortak olan maddeler kullanılarak eşitlenebilmektedir. Veriler TIMSS 2019 uygulamasında bu şekilde toplanarak desenlendiği için araştırmanın eşitleme deseni olarak, denk olmayan gruplarda ortak madde (common item non-equivalent group-CINEG) deseni kullanılmıştır. Araştırmada kullanılan eşitleme deseni Tablo 4'te gösterilmiştir.

Tablo 4

Araştırmada Kullanılan Eşitleme Deseni

Çalışma Grubu	Form X (Yeni/Eşitlenecek Test Formu)		
	Form X'te Ortak Olmayan Maddeler	Ortak Maddeler	Form Y'de Ortak Olmayan Maddeler
Grup 1 (4978 yanıtlayıcı)	✓ (12 Madde)	✓ (11 Madde)	
Grup 2 (5109 yanıtlayıcı)		✓ (11 Madde)	✓ (12 Madde)
Form Y (Eski Test Formu)			

Bu araştırmada Kitapçık 6 yeni/eşitlenecek test formu (Form X), Kitapçık 7 ise eski test formu (Form Y) olarak belirlenmiştir. Grup 1 Form X'i, Grup 2 ise Form Y'yi alan yanıtlayıcı grubunu temsil etmektedir. Eşitleme süreci yeni test formundan elde edilen puanların eski test formundan elde edilen puanlara eşitlenmesi olarak planlanmıştır.

Araştırmada Ele Alınan Koşullar

Araştırmada ele alınan koşullara Tablo 5'te yer verilmiştir.

Tablo 5

Araştırmada Ele Alınan Koşullar

Faktörler	Koşullar	Koşul Sayısı
Örneklem büyüklüğü	750, 1500	2
Kayıp veri oranı	%10, %20, %30	3
Kayıp verilerin bulunduğu test formu	ET, HİT*	2
Kayıp veri ile başa çıkma yöntemleri	SA, HDA, PMM, LOGREG**	4
Toplam koşul sayısı		48+1***=49

* ET: Eşitlenecek test formu, HİT: Her iki test formu

** SA: sıfır atama, HDA: Hot deck atama, PMM: Tahmine dayalı ortalama eşleştirme ile çoklu atama, LOGREG: Lojistik regresyon ile çoklu atama

*** eTIMSS 2019 uygulamasından elde edilen tam veri seti olup araştırmanın referans koşulunu (R) temsil etmektedir. Bu nedenle toplam koşul sayısına +1 olarak eklenmiştir.

Tablo 5'e göre bu çalışmada; örneklem büyüklüğü (2 koşul), kayıp veri oranı (3 koşul), kayıp verilerin bulunduğu test formu (2 koşul) ve kayıp veriyle başa çıkma yöntemleri (4 koşul) olmak üzere toplam 48 koşul (2x3x2x4) incelenmiştir. Çalışmanın başında elde edilen tam veri seti diğer koşullar ile karşılaştırılmak üzere referans koşul (R) olarak ele alınmıştır. Ayrıca araştırma sonuçlarının tutarlı ve genellenebilir olması için her veri seti için yeniden örnekleme (resampling) türlerinden bootstrap yöntemi kullanılmıştır. Çoğu araştırmada ilgilenilen evrenden bir örneklem alınır ve analizler bu örneklem üzerinden

yürütülür. Bootstrap yönteminde ise analiz aşamasına ek bir alt örnekleme (sub-sampling) ve tekrar (replication) adımı eklenir. Bu süreç, orijinal örneklem üzerinden yerine koyma ile rastgele örnekleme yapılarak binlerce yeni alt örneklem oluşturulması ile başlar. Daha sonra ilgili analiz bu yeni alt örneklemelerin her birinde tekrarlanır (Banjanovic ve diğerleri, 2016; Doğan, 2019). Benzer araştırmalarda her veri seti üzerinde 50 (Ertoprak, 2017) ya da 100 tekrarın (Kim, 2015; Ngudgratoke, 2009) yapıldığı görülmüştür. Gerçek veri setinin kullanıldığı bu çalışmada ise sınırlı bir evren içerisinde tekrar tekrar seçilecek örneklem sayısı göz önüne alındığında 50 tekrarın kararlılığa ulaşmada yeterli olduğu düşünülmüştür. Bu nedenle araştırmada 50 tekrar yapılarak toplam 2400 (48x50) veri seti üretilmiştir.

Araştırma kapsamında ele alınan koşullardan biri örneklem büyüklüğüdür. Ulaşılan literatürde test eşitlemede kayıp veri ile başka çıkma yöntemleriyle ilgili yapılan çalışmalarda farklı örneklem büyüklüklerindeki etkilerin incelenmediği belirlenmiştir (Ngudgratoke, 2009; Shin, 2009; Kim, 2015; Shin, 2016; Ertoprak, 2017). Test eşitleme ile ilgili çalışmalarda ise ele alınan örneklem büyüklüğünün 250 ile 10.000 arasında değiştiği görülmüştür (Atalay Kabasakal, 2014; Çokluk Bökeoğlu ve diğerleri, 2022; Lee & Ban, 2010; Speron, 2009). Test eşitleme sürecinde en az 500 kişilik örneklem büyüklüğü ile çalışılması önerilmiş olup (Spence, 1996) MTK'ye dayalı test eşitleme sürecinde 3 PLM için CINEG deseni altında 1500 kişilik örneklem büyüklüğüne ihtiyaç duyulduğu belirtilmiştir (Kolen & Brennan, 2014). Literatürde yer alan çalışmaların bulguları ve MTK varsayımları göz önüne alındığında bu araştırmada her bir test formu için 750 ve 1500 cevaplayıcı olmak üzere iki farklı örneklem büyüklüğü ele alınmıştır.

Araştırmada ele alınan koşullardan diğeri kayıp veri oranıdır. Tamamen rastgele kayıp veri mekanizmasının kullanıldığı kayıp veri ile ilgili çalışmalarda, kayıp veri oranlarının farklılık gösterdiği görülmüştür. Söz konusu farklılığa rağmen araştırmalarda sıklıkla seçilen kayıp veri oranlarının %10 ve %20 olduğu tespit edilmiştir (Ak, 2020; Bayhan, 2018; Bayram, 2020; Ertoprak, 2017; Işıkoğlu, 2017; Karaman, 2022; Ma ve diğerleri, 2011; Öztemür, 2014; Şahin Kürşad, 2014; Tamcı, 2018). Test eşitlemede kayıp veri ile başka çıkma yöntemleriyle ilgili yapılan çalışmalarda ise kayıp veri oranlarının genellikle %1 ile

%45 arasında değiştiği belirlenmiştir (Kim, 2015; Shin, 2016; Ertoprak, 2017). Ayrıca bu araştırmada kullanılan gerçek veri seti için Tablo 1’de verilen kayıp veriye sahip test formları ile tam veriye sahip test formları arasındaki kayıp veri oranları hesaplanmış ve bu oranların %13 ile %72 arasında değiştiği görülmüştür. Kayıp veri oranının %5 ve daha az olduğu durumlarda ihmal edilebileceği (Tabachnick & Fidell, 2013), %40 olduğu durumda ise kestirimlerin tam veri setine ait kestirimlerden uzaklaşacağı ve yanlış sonuçlar elde edilebileceği belirtilmiştir (Ertoprak, 2017). İncelenen tüm çalışmalar göz önüne alınarak araştırmada %10, %20 ve %30 olmak üzere üç farklı kayıp veri oranı seçilmiştir. %10 düşük, %20 orta ve %30 yüksek kayıp veri oranı olarak ele alınmıştır.

Araştırma kapsamında ele alınan bir diğer koşul ise kayıp verilerin bulunduğu test formudur. Ulaşılan literatürde, test eşitlemede kayıp veri ile başka çıkma yöntemleriyle ilgili Ertoprak (2017) tarafından çalışmada kayıp verilerin bulunduğu test formunun etkisi incelenmiştir. Kayıp verilerin bulunduğu test formları; eşitlenecek test (ET), her iki test (HİT) ve ortak test (OT) olarak ele alınmıştır. Çalışma sonucunda, yeni araştırmalarda kayıp veri içerme bakımından OT’nin süreç dışında bırakabileceği önerilmiştir. Bu araştırmada denk olmayan gruplarda ortak madde deseni kullanıldığı için ortak maddeler aracılığıyla ölçek dönüştürme işlemleri gerçekleştirilmektedir. Bu durum da göz önüne alındığında kayıp verilerin bulunduğu test formunun test eşitleme yöntemlerinin performansı üzerindeki etkisinin incelenebilmesi için kayıp veriler, eşitlenecek test (ET – Kitapçık 6) ve her iki test (HİT – Kitapçık 6 ve Kitapçık 7) formu içinde olmak üzere iki farklı şekilde oluşturulmuştur.

Araştırmada; en sık başvuru alan kayıp veri ile başa çıkma yöntemlerinden sıfır atama (SA) yöntemi, büyük ölçekli sınavlarda kullanım kolaylığı sağlayan hot deck atama (HDA) yöntemi ve son yıllarda sıklıkla önerilen en popüler atama yöntemlerinden çoklu atama yöntemleri (PMM ve LOGREG) seçilmiştir.

Veri Setlerinin Oluşturulması

Veri setleri, araştırmada ele alınan koşullara göre üç aşamada oluşturulmuştur:

Birinci aşamada, örneklem büyüklüğü koşulunu karşılayan veri setlerini oluşturmak için; referans gruptan, R kütüphanesinde yer alan “sampling” paketi (Matei & Tillé, 2021) kullanılarak, 750 ve 1500 kişilik örneklemeler rastgele seçilmiştir. Böylece araştırmada kullanılacak iki farklı örneklem büyüklüğüne sahip tam veri setleri elde edilmiştir.

İkinci aşamada, örneklem büyüklüğü $\times$ kayıp veri oranı $\times$ kayıp verilerin bulunduğu test formu koşullarını karşılayan veri setlerini oluşturmak için; TRK veri mekanizması altında, R kütüphanesinde yer alan “missMethods” paketi (Rockel, 2022) kullanılarak tam veri setleri içerisinde hangi verilerin silineceği tamamen rastgele belirlenmiştir. Böylece kayıp verilerin, araştırma kapsamında ele alınan tüm değişken değerlerinden bağımsız olması sağlanmıştır. 750 ve 1500 örneklem büyüklüğüne sahip tam veri setlerinden, ET ve HIT’te %10, %20 ve %30 oranında kayıp veri içeren 12 farklı veri seti elde edilmiştir.

Üçüncü aşamada ise kayıp veri içeren veri setlerine, kayıp veri ile başa çıkma yöntemleri (SA, HDA, PMM, LOGREG) uygulanarak kayıp veri problemi çözülmüş 48 farklı veri seti üretilmiştir. R kütüphanesinde yer alan; SA yöntemi için “base” paketi (R Core Team ve contributors worldwide), HDA yöntemi için “VIM” paketi (Templ ve diğerleri, 2022), ÇA yöntemlerinden PMM ve LOGREG için ise “mice” paketi (van Buuren, 2022) kullanılmıştır. ÇA yöntemleri için R programlama dilinde çoklu atama sayısının varsayılan değeri olan $m=5$ tercih edilmiştir.

Her bir koşul için araştırmacı tarafından yazılan R kodları ile 50 tekrar yapılarak 2400 (48x50) kayıp veri problemi çözülmüş tam veri seti oluşturulmuştur.

Verilerin Analizi

Verileri analiz etmeden önce referans grup olarak ele alınan tam veri setine sahip Kitapçık 6 ve Kitapçık 7’ye ilişkin betimsel istatistikler hesaplanmış ve MTK’nin varsayımları test edilmiştir. Ardından madde ve yetenek parametreleri kestirilmiştir. Bu parametreler MTK’ye dayalı ölçek dönüştürme yöntemleri kullanılarak ortak bir ölçeğe yerleştirilmiş ve eşitleme işlemi gerçekleştirilmiştir. Eşitleme sonuçlarının karşılaştırılması için hata kareleri ortalamasının karekökü (RMSE) ve eşitleme yanlılığı (BIAS) her koşul ve tekrar için ayrı

ayrı hesaplanmıştır. Karşılaştırmalar hem kayıp veriye sahip test formları arasında hem de kayıp veriye sahip test formları ile tam veriye sahip test formları (referans grup) arasında yapılmıştır.

1. Betimsel İstatistikler

Kitapçıklara ilişkin betimsel istatistikler hesaplanmış ve Tablo 6'da verilmiştir.

Tablo 6

Betimsel İstatistikler

	Kitapçık 6			Kitapçık 7		
	Ortak Olmayan Maddeler	Ortak Maddeler	Tüm Maddeler	Ortak Olmayan Maddeler	Ortak Maddeler	Tüm Maddeler
Yanıtlayıcı Sayısı	4978	4978	4978	5109	5109	5109
Madde Sayısı	12	11	23	12	11	23
Arimetik Ortalama ($\bar{X}$)	6.19	5.61	11.8	6.22	5.57	11.79
Ortanca (Medyan)	6	6	12	6	6	12
Tepe Değer (Mod)	6	5	10	6	6	12
Standart Sapma	2.57	2.33	4.41	2.70	2.29	4.48
Varyans	6.63	5.45	19.42	7.30	5.23	20.09
Minimum Değer	0	0	1	0	0	0
Maksimum Değer	12	11	23	12	11	23
Çarpıklık	0.01	0.02	0.07	-0.04	0.03	0
Basıklık	-0.73	-0.72	-0.76	-0.75	-0.59	-0.7

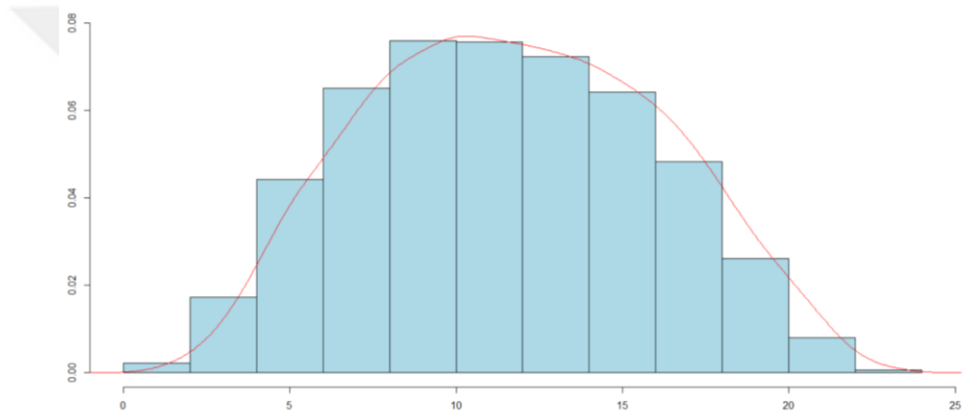
Tablo 6'ya göre; kitapçıkları alan yanıtlayıcı sayıları birbirine yakındır. Her iki kitapçıkta da ortak, ortak olmayan ve tüm madde sayıları birbirine eşittir. Merkezi eğilim ölçüleri incelendiğinde aritmetik ortalama, ortanca ve tepe değerlerin her iki kitapçıkta da birbirine yakın olduğu görülmektedir. Dağılımların standart sapmaları, varyans değerleri, basıklık ve çarpıklık katsayıları da benzerdir. Tabloya göre, çarpıklık ve basıklık katsayılarının -1 ile +1 aralığında değişmektedir. Tabachnick ve Fidell'e (2013) göre -1.5 ile

+1.5 arasındaki bu değerler kabul edilebilir aralıktadır. Bu durumda puanların normalden aşırı bir sapma göstermediği söylenebilir.

Genel olarak değerlendirildiğinde; merkezi eğilim ölçülerinin birbirine yakın olması, basıklık ve çarpıklık katsayılarının -1 ile +1 değerleri arasında yer alması ve puan dağılımlarını gösteren histogram grafiklerinin (Şekil 5, Şekil 6) incelenmesi sonucunda her iki kitapçık için de puanların normal dağılım gösterdiği söylenebilir.

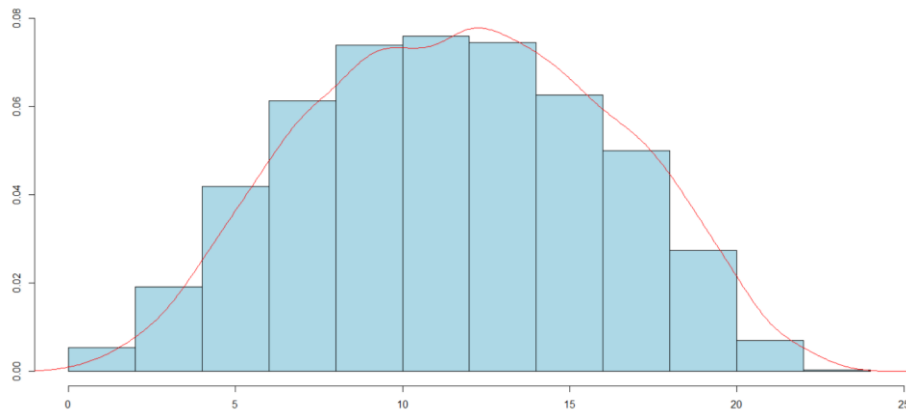
Şekil 5

Kitapçık 6 İçin Puan Dağılımları



Şekil 6

Kitapçık 7 İçin Puan Dağılımları



Kitapçıkların güvenirliklerine ilişkin değerlere Tablo 7'de yer verilmiştir.

Tablo 7**Kitapçıklara İlişkin Güvenirlik Katsayıları**

Güvenirlik Katsayısı	Kitapçık 6	Kitapçık 7
KR-20	0.807	0.814

Tablo 7 incelendiğinde, kitapçıkların güvenirlik katsayılarının birbirine oldukça yakın olduğu görülmektedir. Güvenirlik katsayıları bir iç tutarlılık indeksi olarak yorumlanabilir. Buna göre güvenirlik ölçütünün aralığı; 0.50'den küçükse güvenirlik kötü, 0.50 ile 0.80 arasındaysa güvenirlik orta düzeyde, 0.80'den büyükse güvenirlik iyidir (Salvucci ve diğerleri, 1997). Dolayısıyla her iki kitapçığın da iç tutarlılığının yeterli olduğu söylenebilir.

2. MTK Varsayımlarının Test Edilmesi

Bu aşamada tek boyutluluk, yerel bağımsızlık, model veri uyumu, madde ve yetenek parametre değişmezliği test edilmiş olup madde parametreleri verilmiştir.

2.1. Tek Boyutluluk. Tek boyutluluk, ölçme aracının sadece bir yapıyı ya da özelliği ölçmesidir. Başka bir ifade ile tek boyutluluk, psikolojik özelliklerin ölçüldüğü durumlarda baskın bir boyutun olmasıdır. Ölçme aracında yer alan maddeler için baskın tek bir faktör var ise tek boyutluluk varsayımının karşılandığı söylenebilir. Bu faktörün ölçme aracı ile ölçülmek istenen yetenek olduğu varsayılmaktadır (Hambleton & Swaminathan, 1985; Crocker & Algina, 1986). Araştırmada kullanılan Kitapçık 6 ve Kitapçık 7'de yer alan maddeler 1-0 olarak puanlandıkları için tek boyutluluk varsayımında ilk olarak tetrakorik korelasyon matrisine dayalı açıklayıcı faktör analizi kullanılmıştır. Tetrakorik korelasyon ile normal dağılıma sahip ve yapay süreksiz hale getirilmiş iki kategorili değişkenler arasındaki ilişkileri bulmak amaçlanır (Kirk, 1973).

Faktör analizinden önce veri yapısının analize uygunluğunu belirlemek için Kaiser-Meyer-Olkin (KMO) ve Barlett Küresellik testi yapılmış ve sonuçlara Tablo 8'de yer verilmiştir. Yapının faktör analizine uygun olması için KMO değerinin 0.60 ve bu değerden

büyük olması, Bartlett testinin ise anlamlı ($p<.05$) olması gerekmektedir (Tabachnick & Fidell, 2013).

Tablo 8

KMO ve Bartlett Test Sonuçları

		Kitapçık 6	Kitapçık 7
Kaiser-Meyer-Olkin (KMO) Testi		0.90	0.91
	χ^2	8982.41	9641.67
Bartlett Küresellik Testi	sd	253	253
	p	0.00*	0.00*

* $p<0.05$

Tablo 8 incelendiğinde KMO değerlerinin Kitapçık 6 için 0.90, Kitapçık 7 için 0.91 olduğu görülmektedir. Kaiser'a (1974) göre 0.90'ın üzerindeki KMO değerlerinin örnekleme yeterliliği mükemmeldir. Bartlett küresellik testi sonuçları incelendiğinde ki-kare (χ^2) değerinin Kitapçık 6 ($\chi^2_{(253)} = 8982.41$; $p<.05$) ve Kitapçık 7 ($\chi^2_{(253)} = 9641.67$; $p<.05$) için anlamlı olduğu belirlenmiştir. Bu sonuçlardan hareketle veri yapısının faktörleşmeye uygun olduğu görülmektedir.

R kütüphanesindeki "psych" paketinde (Revelle, 2023) yer alan 'fa' fonksiyonu ile faktör analizi gerçekleştirilmiş ve özdeğerler hesaplanarak sonuçlar Tablo 9'da verilmiştir.

Tablo 9

Faktör Analizi Sonuçları

Faktör	Kitapçık 6 Özdeğer	Kitapçık 7 Özdeğer
1	5.387	5.650
2	1.312	1.202
3	1.064	1.080
4	0.927	0.984

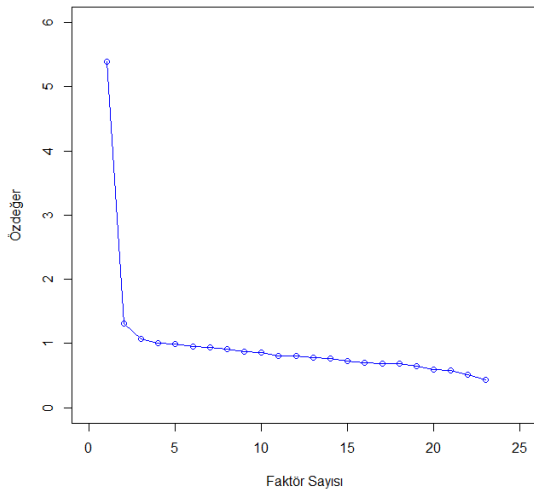
Tablo 9'da ilk dört faktöre ait özdeğerler görülmektedir. Faktör sayısına karar verebilmek için çok sayıda yöntem geliştirilmiştir. En yaygın olarak kullanılan yöntemlerden

Kaiser kriteri ve yamaç birikinti grafiğidir. Kaiser (1960) tarafından geliştirilen yöntemine göre özdeğerin 1'den büyük olması faktör sayısı için kriterdir. Bu kritere göre Kitapçık 6 ve Kitapçık 7 değerlendirildiğinde faktör sayısının her iki kitapçık için de 3 (üç) olduğu söylenebilir. Ancak birinci ve ikinci özdeğerin oranının dörtten büyük olması (Lord, 1980) ya da birinci faktörün özdeğerinin ikinci faktörün özdeğerine oranının çok daha yüksek olması (Tabachnick & Fidell, 2013) tek boyutluluğun kanıtları olarak yorumlanmıştır. Kitapçık 6 için birinci faktörün özdeğeri ikinci faktörün özdeğerinin dört katına, Kitapçık 7 için birinci faktörün özdeğeri ikinci faktörün özdeğerinin yaklaşık beş katına eşittir.

Cattell (1966) tarafından geliştirilen yamaç birikinti grafiği yöntemine göre Y ekseninden (özdeğer) X eksenine (faktör sayısı) doğru eğimli bir düşüş olur ancak bir noktadan sonra eğim iyice azalır ve grafik düz devam eder. Bu nokta kırılma noktasıdır ve kırılma noktasına kadar olan sayı faktör sayısıdır. Kitapçıklara ilişkin hesaplanan özdeğerlere göre çizilen yamaç birikinti grafikleri sırasıyla Şekil 7 ve Şekil 8'de yer almaktadır.

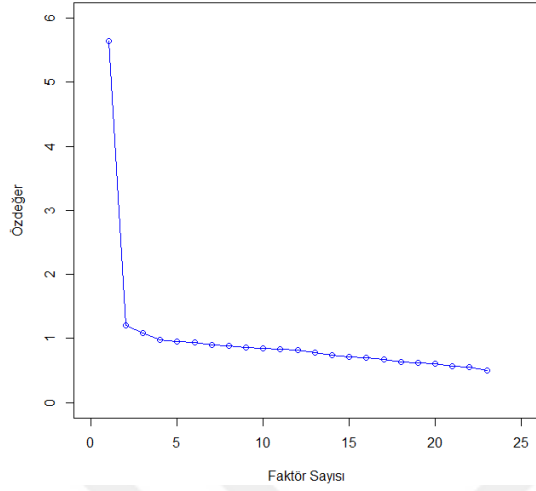
Şekil 7

Yamaç Birikinti Grafiği (Kitapçık 6)



Şekil 8

Yamaç Birikinti Grafiği (Kitapçık 7)



Şekil 7 ve Şekil 8 incelendiğinde, her iki kitapçıkta da birinci faktörden ikinci faktöre doğru keskin bir düşüş olduğu ve diğer faktörlerin ikinci faktöre yakın değerler aldıkları görülmektedir. Bu durum kitapçıklar için tek boyutluluğun sağlandığına işaret etmektedir. Ancak Velicer (1976) tarafından bu yöntemin subjektif (öznel) bir yöntem olduğu belirtilmiş ve bu nedenle farklı yöntemler kullanılarak desteklenmesi önerilmiştir.

Yaygın olarak kullanılan yöntemler olmalarına rağmen literatürde Kaiser kriteri ve yamaç birikinti grafiği yöntemlerinin olması gerekenden fazla sayıda faktör belirleme eğiliminde olduklarına dair çalışmalar mevcuttur (Koçak ve diğerleri, 2016; Revelle & Rocklin, 1979; Tucker ve diğerleri, 1969). Bu nedenle karşılaştırmalı araştırmalarda daha kararlı olduğu ve doğru sayıda faktör belirlediği raporlanan (Crawford ve diğerleri, 2010; Koçak ve diğerleri, 2016) Horn'un paralel analiz yöntemi kullanılmıştır. Horn (1965) tarafından geliştirilen bu yöntemde gerçek veri setine paralel rastgele veriler oluşturulur. Rastgele üretilen veriye ait özdeğerin gerçek veriye ait özdeğerden büyük olduğu nokta faktör sayısını verir. Kitapçık 6 ve Kitapçık 7'ye ilişkin paralel analiz sonuçları Tablo 10'da verilmiştir.

Tablo 10*Paralel Analiz Sonuçları*

	Kitapçık 6 Özdeğerler		Kitapçık 7 Özdeğerler	
	Gerçek Veri	Üretilen Veri	Gerçek Veri	Üretilen Veri
Faktör 1	3.723	1.121	3.862	1.120
Faktör 2	1.098	1.103	1.102	1.109
Faktör 3	1.041	1.077	1.050	1.088
Faktör 4	1.004	1.065	0.992	1.076
Faktör 5	0.993	1.054	0.973	1.065

Tablo 10'da yer alan gerçek verilere ve üretilen verilere ilişkin özdeğerler incelendiğinde her iki kitapçıkta da faktör sayısının 1 (bir) olduğu görülmektedir. Sadece Faktör 1'de gerçek verinin özdeğeri üretilen verinin özdeğerinden büyüktür. Faktör 2'den itibaren gerçek verinin özdeğeri üretilen verinin özdeğerinden küçüktür. Bu nokta kitapçıkların bir faktörlü yapıda olduğunu işaret etmektedir.

Faktör sayısına ilişkin tüm kanıtlar birlikte değerlendirildiğinde Kitapçık 6 ve Kitapçık 7'nin tek boyutluluk varsayımını karşıladığı söylenebilir.

2.2. Model Veri Uyum. Model veri uyumunu test etmek için kullanılan çok sayıda yöntem mevcuttur. Bu araştırmada verilerin hangi modele uygun olduğunu belirlemek için; -2Log Olabilirlik (-2Log Likelihood – -2LL) ve ki-kare (χ^2) fark testi ile Bayesian Bilgi Kriteri (Bayesian Information Criterion – BIC) ve Akaike Bilgi Kriteri (Akaike Information Criterion – AIC) kullanılmıştır. -2 LL değeri verinin modelden ne kadar saptığını göstermektedir. Bu değer ne kadar küçülürse veri modele o kadar yaklaşmıştır (Embretson & Reise, 2000). BIC ve AIC değerlerinin küçülmesi de verinin modele daha iyi uyum sağladığına işaret etmektedir (De Ayala, 2009). Kitapçık 6 ve Kitapçık 7 için model veri uyumunu test etmek amacıyla R kütüphanesinde yer alan “sirt” (Robitzsch, 2022) paketinden yararlanılmıştır. Her bir kitapçık için model veri uyumuna ilişkin değerlere Tablo 11'de yer verilmiştir.

Tablo 11*Model Veri Uyumu*

Model	Kitapçık 6			Kitapçık 7		
	Rasch	2PLM	3PLM	Rasch	2PLM	3PLM
-2LL	144313.1	143322.4	143131.7	146603.8	145716.0	145654.4
AIC	144359.1	143414.4	143269.6	146649.8	145808.0	145792.5
BIC	144508.9	143714.0	143719.0	146800.2	146108.8	146243.6
Parametre Sayısı	23	46	69	23	46	69

Tablo 11 incelendiğinde; Rasch modelden 3PLM'ye doğru gidildikçe -2LL ve AIC değerlerinin düştüğü gözlenmiştir. BIC değerlerinin ise Rasch modelden 2PLM'ye doğru gidildikçe azaldığı ancak 2PLM'den 3PLM'ye doğru gidildikçe arttığı tespit edilmiştir. En küçük -2LL değerlerinin 2PLM ve 3PLM kullanıldığında elde edildiği belirlenmiştir. Bu nedenle χ^2 fark testine bakılmıştır. χ^2 tablosuna göre Kitapçık 6 için, 3PLM ve 2PLM'den elde edilen -2LL değerleri arasındaki fark ($\chi^2 = -2LL_{2PLM} - -2LL_{3PLM} = 144.8$), serbestlik derecesi (23) ile 0.05 anlamlılık düzeyine göre karşılaştırılmıştır. Benzer şekilde χ^2 tablosuna göre Kitapçık 7 için, 3PLM ve 2PLM'den elde edilen -2LL değerleri arasındaki fark ($-2LL_{2PLM} - -2LL_{3PLM} = 61.6$), serbestlik derecesi (23) ile 0.05 anlamlılık düzeyine göre karşılaştırılmıştır. Her iki durumda da modeller arasındaki farkın χ^2 değerinden büyük olduğu görülmüştür. Dolayısıyla hem Kitapçık 6 hem de Kitapçık 7 için, 3PLM'nin 2PLM'ye göre anlamlı bir farklılık oluşturduğu söylenebilir. Bu bilgiler ışığında, araştırmada çoktan seçmeli maddelerin kullanıldığı da göz önüne alınarak her iki kitapçık için de veriye en iyi uyum sağlayan modelin 3PLM olduğuna karar verilmiştir. Bu nedenle araştırmanın devamındaki analizler 3PLM'ye göre yapılmıştır.

2.3. Yerel Bağımsızlık. MTK'nin bu varsayımına göre, bireyin bir testteki maddelere verdiği cevaplar birbirinden bağımsızdır (Lord, 1980; Hambleton & Swaminathan, 1985). Ayrıca tek boyutluluğun sağlanması sonucunda yerel bağımsızlığın da sağlandığı kabul edilir. Yerel bağımsızlığın test edilmesinde kullanılan bir başka kriter ise Yen'in Q_3

istatistiğidir. Bu istatistik madde çiftlerine odaklanır ve her bir madde çifti için üretilen artık değerlerin korelasyonu olarak ifade edilir. Q_3 değerinin 0.20'den büyük olmaması, o testte yer alan ilgili madde çiftleri için yerel bağımsızlık varsayımının sağlandığına kanıt oluşturmaktadır (Demars, 2016). Bu bilgiler ışığında araştırmada R kütüphanesinde yer alan “sirt” (Robitzsch, 2022) paketinden yararlanılmış ve Yen'in Q_3 istatistiği kullanılarak yerel bağımsızlık test edilmiştir. Kitapçıklarda yer alan 23'er maddeye ait 253'er madde çifti için 3PLM altında Q_3 değerleri elde edilmiş ve sonuçlar EK-A'da verilmiştir. EK-A incelendiğinde her iki kitapçık için de elde edilen değerlerin 0,20'yi aşmadığı görülmüştür. Buna göre yerel bağımsızlık varsayımının sağlandığı sonucuna ulaşılmıştır.

2.4. Madde Karakteristik Eğrisi (MKE)'nin Monoton Artışı. Bu varsayım Hambleton vd. (1991) tarafından, yetenek düzeyi arttıkça bireyin maddeyi doğru cevaplanma olasılığının artması olarak ifade edilmiştir. Araştırmada ele alınan Kitapçık 6 ve Kitapçık 7'de yer alan maddeler için madde karakteristik eğrileri R kütüphanesindeki “ltm” (Rizopoulos, 2022) paketinde yer alan “*plot*” fonksiyonu kullanılarak elde edilmiştir. EK-B ve EK-C incelendiğinde kitapçıklarda yer alan tüm maddelerde, bireylerin yetenek düzeyleri yükseldiğinde maddeyi doğru yanıtlama olasılıklarının da yükseldiği gözlenmiştir. Sonuç olarak MKE'nin monoton artışı varsayımının sağlandığı kabul edilmiştir.

3) Madde ve Yetenek Parametrelerinin Kestirilmesi

Madde parametrelerinin kestiriminde, gözlemlenen cevap verilerinin olasılığını en üst düzeye çıkaran bireylerin madde parametre tahminlerini veren ve varsayılan yöntem olarak kullanılan en çok olabilirlik tahmini (Maximum Likelihood Estimation - MLE); yetenek parametrelerinin kestiriminde ise öncül dağılımın ortalamasını kullanması ve maddelerin tamamını doğru ya da tamamını yanlış yanıtlayanlar için de yetenek kestirimi yapılabilmesinden dolayı beklenen sonsal (Expected A Posteriori – EAP) yöntemi (Embretson & Reise, 2000) kullanılmıştır. Madde ve yetenek parametrelerinin kestirilmesi

için R kütüphanesinde yer alan “lrm” (Rizopoulos, 2022) paketi kullanılmıştır. Sonuçlar Tablo 12’de yer almaktadır.

Tablo 12

Madde Parametre Kestirimleri

Kitapçık 6				Kitapçık 7			
Madde	Ayırıcılık Parametresi (a)	Güçlük Parametresi (b)	Şans Parametresi (c)	Madde	Ayırıcılık Parametresi (a)	Güçlük Parametresi (b)	Şans Parametresi (c)
1	0.563	-1.267	0.000	1	0.567	-1.467	0.007
2	0.783	2.021	0.000	2	0.860	2.097	0.000
3	0.711	0.399	0.094	3	0.735	0.243	0.087
4	1.504	0.567	0.239	4	0.860	-0.004	0.059
5	1.114	0.228	0.000	5	1.086	0.208	0.000
6	1.314	-0.045	0.386	6	0.804	-0.435	0.260
7	1.204	0.881	0.207	7	0.887	0.760	0.119
8	0.981	-0.013	0.204	8	0.782	-0.232	0.091
9	0.748	0.186	0.175	9	0.637	-0.025	0.141
10	1.132	0.449	0.267	10	1.485	0.815	0.307
11	1.180	0.342	0.000	11	1.171	0.334	0.000
12	1.486	-0.092	0.336	12	1.026	-0.278	0.342
13	0.461	-1.211	0.000	13	0.525	-0.534	0.001
14	1.847	1.014	0.237	14	1.183	-0.342	0.149
15	1.011	0.737	0.000	15	1.271	1.137	0.000
16	0.763	-0.160	0.171	16	1.134	0.813	0.400
17	1.036	-0.477	0.050	17	1.234	-0.525	0.001
18	1.229	0.418	0.000	18	1.251	-0.076	0.000
19	1.224	0.464	0.350	19	1.015	0.082	0.000
20	2.196	0.303	0.130	20	1.074	-0.179	0.000
21	0.653	0.012	0.000	21	0.882	0.841	0.109
22	0.867	-0.405	0.000	22	0.915	0.742	0.000
23	1.073	1.981	0.198	23	1.365	0.247	0.126

Tablo 12’de parametre değerleri verilen; ilk 11 madde her iki kitapçıkta da ortak olan, son 12 madde ise her iki kitapçıkta da ortak olmayan maddelerdir. Tabloda 11’de verilen a, b ve c parametrelerinin ortalamalarına Tablo 13’te yer verilmiştir.

Tablo 13

Madde Parametrelerinin Ortalamaları

	N	$\bar{X}_a$	$\bar{X}_b$	$\bar{X}_c$
Kitapçık 6	23 (11 ortak + 12 ortak olmayan madde)	1.090	0.275	0.132
Kitapçık 7	23 (11 ortak + 12 ortak olmayan madde)	0.989	0.184	0.096

Tablo 13 incelendiğinde; a parametresinin ortalamasının Kitapçık 6 için $\bar{X}_a = 1.090$, Kitapçık 7 için $\bar{X}_a = 0.989$ olduğu görülmektedir. a parametresi teorik olarak $-\infty$ ile $+\infty$ arasında değer alsa da uygulamada 0 ile 2 arasında değişmektedir. a parametresi 2’ye yaklaştıkça maddenin ayırcılığı artmaktadır (Hambleton ve diğerleri, 1991). Buna göre her iki kitapçığın ayırt ediciliğinin orta düzeyde olduğu söylenebilir. b parametresinin ortalaması Kitapçık 6 için $\bar{X}_b = 0.275$, Kitapçık 7 için $\bar{X}_b = 0.184$ ’tür. b parametresi -2 ile +2 arasında değer almaktadır. Bu değer -2’ye yaklaştıkça maddenin kolay, +2’ye yaklaştıkça maddenin zor olduğunu göstermektedir (Hambleton ve diğerleri, 1991). Buna göre her iki kitapçığın da orta güçlük düzeyinde olduğu söylenebilir. c parametresinin ortalamasının ise Kitapçık 6 için $\bar{X}_c = 0.132$, Kitapçık 7 için $\bar{X}_c = 0.096$ olduğu görülmektedir. c parametresi teorik olarak 0 ile 1 arasında değişen değerler alabilir ancak uygulamada 0 ile 0.35 aralığında değişmektedir (Baker, 2001). Buna göre her iki kitapçık için de şans ile maddeyi doğru cevaplama olasılıklarının kabul edilebilir aralıkta olduğu söylenebilir.

4) Ölçek Dönüştürme Süreci ve Test Eşitleme

Ölçek dönüştürme sürecinde, R kütüphanesindeki “equateIRT” (Battauz, 2022) paketinde yer alan “direc” fonksiyonu kullanılarak ortak maddeler ile iki test formu arasındaki eşitleme katsayıları hesaplanmıştır. Böylece; madde parametre katsayıları kullanılarak

eşitlenecek formların kovaryans matrisini içeren bir nesne ile bir ölçek dönüştürme yöntemi (Haebara, Stocking-Lord) seçilerek test formlarının aynı ölçek üzerine yerleştirilmesi sağlanmıştır. Bu araştırmada denk olmayan gruplarda ortak madde deseni altında 3PL model kullanıldığı için ölçek dönüştürme yöntemi olarak ortak maddelerin a, b ve c parametrelerini birlikte hesaba katarak ele alan karakteristik eğri dönüştürme yöntemleri (Haebara ve Stocking-Lord) tercih edilmiştir. Ardından aynı pakette yer alan “score” fonksiyonu ile önceden hesaplanan eşitleme katsayıları ile bir puanlama yöntemi (gerçek puan eşitleme) kullanılarak theta (θ) değerleri ve eşitlenmiş puanların standart hataları hesaplanmıştır. Her bir koşul için araştırmacı tarafından yazılan R kodları ile 50 tekrar yapılarak toplam 2400 (48x50) eşitleme işlemi gerçekleştirilmiştir. Test puanları gerçek puanlar cinsinden raporlanmıştır. Gerçek puan eşitleme yöntemi kullanılarak elde edilen puanlar gözlenen puan eşitleme yönteminden elde edilen puanlara göre daha kesin ve doğrudur (Li ve diğerleri, 2012).

Gerçek Puan Eşitleme (True Score Equating-TSE). İki ya da daha fazla formun madde parametreleri ölçeklendirildikten sonra temel formdaki puana karşılık gelen hedef formdaki puanı sağlamak için MTK gerçek puan eşitleme yapılabilir. Gerçek puan eşitlemede kişi parametrelerinin formlar arasında eşdeğer olduğu varsayılmaktadır. n: Toplam madde sayısı (Formlar arasında eşit sayıda olmak zorunda değildir), T: Temel form, H: Hedef form olmak üzere:

$$\text{Puan-T}(\theta_i) = \sum_{j=1}^{nT} p_{ij}(\theta_i; a_j, b_j, c_j) \quad (9)$$

$$\text{Puan-H}(\theta_i) = \sum_{j=1}^{nH} p_{ij}(\theta_i; a_j, b_j, c_j) \quad (10)$$

Eşitlik 9 ve Eşitlik 10, temel form ve hedef form için test karakteristik eğrilerini ifade eder.

Gerçek puan eşitleme işleminin adımları aşağıdaki gibidir:

1. Temel formdan olası bir Puan-T seçilir.

2. Karşılık gelen Puan-T kişi parametre değeri (θ_i) temel formun test karakteristik eğrisinden tanımlanır.
3. Hedef formun test karakteristik eğrisinden aynı kişi parametresine (θ_i) karşılık gelen eşdeğer Puan-H bulunur.

5) Eşitleme Sonuçlarının Karşılaştırılması

Eşitleme sonuçlarının doğruluğunu değerlendirmek için birden fazla ölçütün kullanılması önerilmektedir (Harris & Crouse, 1993). Literatür incelendiğinde hem yorumlama hem de kullanım kolaylığı açısından sıklıkla tercih edilen değerlendirme ölçütlerinin RMSE ve BIAS olduğu görülmüştür (Bastari, 2000; Ertoprak, 2017; Han, 2008; He, 2013; Karagül, 2020; Kim & Lee, 2006; Kim 2015; Tian, 2011; Uysal, 2019). Bu araştırmada da gerçek puan eşitleme performansları RMSE ve BIAS açısından değerlendirilmiş ve karşılaştırılmıştır. Eşitlik 11 ve Eşitlik 12’de yer alan formüller için Deng ve Monflis (2017) ile Harris ve Crouse (1993)’ un kaynaklarından yararlanılmıştır.

RMSE (Root Mean Squared Error- Hata Kareleri Ortalamasının Karekökü).

Toplam eşitleme hatası olarak da adlandırılan bu değer kestirilen puanların gerçek puanlara uyum derecesini gösterir. RMSE değeri Eşitlik 11’de gösterilmiştir.

$$RMSE = \sqrt{\frac{\sum_{i=1}^N (\hat{\theta}_i - \theta_i)^2}{N}} \quad (11)$$

N : Sınava katılan toplam birey sayısı

$\hat{\theta}_i$: i bireyinin kullanılan eşitleme yöntemi ile kestirilen yetenek düzeyi

θ_i : i bireyinin gerçek yetenek düzeyi

Araştırmada 50 tekrar sonucu elde edilen RMSE değerleri ele alınan koşullara göre ortalamaları alınarak raporlanmıştır. Daha küçük ortalama RMSE değeri ile daha az hatalı eşitleme sonuçlarına ulaşılabildiği söylenebilir. Ayrıca ortalama RMSE değerinin, bu araştırmada referans grup olarak ele alınan tam veriye sahip test formlarının eşitlenmesi

sonucu elde edilen RMSE değerine yakın olması uyum derecesi daha yüksek eşitleme sonuçlarının elde edildiğini gösterir.

BIAS (Eşitleme Yanlılığı). Sistematik hata olarak da adlandırılan bu değer kestirilen puanlar ile gerçek puanlar arasındaki farkı yani eşitleme sonuçlarının doğruluk düzeyini gösterir. BIAS değeri Eşitlik 12’de gösterilmiştir.

$$BIAS = \frac{\sum_{i=1}^N (\hat{\theta}_i - \theta_i)}{N} \quad (12)$$

Araştırmada 50 tekrar sonucu elde edilen BIAS değerleri ele alınan koşullara göre ortalamaları alınarak raporlanmıştır. Ortalama BIAS değeri eşitlemenin doğruluğu ve yönü hakkında bilgi verir. Ortalama BIAS değeri küçüldükçe elde edilen eşitleme sonuçlarındaki yanlılık azalır. Ayrıca ortalama BIAS değerinin, bu araştırmada referans grup olarak ele alınan tam veriye sahip test formlarının eşitlenmesi sonucu elde edilen BIAS değerine yakın olması doğruluk düzeyi daha yüksek eşitleme sonuçlarının elde edildiğini gösterir.

Daha doğru eşitleme sonuçları için ortalama RMSE ve ortalama BIAS değerleri küçülmelidir. Başka bir ifade ile düşük RMSE ve yanlılık değerlerine sahip olan yöntem iyi bir yöntemdir (Chu & Kamata, 2005).

Araştırma kapsamında oluşturulan veri setleri Tablo 14’te özetlenmiştir.

Tablo 14

Araştırma Kapsamında Oluşturulan Veri Setleri

Kayıp Veri ile Başa Çıkma Yöntemi	Örneklem Büyüklüğü	Kayıp Veri Oranı	Kayıp Verilerin Bulunduğu Test Formu	Veri Seti (V)
Referans (R)	-	-	-	R*
SA	750, 1500	%10	ET	V1**, V7
HDA	750, 1500	%10	ET	V13, V19
PMM	750, 1500	%10	ET	V25, V31
LOGREG	750, 1500	%10	ET	V37, V43
SA	750, 1500	%10	HİT	V2, V8
HDA	750, 1500	%10	HİT	V14, V20

PMM	750, 1500	%10	HİT	V26, V32
LOGREG	750, 1500	%10	HİT	V38, V44
SA	750, 1500	%20	ET	V3, V9
HDA	750, 1500	%20	ET	V15, V21
PMM	750, 1500	%20	ET	V27, V33
LOGREG	750, 1500	%20	ET	V39, V45
SA	750, 1500	%20	HİT	V4, V10
HDA	750, 1500	%20	HİT	V16, V22
PMM	750, 1500	%20	HİT	V28, V34
LOGREG	750, 1500	%20	HİT	V40, V46
SA	750, 1500	%30	ET	V5, V11
HDA	750, 1500	%30	ET	V17, V23
PMM	750, 1500	%30	ET	V29, V35
LOGREG	750, 1500	%30	ET	V41, V47
SA	750, 1500	%30	HİT	V6, V12
HDA	750, 1500	%30	HİT	V18, V24
PMM	750, 1500	%30	HİT	V30, V36
LOGREG	750, 1500	%30	HİT	V42, V48

*R: Referans değer

**V: Veri seti

Araştırmada ele alınan koşullardan;

Örneklem büyüklüğüne göre RMSE ve BIAS değerlerinin nasıl değiştiğini yorumlayabilmek için kayıp veri oranı, kayıp verilerin bulunduğu test formu ve kayıp veri ile başa çıkma yöntemlerinin aynı olduğu durumlarda, farklı örneklem büyüklüklerine göre hesaplanan RMSE ve BIAS değerleri ile referans değerler karşılaştırılmıştır. Tablo 14'e göre aynı satırda verilen veri setleri kendi aralarında ve referans değer ile karşılaştırılarak incelenmiştir.

Kayıp veri oranına göre RMSE ve BIAS değerlerinin nasıl değiştiğini yorumlayabilmek için örneklem büyüklüğü, kayıp verilerin bulunduğu test formu ve kayıp veri ile başa çıkma yöntemlerinin aynı olduğu durumlarda, farklı kayıp veri oranlarına göre hesaplanan RMSE ve BIAS değerleri ile referans değerler karşılaştırılmıştır. Örneğin Tablo 14'e göre SA, 750 örneklem büyüklüğü ve ET için V1, V3, V5 kendi aralarında ve referans

değer ile; SA, 1500 örneklem büyüklüğü ve ET için V7, V9, V11 kendi aralarında ve referans değer ile karşılaştırılarak incelenmiştir. İncelemeler tüm kayıp veri oranlarında ve tüm koşullar için ayrı ayrı yapılmıştır.

Kayıp verilerin bulunduğu test formuna göre RMSE ve BIAS değerlerinin nasıl değiştiğini yorumlayabilmek için örneklem büyüklüğü, kayıp veri oranı ve kayıp veri ile başa çıkma yöntemlerinin aynı olduğu durumlarda, kayıp verilerin bulunduğu test formuna göre hesaplanan RMSE ve BIAS değerleri ile referans değerler karşılaştırılmıştır. Örneğin Tablo 14'e göre SA, 750 örneklem büyüklüğü ve %10 kayıp veri oranı için V1, V2 kendi aralarında ve referans değer ile SA, 1500 örneklem büyüklüğü ve %10 kayıp veri oranı için V7, V8 kendi aralarında ve referans değer ile karşılaştırılarak incelenmiştir. İncelemeler kayıp verilerin bulunduğu tüm test formlarında ve tüm koşullar için ayrı ayrı yapılmıştır.

Karakteristik eğri dönüştürme yöntemlerine göre RMSE ve BIAS değerlerinin nasıl değiştiğini yorumlayabilmek için örneklem büyüklüğü, kayıp veri oranı, kayıp verilerin bulunduğu test formu ve kayıp veri ile başa çıkma yöntemlerinin aynı olduğu durumlarda, farklı karakteristik eğri dönüştürme yöntemlerine göre hesaplanan RMSE ve BIAS değerleri ile referans değerler karşılaştırılmıştır. Örneğin Tablo 14'te elde edilen 48 veri setinin Haebara yöntemine göre oluşturulduğu düşünülürse tüm veri setleri için ayrıca Stocking-Lord yöntemi kullanılarak da oluşturulan 48 farklı veri seti mevcuttur. Bu durumda SA, 750 örneklem büyüklüğü, %10 kayıp veri oranı, ET için V1 ile bu veri setinin Stocking-Lord yöntemine göre karşılığı olarak oluşturulan veri seti kendi aralarında ve referans değer ile karşılaştırılarak incelenmiştir. İncelemeler tüm karakteristik eğri dönüştürme yöntemlerinde ve tüm koşullar için ayrı ayrı yapılmıştır.

Bölüm 4

Bulgular, Yorumlar ve Tartışma

Bu bölümde araştırmanın alt problemlerine ilişkin analiz sonuçlarına göre elde edilen bulgular ile literatür ve ilgili araştırmalar göz önüne alınarak yapılan yorumlar ve tartışmaya yer verilmiştir. Araştırmada elde edilen analiz sonuçları Ek-Ç'de yer alan tabloda raporlanmıştır. Bulguların daha kolay anlaşılabilir olması için Ek-Ç'deki sonuçlar, araştırmacı tarafından yazılan kodlar kullanılarak elde edilen grafikler şeklinde sunulmuştur.

Birinci Alt Probleme İlişkin Bulgular, Yorumlar ve Tartışma

1. Tam veriye sahip test formları ve kayıp veriyle başa çıkma yöntemleri (SA, HDA, PMM, LOGREG) uygulanarak tamamlanan test formları, Haebara ve Stocking-Lord yöntemlerine göre eşitlendiğinde, araştırmada ele alınan koşullara göre;

a. RMSE değerleri,

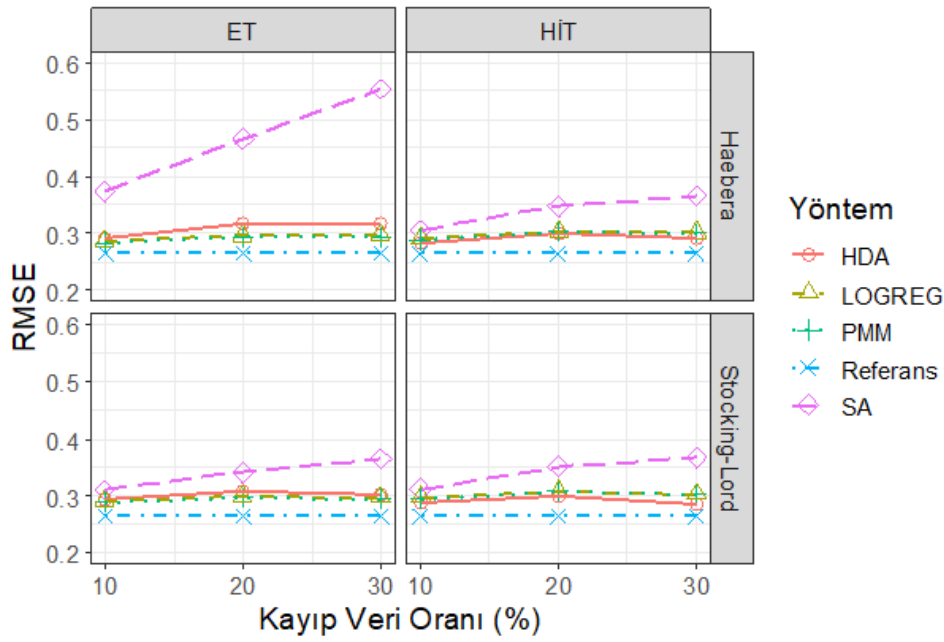
b. BIAS değerleri

nasıl değişmektedir?

a. RMSE değerlerine ilişkin grafikler, 750 örneklem büyüklüğü için ayrı 1500 örneklem büyüklüğü için ayrı olmak üzere, araştırmada ele alınan koşullara göre Şekil 9 ve Şekil 10'da gösterilmiştir. Ardından Şekil 9 ve Şekil 10 birlikte değerlendirilerek karşılaştırma yapılmıştır.

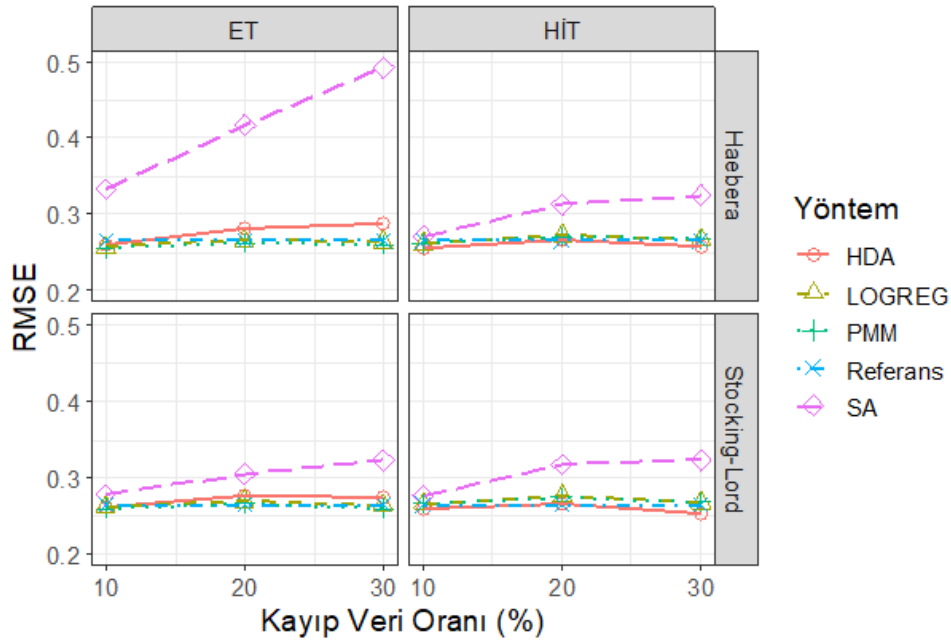
Şekil 9

750 Örneklem Büyüklüğü İçin RMSE Grafiği



Şekil 10

1500 Örneklem Büyüklüğü İçin RMSE Grafiği



Örneklem büyüklüğüne göre değerlendirildiğinde; araştırmada kullanılan kayıp veriyle başa çıkma yöntemlerinin (SA, HDA, PMM, LOGREG) hepsi için, örneklem

büyüklüğü arttıkça RMSE değerlerinin azaldığı ve söz konusu RMSE değerlerinin referans değere yaklaştığı gözlenmektedir. Test eşitlemede kayıp veri ile başka çıkma yöntemleriyle ilgili yapılan çalışmalarda farklı örneklem büyüklüklerindeki etkilerin incelenmediği belirlenmiştir ancak Atalay Kabasakal (2014), Hanson ve Beguin (2002), Kang ve Petersen (2011), Lee ve Ban (2010), Speron (2009), Çokluk Bökeoğlu vd. (2022), Zhang (2010) tarafından yapılan test eşitleme ile ilgili çalışmalarda ele alınan örneklem büyüklüğü arttıkça eşitleme hatalarının azaldığını raporlamışlardır. Bu yönüyle bakıldığında araştırmada elde edilen bulgu ile literatürde yer alan bulguların örtüştüğü görülmektedir. Ayrıca Chu ve Kamata (2005) MTK'ye dayalı eşitleme yöntemlerinde örneklem büyüklüğü arttıkça kararlılığın da artacağını belirtmiştir. Araştırmada kullanılan test eşitleme yöntemlerinden bağımsız olarak örneklem büyüklüğü arttıkça daha az hata ile daha doğru sonuçlara ulaşıldığı görülmektedir. Bununla birlikte araştırmada kullanılan kayıp veri ile başa çıkma yöntemlerinin hepsi için örneklem büyüklüğü arttıkça eşitleme hatasının azaldığı ve daha kararlı kestirimler elde edildiği bulunmuştur.

Kayıp veri oranına göre değerlendirildiğinde; SA yöntemi için kayıp veri oranı arttıkça RMSE değerlerinin de arttığı ve referans değerden uzaklaştığı belirlenmiştir. HDA, PMM ve LOGREG yöntemleri için araştırmada ele alınan %10, %20 ve %30 kayıp veri oranlarının hepsinde genel olarak RMSE değerlerinin birbirine çok yakın olduğu gözlenmiştir. Söz konusu RMSE değerleri ile referans değer arasındaki mesafenin de genel olarak değişmediği tespit edilmiştir. Test eşitlemede kayıp veri ile başka çıkma yöntemleriyle ilgili yapılan çalışmalarda farklı kayıp veri oranlarındaki etkileri inceleyen sınırlı sayıda çalışma bulunmaktadır. Bu çalışmalardan ilkinde Shin (2016), %6-%9 oranında kayıp veriye sahip testin özellikleri ile %1-%2 oranında kayıp veriye sahip testin özellikleri arasındaki farkın daha büyük olduğuna işaret etmiştir. Yani araştırmada kullanılan kayıp veriyle başa çıkma yöntemleri, küçük kayıp veri oranlarında benzer özelliklere sahip ölçekler üretme eğilimindeyken kayıp veri oranındaki artışın ölçeklerin özelliklerinde farklılıklara neden olduğunu belirtmiştir. Bu bulgunun araştırmada ele alınan SA yöntemi için elde edilen sonucu desteklediği görülmekle birlikte araştırmada ele alınan diğer kayıp veri ile başa

çıkma yöntemleri (HDA, PMM, LOGREG) için elde edilen sonuç ile örtüşmediği belirlenmiştir. Sonuçlardaki bu farklılığa, Shin (2016)'in çalışmasının dikey ölçekler üzerinde yapılması ve bu ölçek türünde farklı güçlük düzeyindeki test formlarının farklı yetenek düzeyindeki bireylere uygulanmasının yol açabileceği öngörülmektedir. Ertoprak (2017) ise, TRK veri mekanizması altında %10 ve %20 kayıp veri oranı için kestirilen RMSE değerleri arasında gözle görülür farklılıkların oluşmadığını ancak kayıp veri oranının %40 olduğu koşulda hata miktarının arttığını belirtmiştir. Bu bulgunun ilk kısmı, araştırmada ele alınan HDA, PMM ve LOGREG yöntemleri için kayıp veri oranlarının hepsinde benzer RMSE değerlerinin gözlemlendiği bulgusu ile örtüşmektedir. Bulgunun ikinci kısmı ise araştırmada elde edilen SA yöntemi için kayıp veri oranı arttıkça RMSE değerlerinin de arttığı sonucunu desteklemektedir. Bununla birlikte araştırma sonuçlarının bir kısmında meydana gelen farklılıklara; simülatif olarak üretilen veri seti kullanımı, ele alınan kayıp veri oranı ve kayıp veri ile başa çıkma yöntemlerindeki farklılık sebep olmuş olabilir. Literatürde kayıp veri oranlarının farklı değişkenler üzerindeki etkisini inceleyen çalışmalarda kayıp veri oranı arttıkça kayıp veriyle başa çıkma yöntemlerinin performanslarının ve model veri uyumunun azaldığı, tam veri setinden elde edilen değerlerden farklı sonuçlar ortaya çıktığı, hatanın arttığı bulgulanmıştır (Chen ve diğerleri, 2012; Furlow ve diğerleri, 2007; Sulis & Porcu, 2008; Koçak & Çokluk Bökeoğlu, 2017; Marsh, 1998). Ayrıca kayıp veri oranı arttıkça çoklu atama yöntemi ile elde edilen tahmin değerlerinin diğer yöntemlerden elde edilen değerlere göre daha iyi olduğu (Baraldi & Enders, 2010; Leite & Bretvas, 2010, Pokropek, 2011) ve çoklu atamanın araştırmada ele alınan tüm kayıp veri oranlarında çalıştığı raporlanmıştır (Karaman, 2022). Bu bulgulardan hareketle araştırmada ele alınan kayıp veri oranının artışından çok seçilen kayıp veri ile başa çıkma yönteminin eşitleme hatasında etkin rol oynadığı söylenebilir.

Kayıp verilerin bulunduğu test formuna göre değerlendirildiğinde; SA yöntemi için karakteristik eğri dönüştürme yöntemlerinden Haebara kullanıldığında her iki testte (HİT), eşitlenecek teste (ET) göre RMSE değerlerinin azaldığı belirlenmiştir. Benzer şekilde RMSE değerlerinin HİT'te ET'e göre referans değere olan uzaklığının azaldığı tespit edilmiştir.

Ancak karakteristik eğri dönüştürme yöntemlerinden Stocking-Lord kullanıldığında hem ET'te hem de HİT'te elde edilen RMSE değerlerinin birbirine çok yakın olduğu gözlenmiştir. Dolayısıyla RMSE değerleri ile referans değer arasındaki mesafenin de genel olarak değişmediği tespit edilmiştir. HDA, PMM ve LOGREG yöntemleri için kullanılan karakteristik eğri dönüştürme yöntemlerinden bağımsız olarak hem ET'te hem de HİT'te elde edilen RMSE değerlerinin birbirine çok yakın olduğu gözlenmiştir. Dolayısıyla RMSE değerleri ile referans değer arasındaki mesafenin de genel olarak değişmediği ve bu değerlerin birbirine çok yakın olduğu belirlenmiştir. Test eşitlemede kayıp veri ile başka çıkma yöntemleriyle ilgili yapılan çalışmalarda kayıp verilerin bulunduğu test formuna göre etkilerini inceleyen sadece bir çalışmaya ulaşılmıştır. Ertoprak (2017) tarafından yapılan ve kayıp verinin test formları içindeki yerine göre üç farklı koşulun (her iki test, eşitlenecek test ve ortak test) ele alındığı çalışmada; en küçük RMSE değerlerinin kayıp verilerin HİT'te, en yüksek RMSE değerlerinin ise kayıp verilerin ET'te bulunduğu koşullarda elde edildiği raporlanmıştır. Bu bulgu, araştırmada SA yöntemi için elde edilen bulguyu destekler niteliktedir. Ancak HDA, PMM ve LOGREG yöntemleri için elde edilen bulgu ile örtüşmemektedir. HDA yönteminde her kayıp verinin yerine benzer bir cevap kümesinden gözlemlenen gerçek değerler atandığı için bu yöntem ile elde edilen RMSE değerlerinin, kayıp verilerin bulunduğu test formundan (ET, HİT) bağımsız olması beklenmektedir. PMM ve LOGREG yöntemleri ise çoklu atama yöntemleri olup kayıp verilerin yerine veri setindeki gözlenen değerlerin dağılımından yola çıkılarak çoklu atama yapıldığı ve çok sayıda veri seti oluşturulduğu için bu yöntemler ile elde edilen RMSE değerlerinin, kayıp verilerin bulunduğu test formuna göre değişmediği söylenebilir.

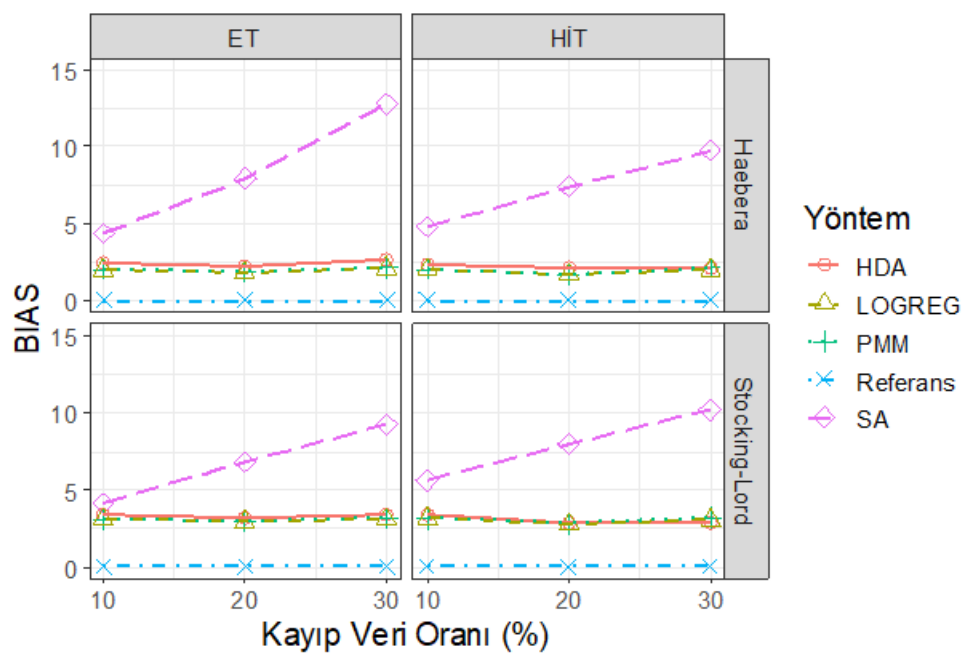
SA yönteminde, karakteristik eğri dönüştürme yöntemlerinden Stocking-Lord kullanıldığında elde edilen RMSE değerlerinin Haebara kullanıldığında elde edilen RMSE değerlerinden genel olarak daha küçük olduğu ve Stocking-Lord yöntemi kullanıldığında elde edilen RMSE değerlerinin referans değere daha yakın olduğu gözlenmektedir. HDA, PMM ve LOGREG için ise Stocking-Lord yöntemi kullanıldığında elde edilen RMSE değerleri ile Haebara yöntemi kullanıldığında elde edilen RMSE değerlerinin birbirine yakın

olduğu ve söz konusu RMSE değerlerinin referans değere çok yakın olduğu tespit edilmiştir. Literatür incelendiğinde karakteristik eğri dönüştürme yöntemlerinin karşılaştırıldığı çalışmaların bazılarında (Kim & Kolen, 2006; Kim & Lee, 2006; Lee & Ban, 2010) Haebara yönteminin, bazılarında ise (Hu ve diğerleri, 2008; Huggins, 2014; Karkee & Wright, 2004; Kilmen, 2010; Yurtçu, 2018) Stocking-Lord yönteminin en az eşitleme hatasına sahip olduğu bulgulanmıştır. Bununla birlikte hem Haebara hem de Stocking-Lord yöntemlerinin benzer performanslara sahip olduklarını ortaya koyan çok sayıda çalışma mevcuttur (Demirus, 2015; Hanson & Béguin, 2002; Kang & Petersen, 2011; Keller & Keller, 2011; Speron, 2009). Ancak test eşitlemede kayıp veri ile başa çıkma yöntemleriyle ilgili yapılan çalışmalarda karakteristik eğri dönüştürme yöntemlerine göre etkilerini inceleyen bir çalışmaya ulaşılamamıştır.

b. BIAS değerlerine ilişkin grafikler, 750 örneklem büyüklüğü için ayrı 1500 örneklem büyüklüğü için ayrı olmak üzere, araştırmada ele alınan koşullara göre Şekil 11 ve Şekil 12’de gösterilmiştir. Ardından şekiller birlikte değerlendirilerek karşılaştırma yapılmıştır.

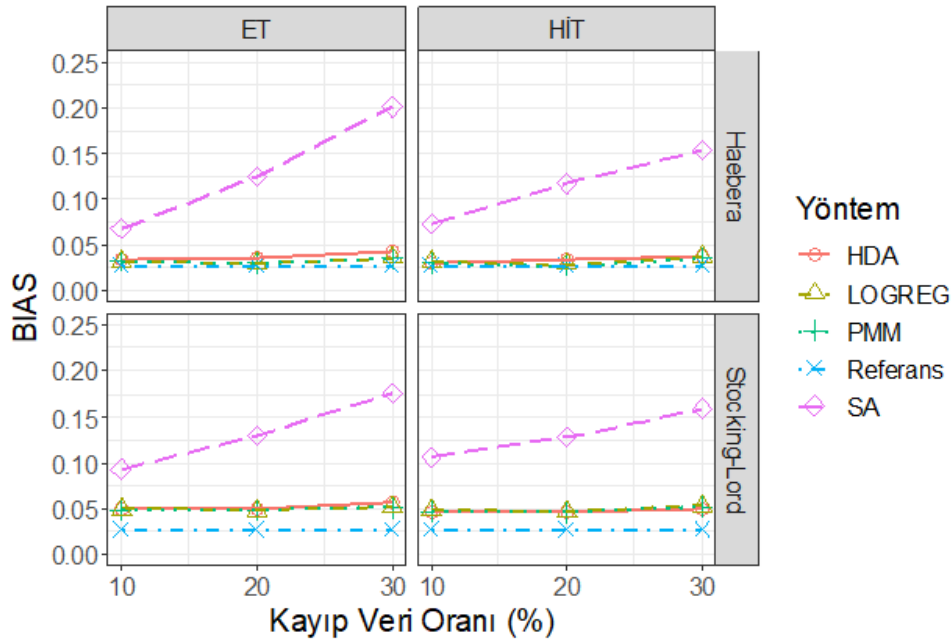
Şekil 11

750 Örneklem Büyüklüğü İçin BIAS Grafiği



Şekil 12

1500 Örneklem Büyüklüğü İçin BIAS Grafiği



Örneklem büyüklüğüne göre değerlendirildiğinde; araştırmada kullanılan kayıp veriyle başa çıkma yöntemlerinin (SA, HDA, PMM, LOGREG) hepsi için, örneklem büyüklüğü arttıkça BIAS değerlerinin azaldığı ve söz konusu BIAS değerlerinin referans değere yaklaştığı bulgulanmıştır. Bu bulgu örneklem büyüklüğü arttıkça eşitleme sonucu kestirilen puanlar ile gerçek puanlar arasındaki farkın azaldığı dolayısıyla eşitleme sonuçlarının doğruluk düzeyinin arttığını göstermektedir. Literatürde kayıp veri ile başa çıkma yöntemlerinin test eşitlemeye etkisinin incelendiği çalışmalarda farklı örneklem büyüklüklerine ilişkin koşulun ele alınmadığı görülmüştür. Ancak test eşitleme ile ilgili Liu ve Kolen (2011) tarafından karma testler için MTK eşitleme yöntemleri ile geleneksel eşitleme yöntemleri arasında bir karşılaştırmaların yapıldığı çalışmada, örneklem büyüklüğü arttıkça eşitleme yanlılığının azaldığı raporlanmıştır. Benzer şekilde Kang ve Petersen (2011) tarafından yapılan ve farklı örneklem büyüklüklerinin (500 ve 2000) ele alındığı araştırmada da örneklem büyüklüğü arttıkça yöntemlerin BIAS değerlerinin azaldığı bulgulanmıştır. Bu bulguların araştırma sonucunu destekledikleri görülmektedir. Bununla birlikte literatürde test eşitleme ile ilgili yapılan diğer çalışmalarda örneklem büyüklüğünün artmasının BIAS

değerlerinde değişikliğe neden olmadığı ve örneklem büyüklüğü arttıkça yanlışlığın değişmediği (Atalay Kabasakal, 2014; Kolen & Brennan, 2014), her koşulda aynı kaldığı veya azaldığı (Akın Arıkan & Gelbal, 2018) yani örneklem büyüklüğünün BIAS üzerinde herhangi bir etkiye sahip olmadığı ortaya konulmuştur. Yanlılığa ilişkin sonuçlardaki bu farklılıklara, söz konusu araştırmalarda ele alınan bağımsız değişkenlerdeki (değişen madde fonksiyonu, küçük ve orta büyüklükteki çapa testler vb.) farklılıkların neden olabileceği öngörülmektedir.

Kayıp veri oranına göre değerlendirildiğinde; SA yöntemi için kayıp veri oranı arttıkça BIAS değerlerinin de arttığı ve referans değerden uzaklaştığı gözlenmiştir. HDA, PMM ve LOGREG yöntemleri için %10, %20 ve %30 kayıp veri oranlarının hepsinde genel olarak BIAS değerlerinin birbirine çok yakın olduğu tespit edilmiştir. Söz konusu BIAS değerleri ile referans değer arasındaki mesafenin de genel olarak değişmediği belirlenmiştir. Test eşitlemede kayıp veri ile başka çıkma yöntemleriyle ilgili yapılan çalışmalarda farklı kayıp veri oranlarındaki etkileri inceleyen sadece bir çalışmaya ulaşılmıştır. Ertoprak (2017) tarafından gerçekleştirilen ve kayıp veri oranlarının %10, %20 ve %40 olarak ele alındığı çalışmada kayıp veri oranı arttıkça BIAS değerlerinin de arttığı bulgulanmıştır. Ancak bu bulgunun kayıt altına alınacak kadar net ve tutarlı olmadığı belirtilmiştir. Ayrıca en yanlış sonucun kayıp veri oranının %40 olduğu koşulda üretildiği raporlanmıştır. Bu bulgunun ilk kısmı, araştırmada ele alınan HDA, PMM ve LOGREG yöntemleri için kayıp veri oranlarının hepsinde benzer yanlışlık değerlerinin gözlemlendiği bulgusu ile örtüşmektedir. Bulgunun ikinci kısmı ise araştırmada elde edilen SA yöntemi için kayıp veri oranı arttıkça yanlışlık değerlerinin de arttığı sonucunu desteklemektedir. Bununla birlikte araştırma sonuçlarının bir kısmında meydana gelen farklılıklara; simülatif olarak üretilen veri seti kullanımı, ele alınan kayıp veri oranı ve kayıp veri ile başa çıkma yöntemlerindeki farklılık sebep olmuş olabilir. Literatürde kayıp verilerin farklı değişkenler üzerindeki etkilerini ele alan çalışmalarda kayıp veri oranı arttıkça kestirimlerdeki yanlışlığın da arttığı bulgulanmıştır (Brown, 1994; Marsh, 1998; Sulis & Porcu, 2008; De Ayala ve diğerleri, 2001; Rose ve diğerleri, 2010). Literatürdeki bu sonucun araştırmada ele alınan SA yöntemi için elde edilen

sonucu desteklediği görülmüştür. Ancak çoklu atama yöntemlerinde TRK veri mekanizması altında kayıp veri oranı arttıkça diğer yöntemlere göre yansız parametre tahminleri ürettiği de raporlanmıştır (Baraldi & Enders, 2010). Tüm bulgular göz önüne alındığında yanlılığa, kayıp veri oranındaki artıştan çok seçilen kayıp veri ile başa çıkma yönteminin yol açtığı söylenebilir.

Kayıp verilerin bulunduğu test formuna göre değerlendirildiğinde; SA yöntemi için karakteristik eğri dönüştürme yöntemlerinden Haebara kullanıldığında HİT'te, ET'e göre BIAS değerlerinin azaldığı belirlenmiştir. Benzer şekilde BIAS değerlerinin, karakteristik eğri dönüştürme yöntemlerinden Haebara kullanıldığında HİT'te, ET'e göre referans değere olan uzaklığının azaldığı tespit edilmiştir. Ancak karakteristik eğri dönüştürme yöntemlerinden Stocking-Lord kullanıldığında hem ET'te hem de HİT'te elde edilen BIAS değerlerinin birbirine çok yakın olduğu gözlenmiştir. Dolayısıyla BIAS değerleri ile referans değer arasındaki mesafenin de genel olarak değişmediği tespit edilmiştir. HDA, PMM ve LOGREG yöntemleri için kullanılan karakteristik eğri dönüştürme yöntemlerinden bağımsız olarak hem ET'te hem de HİT'te elde edilen BIAS değerlerinin birbirine çok yakın olduğu gözlenmiştir. Bununla birlikte söz konusu BIAS değerleri ile referans değer arasındaki mesafenin de genel olarak değişmediği ve bu değerlerin birbirine çok yakın olduğu belirlenmiştir. Test eşitlemede kayıp veri ile başka çıkma yöntemleriyle ilgili yapılan çalışmalarda kayıp verilerin bulunduğu test formuna göre etkilerini inceleyen sadece bir çalışmaya ulaşılmıştır. Ertoprak (2017) tarafından yapılan çalışmada kayıp verilerin bulunduğu test formuna ilişkin üç farklı koşul (her iki test, eşitlenecek test ve ortak test) ele alınmıştır. En küçük yanlılık değerlerinin kayıp verilerin HİT'te, en yüksek yanlılık değerlerinin ise kayıp verilerin ET'te bulunduğu durumda elde edildiği raporlanmıştır. Bu bulgu, araştırmada SA yöntemi için elde edilen bulguyu destekler niteliktedir. Ancak HDA, PMM ve LOGREG yöntemleri için araştırmada elde edilen bulgu ile örtüşmemektedir. HDA yönteminde her kayıp verinin yerine benzer bir cevap kümesinden gözlemlenen gerçek değerler atandığı için bu yöntem ile elde edilen yanlılık değerlerinin, kayıp verilerin bulunduğu test formundan (ET, HİT) bağımsız olması beklenmektedir. PMM ve LOGREG

yöntemleri ise çoklu atama yöntemleri olup kayıp verilerin yerine veri setindeki gözlenen değerlerin dağılımından yola çıkılarak çoklu atama yapıldığı ve çok sayıda veri seti oluşturulduğu için bu yöntemler ile elde edilen yanlılık değerlerinin, kayıp verilerin bulunduğu test formuna göre değişmediği söylenebilir.

SA yöntemi için karakteristik eğri dönüştürme yöntemlerinden Haebara kullanıldığında elde edilen BIAS değerleri ile Stocking-Lord kullanıldığında elde edilen BIAS değerlerinin genel olarak birbirine yakın olduğu ve her iki yöntem için de elde edilen BIAS değerleri ile referans değer arasındaki mesafenin değişmediği, bu değerlerin birbirinden uzak olduğu belirlenmiştir. HDA, PMM ve LOGREG için ise Haebara yöntemi kullanıldığında elde edilen BIAS değerlerinin Stocking-Lord kullanıldığında elde edilen BIAS değerlerinden genel olarak daha küçük olduğu ve Haebara yöntemi kullanıldığında elde edilen BIAS değerlerinin referans değere daha yakın olduğu tespit edilmiştir. Literatür incelendiğinde, Kang ve Petersen (2011) tarafından karakteristik eğri dönüştürme yöntemlerinin karşılaştırıldığı çalışmada, Haebara ve Stocking-Lord yöntemleri kullanıldığında elde edilen eşitleme yanlılıkları arasındaki farkın çok az olduğu bulgulanmıştır. Ancak test eşitlemede kayıp veri ile başka çıkma yöntemleriyle ilgili yapılan çalışmalarda karakteristik eğri dönüştürme yöntemlerine göre etkilerini inceleyen bir çalışmaya ulaşılamamıştır.

İkinci Alt Probleme İlişkin Bulgular, Yorumlar ve Tartışma

2. Referans değer ve her bir kayıp veriyle başa çıkma yöntemi (SA, HDA, PMM, LOGREG) için araştırmada ele alınan koşulların (örneklem büyüklüğü, kayıp veri oranı ve kayıp verilerin bulunduğu test formu) ortalama RMSE ve BIAS değerleri üzerinde anlamlı bir etkisi var mıdır?

Bu alt problemin çözümü için iki faktörlü gruplar arası varyans analizi (ANalysis Of VAriance-ANOVA) kullanılmıştır. Araştırmada ele alınan koşullar (örneklem büyüklüğü, kayıp veri oranı ve kayıp verilerin bulunduğu test formu) ayrı ayrı bağımsız değişken, referans değer ve her bir kayıp veriyle başa çıkma yöntemi kullanıldıktan sonra Stocking-Lord ve Haebara yöntemleri ile yapılan eşitleme sonucu elde edilen ortalama RMSE ve

BIAS değerleri ise ayrı ayrı bağımlı değişken olarak analize dahil edilmiştir. Böylece her bir bağımsız değişken için bireysel etki (ana etki) test edilmiş ve ortak etki (etkileşim etkisi) olasılığı da incelenmiştir. Analiz öncesinde betimsel istatistikler hesaplanmış ve hata varyanslarının eşitliği varsayımı için Levene testi yapılmıştır. Levene testi sonucu ($p < 0.05$) anlamlı çıkmıştır, bu sonuç gruplar bağlamında ortalama RMSE ve BIAS değerlerinin varyansının eşit olmadığını gösterir. Bu durumda analize devam edebilmek ve ANOVA sonuçlarını yorumlamak için daha katı bir anlamlılık düzeyinin belirlenmesi önerilir (Pallant, 2016). Bu nedenle araştırmada anlamlılık düzeyi 0.01 olarak belirlenmiştir. Çok sayıda anlamlılık testi yapıldığı için meydana gelen hata Bonferroni düzeltmesi kullanılarak kontrol edilmiştir. Ayrıca araştırmada bağımsız değişkenlerin bağımlı değişken üzerindeki etki büyüklüklerini göstermek için eta kare değerleri (η^2) verilmiştir. İki faktörlü ANOVA sonuçları ve bu sonuçlarda anlamlı çıkan etkilerin F ve eta kare değerleri Tablo 15'te raporlanmıştır.

Tablo 15

Ortalama RMSE ve BIAS Değerleri İçin ANOVA Sonuçları

		Kayıp Veriyle Başa Çıkma Yöntemi								
			SA		HDA		PMM		LOGREG	
	Etkiler	sd	F	η^2	F	η^2	F	η^2	F	η^2
RMSE (Stocking -Lord)	ÖB	1	75.71*	.79	42.01*	.68	80.01*	.80	108.40*	.84
	KVO	2	85.36*	.83	-	-	-	-	-	-
	KVB	1	-	-	-	-	-	-	-	-
	ÖBXKVO	2	-	-	-	-	-	-	-	-
	ÖBXKVB	1	-	-	-	-	-	-	-	-
	KVOXKVB	2	-	-	-	-	-	-	-	-
RMSE (Haebara)	ÖB	1	21.28*	.52	23.20*	.54	71.70*	.78	88.93*	.82
	KVO	2	26.30*	.59	-	-	-	-	-	-
	KVB	1	41.18*	.67	-	-	-	-	-	-
	ÖBXKVO	2	-	-	-	-	-	-	-	-
	ÖBXKVB	1	-	-	-	-	-	-	-	-
	KVOXKVB	2	-	-	-	-	-	-	-	-
BIAS (Stocking -Lord)	ÖB	1	60.15*	.75	904.31*	.98	2590.57*	.99	2924.08*	.99
	KVO	2	8.76*	.33	-	-	-	-	-	-
	KVB	1	-	-	-	-	-	-	-	-
	ÖBXKVO	2	-	-	-	-	-	-	-	-
	ÖBXKVB	1	-	-	-	-	-	-	-	-
	KVOXKVB	2	-	-	-	-	-	-	-	-
BIAS (Haebara)	ÖB	1	36.14*	.64	807.69*	.98	919.62*	.98	1109.89*	.98
	KVO	2	8.23*	.31	-	-	-	-	-	-
	KVB	1	8.26*	.29	-	-	-	-	-	-
	ÖBXKVO	2	-	-	-	-	-	-	-	-
	ÖBXKVB	1	-	-	-	-	-	-	-	-
	KVOXKVB	2	-	-	-	-	-	-	-	-

* $p < 0.01$

**ÖB: Örneklem büyüklüğü, KVO: Kayıp veri oranı, KVB: Kayıp verilerin bulunduğu test formu, ÖBXKVO: Örneklem büyüklüğü-kayıp veri oranı etkileşimi, ÖBXKVB: Örneklem büyüklüğü-kayıp verilerin bulunduğu test formu etkileşimi, KVOXKVB: Kayıp veri oranı- kayıp verilerin bulunduğu test formu etkileşimi.

Tablo 15'te anlamlı bulunan ana etkiler incelendiğinde; örneklem büyüklüğünün SA, HDA, PMM ve LOGREG yöntemlerinin hepsi için anlamlı bir etkisinin olduğu, kayıp veri oranının ise sadece SA yöntemi için anlamlı bir etkisinin olduğu tespit edilmiştir. Kayıp verilerin bulunduğu test formunun ise sadece SA kullanıldığında ve Haebera yöntemi ile eşitleme yapıldığında elde edilen ortalama RMSE ve BIAS değerleri üzerindeki etkisinin anlamlı olduğu belirlenmiştir. Anlamlı bulunan tüm etkilerin etki büyüklükleri Cohen'in (1988) belirlediği değerlere göre (.01=küçük etki, .06=orta etki, .14=büyük etki) büyük olarak belirlenmiştir. Yani ana değerlerdeki gerçek fark oldukça büyüktür. Etkileşim etkileri incelendiğinde ise referans değer ve her bir kayıp veriyle başa çıkma yöntemi için araştırmada ele alınan koşulların ortalama RMSE ve BIAS değerleri üzerinde anlamlı bir etkileşim etkisinin (ÖBXKVO, ÖBXKVB, KVOXKVB) bulunmadığı görülmektedir.

Tüm bulgular göz önüne alındığında; HDA, PMM ve LOGREG yöntemleri için araştırmada ele alınan koşullardan sadece örneklem büyüklüğünün ortalama RMSE ve BIAS değerleri üzerindeki etkisinin anlamlı olduğu görülmektedir. SA yöntemi için örneklem büyüklüğü ve kayıp veri oranının ortalama RMSE (Stocking-Lord) ve BIAS (Stocking-Lord) değerleri üzerinde; örneklem büyüklüğü, kayıp veri oranı ve kayıp verilerin bulunduğu test formunun ortalama RMSE (Haebera) ve BIAS (Haebera) değerleri üzerinde anlamlı bir etkisinin olduğu belirlenmiştir. Test eşitlemede kayıp veri ile başka çıkma yöntemleriyle ilgili yapılan çalışmalarda ele alınan etkilerin anlamlı olup olmadığı hususunu inceleyen sadece bir çalışmaya ulaşılmıştır. Ertoprak (2017) tarafından yapılan çalışmada kayıp verilerin bulunduğu test formu, kayıp veri mekanizması ve kayıp veri oranı ile bu koşulların etkileşim etkileri ele alınmıştır. Ortalama RMSE ve BIAS değerlerine ilişkin yapılan ANOVA sonuçlarına göre; SA yöntemi hariç diğer yöntemlerin hepsi (uygulanmamış gibi davranma-

UGD, lojistik regresyona dayalı çoklu değer atama-LRÇDA, diskriminant fonksiyonuna dayalı çoklu değer atama- DFÇDA) için kayıp verilerin bulunduğu test formunun ortalama RMSE ve BIAS değerleri üzerindeki etkisinin anlamlı olduğu bulgulanmıştır. LRÇDA yöntemi için; kayıp verilerin bulunduğu test formu ve kayıp veri oranının etkileşim etkisinin ortalama RMSE ve BIAS değerleri üzerinde anlamlı olduğu, ayrıca kayıp veri oranının ortalama BIAS değerleri üzerindeki etkisinin anlamlı olduğu raporlanmıştır. Her iki araştırmada ele alınan ortak koşullara (kayıp verilerin bulunduğu test formu, kayıp veri oranı) ilişkin bulgular incelendiğinde araştırma sonuçlarının örtüşmediği görülmektedir. Ertoprak (2017)'in araştırmasında ele aldığı koşulların detaylarına bakıldığında; kayıp verilerin bulunduğu test formunun “her iki test, eşitlenecek test, ortak test” ve kayıp veri oranının “%10, %20, %40” olarak ele alındığı görülmüştür. “Ortak test” ve “%40” kayıp veri oranı koşulunun etkilerinin araştırmalar arasında farklılığa yol açabileceği düşünülmektedir. Ortak test deseni kullanılarak denk olmayan gruplara uygulanan farklı test formları arasındaki ölçek dönüştürme işlemi bu formlarda yer alan ortak maddeler üzerinden gerçekleştirilmektedir. Ortak testin rolü gruplar arasındaki yetenek farklılıklarını ölçmek, eşitlemedeki yanlılığı ortadan kaldırmak ve eşitleme fonksiyonunun tahmininde kesinliği artırmaktır (Angoff, 1984; Dorans ve diğerleri, 2010). Dolayısıyla ortak testte meydana gelen kayıp veriler ortalama RMSE ve BIAS değerleri üzerinde anlamlı bir etkiye yol açmış olabilir. Ayrıca “%40” kayıp veri oranı, kayıp verilerin yüksek miktarda olduğu durum olarak tanımlanmış ve kayıp veri oranı arttıkça yanlılığın da arttığı belirtilmiştir. Dolayısıyla kayıp veri oranı yüksek miktarda olduğunda BIAS değerleri üzerinde anlamlı bir etkiye yol açtığı düşünülebilir. Son olarak araştırmalar arasındaki farklılığa yol açabileceği düşünülen bir diğer faktörün simülatif olarak üretilen veriler üzerinde araştırma yapılmış olmasıdır. Simülasyon çalışmaları, gerçek veriler ile kolaylıkla ulaşamadığımız çok farklı koşulları sağlayan verilerin türetilmesine imkân veren bir yöntemdir ancak kusursuz değildir. Harris ve Crouse (1993)'a göre, gerçek veriler simüle edilmiş verilere yakın olduğunda, sonuçlar simülasyon çalışmalarının bulguları ile benzerlik göstermektedir.

Üçüncü Alt Probleme İlişkin Bulgular, Yorumlar ve Tartışma

3. Kayıp veriyle başa çıkma yöntemleri (HDA, SA, PMM, LOGREG) uygulanarak tamamlanan test formları, karakteristik eğri dönüştürme yöntemlerine (Haebara ve Stocking-Lord) göre eşitlendiğinde, referans değere en yakın ve en uzak;

a. RMSE ve BIAS değeri üreten kayıp veriyle başa çıkma yöntemi hangisidir?

b. ortalama RMSE ve BIAS değeri üreten kayıp veriyle başa çıkma yöntemi hangisidir?

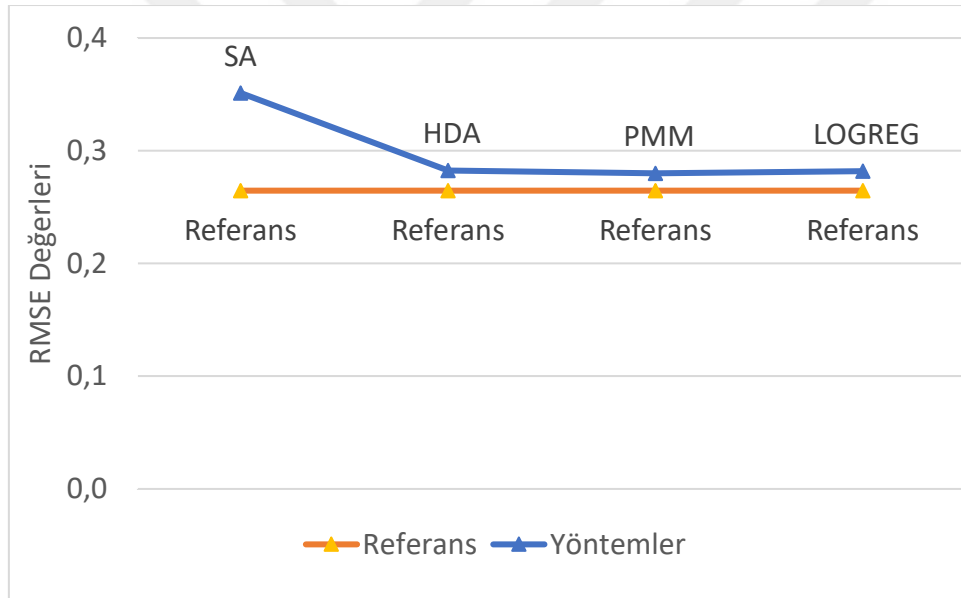
c. ortalama RMSE ve BIAS değeri üreten karakteristik eğri dönüştürme yöntemi hangisidir?

a. Bu alt problem Ek-Ç'de verilen "Analiz Sonucu Elde Edilen RMSE ve BIAS Değerleri" tablosundaki veriler kullanılarak çözümlenmiştir. Referans değere en uzak RMSE değeri üreten kayıp veri ile başa çıkma yöntemi, tüm koşullar için SA yöntemidir. Referans değere en yakın RMSE değeri üreten kayıp veri ile başa çıkma yönteminin ise 1500 örneklem büyüklüğünde, ET'te, %30 kayıp veri oranı olduğunda ve Stocking-Lord'a göre eşitleme yapıldığında LOGREG yöntemi ile elde edildiği gözlenmiştir. Bu yöntemi; 1500 örneklem büyüklüğünde, ET'te, %20 kayıp veri oranı olduğunda ve Haebara'ya göre eşitleme yapıldığında LOGREG yöntemi takip etmektedir. Referans değer olarak ele alınan tam veri setine en uzak BIAS değeri üreten kayıp veri ile başa çıkma yöntemi, genel olarak SA yöntemidir. Referans değere en yakın BIAS değeri üreten kayıp veri ile başa çıkma yönteminin ise 1500 örneklem büyüklüğünde, HİT'te, %20 kayıp veri oranı olduğunda ve Haebara'ya göre eşitleme yapıldığında PMM yöntemi ile elde edildiği gözlenmiştir. Bu yöntemi; 1500 örneklem büyüklüğünde, HİT'te, %20 kayıp veri oranı olduğunda ve Haebara'ya göre eşitleme yapıldığında LOGREG yöntemi takip etmektedir. Yani genel olarak referans RMSE ve BIAS değerine en uzak yöntemin SA, en yakın yöntemin ise LOGREG ve PMM yöntemleri olduğu bulgulanmıştır. Söz konusu bulguya ilişkin yorumlara ve tartışmaya bir sonraki alt problemde detaylı yer verilmiştir.

b. Bu alt problemin çözümlenmesinde öncelikle tam veriye sahip test formları eşitlenerek RMSE ve BIAS değerleri hesaplanmıştır. Karakteristik eğri dönüştürme yöntemlerine (Haebara ve Stocking-Lord) göre elde edilen iki farklı RMSE değerinin ortalaması 0.2646, BIAS değerinin ortalaması ise 0.0271'dir. Bu değerler, kayıp veriyle başa çıkma yöntemleri uygulanarak elde edilen test formları eşitlendiğinde hesaplanan ortalama RMSE ve BIAS değerlerinin nasıl değiştiğinin belirlenebilmesi için referans değer olarak ele alınmıştır. Eşitlemeler ile elde edilen ortalama RMSE değerleri Şekil 13'te, ortalama BIAS değerleri ise Şekil 14'te gösterilmiştir.

Şekil 13

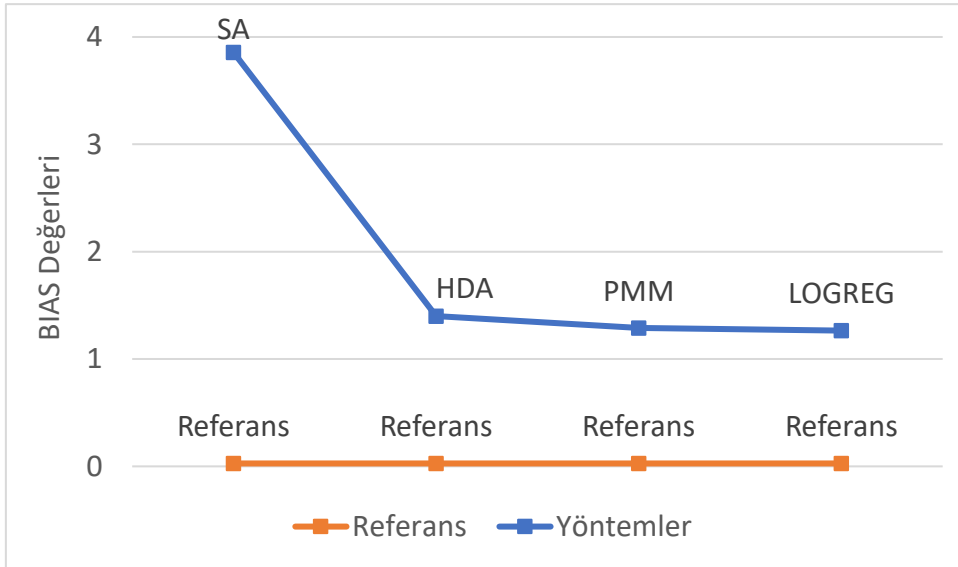
Eşitlemeler ile Elde Edilen Ortalama RMSE Değerleri



Şekil 13 incelendiğinde referans değere en yakın ortalama RMSE değeri üreten kayıp veriyle başa çıkma yönteminin PMM'in kullanıldığı koşullarda (V25-V36) elde edildiği görülmüştür. Bu yöntemi sırasıyla LOGREG (V37-V48) ve HDA'nın (V13-V24) kullanıldığı koşullardan elde edilen ortalama RMSE değerleri takip etmektedir. Referans değere en uzak ortalama RMSE değerinin ise kayıp veriyle başa çıkma yöntemi olarak SA'nın kullanıldığı koşullarda (V1-V12) üretildiği tespit edilmiştir.

Şekil 14

Eşitlemeler ile Elde Edilen Ortalama BIAS Değerleri



Şekil 14'e göre referans değere en yakın ortalama BIAS değeri üreten kayıp veriyle başa çıkma yönteminin LOGREG'in kullanıldığı koşullarda (V37-V48) elde edildiği görülmüştür. Bu yöntemi sırasıyla PMM (V25-V36) ve HDA yönteminin (V13-V24) kullanıldığı koşullardan elde edilen BIAS değerleri takip etmektedir. Referans değere en uzak ortalama BIAS değerinin ise kayıp veriyle başa çıkma yöntemi olarak SA'nın kullanıldığı koşullarda (V1-V12) üretildiği belirlenmiştir.

Şekil 13 ile Şekil 14 birlikte değerlendirildiğinde, araştırmada ele alınan koşullar altında; referans değere en yakın dolayısıyla en az hata ve en az yanlışlık üreten kayıp veriyle başa çıkma yöntemlerinin çoklu atama yöntemleri (PMM ve LOGREG) olduğu, bu yöntemleri HDA yönteminin takip ettiği görülmektedir. Referans değere en uzak dolayısıyla en yüksek hata ile en yanlış sonuçları üreten yöntem ise SA'dır. Literatürde kayıp veri ile başa çıkma yöntemlerinin test eşitlemeye etkisinin incelendiği çalışmalarda ele alınan diğer yöntemlere (tabakalaştırma sonrası eşitleme, uygulanmamış gibi davranma, beklenti maksimizasyonu, liste bazında silme, sıfır atama, kısmi bir puan atama) göre; çoklu atama yönteminin daha iyi performans gösterdiği (Ngudgratoke, 2009; Kim, 2015; Shin, 2016; Ertoprak, 2017), SA yönteminin ise performansının düşük olduğu (Shin, 2009; Shin, 2016,

Ertoprak, 2017) raporlanmıştır. Bu bulgular araştırmadan elde edilen sonucu desteklemektedir. Ayrıca kayıp veriyle başa çıkma yöntemlerinin performanslarını inceleyen literatürdeki diğer çalışmalara bakıldığında (Allison, 2003; Olinsky ve diğerleri, 2003; Finch, 2008; Garret, 2009; Graham, 2009; Finch, 2011; Young ve diğerleri, 2011; Cheema, 2012; Demir, 2013; Akbaş, 2014; Karaman, 2022) belirlenen koşullar altında genellikle en iyi ya da ikinci en iyi yöntemin çoklu atama yöntemi olduğu sonucuna ulaştıkları görülmüştür. Ancak çoklu atama yönteminin kullanımının önerilmediği koşullara (örneklem büyüklüğü çok küçük olduğunda, kayıp veri oranı çok yüksek olduğunda) da yer verilmiştir. Hohensinn ve Kubinger (2011), Demir (2013), Emenogu vd. (2010) ile Finch (2011) ise SA yönteminin hatalı ve/veya yanlı kestirimler yaptığını raporlamışlardır.

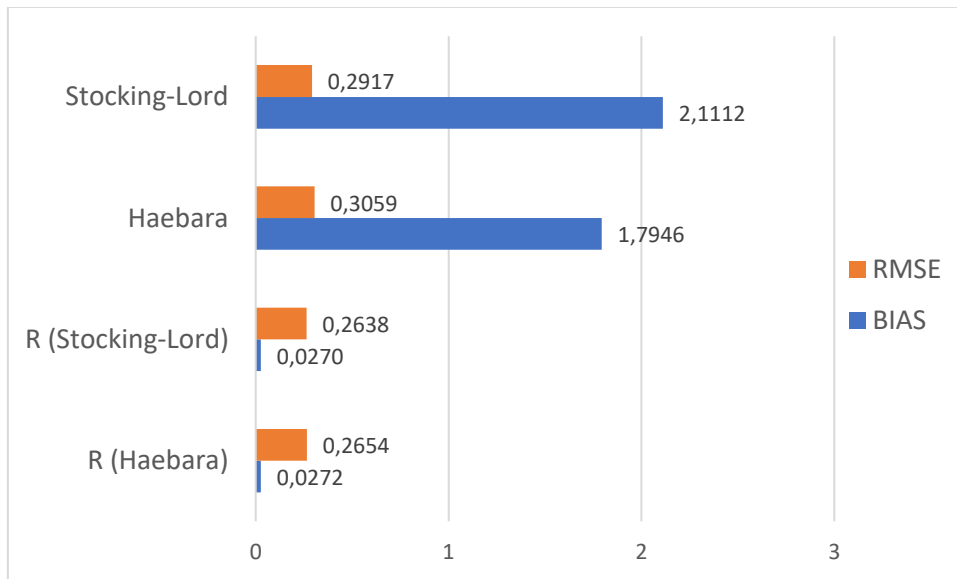
Bu alt problem için önemli bir diğer bulgu ise HDA yönteminin araştırmada ele alınan çoklu atama yöntemlerine çok yakın performans göstermesidir. Kayıp veri ile başa çıkma yöntemlerinin test eşitlemeye etkisinin incelendiği çalışmalarda HDA yöntemi ele alınmamış olsa da kayıp veriyle başa çıkma yöntemlerinin performanslarını inceleyen bazı çalışmalarda bu yönteme yer verilmiştir. Söz konusu çalışmalarda; HDA'nın çoğu koşulda (kayıp veri oranı uygulama aralığı; TRK ve RK veri mekanizması için %1-%20, ROK veri mekanizması için %1-%10) geçerli ve kullanımı kolay bir yöntem olduğu (Roth, 1994), liste bazında silme (Hawthorne & Elliott, 2005; Myers, 2011) ve ikili silme yöntemleri ile ortalama atama yöntemine göre daha iyi performans gösterdiği (Hawthorne & Elliott, 2005), değişkenler arasındaki kısıtlamalar nedeniyle model tabanlı bir yaklaşımın kullanılamadığı durumlara alternatif olabileceği ve tekli değerlendirme yöntemlerinden daha az yanlı sonuçlar sağladığı (Wang ve diğerleri, 2022) belirtilmiştir. Ancak Andridge ve Little (2010) tarafından söz konusu yöntemin teorik özelliklerinin daha fazla geliştirilmesi ve performansının değerlendirilebilmesi için çok sayıda simülasyon karşılaştırmasının yapılması önerilmiştir.

c. Bu alt problemin çözümlenmesi için öncelikle tam veriye sahip test formları karakteristik eğri dönüştürme yöntemlerine (Haebara ve Stocking-Lord) göre eşitlenmiş ve

iki farklı RMSE değeri ile BIAS değeri elde edilmiştir. RMSE(Haebara) 0.2654, RMSE(Stocking-Lord) 0.2638, BIAS(Haebara) 0.0272, BIAS(Stocking-Lord) 0.0270' dir. Bu değerler, araştırmada kayıp veriyle başa çıkma yöntemleri uygulanarak elde edilen test formlarının, tüm koşullar altında Haebara ve Stocking-Lord'a göre eşitlenmesi ile hesaplanan ortalama RMSE(Haebara), RMSE(Stocking-Lord), BIAS(Haebara) ve BIAS(Stocking-Lord) değerlerinin nasıl değiştiğinin belirlenebilmesi için referans değer olarak ele alınmıştır. R (Haebara), tam veriye sahip test formlarının Haebara'ya göre eşitlenmesi ile elde edilen referans RMSE ve BIAS değerlerini; R (Stocking-Lord), tam veriye sahip test formlarının Stocking-Lord'a göre eşitlenmesi ile elde edilen referans RMSE ve BIAS değerlerini; Haebara, kayıp veriyle başa çıkma yöntemleri uygulanarak elde edilen test formlarının, tüm koşullar altında Haebara'ya göre eşitlenmesi ile hesaplanan ortalama RMSE ve BIAS değerlerini; Stocking-Lord ise kayıp veriyle başa çıkma yöntemleri uygulanarak elde edilen test formlarının, tüm koşullar altında Stocking-Lord'a göre eşitlenmesi ile hesaplanan ortalama RMSE ve BIAS değerlerini göstermektedir. Karakteristik eğri dönüştürme yöntemlerine göre elde edilen ortalama RMSE ve BIAS değerleri Şekil 15'te verilmiştir.

Şekil 15

Yöntemlere Göre Elde Edilen Ortalama RMSE ve BIAS Değerleri



Şekil 15 incelendiğinde; hem Stocking-Lord hem de Haebara ile eşitleme yapıldığında elde edilen referans RMSE değerleri birbirine çok yakın olmakla birlikte Stocking-Lord'a göre elde edilen referans RMSE değerinin biraz daha küçük olduğu görülmüştür. Bununla birlikte, Haebara'dan ziyade Stocking-Lord'a göre eşitleme yapıldığında elde edilen referans RMSE değeri ile ortalama RMSE değerinin birbirine daha yakın olduğu tespit edilmiştir. Şekil 15 BIAS değerleri açısından incelendiğinde hem Haebara hem de Stocking-Lord'a göre elde edilen referans BIAS değerlerinin birbirine çok yakın olduğu belirlenmiştir. Ancak Haebara'ya göre eşitleme yapıldığında elde edilen referans BIAS değeri ile ortalama BIAS değerinin, Stocking-Lord'a göre elde edilen değere göre birbirine daha yakın olduğu bulgusuna ulaşılmıştır. Bu bulgulardan hareketle araştırmada ele alınan koşullar dahilinde, Stocking-Lord yöntemi kullanılarak yapılan eşitlemenin daha az hatalı, Haebara yöntemi kullanılarak yapılan eşitlemenin ise daha az yanlış olduğu söylenebilir.

Test eşitlemede kayıp veri ile başka çıkma yöntemleriyle ilgili yapılan çalışmalarda karakteristik eğri dönüştürme yöntemlerine göre etkilerini inceleyen bir çalışmaya ulaşılamamıştır. Ancak literatür incelendiğinde karakteristik eğri dönüştürme yöntemlerinin performanslarını karşılaştıran çalışmaların bazılarında Haebara, bazılarında Stocking-Lord yönteminin daha iyi performans gösterdiği, bazılarında ise her iki yöntemin de benzer performanslar gösterdiği birinci ve ikinci alt problemlerde belirtilmiştir. Farklı bir bakış açısı sunan van der Linden'a (2010) göre eşitleme literatürüne, eşitlemenin standart hatasına yönelik bir ilgi hâkim olmuştur ancak yanlışlık, bir eşitlemenin başarısını değerlendirmek için birincil kriterdir. Çünkü eşitleme, aralarındaki ölçek farklarından dolayı eski test formundaki puanın bir tahmini olarak yeni formdaki puanda oluşabilecek yanlışlığı ortadan kaldırma girişimidir. Dolayısıyla sadece eşitlemenin standart hatasına odaklanmak, eşitlenen puanlarda kalan herhangi bir önyargının hatta eşitleme sürecinde bunlara eklenen olası yeni bir önyargının fark edilmesini engeller. Tüm bulgular göz önüne alındığında

arařtırmada ele alınan kořullar detaylı incelenerek kullanılacak karakteristik eęri dnřtrme yntemine karar verilmesinin uygun olacaęı sylenebilir.



Bölüm 5

Sonuç ve Öneriler

Bu bölümde araştırmadan elde edilen sonuçlara ve bu sonuçlara ilişkin geliştirilen önerilere yer verilmiştir.

Sonuçlar

Bu araştırmanın amacı; kayıp veriyle başa çıkma yöntemlerinin (sıfır atama (SA), hot deck atama (HDA), çoklu atama (PMM ve LOGREG)), farklı koşullar altında, Madde Tepki Kuramı'na (MTK) dayalı test eşitleme yöntemlerinin (karakteristik eğri dönüştürme yöntemleri) performansı üzerindeki etkilerini incelemektir. Araştırmada ele alınan koşullar; örneklem büyüklüğü (750, 1500), kayıp veri oranı (%10, %20, %30) ve kayıp verilerin bulunduğu test formu (her iki test, eşitlenecek test)' dir. Test eşitleme yöntemlerinin performansları; hata kareleri ortalamasının karekökü (RMSE) ve eşitleme yanlılığına (BIAS) göre değerlendirilmiştir. Araştırma sonuçları aşağıda listelenmiştir.

Araştırmada Ele Alınan Koşullara Göre Elde Edilen Sonuçlar

Örneklem büyüklüğü arttıkça araştırmada ele alınan kayıp veriyle başa çıkma yöntemlerinin (SA, HDA, PMM, LOGREG) hepsi için RMSE ve BIAS değerlerinin azaldığı ve söz konusu RMSE ve BIAS değerlerinin referans değere yaklaştığı gözlenmiştir. Dolayısıyla araştırmada kullanılan tüm yöntemler için örneklem büyüklüğü 750'den 1500'e çıktığında eşitleme hatasının ve yanlılığın azaldığı, daha kararlı ve daha doğru kestirimler elde edildiği sonucuna ulaşılmıştır.

Kayıp veri oranı arttıkça araştırmada kullanılan SA yöntemi için RMSE ve BIAS değerlerinin de arttığı ve referans değerden uzaklaştığı belirlenmiştir. HDA, PMM ve LOGREG yöntemleri için ise araştırmada ele alınan %10, %20 ve %30 kayıp veri oranlarının hepsinde genel olarak RMSE ve BIAS değerlerinin birbirine çok yakın olduğu gözlenmiştir. Söz konusu RMSE ve BIAS değerleri ile referans değer arasındaki mesafenin de genel olarak değişmediği tespit edilmiştir. Dolayısıyla kayıp veri oranının %10'dan %20 ve %30'a

doğru arttığı koşullarda SA yöntemi için eşitleme hatası ve yanlılığın arttığı ancak HDA, PMM ve LOGREG yöntemleri için hata ve yanlılık değerlerinin birbirine çok yakın olduğu sonucuna ulaşılmıştır.

Kayıp verilerin bulunduğu test formuna göre değerlendirildiğinde; SA yöntemi için karakteristik eğri dönüştürme yöntemlerinden Haebara kullanıldığında her iki testte (HİT) RMSE ve BIAS değerlerinin azaldığı, Stocking-Lord kullanıldığında ise hem (eşitlenecek test) ET’te hem de HİT’te elde edilen RMSE ve BIAS değerlerinin birbirine çok yakın olduğu belirlenmiştir. HDA, PMM ve LOGREG yöntemleri için kullanılan karakteristik eğri dönüştürme yöntemlerinden bağımsız olarak hem ET’te hem de HİT’te elde edilen RMSE ve BIAS değerlerinin birbirine çok yakın olduğu gözlenmiştir. Genel olarak değerlendirildiğinde araştırmada ele alınan HDA, PMM ve LOGREG yöntemleri için HİT ve ET’in kestirilen yetenek parametrelerinden elde edilen hata ve yanlılık değerlerinin en düşük ve birbirine çok benzer olduğu tespit edilmiştir. En yüksek hata ve yanlılık değerinin ise SA yöntemi için karakteristik eğri dönüştürme yöntemlerinden Haebara kullanıldığında ET’ten elde edildiği sonucuna ulaşılmıştır.

Karakteristik eğri dönüştürme yöntemlerine göre değerlendirildiğinde; SA yöntemi için en küçük hata değerinin Stocking-Lord kullanıldığında, en küçük yanlılık değerinin ise Stocking-Lord ya da Haebara kullanıldığında elde edildiği görülmüştür. HDA, PMM ve LOGREG yöntemleri için en küçük hata değerinin Stocking-Lord ya da Haebara kullanıldığında, en küçük yanlılık değerinin ise Haebara kullanıldığında elde edildiği görülmüştür. Ancak genel olarak değerlendirildiğinde Haebara ve Stocking-Lord yöntemleri kullanıldığında elde edilen eşitleme hataları ya da yanlılıkları arasındaki farkın çok az olduğu dolayısıyla her iki yöntemin de benzer performanslar gösterdiği sonucuna ulaşılmıştır.

Araştırmada Ele Alınan Koşulların Ana Etkilerine Göre Elde Edilen ANOVA Sonuçları

Araştırmada ele alınan kayıp veriyle başa çıkma yöntemlerinin (SA, HDA, PMM ve LOGREG) hepsi için örneklem büyüklüğü (ÖB) ana etkisinin, SA yöntemi için ise kayıp veri oranı (KVO) ana etkisinin ortalama RMSE(Haebara), RMSE(Stocking-Lord) ve

BIAS(Haebara), BIAS(Stocking-Lord) değerleri üzerinde anlamlı olduğu belirlenmiştir. Ayrıca SA yöntemi için kayıp verilerin bulunduğu test formuna (KVB) ilişkin ana etkinin ortalama RMSE(Haebara) ve BIAS(Haebara) değerleri üzerinde anlamlı olduğu gözlenmiştir. Ancak ortalama RMSE ve BIAS değerlerine ilişkin ANOVA sonuçlarına göre etkileşim etkileri (ÖB*KVO, ÖB*KVB, KVO*KVB), araştırmada ele alınan hiçbir kayıp veriyle başa çıkma yöntemi için anlamlı bulunmamıştır. Özetle SA yöntemi için ÖB, KVO ve KVB ana etkilerinin; HDA, PMM ve LOGREG için ise sadece ÖB ana etkisinin ortalama RMSE(Haebara) ve BIAS(Haebara) üzerinde anlamlı olduğu sonucuna ulaşılmıştır.

Tam Veriye Sahip Test Formlarının Eşitleme Sonuçları ve Kayıp Veriyle Başa Çıkma Yöntemleri Kullanılarak Oluşturulan Test Formlarının Eşitleme Sonuçlarının Karşılaştırılması ile Elde Edilen Sonuçlar

Kayıp veriyle başa çıkma yöntemleri (HDA, SA, PMM, LOGREG) uygulanarak elde edilen test formları, karakteristik eğri dönüştürme yöntemlerine (Haebara ve Stocking-Lord) göre eşitlendiğinde;

Referans değere en yakın RMSE değeri üreten kayıp veriyle başa çıkma yönteminin 1500 örneklem büyüklüğünde, ET'te, %30 kayıp veri oranı olduğunda ve Stocking-Lord'a göre eşitleme yapıldığında LOGREG olduğu belirlenmiştir. Referans değere en yakın BIAS değeri üreten kayıp veriyle başa çıkma yönteminin ise 1500 örneklem büyüklüğünde, HİT'te, %20 kayıp veri oranı olduğunda ve Haebara'ya göre eşitleme yapıldığında PMM olduğu gözlenmiştir. Referans değer olarak ele alınan tam veri setine en uzak RMSE ve BIAS değeri üreten kayıp veri ile başa çıkma yöntemi ise SA yöntemidir.

Referans değere en yakın ortalama RMSE değeri üreten kayıp veriyle başa çıkma yöntemlerinin sırasıyla PMM, LOGREG ve HDA olduğu tespit edilmiştir. Referans değere en yakın ortalama BIAS değeri üreten kayıp veriyle başa çıkma yöntemlerinin sırasıyla LOGREG, PMM ve HDA olduğu belirlenmiştir. Referans değere en uzak ortalama RMSE ve BIAS değeri üreten kayıp veriyle başa çıkma yönteminin SA olduğu gözlenmiştir. Özetle araştırmada ele alınan koşullar altında; referans değere en yakın dolayısıyla en az hata ve

en az yanlışlık üreten kayıp veriyle başa çıkma yöntemlerinin çoklu atama yöntemleri (PMM ve LOGREG) olduğu, bu yöntemleri HDA yönteminin takip ettiği görülmektedir. Referans değere en uzak dolayısıyla en yüksek hata ile en yanlış sonuçları üreten yöntem ise SA'dır. Dolayısıyla tüm koşullar için en iyi sonucu veren mutlak bir yöntemin olmadığı, araştırmada ele alınan koşullara göre kayıp veriyle başa çıkma yöntemlerinin performanslarının değiştiği sonucuna varılmıştır.

Referans değere en yakın ortalama RMSE değeri üreten karakteristik eğri dönüştürme yönteminin Stocking-Lord, referans değere en yakın ortalama BIAS değeri üreten karakteristik eğri dönüştürme yönteminin ise Haebara olduğu görülmekle birlikte her iki yöntemden elde edilen ortalama RMSE ve BIAS değerlerinin birbirine yakın olduğu sonucuna ulaşılmıştır.

Öneriler

Bu bölümde araştırmanın sonuçlarına ve yeni araştırmalar için araştırmacılara yönelik önerilere yer verilmiştir.

Araştırmanın Sonuçlarına Yönelik Öneriler

Ulusal (ALES, YÖKDİL, YDS, ABİDE) ya da uluslararası (TIMSS, PISA, PIRLS) sınavlar; bir üst öğrenime öğrenci seçme, bir kuruma personel yerleştirme ya da ülkelerin eğitim politikalarının revize edilmesi amacıyla kullanılmakta olup bireylerin hayatlarında kritik rol oynamakta ve ülkelere önemli dönütler vermektedir. Bu tür büyük ölçekli sınavlar için farklı test formlarından elde edilen puanların anlamlı bir şekilde karşılaştırılması ya da aynı amaç ve benzer yapıları ölçmek için geliştirilen test formlarının birbiri yerine kullanılabilmesi için test eşitleme yapılmaktadır. Ancak doğru ve yansız bir eşitleme çalışması öncesinde farklı test formlarındaki muhtemel kayıp verilere ilişkin detaylar incelenerek uygun başa çıkma yöntemleri ile çözülmeli ve bu yöntemlerin test eşitlemeye yönelik kestirimlerdeki performansları incelenmelidir. Çünkü test formlarının eşit olmaması ya da bu formlarda kayıp verilerin bulunması test eşitleme sürecinin hatalı ve yanlış olmasına

yol açacaktır. Bu nedenle bu çalışmanın, kayıp veriyle başa çıkma yöntemlerinin test eşitleme üzerindeki etkileri konusunda hem test geliştirme hem de veri analizi uzmanlarına ışık tutabileceği ve verilerden maksimum düzeyde yararlanılabilmesi için atama yöntemlerini düşünmeye teşvik edeceği umulmaktadır.

Sunulan tüm öneriler araştırmada ele alınan veri setinin özellikleri (3PLM'ye uygun, tek boyutlu, normal dağılım gösteren, iki kategorili puanlanan, TRK veri mekanizmasına ve ortak test desenine sahip) ve araştırmada ele alınan koşullar için geçerlidir. Bu nedenle araştırmada ele alınacak veri setinin özelliği ve araştırma koşullarına göre kayıp veriyle başa çıkma yöntemlerinin performanslarının da değişiklik gösterebileceği göz önünde bulundurulmalıdır.

Araştırmadan elde edilen bulgu ve sonuçlara dayanarak;

- Eşitleme işleminden önce test formlarında kayıp verilerin bulunması durumunda, kayıp veri oranı ve kayıp verilerin bulunduğu test formu farketmeksizin, en uygun kayıp veriyle başa çıkma yöntemi olarak çoklu atama yöntemlerinden LOGREG ya da PMM'in kullanılması önerilmektedir.
- İki kategorili (1-0) puanlanan maddelerden oluşan test formlarında kayıp veri olması durumunda ve bu test formları TRK veri mekanizması altında eşitlenmek istendiğinde, kayıp veriyle başa çıkma yöntemlerinden HDA, çoklu atama yöntemleri ile benzer performans göstermiş olup bazı durumlarda daha küçük hata ve yanlılık değeri ürettiği görülmüştür. Bu nedenle özellikle 1500 örneklem büyüklüğünde ve kayıp verilerin her iki testte olduğu durumlarda, tüm kayıp veri oranlarında HDA yöntemi de kullanılabilir.
- Normal dağılım gösteren veri setinde; eşitlenecek test formunda özellikle örneklem büyüklüğünün 750 ve kayıp veri oranının %30 olduğu durumda, en hatalı ve en yanlı sonuçlar ürettiği için, kayıp veriyle başa çıkma yöntemi olarak SA yönteminin kullanılmaması önerilir. Ancak normal dağılım gösteren veri setinde; her iki test formunda özellikle örneklem büyüklüğünün 1500 ve kayıp veri oranının %10 olduğu

durumda SA yöntemi kullanılabilir. Uygulamada genellikle kayıp olan verilerin yanlış yanıtlanmış olarak kabul edilme eğilimi yaygın olduğu için eksik olan yanıtların kayıp veri perspektifinden titizlikle ele alınmasının gerekliliği unutulmamalıdır.

- Kayıp veriye sahip test formları eşitlenmek istendiğinde karakteristik eğri dönüştürme yöntemleri benzer performanslar gösterdikleri için herhangi birinin (Haebara ya da Stocking-Lord) kullanılması önerilmektedir.
- Tüm koşullar için en iyi sonucu veren mutlak bir yöntemin olmadığı, araştırmada ele alınan koşullara göre kayıp veriyle başa çıkma yöntemlerinin performanslarının değiştiği sonucuna varılmıştır. Bu nedenle eşitleme çalışmasından önce test formlarında kayıp veriye rastlanması durumunda öncelikle, örneklem büyüklüğü, kayıp veri oranı ve kayıp verilerin bulunduğu test formunun tespit edilmesi önerilmektedir.

Yeni Araştırmalar İçin Araştırmacılara Yönelik Öneriler

- Araştırmada kayıp veriyle başa çıkma yöntemlerinin farklı koşullar altında test eşitlemeye etkisinin incelenmesi amaçlanmış ve örneklem büyüklüğü (750, 1500), kayıp veri oranı (%10, %20, %30), kayıp verilerin bulunduğu test formu (ET, HİT) ile kayıp veriyle başa çıkma yöntemlerine (SA, HDA, PMM, LOGREG) ilişkin 48 farklı koşul ele alınmıştır. Yeni araştırmalarda farklı koşullar (test uzunluğu, ortak madde sayısı, madde parametreleri, model, grup farklılığı vb.) ele alınarak bu koşullara ilişkin kayıp veriyle başa çıkma yöntemlerinin performansları karşılaştırılabilir.
- Araştırmada sabit tutulan değişken kayıp veri mekanizmasıdır (TRK). İleride yapılacak araştırmalara RK ve ROK veri mekanizmaları da dahil edilebilir.
- Araştırma gerçek verileri üzerinde (TIMSS 2019) yapıldığı için verinin doğası gereği denk olmayan gruplarda ortak test deseni kullanılmıştır. Ayrıca her bir kitapçıkta iki kategorili (1-0) puanlanan ve 11'i ortak, 12'si ortak olmayan toplam 23 madde ele alınmıştır. Gelecekte yapılacak araştırmalarda; gerçek verilerin yanı sıra simülasyon verileri de kullanılarak elde edilecek sonuçların kıyaslanması, farklı test desenleri ve

farklı madde sayılarının da işe koşulabileceği simülasyon koşullarının oluşturularak karşılaştırmaların yapılması önerilebilir.

- Araştırmada normal dağılıma sahip bir veri seti kullanılarak kayıp veriyle başa çıkma yöntemlerinin test eşitleme üzerindeki performansları incelenmiştir. Yeni araştırmalarda sağa ve/veya sola çarpık dağılım gösteren veri setleri kullanılarak kayıp veriyle başa çıkma yöntemlerinin performansları test edilebilir.
- Araştırmada sadece 1-0 puanlanan maddeler kullanılmıştır. İleride yapılacak araştırmalarda sadece çok kategorili puanlanan maddeler ele alınabilir ya da 1-0 puanlanan ve çok kategorili puanlanan maddelerin bir arada bulunduğu veriler üzerinde çalışılabilir.
- Araştırmada ölçek dönüştürme yöntemi olarak karakteristik eğri dönüştürme yöntemleri (Haebara, Stocking-Lord) ele alınmıştır. Yeni araştırmalarda moment yöntemler ve/veya eş zamanlı kalibrasyon yöntemi de işe katılarak eşitleme işlemleri gerçekleştirilebilir.
- Araştırmada sadece MTK'ye dayalı yöntemler ele alınmıştır. İleride yapılacak araştırmalarda sadece KTK'ye dayalı yöntemler ele alınabilir ya da MTK ile KTK'ye dayalı yöntemler birlikte ele alınarak sonuçlar kıyaslanabilir.

Kaynaklar

- Ada, S. (2015). *Klasik Test Kuramına göre sayısal ve sözel alanlar için puanlama güvenirliğinin kayıp veri kapsamında incelenmesi* (Yüksek lisans tezi). Gazi Üniversitesi, Ankara.
- Agresti, A. (2007). *An introduction to categorical data analysis* (2nd ed.). John Wiley & Sons, Inc. <https://doi.org/10.1002/9780470114759.fmatter>
- Ak, D. (2020). *Genellenebilirlik Kuramında kayıp veri ile baş etme yöntemleri üzerine bir araştırma* (Yüksek lisans tezi). Ege Üniversitesi, İzmir.
- Akbaş, U. (2014). *Farklı örneklem büyüklüklerinde ve kayıp veri örüntülerinde ölçeklerin psikometrik özelliklerinin kayıp veri baş etme teknikleri ile incelenmesi* (Doktora tezi). Ankara Üniversitesi, Ankara.
- Akın Arıkan, Ç. ve Gelbal, S. (2018). The effect of mini and midi anchor tests on test equating. *International Journal of Progressive Education*, 14(29), 148-160. <https://doi.org/10.29329/ijpe.2018.139.11>
- Aksekioğlu, B. (2017). *Madde Tepki Kuramına dayalı test eşitleme yöntemlerinin karşılaştırılması: PISA 2012 fen testi örneği* (Yüksek lisans tezi). Akdeniz Üniversitesi, Antalya.
- Allison, P. D. (2002). *Missing data*. SAGE Publications.
- Allison, P. D. (2003). Missing data techniques for structural equation modeling. *Journal of Abnormal Psychology*, 112 (4), 545-557. <https://doi.org/10.1037/0021-843X.112.4.545>
- Allison, P. (2015). *Imputation by predictive mean matching: Promise & peril*. <https://statisticalhorizons.com/predictive-mean-matching/>

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME) (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Andridge, R.R., & Little, R.J.A. (2010), A review of hot deck imputation for survey non-response. *International Statistical Review*, 78(1), 40-64.
<https://doi.org/10.1111/j.1751-5823.2010.00103.x>

Angoff, W. H. (1984). *Scales, norms, and equivalent scores*. Educational Testing Service.
<https://www.ets.org/Media/Research/pdf/Angoff.Scales.Norms.Equiv.Scores.pdf>

Aşiret, S. (2014). *Küçük örneklemelerde test eşitleme yöntemlerinin çeşitli faktörlere göre incelenmesi* (Yüksek lisans tezi). Mersin Üniversitesi, Mersin.

Atalay Kabasakal, K. (2014). *Değişen madde fonksiyonunun test eşitlemeye etkisi* (Doktora tezi). Hacettepe Üniversitesi, Ankara.

Austin, P.C., & van Buuren, S. (2022). The effect of high prevalence of missing data on estimation of the coefficients of a logistic regression model when using multiple imputation. *BMC Medical Research Methodology*, 22(196), 1-14.
<https://doi.org/10.1186/s12874-022-01671-0>

Aytekin Kazanç, E. (2019). *Şans başarısının test eşitlemeye etkisinin farklı eşitleme teknikleriyle araştırılması* (Yüksek lisans tezi). Hacettepe Üniversitesi, Ankara.

Baker, F. B. (2001). *The basics of Item Response Theory* (2nd ed.). ERIC Publications.
<https://files.eric.ed.gov/fulltext/ED458219.pdf>

Banjanovic, E. S., & Osborne, J. W. (2016). Confidence intervals for effect sizes: applying bootstrap resampling. *Practical Assessment, Research & Evaluation*, 21(5), 1-20.
<https://doi.org/10.7275/dz3r-8n08>

- Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analysis. *Journal of School Psychology, 48*(1), 5–37. <https://doi.org/10.1016/j.jsp.2009.10.001>
- Barton, M. A., & Lord, F. M. (1981). An upper asymptote for the three-parameter logistic item-response model. *ETS Research Report Series, 1981*, i-8. <https://doi.org/10.1002/j.2333-8504.1981.tb01255.x>
- Bastari, B. (2000). *Linking multiple-choice and constructed-response items to a common proficiency scale* (Doctoral dissertation). Massachusetts Institute of Technology, Amherst. https://scholarworks.umass.edu/dissertations_1/5557
- Başman, M. (2014). *Likert tipi ölçeklerde kayıp verilere değer atamada yapay sinir ağlarının kullanımı* (Yüksek lisans tezi). Ankara Üniversitesi, Ankara.
- Battauz, M. (2022). *IRT equating methods (Version 2.3.0)*. <https://cran.r-project.org/web/packages/equateIRT/equateIRT.pdf>
- Bayhan, A. (2018). *Farklı koşullardaki kayıp veri oranının iç tutarlılığa etkisi* (Yüksek lisans tezi). Hacettepe Üniversitesi, Ankara.
- Bayram, İ. (2020). *Kayıp veri ile baş etme yöntemlerinin güvenilirlik kestirimleri üzerine etkisi* (Yüksek lisans tezi). Hacettepe Üniversitesi, Ankara.
- Bozdağ, S. (2007). *Şans başarısının test eşitlemeye etkisi* (Yüksek lisans tezi). Mersin Üniversitesi, Mersin.
- Bränberg, K., & Wiberg, M. (2011). Observed score linear equating with covariates. *Journal of Educational Measurement, 48*(4), 419-440. <https://doi.org/10.1111/j.1745-3984.2011.00153.x>

- Brown, R. (1994). Efficacy of the indirect approach for estimating structural equation models with missing data: A comparison of five methods. *Structural Equation Modeling*, 1(4), 287-316. <https://doi.org/10.1080/10705519409539983>
- Büyüköztürk, Ş. (2010). *Sosyal bilimler için veri analizi el kitabı*. Pegem Akademi.
- Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö.E., Karadeniz, Ş. ve Demirel, F. (2011). *Bilimsel araştırma yöntemleri* (8. baskı). Pegem Yayınları.
- Cattell, R. B. (1966). The Scree Plot Test for the Number of Factors. *Multivariate Behavioral Research*, 1, 140-161. http://dx.doi.org/10.1207/s15327906mbr0102_10
- Cheema, J. (2012). *Handling missing data in educational research using SPSS* (Doctoral dissertation). George Mason University, Washington, D.C.
- Cheema, J. R. (2014). A review of missing data handling methods in education research. *Review of Educational Research*, 84(4), 487-508. <https://doi.org/10.3102/0034654314532697>
- Chen, H. W. (2001). *Calibration of the ITBS test battery to the complete test battery: a comparison five linking methods* (Doctoral dissertation). University of Iowa, Iowa.
- Chen, S., Wang, S., & Chen, C. (2012). A simulation study using EFA and CFA programs based the impact of missing data on test dimensionality. *Expert Systems with Applications*, 39(4), 4026-4031. <https://doi.org/10.1016/j.eswa.2011.09.085>
- Chu, K. L. (2002). *Equivalent group test equating with the presence of differential item functioning* (Doctoral dissertation). The Florida State University, Florida.
- Chu, K. L., & Kamata, A. (2005). Test equating with the presence of DIF items. *Journal of Applied Measurement*, 6(3), 342-54. <https://pubmed.ncbi.nlm.nih.gov/15942075/>

- Chung, S., & Cai, L. (2019). Alternative multiple imputation inference for categorical Structural Equation Modeling. *Multivariate Behavioral Research*, 54(3), 323–337. <https://doi.org/10.1080/00273171.2018.1523000>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Cook, L.L., & Eignor, D.R. (1991). NCME instructional module: IRT equating methods. *Educational Measurement: Issues and Practice*, 10(3), 37-45.
- Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, 78(1), 98–104. <https://doi.org/10.1037/0021-9010.78.1.98>
- Crawford, A. V., Green, B. S., Levy, R., Lo, W. J., Scott, L., Svetina, D., & Thompson, M. S. (2010). Evaluation of paralel analysis methods for determining the number of factors. *Educational and Psychological Measurements*, 70(6), 885-901.
- Crocker, L., & Algina, J. (1986). *Introduction to classical & modern test theory*. Cengage Learning.
- Çokluk Bökeoğlu, Ö., Uçar, A. ve Balta, E. (2022). Madde Tepki Kuramına dayalı gerçek puan eşitlemede ölçek dönüştürme yöntemlerinin incelenmesi. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 55(1), 1-36. <https://doi.org/10.30964/auebfd.1001128>
- Çüm, S. ve Gelbal, S. (2015). Kayıp veriler yerine yaklaşık değer atamada kullanılan farklı yöntemlerin model veri uyumu üzerindeki etkisi. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, (35), 87-111. <https://dergipark.org.tr/tr/pub/maeuefd/issue/19408/206357>

De Ayala, R. J. (2009). *The theory and practice of Item Response Theory*. The Guilford Press.

De Ayala, R. J., Plake, B. S., & Impara, J. C. (2001). The impact of omitted responses on the accuracy of ability estimation in Item Response Theory. *Journal of Educational Measurement*, 38(3), 213-234. <https://doi.org/10.1111/j.1745-3984.2001.tb01124.x>

Demars, C. (2016). *Madde Tepki Kuramı* (H. Kelecioğlu, Cev.). Nobel Yayıncılık. (2010).

Demir, E. (2013). *Kayıp verilerin varlığında iki kategorili puanlanan maddelerden oluşan testlerin psikometrik özelliklerinin incelenmesi* (Doktora tezi). Ankara Üniversitesi, Ankara.

Demirus, K. B. (2015). *Ortak maddelerin değişen madde fonksiyonu gösterip göstermemesi durumunda test eşitlemeye etkisinin farklı yöntemlerle incelenmesi* (Doktora tezi). Hacettepe Üniversitesi, Ankara.

Deng, W., & Monfils, L. (2017). Long-term impact of valid case criteria on capturing population-level growth under Item Response Theory equating. *Educational Testing Service*, (1), 1-17. <https://doi.org/10.1002/ets2.12144>

Doğan, C. D. (2019). Yeniden örnekleme yöntemleri: Kavram ve R uygulamaları. *Kastamonu Eğitim Dergisi*, 27(6), 2747-2766. <https://doi.org/10.24106/kefdergi.3756>

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/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. *Educational Testing Service*, (2), i-41.
<https://doi.org/10.1002/j.2333-8504.2010.tb02236.x>
- Embretson, S. E., & Reise, S. P. (2000). *Item Response Theory for psychologists*. Lawrence Erlbaum Associates Publishers.
- Emenogu, B. C., Falenchuk, O., & Childs, R. A. (2010). The effect of missing data treatment on Mantel-Haenszel DIF detection. *The Alberta Journal of Educational Research*, 56(4), 459-469. <https://doi.org/10.11575/ajer.v56i4.55429>
- Enders, C. K. (2010). *Applied missing data analysis*. The Guilford Publications, Inc.
- Ertoprak, D. G. (2017). *Kayıp verinin test eşitlemeye etkisinin incelenmesi* (Doktora tezi). Hacettepe Üniversitesi, Ankara.
- Finch, H. (2008). Estimation of Item Response Theory parameters in the presence of missing data. *Journal of Educational Measurement*, 45(3), 225-245.
<https://doi.org/10.1111/j.1745-3984.2008.00062.x>
- Finch, W. H. (2011). The impact of missing data on the detection of nonuniform differential item functioning. *Educational and Psychological Measurement*, 71(4), 663-683.
<https://doi.org/10.1177/0013164410385226>
- Furlow, C. F., Fouladi, R., Gagné, P., & Whittaker, T. (2007). A Monte Carlo study of the impact of missing data and differential item functioning on theta estimates under two polytomous Rasch family models. *Journal of Applied Measurement*, 8(4), 388-403.
- Garrett, P. (2009). *A monte carlo study investigating missing data, differential item functioning and effect size* (Doctoral dissertation). Georgia State Üniversitesi, Atlanta.

- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, 60(1), 549-576.
<https://doi.org/10.1146/annurev.psych.58.110405.085530>
- Gonzalez, E., & Rutkowski, L. (2010). Principles of multiple matrix booklet designs and parameter recovery in large-scale assessments. *IEA-ETS Research Institute Monograph*, 3, 125-156.
- Gonzales, J., & Wiberg, M. (2017). *Applying test equating methods using R*. Springer.
- Gök, B. (2012). *Denk olmayan gruplarda ortak madde deseni kullanılarak madde tepki kuramına dayalı eşitleme yöntemlerinin karşılaştırılması* (Doktora tezi). Hacettepe Üniversitesi, Ankara.
- Gündüz, T. (2015). *Test eşitlemede Madde Tepki Kuramına dayalı yetenek parametresine yönelik ölçek dönüştürme yöntemlerinin karşılaştırılması* (Yüksek lisans tezi). Gazi Üniversitesi, Ankara.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item Response Theory: Principles and applications*. Kluwer, Nijhoff Publishing.
- Hambleton, R. K., Swaminathan, R., & Rogers, H. J. (1991). *Fundamentals of Item Response Theory*. Sage Publication.
- 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.
- Han, K.T. (2008). *Impact of item parameter drift on test equating and proficiency estimates* (Doctoral dissertation). University of Massachusetts, Amherst.

- Hanson, B. A., Zeng, L., & Colton, D. A. (1994). *A comparison of presmoothing and posts smoothing methods in equipercentile equating*. (ACT Research Report Series, 94-4). American College Testing Program.
- Hanson, B. A., & Beguin, A. A. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common-item equating design. *Applied Psychological Measurement*, 26(1), 3-24.
<https://doi.org/10.1177/0146621602026001001>
- Harris, D. J., & Crouse, J. D. (1993). A study of criteria used in equating. *Applied Measurement in Education*, 6 (3), 195–240.
https://doi.org/10.1207/s15324818ame0603_3
- Hawthorne, G., & Elliott, P. (2005). Imputing cross-sectional missing data: Comparison of common techniques. *Australian and New Zealand Journal of Psychiatry*, 39(7), 583–590. <https://doi.org/10.1080/j.1440-1614.2005.01630.x>
- He, Y. (2013). *Robust scale transformation methods in IRT true score equating under common-item nonequivalent groups design* (Doctoral dissertation). University of Missouri, Columbia City.
- Hohensinn, C., & Kubinger K. D. (2011). On the impact of missing values on item fit and the model validness of the Rasch model. *Psychological Test and Assessment Modeling*, 53(3), 380-393.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179-185.
- Hosmer, D.W., & Lemeshow S. (2000). *Applied logistic regression* (2nd ed.). John Wiley and Sons.

- Hu, H., Rover, W. T., & Vukmirovic, Z. (2008). Investigation of IRT-based equating methods in the presence of outlier common items. *Applied Psychological Measurement*, 32(4), 311-333. <https://doi.org/10.1177/0146621606292215>
- Huggins, A. C. (2014). The effect of differential item functioning in anchor items on population invariance of equating. *Educational and Psychological Measurement*, 74(4), 627-658. <https://doi.org/10.1177/0013164413506222>
- Işıkoğlu, M. A. (2017). *Kayıp veri ile baş etme yöntemlerinin ölçme değişmezliğine etkisi açısından karşılaştırılması (Yüksek lisans tezi)*. Hacettepe Üniversitesi, Ankara.
- İnci, Y. (2014). *Örneklem büyüklüğünün test eşitlemeye etkisi (Yüksek lisans tezi)*. Hacettepe Üniversitesi, Ankara.
- Jia, F., & Wu, W. (2019). Evaluating methods for handling missing ordinal data in structural equation modeling. *Behavior Research Methods*, 51, 2337-2355. <https://doi.org/10.3758/s13428-018-1187-4>
- Kaiser, H. F. (1960). The applications of electronic computer to factor analysis. *Educational and Psychological Measurement*, 20, 141-151.
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39, 31-36.
- Kang, T., & Petersen, N. S. (2011). Linking item parameters to a base scale. *Asia Pacific Educ. Rev.*, 13, 311–321. <https://doi.org/10.1007/s12564-011-9197-2>
- Karagül, A. E. (2020). *Küçük örneklemelerde çok kategorili puanlanan maddelerden oluşan testlerde klasik test eşitleme yöntemlerinin karşılaştırılması (Yüksek lisans tezi)*. Ankara Üniversitesi, Ankara.
- Karaman, H. (2022). *Kayıp veri ile başa çıkma yöntemlerinin Yapısal Eşitlik Modellerine etkisi (Doktora tezi)*. Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.

- Karkee, T. B., & Wright, K. R. (2004, April 16). *Evaluation of linking methods for placing three parameter logistic item parameter estimates onto a one parameter scale*. [Conference presentation abstract] American Educational Research, San Diego, California. <https://files.eric.ed.gov/fulltext/ED491663.pdf>
- Keller, L. A., & Keller, R. R. (2011). The long-term sustainability of different Item Response Theory scaling methods. *Educational and Psychological Measurement*, 71(2), 362–379. <https://doi.org/10.1177/0013164410375111>
- Kleinke, K. (2017). Multiple imputation under violated distributional assumptions: A systematic evaluation of the assumed robustness of predictive mean matching. *Journal of Educational and Behavioral Statistics*, 42(4), 371–404. <https://doi.org/10.3102/1076998616687084>
- Kilmen, S. (2010). *Madde Tepki Kuramı'na dayalı test eşitleme yöntemlerinden kestirilen eşitleme hatalarının örneklem büyüklüğü ve yetenek dağılımına göre karşılaştırılması* (Doktora tezi). Ankara Üniversitesi, Ankara.
- Kim, M. S. (2015). *Linking with planned missing data: concurrent calibration with multiple imputation* (Doctoral dissertation). The University of Kansas, Kansas.
- Kim, S., & Kolen, M. J. (2006). Robustness to format effects of IRT linking methods for mixed-format tests. *Applied Measurement in Education*, 19(4), 357-381. https://doi.org/10.1207/s15324818ame1904_7
- Kim, S., & Lee, W. (2006). An extension of four IRT linking methods for mixed format tests. *Journal of Educational Measurement*, 43(1), 53-76. <https://www.jstor.org/stable/20461809>

- Koçak, D. (2016). *Kayıp veriyle baş etme yöntemlerinin Madde Tepki Kuramı bir parametrelili lojistik modelinde model veri uyumuna ve standart hataya etkisi* (Doktora tezi). Ankara Üniversitesi, Ankara.
- Koçak, D., Çokluk, Ö. ve Kayri, M. (2016). Faktör sayısının belirlenmesinde MAP testi, paralel analiz, K1 ve yamaç birikinti grafiği yöntemlerinin karşılaştırılması. *Van Yüzüncü Yıl Üniversitesi Eğitim Fakültesi Dergisi*, 13(1), 330-359.
- Koçak, D. ve Çokluk Bökeoğlu, Ö. (2017). Kayıp veriyle baş etme yöntemlerinin model veri uyumu ve madde model uyumuna etkisi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 8(2), 200-223. <https://doi.org/10.21031/epod.303753>
- Kolen, M. J. (1981). Comparison of traditional and item response theory methods for equating tests. *Journal of Educational Measurement*, 18(1), 1-11. <https://doi.org/10.1111/j.1745-3984.1981.tb00838.x>
- Kolen, M. J. (1988). An NCME instructional module on traditional equating methodology. *Educational Measurement: Issues and Practice*, 7, 29–36.
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd. ed.). Springer.
- Kowarik, A., & Templ, M. (2016). Imputation with the R Package VIM. *Journal of Statistical Software*, 74(7), 1–16. <https://doi.org/10.18637/jss.v074.i07>
- Lee, W. C., & Ban, J. C. (2010). A comparison of IRT linking procedures. *Applied Measurement in Education*, 23(1), 23-48. <https://doi.org/10.1080/08957340903423537>
- Leite, W. & Beretvas, S.N. (2010). The performance of multiple imputation for likert type items with missing data. *Journal of Modern Applied Statistical Methods*, 9(1), 64-74. DOI: 10.22237/jmasm/1272686820

- Li, C. (2013). Little's test of missing completely at random. *The Stata Journal*, 13(4), 795-809. <https://doi.org/10.1177/1536867X1301300407>
- Li, D., Jiang, Y., & von Davier, A. A. (2012). The accuracy and consistency of series of IRT true score equatings. *Journal of Educational Measurement*, 49(2), 167-189.
- Li, J., & Lomax, R. G. (2016). Effects of missing data methods in sem under conditions of incomplete and nonnormal data. *The Journal of Experimental Education*, 85(2), 231-258. <https://doi.org/10.1080/00220973.2015.1134418>
- Lim, A. J. M., & Cheung, M. W. L. (2022). Evaluating FIML and multiple imputation in joint ordinal-continuous measurements models with missing data. *Behavior Research Methods*, 54(3), 1063-1077. DOI: 10.3758/s13428-021-01582-w
- Little, R.J.A. (1988). Missing data adjustments in large surveys. *Journal of Business & Economic Statistics*, 6(3), 287–296. DOI:10.1080/07350015.1988.10509663
- Little, R., & Rubin, D. (2002). *Statistical analysis with missing data* (2nd. ed.). John Wiley & Sons.
- Liu, C., & Kolen, M.J. (2011). Evaluating smoothing in equipercentile equating using fixed smoothing parameters. In M. J. Kolen & W. Lee (Eds.), *Mixed-format tests: Psychometric properties with a primary focus on equating*: Vol. 1. (pp. 213–236). Center for Advanced Studies in Measurement and Assessment (CASMA). <https://education.uiowa.edu/sites/education.uiowa.edu/files/2021-11/casma-monograph-2.1.pdf>
- Liu, Y. ve Sriutaisuk, S. (2019). Evaluation of model fit in Structural Equation Models with ordinal missing data: An examination of the D2 Method. *Structural Equation*

Modeling: A Multidisciplinary Journal, 27(4), 561-583.
doi:10.1080/10705511.2019.1662307

Lord, F. M. (1980). *Applications of Item response theory to practical testing problems*. Erlbaum.

Loyd, B. H., & Hoover, H. D. (1980). Vertical equating using the Rasch model. *Journal of Educational Measurement*, 17(3), 179-193. <https://doi.org/10.1111/j.1745-3984.1980.tb00825.x>

Ma, J., Akhtar-Danesh, N., Dolovich, L., & Thabane, L. (2011). Imputation strategies for missing binary outcomes in cluster randomized trials. *BMC Medical Research Methodology*, 11(18), 1471-2288. <https://doi.org/10.1186/1471-2288-11-18>

Marco, G. L. (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement*, 14(2), 139-160. <https://doi.org/10.1111/j.1745-3984.1977.tb00033.x>

Marsh, H. W. (1998). Pairwise deletion for missing data in structural equation models: Nonpositive definite matrices, parameter estimates, goodness of fit, and adjusted sample sizes. *Structural Equation Modeling: A Multidisciplinary Journal*, 5(1), 22-36. <https://doi.org/10.1080/10705519809540087>

Martin, M. O., von Davier, M., & Mullis, I. V. S. (Eds.). (2020). *Methods and procedures: TIMSS 2019 technical report*. Boston College, TIMSS & PIRLS International Study Center. <https://timssandpirls.bc.edu/timss2019/methods>

Matei, A., & Tillé, Y. (2021). *Survey sampling (Version 2.9)*. <https://cran.r-project.org/web/packages/sampling/sampling.pdf>

McKnight, P. E., McKnight, K. M., Sidani, S., & Figueredo, A. J. (2007). *Missing data: A gentle introduction*. The Guilford Press.

- Meng, Y. (2012). *Comparison of Kernel Equating and Item Response Theory equating methods* (Doctoral dissertation). University of Massachusetts, Amherst.
- Morris, T.P., White, I.R., & Royston, P. (2015). Tuning multiple imputation by predictive mean matching and local residual draws. *BMC Medical Research Methodology*, 14, 1-13. <https://doi.org/10.1186/1471-2288-14-75>
- Mullis, I. V. S., & Martin, M. O. (Eds.). (2017). *TIMSS 2019 assessment frameworks*. Boston College, TIMSS & PIRLS International Study Center. <http://timssandpirls.bc.edu/timss2019/frameworks/>
- Mullis, I. V. S., Martin, M. O., Foy, P., Kelly, D. L., & Fishbein, B. (2020). *TIMSS 2019 international results in mathematics and science*. Boston College, TIMSS & PIRLS International Study Center. <https://timssandpirls.bc.edu/timss2019/international-results/>
- Mutluer, C. (2013). *Yıl içinde farklı Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı (ALES) puanlarına ilişkin bir test eşitleme çalışması* (Yüksek lisans tezi). Abant İzzet Baysal Üniversitesi, Bolu.
- Mutluer, C. (2021). *Klasik Test Kuramına ve Madde Tepki Kuramına dayalı test eşitleme yöntemlerinin karşılaştırılması: Uluslararası Öğrenci Değerlendirme Programı (PISA) 2012 matematik testi örneği* (Doktora tezi). Gazi Üniversitesi, Ankara.
- Myers, T. A. (2011). Goodbye, listwise deletion: Presenting hot deck imputation as an easy and effective tool for handling missing data. *Communication Methods and Measures*, 5(4), 297-310. <https://doi.org/10.1080/19312458.2011.624490>
- Ngudgratoke, S. (2009). *An investigation of using collateral information to reduce equating biases of the post-stratification equating method* (Doctoral dissertation). Michigan State University, Michigan.

- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Ogasawara, H. (2001). Standard errors of Item Response Theory equating/linking by response function methods. *Applied Psychological Measurement*, 25(1), 53-67.
- Olinsky, A., Chen, S., & Harlow, L. (2003). The comparative efficacy of imputation methods for missing data in structural equation modeling. *European Journal of Operational Research*, 151(1), 53-79. [https://doi.org/10.1016/S0377-2217\(02\)00578-7](https://doi.org/10.1016/S0377-2217(02)00578-7)
- Öztemür, B. (2014). *Kayıp veri yöntemlerinin farklı değişkenler altında varyans analizi (T-testi, Anova) parametreleri üzerine etkisinin incelenmesi* (Yüksek lisans tezi). Abant İzzet Baysal Üniversitesi, Bolu.
- Pallant, J. (2016). *SPSS kullanma kılavuzu SPSS ile adım adım veri analizi* (S. Balcı ve B. Ahi, Cev.). Anı Yayıncılık. (2015).
- Patz, R., & Hanson, B. (2002). *Psychometric issues in vertical scaling*. [Conference presentation]. The National Council on Measurement in Education, New Orleans, LA.
- Peng, C.Y., Harwell, M., Liou, S.M., & Ehman, L. H. (2007). Advances in missing data methods and implication for educational research. In S. S. Sawilowski (Ed.), *Real data analysis* (pp. 31-78). Information Age Publishing.
- Pokropek, A. (2011). Missing by design: Planned missing-datadesigns in social science. *Research & Methods*, 20(1), 81–105. <https://www.researchgate.net/publication/250615619>
- Powers, S. J. (2010). *Impact of matched samples equating methods on equating accuracy and the adequacy of equating assumptions* (Doctoral dissertation). The University of Iowa, Iowa.

- R Core Team,. & Contributors worldwide. (2022). *The R base package (Version 4.3.0)*.
<https://stat.ethz.ch/R-manual/R-devel/library/base/DESCRIPTION>
- Revelle, W., & Rocklin, T. (1979). Very simple structure: An alternative procedure for estimating the optimal number of interpretable factors. *Multivariate Behavioral Research*, 14(3), 403-414.
- Revelle, W. (2023). *Psych: Procedures for psychological, psychometric, and personality research*. (Version 2.3.3). <https://cran.r-project.org/web/packages/psych/psych.pdf>
- Rizopoulos, D. (2022). *Latent Trait Models under IRT (Version 1.2-0)*. <https://cran.r-project.org/web/packages/lrm/lrm.pdf>
- Robitzsch, A. (2022). *Supplementary Item Response Theory models (Version 3.12-66)*.
<https://cran.r-project.org/web/packages/sirt/sirt.pdf>
- Rockel, T. (2022). *Methods for missing data (Version 0.4.0)*. <https://cran.r-project.org/web/packages/missMethods/missMethods.pdf>
- Rose, N., von Davier, M., & Xu, X. (2010). Modeling nonignorable missing data with item response theory (IRT). *ETS Research Report Series*, 2010(1), i-53.
<https://doi.org/10.1002/j.2333-8504.2010.tb02218.x>
- Roth, P. L. (1994). Missing data: A conceptual review for applied psychologists. *Personnel Psychology*, 47(3), 537–560. <https://doi.org/10.1111/j.1744-6570.1994.tb01736.x>
- Rubin, D. B., (1976). Inference and missing data. *Biometrika*, 63(3), 581-592.
<https://doi.org/10.1093/biomet/63.3.581>

- Rubin, D. B. (1987). *Multiple imputation for nonresponse in surveys*. John Wiley & Sons.
- Salvucci, S., Walter, E., Conley, V., Fink, S., & Saba, M. (1997). *Measurement error studies at the National Center for Education Statistics*. U. S. Department of Education. <https://files.eric.ed.gov/fulltext/ED410313.pdf>
- Schmeiser, C., & Welch, C.J. (2006). Test development. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 307–353). American Council on Education and Praeger.
- Schafer, J. L., & Graham, J. W. (2002). Missing data: Our view of the state of the art. *Psychological Methods*, 7(2), 147-177. <https://doi.org/10.1037/1082-989X.7.2.147>
- Shin, S. H. (2009). How to treat omitted responses in Rasch model-based equating. *Practical Assessment Research & Evaluation*, 14(1), 1-8. <https://doi.org/10.7275/x9vv-xq85>
- Shin, A. Y. (2016). Investigating the effects of missing data treatments on Item Response Theory vertical scaling (Doctoral dissertation). The University of Iowa, Iowa.
- Siddique, J., & Belin, T. R. (2008). Multiple imputation using an iterative hot-deck with distance-based donor selection. *Statistics in Medicine*, 27(1), 83–102. <https://doi.org/10.1002/sim.3001>
- Spence, P. D. (1996). *The effect of multidimensionality on unidimensional equating with Item Response Theory* (Doctoral dissertation). University of Florida, Florida.
- Speron, E. (2009). *A comparison of metric linking procedures in Item Response Theory* (Doctoral dissertation). University of Illinois, Chicago.

- Sulis, I., & Porcu, M. (2008). *Assessing the effectiveness of a stochastic regression imputation method for ordered categorical data*. Center for North South Economic Research, 4, 1-29. <https://crenos.unica.it/crenos/sites/default/files/wp/08-04.pdf>
- Şahin Kürşad, M. (2014). *Sıklıkla kullanılan kayıp veri yöntemlerinin betimsel istatistik, güvenirlik ve geçerlik açısından karşılaştırılması* (Yüksek lisans tezi). Abant İzzet Baysal Üniversitesi, Bolu.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson.
- Tamcı, P. (2018). *Kayıp veriyle başa çıkma yöntemlerinin değişen madde fonksiyonu üzerindeki etkisinin incelenmesi* (Yüksek lisans tezi). Hacettepe Üniversitesi, Ankara.
- Tanberkan Suna, H. (2018). *Grup değişmezliği özelliğinin farklı eşitleme yöntemlerinde eşitleme fonksiyonları üzerindeki etkisi* (Doktora tezi). Gazi Üniversitesi, Ankara.
- Templ, M., Kowarik, A., Alfons, A., de Cillia, G., Prantner, B., & Rannethbauer, W. (2022). *VIM (Version 6.2.2)*. <https://cran.r-project.org/web/packages/VIM/VIM.pdf>
- Tian, F. (2011). *A comparison of equating/linking using the Stocking-Lord method and concurrent calibration with mixed-format tests in the non-equivalent groups commonitem design under IRT* (Doctoral dissertation). Boston Collage, Boston.
- Toka, O. (2012). *Kayıp veri durumunda sağlam kestirim* (Yüksek lisans tezi). Hacettepe Üniversitesi, Ankara.
- Tong, Y., & Kolen, M. (2010). Scaling: An ITEMS Module. *Educational Measurement: Issues and Practice*, 29(4), 39-48. <https://doi.org/10.1111/j.1745-3992.2010.00192.x>

- Tucker, L.R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika*, 34(4), 421-459.
- Uysal, İ. (2014). *Madde Tepki Kuramı'na dayalı test eşitleme yöntemlerinin karma modeller üzerinde karşılaştırılması* (Yüksek lisans tezi). Abant İzzet Baysal Üniversitesi, Bolu.
- Uysal, İ. (2019). *Açık uçlu maddelerde otomatik puanlamanın güvenilirliği ve test eşitleme hatalarına etkisi* (Doktora tezi). Hacettepe Üniversitesi, Ankara.
- van Buuren, S., Groothuis-Oudshoorn, K. (2011). mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3), 1–67. <https://doi.org/10.18637/jss.v045.i03>
- van Buuren, S. (2018). *Flexible imputation of missing data* (2nd ed.). Chapman and Hall/CRC.
- van Buuren, S. (2022). *Multivariate imputation by chained equations* (Version 3.15.0). <https://cran.r-project.org/web/packages/mice/mice.pdf>
- van der Linden, W. J. (2010). On bias in linear observed-score equating. *Measurement: Interdisciplinary Research and Perspectives*, 8(1), 21-26. <https://doi.org/10.1080/15366361003684711>
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41(3), 321-327.
- Wang, C., Stokes, T., Steele, R. J., Wedderkopp, N., & Shrier, I. (2022) Implementing multiple imputation for missing data in longitudinal studies when models are not feasible: An example using the random hot deck approach. *Clinical Epidemiology*, 14, 1387-1403. <https://doi.org/10.2147/CLEP.S368303>

- Weaver, B., & Maxwell, H. (2014). Exploratory factor analysis and reliability analysis with missing data: A simple method for SPSS users. *The Quantitative Methods for Psychology*, 10(2) 143-152. <https://doi.org/10.20982/tqmp.10.2.p143>
- Wiberg, M., & Branberg, K. (2015). Kernel equating under the non-equivalent groups with covariates design. *Applied Psychological Measurement*, 39(5), 349-361. <https://doi.org/10.1177/0146621614567939>
- Wu, W., Jia, F., & Enders, C. (2015). A comparison of imputation strategies for ordinal missing data on likert scale variables. *Multivariate Behavioral Research*, 50(5), 484–503. <https://doi.org/10.1080/00273171.2015.1022644>
- Young, W., Weckman, G., & Holland, W. (2011). A survey of methodologies for the treatment of missing values within datasets: Limitations and benefits. *Theoretical Issues in Ergonomics Science*, 12(1), 15-43. <https://doi.org/10.1080/14639220903470205>
- Yurtçu, M. (2018). *Parametrik olmayan Bayes yöntemiyle ortak değişkenlere göre yapılan test eşitlemelerinin karşılaştırılması* (Doktora tezi). Hacettepe Üniversitesi, Ankara.
- Zhang, Z. (2010). *Comparison of different equating methods and an application to link testlet-based tests*. (Doctoral dissertation). Villanova University, Philadelphia.
- Zhu, X. (2014). Comparison of four methods for handling missing data in longitudinal data analysis through a simulation study. *Open Journal of Statistics*, 4(11), 933-944. <https://doi.org/10.4236/ojs.2014.411088>

EK-A: Yerel Bağımsızlık Testi Sonuçları

Madde Çifti Sayısı	Madde Çiftleri		Kitapçık 6 için Q_3	Kitapçık 7 için Q_3
1	Madde 1	Madde 2	-0.001	-0.032
2	Madde 1	Madde 3	0.003	-0.016
3	Madde 1	Madde 4	-0.002	-0.007
4	Madde 1	Madde 5	-0.007	-0.034
5	Madde 1	Madde 6	0.001	-0.001
6	Madde 1	Madde 7	-0.004	-0.019
7	Madde 1	Madde 8	-0.043	-0.040
8	Madde 1	Madde 9	-0.055	-0.021
9	Madde 1	Madde 10	-0.036	-0.014
10	Madde 1	Madde 11	-0.032	-0.046
11	Madde 1	Madde 12	-0.030	-0.010
12	Madde 1	Madde 13	0.019	-0.020
13	Madde 1	Madde 14	-0.046	-0.013
14	Madde 1	Madde 15	-0.070	-0.022
15	Madde 1	Madde 15	0.007	-0.013
16	Madde 1	Madde 17	-0.064	-0.061
17	Madde 1	Madde 18	-0.007	-0.024
18	Madde 1	Madde 19	-0.034	-0.042
19	Madde 1	Madde 20	-0.042	-0.029
20	Madde 1	Madde 21	-0.018	-0.013
21	Madde 1	Madde 22	-0.037	-0.024
22	Madde 1	Madde 23	-0.010	-0.004
23	Madde 2	Madde 3	-0.033	-0.017
24	Madde 2	Madde 4	0.015	0.024
25	Madde 2	Madde 5	-0.066	-0.043
26	Madde 2	Madde 6	-0.039	-0.039
27	Madde 2	Madde 7	-0.054	-0.018
28	Madde 2	Madde 8	-0.043	-0.019
29	Madde 2	Madde 9	-0.018	-0.019
30	Madde 2	Madde 10	0.007	-0.021
31	Madde 2	Madde 11	-0.048	-0.076
32	Madde 2	Madde 12	-0.032	-0.034
33	Madde 2	Madde 13	-0.011	-0.026
34	Madde 2	Madde 14	-0.010	-0.031
35	Madde 2	Madde 15	-0.081	-0.035

36	Madde 2	Madde 16	-0.035	-0.003
37	Madde 2	Madde 17	-0.065	-0.050
38	Madde 2	Madde 18	-0.026	-0.046
39	Madde 2	Madde 19	-0.024	-0.013
40	Madde 2	Madde 20	-0.038	-0.002
41	Madde 2	Madde 21	0.032	0.007
42	Madde 2	Madde 22	0.007	-0.064
43	Madde 2	Madde 23	0.029	-0.028
44	Madde 3	Madde 4	-0.005	0.002
45	Madde 3	Madde 5	-0.041	-0.063
46	Madde 3	Madde 6	-0.039	-0.032
47	Madde 3	Madde 7	-0.054	-0.021
48	Madde 3	Madde 8	-0.042	-0.013
49	Madde 3	Madde 9	-0.030	-0.033
50	Madde 3	Madde 10	-0.039	-0.055
51	Madde 3	Madde 11	-0.053	-0.073
52	Madde 3	Madde 12	-0.035	-0.030
53	Madde 3	Madde 13	-0.003	-0.033
54	Madde 3	Madde 14	-0.054	-0.010
55	Madde 3	Madde 15	-0.043	-0.043
56	Madde 3	Madde 16	-0.005	-0.024
57	Madde 3	Madde 17	-0.015	-0.026
58	Madde 3	Madde 18	-0.029	-0.061
59	Madde 3	Madde 19	-0.024	-0.043
60	Madde 3	Madde 20	-0.051	-0.027
61	Madde 3	Madde 21	-0.018	-0.003
62	Madde 3	Madde 22	-0.024	-0.020
63	Madde 3	Madde 23	0.035	0.003
64	Madde 4	Madde 5	-0.054	-0.067
65	Madde 4	Madde 6	-0.029	0.010
66	Madde 4	Madde 7	-0.017	0.004
67	Madde 4	Madde 8	-0.060	-0.058
68	Madde 4	Madde 9	-0.020	-0.045
69	Madde 4	Madde 10	-0.072	-0.064
70	Madde 4	Madde 11	-0.083	-0.073
71	Madde 4	Madde 12	-0.013	-0.060
72	Madde 4	Madde 13	0.021	-0.054
73	Madde 4	Madde 14	-0.033	-0.016
74	Madde 4	Madde 15	-0.108	-0.014
75	Madde 4	Madde 16	-0.059	-0.010

76	Madde 4	Madde 17	-0.034	0.002
77	Madde 4	Madde 18	-0.092	-0.043
78	Madde 4	Madde 19	-0.034	-0.064
79	Madde 4	Madde 20	-0.074	-0.005
80	Madde 4	Madde 21	-0.024	0.017
81	Madde 4	Madde 22	0.007	-0.115
82	Madde 4	Madde 23	0.009	-0.104
83	Madde 5	Madde 6	-0.013	0.024
84	Madde 5	Madde 7	-0.020	-0.046
85	Madde 5	Madde 8	-0.037	-0.032
86	Madde 5	Madde 9	-0.031	-0.012
87	Madde 5	Madde 10	-0.063	-0.029
88	Madde 5	Madde 11	-0.038	-0.045
89	Madde 5	Madde 12	-0.060	-0.078
90	Madde 5	Madde 13	-0.014	-0.020
91	Madde 5	Madde 14	-0.071	-0.020
92	Madde 5	Madde 15	-0.065	-0.069
93	Madde 5	Madde 16	-0.051	-0.031
94	Madde 5	Madde 17	-0.048	-0.070
95	Madde 5	Madde 18	-0.016	-0.049
96	Madde 5	Madde 19	-0.007	-0.062
97	Madde 5	Madde 20	-0.119	-0.079
98	Madde 5	Madde 21	-0.023	-0.026
99	Madde 5	Madde 22	-0.087	-0.042
100	Madde 5	Madde 23	-0.030	-0.077
101	Madde 6	Madde 7	-0.027	-0.040
102	Madde 6	Madde 8	-0.049	-0.044
103	Madde 6	Madde 9	-0.034	0.017
104	Madde 6	Madde 10	-0.007	0.019
105	Madde 6	Madde 11	-0.048	-0.012
106	Madde 6	Madde 12	-0.050	-0.005
107	Madde 6	Madde 13	-0.023	-0.031
108	Madde 6	Madde 14	-0.023	-0.013
109	Madde 6	Madde 15	-0.041	-0.034
110	Madde 6	Madde 16	-0.039	0.009
111	Madde 6	Madde 17	-0.039	-0.071
112	Madde 6	Madde 18	-0.058	-0.079
113	Madde 6	Madde 19	-0.012	-0.065
114	Madde 6	Madde 20	-0.042	-0.030
115	Madde 6	Madde 21	-0.038	-0.040

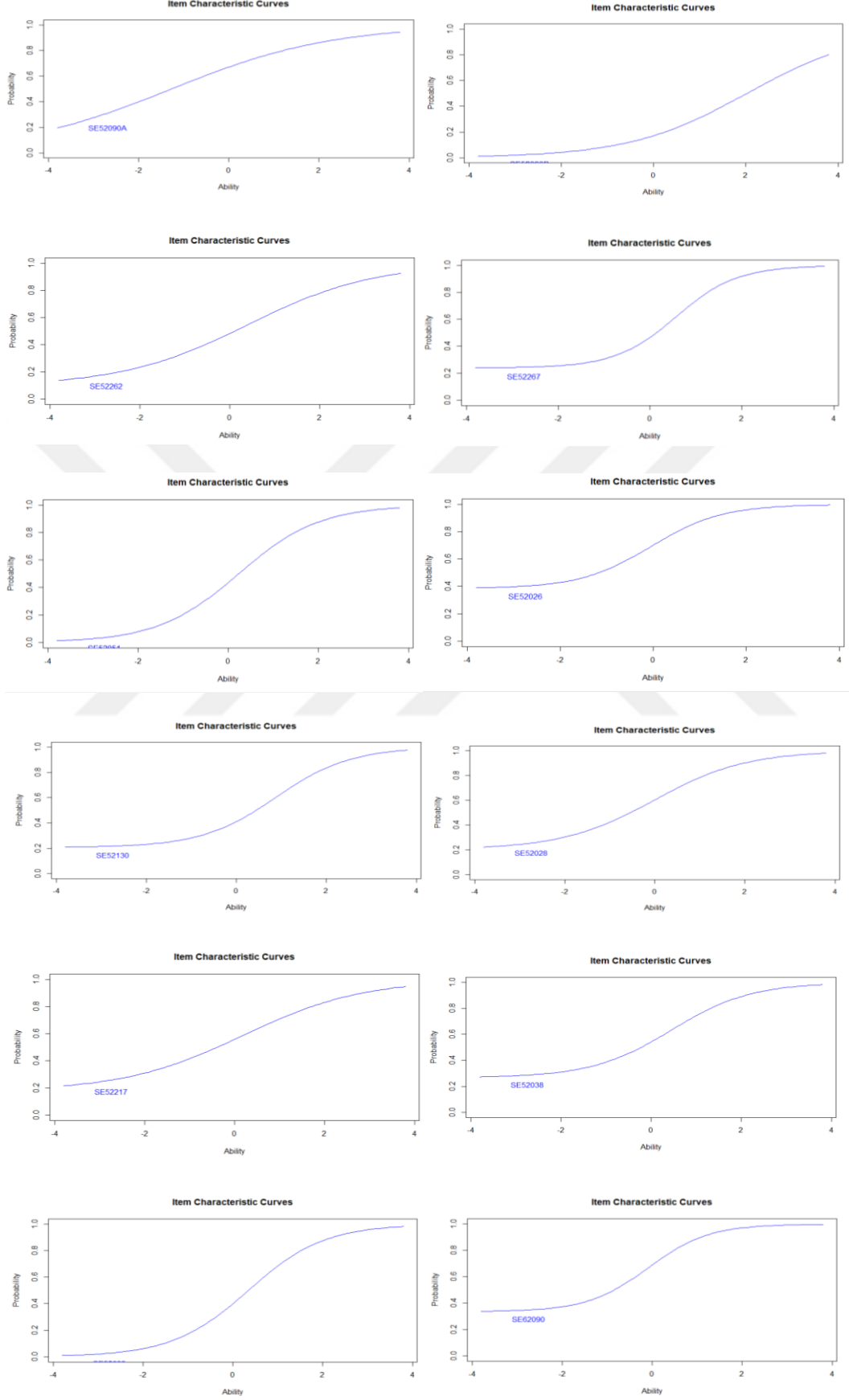
116	Madde 6	Madde 22	-0.056	-0.052
117	Madde 6	Madde 23	-0.037	-0.030
118	Madde 7	Madde 8	-0.038	-0.029
119	Madde 7	Madde 9	-0.020	-0.033
120	Madde 7	Madde 10	-0.038	-0.037
121	Madde 7	Madde 11	-0.040	-0.026
122	Madde 7	Madde 12	-0.038	-0.024
123	Madde 7	Madde 13	-0.018	-0.022
124	Madde 7	Madde 14	0.018	-0.024
125	Madde 7	Madde 15	-0.047	-0.022
126	Madde 7	Madde 16	-0.009	0.001
127	Madde 7	Madde 17	-0.032	-0.059
128	Madde 7	Madde 18	-0.067	-0.024
129	Madde 7	Madde 19	-0.036	-0.039
130	Madde 7	Madde 20	-0.052	-0.024
131	Madde 7	Madde 21	-0.051	-0.029
132	Madde 7	Madde 22	-0.047	-0.043
133	Madde 7	Madde 23	-0.024	-0.092
134	Madde 8	Madde 9	-0.013	-0.023
135	Madde 8	Madde 10	-0.047	-0.042
136	Madde 8	Madde 11	-0.040	-0.017
137	Madde 8	Madde 12	-0.035	0.029
138	Madde 8	Madde 13	-0.061	-0.009
139	Madde 8	Madde 14	-0.043	-0.042
140	Madde 8	Madde 15	0.147	-0.037
141	Madde 8	Madde 16	-0.013	-0.066
142	Madde 8	Madde 17	-0.057	-0.051
143	Madde 8	Madde 18	-0.056	-0.042
144	Madde 8	Madde 19	-0.049	-0.027
145	Madde 8	Madde 20	-0.071	-0.033
146	Madde 8	Madde 21	-0.040	-0.025
147	Madde 8	Madde 22	-0.048	-0.043
148	Madde 8	Madde 23	-0.021	-0.035
149	Madde 9	Madde 10	-0.007	-0.002
150	Madde 9	Madde 11	-0.026	-0.012
151	Madde 9	Madde 12	-0.018	-0.011
152	Madde 9	Madde 13	-0.002	-0.019
153	Madde 9	Madde 14	-0.007	-0.043
154	Madde 9	Madde 15	-0.060	-0.037
155	Madde 9	Madde 16	-0.019	-0.001

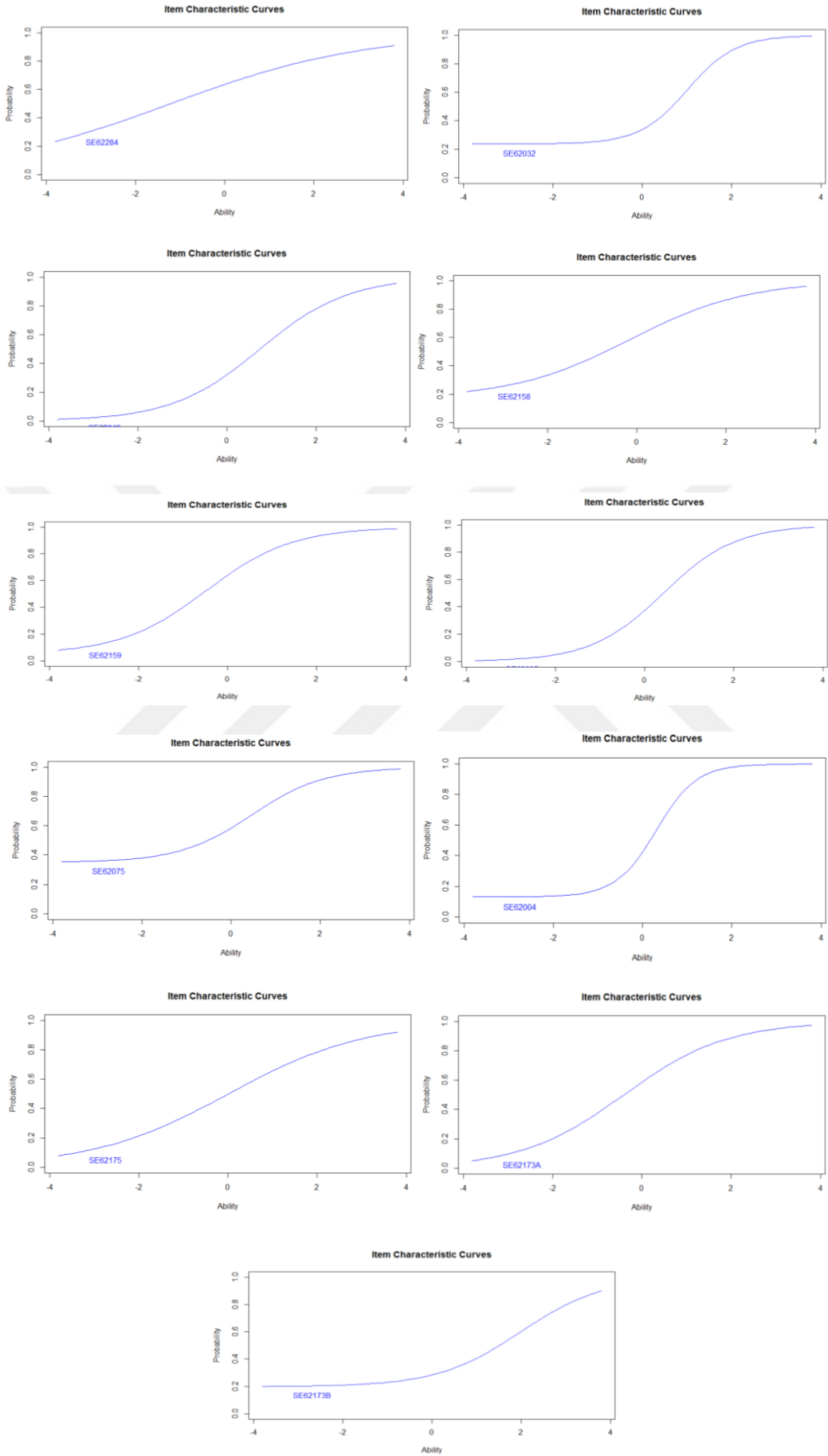
156	Madde 9	Madde 17	-0.027	-0.045
157	Madde 9	Madde 18	-0.076	-0.020
158	Madde 9	Madde 19	-0.018	-0.023
159	Madde 9	Madde 20	-0.053	-0.060
160	Madde 9	Madde 21	-0.032	-0.033
161	Madde 9	Madde 22	-0.005	-0.030
162	Madde 9	Madde 23	-0.037	-0.016
163	Madde 10	Madde 11	0.000	-0.006
164	Madde 10	Madde 12	-0.019	-0.029
165	Madde 10	Madde 13	0.002	0.004
166	Madde 10	Madde 14	-0.042	-0.018
167	Madde 10	Madde 15	-0.045	-0.048
168	Madde 10	Madde 16	-0.016	-0.017
169	Madde 10	Madde 17	-0.036	-0.072
170	Madde 10	Madde 18	-0.047	-0.038
171	Madde 10	Madde 19	-0.023	-0.059
172	Madde 10	Madde 20	-0.025	-0.047
173	Madde 10	Madde 21	-0.043	-0.047
174	Madde 10	Madde 22	-0.077	-0.053
175	Madde 10	Madde 23	-0.004	0.011
176	Madde 11	Madde 12	-0.049	0.029
177	Madde 11	Madde 13	-0.038	0.030
178	Madde 11	Madde 14	-0.073	-0.025
179	Madde 11	Madde 15	0.009	-0.070
180	Madde 11	Madde 16	-0.025	-0.075
181	Madde 11	Madde 17	-0.092	-0.113
182	Madde 11	Madde 18	-0.074	-0.081
183	Madde 11	Madde 19	-0.040	-0.084
184	Madde 11	Madde 20	-0.113	-0.080
185	Madde 11	Madde 21	-0.007	-0.102
186	Madde 11	Madde 22	-0.090	-0.034
187	Madde 11	Madde 23	-0.050	-0.050
188	Madde 12	Madde 13	-0.043	0.017
189	Madde 12	Madde 14	-0.036	-0.039
190	Madde 12	Madde 15	-0.029	-0.050
191	Madde 12	Madde 16	-0.068	-0.045
192	Madde 12	Madde 17	-0.052	-0.035
193	Madde 12	Madde 18	-0.052	-0.093
194	Madde 12	Madde 19	-0.041	-0.039
195	Madde 12	Madde 20	-0.067	-0.091

196	Madde 12	Madde 21	-0.034	-0.031
197	Madde 12	Madde 22	-0.032	0.017
198	Madde 12	Madde 23	-0.019	0.000
199	Madde 13	Madde 14	-0.022	-0.021
200	Madde 13	Madde 15	-0.076	-0.039
201	Madde 13	Madde 16	-0.031	-0.007
202	Madde 13	Madde 17	-0.036	-0.062
203	Madde 13	Madde 18	-0.049	-0.039
204	Madde 13	Madde 19	-0.008	-0.033
205	Madde 13	Madde 20	-0.036	-0.050
206	Madde 13	Madde 21	0.015	-0.031
207	Madde 13	Madde 22	-0.002	0.003
208	Madde 13	Madde 23	0.016	-0.045
209	Madde 14	Madde 15	-0.046	-0.053
210	Madde 14	Madde 16	-0.026	0.001
211	Madde 14	Madde 17	-0.021	-0.084
212	Madde 14	Madde 18	-0.092	-0.094
213	Madde 14	Madde 19	-0.047	-0.067
214	Madde 14	Madde 20	0.006	-0.051
215	Madde 14	Madde 21	-0.045	-0.070
216	Madde 14	Madde 22	-0.020	-0.045
217	Madde 14	Madde 23	0.003	-0.079
218	Madde 15	Madde 16	0.009	-0.022
219	Madde 15	Madde 17	-0.070	-0.041
220	Madde 15	Madde 18	0.022	-0.041
221	Madde 15	Madde 19	-0.072	-0.023
222	Madde 15	Madde 20	-0.070	-0.067
223	Madde 15	Madde 21	-0.021	-0.077
224	Madde 15	Madde 22	-0.087	-0.057
225	Madde 15	Madde 23	-0.126	-0.069
226	Madde 16	Madde 17	-0.038	-0.034
227	Madde 16	Madde 18	-0.003	-0.032
228	Madde 16	Madde 19	-0.021	0.010
229	Madde 16	Madde 20	-0.056	-0.020
230	Madde 16	Madde 21	-0.066	-0.030
231	Madde 16	Madde 22	-0.042	-0.028
232	Madde 16	Madde 23	-0.028	-0.046
233	Madde 17	Madde 18	-0.065	-0.095
234	Madde 17	Madde 19	-0.007	-0.054
235	Madde 17	Madde 20	-0.069	-0.022

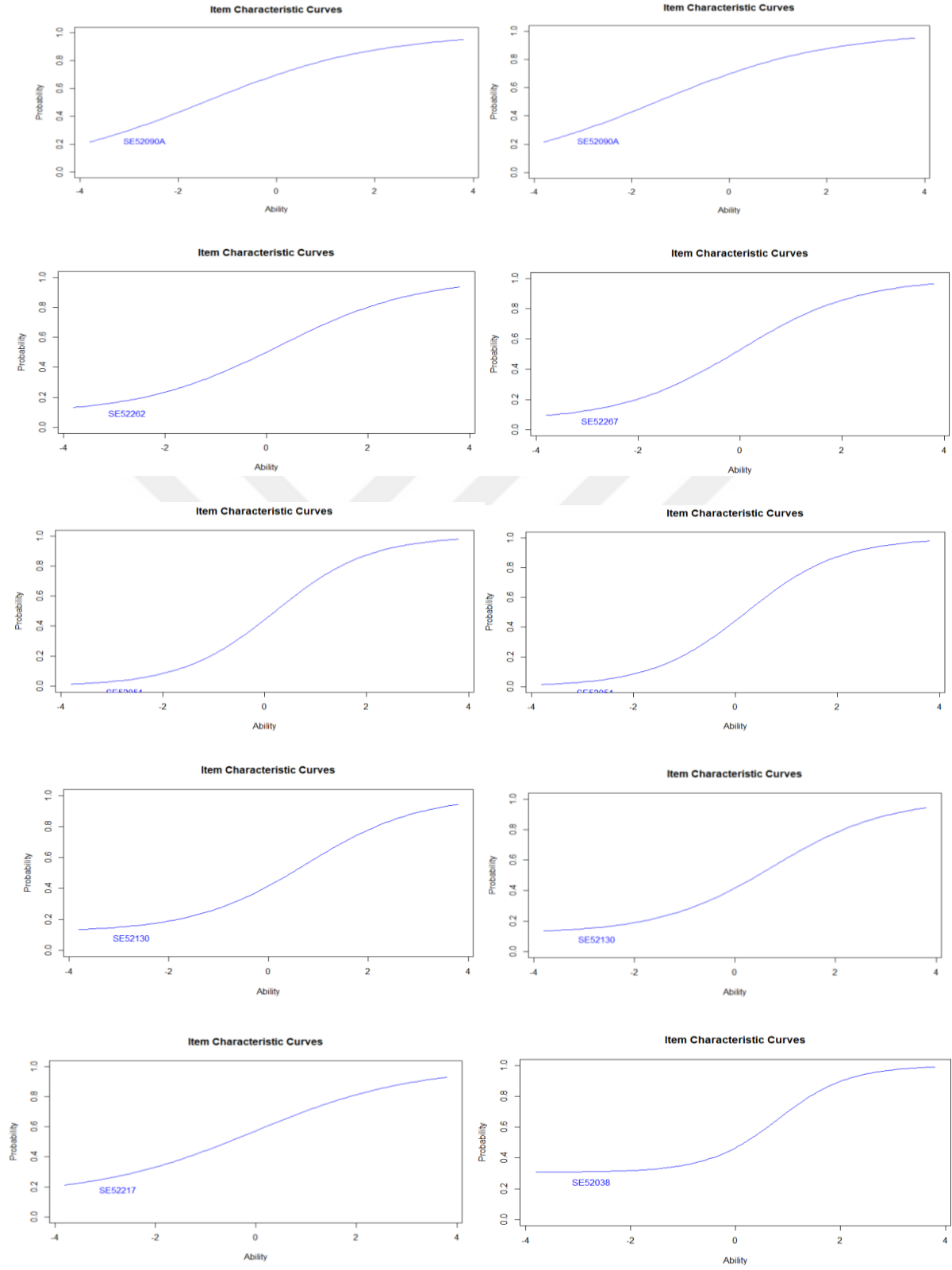
236	Madde 17	Madde 21	-0.030	-0.007
237	Madde 17	Madde 22	-0.015	-0.031
238	Madde 17	Madde 23	0.002	-0.032
239	Madde 18	Madde 19	-0.031	-0.018
240	Madde 18	Madde 20	-0.103	-0.045
241	Madde 18	Madde 21	-0.033	-0.039
242	Madde 18	Madde 22	-0.062	-0.052
243	Madde 18	Madde 23	-0.060	-0.088
244	Madde 19	Madde 20	-0.073	-0.036
245	Madde 19	Madde 21	-0.010	-0.030
246	Madde 19	Madde 22	-0.022	-0.048
247	Madde 19	Madde 23	-0.011	-0.094
248	Madde 20	Madde 21	-0.075	-0.036
249	Madde 20	Madde 22	-0.034	-0.074
250	Madde 20	Madde 23	-0.017	-0.081
251	Madde 21	Madde 22	-0.016	0.001
252	Madde 21	Madde 23	-0.012	-0.008
253	Madde 22	Madde 23	0.026	-0.001

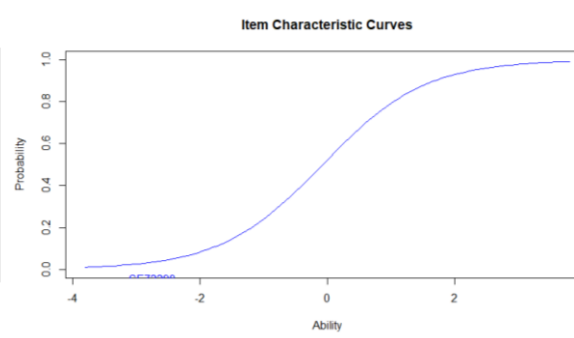
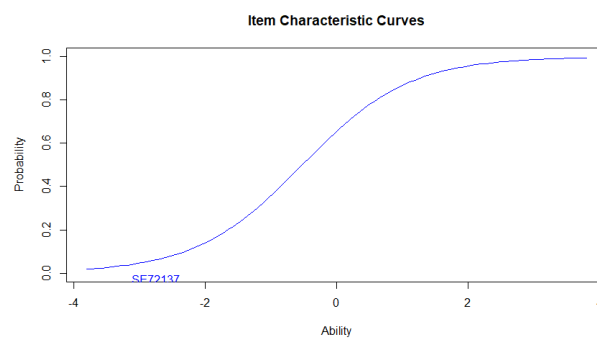
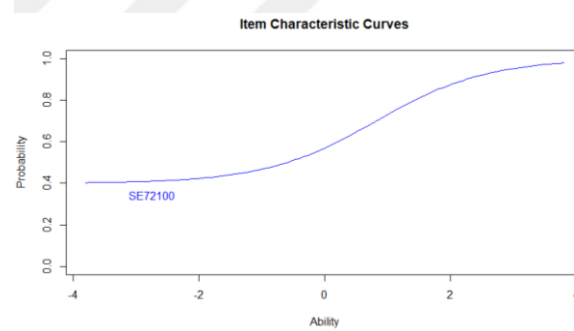
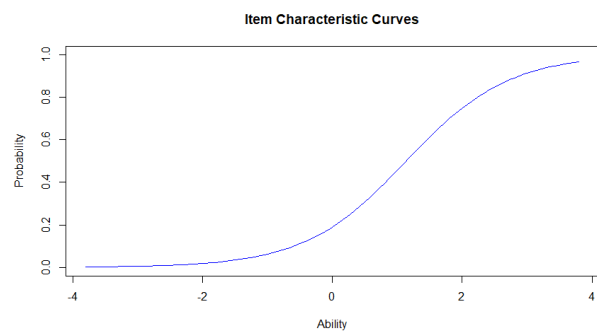
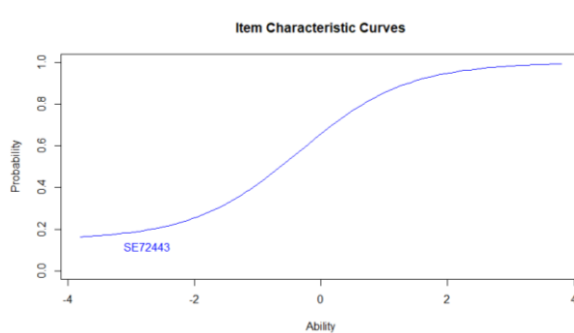
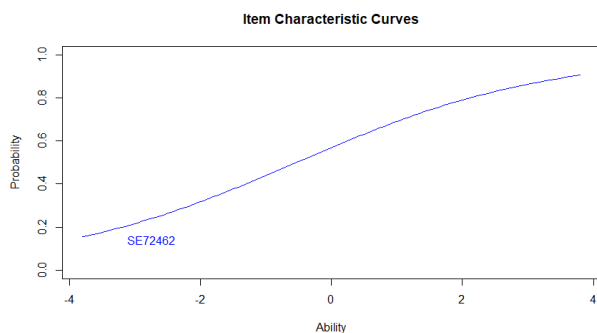
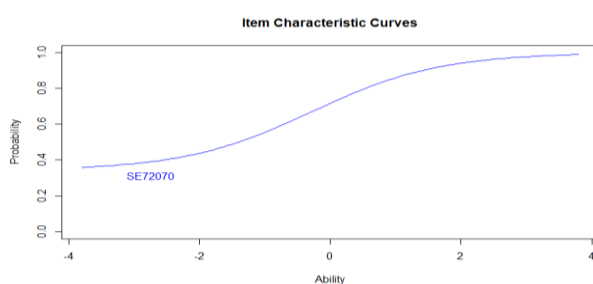
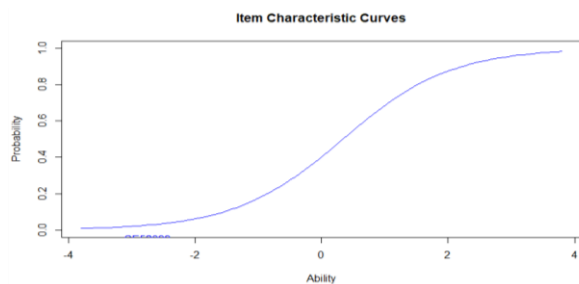
EK-B: Kitapçık 6'da Yer Alan Maddelere Ait Madde Karakteristik Eğrileri

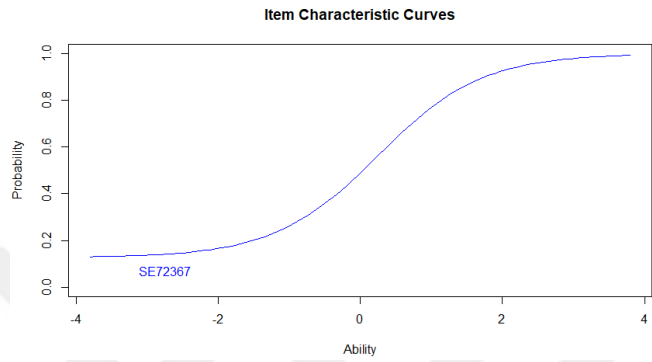
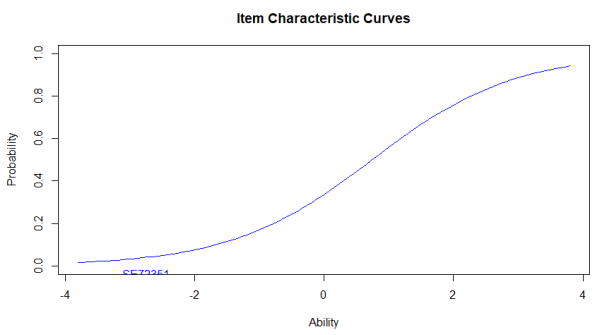
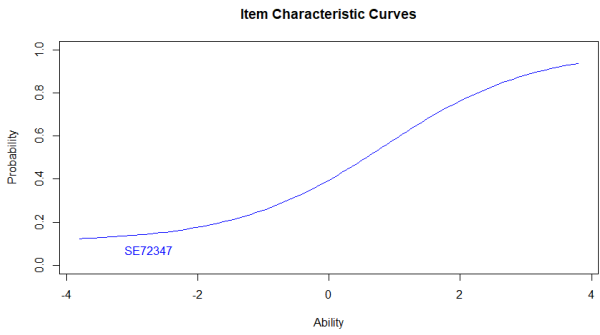




EK-C: Kitapçık 7’de Yer Alan Maddelere Ait Madde Karakteristik Eğrileri







EK-Ç: Analiz Sonucu Elde Edilen RMSE ve BIAS Değerleri

Örneklem Büyükliği	Kayıp Veri Oranı	Kayıp Verilerin Test Formları İçindeki Yeri	Kayıp Veriyle Başa Çıkma Yöntemi	Test Eşitleme Yöntemi	Gerçek Puan RMSE	Gerçek Puan BIAS
Tam veri	-	-	-	SL	0.2638	0.0270
seti*	-	-	-	H	0.2653	0.0272
750	%10	ET	SA	SL	0.3109	4.1274
				H	0.3729	4.3847
				SL	0.2924	3.4047
				H	0.2923	2.4121
			PMM	SL	0.2888	3.1103
				H	0.2842	2.0308
				SL	0.2913	3.1114
				H	0.2868	2.0124
			LOGREG	SL	0.3113	5.5895
				H	0.3058	4.7964
750	%10	HİT	SA	SL	0.2883	3.3534
				H	0.2833	2.4108
				SL	0.2947	3.1453
				H	0.2899	2.0686
			PMM	SL	0.2964	3.1439
				H	0.2916	2.0359
				SL	0.3407	6.8107
				H	0.4663	7.9556
			HDA	SL	0.3083	3.1505
				H	0.3171	2.2233
750	%20	ET	SA	SL	0.2960	2.9812
				H	0.2930	1.8836
				SL	0.2986	2.9219
				H	0.2962	1.8298
			PMM	SL	0.3516	7.9904
				H	0.3476	7.4127
				SL	0.2980	2.8694
				H	0.3008	2.0954
			HİT	SL	0.3072	2.8294
				H	0.3041	1.7507
750	%20	HİT	LOGREG	SL	0.3072	2.8012
				SL	0.3072	2.8012

				H	0.3039	1.7364
			SA	SL	0.3645	9.2492
				H	0.5534	12.7856
			HDA	SL	0.3022	3.4448
				H	0.3178	2.6323
750	%30	ET	PMM	SL	0.2939	3.1771
				H	0.2934	2.1693
			LOGREG	SL	0.2966	3.1148
				H	0.2973	2.1112
			SA	SL	0.3667	10.1800
				H	0.3659	9.7562
			HDA	SL	0.2853	2.8993
				H	0.2915	2.1764
750	%30	HIT	PMM	SL	0.3019	3.2180
				H	0.2994	2.1314
			LOGREG	SL	0.3025	3.0429
				H	0.3020	2.0004
			SA	SL	0.2779	0.0930
				H	0.3324	0.0677
			HDA	SL	0.2606	0.0497
				H	0.2607	0.0345
1500	%10	ET	PMM	SL	0.2599	0.0485
				H	0.2551	0.0319
			LOGREG	SL	0.2615	0.0489
				H	0.2568	0.0321
			SA	SL	0.2767	0.0860
				H	0.2712	0.0733
			HDA	SL	0.2583	0.0451
				H	0.2558	0.0303
1500	%10	HIT	PMM	SL	0.2661	0.0463
				H	0.2625	0.0300
			LOGREG	SL	0.2658	0.0483
				H	0.2620	0.0317
			SA	SL	0.3052	0.0298
				H	0.4171	0.1246
			HDA	SL	0.2755	0.0496
				H	0.2819	0.0354
1500	%20	ET	PMM	SL	0.2659	0.0478
				H	0.2619	0.0299
			LOGREG	SL	0.2693	0.0473

				H	0.2660	0.0300
			SA	SL	0.3173	0,1298
				H	0.3124	0.1173
			HDA	SL	0.2648	0.0464
				H	0.2667	0.0336
1500	%20	HİT	PMM	SL	0.2745	0.0451
				H	0.2711	0.0275
			LOGREG	SL	0.2757	0.0462
				H	0.2723	0.0291
			SA	SL	0.3226	0.1758
				H	0.4921	0.2007
			HDA	SL	0.2738	0.0559
				H	0.2884	0.0424
1500	%30	ET	PMM	SL	0.2600	0.0508
				H	0.2598	0.0353
			LOGREG	SL	0.2640	0.0520
				H	0.2641	0.0362
			SA	SL	0.3229	0.1589
				H	0.3245	0.1535
			HDA	SL	0.2538	0.0497
				H	0.2588	0.0381
1500	%30	HİT	PMM	SL	0.2688	0.0520
				H	0.2675	0.0361
			LOGREG	SL	0.2673	0.0523
				H	0.2670	0.0368

* Referans değeri olarak alınmıştır.

EK-D: Araştırma Etik Komisyon İzin Muafiyeti Formu



Hacettepe Üniversitesi
Eğitim Bilimleri Enstitüsü
Tez Çalışması/Araştırma Etik Komisyon İzin Muafiyeti Formu

F46

05 / 05 / 2022

Hacettepe Üniversitesi
Eğitim Bilimleri Enstitüsü
Eğitimde Ölçme ve Değerlendirme Ana Bilim Dalı Başkanlığına

Tez/Araştırma Başlığı	Kayıp Veriyle Başa Çıkma Yöntemlerinin Test Etkilemeye Etkisinin İncelenmesi
-----------------------	--

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:

- ☐ Kamusal erişime açık (buraya yazınız):
☐ Özel izin ve onaya tabi (buraya yazınız):
☐ Üretilmiş veri (buraya yazınız):

(X) Diğer (buraya yazınız): Bireysel erişime açık olan ve Uluslararası Eğitim Başarılarını Değerlendirme Kuruluşu (IEA-International Association for the Evaluation of Educational Achievement) tarafından resmi internet sayfasında yer alan (<https://timss2019.org/international-database/>) TIMSS 2019 uygulamasına ait veriler kullanılacaktır.

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.

Gülden ÖZDEMİR
(Araştırmacı Adı Soyadı, İmza)

Araştırmacı Bilgileri

Adı Soyadı	Gülden ÖZDEMİR
Öğrenci İse No	N18144972
Ana Bilim Dalı	Eğitim Bilimleri
Programı	Eğitimde Ölçme ve Değerlendirme
Statüsü	☐ Yüksek Lisans ☒ Doktora ☐ Bütünleşik Dr. ☐ Diğer

Danışman Görüşü ve Onayı*

Çalışmada bireysel erişime açık TIMSS 2019 verileri kullanılacağı için etik kurul izni gerekmemektedir.

Prof. Dr. Burcu ATAR
(İmza)
(Danışman Adı Soyadı)

*Tez ve tezden üretilen yayınlarda gerekli

EK-E: Etik Beyanı

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

- * tez içindeki bütün bilgi ve belgeleri akademik kurallar çerçevesinde elde ettiğimi,
- * görsel, işitsel ve yazılı 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.

...../...../2023

Gülden ÖZDEMİR

EK-F: Doktora Tez Çalışması Orijinallik Raporu

...../...../2023

HACETTEPE ÜNİVERSİTESİ
Eğitim Bilimleri Enstitüsü
Eğitim Bilimleri Ana Bilim Dalı Başkanlığına,

Tez Başlığı : Kayıp Veriyle Başa Çıkma Yöntemlerinin Test Eşitlemeye Etkisinin İncelenmesi

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
17/05/2023	97	21690	17/05/2023	%8	2095335443

Uygulanan filtreler:

- Kaynaklar hariç
- Alıntılar dâhil
- 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: Gülden ÖZDEMİR

Öğrenci No.: N18144972

Ana Bilim Dalı: Eğitim Bilimleri

İmza

Programı: Eğitimde Ölçme ve Değerlendirme

Statüsü: ☐ Y.Lisans ☒ Doktora ☐ Bütünleşik Dr.

DANIŞMAN ONAYI

UYGUNDUR.

(Prof. Dr. Burcu ATAR)

EK-G: Dissertation Originality Report

...../...../2023

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

Thesis Title: Investigation of The Impact of Techniques of Handling Missing Data on the Test Equating

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
17/05/2023	97	21690	17/05/2023	8%	2095335443

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: Gülden ÖZDEMİR

Student No.: N18144972

Department: Educational Sciences

Program: Educational Measurement and Evaluation

Status: ☐ Masters ☒ Ph.D. ☐ Integrated Ph.D.

Signature

ADVISOR APPROVAL

APPROVED
(Prof. Dr. Burcu ATAR)

EK-H: Yayımlama 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 fikrî 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 ... ay ertelenmiştir. ⁽²⁾
- ☐ Tezimle ilgili gizlilik kararı verilmiştir. ⁽³⁾

...../...../2023

Gülden ÖZDEMİR

"*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 tez erişime açılmasının ertelenmesine karar verebilir.
- (2) Madde 6.2. Yeni teknik, materyal ve metotların kullanıldığı, henüz makaleye dönüşmemiş veya patent gibi yöntemlerle korunmamış ve internetten 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 tez 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.

