



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

DEĞİŞEN MADDE FONKSİYONU GÖSTEREN ORTAK MADDELERİN
TEST EŞİTLEMeye ETKİSİNİN İNCELENMESİ

Feyzi GÜNEŞ

Yüksek Lisans Tezi

Ankara, 2021

Liderlik, arařtırma, inovasyon, kaliteli eđitim ve deđiřim ile

Daha ileriye ... En İyiyeye ...



HACETTEPE ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı
Eğitimde Ölçme ve Değerlendirme Programı

DEĞİŞEN MADDE FONKSİYONU GÖSTEREN ORTAK MADDELERİN
TEST EŞİTLEMeye ETKİSİNİN İNCELENMESİ

THE STUDY OF EFFECT OF ANCHOR ITEMS SHOWING DIFFERENTIAL
ITEM FUNCTIONING ON TEST EQUATING

Feyzi GÜNEŞ

Yüksek Lisans Tezi

Ankara, 2021

Bu arařtırmada deęiřen madde fonksiyonu (DMF) gsteren ortak maddelerin test eřitlemeye olan etkisinin incelenmesi amalanmıřtır. Testlerin eřdeęer olmayan gruplar ortak test deseni yntemiyle eřitlenmesinde, DMF gsteren ortak test maddelerinin testten ıkarılmasının eřitlenmiř puanlara ve eřitleme hatasına olan etkisi doęrusal ve doęrusal olmayan eřitleme yntemleriyle incelenmiřtir. DMF deęiřkenleri İngilizce, İspanyolca dilleri ve bu dillerde testi alan bireylerin cinsiyetleri olarak belirlenmiřtir. DMF gsteren maddelerin belirlenmesinde Mantel-Haenszel ve SIBTEST teknikleri kullanılmıřtır. Eřitleme alıřması doęrusal eřitleme yntemlerinden Tucker, Levine, zincir eřitleme yntemi, doęrusal olmayan yntemlerde ise frekans kestirim, zincir eřit yzdelikli ve Daire-Yay Levine yntemleri ile yrtlmřtr. alıřma verilerini, PISA 2018 uygulaması 18 ve 24. kitapıkları İngilizce, İspanyolca dillerinde bilgisayar ortamında almıř fen okuryazarlıęı birey yanıtları oluřturmaktadır. Tekniklere gre B ve C dzeyinde DMF gsteren maddelerin ortak testten ıkarılmasının eřitleme hatasına etkisi aęırlıklandırılmıř hata kareleri ortalaması ile deęerlendirilmiřtir. Arařtırma sonucunda DMF’li maddelerin ortak testten ıkarılması ile eřitleme hatalarında tutarlı deęiřimler gzlenmemiřtir. Eřitleme hatalarındaki tutarsızlıęa; kitapıkların ortalamaları arasındaki istatistiksel olarak anlamlı fark, DMF deęiřkenlerine ait alt gruplar arasındaki yetenek farklılařması ve eřitleme yntemlerine ait varsayımların karřılanma derecesinin kaynak olabileceęi dřnlmřtr.

Anahtar szckler: Madde yanlılıęı, deęiřen madde fonksiyonu, test eřitleme, ortak test, eřitleme hatası

Abstract

In this study, it is aimed to examine the effect of anchor items with differential item functioning (DIF) on test equating. The effect of excluding anchor test items which showing DIF on equated scores and equating error was investigated by linear and nonlinear equating methods in non-equivalent groups with anchor test design. DIF variables were determined as English, Spanish languages and the gender of the individuals who took the test in these languages. Mantel-Haenszel and SIBTEST techniques were used to identify items with DIF. Test equating was conducted using the Tucker, Levine, Chained equating method, which are linear equalization methods, Frequency estimation, Chained equipercentile and Circle-Arc Levine methods in nonlinear methods. The study data consists of the science literacy individual responses that have taken the 18th and 24th booklets of the PISA 2018 application in English and Spanish languages in computer based modules. According to the techniques, the effect of removing the items showing DIF at the B and C levels from the anchor test on the equating error was evaluated with the weighted mean square error (WMSE). As a result of the research, the exclusion of DIF items from the anchor test did not consistent changes in the equating errors. Inconsistency in equating errors; It was thought that the statistically significant difference between the averages of the booklets, the skill differentiation between subgroups of DIF variables and the degree of support the assumptions of the equating methods could be the source.

Keywords: Item bias, differential item functioning, test equating, anchor test, equating error

Teşekkür

Yüksek lisans eğitimim ve tez çalışmam süresince bilgisinden, deneyiminden faydalandığım, çalışma disiplinini kendime örnek aldığım değerli danışmanım Prof. Dr. Hülya KELECIÖĞLU'na;

Kıymetli dönütleriyle tezime katkı sunan jüri üyelerim Doç. Dr. Kübra ATALAY KABASAKAL ve Doç. Dr. Kaan Zülfikar DENİZ'e;

Bilgisini, tecrübesini paylaşmaktan hiç çekinmeyen, öğretileri sayesinde tez sürecimi kolaylaştıran hocam Doç. Dr. Burcu ATAR'a;

Eğitimim boyunca gördüğüm samimiyeti ve örnek mütavizi kişiliği için Prof. Dr. Selahattin GELBAL'a ve mezunu olmaktan gurur duyduğum bu bölüme katkı sunan tüm hocalarıma;

Tezimin hazırlanma süresince sıkça fikirlerine başvurduğum çalışma arkadaşlarım Serkan KIRKESER, M. Alper Koçak ve bilge insan Dr. Bahaddin ŞAHİN'e

Yoğun hayat temposuna rağmen desteğini esirgemeyen değerli dostum Arzu IŞIK ve Ferhat SARI'ya

Bu günlere gelmemi sağlayan her şeyimi borçlu olduğum annem Keziban GÜNEŞ, babam Alaattin GÜNEŞ, kardeşim Adem GÜNEŞ'e ve her daim desteklerini esirgemeyen ikinci ailem Sultan TÜTÜNCÜ ve Durdu TÜTÜNCÜ'ye;

Hayat mücadelesi örnek alınacak insan rahmetli babaannem Meryem GÜNEŞ'e

Beraber geçireceğimiz vakitlerinden aldığım oğullarım Tuna ve Ali Ediz'e;

Ezcümle sabır ve anlayışı ile bu tezin gerçek sahibi olduğuna inandığım sevgili eşim Selda GÜNEŞ'e

teşekkür ederim.

İçindekiler

Öz.....	Hata! Yer işareti tanımlanmamış.
Abstract.....	ii
Teşekkür.....	iii
Tablolar Dizini.....	vi
Şekiller Dizini.....	vii
Simgeler ve Kısaltmalar Dizini.....	viii
Bölüm 1 Giriş.....	1
Problem Durumu.....	1
Araştırmanın Amacı ve Önemi.....	5
Araştırma Problemi.....	6
Sayıltılar.....	7
Sınırlılıklar.....	7
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	8
Yanlılık.....	8
Test Bağlama.....	16
İlgili Araştırmalar.....	38
Bölüm 3 Yöntem.....	43
Araştırmanın Türü.....	43
Araştırmanın Evreni ve Örneklemi.....	43
Veri Toplama Süreci.....	44
Veri Toplama Aracı.....	44
Verilerin Analizi.....	45
Bölüm 4 Bulgular ve Yorumlar.....	53
Alt Problem 1'e İlişkin Bulgular.....	53
Alt Problem 2'ye İlişkin Bulgular.....	55
Alt Problem 3'e İlişkin Bulgular.....	61

Bölüm 5 Sonuç, Tartışma ve Öneriler	68
Sonuç ve Tartışma	68
Öneriler	70
Kaynaklar	73
EK-A: PISA 2018 18. Kitapçık Fen Okuryazarlığı Ham Puanlarının Doğrusal ve Doğrusal Olmayan Eşitleme Yöntemleriyle Elde Edilmiş 24. Kitapçığındaki Eşitlenmiş Puanları	86
EK-B: MH Analizine Göre Dil Kaynaklı DMF Gösteren Ortak Maddelerden Arındırılarak Elde Edilen Eşitlenmiş Puanlar	87
EK-C: SIBTEST Analizine Göre Dil Kaynaklı DMF Gösteren Ortak Maddelerden Arındırılarak Elde Edilen Eşitlenmiş Puanlar	88
EK-Ç: MH Analizine Göre Cinsiyet Kaynaklı DMF Gösteren Ortak Maddelerden Arındırılarak Elde Edilen Eşitlenmiş Puanlar	89
EK-D: SIBTEST Analizine Göre Cinsiyet Kaynaklı DMF Gösteren Ortak Maddelerden Arındırılarak Elde Edilen Eşitlenmiş Puanlar	90
EK-E: Etik Komisyonu Onay Bildirimi	91
EK-F: Etik Beyanı	92
EK-G: Yüksek Lisans/Doktora Tez Çalışması Orijinallik Raporu	93
EK-H: Thesis/Dissertation Originality Report.....	94
EK-I: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı	95

Tablolar Dizini

Tablo 1	<i>DMF Belirlemede Teknik Sınıflandırması</i>	11
Tablo 2	<i>Toplam Puanı m Olan Bireylere Ait Olasılık Tablosu</i>	12
Tablo 3	<i>MH Değeri Etki Büyüklüğü Sınıflaması</i>	14
Tablo 4	<i>SIBTEST Değeri Etki Büyüklüğü Sınıflaması</i>	15
Tablo 5	<i>Ortalama Eşitleme</i>	28
Tablo 6	<i>NEAT Desende Gözlenen Puana Dayalı Eşitleme Yöntemleri</i>	33
Tablo 7	<i>Bireylerin Değişkenlere Göre Dağılımı</i>	43
Tablo 8	<i>Maddelerin Kitapçıklara Yerleşimi</i>	44
Tablo 9	<i>PISA 2018 18 ve 24. Kitapçıkların KMO ve Bartlett Değerleri</i>	46
Tablo 11	<i>Kitapçıklara Ait Betimsel İstatistikler</i>	48
Tablo 12	<i>Kitapçıkların Güvenirliklerinin Karşılaştırılmasına İlişkin Bulgular</i>	49
Tablo 13	<i>Kitapçıkların Ortalama Güçlüklerinin Karşılaştırılması İlişkin Bulgular</i> ..	49
Tablo 14	<i>Kitapçıkların Varyans ve Ortalamalarının Karşılaştırılmasına İlişkin Bulgular</i>	50
Tablo 15	<i>Ortak Test Maddelerinin DMF Değişkenlerine Göre Analizi</i>	51
Tablo 16	<i>Doğrusal ve Doğrusal Olmayan Eşitleme Yöntemlerine Ait AHKO Değerleri</i>	55
Tablo 17	<i>Ortak Test Maddelerinin Dil Değişkenine Göre MH Analiz Bulguları</i>	56
Tablo 18	<i>Ortak Test Maddelerinin Dil Değişkenine Göre SIBTEST Analiz Bulguları</i>	57
Tablo 19	<i>Eşitleme Yöntemlerinin Alt Problem 1 ve 2'ye Göre AHKO Değerleri</i> ...	59
Tablo 20	<i>Dil Değişkenine Ait Alt Grupların Varyans ve Ortalamalarının Alt Problemlere Göre Karşılaştırılmasına İlişkin Bulgular</i>	61
Tablo 21	<i>Ortak Test Maddelerinin Cinsiyet Değişkenine Göre MH Analiz Bulguları</i>	62
Tablo 22	<i>Ortak Test Maddelerinin Cinsiyet Değişkenine Göre SIBTEST Analiz Bulguları</i>	63
Tablo 23	<i>Eşitleme Yöntemlerinin Alt Problem 1 ve 3'e Göre AHKO Değerleri</i>	65
Tablo 24	<i>Cinsiyet Değişkenine Ait Alt Grupların Varyans ve Ortalamalarının Alt Problemlere Göre Karşılaştırılmasına İlişkin Bulgular</i>	66

Şekiller Dizini

Şekil 1. Test bağlama yöntemleri	16
Şekil 2. Tek grup deseni	21
Şekil 3. Rastgele gruplar deseni	22
Şekil 4. Dengelenmiş gruplar deseni	23
Şekil 5. Eşdeğer olmayan gruplar ortak test deseni	24
Şekil 6. Eşit yüzdelikli eşitlemenin grafiksel gösterimi	30
Şekil 7. Doğrusal zincir eşitleme yöntemi	36
Şekil 8. Ham puanlar ile eşitlenmiş puanlara ait dağılım grafikleri.....	54
Şekil 9. Dil değişkenine göre belirlenen maddelerin testten çıkarılması sonucunda oluşan ham puanlar ile eşitlenmiş puanlara ait dağılım grafikleri.....	59
Şekil 10. Cinsiyet değişkenine göre belirlenen maddelerin testten çıkarılması sonucunda oluşan ham puanlar ile eşitlenmiş puanlara ait dağılım grafikleri	64

Simgeler ve Kısaltmalar Dizini

AHKO: Ağırlıklandırılmış Hata Kareleri Ortalaması

ALES: Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı

AÖF: Açık Öğretim Fakültesi

AÖL: Açık Öğretim Lisesi

DMF: Değişen Madde Fonksiyonu

KTK: Klasik Test Kuramı

LGS: Liselere Geçiş Sistemi

MEB: Milli Eğitim Bakanlığı

MH: Mantel-Haenszel

MTK: Madde Tepki Kuramı

NEAT: Non Equivalent Groups with Anchor Test (Eşdeğer Olmayan Gruplar Ortak Test)

OECD: Organisation for Economic Co-operation and Development (Ekonomik İşbirliği ve Kalkınma Örgütü)

ÖSYM: Ölçme Seçme ve Yerleştirme Merkezi

PIRLS: Progress in International Reading Literacy Study (Uluslararası Okuma Becerilerinde Gelişim Araştırması)

PISA: Programme for International Student Assessment (Uluslararası Öğrenci Değerlendirme Programı)

TIMSS: Trends in International Mathematics and Science Studies (Uluslararası Matematik ve Fen Eğilimleri Araştırması)

Bölüm 1

Giriş

Araştırmanın bu bölümünde problem durumu, araştırmanın amacı ve önemi, araştırma problemi, sayıtlar ve sınırlılıklara yer verilmiştir.

Problem Durumu

Doğumdan itibaren her dönemde karşılaşılan testlerin insan hayatını şekillendirmede etkili olduğu bir gerçektir. Günümüz insanı üzerindeki etkileri yanı sıra tarih boyunca belirli amaçlar için kullanılmış bireyler hakkında kararlar alma konusunda belirleyici roller üstlenmiştir. Testlerin serüveni M.Ö. 2200'lü yıllarda Çin'de devlet memurlarının seçiminde kullanılmasıyla başlarken günümüzdeki adıyla *psikolojik test* kavramı ise Christian von Wolff'un psikolojiyi bir bilim ve psikolojik ölçmeyi de bu bilim dalında özel bir uygulama alanı olarak öngörmesiyle başlar (Cohen ve Swerdlik, 2010; Turgut ve Baykul, 2015). Psikolojik test kavramı bireylerin önemli olduğu düşünülen kişisel özelliklerini ölçmeyi amaçlayan sistematik uygulamalar bütünü olarak tanımlanır. Ölçülmesi hedeflenen özellikler; başarı, yetenek, zeka, motivasyon, kaygı gibi unsurlar olabilir. Kişilerde bulunan bu özelliklerin varlığı veya yokluğu meslek tercihinden, sürücü belgesi almaya kadar birçok alanda belirleyici rol üstlenir. Psikolojik testler, bu işlevleri sayesinde bireyler hakkında önemli kararların alınmasında birer araç haline dönüşür (Kaplan ve Saccuzzo, 2018; Murphy ve Davidshofer, 2005).

Psikolojik testler ölçülmesi hedeflenen davranış/psikolojik özellik türü bakımından maksimum performans ve tipik tepki testleri olmak üzere iki genel kategoriye ayrılır. Maksimum performans testlerine zeka, yetenek, iş performansı ve başarı testleri örnek olarak gösterilebilir. Bu testlerde bireyin maksimum performansını yansıtması beklenir. Tipik tepki testlerinde tutum, ilgi ve kişilik gibi psikolojik özellikler ölçülür. Tipik tepki testlerinde temel amaç bireylerin belli uyaranlara verdiği tepkiler aracılığıyla sahip oldukları kendine özgü durumların olduğu şekliyle ortaya konmasıdır (Kilmen, 2019).

Psikolojik testin ölçmeyi hedeflediği yapı bazen doğrudan gözlenebilir bir davranışı içeriyor iken bazen de gizil bir özellikten meydana geliyor olabilir. Murphy ve Davidshofer (2005), gizil özelliklerin dolaylı ölçülmesi nedeniyle testleri

mükemmel olarak görmeseler de bireyler hakkında önemli kararların alınmasındaki en iyi, en adil ve en doğru teknoloji ile donatılmış araçlar olarak nitelemişlerdir. Bu yorum aslında testin geçerliğine yönelik bir betimlemedir. Geçerliğe yönelik yapılabilecek en temel tanım, testin ölçmeyi hedeflediği özelliği başka değişkenlerin etkisinde kalmadan ölçebilme derecesidir (Turgut ve Baykul, 2015). Tanımdan anlaşılacağı üzere ölçülmesi amaçlanan değişkenler dışında ölçmeye karışan diğer değişkenler testin geçerliğine gölge düşürmektedir. Klasik Test Kuramına (KTK) göre test puanlarının bir miktar hata içerdiği kabul edilmiş bir olgudur (Hambleton ve Jones, 1993). Bu yüzden test puanlarıyla yapılacak çıkarımların bir dizi istenmeyen değişken etkisi içerdiği varsayılır. Bu etki belli gruptaki bireyler üzerinde sistematik olarak yoğunlaşıyorsa yanlılık kavramı ile açıklanır (Kim, 2003). Yanlılık aynı popülasyondaki belirli alt gruplardan birinin diğerlerine nazaran bazı maddelerin doğru cevabına daha kolay ulaşmasıdır. Bu alt gruplar cinsiyet, kültür, etnik köken olabildiği gibi uluslararası testlerde çeviri kaynaklı problemler de olabilmektedir (Osterlind, 1983). Yanlılığın tespiti, maddenin aynı yetenek seviyesindeki alt gruplardan biri için farklı fonksiyonlaştığının istatistiksel ispatı ile başlayan (Gök, Kelecioğlu ve Doğan, 2010) ve alan uzmanlarının değerlendirmesi ile sonuçlanan bir süreçtir. İstatistiksel ispat süreci alanyazında Değişen Madde Fonksiyonu (DMF) olarak tanımlanır. Birçok test uygulayıcısı test maddelerini kamuoyu ile paylaşmadığı için araştırmalar yanlılığın istatistiksel aşaması olan DMF analizine yönelik çalışmalarla sürdürülmektedir. DMF bulguları yanlılığın varlığını garanti etmese de alt gruplara göre farklı işlevselleşen maddeler hakkında detaylı bilgi sunmaktadır.

Test programcıları ve uygulayıcılar aynı testin birden fazla ve birbiri yerine kullanılabilir formlardan oluşmasını tavsiye eder (Angoff, 1971). Bunun nedenlerinden ilki güvenlik endişesidir. Çünkü yüksek risk taşıyan bu testlerin sonuçları sertifika, lisans almak veya bir eğitim kurumuna kabul edilmek gibi bireyler üzerinde önemli etkilere sahiptir. İkincisi, test maddelerinin kamuoyuyla paylaşılmasının gerekli olduğu düşüncesidir. Paylaşılan maddeler gelecek testlerde kullanılmayacak böylelikle maddelerin sınavı birden fazla alan adaylar için avantaj sağlama olasılığı ortadan kalkacaktır. Üçüncüsü ise değişen sınav içeriği ve koşullardan ötürü maddelerin güncellenmesi ihtiyacıdır (Sansivieri, Wiberg ve Matteucci, 2017). Örneğin, ülkemizde Ölçme Seçme ve Yerleştirme

Merkezi (ÖSYM) tarafından yılda üç kez uygulanan Akademik Personel ve Lisansüstü Eğitimi Giriş Sınavı (ALES) soruları ve doğru cevapları, her uygulamadan sonra sınavı alan adaylarla belirli bir süreliğine paylaşılmakta ve puanlar beş yıl süreyle geçerli kabul edilmektedir (ÖSYM, 2020). Maddelerin adaylarla paylaşılması her formun farklı maddelerden oluşturulmasını, sınav puanının bir yıldan fazla süreyle geçerli olması ise farklı yıllarda alınan puanların eşdeğer kabul edilmesini doğurmuştur. İki farklı uygulamanın puanlarının eşdeğerlik kabulü ancak formların paralel olmasıyla sağlanabilir (Kan, 2011). Zhu'ya (1998) göre iki formun aynı içerikte ve güçlükte olduğunu belirlemek zorken Tanguma'ya (2000) göre ise test geliştiriciler kapsam ve istatistiki özellikler bakımından benzer formlar geliştirmeye çalışsalar da formlar arasında güçlük bakımından farklılıklar olması kaçınılmazdır. Aynı testin farklı oturumlarına katılmış adaylar birbiri ile karşılaştırılmak isteniyorsa puanlar arasında istatistiksel bir dönüşüm süreci geliştirilmelidir. Dönüşüm süreci ancak test eşitleme ile mümkündür. Test eşitleme; bir formun birim sisteminin diğer formun birim sistemine dönüştürülmesidir. Dönüşümden sonra formlardan elde edilen puanlar eşdeğer kabul edilir (Angoff, 1971). Eşdeğer puanlar, formların farklı güçlükte olmasından kaynaklanan puan farklılıklarının minimize edilmesiyle elde edilir. Bu durum testi alanların adil şekilde puanlanması için gereklidir (Dorans, Moses ve Eignor 2010).

Ülkemiz son yıllarda uluslararası ölçekteki araştırmalara katılma kararlılığı göstermektedir. Uluslararası Öğrenci Değerlendirme Programı-PISA (Programme for International Student Assessment), Uluslararası Matematik ve Fen Eğilimleri Araştırması-TIMSS (Trends in International Mathematics and Science Studies) ve Uluslararası Okuma Becerilerinde Gelişim Araştırması-PIRLS (Progress in International Reading Literacy Study) gibi belirli döngülerle uygulanan bu araştırmalar katılımcı ülkeler hakkında detaylı sonuçlar üretmektedir. PISA programı, Ekonomik İşbirliği ve Kalkınma Örgütü (OECD) tarafından 2000 yılından itibaren üçer yıllık döngüler halinde uygulanmaktadır. Zorunlu eğitim evresini bitiren 15 yaş grubu öğrencilerin modern toplumda yer edinebilmek için gerekli bilgi ve becerilere ne derece sahip olduğunu belirlemeyi amaçlar. Ülkeler, öğrencilerinin sahip olduğu yeterlik seviyesini diğer ülkeler ile karşılaştırarak eğitim

politikalarının güçlü ve zayıf yönlerini görme fırsatı yakalar. Bu sayede belirli standartlar geliştirme konusunda fikir edinirler (MEB, 2018).

PISA 2018 uygulamasında tek oturumda ortak maddeler içeren 72 farklı kitapçık kullanılmıştır. Kitapçıklarda birey kaynaklı ortalama performans farklılıklarının oluşmaması için kitapçıklar öğrencilere rastgele dağıtılmıştır. Bu sayede her kitapçık için eşdeğer gruplar elde edilmeye çalışılmıştır. Kitapçıklar farklı maddelerden oluştuğu için güçlük farklılıkları oluşmuş ve bu durum öğrenci performanslarının doğrudan karşılaştırılmasını engellemiştir. Bu yüzden test uygulayıcısı öğrenci puanlarını kıyaslayabilmek için kitapçıkları ortak bir ölçeğe yerleştirme yoluna gitmiştir (OECD, 2019b). Böylelikle puanlar arasında yapılan dönüşümlerle eşitleme çalışması yürütülmüştür.

PISA 2018'e 79 ülke katılmış bilişsel maddeler 50 farklı dile çevrilmiştir. PISA ölçme araçlarının geçerliği ve uluslararası temsilini sağlamak için ülkelerin aktif katılımını önemsemiştir. İçerik üretimine katkıda bulunmak isteyen ülkeler için teşvikler sunmuş hatta 2015 yılında katılımcıları madde geliştirme sürecine dahil etmek üzere 3 atölye çalışması düzenlemiş 28 ülkeden 55 madde yazarı çalışmalardan yararlanmışır. Çalışmalar sonucunda ülkeler kendi dillerinde ürettikleri içeriklerle PISA madde havuzuna katkı sunmuştur. Bahsi geçen çalışmalar 2018 döngüsünde yeni oluşturulan maddeler için uygulanmış olsa da önceki döngüler için de benzer çalışmalar yürütülmüştür. Örneğin bu döngüde kullanılan Fen okuryazarlığı maddeleri 2012 ve 2015 döngülerinde kullanılmış maddelerden oluşmaktadır. Maddelerin gönderilme dili Norveççe, Almanca, İngilizce, Japonca, İspanyolca, Flemenkçe ve Fransızca'dır. Sonrasında ise PISA tarafından tüm maddeler iki kaynak dil olan İngilizce ve Fransızca'ya çevrilerek ülkelere kendi dillerine çevirmesi için gönderilmiştir. Bu aşamada çeviriyi kolaylaştıracak, maddelerin olabildiğince kültürel açıdan tarafsız olmasını sağlayacak yönergeler hazırlamıştır (OECD, 2018b).

Gerekli önlemler alınmaya çalışılmış olsa da hem ulusal hem de uluslararası uygulamalardaki maddelerin yanlılığı üzerine yapılan araştırmalarda test maddelerinin cinsiyet, kültür, okul türü, müfredat farklılıkları ve çeviri sınırlılıkları gibi değişkenlere göre DMF gösterdiği görülmüştür (Abbott, 2007; Allalouf, Hambleton ve Sireci, 1999; Bakan Kalaycıoğlu, 2008; Ercikan, 2002; Gür, 2019; Kabasakal Atalay ve Kelecioğlu, 2012; Yurdugül ve Aşkar, 2004). Test

maddelerinin yanlılığına yönelik yapılacak çalışmalar test puanlarının güvenilirliği, puanlardan yola çıkılarak yapılacak çıkarımlar ve alınacak kararların geçerliği için bir gerekliliktir (Karakaya ve Kutlu, 2012). İki formun eşitlenmesinde kullanılan maddelerin DMF göstermesi, eşdeğer kabul edilen puanların güvenilirliği ve geçerliği konusunda da etkili olacaktır. Dolayısıyla eşitleme hatasının azaltılması için DMF gösteren maddelerin testten çıkarılması bir yol olarak düşünülebilir.

Araştırmanın Amacı ve Önemi

Ülkelerin uluslararası düzeyde yapılan değerlendirmelere ilgisi giderek artmakta, değerlendirme sonuçları ülkelerin eğitim bakanlıkları tarafından daha değerli bulunmaktadır. Çünkü ülkeler eğitim girdilerini, süreç ve başarılarını karşılaştırma imkânı bulmakta böylece eğitim politikalarını gözden geçirme fırsatı yakalamaktadırlar (Ercikan, 1998).

Ülkelerin uluslararası araştırmalara her geçen gün daha fazla önem vermeleri PISA'ya katılımı artırmış ve 2018 döngüsünde en yüksek noktaya ulaşarak 79 ülkenin katılımıyla sonuçlanmıştır. Ülkelerin geniş katılımı, uygulamada dil çeşitliliği sağlasa da uygulamayı alma dili olarak en çok İngilizce ve İspanyolca tercih edilmiştir. Dünyanın farklı yerlerinde yaşamasına karşın aynı dilde sınav olmayı tercih eden bireyler için maddelerin aynı özelliği ölçtüğünden emin olmak gereklidir. Çok dilli uygulamalarda temel endişe maddelerin hedef dile çevrildikten sonra kaynak dildeki haliyle eşdeğer olup olmadığıdır. Çünkü ifadeler herhangi bir kültür veya cinsiyet için aşinalık oluşturduğunda bir DMF kaynağına dönüşecektir (Ercikan, 1998, 2002).

Bu araştırmada PISA kitapçıklarında ortak maddelere dayalı yürütülecek olan bir test eşitleme çalışmasında, cinsiyet ve dil değişkenine göre DMF gösteren ortak maddelerin testten çıkarılmaları durumunda eşitlenmiş puanları ve eşitleme hatasını nasıl etkilediğinin gözlenmesi amaçlanmıştır. Cinsiyet ve dil değişkenine göre DMF gösteren maddelerin belirlenmesinde, DMF belirlemedeki sağlam teorik temellerinden ötürü (Ackerman, 1992) Mantel-Haenszel ve SIBTEST teknikleri kullanılmıştır. Veri yapısı gereği eşitleme deseni olarak eşdeğer olmayan gruplar ortak test deseni kullanılmış, Klasik Test Kuramına (KTK) göre temellendirilmiş eşitleme yöntemleri tercih edilmiştir. Madde Tepki Kuramı (MTK), eşitleme çalışmalarında sunduğu birtakım esnekliklerden dolayı alanda kullanımı daha

yaygındır. Ancak MTK'ya dayalı eşitleme kavram ve uygulama açısından daha karmaşıktır. Yönteminin daha basit, mantığını açıklamanın daha kolay ve altta yatan varsayımların gerçeğe daha yakın olduğu (Livingston, 2014) düşüncesiyle KTK temelli eşitleme yoluna gidilmiştir.

Konuyla ilgili ulusal ve uluslararası literatür tarandığında çok az çalışmaya rastlanmıştır (Atalay Kabasakal, 2014; Chu, 2002; Demirus, 2015; Gübeş ve Uyar, 2020; Huggins, 2014; Turhan, 2006; Yurtçu ve Güzeller, 2017). Yapılan araştırmaların genelinde simülasyon veri kullanılmış ve eşitleme çalışmaları MTK temelli yürütülmüştür.

Bu araştırmayı önemli kılan uluslararası yapılan bir araştırmadan elde edilen gerçek verilerle, KTK temelli yürütülüyor ve etkisinin aynı araştırma üzerinde inceleniyor olmasıdır. Böylece araştırma sonucunda elde edilen bulguların uygulama üzerindeki etkilerinin gözlemlenmesi amaçlanmaktadır.

Araştırma Problemi

PISA 2018 uygulamasındaki fen okuryazarlığı öğrenci puanlarının ortak maddelere dayalı olarak eşitlenmesinde, DMF gösteren maddeler eşitleme puanlarını ve eşitleme hatasını nasıl etkilemiştir?

Alt problemler.

1. PISA 2018 Fen okuryazarlığına ait öğrenci puanlarının ortak maddelere dayalı olarak eşitlenmesinde, doğrusal ve doğrusal olmayan eşitleme yöntemiyle elde edilen eşitlenmiş puanlar ve eşitleme hataları nasıldır?
2. Dil kaynaklı DMF gösterdiği Mantel-Haenszel ve SIBTEST teknikleriyle tespit edilen maddelerin ortak testten çıkarılmaları, doğrusal ve doğrusal olmayan eşitleme yöntemiyle elde edilmiş puanları ve eşitleme hatasını nasıl etkilemiştir?
3. Cinsiyet kaynaklı DMF gösterdiği Mantel-Haenszel ve SIBTEST teknikleriyle tespit edilen maddelerin ortak testten çıkarılmaları, doğrusal ve doğrusal olmayan eşitleme yöntemiyle elde edilmiş puanları ve eşitleme hatasını nasıl etkilemiştir?

Sayıtlar

1. Öğrencilerin test maddelerini maksimum performanslarını yansıtacak şekilde cevapladıkları,
2. PISA 2018 uygulamasının araştırmaya katılan tüm öğrenciler için eşit koşullarda yürütüldüğü varsayılmıştır.

Sınırlılıklar

PISA 2018 uygulamasında 18 ve 24. kitapçıkları İspanyolca veya İngilizce dillerinde bilgisayar ortamında alan bireylerin fen okuryazarlığı madde yanıtları ile sınırlıdır.

Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Araştırmanın bu bölümünde yanlılık, değişen madde fonksiyonu ve belirleme yaklaşımları, test bağlama yöntemleri, eşitleme özellikleri, desenleri, yöntemleri hakkında araştırmanın sınırları kapsamında bilgilendirme yapılmıştır. Sonrasında ise yurt içinde ve yurt dışında konu ile ilgili yapılmış araştırmalar hakkında bilgilendirmeler yer almaktadır.

Yanlılık

Madde yanlılığı çalışmaları tarihte ilk olarak Alfred Binet tarafından 1910 yılında başlamıştır. Binet, düşük sosyo-ekonomik düzeyden çocuklarla yaptığı çalışmalarda bazı maddelerin çocukların bilişsel yeterliklerinden öte evde veya okuldan edindiği kültürel öğeleri de ölçüyor olabileceğini düşünmüştür. Aynı zamanda 1912'de Stern (1914) Almanya'da sosyal sınıf farklılığını incelemiş ve anlamlı ölçüde farklılıklar gözlemlemiştir. Çalışmasında bir sınıf için açıkça avantajlı olarak tanımlanan testlerin belirlenmesi gerekliliğine değinmiştir. Binet ve Stern sosyal sınıf ve IQ arasındaki ilişkiyi belirlemiş olsa da zamanın araştırmacılarının aklındaki soru, gözlenen yetenek farklılıklarının sebebinin genetikle mi yoksa çevreyle mi açıklanabilir olduğu olmuştur. Yanlılık üzerine ilk modern çalışmanın Eells, Davis, Havighurst, Herrick ve Tyler (1951) tarafından başlatıldığı söylenebilir. Eells ve diğerleri, ölçülen bazı farklılıkların net olarak yeteneğin yansıması olmayacağı olasılığına odaklanmışlar ve test maddelerinin de kaynak olabileceği üzerine ışık tutmuşlardır (akt. Camilli ve Shepard, 1994).

Yanlılığı; Camilli ve Shepard (1994) belirli bir grup üyelerinin bir testten elde ettiği puanların sistematik hata içermesi veya geçersiz olması, Lord (1980) bir maddenin bir grup için diğerinden farklı yanıt işlevi göstermesi, Shepard, Camilli ve Averill (1981) ise bir gruba diğerinden daha fazla zarar veren bir tür geçersizlik olarak tanımlamışlardır. Geçerlik için potansiyel bir tehdit olan yanlılık, aslında bir grubun diğerine nazaran adil olmayacak ölçüde desteklenmesidir (Clauser ve Mazor, 1998). Bu noktada sistematik hatanın varlığından söz edilebilmekte çünkü maddenin lehine davrandığı grup (etnik köken, cinsiyet, sosyo-ekonomik düzey, kültür gibi) değişkenine göre hatanın yönü ve büyüklüğü belirlenebilmektedir (Kelecioğlu, Karabay ve Karabay, 2014).

Yanılığın tespiti istatistiksel ve yargısal olmak üzere iki aşamalı bir süreçte gerçekleşmektedir (Karakaya ve Kutlu, 2012; Zumbo, 1999). İlk aşama, maddeyi doğru yanıtlama olasılığının aynı yetenek seviyesinde ve farklı gruplardaki bireyler için değişip değişmediğini sorgulayan değişen madde fonksiyonu (DMF) analizi ile başlar. İkinci aşamada ise DMF varlığı tespit edilen maddelerin gruplar için farklı işlev gösterme gerekçesi yetenek seviyesinden mi yoksa sistematik hataya dayalı yanlılıktan mı kaynaklandığı alan uzmanlarınca değerlendirilir (Zumbo, 1999). Değerlendirme sonucunda maddenin yanlı olduğuna karar verilmişse geçerliği artırmak için madde testten çıkarılmalıdır (Lord, 1980).

Ancak kafa karışıklığına sebep olabilecek bazı durumlar da olabilir. Örneğin Amerikalı ve Kanadalı çocukların bulunduğu bir topluluğa uygulanan bilgi testinde “*Amerikan kongresinde kaç senatör vardır?*” şeklinde bir soru sorulmuş olursa bu maddeyi doğru yanıtlama olasılığı, iki milliyette eşit yeterlik düzeyindeki çocuklar için aynı olmayacak ve Kanadalı çocuklar aleyhine yanlılık oluşturacaktır (Ironson ve Craig, 1982). Bunun yanısıra doğru cevap vermenin hem sayısal işlem yeteneğine hem de okuma becerisine dayalı olduğu bir matematik maddesi var olsun ve gruplar ise sayısal işlem yeteneğine göre eşleştirilsin. Eğer bir grup diğerine göre okuduğunu anlama becerisi bakımından yeterliği düşük olursa gruplar arasındaki performans farklılığı olası bir DMF kaynağı olacaktır. Ancak bu farklılık yanlılıktan öte bir yetenek farkının da habercisi olabilir. Alanyazında madde etkisi (item impact) olarak adlandırılan bu durum, özünde gruplar arasındaki yeterlik farklarından kaynaklı ortaya çıkan performans farklılığıdır. Sonuç olarak DMF, yanlılık için gerekli bir koşul olsa da tek başına yeterli değildir (Clauser ve Mazor, 1998). Uzman değerlendirmesi belirleyici rol oynamaktadır.

Değişen madde fonksiyonu (Differential item functioning). Camilli ve Shepard, (1994) yanlılık ve DMF terimleri arasındaki farklılığının arkasında yatan gerekçeden habersiz olarak alanyazında sıklıkla birbiri yerine kullanıldığını belirtmiştir. Oysaki DMF, aynı yeterlik düzeyinde farklı alt gruplardaki bireylerin ilgili maddeyi doğru cevaplama olasılığının farklılaşmasıdır (Angoff, 1993). DMF istatistiği tüm maddelerin gruplar için farklı işlev gösterip göstermediği ile ilgilenirken, yanlılık ise DMF’nin tespitinden sonra maddelerin niçin farklı gruplar için göreceli olarak daha zor olduğunu tanımlamaya yarayan mantıksal bir analizi daha barındırır (Camilli ve Shepard, 1994).

DMF, tek biçimli (TB) ve tek biçimli olmayan (TBO) olmak üzere iki başlık altında incelenebilir. TB-DMF grup üyeliği ve yetenek düzeyi arasında etkileşim olmadığına yani bir madde tüm yetenek düzeyleri boyunca sadece tek bir gruba avantaj/dezavantaj sağlıyor olduğunda ortaya çıkar. TBO-DMF ise iki grup arasındaki maddeyi doğru yanıtlama olasılığının tüm yetenek seviyelerinde aynı olmaması olarak tanımlanır yani grup üyeliği ve yetenek seviyeleri etkileşimlidir. Yetenek düzeyinin belirli bir kısmına kadar bir grup için, sonrasında diğer grup için bir farklılaşmadan bahsedilir (Ellis ve Raju, 2003; Mellenbergh, 1982; Swaminathan ve Rogers, 1990).

DMF gösteren maddelerin tespit sürecinde, bireylerin belirli yetenek düzeylerinde eşleştirilmeleri için kriter olarak toplam test puanları kullanılmaktadır. Toplam puanlar kusursuz bir kriter olmasa da, her bir maddenin ölçmek istediği özelliği daha güvenilir olarak ölçmesi, tüm bireyler için aynı koşullarda elde edilmiş olması gibi nedenlerden ötürü en iyi kriter olarak düşünülebilir. Ancak DMF analizi eşleştirme kriterinin tüm gruplar için geçerli ve yansız olduğunu varsaymakla başlamaktadır (Dorans ve Holland, 1993). Bu nedenle DMF'li maddeler toplam puanın hesaplanmasına katkı sunuyorsa puanların geçerliği hakkında ciddi soru işaretleri olacaktır.

Bir DMF analizinde eşleştirme kriterinin arındırılmasının gerekliliği akıldan çıkarılmaması gereken bir durumdur (Zumbo, 1999). Arındırma (purification), DMF belirleme sürecinde eşitleme kriterinin DMF içermeyen maddelerden hesaplanıyor olmasıdır (French ve Maller, 2007). Eşleştirme kriterinin arındırılmasında iki temel yaklaşım ele alınmaktadır. Bunlar iki aşamalı (two-step) arındırma ve yinelemeli (iterative) arındırma. Yaklaşımlar DMF'li maddeler hakkında karara varmak için kullanılan temel DMF analizinden önce yapılan ön analiz sayısına göre ayrışmaktadır (Lee ve Geisinger, 2016).

İki aşamalı yaklaşımda, ilk aşamada ön analizle DMF'li maddeler belirlenir ve toplam puandan çıkarılır, ikinci aşamada ise arındırılmış toplam puan üzerinden temel DMF analizi yürütülür (Holland ve Thayer, 1988). Holland ve Thayer bu yaklaşımı Mantel-Haenszel (MH) tekniği için önermişler ve arındırma yapılmamış duruma göre daha yüksek doğruluk düzeyinde DMF tanımlaması yapmışlardır.

Yinelemeli arındırma prosedürü Log-lineer ve Madde Tepki Kuramına (MTK) dayalı teknikler için geliştirilmiştir. Analiz sürecinde, öncelikle ön analiz yürütülür ve DMF gösteren maddeler belirlenir sonrasında ise belirlenen maddeler toplam puana dahil edilmeden toplam puanlar tekrar hesaplanır. Yeni toplam puanlar eşitleme kriteri olarak baz alınır ve analiz tüm maddeler için tekrar edilir, DMF gösteren madde tükeninceye veya en son iki analizde aynı DMF'li veri seti elde edilinceye kadar analize devam edilir (French ve Maller, 2007). Böylelikle olabildiğince yanlı madde etkisinden temizlenmiş toplam puanlar ile DMF analizi yapmaya imkan sağlanmış olur. French ve Maller lojistik regresyon (LR) tekniğine dayalı yürüttüleri çalışmalarında yinelemeli arındırmayla yinelemeli olmayan prosedüre göre daha yüksek oranda DMF varlığı tespit edilebildiğini gözlemlemişlerdir.

DMF belirleme yaklaşımları. Test maddelerin farklı alt gruplar için denkliliğini ve karşılaştırılabilirliğini sınamak için DMF analizi yapılması gerekliliği bilinmekte ve DMF varlığının tespiti için birçok teknik uygulanabilmektedir (Ercikan, 2002). Potenza ve Dorans, (1995) çalışmalarında DMF belirleme yöntemlerini; madde puanının iki veya çok kategorili olması, gözlenen veya örtük puan yaklaşımını kullanma, parametrik olma veya olmama durumlarına göre ayırtmışlardır. Potenza ve Dorans'ın (1995) sınıflandırması Tablo 1'de verilmiştir.

Tablo 1

DMF Belirlemede Teknik Sınıflandırması

Madde Puanı	Gözlenen/Örtük	Parametrik	Parametrik Olmayan
	Gözlenen	Lojistik Regresyon (LR)	Mantel-Haenszel (MH) Standartlaştırma
İki Kategorili	Örtük	Genel LR Sınırlanmış LR Loglineer LR Olasılık Fark İndeksi Lord'un Ki-karesi	SIBTEST
	Gözlenen	LR	Standartlaştırma HW1 HW3 Genelleştirilmiş MH
Çok Kategorili	Örtük	Genel LR Kısmi Kredi Modeli (PCM)	SIBTEST Genelleştirilmiş PCM

DMF belirlemeye yönelik çalışmalarda kararların bir teknikle elde edilen sonuçlara dayandırılması yanlış kararlar almaya neden olabilir. Bu bakımdan analizlerin birden fazla teknikle yürütülmesi önerilir (Camilli ve Shepard, 1994). Bu çalışmada Mantel-Haenszel (MH) ve Simultaneous Item Bias Test (SIBTEST) teknikleri kullanılmıştır. MH ve SIBTEST teknikleri yanlılığı madde etkisinden ayırmada güçlü olan teorik temelleri nedeniyle tercih edilmiştir (Ackerman, 1992). Çalışmanın kapsamı gereği bu teknikler hakkında detaylı bilgi verilmiştir.

Mantel-Haenszel (MH) tekniği. Gözlenen puana dayalı parametrik olmayan teknik ilk olarak Mantel ve Haenszel (1959) tarafından retrospektif bir çalışmaya ait verilerin analizi için önerilmiştir. Sonrasında Holland (1995) sınavı alan alt gruptaki bireylere farklı işlev gösteren maddelerin belirlenmesinde pratik ve güçlü bir teknik olarak kullanılabileceğini göstermiştir (akt. Holland ve Thayer, 1988).

DMF analizleri bahsedildiği üzere maddelerin farklı işlev gösterip göstermediğinin tespiti için iki alt grup üzerinde incelenir. Bu gruplardan birincil olarak ilgili olunan gruba odak (O), odak grubun performansının karşılaştırıldığı bir standart olarak değerlendirilen diğer gruba ise referans grup (R) denir (Holland ve Thayer, 1988).

MH, ki-kare istatistiğine dayananan, odak ve referans gruptaki bireylerin toplam test puanlarınca eşleştirilerek her bir madde için 2×2 ' lik olasılık tablosunun oluşturulduğu bir DMF analiz tekniğidir (Bakan Kalaycıoğlu, 2008). Tablo 2'de Dorans ve Holland'ın (1993) toplam puanı m olan bireylerin bir maddeye verdikleri cevaplar ile oluşturulmuş olasılık tablosu verilmiştir.

Tablo 2

Toplam Puanı m Olan Bireylere Ait Olasılık Tablosu

Grup	Madde Puanı		
	Doğru	Yanlış	Toplam
Odak	D_{om}	Y_{om}	N_{om}
Referans	D_{rm}	Y_{rm}	N_{rm}
Toplam	D_{tm}	Y_{tm}	N_{tm}

Olasılık tablosundan yararlanılarak yokluk ve alternatif hipotezler aşağıdaki gibi oluşturulabilir.

$$H_0: [D_{rm} / Y_{rm}] = [D_{om} / Y_{om}] \quad m = 1, \dots, M$$

$$H_1: [D_{rm} / Y_{rm}] = \alpha [D_{om} / Y_{om}] \quad m = 1, \dots, M \text{ ve } \alpha \neq 1$$

Burada alternatif hipotez tüm M eşleştirme seviyelerinde odak ve referans grubun ilgili maddeyi doğru cevaplama oranlarının aynı olduğunu savunan yokluk hipotezine karşı ki-kare analizi ile test edilir (Holland ve Thayer, 1986). α parametresi 2×2 olasılık tablosunun ortak olasılık oranı (common odds ratio) olarak adlandırılır (Zwick ve Ercikan, 1989).

$$\alpha_m = [D_{rm} / Y_{rm}] / [D_{om} / Y_{om}]$$

“Çünkü alternatif hipotez altında α 'nın değeri, tüm m için aynı olan olasılık oranıdır” (Dorans ve Holland, 1993). MH aynı zamanda sabit olasılık oranının bir kestirimini sağlar.

$$\alpha_{MH} = \frac{\sum_m D_{rm} Y_{om} / N_{tm}}{\sum_m D_{om} Y_{rm} / N_{tm}}$$

Yalnız bu eşitlikle, referans aralığı $0 - \infty$ arasında olan bir DMF etki büyüklüğü kestirimi elde edilir. Bu değer test geliştiriciler için çok anlamlı bulunamadığından, yorumlama kolaylığı sağlamak adına olasılık oranının doğal logaritması alınır. Bu sayede olası değerler merkezi 0 olacak şekilde simetrik yerleşim gösterir (Dorans ve Holland, 1993).

$$MH D - DIF(\Delta MH) = -2,35 \ln[\alpha_{MH}]$$

Etki büyüklüğüne yönelik elde edilen $MH D - DIF$ değeri; negatif ise maddenin referans grup lehine, pozitif ise odak grup lehine DMF gösterdiği şeklinde yorumlanmaktadır (Ackerman ve Evans, 1992; Kim, 2003). MH D-DIF'e dayanarak elde edilen değerlerle maddeler üç kategori içinde ayrıştırılır. Bir maddenin hangi kategoriye yerleştirileceğini belirleyen iki unsur vardır. Bunlar mutlak değer

büyüklüğü ve istatistiksel olarak anlamlı olup olmama durumudur. Tablo 3’de Zieky’den (1993) alınan sınıflamaya yer verilmiştir.

Tablo 3

MH Değeri Etki Büyüklüğü Sınıflaması

DMF düzeyi	Değer Aralığı	Açıklama
A	$ \Delta MH < 1$	Yok veya ihmal edilebilir
B	$1 \leq \Delta MH < 1,5$	Orta
C	$ \Delta MH \geq 1,5$	Yüksek

İstatistiksel anlamlılık yanında mutlak değer büyüklüğünü kullanmak önemlidir çünkü katılımcı sayısının çok fazla olduğu durumlarda oldukça küçük bir DMF değerinin anlamlı olduğu gözlemlenebilir (Zieky, 1993).

SIBTEST tekniği (Simultaneous item bias test). Shealy ve Stout, (1993) örtük puan modeline dayalı parametrik olmayan bu tekniği, standartlaştırma indeksinin bir uyarlaması olarak DMF tespiti için önermiştir. Yaklaşım çok boyutlu yapılar için geliştirilmiş olsa da tek boyutlu veri yapısı için de uygunluk göstermiştir (Clauser ve Mazor, 1998; Wiberg, 2007). Bireyleri eşleştirmek için gözlenen puan yerine gerçek puanın bir regresyon kestirimini kullanır. SIBTEST adından da anlaşılacağı üzere birçok maddeyi aynı anda analiz etmeye elverişli bir tekniktir. Analizi madde kümeleri üzerinden gerçekleştirmek testin istatistiksel gücünü artırdığı gibi yapılan analiz sayısını da azaltır. Bu sayede I. tip hata kontrol altına alınmış olur (Nandakumar, 1993).

SIBTEST yinelemeli arındırma yapan bir DMF analizidir. Öncelikli olarak tüm maddeler eşleştirme kriteri için işleme alınır, sonra DMF içeren maddeler tespit edildikçe maddeler eşleştirme kriterinden çıkarılır. Bu işlem DMF içermeyen maddelerden oluşan bir alt madde seti oluşturulana kadar devam edilir. Böylece madde seti DMF içerdiğinden şüphelenilen ve şüphelenilmeyen olmak üzere iki alt sete bölünmüş olur (Wiberg, 2007). SIBTEST tekniğinde sınanan yokluk hipotezi,

$$H_0 : \beta = \pi_R(\tau) - \pi_O(\tau) = 0$$

“ β , τ gerçek puanı üzerinde eşleşen referans ve odak gruba ait bireylerin belirli bir maddeyi doğru cevaplandırma olasılıkları (π) arasındaki fark 0’dır.” şeklinde kurulur. β ’nin sezgisel kestirimi olan $\hat{\beta}$ eşleştirme değişkeninde ağırlıklandırılmış madde güçlükleri farkının ortalamasını ifade etmektedir. $\hat{\beta}$,

$$\hat{\beta} = \sum_{k=0}^k p_k d_k$$

yoluyla hesaplanır. Burada p_k , k puanında odak gruptaki bireylerin oranı iken, d_k yine k puanında odak ve referans gruptaki bireylerin düzeltilmiş ortalamaları arasındaki farkı ifade etmektedir. Teknik,

$$SIBTEST = \frac{\hat{\beta}}{\hat{\sigma}(\hat{\beta})}$$

$\hat{\beta}$ ’nin standart hata kestirimi $\hat{\sigma}(\hat{\beta})$ ’ya bölünmesiyle hesaplanır (Abbott, 2007; Kim, 2003; Wiberg, 2007). SIBTEST’te DMF indeksi hesaplanmasında referans ve odak grubun belirli bir eşleştirme seviyesindeki ilgili maddenin güçlük değerleri farkı temel alınır. Devamında odak grubun her eşleştirme seviyesindeki oranına göre hesaplanan ağırlıklı madde güçlük değerleri toplanır. Elde edilen bu değer $\hat{\beta}$ eşleştirme kriterinin kontrolünde kullanılan ağırlıklandırılmış madde güçlüğü farklarının ortalamasıdır (Kim, 2003). Eşleştirme kriterinde gerçek puanlar regresyon düzeltme katsayısı ile kestirilir. Kestirilen $\hat{\beta}$ değeri ile DMF etki büyüklüğü miktarını göstermek mümkündür (Wiberg, 2007). Roussos ve Stout (1996) tarafından hazırlanmış DMF etki büyüklüğü kategorileri Tablo 4’te verilmiştir.

Tablo 4

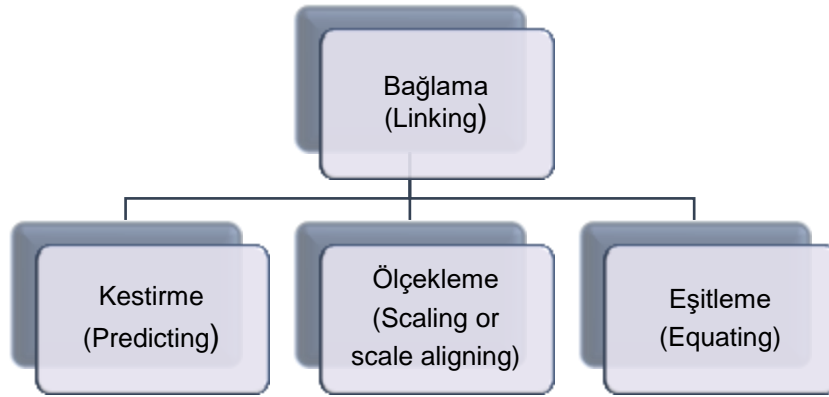
SIBTEST Değeri Etki Büyüklüğü Sınıflaması

DMF düzeyi	Değer Aralığı	Açıklama
A	$ \hat{\beta} < 0,059$	Yok veya ihmal edilebilir
B	$0,059 \leq \hat{\beta} < 0,088$	Orta
C	$ \hat{\beta} \geq 0,088$	Yüksek

İstatistiksel olarak anlamlı; negatif bir β değeri maddenin odak grup lehine, pozitif değeri ise referans grup lehine DMF gösterdiğini belirtir (Abbott, 2007; Clauser ve Mazor, 1998).

Test Bağlama (Linking)

Bir testten veya değerlendirme sonucundan elde edilen puanların bir başka testten ve değerlendirmeden elde edilen puanlarla karşılaştırması psikometrinin konusu olmuştur (Linn, 1993). Holland ve Dorans, (2006) bağlamayı iki test puanının karşılaştırmasına ve birbirine dönüştürmesine yönelik yaklaşımlar içeren genel bir terim olarak tanımlamış ve yaklaşımları çeşitli şekillerde farklı amaçlara hizmet eder nitelikte üç başlık altında sınıflandırmıştır. Holland ve Dorans'ın (2006) sınıflaması Şekil 1'de verilmiştir.



Şekil 1. Test bağlama yöntemleri

Farklı amaçlarından dolayı sınıflanan bu yöntemlerin sıklıkla benzer veya aynı kavramlarmış gibi birbirleri yerine kullanıldığı görülmektedir. Kategoriler arasındaki temel farklılıkların bilinmesi hangi durumda hangi yöntemin daha ilişkili ve gerekli olduğunu anlamak açısından önemlidir (Holland, 2007). Bunun yanı sıra ayrımları çok net olmamakla beraber bir sınıflama daha önerilmektedir. Doğrudan (direct) ve dolaylı (indirect) bağlama şeklinde ayrılan sınıflamada; puanlar arasında direkt fonksiyonel bir bağlantı kuruluyorsa doğrudan, iki puan arasındaki bağlantı 3. bir test veya ölçek aracılığıyla kuruluyorsa dolaylı bağlama olarak tanımlanmaktadır. Kestirme ve eşitleme doğrudan bağlantı yoluyla sağlanırken, ölçeklemenin çeşitli alt kategorileri ise dolaylı bağlantıyla üretilir (Holland ve Dorans, 2006).

Kestirme (predicting) bağlama yöntemlerinden en eski ve en zayıf olanı olarak tabir edilmektedir. Testlerin aynı yapıyı ölçmeleri beklenmez. Ölçümler arasında ilişkinin gücü, kestirimin hassasiyeti açısından önemli bir etkidir. Çünkü kestirim puanları arasındaki ilişkinin deneysel bir tahminidir. Kestirimler büyük ölçüde bağlama, gruba ve zamana bağlı kalınarak yapılır. Örneğin bir alt grup için elde edilen kestirim, diğer altgrup veya bu alt grupların birleşiminden elde edilen grup için geçerli olmayabilir (Linn, 1993; Mislevy, 1992).

Ölçekleme (scaling) genel olarak farklı iki testten elde edilen puanların ortak bir ölçek üzerine yerleştirilmesini amaçlar. Burada temel amaç karşılaştırılabilir puanlar elde etmektir. Bu puanlar farklı kaynaklar (puanlayıcı) veya konu alanlarından elde edilmiş olabilir. Ölçekleme işleminde; karşılaştırılabilir puanların dağılımları kullanılarak X üzerinden alınan bir puan öncelikle ortak ölçeğe dönüştürülür, sonrasında ise Y ölçeğine dönüştürülerek, X'in Y üzerindeki karşılığı elde edilmiş olur. Ölçekleme yöntemi X'ten Y'ye dolaylı bir bağlantı gerektirir. Ancak bu dolaylı bağlantının anlamlılığı birçok faktöre bağlı olarak değişebilmektedir. Ortak ölçek üzerinden dolaylı bağlantı sağlamanın, ölçeklemede nadiren temel amaç olarak kullanıldığı akıldan çıkarılmamalıdır (Holland, 2007; Linn, 1993).

Eşitleme (equating) bağlama yöntemlerinden en güçlü olanı ve en fazla varsayım gerektirenidir. Bir bağlama işleminde; testler eğer aynı yapıyı ölçer, eşit güvenilirlikte puanlar üretir ve uygulandığı gruplardan bağımsız olarak çalışırsa, bu bağlantıdan elde edilen puanlar birbiri yerine kullanılabilir. Bu durumda bu bağlama işlemi eşitleme olarak adlandırılabilir (Huggins ve Penfield, 2012; Kolen ve Brennan, 2014; Dorans, Moses ve Eignor, 2010). Eşitleme işlemi aynı veya farklı zamanlarda bir testin farklı formlarını alan bireylerin puanlarının karşılaştırılmasında kullanılır. Başarılı bir eşitleme işlemi sonrasında bireylerin hangi formu aldığına bakılmaksızın aynı puanı elde etmeleri beklenir (Kolen, 1988).

Eşitleme prosedürü genel olarak yatay (horizontal) ve dikey (vertical) eşitleme olmak üzere iki kategoriye ayrılabilir. Yatay eşitleme, güçlük ve güvenilirlik bakımından benzer olan ve aynı yapıyı ölçen birden fazla test formunun uygulandığı durumlarda kullanılır. Formlardan elde edilen puanların dağılımları da yaklaşık olarak aynıdır (Hambleton ve Swaminathan, 1985). Dikey eşitleme de

aynı genel bilgi ve beceri alanından güçlük düzeyleri istendik şekilde farklı hazırlanmış testler kullanılır. Bu testleri alan farklı yeterlik düzeyindeki birey puanlarının eşitlenmesi hedeflenir (Loyd ve Hoover, 1980). Eşitleme işleminin farklı boyut sayısındaki testler için bir düzenlemeye gidememesi ve dikey eşitleme sonucunda puanların birbiri yerine kullanılamaması gibi sebeplerden dolayı dikey eşitleme yerine dikey ölçekleme (vertical scaling) daha yerinde bir kullanım olacaktır (Felan, 2002). Çalışmanın kapsamı gereği, bu aşamadan itibaren test eşitleme literatürü hakkında bilgilendirme yapılmıştır.

Eşitleme Süreci. Kolen ve Brennan (2014) test eşitleme sürecini istatistiksel işlemlerin uygulanması olarak tanımlamış ve bunları basamaklar halinde aşağıdaki şekilde sıralamıştır:

1. Eşitlemenin niçin yapılacağına karar verilmesi
2. Aynı yapı ve istatistiksel özellikleri içeren alternatif test formlarının yapılandırılması.
3. Formların istatistiksel olarak nasıl ayrıştığı hakkında bilgi verecek veri toplama deseninin seçilmesi
4. Testlerin uygulanması ve belirlenen veri toplama deseni ile verilerin toplanması
5. Kestirim yapılacak formlar arasındaki ilişkinin türüne göre bir veya daha fazla işlevsel eşitleme tanımı hakkında seçim yapılması
6. Formlar arasındaki ilişkiye göre bir veya daha fazla kestirim yöntemine karar verilmesi
7. Eşitleme işlemi yapıldıktan sonra sonuçları değerlendirecek kritere karar verilmesidir.

Kolen ve Brennan, bu adımların yanı sıra eşitlemede süreci yönetenlerin testi alan bireyler adına yargılarını içerdiğini ve bu yüzden süreç yöneticilerine ait deneyim ve birikimlerin karar alınmasında önemli rol oynadığını belirtmişlerdir.

Eşitleme Koşulları. Eşitleme işleminin gerçekleştirilebilmesi birtakım koşulların yerine getirilmesine bağlıdır. Yaygın olarak gerekli görülen beş koşul ve koşulların literatürdeki bulunuşlarının tarihsel sırasına göre yerleşimi; aynı yapıyı

ölçme, eşit güvenilirlik, simetriklik, eşitlik ve grup değişmezliği şeklindedir (Holland ve Dorans, 2006).

Aynı yapıyı ölçme. Eşitlenmesi düşünülen formlar öncelikli olarak aynı yapıyı ölçmeli ve aynı yapının içerik özellikleriyle donatılmalıdır (Bandalos, 2018). Formlar sadece aynı yapıyı ölçmenin yanısıra yetenek, beceri gibi tek bir özelliği ölçmeli, yani tek boyutlu olmalıdır (Woldbeck, 1998). Eğer testler, bireysel puanlanan birden çok homojen alt boyut barındırıyor ise eşitleme alt boyutlar seviyesinde yürütülmelidir (Holmes, 1986).

Eşit güvenilirlik. Araştırmacılar arasında eşit güvenilirlik özelliğinin eşitleme için bir koşul olduğu üzerine tam bir uzlaşma olmadığı görülmüştür. Buna rağmen Dorans ve Holland (2000) eşit güvenilirlik özelliği olmasa güvenilir bir testin tek bir test maddesine eşitlenmenin olası olduğunu belirtmiş lakin güvenilirlik özelliğinin bir koşul olmaktan öte ikincil bir kontrol mekanizması olarak kabul edilmesinin de doğru olacağını eklemiştir. Sonuç olarak temel odaklanılması gereken noktanın güvenilirlik değerinin büyüklüğü olduğu ve güvenilirlik arttıkça eşitleme çalışmasının daha iyi sonuçlar vereceğidir.

Simetriklik. Eşitleme sürecinde hangi forma X hangi forma Y denildiğine bakılmaksızın aynı sonuçların elde edilmesidir (Lord, 1980). Dönüşümün tersinin alınabilir olduğu X formundan Y formuna kurulan eşitleme fonksiyonun tersinden de aynı sonucu vermesi olarak nitelendirilir (Tanguma, 2000). Bu özelliğin regresyon yoluyla kurulabilecek eşitlik ilişkisi için uygun olmadığı genelde regresyonla kurulan eşitleme fonksiyonlarında tersi için farklı fonksiyon elde edildiğine dikkat edilmelidir (Kolen ve Brennan, 2004).

Eşitlik. Bu özellik sınava giren bireyin her hangi bir test formunu almasının farklılık oluşturmadığını ifade eder. Başka bir ifade ile eski form puan dağılımı ile yeni formdaki eşitlenmiş puanların aynı dağılıma sahip olması gerekliliğidir (Lord, 1980). Buna karşılık Kolen ve Brennan (2004), iki formun aynı dağılımı göstermesinin ancak testlerin özdeş olduğu durumda mümkün olacağı, pratikte birbirini aynı testleri yapılandırmanın mümkün olmamasının yanısıra oluşturulduğu varsayıldığında eşitlemeye gerek kalmayacağını, sonuç olarak Lord'un eşitlik koşulunu kullanarak eşitlemenin ya imkansız ya da gereksiz olduğunu belirtmişlerdir. Bu durumdan yola çıkarak, eşit güvenilirlik koşulunda olduğu gibi

eşitlik koşulunun da en mümkün ölçüde karşılandığından emin olunması gerekmektedir (Karagül, 2020).

Grup değişmezliği. Eşitleme fonksiyonunun testi alan bireylerin oluşturduğu alt gruplardan bağımsız olarak işlev göstermesidir (Dorans ve diğerleri, 2010). Grup değişmezliği eşit yapı ve eşit güvenilirlik özelliklerinin niçin gerekli olduğunu açıklamak için de kullanılabilir. Eğer iki test farklı yapıları ölçüyor ve eşit güvenilirlik özelliğini sağlamıyorsa, standart eşitleme yöntemleri sınavı alan belirli alt gruplara göre tutarsız sonuçlar üretecektir (Holland ve Dorans, 2006). Grup değişmezlik özelliğinin derecesi birtakım istatistiki işlemlerle kestirilebilir ve elde edilen değer eşitleme fonksiyonunun cinsiyet, kültür vb. alt gruplar arasında puanların birbirleri yerine kullanılabilir olduğu konusunda fikir verir. Grup değişmezlik özelliği bir bakıma elde edilen eşitleme puanlarının yansızlığı üzerine yapılacak yorumları barındırır. Örneğin sınava giren bir birey, farklı bir grupta aynı performans için farklı eşitleme puanı elde ediyorsa eşitleme fonksiyonunun gruplar arasında adaletli işlediği söylenemez (van Der Linden, 2000). Eşitleme çalışmalarını içeren alanyazın incelendiğinde, gözlenen puana dayalı eşitleme yöntemlerinde grup değişmezlik özelliğinin pek kontrol edilmediği görülmüştür. Lord ve Wingersky'nin (1984) gözlenen puana dayalı eşitleme özelliklerinin grup değişmezliğini sağlayamadığı görüşü alanyazındaki bu görüş ile uyuşmaktadır.

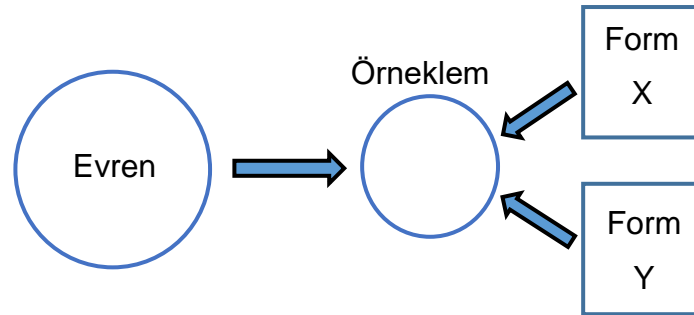
Bu tezin kapsamı gereği test eşitleme süreci gözlenen puana dayalı yöntemlerle yürütülmüştür. Eşitlemenin grup değişmezlik özelliği Lord ve Wingersky'nin görüşü doğrultusunda hareket edilerek incelenmemiştir. Ancak farklı bir yanlılık yaklaşımı olan değişen madde fonksiyonu ile maddelerin cinsiyet ve dil gibi alt gruplarda farklı işleyip işlemediği araştırılmıştır.

Veri Toplama Desenleri. Test geliştirmenin diğer yaklaşımlarında olduğu gibi test eşitleme de veri toplamayla başlar (Zhu, 1998). Başarılı bir eşitleme çalışmasında, test formları arasındaki güçlük farklılığının, testi alan bireylerin olası yetenek farklılıklarından ayrıştırılmasında veri toplama desenleri hayati öneme sahiptir (Deborah, 2018; Holland ve diğerleri, 2006). Hayali bir düşünce tarzıyla aslında test eşitleme için hedef evrenden elde edilmiş iki (daha fazla da olabilir) puan dağılımı gereklidir. Lakin gerçek uygulamalarda verilerin ancak aynı/farklı örneklem ve aynı/farklı formlar üzerinden elde edilebildiği görülür. Bu durumda veri toplama desenleri üç farklı duruma göre şekillenmektedir.

1. Formların aynı gruba uygulanması
2. Formların yetenek bakımından benzer olan gruplara uygulanması
3. Gruplar arasındaki farklılıkları düzenleyecek ortak maddeler içeren formların farklı gruplara uygulanmasıdır.

Bu üç durum, kendi içinde avantaj ve sınırlılıkları olan beş farklı eşitleme deseninin oluşumuna imkan tanımaktadır (Livingston, 2014). Alanyazında yaygın kullanılan veri toplama desenleri şu şekildedir:

Tek grup deseni (Single group design). Eşileme uygulamalarında nadiren tercih edilen bir desendir. İki ya da daha çok test formu aynı cevaplayıcı grubuna uygulanır bu yüzden de örneklem kaynaklı ölçme hatası olabildiğince küçülür (Felan, 2002). Bu desenin gizil varsayımı örneklemden elde edilen sonuçların hedef evrene genellenebileceğidir, tabi bu da birtakım şartları beraberinde getirir. Aslında örneklemin evreni temsil etmesi gerekmez ancak referans formda olan evren örneklem farklılaşmasının yeni formda da korunması gereklidir (Livingston, 2014). Şekil 2.'de birey form dağılımı modellenmiştir.

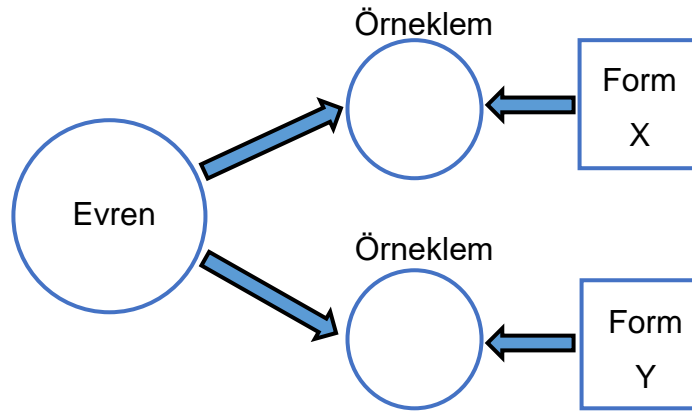


Şekil 2. Tek grup deseni

Tek grup deseninde bireylerin iki formu da almasının doğurduğu bazı dezavantajlar vardır. Uygulanan formlardan ilkinin nispeten zor olması sonraki formun uygulama aşamasında yorgunluk etkisi oluşturabileceği gibi ilk formdan kaynaklı oluşan aşinalık sonraki formdaki başarıyı artırabilir (Kolen, 1988).

Rastgele gruplar deseni (Random groups design). Çoğu eşitleme durumunda bireylerin aynı özelliği ölçen farklı formları alma fırsatı olmaz (Holland ve diğerleri, 2006; Livingston, 2014). Bu desenin uygulandığı test uygulamalarında bireylere formlar sarmallama (spiralling) yöntemiyle atanır. Örneğin ilk bireye A formu, ikinci bireye B formu, bir sonraki bireye A formu verilmek suretiyle devam

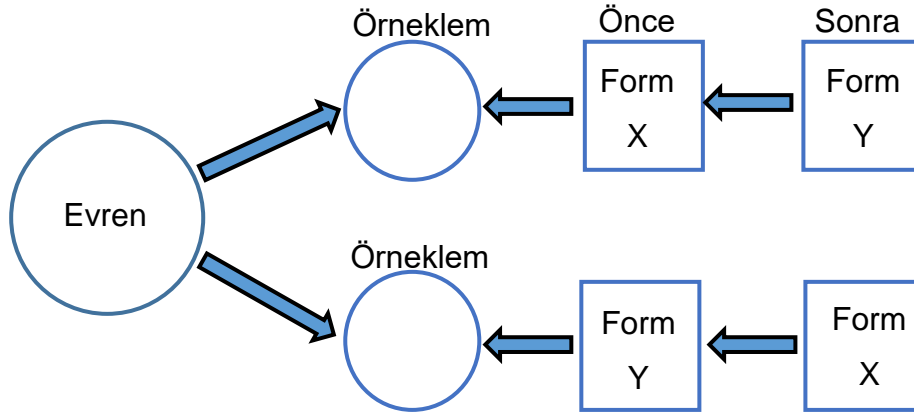
ettirilir ve doğal olarak iki farklı karşılaştırılabilir eşdeğer grup oluşturulmuş olur (Kolen ve Brennan, 1995). Örneklem büyüklüğünün artması ile birlikte grupların eşdeğerlerinin artacağı düşüncesinden dolayı bu desen farklı araştırmacılar tarafından “eşdeğer gruplar deseni” olarak adlandırılmıştır (Dorans, 1990; Dorans ve diğerleri, 2010; Holland ve Dorans, 2006; Holmes, 1986; Livingston, 2014). Rastgele gruplar deseninde formların bireylere nasıl dağıldığı Şekil 3’te gösterilmiştir.



Şekil 3. Rastgele gruplar deseni

Bu desende grup performansları arasındaki değişkenliğin kaynağı olarak formların güçlük farkının olduğu düşünülür (Kolen ve Brennan, 2014). Rastgele gruplar deseninin önemli avantajı bireylerin tek form alması ve tüm bireylerin tüm formları aldığı desenlere göre daha az zamanda uygulanmasıdır (Kolen, 1988). Bunun yanı sıra sıklıkla tercih edilmesine rağmen uygulamada doğurduğu temel kısıtlama ise çok sayıda birey gerektirmesidir hatta aynı kesinlik derecesinde sonuçlar için dengelenmiş grup deseninin 5 ile 15 katı büyüklüğünde örneklem gerektirir (Livingston, 2014).

Dengelenmiş gruplar deseni (Counterbalanced groups design). Tek grup deseni ve rastgele gruplar desenlerinin her ikisini de içeren bir desen yaklaşımıdır. Tek grup deseninde oluşan sıra ve aşinalık etkisini kırmak amaçlanır. Büyük örneklem grubu rastgele olarak iki yarıya ayrılır, bir yarıya önce X sonra Y formu uygulanırken diğer yarıya ise önce Y sonra X formu uygulanır (Dorans ve diğerleri, 2011). Uygulama şeması Şekil 4.’te gösterilmiştir.

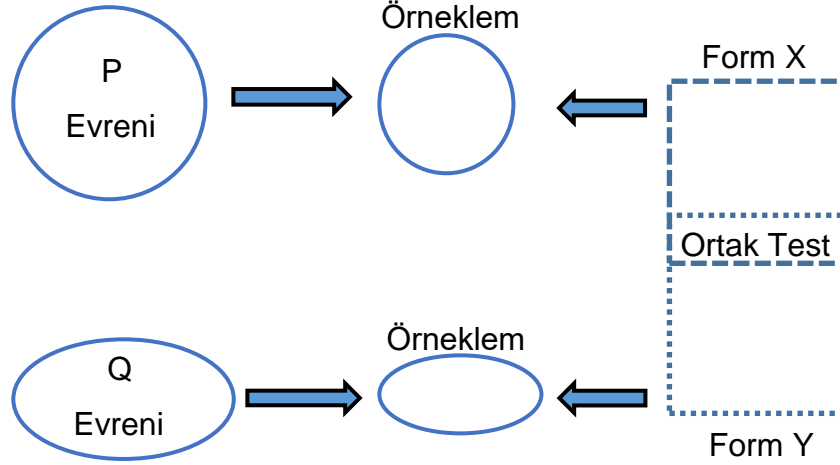


Şekil 4. Dengelenmiş gruplar deseni

Testi alan bireylerin arasındaki olası iletişim durumu gibi uygulama kaynaklı doğabilecek hataları engellemek adına formlar peşi sıra olarak tek sınav oturumunda uygulanabilir (Angoff, 1971). Test uygulayıcısı bu sayede dengelenmiş grup deseninin avantajından yararlanırken testi alan bireyler ise iki ayrı formu bitirmek için uzatılmış bir sınav oturumunun dezavantajı ile karşı karşıyadır (González ve Wiberg, 2017).

Eşdeğer olmayan gruplar ortak test deseni-NEAT (Non-Equivalent groups with anchor test desing). Eşitleme çalışmalarında bireylerin yetenek düzeyleri arasındaki farklılığın önüne geçmek için ideal olan eşdeğer gruplara iki ayrı formu uygulamak olacaktır (von Davier, Holland ve Thayer, 2004). Lakin iki formun gerçekte asla mükemmel düzeyde paralel, eşit güvenilirlikte ve tek boyutlu olamayacağı gibi, grupların eşdeğerliğini sağlayacak örneklem büyüklüğü ise çok nadir durumlarda sağlanacaktır. Bunun yanı sıra her iki gruba da tüm formları uygulamak güvenlik ve uygulamadan doğan birtakım endişelerden dolayı sıkça uygulanabilir de değildir (Cook ve Paterson, 1987). Bu endişeler yüksek risk içeren yılın belli dönemlerinde yapılan uygulamalar şeklinde örneklendirilebilir. Bu sınavlarda herhangi iki dönemde sınavı alan grupların aynı evrene ait olduğu garanti edilemez. Böylelikle eşdeğer olmayan grupların tek bir formu aldığı bir desende eşitleme çalışması yapmak zorunlu hale gelir. Formlara yerleştirilen ortak maddelerle ortak bir bağıntı testi oluşturulur, bu test formlar arasındaki eşitleme için köprü görevi üstlenir. Bu sayede gruplar arasındaki yetenek farkı kontrol altına alınmış olur (Angoff, 1984). Alanyazında bu deseni von Davier ve diğerleri (2004) “eşdeğer olmayan gruplar ortak test deseni (NEAT)” olarak adlandırırken, Kolen

ve Brennan (1995) “ Ortak madde eşdeğer olmayan gruplar (CINEG) ” veya “Ortak test” , “Ortak madde” deseni olarak kullanmışlardır. Grup ve formların desendeki yerleşimi Şekil 5.’te gösterilmiştir.



Şekil 5. Eşdeğer olmayan gruplar ortak test deseni

Yukarıda da belirtildiği gibi bu desen bireylerin yetenek farklılıklarını düzenlemek için kullanılmaktadır (Kolen, 1990). Örneklendirilmesi gerekirse, iki grubun ortak test puanları üzerinde eşit düzeyde başarı sağlaması durumunda, gruplar arasındaki ortalama puan farklılığı ancak formların güçlük düzeyi farkı ile açıklanabilir. Ya da gruplardan birinin ortak testten daha başarılı olması durumunda, tüm testten elde edilecek puanda da başarılı olması beklenir. Eğer ortak test puanları arasında elde edilen fark oransal olarak tüm test puanlarında da devam ediyorsa buradan testlerin paralel olduğu yorumu yapılabilir (Harris, 2003). Desenin bu şekilde işlemesi için ortak maddelerin birtakım özellikleri sağlıyor olması gereklidir. Ortak maddeler form ile aynı yapıyı ölçen, ortalama aynı güçlükte mini bir versiyonudur (Dorans, Pommerich ve Holland, 2007; Kolen ve Brennan, 2014) ve formdan daha kısa olmaları dolayısıyla formdan daha az ve ona yakın güvenirlilikte olmak durumundadırlar. Bu sayede ortak maddeler istatistiksel olarak yanlılığı ortadan kaldırarak, eşitleme puanlarının kestirim kesinliğine katkı sağlarlar (Holland ve diğerleri, 2006). Aynı zamanda ortak test ile tüm test puanları arasındaki yüksek korelasyon gruplar arasındaki yetenek farkının ayırt edilmesinde ortak maddelere olan güveni artırır (Livingston, 2014) .

Bu desenin iki farklı kullanımı bulunmaktadır. Eğer ortak maddelerden elde edilen puanlar bireylerin test puanına dahil ediliyorsa “iç ortak”, ortak maddeler tipik olarak testten ayrı bir bölüm olarak uygulanıyor ve test puanına katkı sunmuyorsa “dış ortak” test olarak adlandırılır (Kelecioğlu, 1994; Kolen ve Brennan, 1995). Bir iç ankor test düzeneği oluşturulacaksa ortak maddelerin her iki formda da yerleşimi, aynı konumda olmalı aynı zamanda maddeleri doğru cevaplama olasılıklarını etkileyecek yorgunluk, motivasyon ve sıra etkisi her grup için eşit düzeyde olmalıdır (Holland ve Dorans, 2006). Ayrıca her iki kullanım durumu için de testteki ortak madde sayısını Angoff (1971) ya en az 20 ya da formdaki madde sayısının % 20'sinden az olmayacak sayıda olmasını ve hangi durumda daha fazla madde elde ediliyorsa onun tercih edilmesini önermiştir. Lakin Budescu (1985) araştırmasında Angoff' un pratik kuralını kullanmak yerine ortak testlerin eşitlenecek formların güvenilirliğini yansıtacak ölçüde madde içermesi konusunda uyarıda bulunmuştur.

Eşitleme Yöntemleri. Eşitleme yöntemleri, yeni formdan elde edilen ham puanları eski form ölçeğine bağlamaya yarayan matematiksel prosedürlerdir (Aşiret, 2014; Dorans ve diğerleri, 2011). Holland ve Dorans (2006) bu prosedürleri geliştirirken üç temel değişkenin belirleyici olduğunu belirterek eşitleme yöntemlerinde sistematik bir sınıflama yoluna gitmiştir. Sınıflama;

1. Veri toplama deseninin ortak bireylerle mi yoksa ortak maddelerle mi oluşturulduğu,
2. Gözlenen mi yoksa gerçek puanlarla mı modellendiği,
3. Formlardan elde edilen ham puanlar arası ilişkinin doğrusal veya doğrusal olmayan yöntemlerle mi kurulduğu şeklinde sorularla yöntemler geliştirilmiştir.

Bunun yanı sıra günümüze kadar bu sınıflandırmalar iki temel kuram altında yapılmıştır. Bunlar “Klasik Test Kuramı(KTK)” ve “Madde Tepki Kuramı(MTK)”'dır.

Klasik test kuramında gözlenen puana dayalı eşitleme yöntemleri. KTK gerçek puan ve hata puanı olmak üzere iki gözlemlenemeyen bileşen içerir. Gerçek puan, testin tekrarlanan uygulamaları sonucunda bireyin elde edeceği beklenen test puanı olarak tanımlanırken, hata puanı ise gözlenen puan ile gerçek

puan arasındaki farktır. KTK, hata puanlarının ortalaması 0 olacak şekilde rastgele olduğunu ve paralel bir testteki hata puanları ile gerçek puanların da ilişkisiz olduğunu varsayar (Hambleton ve Swaminathan, 1985). Eşitleme ilişkilerinde bir gerçek puan, bir ölçmede bir bireyin puan dağılımındaki beklenen gözlenen puanı olarak ele alınır. Sonuç olarak gerçek puan ölçeğinin tanımlanması da gözlenen puan ölçeği üzerinden olmaktadır (Lord ve Novick, 1968). Gözlenen puana dayalı eşitlemenin amacı, yeni testte doğru cevaplanan madde sayısı olarak hesaplanan puanların sanki referans testten elde edilmiş gibi ayırt edilemeyecek biçimde ayarlamaktır (van Der Linden, 2006).

Bu çalışmanın kapsamı KTK'ya göre temellendirilmiş NEAT desende gözlenen puana dayalı eşitleme yöntemlerini içerdiği için onlar hakkında detaylı bilgilendirme yapılmıştır. Lakin okuyucuya temel bir giriş sağlamak adına ortak popülasyonu kullanan desenlerde gözlenen puana dayalı yöntemler hakkında da bilgi verilmiştir. Bu yöntemlerin alanyazında bazı araştırmacılar tarafından "geleneksel eşitleme yöntemleri" olarak adlandırıldığı görülmüştür (González ve Wiberg, 2017; Han ve diğerleri, 1997; Kolen ve Brennan, 2014)

Ortak popülasyonu kullanan desenler için yöntemler. Eşitleme verileri ortak popülasyondan çekilmiş grup(lar)ı içeren tek grup, rastgele gruplar ve dengelenmiş gruplar desenlerinin herhangi birinden elde ediliyorsa eşitleme işlemi nispeten daha basittir. Çünkü bir veya iki değişkenli puan dağılımları doğrudan elde edilebilir bu da eşitleme dönüşümlerinin direk kestirilmesine imkan tanır. Yani tüm veri ortak popülasyondan geldiği için doğrudan gözlemlenebilir, kestirilebilir ve hedef evrene genellebilir (Livingston, 2014) eşitleme dönüşümü için özel bir varsayıma ihtiyaç duyulmaz (González ve Wiberg, 2017).

Özdeş eşitleme (Identity equating). X ve Y formları arasında her hangi bir ayarlama yapmadan eşitleme yapılmasıdır. Tek bir varsayımı vardır; formlar arasındaki güçlük farkının ölçek boyunca aynı olduğudur. Yöntem puanların direk karşılaştırılabilir olduğu üzerine kuruludur (Brossman, 2013; Holland ve diğerleri, 2006; Kolen ve Brennan, 2014) . Albano (2016) özdeş eşitlemenin bir doğrusal eşitleme yaklaşımı olduğunu belirtmiş ve aşağıda verildiği şekilde formülize etmiştir. Denklemden, (x_1, y_1) ve (x_2, y_2) X ve Y ölçeğinde tanımlanmış doğru üzerindeki her hangi iki puana ait koordinatlarıdır.

$$ide_Y(x) = y = \frac{y_2 - y_1}{x_2 - x_1} x + y_1 - \frac{y_2 - y_1}{x_2 - x_1} x_1$$

burada a denklemin eğimi, b denklemin sabiti olmak üzere

$$a = \frac{y_2 - y_1}{x_2 - x_1} \quad ve \quad b = y_1 - \frac{y_2 - y_1}{x_2 - x_1} x_1 \text{ olur.}$$

Özdeş eşitlemenin varsayımı gereği X ve Y formlarının puan ölçekleri özdeştir. Bu bakımdan $a = 1$ ve $b = 0$ olacaktır. Sonuç olarak özdeş eşitleme fonksiyonu:

$$ide_Y(x) = x$$

şeklinde elde edilir. Basit bir örnekle anlatılacak olursa X formundan alınan 40 puan varsayım gereği Y formundan alınan 40 puana eşdeğer kabul edilir (Kolen ve Brennan, 1995). Ayrıca Kolen ve Brennan, iki forma ait güçlük düzeylerinin ölçek boyunca eşit olduğunda özdeş eşitlemenin ortalama ve doğrusal eşitleme ile aynı işlevi göreceğini belirtmişlerdir.

Ortalama eşitleme (Mean equating). Yöntem formların madde güçlükleri arasındaki farkın tüm ölçek boyunca aynı olduğu varsayar, bu varsayımın getirdiği kolaylıktan dolayı uygulaması en basit yöntemlerden kabul edilir (Bandalos, 2018). Kolen ve Brennan, (2004) ortalama eşitleme yöntemini X yeni form, Y eski form, x : X formundan alınan belirli bir puan, y : Y formundan alınan belirli bir puan, $\mu(X)$ ve $\mu(Y)$ ' de sırasıyla X ve Y formlarından elde edilen puanların ortalaması olması kaydıyla şöyle formülleştirmişlerdir:

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

denklem y' ye göre çözüldüğünde

$$mean_Y(x) = y = x - \mu(X) + \mu(Y)$$

Burada $mean_Y(x)$ X formundan elde edilen x puanının ortalama eşitleme yöntemiyle dönüştürülmüş Y formundaki eşitidir.

Tablo 5'te Deborah 'tan (2018 s.555) dönüşümü özetleyen bir örnek durum gösterilmiştir. Tabloda formlardan elde edilmiş betimleyici istatistikler farklı

sütunlarda verilmiştir. X formundan elde edilen ortalama puanın Y formunununkinden 0,4 puan düşük olduğu görülmektedir.

Tablo 5

Ortalama Eşitleme

Test formu	Ortalama	S. sapma	Çarpıklık	Basıklık
X	35,4	10,0	-0,64	-0,42
Y	35,8	10,7	-0,51	-0,68
$mean_Y(x)$	35,8	10,0	-0,64	-0,42

Ortalama eşitleme sonucunda formların ortalamalarının eşitlendiği standart sapma, çarpıklık ve basıklık gibi diğer betimleyici istatistiklerinin değişmediğine dikkat edilmelidir. Sonuç olarak formlar arasındaki ortalama farkını kullanarak eşitleme birçok test durumu için fazla kısıtlayıcı olabilir lakin yöntem eşitleme kavramlarını göstermek adına yararlı olabilir (Kolen ve Brennan, 2014).

Doğrusal eşitleme (Linear equating). Bu yöntem aritmetik ortalama ve standart sapma dışında X ve Y formlarından elde edilen puanların dağılımlarının benzer olduğunu varsayar. Bu da eşdeğer kabul edilen puanların aynı z (standart) puana sahip olmaları anlamına gelir (Crocker ve Algina, 1986). Farklı bir ifade ile formlar arasındaki güçlük farkının ölçek boyunca sabit bir katsayı miktarınca değiştiği düşünülür. Burada sabit olarak düşünülen katsayı, eşitleme denkleminin eğimidir ve eğim X ve Y' nin standart sapmalarının oranından oluşur (Albano, 2016). Denklem şu şekilde tanımlanır:

$$lin_Y(x) = \frac{\sigma_Y}{\sigma_X} x + \mu_Y - \frac{\sigma_Y}{\sigma_X} \mu_X$$

Denklem sadeleştirildiğinde,

$$\frac{x - \mu_X}{\sigma_X} = \frac{y - \mu_Y}{\sigma_Y}$$

$$z_x = z_y$$

olduğu görülür. Tablo 5'deki veriler denklemde yerine konulacak olursa, X formundan elde edilen belirli bir puan x'i ortalama puan olarak aldığımızda;

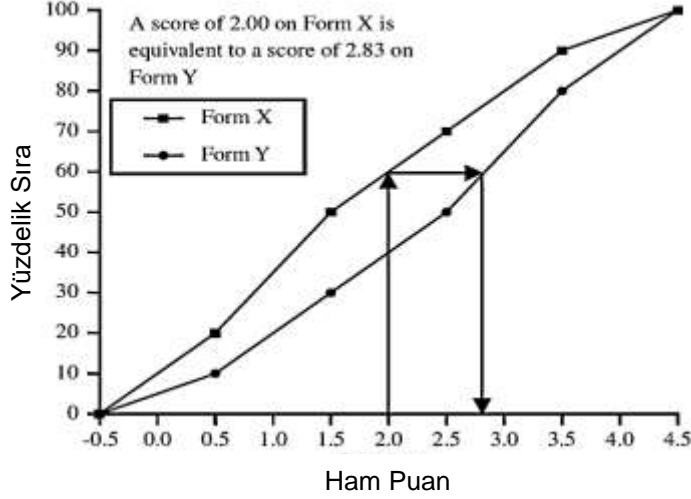
$$lin_Y(35,4) = \frac{10,7}{10,0} 35,4 + 35,8 - \frac{10,7}{10,0} 35,4$$

$$lin_Y(35,4) = 35,8$$

İşlemden de anlaşılacağı üzere, doğrusal eşitlemeden sonra formlar aynı ortalamaya ve standart sapmaya sahip olur (Bandalos, 2018). Yukarıdaki örneği X formundan elde edilmiş belirli bir x puanı için tekrar uygulansa ve bu x puanı X formundan elde edilebilecek maksimum puan olsaydı Y formundaki eşiti için anormal bir durum elde edilebilirdi. Bu durumda Y formundaki eşdeğer puan ölçek için belirlenmiş sınırın dışına çıkacaktı. Böyle durumlar için Kolen ve Brennan, (2004) iki yol önermiştir: ilki eşit puanların referans form ölçeğinin dışına çıkmasına izin verme, ikincisi ise referans form sınırlarını aşan puanları sınırlar içinde tutmak için budama (truncation) yöntemini uygulama. Bunun yanı sıra doğrusal eşitleme sonucunda doğal sayılardan oluşan ham puanların eşiti neredeyse hiçbir zaman bir doğal sayı olmayacaktır. Eğer referans puan ölçeği doğal sayılarda tanımlı ise böyle bir durumda ayrıklık sorunu (discreteness problem) oluşacak ve sorunun aşımı için ara değer atama (interpolate) yoluna gidilmelidir (Livingston, 2014).

Eşit yüzdellikli eşitleme (Equipercetile equating). Doğrusal eşitleme iki puan dağılımının aynı olduğunu varsayar, formların standart sapma ve ortalamalarını kullanarak standart puanlar üzerinden eşitleme işlemini gerçekleştirir. Peki ya puan dağılımları farklı olduğunda ne gibi bir yol izlenmelidir? Puan dağılımları bazı noktalarda birbirine yaklaşıyorken bazı noktalarda fazlaca ayrışıyor ve değişken bir tavır sergiliyorsa eşitleme için seçilecek yöntem eşit yüzdellikli eşitleme olmalıdır (Bandalos, 2018). Eşit yüzdellikli eşitleme çalışması iki basamakta gerçekleştirilebilir: ilk olarak formlara ait puanların göreceli yığılmalı frekans dağılımları tablollaştırılır veya grafiğe dökülür, ikinci aşamada ise göreceli yığılmalı frekans dağılımlarından eşdeğer puanlar elde edilir (Kolen, 1984). Böylelikle aynı yüzdellik sıraya denk gelen puanlar eşdeğer kabul edilir (Albano,

2016; Deborah, 2018; Kolen ve Brennan, 1995; Livingston, 2014). Örnek bir durum Şekil 6.'da grafiksel yöntemle gösterilmiştir. Her iki formda yüzdelik sırası 60 olan puanlar belirlenmiş ve X formunda 2,0 olan puanın, Y formu puan ölçeğindeki eşdeğeri 2,8 olarak kabul edilmiştir.



Şekil 6. Eşit yüzdelikli eşitlemenin grafiksel gösterimi

Kaynak. Kolen ve Brennan, 2014

Grafiksel yaklaşımların çok sayıda gerçek formda uygulanması ve kesin sonuçlar elde edilmesi güç görünmektedir. Bu bağlamda yüzdelik sıraları daha formal bir tanımlamayla açıklayacak analitik prosedüre ihtiyaç duyulmuştur (Kolen ve Brennan, 2004).

Albano (2016) eşit yüzdelikli eşitleme fonksiyonunu X ve Y formlarındaki yığılmalı dağılım fonksiyonlarını birbirine eşitleyerek (Michael, Paul, Meryl, Cadelle, 1999) eğrisel bir tanımlama yapar ve şöyle formüle eder: $F(x)$ ve $F(y)$ sırasıyla X ve Y formlarının yığılmalı dağılım fonksiyonu olmak üzere $F(x) = F(y)$ 'dir. Bu durumda eşitleme fonksiyonu:

$$equip_Y(x) = G^{-1}[F(x)]$$

olur. Formüldeki genel varsayım puanların sürekli, rastgele birer değişken olduğudur.

Eşit yüzdelikli eşitleme yönteminin temel problemi, örneklem kaynaklı olarak puan dağılımında düzensizliklerin olmasıdır. Belirli bir puana sahip sınav katılımcılarının yüzdesi puanlarla beraber kademeli olarak değişmez,

dalgalanmalar oluşturur. Dalgalanmalar neticesinde eşit yüzdelikli, eşitleme sonuçları düzensizlikler üretir ve bu da eşitleme ayarlamasının diğer sınav katılımcı gruplarına genellenmesine engel olur. Böyle durumlarda dağılımı aynı şekle, konuma ve yayılıma sahip ama düzensiz olmayan bir dağılım ile değiştirmektir. Bu tekniğe düzgünleştirme (smoothing) denilmektedir (Livingston, 2014). Düzgünleştirme işlemi eşitlemeden önce ham puan dağılımına yapılıyorsa ön-düzgünleştirme (presmoothing), eşitleme süreciyle elde edilmiş eşitlenmiş puanlara yapılıyorsa son-düzgünleştirme (postsmoothing) olarak adlandırılır (Hanson, Zeng, Colton, 1994). Düzgünleştirme yöntemiyle elde edilen eşitlenmiş puanların düzgünleştirilmemiş eşitleme çalışmasına göre daha az örnekleme hatası içermesi beklenir (Albano, 2016; Kelecioğlu ve Gübeş, 2013; Kolen, 1984). Bu çalışmanın kapsamı gereği düzgünleştirme yöntemleri hakkında detaylı bilgilendirme yapılmamıştır.

Eşit yüzdelikli eşitlemede diğer bir problem ise ayrıklık (discreteness) sorunudur. Ayrıklık, yeni formda belirli bir puana ait yüzdelik sıranın, referans formda elde edilmiş bir puana karşılık bulunmamasıdır (Kolen ve Brennan, 2014). Livingston (2014) ayrıklık sorunu için ara değer atama (interpolation) yöntemini önermiştir. Eşit yüzdelikli eşitlemenin doğası gereği ayrıklık sorunu olacağını lakin ara değer atama yönteminin ürettiği eşitlenmiş puanların yeni formdaki ham puanlarla neredeyse aynı ortalamaya, standart sapmaya ve çarpıklığa sahip olduğunu belirtmiştir.

Daire-yay eşitleme yöntemi (Circle-arc equating). Daire-yay eşitleme yöntemi doğrusal olmayan eşitleme yöntemlerinden biridir. Altında yatan temel mantık ilk olarak Divgi (1987) tarafından eşitleme ilişkisinin doğrusal olmadığı varsayımıyla ortaya çıkmıştır. Divgi' ye göre eşitleme ilişkisini gösteren eğri önceden belirlenmiş iki uç noktadan ve deneysel olarak elde edilmiş bir orta noktadan geçecek şekilde sınırlandırılmıştır. Burada uç noktalar eşitlenecek testteki olası en düşük ve en yüksek puanlar iken, orta nokta ise ortalama puanlara göre belirlenmiştir (akt. von Davier, 2011). Daha sonra Livingston ve Kim (2009) daire-yay yöntemini şu şekilde tanımlamışlardır; üst uç nokta testte alınabilecek olası maksimum puan, alt uç nokta ise bireyin hiç okumadan şansla cevaplayabileceği minimum puandır. Orta nokta ise puan dağılımının ortasındaki puanı işaret etmektedir. Eğer bu üç nokta doğrusal değilse bir dairenin yayı

üzerindeki üç nokta olarak düşünülebilir. Eşitlenecek testlerdeki olası maksimum ve minimum puanlar birbiri ile direk eşitlendikleri için uç noktalar özdeşlik eşitleme fonksiyonu ile eşitlenmiş olurlar. Orta nokta eğer uç noktalar arasında çizilen bir doğru üzerinde değilse, başlangıç noktaları $(x_1, y_1), (x_2, y_2)$ merkezi (x_c, y_c) ve yarıçapı r , olan bir yayın üzerinde yer alır. Eşitleme fonksiyonu bu değerlerle şöyle tanımlanır:

$$circ_Y(x) = y_c \pm \sqrt{r^2 - (x - x_c)^2}$$

karekök içindeki değer, eğer $\mu_Y > id_Y(\mu_x)$ ise y_c 'ye eklenir, $\mu_Y < id_Y(\mu_x)$ ise y_c 'den çıkarılır. Eşitlik durumunda ise daire-yay yöntemi özdeşlik fonksiyonu ile aynı sonucu verir (Albano, 2015).

Alanyazında daire-yay yöntemi daha çok küçük örneklerle yapılan araştırmalarda sıklıkla tercih edilmiş, diğer yöntemlere göre anlamlı düzeyde daha düşük eşitleme hatası ürettiği gözlemlenmiştir (Aşiret, 2014; Livingston ve Kim, 2009, 2010; Özdemir, 2017). Bu araştırmada örneklem büyüklüğü küçük seviyelerde olmasa da daire-yay yönteminin etkinliğinin görülmesi amaçlanmıştır.

Ortak maddeleri kullanan (NEAT) desen için yöntemler. Ortak test kullanılarak yapılan eşitleme yöntemleri ortak popülasyon kullanılan yöntemlere göre daha karmaşık bir yapıya sahiptir. Herhangi bir test eşitleme çalışmasında farklı formları alan farklı evrenden çekilmiş grupların testin ölçmeyi amaçladığı bilgi ve beceriler konusunda eşdeğer olduğu varsayılmaz (Livingston, 2014). Bu durumda gruplar arasındaki yetenek farkını kontrol altına almak için ortak maddeler kullanılır (Kan, 2011). Ortak maddeleri kullanarak oluşturulan yöntemler diğer yöntemlere göre uygulamada esneklik sağlar (Holland ve Dorans, 2006). Lakin bu esnekliğin de bir bedeli vardır; bilindiği üzere tüm bağlama ve eşitleme fonksiyonları tek bir hedef evren üzerinden tanımlanır. NEAT desende ise yapısı gereği P ve Q gibi iki farklı evrenden örneklem içerir, eşitleme ilişkisinin kurulabilmesi için her iki evrenden de bireyleri içeren tek bir evrenin tanımlanması gerekir (Albano, 2016; Kolen ve Brennan, 2014). Bu evrenin tanımlanmasına yönelik Braun ve Holland (1982) sentetik evren (synthetic population) kavramını önermişlerdir. Sentetik evren P ve Q evrenlerinden alınan örneklemelerin ağırlıklı kombinasyonlarından oluşur. w_1 ve w_2 sırasıyla P ve Q popülasyonlarının ağırlıkları olmak üzere, $w_1 + w_2 = 1$ ve $w_1, w_2 \geq 0$ 'dir. Buradan, sentetik evren

$$T = w_1 P + w_2 Q$$

şeklinde tanımlanır (akt. von Davier ve diğerleri, 2004).

NEAT desende gözlenen puana dayalı eşitleme yöntemleri kurulan istatistiksel ilişkinin hangi açıdan evrene genellenebileceği konusunda iki farklı varsayım içerir. İlki eşitleme ilişkisi iken ikincisi koşul ilişkisi sunar. Eşitleme ilişkisi zincir eşitleme (chained equating) yöntemiyle kurulur. Yeni form öncelikle ortak teste, ardından da ortak testte referans forma eşitlenir. Burada ortak test bağlayıcı zincir görevi üstlenir. Böylelikle her zincir eşitleme hedef evrene genelleştirilerek istatistiksel bir eşitleme ilişkisi kurulduğu varsayılır. Koşul ilişkisini ise tabakalama sonrası eşitleme (post-stratification) yöntemi kurar. “Starafication” kavramı ortak test puanının, testi alan her bir grubu sınıflaması için kullanılırken, “post” da bireylerin sınıflamasının ancak tüm veriler elde edildikten sonra yapılıyor olmasını tanımlar. Ortak test üzerindeki her olası puan için hedef evren üzerindeki yeni form ve referans form için puan dağılımları kestirilir. Sonrasında ise sanki gözlemler hedef evrende yapılmış gibi eşitleme ilişkisi kurulur. Kurulan eşitleme ilişkisi, her eşitleme örneğini hedef evrene genelleştiren koşullu bir ilişki olduğunu varsayar (Livingston, 2014). Tablo 6’da eşitleme yöntemleri doğrusal ve doğrusal olmayan ilişkiye bağlı olarak sınıflandırılmıştır.

Tablo 6

NEAT Desende Gözlenen Puana Dayalı Eşitleme Yöntemleri

Yöntemler	Eşitleme ilişkisine dayalı (chained)	Koşullu ilişkiye dayalı (poststratification)
Doğrusal yöntemler	Zincir eşitleme	Tucker Levine Braun–Holland Düşük ağırlıklar
Doğrusal olmayan yöntemler	Zincir eşit yüzdelli eşitleme	Frekans kestirim

Alanyazı incelendiğinde NEAT desende gözlenen puana dayalı farklı eşitleme yöntemlerinin varlığı görülmüştür. Ancak çalışmanın kapsamı gereği Tablo 6’da verilen bazı yöntemler hakkında bilgilendirme yapılacaktır. Daha detaylı bilgi için Albano'ya (2016) bakılabilir.

Tucker yöntemi (Tucker method). Bilindiği gibi grupların eşdeğer olduğu desenlerde doğrusal eşitleme standart puanlar üzerinden yürütülür. Ancak bir NEAT desende doğrusal eşitleme için eşitlenecek formları alan gruplar birleşik bir form oluşturulmak üzere ağırlıklandırılır (Kolen, 1985). Sentetik evren denilen bu ağırlıklandırılmış evrende gözlenen puana dayalı doğrusal eşitleme denklemi; yeni form X, referans form Y ve sentetik evren s olmak üzere,

$$lin_{Y_s}(x) = \frac{\sigma_s(Y)}{\sigma_s(X)} [x + \mu_s(X)] + \mu_s(Y)$$

olur.

Denklemdaki sentetik evren parametreleri evren 1 ve 2 türünden tanımlanacak olursa:

$$\mu_s(X) = w_1\mu_1(X) + w_2\mu_2(X),$$

$$\mu_s(Y) = w_1\mu_1(Y) + w_2\mu_2(Y),$$

$$\sigma_s^2(X) = w_1\sigma_1^2(X) + w_2\sigma_2^2(X) + w_1w_2[\mu_1(X) - \mu_2(X)]^2$$

$$\sigma_s^2(Y) = w_1\sigma_1^2(Y) + w_2\sigma_2^2(Y) + w_1w_2[\mu_1(Y) - \mu_2(Y)]^2$$

şeklinde olur. NEAT desende X formu 2. evrene uygulanmadığı gibi Y formu da 1. evrene uygulanmaz bu nedenle denklemler üzerinde de koyu şekilde belirtilmiş olan $\mu_2(X), \mu_1(Y), \sigma_2^2(X)$ ve $\sigma_1^2(Y)$ parametreleri doğrudan kestirilemez. Tucker ve Levine gözlenen puana dayalı yöntemler yukarıdaki parametreleri kestirilebilir gözlenen parametreler olarak açıklayan birtakım istatistiksel varsayımlar içerir (Kolen ve Brennan, 2014). Tucker yönteminin içerdiği iki varsayım şu şekildedir:

Varsayım 1: Ortak maddelerin üzerinden formların puanlarının kestirimini tanımlayan doğrusal regresyon fonksiyonunun sabiti ve eğimi her iki popülasyon için aynıdır.

Varsayım 2: Her formdaki ortak testin koşullu varyansı her iki popülasyon için aynıdır (Kolen ve Brennan, 1987).

Levine gözlenen puan yöntemi (Levine observed-score method). Bir diğer doğrusal eşitleme yöntemi olan Levine eşitleme yöntemi Levine (1955) tarafından her iki popülasyon içinde gerçek puanlar üzerinden açıklanmıştır. Yöntem özünde gözlenen puanın, hata puanı ve gerçek puanın toplamından oluştuğu ve hata puanlarının gerçek puanlarla ilişkisiz olduğu Klasik Test Kuramına (Lord ve

Novick, 1968) dayalıdır. Bu sayede eşitleme için gözlenen puanlar kullanılıyor olsa da varsayımlar gerçek puanlar üzerinden kurulmuştur (González ve Wiberg, 2017). Levine yöntemi; korelasyonel varsayım, doğrusal regresyon varsayımı ve hata varyansı varsayımı olmak üzere üç varsayım içerir (Chen,Cui,Zhu ve Gao, 2010).

X, Y formlar, V ortak test olmak üzere;

Varsayım 1: V'nin X ve Y ile gerçek puan korelasyonları her iki evren için 1'e eşittir.

Varsayım 2: X'in V üzerinden veya Y'nin V üzerinden elde edilen gerçek puan doğrusal regresyon katsayıları her iki evren için eşittir.

Varsayım 3: X, Y ve V'nin hata varyansları her iki evren için eşittir.

Levine gerçek puan eşitlemesi aynı varsayımlara sahip olsa da kurulan doğrusal denklem açısından az bir farklılık gösterir (Albano, 2016).

Frekans kestirim yöntemi (Frequency estimation method). Frekans kestirim yöntemi tabakalama sonrası eşitleme altında doğrusal olmayan eşitleme yöntemlerinden biridir. Bu yapısından ötürü ilgili alanyazında frekans kestirim eşityüzdelikli eşitleme yöntemi (Frequency estimation equipercentile equating method) olarak da tabir edilmektedir (Jarjoura ve Kolen, 1985). Yöntemsel yaklaşım olarak farklı ancak temel aldığı eşitleme fonksiyonu aynı olması kaynaklı zincir eşit yüzdelikli eşitleme yöntemine rakip olarak kabul edilir (Liou ve Cheng, 1995). Frekans kestirim yöntemi, ortak test puanları ile kestirilen X ve Y formlarına ait koşullu puan dağılımlarının popülasyonlara göre değişmez olduğunu varsayar (Von Davier, Holland, Thayer, 2004a; Wang, Lee, Brennan ve Kolen, 2008) . Varsayım cebirsel yolla izah edilecek olursa:

X, Y formlar, V ortak test ve x, y, v sırasıyla bu testlerden elde edilmiş toplam puanlar olmak üzere;

$$f_1(x|v) = f_2(x|v)$$

$$g_1(y|v) = g_2(y|v)$$

'dir. Varsayım sentetik grup fonksiyonu üzerinden sürdürülürse,

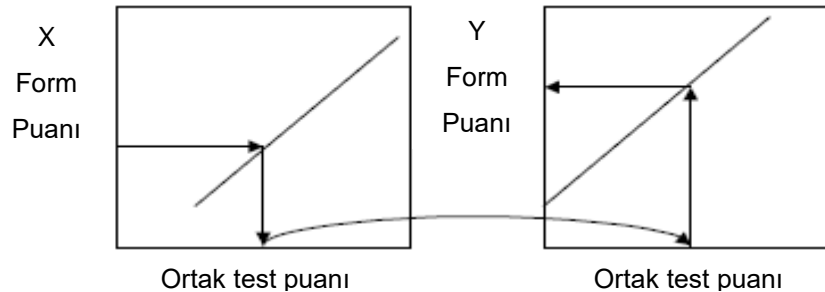
$$f_s(x) = w_1 f_1(x) + w_2 \sum_v f_1(x|v) h_2(v)$$

$$g_s(y) = w_1 \sum_v g_2(y|v)h_1(v) + w_2 g_2(y)$$

olur. Burada $h_1(v)$ ve $h_2(v)$ 1 ve 2. evrendeki ortak test puanlarının marjinal dağılımıdır. Eşitliklerdeki parametreler frekans kestirim yönteminin koşullu puan dağılımı popülasyon değişmezliği varsayımıyla gözlenebilir duruma gelmiştir. Hesaplanabilir olan $f_s(x)$ ve $g_s(x)$ değerleri X ve Y formlarının yığılmalı dağılım fonksiyonlarını üretmek için kullanılacak, sonrasında ise geriye sadece yığılmalı dağılım fonksiyonlarını eşit yüzdelikli eşitleme yönteminde yerine koymak kalacaktır (Wang ve diğerleri, 2008).

Zincir eşitleme (Chained method). Yöntem, X puanlarının Y'deki eşdeğerinin elde edilmesinde iki aşamalı dönüşüm kullanır (von Davier ve diğerleri, 2004a). Önce X formundaki puanlar 1. Evrenden elde edilen ortak test puanlarına dönüştürülür, sonrasında ise ortak test puanları 2. Evrendeki bireylerin oluşturduğu Y puanlarına dönüştürülür. Bu zincirleme yoluyla X puanları Y puanlarına eşitlenmiş olur (Kolen ve Brennan, 2014). Bundan olacaktır ki Dorans, (1990) ve Livingston, Dorans, Wright, (1990) yönteme zincir eşitleme adını vermişlerdir.

Zincir eşitleme yönteminde örneklemden hedef evrene genelleştirileceği varsayılan ilişki hem doğrusal hem de doğrusal olmayan bir eşitleme ilişkisi olarak kurulabilir. Temelinde yatan mantık aynıdır hatta doğrusal zincir eşitleme yönteminin uygulamada daha kolay olduğu söylenebilir (Livingston, 2014). Şekil 7.'de X formundaki belirli bir puanın ortak teste bağlı Y formundaki eşdeğeri doğrusal zincir eşitleme yöntemi ile gösterilmiştir.



Şekil 7. Doğrusal zincir eşitleme yöntemi

Kaynak. Livingston, 2014

Zincir eşitleme yönteminin temelinde yatan varsayımlar:

X, Y formlar, V ortak test, P ve Q evrenler olmak üzere;

Varsayım 1: X'ten V'ye yapılan eşitleme her iki evren için aynıdır,

Varsayım 2: V'den Y'ye yapılan eşitleme her iki evren için aynıdır.

Ortak testin her iki form arasında zincir görevi üstlenmesi ancak eşitleme ilişkisinin her iki evren için de aynı olmasındandır. Bu varsayımla temellendirilen zincir eşitleme yönteminin cebirsel tanımı,

$$chain_Y(x) = link_{Y_Q}[link_{V_P}(x)]$$

şeklindedir (Albano, 2016).

Eşitleme Hatası. Eşitleme sürecinde elde edilen eşdeğer puanlar tipik olarak kestirim hatası içerir. Eşitleme çalışmalarında aslolan en az eşitleme hatası içerecek düzeyde kestirim yapabilmektir. Rastgele ve sistematik olarak adlandırılan bu hatalar iki temel kaynaktan beslenir. Rastgele hatanın kaynağı eşitleme çalışmasının yürütüldüğü örneklemdir. Eşitleme ilişkisiyle elde edilen ortalama, standart sapma ve yüzdelik sıra gibi parametreler bir evrenden çekilen örneklemden elde edilir (Kolen ve Brennan, 2014; Zeng, 1991). Örneklemin evreni ne derece temsil ettiği burada belirleyici bir unsurdur ve azalan temsil gücü artan eşitleme hatası olarak yansımaktadır. Hal böyle olunca akla ilk gelen örneklem büyüklüğünü artırarak temsil gücünü artırmak olmalıdır. Böylelikle örneklem büyüdükçe rastgele hatanın azaldığı gözlemlenecektir. Eşitleme ilişkisinde diğer hata türü olan sistematik hata, eşitleme yöntemindeki varsayımların ve koşulların ihlalden kaynaklanır. Örneğin NEAT desende ortak test içerik ve istatistiksel özellikler bakımından formlardan farklıysa, bir uygulamadan diğerine farklı şekilde işlev gösteriyorsa veya gruplar arasında bariz yetenek farklılığı varsa artan sistematik hatadan söz etmek yerinde olacaktır (Kolen ve Brennan, 1995).

Bu araştırmada eşitleme hatasının tespitinde *ağırlıklandırılmış hata kareleri ortalaması-AHKO (weighted mean square error- WMSE)* kullanılmıştır. AHKO'nun hesaplanmasına yönelik bağıntı aşağıda verildiği gibidir.

$$AHKO (WMSE) = \sum_i^{k-1} f_i (X_E - X_{crit})^2 / \sum_i^k f_i S_Y^2$$

Burada, *k*: Y testindeki madde sayısı

S_Y^2 : Y formundaki ham puanların varyansı

X_{crit} : Y formundaki i . ham puan

X_E : Tek bir eşitleme yöntemiyle elde edilmiş X formundaki i . ham puana eşdeğer puan

f_i : Y formundaki i . ham puanın frekansdır.

AHKO eşitleme yöntemlerinin değerlendirilmesi bir kriter görevi görür ve uygulanan yöntemler arasında en az eşitleme hatasına sahip yöntemin kabul edilebilir olduğunu belirtir (Skaggs ve Lissitz, 1986).

İlgili Araştırmalar

Araştırmanın bu kısmında DMF içeren maddeler barındıran testlerin eşitlenmesinde DMF'nin test eşitleme üzerine ülkemizde ve diğer ülkelerde yapılmış olan araştırmalara dair bilgiler verilmiştir.

Chu, (2002) çalışmasında hiyerarşik Rasch modelin test eşitleme ve madde parametre kestirimindeki performansını DMF gösteren maddelerin varlığında araştırmıştır. Araştırmada DMF'li maddelerin varlığında ve yokluğunda tek grup eş zamanlı eşitleme deseniyle yöntemin performansı kıyaslanmıştır. Çalışmada hiyerarşik Rasch modelin performansı 3 DMF koşulu üzerine konumlandırılarak simülatif veriler üzerinde değerlendirilmiştir. Koşullar; madde sayısı (2 veya 5), iki alt grup arasında madde güçlük farkları (0,2, 0,6 ve 0,8) ve son olarak DMF gösteren maddelerin türü (ortak madde, ortak olmayan madde veya her ikisi de) olarak belirlenmiştir. Araştırma sonucunda, testlerden DMF varlığını silmek veya yok saymak yerine DMF varlığından haberdar olarak eşitleme işlemini ve madde parametre kestirimini yapmanın daha iyi sonuçlar verdiği görülmüştür. Kısa testlerden DMF nedeniyle madde çıkarmanın DMF varlığını görmezden gelmeye göre eşitlemede ve parametre kestiriminde daha yüksek hata riski doğurduğu belirtilmiştir. DMF'li madde sayısı ve etki büyüklüklerinin madde parametre kestirimi ile negatif yönde ilişki gösterdiği gözlenmiş bunun yanı sıra anlamlılık testleriyle de hiyerarşik Rasch modelin DMF parametre kestiriminde DMF etki büyüklüğü dışındaki diğer koşullara karşı güçlü durduğunu doğrulamıştır.

Turhan, (2006) araştırmasında çok düzeyli iki parametrelili madde tepki modelinin madde güçlük ve ayırt edicilik parametrelerinin kestirim performansını,

dikey eşitleme üzerinde DMF varlığında incelenmiştir. Farklı sayıdaki DMF'li maddelerle oluşturulan koşulların kontrolü için Florida Geniş Kapsamlı Başarı Testi verileri beraberinde simülatif veriler kullanılmıştır. Dikey eşitleme çalışması üç farklı yeterlilik düzeyi üzerindeki verilerle yürütülmüş bunun yanı sıra ölçeğin geliştirilmesinde DMF ve yeterlik düzeyi değişkenlerinin varlığında çok düzeyli 2PLM'in performansı geleneksel Madde Tepki Kuramı 2 PLM ile karşılaştırılmıştır. Aynı zamanda küçük bir simülasyon çalışması ile TB, TBO ve her ikisinde olduğu DMF'li ortak maddelerin dikey eşitleme çalışması üzerindeki etkisi araştırılmıştır. Sonuç olarak DMF yokluğunda çok düzeyli 2 PLM ile geleneksel 2 PLM madde güçlük ve ayırt ediciliğinde aynı kestirimleri üretmişler hatta dikey ölçeklemede aynı puan ölçeğini oluşturmuşlardır. Öğrenci düzey değişkeni ve TBO DMF değişkeni beraberinde çok düzeyli 2 PLM için daha iyi bir dikey ölçekleme elde edilirken, TB DMF varlığında ise ölçeğin bozulmasına yol açmıştır. TB ve TBO DMF etkisinin dikey eşitleme üzerindeki etkisi için yapılan simülasyon çalışmasında elde edilenlere göre, TBO DMF'li maddelerin sayısı arttıkça ölçekteki bozulmalar artarken, TB DMF'li maddeler arttıkça ölçekteki bozulmalar azalma eğiliminde olmuştur. Bunun beraberinde ortak maddeler içinde sadece tek tip DMF içeren maddelerin iki tip DMF olması durumuna göre ölçekte daha çok bozulma etkisi meydana getirmiştir. Araştırmacı bu sonucun, simülasyonun çalışmasındaki sınırlı örneklem büyüklüğü kaynaklı rastgele hatadan olabileceği şeklinde yorumlamıştır.

Atalay Kabasakal, (2014) eşitleme çalışmasında testlerde DMF'li maddelerin olup olmama durumuna göre eşitleme hatasındaki değişimi çok düzeyli madde tepki modeli (ÇDMTM) ve geleneksel madde tepki modeline (MTM) göre ele almıştır. Eşitleme yöntemi olarak eş zamanlı kalibrasyon (MTM-EK), ayrı kalibrasyon (MTM-AK) ve Stoking-Lord (LR) kullanılmıştır. 3 yöntem ile 24 farklı koşulun simule edildiği çalışmada; DMF'nin bulunduğu test, DMF etki büyüklüğü, örneklem büyüklüğü ve test uzunluğu değişkenlerinin eşitleme hatasına etkisi ele alınmıştır. Araştırma sonucunda ÇDMTM, MTM'ye göre eşitleme sürecinde ve yanlılığın modelden çıkarılmasında iyi sonuçlar verse de değişen örneklem büyüklüğü ve test uzunluğu gibi koşullardan MTM daha olumlu etkilenmiştir. Ayrıca küçük örneklemelerde ve kısa testlerde DMF'li maddelerin testten çıkarılmaları eşitleme hatasını artırırken, aksi koşullarda ise eşitleme hatasının azaldığı

gözlemlenmiştir. Burada, çıkarılan maddelerin DMF etki büyüklüğü çok yüksek düzeyde olursa durumun tersine işleyebileceği ayrıca testten madde çıkarılmasının geçerlik üzerindeki olumsuz etkileri olacağı konusunda araştırmacının hatırlatmaları bulunmaktadır.

Huggins, (2014) araştırmasında ortak maddelerin DMF içermesinin test eşitlemenin grup değişmezliği özelliği üzerine etkisini incelemiştir. Eşitleme işlemi NEAT desende MTK 3PL modelde gerçek puana dayalı yöntemler üzerinden yapılmıştır. Yöntem olarak ortalama-ortalama, ortalama-standart sapma, Stocking-Lord ve Haebara tercih edilmiştir. Koşullar beş ana faktör üzerine şekillendirilmiştir. Faktörler: maddelerin DMF etki büyüklüğü, DMF'li madde sayısı, DMF'li maddelerin TB/TBO durumu, alt grupların ortalama yetenek düzeyleri arasında fark olması/olmaması, DMF'li maddelerin sadece bir ortak formda olması veya her ikisinde de olması. Çalışmanın sonucunda, ortak testlerdeki eşit sayıda DMF'li madde varlığı, eşitlemenin grup değişmezliği üzerinde bir etki göstermemiştir hatta sonuçlar ortak testin DMF'li madde içermediği durumdaki eşitleme sonuçlarına yakındır. Değişen koşullarda da sonuçlar birbiri ile tutarlı çıkmıştır. Ancak formlar arasında DMF kaynaklı farklılar olduğunda grup değişmezliği göstergeleri sonuçların problemliliğini işaret etmiştir.

Demirus, (2015) araştırmasında ortak maddelerinin tamamının tek biçimli DMF'li olma ve DMF'li olmama durumlarında eşitleme yöntemlerinin performanslarını kıyaslamayı amaçlamıştır. DMF değişkeni olarak cinsiyet, yöntem olarak ise Mantel-Haenszel(MH) ve lojistik regresyon(LR) kullanılmıştır. Eşitleme süreci eşdeğer gruplar ortak test deseni altında, MTK'ya dayalı ortalama-ortalama, ortalama-standart sapma, Haebara ve Stocking-Lord yöntemleri ile yürütülmüştür. Eşitleme yöntemlerinin performanslarının karşılaştırılmasında kriter olarak RMSD tercih edilmiştir. Araştırma verileri DMF ve eşitleme için ayrı planlanmıştır. DMF analizinde 2000-2012 yılları arasında yapılan OKS ve SBS 8. Sınıf verileri kullanılırken, eşitleme için ise bu sınavlardan alınan belirli maddelerle oluşturulmuş testlerin uygulanması sonucunda elde edilen öğrenci yanıtları kullanılmıştır. Araştırma sonucunda en küçük eşitleme hatasını; ortak maddeler erkekler lehine TB-DMF'li olduğunda ortalama-standart sapma, ortak maddeler DMF'siz olduğunda Haebara ve Stocking-Lord yöntemleri üretmiştir.

Yurtçu ve Güzeller, (2017) çalışmalarında DMF'li maddeler içeren testlerin farklı koşullar altında eşitleme hatalarını incelemeyi amaçlamışlardır. Test içindeki DMF'li madde sayısı ve dağılımının eşitleme hatası üzerine etkisini 6 farklı simulasyon koşulu altında incelemişlerdir. Eşitleme çalışmasını eşdeğer gruplar ortak madde test deseni altında MTK'ya dayalı ortalama-ortalama, ortalama-standart sapma, Haebara ve Stocking Lord yöntemleriyle yürütmüşlerdir. Çalışmada 2013-2014 TEOG matematik testi madde parametreleri kullanılarak farklı matematik test formları oluşturulmuştur. Koşullarda DMF'li maddeler ortak test ve ortak olmayan test maddeleri içine farklı sayıda ve dağılımda yerleştirilmiştir. Eşitleme hataları hem koşullar arasında hem de formların DMF'li madde içermeyen durumu ile kıyaslanmıştır. Sonuç olarak formların DMF'li madde içermemesi durumunda en az eşitleme hatasını ortalama-ortalama yöntemi üretmiştir. Formların DMF'li madde içermesi, içermemesi durumuna göre her yöntemde hataların artmasına sebep olmuştur. Ortak testte veya temel formda DMF'li maddeler arttıkça tüm yöntemlerde eşitleme hatasını arttığı gözlenmiştir. Ortak teste ve temel forma eşit sayıda DMF'li madde yerleştirilerek yapılan karşılaştırmada, temel formda olan durumda daha fazla eşitleme hatası elde edilmiştir.

Gübeş ve Uyar, (2020) ortak testteki maddelerin DMF gösterip göstermeme durumlarında farklı eşitleme yöntemlerinin küçük örneklemeler üzerinde performanslarını karşılaştırmışlardır. Araştırma verisi TIMMS 2015 uygulaması Türkiye örnekleme 8. sınıf 1 ve 14. kitapçıklar matematik yanıtlarından oluşmaktadır. Karma Rasch modeli üzerinden belirlenen alt gruplara göre DMF gösteren maddeler MH ve LR teknikleri ile tespit edilmiştir. Bu maddelerin varlığında ve yokluğunda elde edilen RMSE katsayıları karşılaştırılarak değerlendirme yapılmıştır. Eşitleme süreci, küçük örneklemeler üzerinden NEAT desen altında Tucker, Levine, Daire-yay ve zincir eşitleme yöntemleri ile yürütülmüştür. Sonuç olarak her iki eşitleme durumunda da en az eşitleme hatası üreten yöntem ön düzgülendirme yapılan eşit yüzdelikli zincir eşitleme yöntemi olmuştur. Çalışmanın ana teması olarak belirlenen ortak testte DMF'li maddelerin varlığında ve yokluğunda yapılan eşitleme karşılaştırılmasında, değişkenin eşitleme hataları üzerinde dikkate değer bir etkisi gözlenmemiştir.

Ulusal ve uluslararası literatürde DMF'nin test eşitleme üzerine etkisinin incelendiği çalışmanın çok az olduğu görülmektedir. Ayrıca yapılan çalışmaların genelinde veriler simülasyona dayalı üretilmiş aynı zamanda eşitleme yöntemleri de MTK' ya göre temellendirilmiştir. Bu çalışmanın, KTK'ya göre şekillendirilmiş olması ve bizzat gerçek uygulamadan alınmış verilerle değişken etkisinin inceleniyor olması sebebiyle alanyazına katkı sunacağı düşünülmektedir.

Bölüm 3

Yöntem

Bu bölümde; araştırmanın türüne, evren ve örnekleme, veri toplama sürecine ve verilerin analizine yer verilmiştir.

Araştırmanın Türü

Bu araştırmada DMF gösteren ortak maddelerin testten çıkarılması durumunda, eşitlenmiş puanlarda ve eşitleme hatasındaki değişimin gözlenmesi amaçlanmıştır. Araştırma bu yönüyle tarama türündedir. Tarama araştırmalarında var olan bir durumun fotoğrafı çekilerek betimleme yapılırken aynı zamanda ölçülen değişkenler arasındaki ilişkiler de incelenebilir (Büyüköztürk, Kılıç Çakmak, Akgün, Karadeniz ve Demirel, 2019).

Araştırmanın Evreni ve Örnekleme

Araştırmanın evrenini PISA 2018 uygulamasını bilgisayar ortamında İngilizce ve İspanyolca dillerinde almış yaklaşık 186 bin öğrenci oluşturmaktadır (OECD, 2019a). Örnekleme ise 18 ve 24. kitapçıklar fen okuryazarlığı maddelerinde kayıp ve eksik verisi olmayan 5924 öğrenci meydana getirmektedir. 18. Kitapçık 3269, 24. Kitapçık ise 2655 öğrenci verisini içermektedir. DMF kaynağı olarak düşünülen dil ve cinsiyet değişkenlerinin kitapçıklardaki sayıları ve yüzdeleri Tablo 7’de verilmiştir.

Tablo 7

Bireylerin Değişkenlere Göre Dağılımı

Kitapçık	Örnekleme		Dil				Cinsiyet			
			İngilizce		İspanyolca		Kadın		Erkek	
	N	%	f	%	f	%	f	%	f	%
18	3269	55	2223	37	1046	18	1568	26	1701	29
24	2655	45	1993	34	662	11	1260	21	1395	24
Toplam	5924	100	4216	71	1708	29	2828	47	3069	53

Tablo 7 incelendiğinde deęişkenleri oluřturan alt grupların kitapçıklardaki sayısı ve daęılımının DMF belirleme yöntemleri için gerekli en az 200 örneklem büyüklüğünü sağladığı görülmüřtür (Zieky, 1993).

Veri Toplama Süreci

Arařtırma kapsamında kullanılan PISA 2018 biliřsel maddelere ait uygulama verileri OECD'nin resmi web adresinden (<http://www.oecd.org/pisa/data/2018database/>) temin edilmiřtir.

Veri Toplama Aracı

Arařtırmanın verilerinin elde edildiđi PISA 2018 uygulamasında 18 ve 24. Kitapçıklarındaki fen okuryazarlığı maddeleri ikiřer madde kümesinden meydana gelmektedir. 18. Kitapçık S1 ve S6, 24. Kitapçık S2 ve S6 madde kümelerinden oluřmaktadır (OECD, 2018a). S6 madde kümesi her iki kitapçık için ortaktır. Bu arařtırmada iki kategorili puanlanan maddelere yönelik DMF belirleme teknikleri tercih edildiđi için S1 kümesinde bulunan 3 madde, S2 kümesinde bulunan 1 madde ve S6 kümesinde bulunan 1 madde kapsam dıřında tutulmuřtur. Son haliyle arařtırmaya dâhil olan madde sayıları ve madde kümelerinin kitapçıklara yerleřimi Tablo 8'de belirtilmiřtir.

Tablo 8

Maddelerin Kitapçıklara Yerleřimi

Kitapçıklar	Madde Kümeleri			Toplam (K)
	S1 (K)	S6 (K)	S2 (K)	
18	17	18		35
24		18	17	35

Not: K:Madde sayısı

Tablo 8'deki bilgilere göre kitapçıklarda iki kategorili olarak puanlanmıř 18'i ortak toplam otuz beřer madde bulunmaktadır. Kitapçıklar arasındaki eřitleme çalıřması ortak madde kümesinde bulunan 18 madde aracılıđı ile yürütülmüřtür.

Verilerin Analizi

Verilerin analizi 4 aşamada gerçekleştirilmiştir. İlk aşamada DMF analizi ve test eşitleme yapmak için ortak koşul testin tek boyutluluk özelliğini sağlayıp sağlamadığının tespitinde Tetrakorik korelasyona dayalı açımlayıcı faktör analizi yapılmış, analiz için FACTOR (Versiyon 10.10.03) programı kullanılmıştır. İkinci aşamada kitapçıkların betimsel istatistikleri hesaplanmış ve test eşitleme koşullarını sağlayıp sağlamadığı yoklanmıştır. Betimsel istatistiklerde; ortalama, mod, medyan, standart sapma, ortalama güçlük, KR-20 güvenilirlik katsayısı, çarpıklığı ve basıklığı SPSS (Versiyon 23) programı ile hesaplanmıştır. Test eşitleme koşulları olarak; kitapçıkların aritmetik ortalamalarının eşitliği t testi, varyanslarının karşılaştırılması Levene's testi, eşit güvenilirlik koşulu Fischer-z istatistiği, ortalama güçlüklerinin kıyaslanması z istatistiği ile gerçekleştirilmiştir. Üçüncü aşamada ortak maddelerin dil ve cinsiyet değişkenine göre madde istatistikleri hesaplanmıştır. Madde istatistikleri Test Analiz Programı (TAP Versiyon 19.1.4.) ile gerçekleştirilmiştir. Son aşamada her iki kitapçıktaki ortak olan maddeler dil ve cinsiyet değişkenine göre DMF analizine tabi tutulmuştur. DMF analizinde Mantel-Haenszel ve SIBTEST teknikleri kullanılmış, analizler R programı "difR" paketi ile yürütülmüştür. Sonrasında kitapçıklar eşdeğer olmayan gruplar ortak test deseninde gözlenen puana dayalı doğrusal ve doğrusal olmayan yöntemlerle eşitlenmiştir. Eşitleme doğrusal yöntemlerde; Tucker, Levine, zincir eşitleme, doğrusal olmayan yöntemlerde; frekans kestirim, Daire-Yay Levine ve zincir eşit yüzdelikli eşitleme yöntemleri kullanılmıştır. Eşitleme çalışmaları R programı "equate" paketi ile yürütülmüştür. Eşitleme kriteri olarak AHKO (WMSE) kullanılmış, AHKO Microsoft Office Excel programı ile hesaplanmıştır. DMF gösterdiği tespit edilen maddeler ortak testten çıkarıldığında, öğrencilerin ortak test puanları ile toplam test puanları arasındaki ilişkinin değişimi Pearson momentler çarpımı korelasyon katsayısı ile izlenmiştir. Korelasyon katsayısı için SPSS (Versiyon 23) programı kullanılmıştır. Her alt problem durumuna göre eşitleme tekrar edilmiş, eşitleme hatalarındaki farklılıklar ilk durumla karşılaştırılmıştır.

Veri analiz sürecinin ilk aşamasında hem DMF analizi hem de test eşitleme için ön koşul olan testlerin tek boyutluluğunun analizi yapılmıştır. Woldbeck (1998) eşitlenecek testlerin tek boyutlu olmasının önemli bir koşul olduğu yanı sıra her iki testin aynı ve tek bir yapıyı ölçtüğünden de emin olunması gerektiğini belirtmiştir.

Kitapçıklardaki fen okuryazarlığı testlerinin faktör yapısını ortaya koymak için açımlayıcı faktör analizi yapılmıştır. Veriler 0 ve 1 şeklinde kodlanmış iki kategorili yapıda olduğundan dolayı faktör analizi tetrakorik korelasyon matrisine dayalı olarak yürütülmüştür (Kelecioğlu ve Gübeş, 2013). Faktör analizine geçmeden önce iki koşul: veri yapısının örneklem büyüklüğü bakımından uygunluğu ve verilerin çok değişkenli normal dağılım özelliği test edilmiştir. Örneklem yeterliğine yönelik ölçütlerden biri Kaiser-Meyer-Olkin (KMO) testidir. KMO değerinin büyük olması değişkenlerin her birinin diğer değişkenler tarafından iyi bir şekilde tahmin edilebileceği anlamını taşır. Verilerin çok değişkenli normal dağılımdan gelip gelmediği ise Bartlett Küresellik Testi ile yorumlanır. Bartlett küresellik testi bir Ki-kare (χ^2) analizidir. Ki-kare değerinin 0,05'ten küçük bir anlamlılık değeri, verilerin çok değişkenli normallik sayılığını sağladığı şeklinde yorumlanır (Çokluk, Şekercioğlu ve Büyüköztürk, 2016). Kitapçıklardan elde edilen KMO ve Bartlett küresellik testi sonuçları Tablo 9 verilmiştir.

Tablo 9

PISA 2018 18 ve 24. Kitapçıkların KMO ve Bartlett Değerleri

Kitapçıklar	Kaiser-Meyer-Olkin (KMO)	Bartlett Küresellik Testi		
		χ^2	df	p
18	0,96	37480,3	595	0,00*
24	0,96	30411,4	595	0,00*

Not: * $p < 0,05$

Tablo 9 incelendiğinde tüm kitapçıklarda KMO değerlerinin 0,90'dan büyük olduğu görülmüştür. Kaiser (1974) açımlayıcı faktör analizinde 0,90 üzerindeki KMO değerini, veri yapısının mükemmel yeterlik düzeyinde olduğu şeklinde yorumlamıştır. Bartlett küresellik testi sonuçları incelendiğinde kitapçıklarda elde edilen χ^2 değerinin 0,05 düzeyinde istatistiksel olarak anlamlı olması verilerin çok değişkenli normal dağılımdan geldiğini doğrulamıştır. İki sayılıtlının karşılanması verilerin açımlayıcı faktör analizi için uygunluğunu doğrular niteliktedir.

Kitapçıkların faktör analizine uygunluğu sınıandıktan sonra faktör sayısı 3 ile sınırlandırılarak açımlayıcı faktör analizi yapılmıştır. Analiz sonucunda

kitapçıklardaki ilk üç faktöre yönelik özdeğer ve varyans açıklama yüzdeleri Tablo 10'da verilmiştir.

Tablo 10

PISA 2018 18 ve 24. Kitapçıklar Fen Okuryazarlığı Testlerine Ait Faktör Analizi Sonuçları

Faktör	Kitapçıklar			
	18		24	
	Özdeğer	Varyans Yüzdesi	Özdeğer	Varyans Yüzdesi
1	11,27	32,2	12,05	34,42
2	1,23	3,52	1,30	3,70
3	1,11	3,17	1,19	3,40

Tablo 10'daki verilere göre toplam varyansın 18. Kitapçıkta %32,2'si 24. Kitapçıkta ise %34,42'si birinci faktör tarafından açıklanmaktadır. Büyüköztürk (2017) tek faktörlü ölçeklerde açıklanan varyansın %30 olmasını kabul edilebilir bir sınır olarak belirtmiş ayrıca birinci faktöre ait özdeğerin ikinci faktör özdeğerinin 3 katından fazla olmasını da ölçeğin tek boyutluluğuna yönelik karar almak için önemli bir kriter olarak eklemiştir. Ayrıca analizin yapıldığı FACTOR programı yapıların tek faktörlü olarak değerlendirilmesini önermiştir. Kitapçıkların tek boyutlu yapısı için uyum iyiliği indekslerini ($GFI_{18} = 0,99$; $GFI_{24} = 0,99$) mükemmel uyum (akt. Çokluk ve diğerleri, 2016) düzeyinde göstermiştir. PISA uygulayıcılarının test maddelerini ilgililerle paylaşmıyor olmasından dolayı kitapçıkların aynı yapıyı ölçüp ölçmediği ile ilgili uzman kanısına başvurulamamıştır. Konuya çözüm olacağı düşünülerek her iki kitapçıkta ortak test ile toplam test puanları arasındaki Pearson momentler çarpımı korelasyon katsayısına bakılmıştır. Ortak ve toplam test puanları arası korelasyon katsayısının her iki kitapçıkta da pozitif yönde çok yüksek düzeyde (Tan, 2016, s.145) istatistiksel olarak anlamlı olduğu görülmüştür ($r_{18} = 0,93$ $p < 0,01$; $r_{24} = 0,94$ $p < 0,01$). Tüm bu sonuçlar ışığında 18 ve 24. kitapçıkların aynı ve tek bir yapıyı ölçtüğüne karar verilerek araştırmaya devam edilmiştir. DMF analizi için gerekli koşul sağlandıktan sonra analizlerin ikinci aşamasına geçilmiştir.

Analizlerin ikinci aşamasında ilk olarak 18 ve 24. kitapçıkları fen okuryazarlığı testlerinin betimsel istatistikleri hesaplanmış ulaşılan sonuçlara Tablo 11’de yer verilmiştir.

Tablo 11

Kitapçıklara Ait Betimsel İstatistikler

İstatistikler	Kitapçıklar	
	18	24
N	3269	2655
K	35	35
Minimum Puan	2	0
Maksimum Puan	35	35
Ortalama	19,11	20,07
Mod	24	23
Medyan	19	21
Standart Sapma	7,12	7,11
Ort. Güçlük	0,55	0,57
Ort. Ayırtedicilik (r_{pb})	0,44	0,48
KR-20	0,878	0,883
Çarpıklık K.	- 0,15	- 0,30
Basıklık K.	- 0,92	- 0,76

Not: N:Örneklem büyüklüğü K:Madde sayısı r_{pb} :Nokta çift serili korelasyon katsayısı

Tablo 11’deki veriler incelendiğinde kitapçık ortalamalarının ve standart sapmalarının birbirine yakın olduğu görülmüştür. Her iki kitapçık için çarpıklık ve basıklık katsayılarının ± 1 aralığında kalması dağılımın normalden aşırı sapmadığı şeklinde yorumlanabilir (Çokluk ve diğerleri , 2016).

İkinci aşamanın devamında eşitleme çalışmasını yürütebilmek için gerekli olan koşulların sağlanıp sağlanmadığına bakılmıştır. Alanyazında Klasik Test Kuramı temelli bir yatay eşitleme çalışmasında; testlerin tek ve aynı yapıyı ölçmesi, eşit güvenilirliğe ve benzer düzeyde ortalama güçlüğüne sahip olmasının gerekliliği belirtilmiştir (Crocker ve Algina, 1986; Dorans ve Holland, 2000; Lord,

1950). Tek boyutluluk DMF analizi için de bir önkoşul olduğundan dolayı tek boyutluluk tespiti analizin birinci aşamasında yapılmıştır.

18 ve 24. Kitapçıkların KR-20 güvenilirlik katsayıları birer ilişkisiz örneklem korelasyon katsayısı olarak kabul edilmiş (Akhun, 1984) ve katsayılara Fischer'in Z_r dönüşümü uygulanarak aradaki farkın istatistiksel anlamlılığı test edilmiştir (Cohen ve Cohen, 1983). Elde edilen bulgular Tablo 12'de paylaşılmıştır.

Tablo 12

Kitapçıkların Güvenirliklerinin Karşılaştırılmasına İlişkin Bulgular

Kitapçıklar	KR-20	Z_r	z	p
18	0,878	1,37	-0,85	0,39
24	0,883	1,39		

Tablo 12'deki bulgular incelendiğinde 18 ve 24. kitapçıkların güvenilirlik katsayıları arasında 0,05 düzeyinde istatistiksel olarak anlamlı bir fark olmadığı görülmüştür ($z=-0,85$; $p>0,05$). Bu bulguyla, kitapçıkların eşit güvenirlikte olması koşulunun sağlandığı söylenebilir.

18 ve 24. kitapçıkların ortalama güçlükleri arasındaki farkın istatistiksel anlamlılığı iki oran farkı testi ile incelenmiştir (Baykul, 1999; Can, 2019). Ortalama güçlüklerin karşılaştırılmasına ilişkin sonuçlar Tablo 13'te sunulmuştur.

Tablo 13

Kitapçıkların Ortalama Güçlüklerinin Karşılaştırılması İlişkin Bulgular

Kitapçıklar	\bar{p}	K	z	p
18	0,55	35	1,5	0,12
24	0,57	35		

Not: \bar{p} : Ortalama güçlük K: Madde sayısı

Tablo 13'teki z testi sonuçları incelendiğinde kitapçıkların ortalama güçlükleri arasında 0,05 düzeyinde istatistiksel olarak anlamlı bir farklılık görülmemiştir. ($z= 1,5$; $p>0,05$). Bu sonuçla kitapçıkların eşit güçlükte olduğu yargısına varılabilir.

Analizin ikinci aşamasında son olarak kitapçıktan elde edilen puanların varyanslarının eşitliği Levene testi, aritmetik ortalamaları arasındaki farkın istatistiksel anlamlılığı ise bağımsız gruplar *t* testi ile incelenmiştir. İnceleme sonuçları Tablo 14’de paylaşılmıştır.

Tablo 14

Kitapçıkların Varyans ve Ortalamalarının Karşılaştırılmasına İlişkin Bulgular

Kitapçıklar	\bar{X}	S_x^2	Levene Testi		t testi		
			F	p	t	p	d
18	19,10	50,74	0,61	0,44	-5,19	0,00*	0,14
24	20,07	50,51					

Not: * $p < 0,05$

Tablo 14’deki bulgular incelendiğinde Levene testi sonucunda kitapçıklardan elde edilen puanların varyansları arasında 0,05 düzeyinde istatistiksel olarak anlamlı bir fark olmadığı görülmüştür ($F=0,61$; $p>0,05$). Ortalamalar arasındaki 0,87 puanlık fark 0,05 düzeyinde istatistiksel olarak anlamlı bulunmuştur ($t=-5,19$; $p<0,05$). Ancak *t* testi ortalamalar arasındaki farkın istatistiksel olarak anlamlılığını ortaya koymakta, farkın büyüklüğü hakkında bilgi vermemektedir. Bu yüzden istatistiksel olarak anlamlılık etki büyüklüğü ile beraber değerlendirilmelidir (Can, 2019). Cohen’in *d* değeri karşılaştırılan ortalamalar arasında kaç standart sapma fark olduğu hakkında bilgi verir (Büyüköztürk, 2017). Etki büyüklüğünün 0 olması ortalamalar arasında fark olmadığı şeklinde yorumlanırken, işaretlerine bakılmaksızın 0,2, 0,5, 0,8 etki büyüklükleri sırasıyla küçük, orta, büyük etki büyüklüğü olarak yorumlanır (Green ve Salkind, 2009). İlişkisiz örneklem t testi sonucunda kitapçıkların ortalamaları arasındaki fark anlamlı bulunmuş olsa da etki büyüklüğü küçük etki büyüklüğünün altında bir değer almıştır ($d=0,14$). Yapılan analizler neticesinde kitapçıkların eşitleme koşullarını sağladığı sonucuna varılmıştır.

Analizlerin 3. aşamasında 18 ve 24. kitapçıklarda ortak olarak kullanılan 18 fen okuryazarlığı maddesinin DMF değişkenlerine göre güçlük ve ayırt edicilik değerleri incelenmiş ve DMF kaynaklarına ait alt gruplar arasında yetenek

farklılıkları incelenmiştir. Madde analizlerine yönelik bulgular Tablo 15'te paylaşılmıştır.

Tablo 15

Ortak Test Maddelerinin DMF Değişkenlerine Göre Analizi

Maddeler	Değişkenler							
	Dil				Cinsiyet			
	İngilizce		İspanyolca		Kadın		Erkek	
	p	r _{pb}	p	r _{pb}	p	r _{pb}	p	r _{pb}
CS478Q01S	0,46	0,38	0,36	0,42	0,42	0,4	0,44	0,41
CS478Q02S	0,59	0,53	0,47	0,55	0,55	0,52	0,56	0,56
CS478Q03S	0,63	0,44	0,57	0,44	0,62	0,44	0,6	0,45
CS415Q07S	0,78	0,42	0,73	0,3	0,77	0,38	0,76	0,39
CS415Q02S	0,86	0,45	0,85	0,4	0,86	0,41	0,86	0,45
CS415Q08S	0,55	0,44	0,57	0,48	0,57	0,45	0,54	0,43
CS627Q01S	0,44	0,41	0,31	0,37	0,35	0,41	0,44	0,41
CS627Q03S	0,81	0,43	0,75	0,39	0,79	0,44	0,79	0,4
CS627Q04S	0,76	0,51	0,64	0,51	0,74	0,49	0,71	0,55
CS607Q01S	0,93	0,38	0,86	0,4	0,9	0,39	0,93	0,39
CS607Q02S	0,7	0,57	0,48	0,6	0,6	0,6	0,67	0,59
CS638Q01S	0,62	0,58	0,5	0,56	0,54	0,57	0,62	0,58
CS638Q02S	0,86	0,5	0,8	0,44	0,83	0,48	0,85	0,48
CS638Q04S	0,58	0,55	0,32	0,43	0,53	0,56	0,48	0,52
CS615Q07S	0,29	0,45	0,21	0,43	0,26	0,45	0,27	0,45
CS615Q01S	0,85	0,41	0,81	0,39	0,81	0,4	0,87	0,41
CS615Q02S	0,51	0,55	0,37	0,58	0,49	0,6	0,46	0,55
CS615Q05S	0,19	0,32	0,15	0,17	0,18	0,25	0,18	0,31
Ortalama	0,63	0,46	0,54	0,44	0,60	0,46	0,61	0,46

Not: p: Madde güçlüğü r_{pb}: Madde ayırt ediciliği

PISA 2018, 18 ve 24. kitapçıklarda ortak olan fen okuryazarlığı maddeleri dil değişkenine göre analiz edildiğinde; İngilizce dilinde yazılmış kitapçıklarda ortalama madde güçlüğü 0,63 iken İspanyolca dilinde yazılmış olanlarda 0,54 olarak hesaplanmıştır. Bilişsel testleri İngilizce dilinde alanların, İspanyolca dilinde alanlara göre maddeleri daha kolay bulduğu söylenebilir. Ayrıca maddelerin ortalama ayırt ediciliği incelendiğinde ise neredeyse her iki grup içinde benzer ayırt

edicilikte olduđu şekilde yorumlanabilir. Maddeler cinsiyet deęiřkenine gre kıyaslandıęında, hem ortalama madde glğnn hem de ortalama ayırt edicilięin kadın ve erkekler iin benzer istatistiki sonular verdięi grlmřtr. Madde zelinde incelendięinde sadece CS615Q05S kodlu maddenin tm gruplar iin ok zor olarak algılandıęı beraberinde ayırt edicilik zellięinin testten ıkarmaya varacak lde dřk ($r_{pb} \leq 0,19$) olduđu gzlemlenmiřtir. Harici maddelerin madde glk ve ayırt edicilik istatistikleri istendik dzeydedir (Crocker ve Algina, 1986).

Verilerin analizinin ilk  ařamasında; kitapıkların betimsel istatistiklerine deęinilmiř, verilerin DMF ve test eřitleme analizleri iin uygunluęu kontrol edilmiř, ortak test maddelerinin DMF deęiřkenlerine gre madde istatistikleri incelenmiřtir. Veri analizinin son ařamasına ynelik bulgular alt problem durumları altında raporlanmıřtır.

Bölüm 4

Bulgular ve Yorumlar

Araştırmanın bu bölümünde araştırmanın amacı doğrultusunda elde edilen bulgulara ve yorumlara yer verilmiştir.

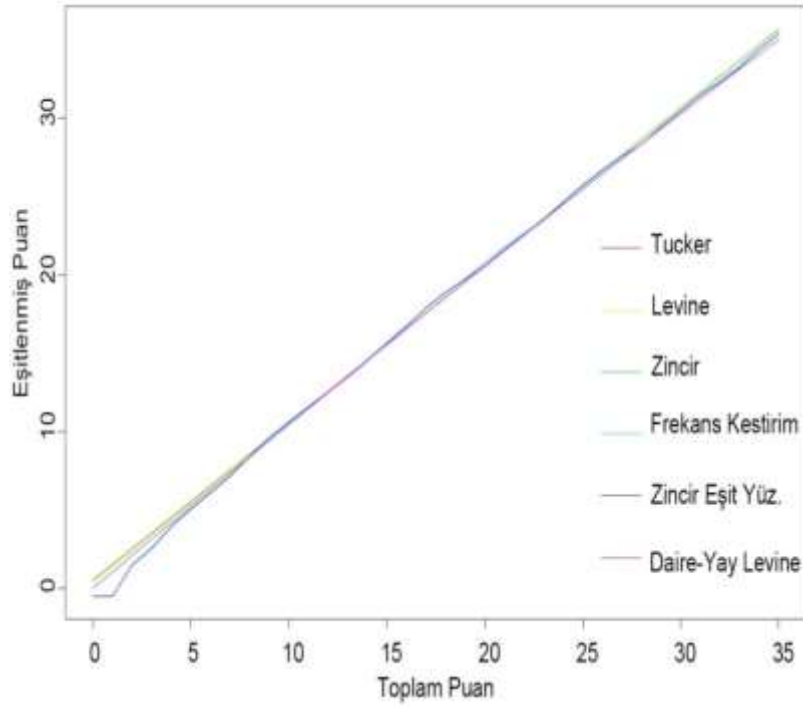
Alt Problem 1'e İlişkin Bulgular

PISA 2018 Fen okuryazarlığına ait öğrenci ham puanlarının ortak maddelere dayalı olarak eşitlenmesinde, doğrusal ve doğrusal olmayan eşitleme yöntemiyle elde edilen eşitlenmiş puanlar ve eşitleme hataları nasıldır?

PISA 2018 18 ve 24. Kitapçıklar fen okuryazarlığı ham puanlarının eşitlenmesi NEAT deseni KTK temelli gözlenen puan eşitleme yöntemleriyle yürütülmüştür. Doğrusal yöntemlerde; Tucker, Levine, zincir eşitleme yöntemi tercih edilirken doğrusal olmayan yöntemlerde frekans kestirim, zincir Eşit yüzdelikli ve Daire-yay Levine yöntemleri tercih edilmiştir. 18. Kitapçığın 24. kitapçığa eşitlenmesinde yöntemlere göre elde edilen eşitlenmiş puanlar EK-A'da verilmiştir.

EK-A'daki verilere göre doğrusal yöntemlerden elde edilen eşitlenmiş puanlar 18. kitapçık puan ölçeğindeki ham puanlardan her puan düzeyinde yüksek çıkmıştır. Ham puanlar ile eşitlenmiş puanlar arasındaki farklar 0,45 ile 0,67 arasında değişmiş ve ölçek boyunca farklar en düşük puandan en yüksek puana doğru artış göstermiştir. Doğrusal olmayan yöntemlerde ise Daire-yay Levine yönteminin altında yatan mantığı gereği eşitlenmiş puanlar uç noktalarda olası maksimum ve minimum ham puanlara eşitlenmiştir (Livingston ve Kim, 2009). Frekans kestirim ve zincir eşit yüzdelikli eşitlemede eşitlenmiş puanlar 0'dan 3'e kadar olan ham puanlardan düşük, sonraki puanlarda daha yüksek değerler almıştır. Ham puan ile eşitlenmiş puanlar arası farklılıklar -0,50 ile 0,90 arasında dağılmıştır. Puan farklılıkları uç noktalardan merkeze yaklaştıkça artış göstermiştir.

Yöntemler arasında grafiksel bir kıyaslama için her bir yöntem özelinde eşitlenmiş puanların ve ham puanlarla ilişkisi Şekil 8.'de verilmiştir.



Şekil 8. Ham puanlar ile eşitlenmiş puanlara ait dağılım grafikleri

Şekil 8 incelendiğinde ölçeğin ilk ve son çeyreğinde grafikler arasında belirgin ayrımlar görülürken, 10-15 ve 20-25 toplam puan aralıklarında tüm yöntemlerden elde edilen grafikler neredeyse aynı iz üzerinde yer almıştır. Zincir eşit yüzdelikli yöntemle ait grafik 0-5 puan aralığında diğer grafiklerden bariz ayrım göstermiş, ölçek boyunca da diğer grafikler ile çakışma eğiliminde olmamıştır.

Ortak test ve toplam test puanları arasındaki korelasyon eşitlik etkinliğinin bir ölçüsü olarak kabul edilip kararlı ve hassas eşdeğer puanlar için gerekli kabul edilmektedir (Budescu, 1985). Ortak ve toplam test puanları arasında yapılan Pearson momentler çarpımı korelasyonu sonucunda istatistiksel olarak anlamlı pozitif yönde çok yüksek düzeyde ilişkiye rastlanmıştır ($r = 0,93$ $p < 0,01$). Lord'a (1955) göre ortak test ve toplam test puanları arasındaki korelasyon katsayısı büyüdükçe eşitleme hataları azalma eğiliminde olacaktır. Araştırmada eşitleme hatalarının tespitinde ağırlıklandırılmış hata kareleri ortalaması (AHKO)'dan yararlanılmıştır. Eşitleme yöntemlerinin performansı AHKO değerleri üzerinden değerlendirilmiş elde edilen bulgular Tablo 16'da paylaşılmıştır.

Tablo 16

Doğrusal ve Doğrusal Olmayan Eşitleme Yöntemlerine Ait AHKO Değerleri

Değerlendirme Kriteri	Doğrusal Eşitleme Yöntemleri			Doğrusal Olmayan Eşitleme Yöntemleri		
	Tucker	Levine	Zincir	Frekans Kestirim	Zincir Eşit Yüzdellikli	Daire-Yay Levine
AHKO	0,0074	0,0063	0,0068	0,0075	0,0070	0,0044

Skaggs ve Lissitz (1986) AHKO istatistiği için standart hataların ulaşılabilir olmadığını bu yüzden AHKO değerleri arasındaki farkın istatistiksel anlamlılığı konusunda bir değerlendirme yapmanın zor olduğunu ancak değerler arasındaki 0,05'lik farkın pratik olarak anlamlı kabul edilebileceğini belirtmiştir. Tablo 16'daki bulgular değerlendirildiğinde AHKO değerlerinin 0,0074 ile 0,0044 arasında değiştiği ve hiçbir değer arasında 0,05'lik bir fark olmadığı gözlemlenmiştir. AHKO formül üzerinden incelenirse AHKO değerini oluşturan temel unsurun eşitlenmiş ve ham puanlar arasındaki fark olduğu görülür. Hata olarak tanımlanan bu farklar ne kadar küçük olursa yöntemden elde edilen ağırlıklandırılmış hata kareleri ortalaması da düşük olacaktır. Bu yüzden en düşük AHKO değerine sahip eşitleme yöntemi kullanılanlar içinde en uygun yöntem olarak düşünülebilir. Tablo 16'daki bulgular incelendiğinde en çok eşitleme hatası görülen yöntem frekans kestirim iken en az eşitleme hatası Daire-Yay Levine yönteminde görülmüştür. PISA 2018 18 ve 24. kitapçıklar Fen okuryazarlığı ham puanlarının eşitlenmesinde kullanılan yöntemler arasında en uygun yöntemin Daire-Yay Levine olduğu, Levine ve zincir doğrusal eşitleme yöntemlerinin izlediği söylenebilir.

Alt Problem 2'ye İlişkin Bulgular

Dil kaynaklı DMF gösterdiği Mantel-Haenszel ve SIBTEST teknikleriyle tespit edilen maddelerin ortak testten çıkarılmaları, doğrusal ve doğrusal olmayan eşitleme yöntemiyle elde edilmiş puanları ve eşitleme hatasını nasıl etkilemiştir?

Bu alt problemin çözümüne yönelik olarak PISA 2018 uygulaması 18 ve 24. kitapçıklardaki fen okuryazarlığı ortak testinde bulunan 18 maddeye dil değişkenine göre DMF analizi yapılmıştır. Mantel-Haenszel ve SIBTEST tekniği kullanılarak İngilizce ve İspanyolca dillerinden her hangi biri için DMF gösterdiği

tespit edilen maddeler ortak testten çıkarılarak eşitleme işlemi kalan ortak maddeler üzerinden tekrar yürütülmüştür. Elde edilen eşitlenmiş puanlar ve AHKO değerleri ortak maddelerin çıkarılmadığı durumdaki eşitleme sonuçları ile kıyaslanmıştır. Ayrıca tekniklere göre DMF'li bulunan maddelerin çıkarılması sonucunda, ortak test ile toplam test puanları arasındaki korelasyon katsayısındaki değişimler incelenirken dil değişkenini oluşturan alt grupların yetenek farklılıklarındaki değişimler de gözlemlenmiştir. Dil değişkenine göre yapılan DMF analizlerinde, testi İspanyolca alanlar odak grup, İngilizce alanlar ise referans grup olarak tanımlanmıştır. DMF analizi ilk olarak MH tekniğine göre yapılmış elde edilen bulgular Tablo 17'de paylaşılmıştır.

Tablo 17

Ortak Test Maddelerinin Dil Değişkenine Göre MH Analiz Bulguları

Madde Kodu	ΔMH	p	DMF Düzeyi	Avantaj Sağladığı Grup
CS478Q01S	-0,22	0,18		
CS478Q02S	-0,15	0,38		
CS478Q03S	0,34	0,04*	A	
CS415Q07S	0,23	0,20		
CS415Q02S	0,69	0,00*	A	
CS415Q08S	1,32	0,00*	B	Odak (İspanyolca)
CS627Q01S	-0,53	0,00*	A	
CS627Q03S	0,17	0,38		
CS627Q04S	-0,47	0,01*	A	
CS607Q01S	-1,10	0,00*	B	Referans (İngilizce)
CS607Q02S	-1,47	0,00*	B	Referans (İngilizce)
CS638Q01S	-0,09	0,64		
CS638Q02S	0,03	0,93		
CS638Q04S	-1,75	0,00*	C	Referans (İngilizce)
CS615Q07S	0,09	0,69		
CS615Q01S	0,24	0,27		
CS615Q02S	-0,34	0,06		
CS615Q05S	0,27	0,19		

Not: * $p < 0,05$ $|\Delta MH| < 1$ A düzeyi, $1 < |\Delta MH| < 1,5$ B düzeyi, $1,5 < |\Delta MH|$ C düzeyi; $\Delta MH < 0$ Referans grup lehine, $\Delta MH > 0$ Odak grup lehine

Ortak maddelerin MH analizi aşamasında, DMF'li maddelerin toplam puana olan katkısının önüne geçmek için Holland ve Thayer' in (1988) MH analizi için önermiş olduğu iki aşamalı arındırma yaklaşımı benimsenmiştir. 0,05 hata

düzeyinde istatistiksel olarak anlamlı olan maddeler belirlendikten sonra Zieky'nin (1993) DMF etki büyüklüğü sınıflamasından yararlanılmıştır. Maddelerin hangi grup lehine DMF sağladığının tespitinde ΔMH 'nin işaretine bakılmış, $\Delta MH < 0$ olduğunda maddenin referans grup lehine işlev gösterdiği şeklinde yorumlanmıştır (Ackerman ve Evans, 1992; Kim, 2003).

Tablo 17'deki verilere göre 3 maddede (CS415Q08S, CS607Q01S, CS607Q02S) B düzeyinde, 1 madde de ise (CS638Q04S) C düzeyinde DMF tespit edilmiştir. CS415Q08S maddesinin İspanyolca lehine, CS607Q01S, CS607Q02S, CS638Q04S maddelerinin ise İngilizce lehine DMF gösterdiği görülmüştür. MH tekniğine göre DMF gösteren maddelerin ortak testten çıkarılması sonucunda ortak madde sayısı 14, toplam madde sayısı 31 olmuştur. Sonraki aşamada dil değişkeni kaynaklı DMF analizi SIBTEST tekniğine göre yapılmış elde edilen bulgular Tablo 18'de sunulmuştur.

Tablo 18

Ortak Test Maddelerinin Dil Değişkenine Göre SIBTEST Analiz Bulguları

Madde Kodu	$\hat{\beta}$	p	DMF Düzeyi	Avantaj Sağladığı Grup
CS478Q01S	0,02	0,09		
CS478Q02S	0,00	0,84		
CS478Q03S	- 0,03	0,02*	A	
CS415Q07S	0,00	0,84		
CS415Q02S	- 0,04	0,00*	A	
CS415Q08S	- 0,12	0,00*	C	Odak (İspanyolca)
CS627Q01S	0,05	0,00*	A	
CS627Q03S	- 0,01	0,36		
CS627Q04S	0,03	0,02*	A	
CS607Q01S	0,03	0,00*	A	
CS607Q02S	0,08	0,00*	B	Referans (İngilizce)
CS638Q01S	- 0,01	0,55		
CS638Q02S	- 0,01	0,25		
CS638Q04S	0,16	0,00*	C	Referans (İngilizce)
CS615Q07S	- 0,01	0,33		
CS615Q01S	- 0,02	0,049*	A	
CS615Q02S	0,01	0,62		
CS615Q05S	- 0,01	0,60		

Not: * $p < 0,05$; $|\hat{\beta}| < 0,059$ A düzeyi, $0,059 \leq |\hat{\beta}| < 0,088$ B düzeyi, $|\hat{\beta}| \geq 0,088$ C düzeyi; $\hat{\beta} < 0$ Odak grup lehine, $\hat{\beta} > 0$ Referans grup lehine

Dil deęişkenine göre yapılan DMF analizinde, SIBTEST teknięi ile 0,05 düzeyinde istatistiksel olarak anlamlı bulunan maddelerin DMF etki büyüklüğünün belirlenmesinde Roussos ve Stout'un (1996) sınıflamasından yararlanılmıştır. Maddelerin hangi grup lehine çalıştığının tespiti için $\hat{\beta}$ parametresinin işaretine bakılmış, $\hat{\beta} < 0$ durumu odak grup lehine DMF'li olarak değerlendirilmiştir (Abbott, 2007; Clauser ve Mazor, 1998) .

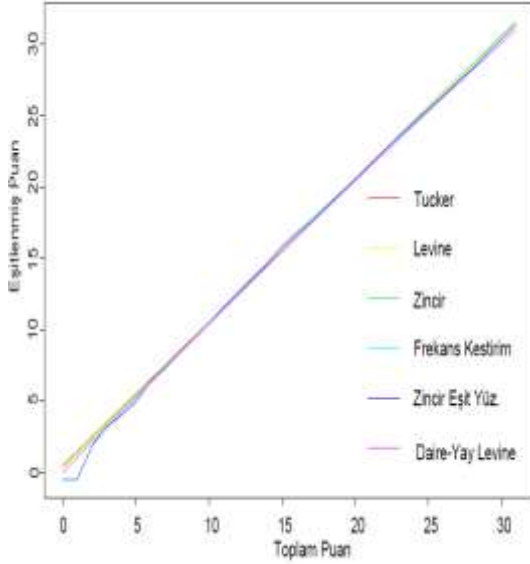
Tablo 18'deki bulgular incelendiğinde 2 maddenin (CS415Q08S, CS638Q04S) C düzeyinde, 1 maddenin (CS607Q02S) B düzeyinde DMF gösterdiği tespit edilmiştir. CS415Q08S maddesi İspanyolca dili lehine, CS638Q04S ve CS607Q02S maddeleri İngilizce dili lehine DMF göstermiştir. SIBTEST tekniğine göre DMF gösteren maddelerin ortak testten çıkarılması sonucunda ortak madde sayısı 15, toplam madde sayısı 32 olmuştur. MH ve SIBTEST analiz bulguları karşılaştırıldığında tekniklerin CS415Q08S, CS607Q02S, CS638Q04S maddelerini ortak olarak tespit ettikleri gözlemlenmiştir.

Dil kaynaklı DMF gösterdiği tekniklerce tespit edilen maddeler testten çıkarıldıktan sonra eşitleme doğrusal ve doğrusal olmayan yöntemlerle yürütülmüştür. MH tekniğince belirlenen maddelerin çıkarılması ile oluşan eşitlenmiş puanlar EK-B'de, SIBTEST tekniğine göre saptanan maddelerin çıkarılması ile elde edilen eşitlenmiş puanlar EK-C'de paylaşılmıştır.

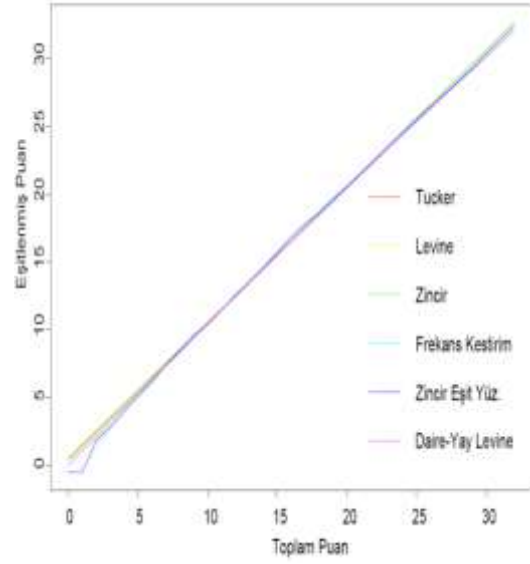
EK-A, EK-B ve EK-C'deki bulgular birlikte değerlendirildiğinde doğrusal yöntemlerde eşitlenmiş puanlar ham puanlardan yüksek hesaplanmış ayrıca ham puan ve eşitlenmiş puanlar arasındaki farkın ölçek boyunca arttığı görülmüştür. Doğrusal olmayan yöntemlerde frekans kestirim ve zincir eşit yüzdelikli eşitleme yöntemlerinde ham puanlar 4 puan seviyesinde eşitlenmiş puanlardan yüksek hesaplanmış diğer puan seviyelerinde düşük kalmıştır. Ham ve eşitlenmiş puanlar arasındaki farklar uç noktalarda azalırken, merkeze doğru artış göstermiştir. Farklar, doğrusal yöntemlerde EK-A'da 0,45 ile 0,67 aralığında iken madde çıkarılınca EK-B'de 0,31 ile 0,62, EK-C'de 0,35 ile 0,62, aralığına yerleşmiştir. Doğrusal olmayan yöntemlerde EK-A'da -0,50 ile 0,90 aralığında yer alırken maddeler çıkarılınca EK-B'de -0,50 ile 0,85, EK-C'de -0,50 ile 0,84 arasında deęişkenlik göstermiştir. Doğrusal yöntemlerde fark puanlarının alt ve üst sınırları deęişip ranjı genişlemiştir. Doğrusal olmayan yöntemlerde alt sınır sabit kalmış üst sınır küçülmüş böylelikle fark puanlarının ranjı daralmıştır. MH ve SIBTEST

tekniklerine göre DMF gösterdiği tespit edilen maddeler ortak testten çıkarıldıktan sonra ulaşılan eşitleme sonuçlarının yöntemlere göre grafiksel gösterimi Şekil 9'da sunulmuştur.

a) MH tekniğine göre



b) SIBTEST tekniğine göre



Şekil 9. Dil değişkenine göre belirlenen maddelerin testten çıkarılması sonucunda oluşan ham puanlar ile eşitlenmiş puanlara ait dağılım grafikleri

Şekil 9 incelendiğinde her iki DMF tekniğine göre maddelerin testten çıkarılması ile benzer iki grafik elde edilmiştir. Yalnızca zincir eşit yüzdelikli eşitlemeye ait çizginin yaklaşık 3-7 puan aralığında tekniklere göre az da olsa farklılaştığı görülmektedir. Onun haricinde her iki grafikte de çizgiler alt ve üst puanlardan merkeze doğru tek çizgi üzerinde seyretme eğiliminde olmuştur. Grafikler DMF'li madde varlığında yapılan Şekil 8'deki grafikle kıyaslandığında 1-7 puan arasında zincir eşit yüzdelikli eşitleme kaynaklı farklılıklar oluşmuş ancak zincir eşit yüzdelikli eşitleme yöntemine ait çizginin maddelerin çıkarılması ile puan ölçeği boyunca daha doğrusal bir yapıda ilerlediği görülmüştür. Diğer yöntemlere ait çizgiler her iki durumda da benzer görünüm sergilemiştir. Grafiksel karşılaştırmanın yanı sıra istatistiksel bir değerlendirme için her iki durumda yöntemlerden elde edilen eşitleme hataları AHKO değeri cinsinden Tablo 19'de paylaşılmıştır.

Tablo 19

Eşitleme Yöntemlerinin Alt Problem 1 ve 2'ye Göre AHKO Değerleri

Eşitleme Durumu	Doğrusal Eşitleme Yöntemleri			Doğrusal Olmayan Eşitleme Yöntemleri		
	Tucker	Levine	Zincir	Frekans Kestirim	Zincir Eşit Yüzdellikli	Daire-Yay Levine
DMF'li Maddelerin Varlığında	0,0074	0,0063	0,0068	0,0075	0,0070	0,0044
DMF'li Maddeler Çıkarıldığında (MH Tekniği)	0,0077	0,0057	0,0067	0,0081	0,0068	0,0039
DMF'li Maddeler Çıkarıldığında (SIBTEST Tekniği)	0,0078	0,0059	0,0068	0,0081	0,0069	0,0040

Dil değişkeni kaynaklı DMF gösterdiği MH tekniğine göre tespit edilen 4 madde, SIBTEST tekniğine göre saptanan 3 madde ayrı ayrı ortak testten çıkarılarak eşitleme kalan ortak maddelere dayalı olarak aynı yöntemler üzerinden yürütülmüştür. Tablo 19'daki veriler incelendiğinde en düşük eşitleme hatasını tüm durumlarda Daire-Yay Levine yöntemi üretmiş hatta düşen eşitleme hataları oransal değerlendirildiğinde yöntemin değişen durumlara duyarlı olduğu görülmüştür. DMF'li maddeler çıkarıldığında Daire-Yay Levine, zincir eşit yüzdellikli eşitleme ve Levine yöntemlerinde eşitleme hataları düşmüş diğer yöntemlerde yükselmiştir. Zincir doğrusal eşitleme yöntemine ait değerler incelendiğinde madde çıkarılmasından en az etkilenen yöntem olduğu söylenebilir. DMF'li maddelerin çıkarılması sonucunda yapılan eşitleme çalışmaları kendi içinde kıyaslandığında: ilk durumda daha çok madde çıkarılmış, eşitleme daha az maddeye dayalı olarak yürütülmüştür. Buna rağmen AHKO değerleri tüm yöntemlerde daha düşük çıkmıştır.

Dil kaynaklı DMF'li bulunan maddelerin testten çıkarılmasıyla eşitleme hatalarındaki değişimler yöntemler özelinde incelendikten sonra ortak ve toplam test puanları arasındaki ilişkinin değişimi gözlemlenmiş beraberinde dil değişkenini meydana getiren alt gruplar arası yetenek farklılıklarındaki değişimler de izlenmiştir. Madde çıkarılmadan önce 0,93 olan ortak test ve toplam test puanları arası Pearson momentler çarpımı korelasyon katsayısının, her iki tekniğine göre DMF'li bulunan maddelerin çıkarılmasıyla düştüğü görülmüştür ($r_{MH} = 0,91, p < 0,01; r_{SIBTEST} = 0,91 p < 0,01$).

Eşitleme üzerinde etkisi incelenen DMF kaynaklarından dil değişkeninin İngilizce ve İspanyolca alt grupları arasındaki puan ortalamaları farkının istatistiksel anlamlılığı t testi ile sınanmış, bulgular Tablo 20’de paylaşılmıştır.

Tablo 20

Dil Değişkenine Ait Alt Grupların Varyans ve Ortalamalarının Alt Problemlere Göre Karşılaştırılmasına İlişkin Bulgular

DMF Değişkeni	DMF Tekniği	Alt Gruplar	\bar{X}	S_x^2	Levene Testi		t testi	
					F	p	t	p
Dil	-	İspanyolca	9,77	12,63	0,07	0,79	-15,69	0,00*
		İngilizce	11,39	13,25				
	MH	İspanyolca	7,52	7,67	0,55	0,46	-13,67	0,00*
		İngilizce	8,64	8,24				
	SIBTEST	İspanyolca	8,39	8,37	0,01	0,92	-14,03	0,00*
		İngilizce	9,57	8,76				

Not: * $p < 0,05$

Tablo 20’deki bulgular değerlendirildiğinde dil değişkenine ait grupların ortak test puan varyansları arasında 0,05 düzeyinde istatistiksel olarak anlamlı bir fark görülmemiştir ($F_{dil} = 0,07$ $p > 0,05$). Testi İspanyolca ve İngilizce dillerinde alan bireylerin ortalamaları arasında oluşan fark 0,05 düzeyinde istatistiksel olarak anlamlı bulunmuştur ($t = -15,69$; $p < 0,05$). Bu bulgu fen okuryazarlığı testini İngilizce ve İspanyolca dillerinde alan bireyler arasında yetenek farklılığı olduğu şeklinde yorumlanabilir. DMF’li bulunan maddelerin testten çıkarılması t değerlerinin küçülmesine sebep olmuş ancak İspanyolca ve İngilizce dilleri ortalamaları arasındaki farkın istatistiksel anlamlılığı devam etmiştir.

Alt Problem 3’e İlişkin Bulgular

Cinsiyet kaynaklı DMF gösterdiği Mantel-Haenszel ve SIBTEST teknikleriyle tespit edilen maddelerin ortak testten çıkarılmaları, doğrusal ve doğrusal olmayan eşitleme yöntemiyle elde edilmiş puanları ve eşitleme hatasını nasıl etkilemiştir?

Bu alt problemin çözümünde araştırmanın DMF değişkenlerinden ikincisi cinsiyet değişkeni üzerinde yoğunlaşmıştır. PISA 2018 uygulaması 18 ve 24.

Kitapçıkların Fen okuryazarlığı ham puanlarının ortak teste dayalı eşitlemesi aşamasında MH tekniği ile cinsiyet lehine DMF gösteren maddelerin ortak testten çıkarılmasının eşitlenmiş puanlar ve eşitleme hatası üzerindeki etkisi incelenmiştir. Yöntemlerin eşitleme performansları AHKO değerleri üzerinden kıyaslanmıştır. Cinsiyete göre DMF'li olduğu düşünülen maddelerin ortak testten çıkarılmasının, ortak test ile toplam test puanları arasındaki korelasyon katsayısına etkisi incelenmiş aynı zamanda cinsiyet değişkenini meydana getiren kadın ve erkek alt grupları arasındaki yetenek farklılıklarına olan etkisi de gözlemlenmiştir. DMF analizi öncesinde kadınlar odak grup olarak, erkekler ise referans grup olarak tanımlanmıştır. MH tekniği ile ulaşılan bulgular Tablo 21'de paylaşılmıştır.

Tablo 21

Ortak Test Maddelerinin Cinsiyet Değişkenine Göre MH Analiz Bulguları

Madde Kodu	ΔMH	p	DMF Düzeyi	Avantaj Sağladığı Grup
CS478Q01S	-0,06	0,73		
CS478Q02S	-0,17	0,29		
CS478Q03S	0,33	0,03*	A	
CS415Q07S	0,19	0,26		
CS415Q02S	-0,11	0,64		
CS415Q08S	0,33	0,03*	A	
CS627Q01S	-0,89	0,00*	A	
CS627Q03S	-0,05	0,79		
CS627Q04S	0,50	0,00*	A	
CS607Q01S	-1,15	0,00*	B	Referans (Erkek)
CS607Q02S	-0,94	0,00*	A	
CS638Q01S	-1,04	0,00*	B	Referans (Erkek)
CS638Q02S	-0,50	0,02*	A	
CS638Q04S	0,61	0,00*	A	
CS615Q07S	-0,02	0,94		
CS615Q01S	-1,36	0,00*	B	Referans (Erkek)
CS615Q02S	0,38	0,02*	A	
CS615Q05S	0,25	0,20		

Not: * $p < 0,05$ $|\Delta MH| < 1$ A düzeyi, $1 < |\Delta MH| < 1,5$ B düzeyi, $1,5 < |\Delta MH|$ C düzeyi; $\Delta MH < 0$ Referans grup lehine, $\Delta MH > 0$ Odak grup lehine

Tablo 21'deki bulgular incelendiğinde MH tekniğine göre 3 madde (CS607Q01S, CS638Q01S, CS615Q01S) B düzeyinde erkekler lehine DMF

göstermektedir. Erkekler lehine avantaj sağladığı düşünülen maddelerin ortak testten çıkarılması sonucunda ortak madde sayısı 15'e, toplam madde sayısı 32'ye düşmüştür. Cinsiyet değişkenine göre DMF analizi SIBTEST tekniğine göre yapılmış bulgular Tablo 22'de verilmiştir.

Tablo 22

Ortak Test Maddelerinin Cinsiyet Değişkenine Göre SIBTEST Analiz Bulguları

Madde Kodu	β	p	DMF Düzeyi	Avantaj Sağladığı Grup
CS478Q01S	0,00	0,78		
CS478Q02S	0,02	0,23		
CS478Q03S	- 0,03	0,01*	A	
CS415Q07S	- 0,01	0,39		
CS415Q02S	0,01	0,48		
CS415Q08S	- 0,03	0,01*	A	
CS627Q01S	0,08	0,00*	B	Referans (Erkek)
CS627Q03S	0,00	0,76		
CS627Q04S	- 0,03	0,00*	A	
CS607Q01S	0,03	0,00*	A	
CS607Q02S	0,07	0,00*	B	Referans (Erkek)
CS638Q01S	0,08	0,00*	B	Referans (Erkek)
CS638Q02S	0,02	0,01*	A	
CS638Q04S	- 0,05	0,00*	A	
CS615Q07S	0,00	0,81		
CS615Q01S	0,07	0,00*	B	Referans (Erkek)
CS615Q02S	- 0,03	0,02*	A	
CS615Q05S	- 0,01	0,17		

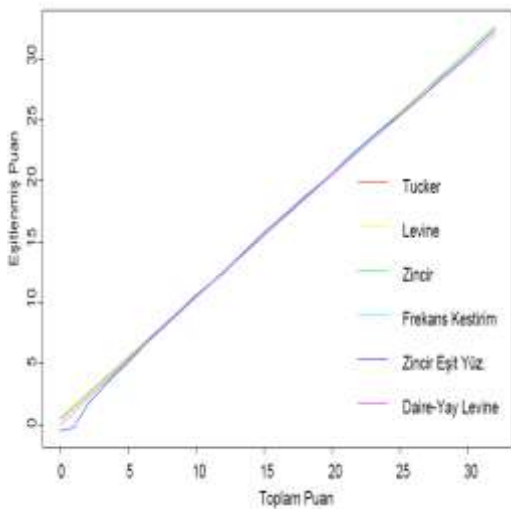
Not: * $p < 0,05$; $|\hat{\beta}| < 0,059$ A düzeyi, $0,059 \leq |\hat{\beta}| < 0,088$ B düzeyi, $|\hat{\beta}| \geq 0,088$ C düzeyi; $\hat{\beta} < 0$ Odak grup lehine, $\hat{\beta} > 0$ Referans grup lehine

Tablo 22 incelendiğinde SIBTEST tekniğine göre 4 maddenin (CS627Q01S, CS607Q02S, CS638Q01S, CS615Q01S) B düzeyinde erkekler lehine DMF gösterdiği tespit edilmiştir. SIBTEST tekniği ile saptanan maddelerin ortak testten çıkarılması ile kitapçıklardaki fen okuryazarlığı ortak madde sayısı 14, toplam madde sayısı 31 olmuştur. Her iki teknikle de elde edilen maddeler erkekler lehine DMF göstermiş tekniklerin üzerinde uzlaştığı maddeler CS638Q01S, CS615Q01S olmuştur.

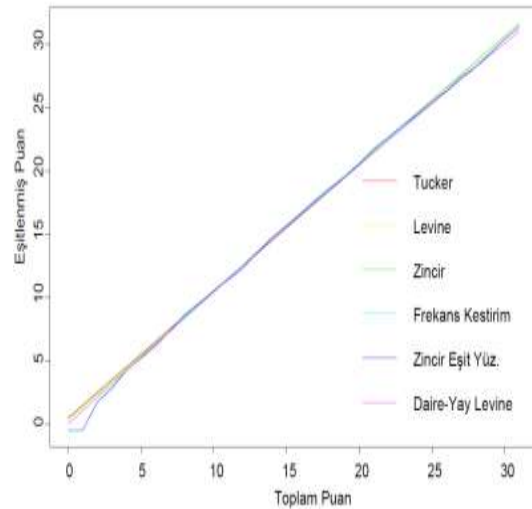
Cinsiyet kaynaklı DMF gösterdiği tekniklerce saptanan maddeler ortak testten çıkarılmış eşitleme işlemi doğrusal ve doğrusal olmayan yöntemlerle yinelenmiştir. MH tekniğine göre belirlenen maddelerin çıkarılması ile oluşan eşitlenmiş puanlar EK-Ç'de, SIBTEST tekniğine göre tespit edilen maddelerin çıkarılması ile elde edilen eşitlenmiş puanlar EK-D' sunulmuştur.

EK-A, EK-Ç ve EK-D'deki veriler karşılaştırıldığında doğrusal ve doğrusal olmayan yöntemlerde ham puan ve eşitlenmiş puanlar arası farklar doğrusal yöntemlerde alt uçtan üst uca doğru artarken doğrusal olmayan yöntemlerde puan ölçeğinin merkezine doğru artış göstermiştir. DMF'li maddelerin varlığında yapılan eşitlemede fark puanları doğrusal yöntemlerde 0,45 ile 0,67 aralığında seyrederken maddeler çıkarılınca EK-Ç'de 0,36 ile 0,67, EK-D'de 0,30 ile 0,60 aralığında değişkenlik göstermiştir. Maddelerin çıkarılması sonucunda fark puanlarına ait ranj genişlemiştir. Doğrusal olmayan yöntemlerde fark puanları -0,50 ile 0,90 aralığında değişirken DMF gösteren maddeler çıkarılınca farklar EK-Ç'de -0,50 ile 0,79, EK-D'de -0,50 ile 0,78 arasında dağılmıştır. Doğrusal olmayan yöntemlerde fark puanlarına ait ranjin daraldığı görülmüştür. Cinsiyet kaynaklı DMF gösterdiği MH ve SIBTEST tekniklerince tespit edilen maddelerin ortak testten çıkarılmasıyla ulaşılan eşitleme sonuçlarının grafiksel gösterimi yöntemlere göre Şekil 10'da sunulmuştur.

a) MH tekniğine göre



b) SIBTEST tekniğine göre



Şekil 10. Cinsiyet değişkenine göre belirlenen maddelerin testten çıkarılması sonucunda oluşan ham puanlar ile eşitlenmiş puanlara ait dağılım grafikleri

Şekil 10 incelendiğinde tekniklerce DMF gösterdiği tespit edilen maddelerin çıkarılması sonucunda elde edilen grafikler oldukça benzerdir. Alt ve üst puanlarda çizgiler az da olsa ayrışma gösterirken uçlardan merkeze doğru tek bir çizgi üzerinde seyretme eğiliminde olmuşlardır. Grafikler incelendiğinde sadece zincir eşit yüzdelikle eşitleme yöntemine ait çizginin 0-5 puan aralığında farklılaştığı görülmektedir. Grafikler DMF'li maddelerin varlığında yapılan eşitleme sonuçlarına ait grafiklerle kıyaslandığında 0-5 puan aralığında oluşan küçük dalgalanmalar dışında benzer görünüm sergilemişlerdir. Grafikselle mukayesenin yanı sıra istatistiksel kıyaslama için eşitleme yöntemlerinden elde edilen AHKO değerleri Tablo 23'te paylaşılmıştır.

Tablo 23

Eşitleme Yöntemlerinin Alt Problem 1 ve 3'e Göre AHKO Değerleri

Eşitleme Durumu	Doğrusal Eşitleme Yöntemleri			Doğrusal Olmayan Eşitleme Yöntemleri		
	Tucker	Levine	Zincir	Frekans Kestirim	Zincir Eşit Yüzdelikli	Daire-Yay Levine
DMF'li Maddelerin Varlığında	0,0074	0,0063	0,0068	0,0075	0,0070	0,0044
DMF'li Maddeler Çıkarıldığında (MH Tekniği)	0,0081	0,0066	0,0074	0,0083	0,0076	0,0044
DMF'li Maddeler Çıkarıldığında (SIBTEST Tekniği)	0,0081	0,0059	0,0070	0,0082	0,0071	0,0039

MH tekniğine göre 3 madde, SIBTEST tekniğine göre 4 maddenin cinsiyet değişkeni kaynaklı DMF gösterdiği tespit edilmiş, maddeler sırasıyla ortak testlerden çıkarılarak eşitleme işlemleri tekrar edilmiştir. Tablo 23'teki veriler incelendiğinde MH tekniğine göre madde çıkarılması Daire-yay Levine yöntemi dışındaki tüm yöntemlerde eşitleme hatasını yükselmiştir. SIBTEST tekniğine göre çıkarılan maddeler Levine ve Daire-yay Levine yöntemlerinde eşitleme hatalarını azaltırken diğer yöntemlerde artırmıştır. Tüm durumlarda elde edilen en düşük AHKO değeri sebebiyle Daire-yay Levine yöntemi uygulanan yöntemler içinde en uygunu olarak düşünülebilir. Çıkarılan maddeler tekniklere göre farklılaşma gösterse de Tucker yönteminde aynı AHKO değeri elde edilmiştir. SIBTEST tekniğinde daha çok madde saptanmış ve eşitleme daha az madde üzerinden

yürütülmüştür. Buna rağmen yöntemlerin genelinde daha az eşitleme hatası elde edilmiştir.

Cinsiyet kaynaklı DMF gösteren maddelerin ortak testten çıkarılması ortak test ile toplam test puanları arasındaki ilişkiyi etkilemiştir. DMF'li maddelerin varlığında 0,93 olan Pearson momentler çarpımı korelasyon katsayısı tekniklerce belirlenen maddeler çıkarılınca düşüş göstermiştir ($r_{MH} = 0,92$ $p < 0,01$; $r_{SIBTEST} = 0,90$ $p < 0,01$).

Cinsiyet değişkeni kaynaklı DMF gösteren maddelerin varlığı ve yokluğu durumlarına göre alt grupların yetenek farklılığındaki değişimler gözlemlenmiştir. Alt grupların puan ortalamaları arasındaki farkın istatistiksel anlamlılığı t testi ile sınanmış, bulgular Tablo 24'te paylaşılmıştır.

Tablo 24

Cinsiyet Değişkenine Ait Alt Grupların Varyans ve Ortalamalarının Alt Problemlere Göre Karşılaştırılmasına İlişkin Bulgular

DMF Değişkeni	DMF Tekniği	Alt Gruplar	\bar{X}	S_x^2	Levene Testi		t testi		
					F	p	t	p	
Cinsiyet	-	Kadın	10,80	13,63	1,234	0,27	-2,45	0,01*	
		Erkek	11,04	13,54					
	MH	Kadın	8,56	10,13	0,02	0,89	-0,68	0,49	
		Erkek	8,62	10,24					
	SIBTEST	-	Kadın	8,50	8,12	0,01	0,92	0,94	0,35
			Erkek	8,43	8,18				

Not: * $p < 0,05$

Tablo 24'deki bulgular değerlendirildiğinde cinsiyet değişkenine ait grupların ortak test puan varyansları arasında 0,05 düzeyinde istatistiksel olarak anlamlı bir fark görülmemiştir ($F_{cins} = 1,23$ $p > 0,05$). Testi alan kadın ve erkek bireyler arasında erkekler lehine olan 0,24'lük fark 0,05 düzeyinde istatistiksel olarak anlamlı bulunmuştur ($t = -2,45$; $p < 0,05$). Bu bulgu cinsiyet değişkenini oluşturan alt gruplar arasında yetenek farklılığı olduğu şeklinde yorumlanabilir. MH tekniğine göre belirlenen maddeler testten çıkarıldığında grupların puan ortalamaları arasındaki fark erkekler lehine 0,06'ya, SIBTEST tekniğinde 0,07'ye gerilemiştir. SIBTEST

tekniginde fark kadınlar lehine yön deęiřtirmiřtir. Dil deęiřkeninde oluřan durumun aksine cinsiyet deęiřkeninde madde ıkarılması gruplar arasındaki yetenek farklılıęını ortadan kaldırmıřtır. Kadın ve erkek puan ortalamaları arasındaki fark 0,05 düzeyinde istatistiksel olarak anlamlı bulunmamıřtır.

Bölüm 5

Sonuç, Tartışma ve Öneriler

Araştırmanın bu bölümünde elde edilen bulgular ilgili alanyazın ile karşılaştırılmış ve yorumlanmıştır. Ulaşılan sonuçlara dayalı olarak araştırmacılara önerilerde bulunulmuştur.

Sonuç ve Tartışma

Bu araştırmada, dil ve cinsiyet değişkenine göre DMF gösterdiği tespit edilen ortak maddelerin testten çıkarılması doğrusal ve doğrusal olmayan eşitleme yöntemlerinde eşitlenmiş puanları ve eşitleme hatasını nasıl etkilediği gözlemlenmiştir. DMF değişkenlerinin eşitleme sonuçları üzerindeki etkisi PISA 2018 verileri üzerinden sınanmıştır. Böylelikle eşitleme yöntemlerinin performanslarındaki gerçek değişimlerin gözlemlenmesi amaçlanmıştır. Dil ve cinsiyet değişkenine göre, MH ve SIBTEST teknikleriyle tespit edilen maddelerin ayrı ayrı ortak testten çıkarılması neticesinde kitapçıklar kalan ortak maddelere dayalı olarak her durum için tekrar eşitlenmiştir. Elde edilen sonuçlar maddeler halinde sunulmuştur.

1. Dil ve cinsiyet değişkenine göre DMF gösteren maddelerin ortak testten çıkarılması sonucunda elde edilen eşitleme hataları DMF'li maddelerin varlığında yapılan durumla karşılaştırıldığında hatalardaki değişimin yönü eşitleme yöntemlerine göre farklılaşmıştır. Eşitleme hataları Tucker ve frekans kestirim yöntemlerinde artarken Daire-yay Levine yönteminde azalma eğiliminde olmuştur. Etkisi incelenen diğer yöntemlerde ise sonuçlar DMF kaynağına göre değişkenlik göstermiştir. DMF'li maddelerin ortak testten çıkarılmasıyla eşitleme yöntemlerinde görülen farklılaşmalar DMF'nin test eşitleme üzerine etkilerinin çalışıldığı araştırmalarla benzerdir (Atalay Kabasakal, 2014; Demirus, 2015; Gübeş ve Uyar, 2020).
2. Dil kaynaklı DMF gösteren maddelerin tespitinde MH tekniği bir maddeyi fazladan B düzeyinde İngilizce lehine DMF'li olarak saptamıştır. MH tekniği ile DMF'li bulunan 1 maddenin daha testten çıkarılmasıyla yapılan eşitlemede yöntemlerin hepsinde eşitleme hataları azalış göstermiştir. Bu sonuç, Atalay Kabasakal'ın (2014) 40 maddelik test uzunluğunda DMF gösteren maddelerin testten çıkarılmasının eşitlemeyi olumsuz etkilemediği

hatta bazı durumlarda düşürdüğünü belirttiği araştırması, Turhan'ın (2006) DMF'li ortak madde sayısının artmasının ortak testin tersine etki göstermeye başlayacağı ifadesi, Yurtçu ve Güzeller'in (2017) daha az DMF'li ortak madde ile daha düşük eşitleme hatası elde ettiği araştırma bulgusu ve Huggins'in (2014) ortak testin 4 ve 6 DMF'li madde içerdiği durumlarda DMF'nin eşitlemenin grup değişmezliği özelliği üzerinde anlamlı etkisinin görülmediği sonucu ile tutarlılık göstermiştir. Ancak daha çok madde çıkarmak eşitleme işleminin daha az ortak madde üzerinden yürütülmesini gerektirir. Kolen ve Brennan'a (2004) göre ortak madde sayısındaki artışın eşitleme hatasını düşüreceği görüşü araştırma sonuçları ile tam bir tutarlılık göstermemiştir. Bir başka açıdan bakıldığında; Klein ve Kolen (1985) ortak test uzunluğunun eşitleme kesinliği üzerine etkilerini geleneksel eşitleme yöntemleri üzerinden inceledikleri araştırmalarında; formları alan bireyler yetenek düzeyleri bakımından farklılık gösteriyorsa ortak test uzunluğunun eşitleme sonuçları üzerinde önemli etkileri olduğunu belirtmiştir (akt. Cook ve Paterson, 1987). Nitekim bu araştırmada kitapçıklar arası puan ortalamaları farkı istatistiksel olarak anlamlı çıkmasına rağmen dil değişkenine dayalı eşitlemelerde daha az ortak madde ile her eşitleme yönteminde daha düşük eşitleme hatası elde edilmiştir. Bu bulgu Klein ve Kolen'in bulgusuna ters düşmüştür.

3. Chu (2002) "Eşdeğer gruplarda DMF varlığında test eşitleme" adlı araştırmasında, formlarda alt gruplar seviyesinde ölçek genelinde görülen madde güçlükleri farklılaşmasının DMF etkisinin incelendiği bir çalışmayı olumsuz etkileyebileceğini belirtmiştir. Bu araştırmada DMF kaynaklarını oluşturan alt grupların yetenek düzeyleri farklılaşması istatistiksel olarak anlamlı bulunmuştur. Bu bulgu Chu'nun bulgusunu destekler niteliktedir. Dolayısıyla formların genelinde görülen madde güçlükleri farklılaşması DMF gösteren maddelerin belirlenmesini etkilemiş, neticesinde eşitleme hatalarındaki tutarsızlığın sebebi olmuş olabilir. Bu sonuç Camilli ve Shepard'ın (1994) yanlılık üzerine yapılan araştırmalarda DMF belirlemede birkaç teknikten yararlanılması gerekliliğini doğrular niteliktedir.
4. Dil kaynaklı SIBTEST tekniği sonucu ile cinsiyet kaynaklı MH tekniği sonucunda DMF gösterdiği tespit edilen farklı üçer madde ortak testten

çıkarılmıştır. İki durum arasındaki eşitleme hataları arasında ciddi farklar gözlenmiştir. Bu sonucun temel sebebi cinsiyet değişkenini meydana getiren alt gruplar arasındaki yetenek farklılığının dil değişkenini meydana getiren alt gruplar arasındakinden oldukça küçük olması olabilir.

5. Bulgular eşitleme yöntemleri üzerinden değerlendirildiğinde, DMF gösteren maddelerin ortak testten çıkarılması sonucunda doğrusal ve doğrusal olmayan yöntemlerde tutarlı değişimler görülmüştür. Bu sonuç, Gübeş ve Uyar'ın (2020) ortak testin DMF'li madde içerdiği ve içermediği durumlarda eşitleme yöntemlerini karşılaştırdıkları araştırma sonuçları ile benzerlik göstermektedir. Bunun yanı sıra doğrusal yöntemlerde Tucker, doğrusal olmayan yöntemlerde frekans kestirim yönteminde her durumda eşitleme hataları artış göstermiştir. Livingston (2014) eğer ortak test puanları toplam test puanları ile mükemmel düzeyde ilişki gösterirse Tucker eşitleme yönteminin zincir doğrusal eşitleme yöntemi ile özdeş hale geleceğini belirtmiştir. Araştırmada puanlar arasında çok yüksek düzeyde istatistiksel olarak anlamlı pozitif yönde korelasyona rastlansa da ilişki Livingston'un tanımladığı mükemmellik düzeyine erişememiştir. Tucker ve zincir doğrusal eşitleme yöntemleri arasında alt problemler düzeyinde tutarlı bir ilerleme gözlemlenmemesi Livingston'un bu yaklaşımının bir sonucu olabilir. Livingston aynı zamanda ortak puanlar ile toplam puanlar arasındaki korelasyonun 1'den ayrıldığı ölçüde frekans kestirim denklemi eşitleme örneklemelerini yetenek bakımından olduğundan daha fazla benzemiş gibi ayarlama yapacağını ifade etmiştir. Araştırmada DMF'li maddelerin atılmasıyla ilk duruma göre formlar arasındaki yetenek farklılıkları artmış ortak ve toplam test puanları arasındaki korelasyonlar da azalmıştır. DMF'li maddelerin testten çıkarılması ile frekans kestirim yönteminde artan eşitleme hataları bu yaklaşımla açıklanabilir.

Öneriler

Bu araştırmanın, DMF'li maddelerin varlığında test eşitlemesi yapacak araştırmacılara yöntem, teknik ve yaklaşım belirlemede katkı sunacağı düşünülmektedir.

Araştırma sonuçlarına dayalı öneriler.

1. Eşitleme yöntemlerinde farklılaşan sonuçlar, DMF gösterdiği düşünülen maddeleri testten çıkarmanın ilk akla gelen yol olmaması gerektiğini göstermiştir. Eşitleme işleminde DMF'li maddelerin testten çıkarılması daha yüksek eşitleme hatasına sebep oluyorsa Chu'nun (2002) da belirttiği gibi daha düşük eşitleme hatası ve kararlı madde parametre kestirimi için eşitleme işlemini DMF'li maddeleri çıkarmadan yürütmek önerilebilir.
2. Daire-yay Levine yöntemi DMF'li maddelerin yokluğunda yapılan tüm eşitleme durumlarında DMF'li maddelerin varlığında yapılan eşitlemeye göre daha az eşitleme hatası üretmiştir. Yöntem, DMF'li maddelerden arındırılarak yapılacak eşitleme çalışmaları için önerilebilir.
3. DMF kaynağını oluşturan alt gruplar arasındaki yetenek farklılaşmasının eşitleme sonuçları üzerinde etkili olabileceği sonucundan hareketle yetenek farklılaşmasının olabildiğince az olduğu alt gruplarla çalışılmasının uygun olacağı önerilebilir.

Sonraki araştırmalara yönelik öneriler.

1. Bu araştırmanın örneklemini 26 farklı ülkeye ait birey verilerinden oluşmaktadır. Öğrenciler arası kültür farklılığı bir DMF değişkeni olarak eşitlemenin grup değişmezliği koşulunu etkilemiş olabilir. Benzer bir çalışma aynı kültüre ait örneklem üzerinden yürütülebilir.
2. Bu araştırma PISA 2018 Fen okuryazarlığı maddeleri öğrenci verileri üzerinden yürütülmüştür. Benzer bir çalışma matematik okuryazarlığı ve okuma becerileri maddeleri kullanılarak yapılabilir.
3. Bu araştırmada gerçek uygulamadan elde edilmiş veriler kullanılmıştır. Bu durum eşitleme üzerinde etkisi incelenecek olan DMF koşullarını sınırlandırmıştır. İleriki çalışmalarda ortak test uzunluğu, DMF'li madde sayısı, DMF etki büyüklüğü, DMF'li maddelerin türü, DMF'li maddelerin tek biçimli olması veya olmaması gibi değişkenlerin eşitleme hataları üzerindeki etkilerinin gerçek veri üzerinden sınanabileceği araştırmalara yer verilebilir.

4. Arařtırmada DMF teknikleri ve eřitleme ynlemleri KTK'ya dayalı yrtlmřtr. DMF'nin eřitleme zerindeki etkisi MTK temelli ynlem ve tekniklerle yrtlebilir.
5. DMF kaynađı olarak dil ve cinsiyet deđiřkeni kullanılmıřtır. Farklı DMF kaynaklarının eřitleme zerindeki etkileri incelenebilir. Hatta DMF'li maddelerin belirlenmesi gizil sınıf yaklařımıyla srdrlebilir.
6. PISA uygulamasında maddeler kamuoyu ile paylařılmadıđı iin alıřma DMF ile yrtlmřtr. lkemizde MEB tarafından uygulanan LGS, AL, AF gibi veya niversitelerin dil yeterliliđi iin uyguladıkları sınavlarda yanlılık gsteren maddelerin eřitleme zerindeki etkileri incelenebilir.
7. Bu arařtırmada rneklem byklđ byk rneklem seviyesindedir. İleriki arařtırmalarda sadece lkemiz veya bařka bir lkeye ait kitapıklarda aynı temalı alıřma kk rneklem boyutunda yapılabilir.

Kaynaklar

- Abbott, M. L. (2007). A confirmatory approach to differential item functioning on an ESL reading assessment. *Language Testing*, 24(1), 7–36. <https://doi.org/10.1177/0265532207071510>
- Ackerman, T. A. (1992). A Didactic Explanation of Item Bias, Item Impact, and Item Validity From a Multidimensional Perspective. *Journal of Educational Measurement*, 29(1), 67–91. <https://doi.org/10.1111/j.1745-3984.1992.tb00368.x>
- Ackerman, T. A., & Evans, J. A. (1992). An investigation of the relationship between reliability , power and the type I error rate of the Mantel-Haenszel and simultaneous item bias detection procedures. *Paper Presented at the 32nd Annual Meeting of the National Council on Measurement in Education*. <https://files.eric.ed.gov/fulltext/ED344937.pdf>
- Albano, A. D. (2015). A general linear method for equating with small samples. *Journal of Educational Measurement*, 52(1), 55–69.
- Albano, A. D. (2016). Equate: An R package for observed-score linking and equating. *Journal of Statistical Software*, 74(8), 1–36. <https://doi.org/10.18637/jss.v074.i08>
- Allalouf, A., Hambleton, R. K., & Sireci, S. G. (1999). Identifying the causes of dif in translated verbal items. *Journal of Educational Measurement*, 36(3), 185–198. <https://doi.org/10.1111/j.1745-3984.1999.tb00553.x>
- Angoff, W. H. (1971). Scales, norms and equivalent scores. In R. L. Thorndike (Ed.), *Educational Measurement* (2nd ed., pp. 508–600). American Council on Education.
- Angoff, W. H. (1984). *Scales, norms, and equivalent scores*. Princeton, NJ: Educational Testing Service.
- Angoff, W. H. (1993). Perspectives on differential item functioning methodology. In P. W. Holland & H. Wainer (Eds.), *Differential Item Functioning* (pp. 3–23). Lawrence Erlbaum Associates.
- Aşiret, S. (2014). *Küçük örneklemelerde test eşitleme yöntemlerinin çeşitli faktörlere göre incelenmesi* (Yüksek Lisans Tezi, Mersin Üniversitesi).

- <https://tez.yok.gov.tr/UlusalTezMerkezi/tarama.jsp>
- Atalay Kabasakal, K. (2014). *Değişen madde fonksiyonunun test eşitlemeye etkisi* (Doktora tezi, Hacettepe Üniversitesi).
<https://tez.yok.gov.tr/UlusalTezMerkezi/tarama.jsp>
- Bakan Kalaycıoğlu, D. (2008). *Öğrenci Seçme Sınavı'nın madde yanlılığı açısından incelenmesi* (Doktora tezi, Hacettepe Üniversitesi).
<https://tez.yok.gov.tr/UlusalTezMerkezi/tarama.jsp>
- Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. Guilford Press.
- Baykul, Y. (1999). *İstatistik: Metodlar ve uygulamalar* (4. Baskı). Anı Yayıncılık.
- Brossman, B. G. (2013). Observed score and true score equating procedures for multidimensional item response theory. In *Applied Psychological Measurement* (Vol. 37, Issue 6). <https://doi.org/10.1177/0146621613484083>
- Budescu, D. (1985). Efficiency of linear equating as a function of the length of the anchor test. *Journal of Educational Measurement*, 22(1), 13–20.
<https://doi.org/10.1111/j.1745-3984.1985.tb01045.x>
- Büyüköztürk, Ş. (2017). *Sosyal bilimler için veri analizi el kitabı: İstatistik, araştırma deseni, SPSS uygulamaları ve yorum* (23. Baskı). Pegem Akademi.
- Büyüköztürk, Ş., Kılıç Çakmak, E., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2019). *Eğitimde bilimsel araştırma yöntemleri* (27. Baskı). Pegem Akademi.
- Camilli, G., & Shepard, L. A. (1994). *Methods for identifying biased test items* (4th ed.). Sage Publications.
- Can, A. (2019). *SPSS ile bilimsel araştırma sürecinde nicel veri analizi* (7. Baskı). Pegem Akademi.
- Chen, H., Cui, Z., Zhu, R., & Gao, X. (2010). Evaluating the effects of differences in group abilities on the Tucker and the Levine observed-score methods for common-item nonequivalent groups equating. *ACT Research Report Series 2010-1*, Inc. <https://files.eric.ed.gov/fulltext/ED527217.pdf>
- Chu, K.-L. (2002). *Equivalent group test equating with the presence of differential item functioning* (Doctoral dissertation, Florida State University).

<https://search.proquest.com/docview/275897153/9A8021911A4F4DE5PQ/1?accountid=11248>

- Clauser, B. E., & Mazor, K. M. (1998). Using statistical procedures to identify differentially functioning test items. *Educational Measurement: Issues and Practice*, 17(1), 31–44. <https://doi.org/10.1111/j.1745-3992.1998.tb00619.x>
- Cohen, J., & Cohen, P. (1983). Applied multiple regression/correlation analysis for the behavioral sciences. In *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences* (2nd ed.). Lawrence Erlbaum Associates. <https://doi.org/10.4324/9780203774441>
- Cohen, R. J., & Swerdlik, M. E. (2010). *Psychological testing and assessment: An introduction to tests and measurement* (7th ed.). McGraw–Hill Primis.
- Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2016). *Sosyal bilimler için çok değişkenli istatistik SPSS ve LISREL uygulamaları* (4. Baskı). Pegem Akademi.
- Cook, L. L., & Paterson, N. S. (1987). Problems related to the use of conventional and item response theory equating methods in less than optimal circumstances. *Applied Psychological Measurement*, 11(3), 225–244. <https://doi.org/10.1177/014662168701100302>
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. CA: Harcourt Brace Jovanovich College Publishers.
- Demirus, K. B. (2015). *Ortak maddelerin değişen madde fonksiyonu gösterip göstermemesi durumunda test eşitleme etkisinin farklı yöntemlerle incelenmesi* (Doktora Tezi, Hacettepe Üniversitesi). <https://tez.yok.gov.tr/UlusalTezMerkezi/tarama.jsp>
- Dorans, N. J. (1990). Equating methods and sampling designs. *Applied Measurement in Education*, 3(1), 3–17. https://doi.org/10.1207/s15324818ame0301_2
- Dorans, N. J., & Holland, P. W. (1993). DIF detection and description: Mantel-Haenszel and standardization. In P. W. Holland & H. Wainer (Eds.), *Differential Item Functioning* (pp. 35–66). Lawrence Erlbaum Associates.
- 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. *ETS Research Report Series*, 2010(2), i–41. <https://doi.org/10.1002/j.2333-8504.2010.tb02236.x>
- Dorans, N. J., Moses, T. P., & Eignor, D. R. (2011). Equating Test Scores: Toward Best Practices. In A. A. Von Davier (Ed.), *Statistical Models for Test Equating, Scaling, and Linking* (pp. 21–42). https://doi.org/10.1007/978-0-387-98138-3_2
- Dorans, N. J., Pommerich, M., & Holland, P. W. (2007). Linking and aligning scores and scales. In *Nonparametric Statistics for Social and Behavioral Sciences*. <https://doi.org/10.1201/b16188>
- Ellis, B. B., & Raju, N. S. (2003). Test and item bias : What they are, what they aren't, and how to detect them. *Educational Resources Information Center*, 88–98. <https://files.eric.ed.gov/fulltext/ED480042.pdf>
- Ercikan, K. (1998). Translation effects in international assessments. *International Journal of Educational Research*, 29(6), 543–553. [https://doi.org/10.1016/S0883-0355\(98\)00047-0](https://doi.org/10.1016/S0883-0355(98)00047-0)
- Ercikan, K. (2002). Disentangling sources of differential item functioning in multilanguage assessments. *International Journal of Testing*, 2(3–4), 199–215. <https://doi.org/10.1080/15305058.2002.9669493>
- Felan, G. D. (2002). Test equating: Mean, linear, equipercentile, and item response theory. *Paper Presented at the Annual Meeting of the Southwest Educational Research Association*. <https://eric.ed.gov/?id=ED462436>
- French, B. F., & Maller, S. J. (2007). Iterative purification and effect size use with logistic regression for differential item functioning detection. *Educational and Psychological Measurement*, 67(3), 373–393. <https://doi.org/10.1177/0013164406294781>
- Gök, B., Kelecioğlu, H., & Dogan, N. (2010). Değişen madde fonksiyonunu belirlemede Mantel-Haenszel ve Lojistik Regresyon tekniklerinin

- karşılaştırılması. *Eğitim Ve Bilim*, 35(156), 3–16.
- González, J., & Wiberg, M. (2017). *Applying test equating methods using R*. Springer International Publishing.
- Green, S. B., & Salkind, N. J. (2009). *Using SPSS for Windows and Macintosh: Analyzing and Understanding Data* (7th ed.). Pearson.
- Gübeş, N., & Uyar, Ş. (2020). Comparing performance of different equating methods in presence and absence of DIF items in anchor test. *International Journal of Progressive Education*, 16(3), 111–122. <https://doi.org/10.29329/ijpe.2020.248.8>
- Gür, E. (2019). *PISA 2015 uygulamasındaki maddelerin kültüre göre değişen madde fonksiyonu açısından incelenmesi*. (Yüksek lisans tezi) Hacettepe Üniversitesi.
- Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice*, 12(3), 38–47.
- Hambleton, R. K., & Swaminathan, H. (1985). *Item response theory: Principles and applications* (1st ed.). Springer Netherlands.
- Han, T., Kolen, M., & Pohlmann, J. (1997). A comparison among IRT true-and observed-score equatings and traditional equipercentile equating. *Applied Measurement in Education*, 10(2), 105–121. https://doi.org/10.1207/s15324818ame1002_1
- Hanson, B. A., Zeng, L., & Colton, D. (1994). *A comparison of presmoothing and postsmoothing methods in equipercentile equating* (No.94). American College Testing Program.
- Harris, D. J. (2003). Equating the multistate bar examination. *The Bar Examiner*, 72(3), 12–18.
- Holland, P. W. (2007). A framework and history for score linking. In N. J. Dorans, M. Pommerich, & P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 5–30). https://doi.org/10.1007/978-0-387-49771-6_2
- Holland, P. W., & Dorans, N. J. (2006). Linking and equating. In R. L. Brennan

- (Ed.), *Educational measurement* (4th ed., pp. 189–220). Praeger Publishers.
- Holland, P. W., Dorans, N. J., & Petersen, N. S. (2006). 6 equating test scores. In C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics* (pp. 169–203). Elsevier. [https://doi.org/10.1016/S0169-7161\(06\)26006-1](https://doi.org/10.1016/S0169-7161(06)26006-1)
- Holland, P. W., & Thayer, D. T. (1986). *Differential item functioning and the Mantel-Haenszel procedure* (No.86-89). Educational Testing Service.
- Holland, P. W., & Thayer, D. T. (1988). Differential item performance and the Mantel-Haenszel procedure. In H. Wainer & H. I. Braun (Eds.), *Test validity* (pp. 129–143). Routledge.
- Holmes, S. E. (1986). Test equating and credentialing examinations. *Evaluation & The Health Professions*, 9(2), 230–249.
- 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>
- Huggins, A. C., & Penfield, R. D. (2012). An NCME Instructional Module on Population Invariance in Linking and Equating. *Educational Measurement: Issues and Practice*, 31(1), 27–40. <https://doi.org/10.1111/j.1745-3992.2011.00225.x>
- Ironson, G. H., & Craig, R. (1982). Item Bias Techniques When Amount of Bias Is Varied and Score Differences between Groups Are Present. Final Report.
- Jarjoura, D., & Kolen, M. J. (1985). Standard Errors of Equipercentile Equating for the Common Item Nonequivalent Populations Design. *Journal of Educational Statistics*, 10(2), 143–160. <https://doi.org/10.3102/10769986010002143>
- Kabasakal Atalay, K., & Kelecioğlu, H. (2012). Evaluation of attitude items in PISA 2006 student questionnaire in terms of differential item functioning. *Ankara Universitesi Egitim Bilimleri Fakultesi Dergisi*, 45(2), 77–96. https://doi.org/10.1501/egifak_0000001254
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31–36. <https://doi.org/10.1007/BF02291575>
- Kan, A. (2011). Test Eşitleme : OKS testlerinin istatistiksel eşitliğinin sınanması.

Eğitim ve Bilim, 36(160), 38–51.

Kaplan, R. M., & Saccuzzo, D. P. (2018). *Psychological testing: Principles, applications, and issues* (9th ed.). Cengage Learning.

Karagül, A. E. (2020). *Küçük örneklerde çok kategorili puanlanan maddelerden oluşan testlerde klasik test eşitleme yöntemlerinin karşılaştırılması* (Yüksek lisans tezi, Ankara Üniversitesi).
<https://tez.yok.gov.tr/UlusalTezMerkezi/tarama.jsp>

Karakaya, I., & Kutlu, Ö. (2012). Seviye belirleme sınavındaki Türkçe alt testlerinin madde yanlılığının incelenmesi. *Eğitim ve Bilim*, 37(165), 348–362.

Kelecioğlu, H. (1994). *Öğrenci seçme sınavı puanlarının eşitlenmesi üzerine bir çalışma* (Doktora Tezi, Hacettepe Üniversitesi).
<https://tez.yok.gov.tr/UlusalTezMerkezi/tarama.jsp>

Kelecioğlu, H., & Gübeş, N. Ö. (2013). Comparing linear equating and equipercentile equating methods using random groups design. *International Online Journal of Educational Sciences*, 5(1), 227–241.

Kelecioğlu, H., Karabay, B., & Karabay, E. (2014). Investigation of placement test in terms of item biasness. *Elementary Education Online*, 13(3), 934–953.

Kilmen, S. (2019). *Ölçme ve değerlendirmede temel kavramlar*. R. N. Demirtaşlı (Ed.), *Eğitimde Ölçme ve Değerlendirme* (s. 25-56) içinde. Anı Yayıncılık

Kim, W. (2003). *Development of a differential item functioning (DIF) procedure using the hierarchical generalized linear model: A comparison study with logistic regression procedure* (Doctoral Dissertation, Pennsylvania State University). <https://etda.libraries.psu.edu/catalog/6080>

Klein, L. W., & Kolen, M. J. (1985). Effect of number of common items in common-item equating with nonrandom groups. *In Annual Meeting of the American Educational Research Association. Chicago.*

Kolen, M. J. (1984). Effectiveness of analytic smoothing in equipercentile equating. *Journal of Educational Statistics*, 9(1), 25–44.

Kolen, M. J. (1985). Standard errors of Tucker equating. *Applied Psychological Measurement*, 9(2), 209–223. <https://doi.org/10.1177/014662168500900209>

- Kolen, M. J. (1988). Traditional equating methodology. *Educational Measurement: Issues and Practice*, 7(4), 29–37. <https://doi.org/10.1111/j.1745-3992.2011.00208.x>
- Kolen, M. J. (1990). Does matching in equating work: A discussion. *Applied Measurement in Education*, 3(1), 97–104. https://doi.org/10.1207/s15324818ame0301_7
- Kolen, M. J., & Brennan, R. L. (1987). Linear equating models for the common-item nonequivalent-populations design. *Applied Psychological Measurement*, 11(3), 263–277. <https://doi.org/10.1177/014662168701100304>
- Kolen, M. J., & Brennan, R. L. (1995). *Test equating: Methods and practices* (1st ed.). Springer.
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking: Methods and practices* (2nd ed.). Springer.
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling, and linking: Methods and practices* (3rd ed.). Springer.
- Lee, H. S., & Geisinger, K. F. (2016). The matching criterion purification for differential item functioning analyses in a large-scale assessment. *Educational and Psychological Measurement*, 76(1), 141–163. <https://doi.org/10.1177/0013164415585166>
- Levine, R. (1955). Equating the score scales of alternate forms administered to samples of different ability. *ETS Research Bulletin Series*, 1955(2), i–118.
- Linn, R. L. (1993). Linking results of distinct assessments. *Applied Measurement in Education*, 6(1), 83–102. https://doi.org/10.1207/s15324818ame0601_5
- Liou, M., & Cheng, P. E. (1995). Asymptotic standard error of equipercentile equating. *Journal of Educational and Behavioral Statistics*, 20(3), 259–286. <https://doi.org/10.3102/10769986020003259>
- Livingston, S. A. (2014). Equating Test Scores (without IRT). *Educational Testing Service*.
- Livingston, S. A., Dorans, N. J., & Wright, N. K. (1990). What combination of sampling and equating methods works best? *Applied Measurement in*

- Education*, 3(1), 73–95. https://doi.org/10.1207/s15324818ame0301_6
- Livingston, S. A., & Kim, S. (2009). The circle-arc method for equating in small samples. *Journal of Educational Measurement*, 46(3), 330–343. <https://doi.org/10.1111/j.1745-3984.2009.00084.x>
- Livingston, S. A., & Kim, S. (2010). Random-groups equating with samples of 50 to 400 test takers. *Journal of Educational Measurement*, 47(2), 175–185. <https://doi.org/10.1111/j.1745-3984.2010.00107.x>
- Lord, F. M. (1950). Notes on comparable scales for test scores. *ETS Research Bulletin Series*, 1950(2), i–21.
- Lord, F. M. (1955). Equating test scores-A maximum likelihood solution. *Psychometrika*, 20(3), 193–200. <https://doi.org/10.1007/BF02289016>
- Lord, F. M. (1980). *Applications of item response theory to practical testing problems*. Routledge.
- Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores 1968*. Reading, MA Addison-Wesley.
- Lord, F. M., & Wingersky, M. S. (1984). Comparison of IRT true-score and equipercentile observed-score " equatings ". *Applied Measurement in Education*, 8(4), 453–461.
- 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>
- Mantel, N., & Haenszel, W. (1959). Statistical aspects of the analysis of data from retrospective studies of disease. *Journal of the National Cancer Institute*, 22(4), 719–748. <https://doi.org/10.1093/jnci/22.4.719>
- MEB. (2018). *Uluslararası Öğrenci Değerlendirme Programı: Pisa 2015 Ulusal Raporu*. M.E.B. Ölçme Değerlendirme ve Sınav Hizmetleri Genel Müdürlüğü.
- Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. *Journal of Educational Statistics*, 7(2), 105–118. <https://doi.org/10.3102/10769986007002105>
- Michael, J., Paul, W., Meryl, W., & Cadelle, F. (1999). *Uncommon measures:*

- Equivalence and linkage among educational tests.* National Academy Press.
- Mislevy, R. J. (1992). Linking educational assessments: Concepts, issues, methods, and prospects. *Educational Testing Service.*
- Murphy, K. R., & Davidshofer, C. O. (2005). *Psychological testing: Principles and applications* (6th ed.). Pearson Education International.
- Nandakumar, R. (1993). Simultaneous DIF amplification and cancellation: Shealy-Stout's test for DIF. *Journal of Educational Measurement*, 30(4), 293–311. <https://doi.org/10.1111/j.1745-3984.1993.tb00428.x>
- OECD. (2018a). *PISA 2018 Technical Report (Chapter 2): Figures and tables.* <https://www.oecd.org/pisa/data/pisa2018technicalreport/PISA2018TecReport-Ch-02-Test-Design.pdf>
- OECD. (2018b). *PISA 2018 Technical Report (Chapter 2): Test design and test development.* <https://www.oecd.org/pisa/data/pisa2018technicalreport/PISA2018TecReport-Ch-02-Test-Design.pdf>
- OECD. (2019a). *PISA 2018 Results (Volume I): What Students Know and Can Do.* OECD Publishing. <https://doi.org/10.1787/5f07c754-en>
- OECD. (2019b). *PISA 2018 Technical Report (Chapter 9): Scaling PISA data.* <https://www.oecd.org/pisa/data/pisa2018technicalreport/Ch.09-Scaling-PISA-Data.pdf>
- Osterlind, S. (1983). *Test item bias.* Sage Publications. <https://doi.org/10.4135/9781412986090>
- ÖSYM. (2020). *2020 akademik personel ve lisansüstü eğitimi giriş sınavı (ALES) kılavuzu.* <https://dokuman.osym.gov.tr/pdfdokuman/2020/ALES-1/aleskilavuz15052020.pdf>
- Ozdemir, B. (2017). Equating TIMSS mathematics subtests with nonlinear equating methods using NEAT design: Circle-Arc equating approaches. *International Journal of Progressive Education*, 13(2), 116–132.
- Potenza, M. T., & Dorans, N. J. (1995). A Framework for Classification and Evaluation. *Applied Psychological Measurement*, 19(1), 23–37.

- Roussos, L., & Stout, W. (1996). A multidimensionality-based DIF analysis paradigm. *Applied Psychological Measurement*, 20(4), 355–371. <https://doi.org/10.1177/014662169602000404>
- Sansivieri, V., Wiberg, M., & Matteucci, M. (2017). A review of test equating methods with a special focus on IRT-based approaches. *Statistica*, 77(4), 352.
- Shepard, L. A., Camilli, G., & Averill, M. (1981). Comparison of procedures for detecting test-item bias with both internal and external ability criteria. *Journal of Educational Statistics*, 6(4), 317–375.
- Skaggs, G., & Lissitz, R. W. (1986). An exploration of the robustness of four test equating models. *Applied Psychological Measurement*, 10(3), 303–317. <https://doi.org/10.1177/014662168601000308>
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement*, 27(4), 361–370. <https://doi.org/10.1111/j.1745-3984.1990.tb00754.x>
- Tan, Ş. (2016). *SPSS ve Excel uygulamalı temel istatistik-1* (1. Baskı). Pegem Akademi.
- Tanguma, J. (2000). Equating test scores using the linear method: A primer. *Paper Presented at the Annual Meeting of the Southwest Educational Research Association*. <https://files.eric.ed.gov/fulltext/ED449169.pdf>
- Turgut, F. M., & Baykul, Y. (2015). *Eğitimde ölçme ve değerlendirme* (7. Baskı). Pegem Akademi.
- Turhan, A. (2006). *Multilevel 2PL item response model vertical equating with the presence of differential item functioning* (Doctoral Dissertations, Florida State University). <https://diginole.lib.fsu.edu/islandora/object/fsu%3A175995/>
- van Der Linden, W. J. (2000). A test-theoretic approach to observed-score equating. *Psychometrika*, 65(4), 437–456. <https://doi.org/10.1007/BF02296337>
- van Der Linden, W. J. (2006). Equating error in observed-score equating. *Applied Psychological Measurement*, 30(5), 355–378. <https://doi.org/10.1177/0146621606289948>

- von Davier, A. A. (2011). A statistical perspective on equating test scores. In A. A. Von Davier (Ed.), *Statistical models for test equating, scaling, and linking* (pp. 1–17). Springer.
- von Davier, A. A., Holland, P. W., & Thayer, D. T. (2004). The chain and post-stratification methods for observed-score equating: Their relationship to population invariance. *Journal of Educational Measurement*, *41*(1), 15–32. <https://doi.org/10.1111/j.1745-3984.2004.tb01156.x>
- Von Davier, A. A., Holland, P. W., & Thayer, D. T. (2004). *The kernel method of test equating*. Springer Science & Business Media.
- Wang, T., Lee, W. C., Brennan, R. L., & Kolen, M. J. (2008). A comparison of the frequency estimation and chained equipercentile methods under the common-item nonequivalent groups design. *Applied Psychological Measurement*, *32*(8), 632–651. <https://doi.org/10.1177/0146621608314943>
- Wiberg, M. (2007). *Measuring and detecting differential item functioning in criterionreferenced licensing test* (EM No 60). Umea University.
- Woldbeck, T. (1998). Basic concepts in modern methods of test equating. *Paper Presented at the Annual Meeting of the Southwest Psychological Association*. <https://files.eric.ed.gov/fulltext/ED417215.pdf>
- Yurdugül, H., & Aşkar, P. (2004). Ortaöğretim kurumları öğrenci seçme ve yerleştirme sınavı'nın cinsiyete göre madde yanlılığı açısından incelenmesi. *Eğitim Bilimleri ve Uygulama*, *3*(5), 3–20.
- Yurtçu, M., & Güzeller, C. O. (2017). Investigation of equating error in tests with differential item functioning. *International Journal of Assessment Tools in Education*, *5*(1), 50–57. <https://doi.org/10.21449/ijate.316420>
- Zeng, L. (1991). Standard errors of linear equating for the single-group design. *ACT Research Report Series 91-4*, Inc. <https://files.eric.ed.gov/fulltext/ED344944.pdf>
- Zhu, W. (1998). Test equating: What, why, how? *Research Quarterly for Exercise and Sport*, *69*(1), 11–23. <https://doi.org/10.1080/02701367.1998.10607662>
- Zieky, M. (1993). Practical questions in the use of DIF statistics in test development. In P. W. Holland & H. Wainer (Eds.), *Differential Item*

Functioning (pp. 337–347). Lawrence Erlbaum Associates.

Zumbo, B. D. (1999). Logistic regression modeling as a unitary framework for binary and likert-type (ordinal) item scores. *A Handbook on the Theory and Methods of Differential Item Functioning (DIF)*.

Zwick, R., & Ercikan, K. (1989). Analysis of Differential Item Functioning in the NAEP History Assessment. *Journal of Educational Measurement*, 26(1), 55–66. <https://doi.org/10.1111/j.1745-3984.1989.tb00318.x>

EK-A: PISA 2018 18. Kitapçık Fen Okuryazarlığı Ham Puanlarının Doğrusal ve Doğrusal Olmayan Eşitleme Yöntemleriyle Elde Edilmiş 24. Kitapçıkta Eşitlenmiş Puanları

18. Kitapçık Puanı	24. Kitapçıkta Eşdeğer Puan					
	Doğrusal Eşitleme Yöntemleri			Doğrusal Olmayan Eşitleme Yöntemleri		
	Tucker	Levine	Zincir	Frekans Kestirim	Zincir Eşit Yüzdelikli	Daire-Yay Levine
0	0,53	0,45	0,49	-0,5	-0,5	0
1	1,54	1,45	1,5	-0,5	-0,5	1,06
2	2,54	2,46	2,5	1,5	1,5	2,12
3	3,54	3,47	3,51	2,51	2,6	3,18
4	4,55	4,47	4,51	3,92	3,96	4,23
5	5,55	5,48	5,52	5,03	5,08	5,28
6	6,56	6,48	6,52	6,13	6,09	6,32
7	7,56	7,49	7,53	7,16	7,17	7,36
8	8,56	8,49	8,53	8,37	8,42	8,4
9	9,57	9,5	9,53	9,55	9,62	9,43
10	10,57	10,51	10,54	10,61	10,62	10,46
11	11,57	11,51	11,54	11,6	11,59	11,48
12	12,58	12,52	12,55	12,53	12,51	12,51
13	13,58	13,52	13,55	13,53	13,47	13,52
14	14,59	14,53	14,56	14,63	14,51	14,54
15	15,59	15,53	15,56	15,7	15,61	15,55
16	16,59	16,54	16,57	16,77	16,72	16,56
17	17,6	17,54	17,57	17,86	17,89	17,56
18	18,6	18,55	18,58	18,82	18,9	18,56
19	19,61	19,56	19,58	19,75	19,72	19,56
20	20,61	20,56	20,59	20,8	20,6	20,55
21	21,61	21,57	21,59	21,81	21,64	21,54
22	22,62	22,57	22,6	22,66	22,56	22,52
23	23,62	23,58	23,6	23,53	23,48	23,51
24	24,62	24,58	24,6	24,59	24,65	24,48
25	25,63	25,59	25,61	25,64	25,73	25,46
26	26,63	26,6	26,61	26,61	26,69	26,43
27	27,64	27,6	27,62	27,54	27,57	27,4
28	28,64	28,61	28,62	28,35	28,33	28,36
29	29,64	29,61	29,63	29,46	29,43	29,32
30	30,65	30,62	30,63	30,43	30,43	30,28
31	31,65	31,62	31,64	31,48	31,47	31,23
32	32,66	32,63	32,64	32,36	32,28	32,18
33	33,66	33,63	33,65	33,32	33,22	33,12
34	34,66	34,64	34,65	34,39	34,37	34,06
35	35,67	35,65	35,66	35,37	35,37	35

**EK-B: MH Analizine Göre Dil Kaynaklı DMF Gösteren Ortak Maddelerden
Arındırılarak Elde Edilen Eşitlenmiş Puanlar**

18. Kitapçık Puanı	Eşitleme Yöntemleri ile Elde Edilmiş 24. Kitapçıktaki Eşdeğer Puan					
	Doğrusal Eşitleme Yöntemleri			Doğrusal Olmayan Eşitleme Yöntemleri		
	Tucker	Levine	Zincir	Frekans Kestirim	Zincir Eşit Yüzdelikli	Daire-Yay Levine
0	0,48	0,31	0,4	-0,5	-0,5	0
1	1,48	1,32	1,41	-0,5	-0,5	1,06
2	2,49	2,33	2,41	1,8	1,85	2,11
3	3,49	3,34	3,42	2,99	3,23	3,16
4	4,49	4,35	4,43	3,96	4	4,21
5	5,5	5,36	5,43	5,08	4,93	5,25
6	6,5	6,37	6,44	6,31	6,37	6,29
7	7,51	7,38	7,45	7,22	7,3	7,33
8	8,51	8,39	8,45	8,37	8,41	8,36
9	9,52	9,4	9,46	9,41	9,4	9,39
10	10,52	10,41	10,47	10,5	10,41	10,41
11	11,53	11,41	11,47	11,62	11,51	11,43
12	12,53	12,42	12,48	12,7	12,55	12,44
13	13,54	13,43	13,49	13,71	13,61	13,45
14	14,54	14,44	14,49	14,67	14,63	14,46
15	15,54	15,45	15,5	15,84	15,79	15,47
16	16,55	16,46	16,51	16,85	16,74	16,47
17	17,55	17,47	17,51	17,76	17,59	17,46
18	18,56	18,48	18,52	18,69	18,57	18,45
19	19,56	19,49	19,53	19,62	19,54	19,44
20	20,57	20,5	20,53	20,53	20,48	20,43
21	21,57	21,51	21,54	21,55	21,54	21,41
22	22,58	22,52	22,55	22,53	22,54	22,39
23	23,58	23,53	23,55	23,49	23,55	23,36
24	24,58	24,53	24,56	24,41	24,5	24,33
25	25,59	25,54	25,57	25,4	25,42	25,29
26	26,59	26,55	26,57	26,39	26,34	26,25
27	27,6	27,56	27,58	27,37	27,31	27,21
28	28,6	28,57	28,59	28,33	28,26	28,16
29	29,61	29,58	29,59	29,36	29,32	29,11
30	30,61	30,59	30,6	30,36	30,35	30,06
31	31,62	31,6	31,61	31,37	31,37	31

EK-C: SIBTEST Analizine Göre Dil Kaynaklı DMF Gösteren Ortak Maddelerden Arındırılarak Elde Edilen Eşitlenmiş Puanlar

18. Kitapçık Puanı	Eşitleme Yöntemleri ile Elde Edilmiş 24. Kitapçıktaki Eşdeğer Puan					
	Doğrusal Eşitleme Yöntemleri			Doğrusal Olmayan Eşitleme Yöntemleri		
	Tucker	Levine	Zincir	Frekans Kestirim	Zincir Eşit Yüzdellikli	Daire-Yay Levine
0	0,49	0,35	0,42	-0,5	-0,5	0
1	1,5	1,35	1,43	-0,5	-0,5	1,06
2	2,5	2,36	2,44	1,68	1,84	2,11
3	3,5	3,37	3,44	2,77	2,79	3,17
4	4,51	4,38	4,45	3,96	3,97	4,21
5	5,51	5,39	5,45	5,07	5,08	5,26
6	6,52	6,39	6,46	6,24	6,13	6,3
7	7,52	7,4	7,46	7,32	7,37	7,33
8	8,52	8,41	8,47	8,32	8,43	8,36
9	9,53	9,42	9,48	9,53	9,58	9,39
10	10,53	10,42	10,48	10,37	10,39	10,42
11	11,54	11,43	11,49	11,49	11,42	11,44
12	12,54	12,44	12,49	12,63	12,51	12,46
13	13,54	13,45	13,5	13,68	13,46	13,47
14	14,55	14,46	14,5	14,7	14,53	14,48
15	15,55	15,46	15,51	15,73	15,68	15,48
16	16,56	16,47	16,52	16,84	16,83	16,49
17	17,56	17,48	17,52	17,84	17,77	17,48
18	18,56	18,49	18,53	18,75	18,61	18,48
19	19,57	19,5	19,53	19,69	19,59	19,47
20	20,57	20,5	20,54	20,61	20,54	20,46
21	21,58	21,51	21,54	21,55	21,5	21,44
22	22,58	22,52	22,55	22,56	22,54	22,42
23	23,58	23,53	23,56	23,54	23,55	23,39
24	24,59	24,53	24,56	24,5	24,56	24,36
25	25,59	25,54	25,57	25,42	25,51	25,33
26	26,6	26,55	26,57	26,4	26,42	26,3
27	27,6	27,56	27,58	27,4	27,35	27,26
28	28,6	28,57	28,58	28,38	28,31	28,21
29	29,61	29,57	29,59	29,33	29,26	29,17
30	30,61	30,58	30,6	30,36	30,32	30,11
31	31,61	31,59	31,6	31,36	31,35	31,06
32	32,62	32,6	32,61	32,37	32,37	32

EK-Ç: MH Analizine Göre Cinsiyet Kaynaklı DMF Gösteren Ortak Maddelerden Arındırılarak Elde Edilen Eşitlenmiş Puanlar

18. Kitapçık Puanı	Eşitleme Yöntemleri ile Elde Edilmiş 24. Kitapçıktaki Eşdeğer Puan					
	Doğrusal Eşitleme Yöntemleri			Doğrusal Olmayan Eşitleme Yöntemleri		
	Tucker	Levine	Zincir	Frekans Kestirim	Zincir Eşit Yüzdelikli	Daire-Yay Levine
0	0,5	0,36	0,43	-0,5	-0,5	0
1	1,5	1,37	1,44	-0,23	-0,23	1,06
2	2,51	2,38	2,45	1,7	1,63	2,12
3	3,51	3,39	3,46	2,87	2,89	3,18
4	4,52	4,4	4,46	3,95	4,11	4,23
5	5,52	5,41	5,47	5,08	5,12	5,28
6	6,53	6,42	6,48	6,29	6,28	6,32
7	7,54	7,43	7,49	7,51	7,5	7,36
8	8,54	8,44	8,49	8,5	8,5	8,39
9	9,55	9,45	9,5	9,56	9,57	9,42
10	10,55	10,46	10,51	10,6	10,64	10,45
11	11,56	11,47	11,52	11,56	11,51	11,47
12	12,56	12,48	12,52	12,59	12,4	12,49
13	13,57	13,49	13,53	13,69	13,56	13,51
14	14,57	14,5	14,54	14,7	14,65	14,52
15	15,58	15,51	15,55	15,77	15,75	15,52
16	16,59	16,52	16,55	16,76	16,74	16,53
17	17,59	17,53	17,56	17,74	17,67	17,52
18	18,6	18,54	18,57	18,79	18,69	18,52
19	19,6	19,55	19,58	19,71	19,57	19,51
20	20,61	20,56	20,58	20,6	20,54	20,49
21	21,61	21,57	21,59	21,66	21,71	21,47
22	22,62	22,57	22,6	22,64	22,7	22,45
23	23,62	23,58	23,6	23,58	23,67	23,42
24	24,63	24,59	24,61	24,53	24,57	24,39
25	25,64	25,6	25,62	25,42	25,41	25,36
26	26,64	26,61	26,63	26,42	26,39	26,32
27	27,65	27,62	27,63	27,37	27,33	27,28
28	28,65	28,63	28,64	28,46	28,43	28,23
29	29,66	29,64	29,65	29,36	29,35	29,18
30	30,66	30,65	30,66	30,34	30,34	30,12
31	31,67	31,66	31,66	31,44	31,44	31,06
32	32,67	32,67	32,67	32,39	32,39	32

EK-D: SIBTEST Analizine Göre Cinsiyet Kaynaklı DMF Gösteren Ortak Maddelerden Arındırılarak Elde Edilen Eşitlenmiş Puanlar

18. Kitapçık Puanı	Eşitleme Yöntemleri ile Elde Edilmiş 24. Kitapçıktaki Eşdeğer Puan					
	Doğrusal Eşitleme Yöntemleri			Doğrusal Olmayan Eşitleme Yöntemleri		
	Tucker	Levine	Zincir	Frekans Kestirim	Zincir Eşit Yüzdelikli	Daire-Yay Levine
0	0,49	0,3	0,4	-0,5	-0,5	0
1	1,49	1,31	1,41	-0,5	-0,5	1,06
2	2,49	2,32	2,41	1,71	1,73	2,11
3	3,5	3,33	3,42	2,91	2,79	3,16
4	4,5	4,34	4,43	4,27	4,21	4,21
5	5,51	5,35	5,43	5,14	5,22	5,25
6	6,51	6,36	6,44	6,19	6,17	6,29
7	7,52	7,37	7,45	7,36	7,36	7,33
8	8,52	8,38	8,45	8,54	8,62	8,36
9	9,52	9,39	9,46	9,51	9,55	9,38
10	10,53	10,4	10,47	10,52	10,49	10,41
11	11,53	11,41	11,48	11,5	11,38	11,43
12	12,54	12,42	12,48	12,56	12,29	12,44
13	13,54	13,43	13,49	13,66	13,52	13,45
14	14,55	14,44	14,5	14,73	14,69	14,46
15	15,55	15,45	15,5	15,7	15,65	15,47
16	16,55	16,46	16,51	16,71	16,61	16,47
17	17,56	17,47	17,52	17,78	17,64	17,46
18	18,56	18,48	18,52	18,71	18,59	18,45
19	19,57	19,49	19,53	19,6	19,47	19,44
20	20,57	20,5	20,54	20,6	20,54	20,43
21	21,58	21,51	21,54	21,66	21,72	21,41
22	22,58	22,52	22,55	22,51	22,63	22,38
23	23,58	23,53	23,56	23,49	23,57	23,36
24	24,59	24,54	24,56	24,41	24,45	24,33
25	25,59	25,55	25,57	25,4	25,4	25,29
26	26,6	26,56	26,58	26,37	26,31	26,25
27	27,6	27,56	27,58	27,46	27,37	27,21
28	28,61	28,57	28,59	28,28	28,19	28,16
29	29,61	29,58	29,6	29,33	29,25	29,11
30	30,62	30,59	30,6	30,39	30,38	30,06
31	31,62	31,6	31,61	31,37	31,37	31

EK-E: Etik Komisyonu Onay Bildirimi



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

F46

05 / 05 / 2021

Hacettepe Üniversitesi
Eğitim Bilimleri Enstitüsü
Eğitim Bilimleri Ana Bilim Dalı Başkanlığına

Tez/Araştırma Başlığı	Değişen Madde Fonksiyonu Gösteren Ortak Maddelerin Test Eşitlemeye 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:

(X) Kamusal erişime açık (buraya yazınız): PISA 2018 bilişsel test verileri

() Özel izin ve onaya tabi (buraya yazınız):

() Üretilmiş veri (buraya yazınız):

() Diğer (buraya yazınız):

Yükseköğretim Kurumları Etik Kurulları ve Komisyonlarının Yönergelerini inceledim ve bunlara göre çalışmamın yürütülebilmesi için herhangi bir Etik Komisyondan/Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.

Gereğini saygılarımla arz ederim.

Feyzi GÜNEŞ
(Araştırmacı Adı Soyadı, İmzası)

Araştırmacı Bilgileri

Adı Soyadı	Feyzi GÜNEŞ
Öğrenci ise No	N19134741
Ana Bilim Dalı	Eğitim Bilimleri
Programı	Eğitimde Ölçme ve Değerlendirme
Statüsü	<input checked="" type="checkbox"/> Yüksek Lisans <input type="checkbox"/> Doktora <input type="checkbox"/> Bütünleşik Dr. <input type="checkbox"/> Diğer

Danışman Görüşü ve Onayı*

Tezde PISA 2018 verileri kullanılmıştır. Veriler OECD'nin sitesinden indirilmiştir.

Prof. Dr. Hülya KELECİOĞLU
(İmza)
(Danışmanın İmzası, Adı ve Soyadı)

*Tez ve tezden üretilen yayınlarda gerekli

EK-F: 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.

02/06/2021

Feyzi GÜNEŞ

EK-G: Yüksek Lisans Tez Çalışması Orijinallik Raporu

02/06/2021

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

Tez Başlığı: Değişen Madde Fonksiyonu Gösteren Ortak Maddelerin 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
02/06/2021	70	116.210	28/05/2021	%4	1598956005

Uygulanan filtreler:

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

Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Tez Çalışması Orijinallik Raporu Alınması ve Kullanılması Uygulama Esasları'nı 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: Feyzi GÜNEŞ

Öğrenci No.: N19134741

Ana Bilim Dalı: Eğitim Bilimleri

Programı: Eğitimde Ölçme ve Değerlendirme

Statüsü: Y.Lisans Doktora Bütünleşik Dr.

İmza

DANIŞMAN ONAYI

UYGUNDUR.

(İmza)

Prof. Dr. Hülya KELECİOĞLU

EK-H: Thesis Originality Report

02/06/2021

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

Thesis Title: The Study of Effect of Anchor Items Showing Differential Item Functioning on 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
02/06/2021	70	116.210	28/05/2021	%4	1598956005

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: Feyzi GÜNEŞ
Student No.: N19134741
Department: Educational Sciences
Program: Educational Measurement and Evaluation
Status: Masters Ph.D. Integrated Ph.D.

Signature

ADVISOR APPROVAL

APPROVED
(Signature)

Prof. Dr. Hülya KELECİOĞLU

EK-I: Yayınlama ve Fikrî Mülkiyet Hakları Beyanı

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

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

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

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

02/06/2021

Feyzi GÜNEŞ

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

(1) Madde 6. 1. Lisansüstü tezle 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 tezin 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 tezin erişime açılması engellenebilir.

(3) Madde 7. 1. Ulusal çıkarları veya güvenliği ilgilendiren, emniyet, istihbarat, savunma ve güvenlik, sağlık vb. konulara ilişkin lisansüstü tezlerle ilgili gizlilik kararı, tezin yapıldığı kurum tarafından verilir*. Kurum ve kuruluşlarla yapılan işbirliği protokolü çerçevesinde hazırlanan lisansüstü tezlere ilişkin gizlilik kararı ise, ilgili kurum ve kuruluşun önerisi ile enstitü veya fakültenin uygun görüşü üzerine üniversite yönetim kurulu tarafından verilir. Gizlilik kararı verilen tezler Yükseköğretim Kuruluna bildirilir.

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

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

