



HACETTEPE ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

Eğitim Bilimleri Ana Bilim Dalı

Eğitimde Ölçme ve Değerlendirme Programı

BİLİŞSEL TANI MODELLERİ İLE ELDE EDİLEN BİREYSEL BECERİ PUANLARININ BENZERLİĞİNİN PARALEL FORMLARLA İNCELENMESİ

Melek ERDOĞAN

Doktora Tezi

Ankara, 2024

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

BİLİŞSEL TANI MODELLERİ İLE ELDE EDİLEN BİREYSEL BECERİ PUANLARININ
BENZERLİĞİNİN PARALEL FORMLARLA İNCELENMESİ

EXAMINING THE COMPARABILITY OF INDIVIDUAL SKILL SCORES OBTAINED
THROUGH COGNITIVE DIAGNOSTIC MODELS WITH PARALLEL FORMS

Melek ERDOĞAN

Doktora Tezi

Ankara, 2024

Kabul ve Onay

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼ne,

Melek ERDOđAN'ın hazırladıđı “Bilişsel Tanı Modelleri ile Elde Edilen Bireysel Beceri Puanlarının Benzerliđinin Paralel Formlarla İncelenmesi” bařlıklı bu alıřma j¼rimiz tarafından **Eđitim Bilimleri Ana Bilim Dalı, Eđitimde Ölme ve Deđerlendirme Bilim Dalında Doktora Tezi** olarak kabul edilmiřtir.

J¼ri Bařkanı	Prof. Dr. SelahattinGELBAL	İmza
J¼ri Üyesi (Danıřman)	Prof. Dr.BurcuATAR	İmza
J¼ri Üyesi	Do. Dr. Derya OBANOđLU	İmza
J¼ri Üyesi	Do. Dr. Lokman AKBAY	İmza
J¼ri Üyesi	Do. Dr. Melek G¼lřah řAHİN	İmza

Bu tez Hacettepe Üniversitesi Lisans¼st¼ Eđitim, Öğretim ve Sınav Yönetmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri üyeleri tarafından 29 / 01 / 2024 tarihinde uygun gör¼lm¼ř ve Enstit¼ Yönetim Kurulunca / /tarihi itibarıyla kabul edilmiřtir.

Prof. Dr. İsmail Hakkı MİRİCİ
Eđitim Bilimleri Enstitüsü M¼d¼r¼

Öz

Tanılayıcı değerlendirme öğrencilerin güçlü ve zayıf yönlerini belirleme ve öğrenme gereksinimlerini tespit etmede etkilidir. Bilişsel Tanı Modelleri (BTM) ölçme hedefi çerçevesinde nitelik olarak belirlenen kazanım veya beceriler ile ilgili geri bildirim sağlayan önemli bir psikometrik modeldir. Test eşitleme, farklı test formlarının puanlarını aynı ölçekte birleştirerek, bu puanların birbirinin yerine geçebildiğini ve aynı anlama geldiğini ifade eder. Bu çalışmanın amacı, BTM çerçevesinde oluşturulan test formlarının karma hiyerarşik yapıları GDINA model analizleri ile elde edilen yetkinlik düzeyi olasılık değerlerinin beceri puanlarına çevrilerek geleneksel yöntemlerle test eşitleme uygulamaları ile testler arası karşılaştırılabilir beceri puanlarının elde edilebilirliğinin incelenmesidir. Bu amaç çerçevesinde altı cebirsel düşünme becerisini (genelleştirilmiş aritmetik, fonksiyonel düşünme, niceliksel mantık yürütme, temsilleştirme, genelleme ve mantık yürütme, gerekçelendirme) ölçmek üzere hazırlanan 20 maddelik çoktan seçmeli tanılayıcı üç testin (Test A, Test B, Test C) toplamda 2600 lise öğrencisine uygulanarak elde edilen cevap örüntüsü kullanılmıştır. Bu cevap örüntüsüne uyum sağlayan karma hiyerarşik yapıları GDINA model analizleri ile elde edilen beceri yetkinlik düzey olasılıkları beceri puanlarına çevrilmiştir. Testler arası beceri puanlarının eşitlenebilirlik durumu genelleştirilmiş aritmetik, fonksiyonel düşünme ve niceliksel mantık yürütme becerileri ile sınırlandırılarak denk gruplar (DG), denk olmayan gruplarda ortak madde (DOG-OM) ve denk olmayan gruplarda ortak değişken (DOG-OD) desenleri çerçevesinde ortalama, doğrusal, eşit yüzdelikli ve kernel eşitleme uygulamalarında farklı metodlarla incelenmiştir. Eşitlemelere ait eşitlenmiş puan farkları, eşitleme hataları ve yanlılık gibi istatistiksel bilgiler paylaşılmıştır. Yapılan inceleme bulgularına göre grup momentleri anlamlı bir farklılık göstermeyen ve puan dağılım biçimleri benzer olan gruplarda eşitlemenin yapılabildiği görülmektedir.

Anahtar sözcükler: bilişsel tanı modelleri, cebirsel düşünme becerileri, GDINA model, karma hiyerarşik yapıları GDINA model, test eşitleme

Abstract

Diagnostic assessment identifies students' strengths and weaknesses, and learning needs. Cognitive Diagnostic Models (CDM) provide feedback on objectives or skills identified as attributes within the framework of measurement aims. Test equating refers to the process of combining scores from different test forms on the same scale, indicating that these scores can be used interchangeably and convey the same meaning. The aim of this study is to investigate the feasibility of obtaining comparable skill scores between tests through traditional test equating practices by converting mastery level probability values obtained through mixed hierarchically structured GDINA model analyses of test forms created within the framework of Cognitive Diagnostic Models (CDMs). In this context, three 20-item multiple-choice diagnostic tests constructed to measure six algebraic thinking abilities were administered to a total of 2600 high school students, and the response patterns obtained were used. Mastery level probabilities obtained through mixed hierarchically structured GDINA model analyses that fit the response patterns were converted into ability scores. Equating ability scores between tests was examined through various methods, including mean, linear, equipercentile, and kernel equating, by restricting the equating to generalized arithmetic, quantitative reasoning, and functional thinking abilities, within the frameworks of equivalent groups, non-equivalent groups with anchor test, and non-equivalent groups with covariates designs. Statistical information such as equated score differences, equating errors, and bias were shared. According to the findings of the study, it is observed that equating can be performed in groups showing no significant differences in group moments and having similar score distribution patterns.

Keywords: cognitive diagnostic models, algebraic thinking abilities, GDINA model, mixed hierarchically structured GDINA model, test equating

Teşekkür

Bu tez çalışmamın başarıyla tamamlanmasında katkısı olan herkese içten teşekkürlerimi sunmak isterim.

Öncelikle, tez danışmanım Prof. Dr. Burcu Atar Hocamın benim için kıymetini ifade etmek isterim. Bu çalışmanın temelleri atılan ODTÜ'deki çalışmalardan bu yana rehberliği, yönlendirmeleri ve sabrı olmadan şimdiki çalışma hayata geçmezdi. Tez jürisinde yer alan kıymetli akademisyenler, Prof. Dr. Selahattin Gelbal, Doç. Dr. Derya Çobanoğlu, Doç. Dr. Lokman Akbay ve Doç. Dr. Melek Gülşah Şahin Hocalarıma içtenlikle teşekkür ederim. Zamanlarını ayırarak tezimi değerlendirmeleri ve yol göstermeleri benim için büyük bir değer taşıyor.

Bu tezde kullanılan test formlarının geliştirilmesi aşamasında akademik destek sağlayan değerli hocalarıma da teşekkürümü sunmak isterim. Prof. Dr. Ayhan Kürşat Erbaş, Doç. Dr. Bülent Çetinkaya ve Dr. Murat Kol Hocalarımla desteği ile bu tez sağlam bir temele dayandı. Yine bu tezin temellerini oluşturan konularda matematik eğitimi çerçevesinde katkılar sunan ve beni daha geniş açılardan bakmaya yönlendiren kıymetli Prof. Dr. Erdinç Çakıroğlu Hocama da katkılarından dolayı teşekkür ederim. Ayrıca, bana hem ODTÜ'deki doktora çalışmamda hem de buradaki çalışmamda zaman ayıran Doç. Dr. Lokman Akbay Hocama benim için kullandığı kıymetli vakitleri için teşekkür ederim. Test formlarının geliştirilmesinde ÖSYM'deki deneyimleri ile katkı sağlayan meslektaşlarıma ve matematik eğitimi alanı olsun, eğitimde ölçme ve değerlendirme alanı olsun çeşitli üniversitelerde görev yapan tüm arkadaşlarımla test geliştirme aşamasındaki emeklerine karşı da teşekkürlerimi sunarım.

Araştırmanın uygulama sürecinde katkı sağlayan tüm öğrencilere, öğretmenlere ve okul müdürlerine teşekkür ederim. Onların yardımlarıyla bu çalışma daha zengin ve kapsamlı hale geldi.

Son olarak Buğlem ve Yaren kızlarıma, anneme, babama, kardeşlerime ve tüm yakınlarıma içtenlikle teşekkür ederim. Onların sonsuz desteği ve sevgisi bu yolculuğun her anında benimle oldu.

Bütün geçen bu koskaca uzun yıllar içerisinde her doktora öğrencisi gibi ben de çeşitli duygu durumları yaşadım. Bir taraftan öğrenci olmanın sorumluluğu yanında bir anne, bir evlat, bir kardeş, bir komşu, bir vatandaş, bir devlet memuru ve ruhindaki eğitimci olma hissi ile topluma ve dünyaya olan görev ve sorumluluklarım dahilinde stresimi yönettiğim anlar olduğu kadar yönetemediğim zamanlar da fazlaca oldu. Bazen koştum, bazen yoruldum, bazen durdum, bazen de harekete geçmek için kendimi zorladım, bazen yol bulamadım derken yeni yollarla karşılaştım, fakat çoğu zaman bu tezin tamamlanma aşamalarının yani içinde bulunduğum bugünlerin heyecanını duydum. Bu süreç beni sadece akademik gelişim anlamında değil yaşamın anlamını fark etme yönünden de şekillendiren ve ilerleten bir süreç oldu.

Bilimin ışığında yürüyen bizler için ülkemizin çağdaş medeniyetler seviyesinde olması hedefiyle eğitim ve öğretime akademik temeldeki katkılarımızın ve yaşamımızın güzelliklerini zenginleştirmek yolunda ilerleyişlerimizin hep devam etmesini dilerim.

İçindekiler

Kabul ve Onay.....	ii
Öz.....	iii
Abstract.....	iv
Teşekkür.....	v
Tablolar Dizini.....	x
Şekiller Dizini.....	xiv
Simgeler ve Kısaltmalar Dizini.....	xvii
Bölüm 1 Giriş.....	1
Problem Durumu.....	1
Araştırmanın Amacı ve Önemi.....	5
Araştırma Problemi.....	7
Sayıtlar.....	9
Sınırlılıklar.....	9
Tanımlar.....	9
Bölüm 2 Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar.....	11
Bilişsel Tanı Modelleri.....	11
GDINA Modeli.....	19
Bilişsel Tanı Modelleri Uygulamasına Yönelik Yapılan Araştırmalar.....	23
Test Eşitleme.....	29
Eşitlemede Veri Toplama Desenleri.....	32
Klasik Test Kuramına Dayalı Eşitleme Yöntemleri.....	35
Test Eşitleme Desenlerine Göre Varsayımlar.....	39
Test Eşitleme Desenlerine Göre KTK'ya Dayalı Eşitleme Yöntemleri.....	40
Kernel Eşitleme.....	48
Test Eşitleme ile İlgili Araştırmalar.....	51
Bölüm 3 Yöntem.....	62

Araştırmanın Türü	62
Çalışma Grubu	62
Veri Toplama Süreci.....	65
Veri Toplama Araçları	65
Verilerin Analizi	76
Çalışma Grubuna Ait Betimleyici Bilgiler	94
Bölüm 4 Bulgular, Yorum ve Tartışma.....	105
Birinci Araştırma Sorusu Kapsamındaki Bulgular.....	105
Genelleştirilmiş Aritmetik (GA) Becerisinin Denk Grup (DG) Desenine Göre Eşitleme İncelemeleri.....	105
GA Becerisinin Denk Olmayan Gruplarda Ortak Madde (DOG-OM) Desenine Göre Eşitleme İncelemeleri.....	116
GA Becerisinin Denk Olmayan Gruplarda Ortak Değişken (DOG-OD) Desenine Göre Eşitleme İncelemeleri.....	126
İkinci Araştırma Sorusu Kapsamındaki Bulgular	134
Fonksiyonel Düşünme (FD) Becerisinin Denk Grup (DG) Desenine Göre Eşitleme İncelemeleri.....	134
FD Becerisinin Denk Olmayan Gruplarda Ortak Madde (DOG-OM) Desenine Göre Eşitleme İncelemeleri.....	144
FD Becerisinin Denk Olmayan Gruplarda Ortak Değişken (DOG-OD) Desenine Göre Eşitleme İncelemeleri.....	156
Üçüncü Araştırma Sorusu Kapsamındaki Bulgular	164
Niceliksel Mantık Yürütme (NMY) Becerisinin DG, DOG-OM ve DOG-OD Desenlerine Göre Eşitleme İncelemeleri.....	164
Yorum ve Tartışma.....	166
Bölüm 5 Sonuç ve Öneriler.....	179
Araştırmacılara Yönelik Öneriler	180
Uygulayıcılara Yönelik Öneriler.....	182
Kaynaklar	184

EK-A: Niceliksel Mantık Yürütme (NMY) Becerisinin Denk Grup (DG) Desenine Göre Eşitleme İncelemeleri	203
EK-B: NMY Becerisinin Denk Olmayan Gruplarda Ortak Madde (DOG-OM) Desenine Göre Eşitleme İncelemeleri	207
EK-C: NMY Becerisinin Denk Olmayan Gruplarda Ortak Değişken (DOG-OD) Desenine Göre Eşitleme İncelemeleri	212
EK-D:Araştırma Etik Komisyon İzin Muafiyeti Formu/Araştırma Etik Komisyonu Onay Bildirimi	216
EK-E: MEB Araştırma İzni Onay Bildirimi	217
EK-F: Etik Beyanı	218
EK-G: Doktora Tez Çalışması Orijinallik Raporu	219
EK-H: Thesis/Dissertation Originality Report.....	220
EK-I: Yayımlama ve Fikrî Mülkiyet Hakları Beyanı	221

Tablolar Dizini

Tablo 1 <i>Madde-Nitelik Dizilimi Örneği</i>	16
Tablo 2 <i>BTM'nin Sınıflandırılması (Rupp ve diğerleri, 2010, s.98)</i>	17
Tablo 3 <i>Okul Türlerine ve Kabul Koşullarına Göre Katılımcı Sayısı</i>	63
Tablo 4 <i>Testlerdeki Okul Türlerine Göre Katılımcı Sayısı</i>	64
Tablo 5 <i>Okulların Kabul Koşullarına Göre Testlerdeki Katılımcı Sayısı</i>	64
Tablo 6 <i>Her Teste Göre Cinsiyet Bilgisi</i>	65
Tablo 7 <i>Ortaokul Cebir Programı Konu Kazanım ve Süre Bilgisi</i>	67
Tablo 8 <i>Konu ve Niteliklere Göre Madde Dağılımları</i>	68
Tablo 9 <i>Test Formlarına ait Q-matris</i>	78
Tablo 10 <i>Test Formlarında Model Uyumu</i>	83
Tablo 11 <i>Testlere Formlarında Absolute Fit ve Item Fit İndeksleri</i>	84
Tablo 12 <i>Test Formlarının ve Niteliklerin Sınıflandırma Doğruluk Düzeyleri</i>	88
Tablo 13 <i>Eşitleme Bulguları Tablosu</i>	94
Tablo 14 <i>Genel Test Puanlarına Ait Betimsel İstatistikler</i>	96
Tablo 15 <i>Ortak Maddelere Ait Betimsel İstatistikler</i>	97
Tablo 16 <i>KTK'ya göre Testlerin Ortalama Güçlük ve Güvenirlik İndeksleri</i>	98
Tablo 17 <i>KTK'ya Göre Testlerin Madde Güçlük ve Ayırt Edicilik İndeksleri</i>	99
Tablo 18 <i>Beceri Puanlarının Testlere Göre Betimsel İstatistikleri</i>	103
Tablo 19 <i>DG Deseni GA Becerisi Ortalama Eşitleme Puanı</i>	106
Tablo 20 <i>DG Deseni GA Becerisi Doğrusal Eşitleme Puanı</i>	107
Tablo 21 <i>DG Deseni GA Becerisi EYE Puanı</i>	107
Tablo 22 <i>DG Deseni GA Becerisi Eşitlenmiş Puan Bootstrap Hata, Yanlılık ve RMSE Değerleri</i>	110
Tablo 23 <i>DG Deseni GA Becerisi Logistic Kernel EYE ve Doğrusal Eşitleme Puan ve Hataları</i>	110
Tablo 24 <i>DG Deseni GA Becerisi Uniform Kernel EYE ve Doğrusal Eşitleme Puan ve Hataları</i>	112
Tablo 25 <i>DG Deseni GA Becerisi Uniform ve Logistic Kernel Yüzde Göreceli Hata Değerleri</i>	114
Tablo 26 <i>DG Deseninde GA Becerisinin Testlere Göre Eşitlenebilme Durumları</i>	115
Tablo 27 <i>DOG-OM Deseni GA Becerisi Doğrusal Eşitleme Puanları</i>	116

Tablo 28 <i>DOG-OM Deseni GA Becerisi EYE Puanları</i>	118
Tablo 29 <i>DOG-OM Deseni GA Becerisi Eşitlenmiş Puan Bootstrap Hata, Yanlılık ve RMSE Değerleri</i>	121
Tablo 30 <i>DOG-OM Deseni GA Becerisi Kernel PSE EYE ve Doğrusal Eşitleme Puan ve Hataları</i>	121
Tablo 31 <i>DOG-OM Deseni GA Becerisi Kernel PSE ve CE Yüzde Göreceli Hata Değerleri</i>	124
Tablo 32 <i>DOG-OM Deseninde GA Becerisinin Testlere Göre Eşitlenebilme Durumları</i>	125
Tablo 33 <i>DOG-OD Deseni GA Becerisi Kernel EYE ve Doğrusal Eşitleme Puan ve Hataları</i>	126
Tablo 34 <i>DOG-OD Deseni GA Becerisi Kernel PSE EYE Yüzde Göreceli Hata Değerleri</i>	129
Tablo 35 <i>DOG-OD Deseni GA Becerisi Kernel EYE ve Doğrusal Eşitleme Puan ve Hataları</i>	130
Tablo 36 <i>DOG-OD Deseni GA Becerisi Kernel PSE EYE Yüzde Göreceli Hata Değerleri</i>	132
Tablo 37 <i>DOG-OD Deseninde GA Becerisinin Testlere Göre Eşitlenebilme Durumları</i>	133
Tablo 38 <i>DG Deseni FD Becerisi Ortalama Eşitleme Puanı</i>	134
Tablo 39 <i>DG Deseni FD Becerisi Doğrusal Eşitleme Puanı</i>	135
Tablo 40 <i>DG Deseni FD Beceri Puanları EYE Puanı</i>	136
Tablo 41 <i>DG Deseni FD Becerisi Eşitlenmiş Puan Bootstrap Hata, Yanlılık ve RMSE Değerleri</i>	139
Tablo 42 <i>DG Deseni FD Becerisi Logistic Kernel EYE ve Doğrusal Eşitleme Puan ve Hataları</i>	139
Tablo 43 <i>DG Deseni FD Becerisi Uniform ve Logistic Kernel Yüzde Göreceli Hata Değerleri</i>	142
Tablo 44 <i>DG Deseninde FD Becerisinin Testlere Göre Eşitlenme Durumları</i>	143
Tablo 45 <i>DOG-OM Deseni FD Becerisi Doğrusal Eşitleme Puanları</i>	144
Tablo 47 <i>DOG-OM Deseni FD Becerisi EYE Puanları</i>	146
Tablo 48 <i>DOG-OM FD Becerisi Eşitlenmiş Puan Bootstrap Hata, Yanlılık ve RMSE Değerleri</i>	149

Tablo 49 <i>DOG-OM Deseni FD Becerisi PSE EYE ve Doğrusal Eşitleme Puan ve Hataları</i>	150
Tablo 50 <i>DOG-OM Deseni FD Becerisi Kernel CE EYE ve Doğrusal Eşitleme Puan ve Hataları</i>	151
Tablo 51 <i>DOG-OM Deseni FD Becerisi Kernel PSE ve CE Yüzde Göreceli Hata Değerleri</i>	154
Tablo 52 <i>DOG-OM Deseninde FD Becerisinin Testlere Göre Eşitlenme Durumları</i>	155
Tablo 53 <i>DOG-OD Deseni FD Becerisi Kernel EYE ve Doğrusal Eşitleme Puan ve Hataları</i>	156
Tablo 54 <i>DOG-OD Deseni FD Becerisi Kernel PSE EYE Yüzde Göreceli Hata Değerleri</i>	159
Tablo 55 <i>DOG-OD Deseni FD Becerisi Kernel EYE ve Doğrusal Eşitleme Puan ve Hataları</i>	160
Tablo 56 <i>DOG-OD Deseni FD Becerisi Kernel PSE EYE Yüzde Göreceli Hata Değerleri</i>	162
Tablo 57 <i>DOG-OD Deseninde FD Becerisinin Testlere Göre Eşitlenme Durumları</i>	163
Tablo 58 <i>DG Deseninde NMY Becerisinin Testlere Göre Eşitlenme Durumları</i> .	164
Tablo 59 <i>DOG-OM Deseninde NMY Becerisinin Testlere Göre Eşitlenme Durumları</i>	165
Tablo 60 <i>DOG-OD Deseninde NMY Becerisinin Testlere Göre Eşitlenme Durumları</i>	165
Tablo 61 <i>Genelleştirilmiş Aritmetik Beceri Puanlarının Testlere Göre Genel Eşitlenme Durumu</i>	166
Tablo 62 <i>Fonksiyonel Düşünme Beceri Puanlarının Testlere Göre Genel Eşitlenme Durumu</i>	168
Tablo 63 <i>Niceliksel Mantık Yürütme Beceri Puanlarının Testlere Göre Genel Eşitlenme Durumu</i>	170
Tablo 64 <i>DG Deseni NMY Becerisi Ortalama Eşitleme Puanı</i>	203
Tablo 65 <i>DG Deseni NMY Becerisi Eşitlenmiş Puan Bootstrap Hata, Yanlılık ve RMSE Değerleri</i>	204
Tablo 66 <i>DG Deseni NMY Becerisi Logistic Kernel EYE ve Doğrusal Eşitleme Puan ve Hataları</i>	205

Tablo 67 DG Deseni NMY Becerisi Uniform ve Logistic Kernel Yüzde Göreceli Hata Değerleri	206
Tablo 68 DOG-OM Deseni NMY Becerisi Doğrusal Eşitleme Puanları.....	207
Tablo 69 DOG-OM Deseni NMY Becerisi Levine OS Doğrusal Eşitleme Puanı	207
Tablo 70 DOG-OM Deseni NMY Becerisi EYE ve CE Puanları.....	208
Tablo 71 DOG-OM Deseni GA Becerisi Eşitlenmiş Puan Bootstrap Hata, Yanlılık ve RMSE Değerleri.....	209
Tablo 72 DOG-OM Deseni NMY Becerisi Logistic Kernel EYE ve Doğrusal Eşitleme Puan ve Hataları	209
Tablo 73 DOG-OM Deseni NMY Becerisi Kernel PSE ve CE Yüzde Göreceli Hata Değerleri.....	211
Tablo 74 DOG-OD Deseni NMY Becerisi Kernel EYE ve Doğrusal Eşitleme Puan ve Hataları	212
Tablo 75 DOG-OD Deseni NMY Becerisi Kernel PSE EYE Yüzde Göreceli Hata Değerleri.....	213
Tablo 76 DOG-OD Deseni NMY Becerisi Kernel EYE ve Doğrusal Eşitleme Puan ve Hataları	214
Tablo 77 DOG-OD Deseni NMY Becerisi Kernel PSE EYE Yüzde Göreceli Hata Değerleri.....	215

Şekiller Dizini

Şekil 1 <i>GDINA Modeli Örtürk Sınıflama ve Cevaplama Olasılıkları Örneği</i>	22
Şekil 2 <i>Testlere Ait Madde 8</i>	71
Şekil 3 <i>Testlere Ait Madde 10</i>	72
Şekil 4 <i>Testlere Ait Madde 11</i>	73
Şekil 5 <i>Testlere Ait Madde 12</i>	74
Şekil 6 <i>Test A'dan PVAF Değerlerine Göre Mesa Plot Örnekleri</i>	78
Şekil 7 <i>Sadece Birinci Faktörün Hiyerarşik Yapılı Olma Durumunun Test Edilmesi</i>	80
Şekil 8 <i>Sadece İkinci Faktörün Hiyerarşik Yapılı Olma Durumunun Test Edilmesi</i>	81
Şekil 9 <i>Her İki Faktörün Hiyerarşik Yapılı Olma Durumunun Test Edilmesi</i>	82
Şekil 10 <i>Her Test Formundaki Madde Çiftleri İçin Isı Haritası</i>	86
Şekil 11 <i>Kişi Uyumunu Gösteren FHCI Değerlerinin Saçılım Grafikleri</i>	87
Şekil 12 <i>Bireylere Ait Test Puanı ve Beceri Yetkinlik Düzeyleri Örneği</i>	89
Şekil 13 <i>Bireysel Beceri Yetkinlik Düzeylerine Karşılık Gelen Beceri Puanları Örneği</i>	91
Şekil 14 <i>Test Puanları Dağılımı</i>	95
Şekil 15 <i>GDINA Modele Göre Yetkin ve Yetkin Olmayan Gruplarda Maddelerin Cevaplanma Olasılıkları</i>	101
Şekil 16 <i>DG Deseni GA Becerisi Eşitlenmiş Puan Fark ve Bootstrap Standart Hataları</i>	108
Şekil 17 <i>DG Deseni GA Becerisi Eşitlenmiş Puan Yanlılık ve RMSE Değerleri.</i>	109
Şekil 18 <i>DG Deseni GA Becerisi Kernel Eşitleme Puan Farkları ve Standart Hataları</i>	113
Şekil 19 <i>DG Deseni GA Becerisi Logistic ve Uniform Kernel SEED Değerleri</i>	114
Şekil 20 <i>DOG-OM Deseni GA Becerisi FE ve Chain Eşitleme Puan Farkları</i>	118
Şekil 21 <i>DOG-OM Deseni GA Becerisi Eşitlenmiş Puanve Standart Hatalar</i>	119
Şekil 22 <i>DOG-OM Deseni GA Becerisi Eşitlenmiş Puan Yanlılık ve RMSE Değerleri</i>	120
Şekil 23 <i>DOG-OM Deseni GA Becerisi Kernel PSE ve CE Eşitleme Puan Fark ve Standart Hataları</i>	123
Şekil 24 <i>DOG-OM Deseni GA Becerisi Kernel PSE ve CE SEED Değerleri</i>	124

Şekil 25 DOG-OD Deseni GA Becerisi Kernel PSE EYE ve Doğrusal Eşitleme Puan Fark ve Standart Hataları	127
Şekil 26 DOG-OD Deseni GA Becerisi Kernel PSE EYE ve Doğrusal Eşitleme SEED Değerleri	128
Şekil 27 DOG-OD Deseni GA Becerisi PSE Kernel EYE ve Doğrusal Eşitleme Puan Fark ve Standart Hataları	131
Şekil 28 DOG-OD Deseni GA Becerisi PSE Kernel EYE ve Doğrusal Eşitleme SEED Değerleri.....	132
Şekil 29 DG Deseni FD Becerisi Eşitlenmiş Puan Fark ve Bootstrap Standart Hataları.....	137
Şekil 30 DG Deseni FD Becerisi Eşitlenmiş Puan Yanlılık ve RMSE Değerleri .	138
Şekil 31 DG Deseni FD Becerisi Kernel Eşitleme Puan Fark ve Standart Hataları	141
Şekil 32 DG Deseni FD Becerisi Uniform ve Logistic Kernel SEED Değerleri ...	142
Şekil 33 DOG-OM Deseni FD Becerisi FE ve Chain Eşitleme Puan Farkları.....	146
Şekil 34 DOG-OM Deseni FD Becerisi Eşitlenmiş Puan Farkları.....	147
Şekil 35 DOG-OM Deseni FD Becerisi Eşitlenmiş Puanlar ve Standart Hataları	148
Şekil 36 DOG-OM Deseni FD Becerisi Eşitlenmiş Puan Yanlılık ve RMSE Değerleri.....	148
Şekil 37 DOG-OM Deseni FD Becerisi Kernel PSE ve CE Eşitleme Puan Fark ve Standart Hataları	153
Şekil 38 DOG-OM Deseni GA Becerisi Kernel PSE ve CE SEED Değerleri	154
Şekil 39 DOG-OD Deseni FD Becerisi Kernel PSE EYE ve Doğrusal Eşitleme Puan Fark ve Standart Hataları	158
Şekil 40 DOG-OD Deseni FD Becerisi Kernel PSE EYE ve Doğrusal Eşitleme SEED Değerleri	158
Şekil 41 DOG-OD Deseni FD Becerisi Kernel PSE EYE ve Doğrusal Eşitleme Puan Fark ve Standart Hataları	161
Şekil 42 DOG-OD Deseni FD Becerisi Kernel PSE EYE ve Doğrusal Eşitleme SEED Değerleri	162
Şekil 43 Genelleştirilmiş Aritmetik Beceri Puanları Dağılımları	167
Şekil 44 Fonksiyonel Düşünme Beceri Puanlarının Dağılımları	169
Şekil 45 Niceliksel Mantık Yürütme Beceri Puanları Dağılımları	171
Şekil 46 DG Deseni NMY Becerisi Eşitlenmiş Puanlar	203

Şekil 47 DG Deseni NMY Becerisi Eşitlenmiş Puan Fark ve Bootstrap Standart Hataları.....	204
Şekil 48 DG Deseni NMY Becerisi Eşitlenmiş Puan Yanlılık ve RMSE Değerleri	204
Şekil 49 DG Deseni NMY Becerisi Kernel Eşitleme Puan Fark ve Standart Hataları	205
Şekil 50 DG Deseni NMY Becerisi Logistic ve Uniform Kernel SEED Değerleri	206
Şekil 51 DOG-OM Deseni NMY Becerisi Eşitlenmiş Puanlar ve Standart Hataları	208
Şekil 52 DOG-OM Deseni NMY Becerisi Eşitlenmiş Puan Yanlılık ve RMSE Değerleri.....	209
Şekil 53 DOG-OM Deseni NMY Becerisi Kernel PSE ve CE Eşitleme Puan Fark ve Standart Hataları	210
Şekil 54 DOG-OM Deseni NMY Becerisi Kernel PSE ve CE SEED Değerleri...	210
Şekil 55 DOG-OD Deseni NMY Becerisi Kernel PSE EYE ve Doğrusal Eşitleme Puan Fark ve Standart Hataları	212
Şekil 56 DOG-OD Deseni NMY Becerisi Kernel PSE EYE ve Doğrusal Eşitleme SEED Değerleri	213
Şekil 57 DOG-OD Deseni NMY Becerisi Kernel PSE EYE ve Doğrusal Eşitleme Puan Fark ve Standart Hataları	214
Şekil 58 DOG-OD Deseni NMY Becerisi Kernel PSE EYE ve Doğrusal Eşitleme SEED Değerleri	215

Simgeler ve Kısaltmalar Dizini

- ALES:** Akademik Personel ve Lisansüstü Eğitim Sınavı
- BTM:** Bilişsel Tanı Modelleri
- CD-CAT:** Cognitive Diagnostic Computer Adaptive Testing
- DFA:** Doğrulayıcı Faktör Analizi
- DG:** Denk Grup Deseni
- DOG-OD:** Denk Olmayan Gruplar Ortak Değişken Deseni
- DOG-OM:** Denk Olmayan Gruplar Ortak Madde Deseni
- e-YDS:** Elektronik Yabancı Dil Sınavı
- e-YÖKDİL:** Elektronik Yükseköğretim Kurumları Yabancı Dil Sınavı
- Fonksiyonel Düşünme:** FD
- GDINA:** Generalized DINA Model
- Genelleştirilmiş Aritmetik:** GA
- KTK:** Klasik Test Kuramı
- KVA:** Kategorik Veri Analizi
- MTK:** Madde Tepki Kuramı
- NCME:** The National Council on Measurement in Education
- Niceliksel Mantık Yürütme:** NMY
- YDS:** Yabancı Diller Sınavı
- YEM:** Yapısal Eşitlik Modelleri
- YÖKDİL:** Yükseköğretim Kurumları Yabancı Dil Sınavı

Bölüm 1

Giriş

Bu bölümde araştırmada ortaya konulan problem durumu, araştırmacının amacı ve önemi, araştırma problemi ve alt problemleri, araştırmaya ait sayıtlılar ve sınırlılıklar ile araştırma boyunca kullanılan bazı tanımların açıklaması sunulmuştur.

Problem Durumu

Eğitimde uygulanan test ve değerlendirmeler, bireyleri, kurumları ve toplumu önemli ölçüde etkileyen konulardır (NCME, 2014). Eğitim ve öğretim süreçlerinin değerlendirilmesi, öğrenci başarısını anlamak ve öğrenme çıktılarını artırmak amacıyla önemli bir rol oynamaktadır. NCTM (2000), "*değerlendirme öğrencilerin öğrenmelerini artırmalı*" ve "*öğretim kararları için değerli bir araç olmalıdır*" (ss. 22-23) vurgusu yapmaktadır. Geleneksel olarak değerlendirme genellikle bir öğretim biriminin sonunda yapılan testler ve sınavlar aracılığıyla öğrenci başarısını ölçer. Öğrenciler genel bir başarı düzeyini gösteren bir not veya derece gibi geri bildirimler alırlar, ancak bu tür geri bildirimler genellikle öğrencinin güçlü ve zayıf yönleri hakkında özel çıkarımlar yapamaz. Geleneksel değerlendirmeler genellikle özet amaçlar için kullanıldığından, öğrencinin performansını nasıl geliştirebileceğine dair sınırlı geri bildirim sağlar, genellikle pasif olarak görülür ve öğrenmeye hemen etki etmez (Anderson, 1998; Sadler, 1989; Struyven et al., 2003). Bu durumun eğitimin paydaşları açısından daha etkili ve verimli bir duruma dönüşebilmesi amacıyla yararlanılan ve öğretim süreçlerine dahil edilen tanılayıcı ve biçimlendirici değerlendirmeler öğrenme sürecini daha da güçlendirmektedir.

Tanılayıcı değerlendirmeye öğrenci başarısını anlama ve geliştirmenin anahtarı olarak bakmak doğru bir benzetme olur. Geleneksel ölçme ve değerlendirme pratiğinden farklı olarak, tanılayıcı değerlendirme öğrencilerin güçlü ve zayıf yönlerini belirleme, öğrenme gereksinimlerini tespit etme ve bu ihtiyaçlara özelleştirilmiş çözümler sunma konusunda daha ayrıntılı bir bakış açısı sunar (Bejar, 1984; Huff & Goodman, 2007; Sun

& Suzuki, 2013). Bu yöntem, genel başarı puanlarından ziyade öğrencilerin belirli beceri veya kazanımları anlama düzeylerini belirlemeye odaklanır. Öğrencinin neden bir konuyu anlamadığını, hangi becerilerde zayıf olduğunu veya hangi konularda güçlü olduğunu belirlemek için kullanılır. Bu detaylı bakış açısı, öğretmenlere öğrencilerin öğrenme ihtiyaçlarına göre daha özel ve etkili bir şekilde müdahale etme fırsatı sunar. Tanılayıcı değerlendirmenin bir avantajı öğrencilere geri bildirim sağlama kapasitesidir. Bu, öğrencilerin öğrenme stratejilerini geliştirmelerine, eksikliklerini anlamalarına ve gelecekteki başarıları için daha etkili bir öğrenme yol haritası çizmelerine olanak tanır. Eğitimciler, tanılayıcı değerlendirmenin sonuçlarına dayalı olarak öğrencilere özelleştirilmiş öğrenme planları oluşturabilir ve müfredatı bireysel ihtiyaçlara uyacak şekilde düzenleyebilirler. Bu, öğrencilere genel başarı seviyelerinin ötesinde, kişisel öğrenme hedeflerine daha etkili bir şekilde ulaşma fırsatı sunar. Sonuç olarak, tanılayıcı değerlendirme, öğrenci başarısını daha derinlemesine anlama ve geliştirmenin kritik bir aracıdır.

Tanılayıcı değerlendirme kapsamında psikometrik model ve istatistiksel analizlere bağlı olarak yapılan ölçme, literatürde bilişsel ölçme olarak karşımıza çıkar ve bilişsel psikoloji ile eğitimde ölçmenin kesişimini temsil eder (Rupp ve diğerleri,2010). Bilişsel ölçme, öğrenci yanıtlarının ölçülmek istenen ilgili konu, kazanım veya becerilerde güçlü ve zayıf yönlerini teşhis etmeyi amaçlar ve bireysel geri bildirim sunar (de la Torre & Minchen, 2014). İlerleyen teknolojinin bir katkısı olarak, daha öncesi olmakla birlikte 2010 yıllarından günümüze bilişsel ölçmede tanınan, yaygınlaşan ve gelişimini sürdüren Bilişsel Tanı Modelleri (BTM), bilişsel teorileri ve istatistiksel modelleri birleştirerek testi alan katılımcıların yetkinlik profilleri hakkında tanısal bilgi sağlar. Eğitimde günümüzde özellikle çoktan seçmeli test maddeleri ve kısa cevap soruları ile daha az madde ve kısa süreli uygulama ile güçlü tanısal bilgi sunan testlere yüksek talep bulunmaktadır (Leighton & Gierl, 2007). Bu talep eğitimin her alanında dikkate değerdir. Bu talebe de cevap verdiği görülen BTM, öğrenme süreçlerini ve bireyler arasındaki farklılıkları anlama amacı taşıyan

önemli bir istatistiksel yaklaşımı temsil eder (Tatsuoka, 2009). Bu modeller, öğrencilerin bilişsel beceri düzeylerini belirlemek, elde edilen bulgular üzerinden öğrenme süreçlerine dair ayrıntılı iç görüleri olanak sağlayarak öğrenme süreçlerini analiz etmek ve eğitimdeki öğrenci çeşitliliğine uygun müdahaleler geliştirmek amacıyla kullanılan güçlü araçlar olarak bilişsel ölçmede yerini alır (Gierl ve diğerleri, 2007; Henson ve diğerleri, 2007; Huff & Goodman, 2007; Izsák & Templin, 2016; Rupp ve diğerleri, 2010). Sonuç olarak, BTM, eğitim alanında kullanılan önemli bir istatistiksel yaklaşımı temsil etmekle beraber öğrencilerin bilişsel yetkinlik düzeyleriyle ilgili olarak geri bildirimde bulunması, bunlara dayalı olarak eğitim-öğretim süreç ve stratejilerinin geliştirilmesine katkıda bulunması ile eğitimde daha etkili ve özelleştirilmiş bir yaklaşımın önünü açar. Bu modeller, gelecekteki eğitim uygulamalarının daha çeşitli, adil ve etkili olması için önemli bir temel oluşturabilir.

Günümüzde teknolojinin sunduğu imkanların büyük katkısıyla BTM'nin istatistiksel yöntem ve yaklaşımlardaki gelişmelerinin bir sonucu olarak eğitimdeki uygulama çalışmalarının yaygınlaştığı görülmektedir. BTM'nin bir psikometrik model olarak geliştirilmesi çalışmalarından farklı olarak var olan diğer yaklaşımlara ait psikometrik modellerin özelliklerinin kullanılarak yeni BTM modellerinin geliştirilmesi (MTK temelli BTM gibi) çalışmalarının yanı sıra bireye uyarlanmış test uygulama çalışmaları da kendisini göstermektedir (CD-CAT gibi). Ancak, eğitimde ölçme ve değerlendirme çalışmalarında önemli bir yeri olan test eşitleme çalışmalarının bilişsel tanı modellerinde kullanılması ile ilgili çalışmalar yönünden ilgili literatürün şimdilik sığ kaldığı görülmektedir ve bu çalışmada BTM'nin istatistiksel analizler sonucu sunduğu bireysel yetkinlik düzeylerinin test eşitleme yaklaşımlarında kullanımı mercek altına alınmıştır.

Test eşitleme, farklı testleri aynı ölçek üzerinde yerleştirme işlemi olarak tanımlanmakta ve bunun yanı sıra belirli gereksinimleri de karşılayarak, test puanlarını birbirinin yerine kullanılabilir kılmayı amaçlamaktadır (Kolen & Brennan, 2014). Yani, eşitleme, farklı test formlarının puanlarını aynı ölçekte birleştirerek, bu puanların birbirinin yerine geçebildiğini ve aynı anlama geldiğini ifade eder. Bu işlemin temel amacı, farklı

testler arasında tutarlı bir karşılaştırma yapabilmek ve test puanlarının aynı anlamı taşıdığına güvenle karar verebilmektir (Ryan & Brockmann, 2009). Aynı hedef ve nitelikler çerçevesinde hazırlanan farklı test formlarının uygulandığı adayların başarılarında objektif, adil ve karşılaştırılabilir bir şekilde değerlendirme yapabilmek için test eşitleme ihtiyacının ortaya çıkmasına sebep olmuştur. Böylelikle eğitimdeki değerlendirme süreçlerinin daha güvenilir ve geçerli olmasına, eşitsizliklerin önlenmesine ve öğrenci başarılarının objektif bir şekilde değerlendirilmesine katkı sağlanmıştır. Örneğin ülkemizde ÖSYM'nin 2024 yılı için belirttiği sınav takviminde ALES üç kez, yabancı dil sınavlarından YDS iki kez, e-YDS/İngilizce sekiz kez, YÖKDİL iki kez ve e-YÖKDİL üç kez yer almıştır. Sonraki yıllar için bu sınavların yine en az aynı sayıda uygulanacağı ve her sınavın geçerli olduğu süre de hesaba katılırsa bu sınavlara ait sonuç belgesi isteyen bir kuruluş başvuran adaylarda geniş yelpazede farklı test formlarından elde edilmiş sınav sonuç belgeleri ile karşı karşıya kalacaktır. Farklı formlara ait test puanlarının aynen alınıp adaylar arasında bir karşılaştırmanın yapılması ise adil bir değerlendirme konusunu gündeme getirmektedir. Kısacası test eşitlemenin eğitimdeki değerlendirme süreçlerinin daha güvenilir hale getirilmesinde önemli bir araç olduğunu belirtmekte fayda vardır.

Günümüzde bireylerin sahip olduğu potansiyellerin birbirinden farklı olduğu ve bu potansiyellerin bireye özgü biçimde geliştirilmesi kabulü yaygın ve bir gerçek olarak karşımıza çıkmaktadır. Eğitim ve öğretimde disiplin alanlarının ilgili konu kazanımlarının öğrenci tarafından edinilmesi hedeflerinin yanında bu hedeflerin desteklenmesi ve gerçekleşmesinde bilişsel becerilerinin etkin bir biçimde kullanımı ve geliştirilmesi de eğitim ve öğretimde kendini kabul ettiren bir durumdur. Hatta bir testten alınan toplam puanlara göre yapılan değerlendirmelerin yanında becerilerdeki yetkinlik düzeylerinin saptanmasına yönelik yapılan seçme sınavları da önem kazanmaktadır. Bu beceriye dayanan ölçmeler günümüzde hem psikolojik testlerde hem de başarı testlerinde kolaylıkla mümkündür. BTM ile yapılan ölçmeler hedefe göre kazanımların edinilmesi veya beceri ölçmeleriyle ilgili geri bildirim sağlayan önemli bir psikometrik modeldir. Bunun

yanında bu testlerden elde edilen yetkinlik düzeylerinin farklı testler için eşitleme durumu ise akademik çalışma alanı olarak beklemektedir.

Sonuç olarak, psikometrik bir yaklaşım olarak BTM'nin eğitim ve öğretime sunduğu geribildirimler ile olan katkısı son derece önemlidir. Test eşitlemenin de özellikle karşılaştırılabilir test puanları elde etmekteki katkısı adillik açısından kıymetlidir. Klasik test teoremi, madde tepki kuramı gibi yaklaşımlarda pek çok eşitleme yöntem ve uygulamaları şu an ilgili alanyazında belli bir olgunluk dönemine ulaşmış olduğunu belirtmek yanlış olmaz. Ancak BTM ile eşitleme henüz başlangıç aşamasındadır denilebilir. Bu çalışma kapsamında saptanan temel bir problem, BTM analizlerinden elde edilen yetkinlik düzeylerinin farklı testler arasında eşitlenmesinin incelenmesidir. Yani, farklı testlerle ölçülen beceri düzeylerinin karşılaştırılabilir hale getirilebilmesi için etkili bir eşitleme sürecinin geliştirilmesi gerekmektedir. Bu, öğrencilerin gerçek yeteneklerini objektif bir biçimde değerlendirmenin yanı sıra, eğitim programlarının etkililiğini daha kapsamlı bir şekilde değerlendirebilmenin önünü açacaktır. Kısacası bu çalışmada, BTM çerçevesinde aynı hedeflerle hazırlanan ve aynı becerileri ölçmek üzere uygulanan farklı test formlarının yine BTM çerçevesinde analiz edilerek elde edilen beceri yetkinlik düzeylerinin eşitlenebilme durumlarının nasıl olduğu problemi üzerinde durulmuştur.

Araştırmanın Amacı ve Önemi

Bu çalışmanın amacı, BTM çerçevesinde oluşturulan test formlarının BTM çerçevesinde analizleri ile elde edilen yetkinlik düzeyi değerlerinin kullanılarak geleneksel yöntemlerle test eşitleme uygulamaları ile testler arası karşılaştırılabilir beceri puanlarının elde edilebilirliğinin incelenmesidir.

BTM ile ilgili yapılan çalışmaların büyük bir çoğunluğu genellikle modelin istatistiksel verim ve etkinliğini artırmak ve model gelişimine katkı sağlamak amacıyla üretilmiş verilerle yapılan çalışmalar sunmaktadırlar (Chen & Chen, 2016; de la Torre & Karelitz, 2009; Ravand, 2016). Yine ilgili diğer çalışmalarda farklı sınıflandırma modelleri

karşılaştırılmakta ve genellikle sınıflandırma ve geri bildirim yerine farklı amaçları dikkate alarak uygulanan büyük ölçekli sınavlardan elde edilen cevap örüntülerinin BTM'ye uyumlandırılarak (*retrofitting*) yapılan çalışmalar yer almaktadır (Gierl & Cui, 2008; Ketabi ve diğerleri, 2021; Lee & Luna-Bazaldua, 2019; Tatsuoka et al., 2016). Bu çalışmadaki kullanılan veriler ise tanılayıcı değerlendirme kapsamında beceri ölçmesi ve geri bildirim sunabilmeyi hedefleyen BTM temelli bir çerçeve ile oluşturulmuş tanılayıcı bir testin uygulanmasından elde edilen gerçek verilerdir. Bu gerçek veriler BTM'nin eğitimdeki rolünün anlaşılması ve geliştirilmesine katkı sağlaması bakımından önemlidir. Ayrıca, BTM'nin aynı zamanda gelişmeye devam eden bir model olması nedeniyle, alanyazında belirli ders kazanımları ile yapılan uygulamalar baskındır (Gierl & Leighton, 2007; Norris ve diğerleri; Rupp ve diğerleri, 2010). Bu çalışmadaki önemli bir bakış açısı ise BTM çerçevesinde tanımlanan ve analizlerde kilit rol oynayan nitelikler için ders konu kazanımları kullanılmamış, pek çok alt kazanımın edinilmesini destekleyen beceriler ilgili testlerde nitelik olarak tercih edilmiştir. Araştırma, kullanılan niteliklerin kapsamı bakımından da alanyazındaki çalışmalardan farklı bir çerçeve ile ortaya çıkmaktadır. Bu durum, özellikle geniş ölçekli ölçme uygulamalarına büyük katkı sağlar. Bilindiği gibi, bu gibi sınavlarda tüm ders kazanımlarını değerlendirmek kolay değildir. Ancak, becerilere dayalı çerçeve aynı zamanda her bir hedefin elde edilmesi hakkında ipuçları sunmaktadır.

Bu çalışma BTM'nin eğitimin öğretim boyutuna katkı sağladığı gibi eğitimin ölçme boyutu açısından elde edilen beceri puanlarının eşitlenebilme durumunun incelenerek eğitimin adillik-eşitlik anlayışı çerçevesinde katkı sunmaktadır. Çalışma kapsamında BTM çerçevesindeki farklı test formları arasında eşitlenmiş beceri puanları elde etme amacıyla yapılan inceleme BTM ve test eşitleme konusu üzerine araştırmacıların dikkatini çekerek eğitimde ölçme ve değerlendirme alanında önemli bir role sahip olmuştur. Böylelikle BTM çerçevesinde maddelerin yazımı ve test oluşturma safhasından itibaren oluşturulan testlerin güvenilirliğini ve geçerliliğini artırmak, adil değerlendirme sağlamak ve bu hedef doğrultusunda kullanılan ölçme araçlarını daha geniş kitlelere uygulanabilir kılmak

açısından önem arz etmektedir. Araştırmanın temel bulguları, ölçme ve değerlendirme merkezleri ile bu alanda çalışan araştırmacılara BTM ve test eşitleme ile ilgili değerli ipuçları sunacaktır. Bu ipuçları, özellikle geniş ölçekli test uygulamaları yapan kurumlar için test eşitleme süreçlerini daha etkili ve güvenilir hale getirebilmek adına bir katkı sunmaktadır. Bu çalışma, sadece uygulayıcılara yönelik pratik çözümler sunmakla kalmamakta, aynı zamanda BTM ve eşitleme ile ilgili olarak teorik düzeyde de katkısı olacaktır. Elde edilen bulgular, BTM ve test eşitleme ile ilgili çalışmalara ipucu niteliği taşıyarak en doğru eşitleme sonuçlarını elde edebilmek için gerekli koşulları belirlemeye yönelik yapılan araştırmalara yeni bir bakış açısı getirebilecektir. Bu bağlamda, BTM ve test eşitleme süreçleri, ölçme ve değerlendirme alanında çalışan profesyonellerin dikkatle üzerinde durması gereken kritik bir konudur. Hem uygulamada hem de teoride yapılan bu tür araştırmalar, ölçme araçlarının daha güvenilir, adil ve etkili bir şekilde kullanılabilmesi için önemli bir rehberlik sağlamaktadır.

Araştırma Problemi

BTM çerçevesinde aynı hedeflerle hazırlanan ve cebirsel düşünme becerilerini ölçmek üzere uygulanan farklı test formlarının yine BTM çerçevesinde analiz edilerek saptanan beceri yetkinlik düzeyi değerlerinin kullanılarak elde edilen beceri puanlarının farklı grup desenlerinde ve eşitleme yaklaşımlarında eşitleme durumu eşitleme puanları ve hataları açısından nasıldır?

Alt Problemler

Bu çalışmada aşağıdaki sorulara cevap aranmaktadır.

1. BTM çerçevesinde aynı hedeflerle hazırlanan ve cebirsel düşünme becerilerini ölçmek üzere uygulanan tanılayıcı paralel formların BTM çerçevesinde analiz edilerek saptanan genelleştirilmiş aritmetik (GA) becerisi yetkinlik düzeyi değerlerinden elde edilen GA beceri puanlarının Test A-Test B, Test A-Test C ve Test B-Test C karşılaştırmalarına göre

- i. denk gruplar (DG) deseni kapsamında Klasik Test Kuramı'na (KTK) dayalı geleneksel eşitleme (ortalama, doğrusal, eşit yüzdelikli) ve Kernel eşitleme yöntemleri ile eşitlenme durumu eşitleme puanları ve hataları açısından nasıldır?
- ii. denk olmayan gruplarda ortak madde (DOG-OM) deseni kapsamında KTK'ya dayalı geleneksel eşitleme (ortalama, doğrusal, eşit yüzdelikli) ve Kernel eşitleme yöntemleri ile eşitlenme durumu eşitleme puanları ve hataları açısından nasıldır?
- iii. denk olmayan gruplarda ortak değişken (DOG-OD) deseni kapsamında KTK'ya dayalı Kernel eşitleme yöntemi ile eşitlenme durumu eşitleme puanları ve hataları açısından nasıldır?

2. BTM çerçevesinde aynı hedeflerle hazırlanan ve cebirsel düşünme becerilerini ölçmek üzere uygulanan tanılayıcı paralel formların BTM çerçevesinde analiz edilerek saptanan fonksiyonel düşünme (FD) becerisi yetkinlik düzeyi değerlerinden elde edilen FD beceri puanlarının Test A-Test B, Test A-Test C ve Test B-Test C karşılaştırmalarına göre

- i. DG deseni kapsamında KTK'ya dayalı geleneksel eşitleme (ortalama, doğrusal, eşit yüzdelikli) ve Kernel eşitleme yöntemleri ile eşitlenme durumu eşitleme puanları ve hataları açısından nasıldır?
- ii. DOG-OM deseni kapsamında KTK'ya dayalı geleneksel eşitleme (ortalama, doğrusal, eşit yüzdelikli) ve Kernel eşitleme yöntemleri ile eşitlenme durumu eşitleme puanları ve hataları açısından nasıldır?
- iii. DOG-OD deseni kapsamında KTK'ya dayalı Kernel eşitleme yöntemi ile eşitlenme durumu eşitleme puanları ve hataları açısından nasıldır?

3. BTM çerçevesinde aynı hedeflerle hazırlanan ve cebirsel düşünme becerilerini ölçmek üzere uygulanan tanılayıcı paralel formların BTM çerçevesinde analiz edilerek saptanan niceliksel mantık yürütme (NMY) becerisi yetkinlik düzeyi değerlerinden elde

edilen NMY beceri puanlarının Test A-Test B, Test A-Test C ve Test B-Test C karşılaştırmalarına göre

- i. DG deseni kapsamında KTK'ya dayalı geleneksel eşitleme (ortalama, doğrusal, eşit yüzdelikli) ve Kernel eşitleme yöntemleri ile eşitlenme durumu eşitleme puanları ve hataları açısından nasıldır?
- ii. DOG-OM deseni kapsamında KTK'ya dayalı geleneksel eşitleme (ortalama, doğrusal, eşit yüzdelikli) ve Kernel eşitleme yöntemleri ile eşitlenme durumu eşitleme puanları ve hataları açısından nasıldır?
- iii. DOG-OD deseni kapsamında KTK'ya dayalı Kernel eşitleme yöntemi ile eşitlenme durumu eşitleme puanları ve hataları açısından nasıldır?

Sayıtlılar

Katılımcıların; test uygulamalarına aynı koşullar altında katıldıkları kabul edilmiştir.

Sınırlılıklar

Araştırma Ankara'nın Çankaya bölgesindeki kamu okullarında liseye yeni başlamış (9. sınıf veya lise hazırlık sınıfı) olan öğrencilerin katılımıyla cebirsel düşünme becerisi testlerine verdikleri cevapların BTM kapsamında karma hiyerarşik yapılı GDINA modeli uygulamasına dayalı test istatistiğinde saptanan cebirsel düşünme becerileri yetkinlik düzeyi değerleri kullanılarak elde edilen genelleştirilmiş aritmetik, fonksiyonel düşünme ve niceliksel mantık yürütme beceri puanlarının kullanımıyla sınırlıdır.

Tanımlar

Cebirsel Düşünme Becerileri Testi: Geniş bir alanyazın çalışmasına bağlı olarak belirlenen altı adet cebirsel düşünme becerisini (i. genelleştirilmiş aritmetik, ii. fonksiyonel

düşünme, iii. niceliksel mantık yürütme, iv. genelleme ve mantık yürütme, v. temsilleştirme, vi. gerekçelendirme) bilişsel tanı modelleri çerçevesinde ölçme yapmayı amaçlayarak tasarlanan, 20 test maddesinden oluşan ve ortaokulu tamamlamış veya liseye yeni başlayan öğrencilerin seviyesine uygun olarak hazırlanmış olan bir ölçme aracıdır.

Bölüm 2

Araştırmanın Kuramsal Temeli ve İlgili Araştırmalar

Bu bölümde araştırmanın kuramsal temeli çerçevesinde Bilişsel Tanı Modelleri (BTM) ile başlangıç yapılmış, bu modellerin eğitimdeki uygulamalarına dair uluslararası ve ulusal çalışmalardan örnekler sunulmuş, araştırmanın diğer bir kuramsal temeli olan Klasik Test Kuramı (KTK) kapsamında test eşitleme yöntem ve uygulamalarına değinilmiştir.

Bilişsel Tanı Modelleri

Huff ve Goodman (2007) bilişsel değerlendirme tarihinin Embretson'un 1983 yılı Psikolojik Bülten'deki yayımlanan makalesi ile başlatılmaktadır. Bu makalede Embretson'un bilişsel psikolojideki gelişmeler ile güncel yapı geçerliliği kavramlarını etkili bir biçimde birleştirdiğini ve madde zorluğu ile bilişsel süreçler arasındaki ilişkinin modellenme ihtiyacına vurgu yaptığını belirtir. İşte bu sebeple Embretson'un yayını bilişsel psikolojideki gelişmelerin ölçme teorisine uygulanması açısından önemli bir adım olduğuna dikkat çeker. İşte bu yıllar itibariyle eğitimde ölçme ile ilgilenen araştırmacılar tarafından test teorisi ve uygulamalarının öğrenci bilişini tek bir gizil özellikten ziyade daha geniş bir şekilde yansıtmak üzere kavramsallaştırmaya ve testi öğretimle ve öğrenmeyle daha etkili bir biçimde birleştirmeye olan gereksinimler belirtilerek eğitimde ölçme uzmanları yani psikometristler ile bilişsel psikologların işbirliği yapmaları talep edilmiştir (Leighton & Gierl, 2007). Nichols (1994) geleneksel test uygulamalarının performansın psikolojik yönlerini değerlendirmekte başarısız olması ve yapı temelindeki psikoloji yerine daha çok işlemlere ve istatistiksel yöntemlere odaklanması nedeniyle eleştiri aldığını belirterek bilişsel bilim ve psikometriyi birleştirmeyi amaçlayan "*bilişsel tanılayıcı değerlendirme*" terimini tanıtmıştır. Bu perspektif, öğrenmeyi bir özellik sürekliliği boyunca kademeli değişiklikler olarak gören geleneksel görüşten farklı olarak öğrenmenin bilişsel süreçlerde ve yapılarında meydana gelen değişikliklerle gerçekleştiğini kabul eder.

Bilişsel tanılayıcı değerlendirmeler temelde tanılayıcı olduğundan, bu seviyedeki bilgiyi veriden çıkartabilen istatistiksel modellere ihtiyaç duyarlar ve bu tür modeller Bilişsel Tanı Modelleri olarak adlandırılırlar. MTK ve KTK modellerinin betimleyici doğasına karşı duran BTM, bir alternatif psikometrik çerçeve olarak geliştirilmiştir ve testi alan katılımcıların belirli nitelikler kümesine göre sınıflandırılması şeklinde tanılayıcı bilgi sağlamayı amaçlar. Bunun yanında BTM'nin, kesikli örtük değişkenlere uygulanan MTK olduğu hatırlatılmakta ve bu nedenle, uygun düzenlemelerle MTK'daki mevcut yöntemler, işlem ve uygulamaların (örneğin, tahmin, model uygunluk değerlendirmesi, değişen madde fonksiyonu analizi, bilgisayar destekli bireye uyarlanmış test) kolayca BTM bağlamında uygulanabilir olduğu da vurgulanmaktadır (de la Torre & Minchen, 2014).

Bilişsel teori, bilgi yapılarının nasıl geliştiği, düşünme ve akıl yürütme süreçlerinin nasıl gerçekleştiği ve karşılaşılan sorunları çözmek için kullanılan stratejilerle ilgilenir. Bilgi işleme teorisi ise insanların bilgiyi nasıl aldıklarını, düzenlediklerini, temsil ettiklerini, sembolize ettiklerini, hatırladıklarını ve kullanmaya çalıştıklarını açıklamaya çalışır. Bu teori, insan zihninin bilgiyi nasıl işlediğini ve hangi türden birimlerle işlem yaptığını inceler (Lohman, 2000). BTM de tanısal değerlendirme bağlamında kullanıldığı ve geri bildirim önemli bir rol oynadığı için bilişsel teorinin içinde yer alan bilgi işleme teorilerinden faydalandığı görülmektedir (Leighton ve diğerleri, 2004). Bilgi işleme teorileri, çevresel olaylara odaklanma, bilgiyi kodlama, yeni bilgiyi mevcut bilgiyle bağlama, bilgiyi yeni bilgi olarak depolama ve bellekten bilgiyi çekme gibi zihinsel süreçlere odaklanır (Sternberg, 1977). Doğrudan gözlemlenemeyen zihinsel süreçlerin (genelleme, akıl yürütme, temsil etme, gerekçelendirme, planlama, hayal etme, hatırlama ve kodlama gibi) daha iyi anlaşılması için zihindeki süreçlere bir metafor olarak bilgisayarlar kullanılabilir. Sonuç olarak, BTM önceden belirlenmiş ancak gözlemlenemeyen süreçler aracılığıyla çözülebilen madde yanıtlarını kullanarak bu gözlemlenemeyen süreçleri ölçmeyi amaçlamıştır.

BTM'nin temeli, Klasik Test Kuramı (KTT), Madde Tepki Kuramı (MTK), Doğrulayıcı Faktör Analizi (DFA), Yapısal Eşitlik Modelleri (YEM), Kategorik Veri Analizi (KVA) ve Bayes istatistiği gibi önemli psikometrik ve istatistiksel çerçeveleri kullanır (Rupp ve diğerleri, 2010). BTM parametrik ve doğrulayıcı yaklaşımları kullanır ve sınıflandırma işlemi istatistiksel çıkarımlar üzerine kurulmuştur. Bu, BTM'nin etkililiği, yapılan işlemlerin gerçek durumlara yakınlığı yani kesinliği, doğruluğu ve güvenilirliği açısından güçlü yönüdür. Diğer bir güçlü yanı da, sıralama veya karşılaştırma amaçlarıyla ön plana çıkan ve öğrenciler hakkında tanısal bilgide zayıf kalan KTK ve MTK modellerindeki gibi öğrencileri tek bir sürekli ölçekte konumlandırmak yerine çoklu özelliklere göre sınıflandırmasıdır. Öğrencilerin, öğretmenlerin ve eğitimin diğer paydaşlarının tanısal bilgiye duyduğu ihtiyaç kabul gören bir durumdur. BTM, ölçülmek istenen özelliğin cevaplayıcıdaki durumunu açıklamak, yetkinlik düzeyi hakkında bilgi vermek, yeterli olup olmadığını belirtmek suretiyle bu boşluğu doldurmayı amaçlar. Bu açıklama, bir test için belirlenmiş nitelikleri kullanarak yapılır. BTM'de nitelikler, katılımcıların örtük özellikleridir ve bir madde çözümü için gerekli becerileri, bilgiyi, stratejiyi veya işlemsel sürecini tanımlayan gizil değişkenlere işaret eder (de la Torre & Minchen, 2014; Ma & de la Torre, 2019; Nichols ve diğerleri, 1995). Her bir nitelik ile ilgili olarak, katılımcıların o nitelikte yetkin olma veya olmama durumu hakkında geri bildirim verilir. Genel olarak, öğrencilerin güçlü ve zayıf yönleri kolayca görülebilir ve tüm paydaşlar bu bilgiyi kullanabilir. Kısacası, özellikleri açısından BTM doğrulayıcıdır, bir madde çözüm sürecinde birden fazla nitelik içerebilir, kesikli ve çok boyutlu gizil değişkenlere sahiptir ve sınıflandırma, niteliklerden elde edilen yetkinlik düzeyine uygun olarak yapılır.

Rupp ve Templin (2008) BTM için dokuz belirleyici özellikten bahseder. Bunlar; i) çok boyutlu doğası (bir modelin birden fazla değişkeni içerebilmesine olanak tanır), ii) doğrulayıcı doğası (belirli hipotezleri test edebilme özelliği sunar), iii) yüklenme yapısının karmaşıklığı (değişkenler arasındaki ilişkilerin ayrıntılı bir şekilde incelenebileceğini gösterir), iv) Uygun oldukları gözlemlenebilir yanıt değişken türleri (bu modellerin geniş bir

uygulama alanına sahip olduğunu belirtir), v) içerdikleri gözlemlenemeyen yordayıcı değişken türleri (bu değişkenler, modelin temel özelliklerini belirlemede kullanılan ancak doğrudan gözlemlenemeyen faktörleri ifade eder), vi) gizil yordayıcı değişkenlerin etkileşim şekli (modeldeki gizli faktörler arasındaki karşılıklı ilişkileri ve bağlantıları anlamaya odaklanır), vii) izin verdikleri ölçütleme ile ilgili yorumlamalar (kriter-tabanlı yorumlamalara izin verme özelliği bulunur), viii) yorumlamaların tanısal doğası (modelin bireylerin öğrenme güçlükleriyle ilgili detaylı bilgiler sağlama yeteneğini vurgular), ix) modelleyebilecekleri heterojenlik türleri (farklı birey gruplarının özelliklerini anlama kapasitesini yansıtır).

BTM, kategorik yanıt değişkenler için uygundur. Yanıtlar gözlemlenebilir değişkenleri içerirken, aynı zamanda gözlemlenemeyen gizil değişkenleri de içermektedir. BTM, bu tür gözlemlenemeyen kategorik değişkenleri ortaya çıkarmaya çalışır. Bu tür gözlemlenemeyen değişkenler, konuya özgü geniş veya dar kapsamlı farklı seviyedeki tanımlayıcıların veya özelliklerin bir kümesi olarak adlandırılır ve Tatsuoka'nun (1983) belirttiği üzere bunların her biri "*nitelik (attribute)*" olarak adlandırılan unsurlardır. Katılımcının belirli niteliklerdeki yetkinlik durumuna (*mastery profile*) göre sınıflandırma gerçekleştirilir. Bilindiği üzere, bir nitelikte yetkinlik düzeyine ulaşan bir kişi o nitelik için "yetkindir" ve "1" olarak sembolize edilirken, yetkinlik düzeyine ulaşamayanların "yetkin değildir" ve "0" olarak sembolize edilir. Örneğin, bir sınav katılımcısı altı nitelik açısından sınıflandırılırsa ve katılımcının yetkinlik durumu $q = (1, 1, 1, 0, 1, 0)$ vektörüyle gösterilirse, bu katılımcının ilk üç niteliği ve beşinci niteliği başarıyla tamamladığı anlamına gelir, ancak diğer iki nitelik (dördüncü ve altıncı nitelik) için durum tersidir. Bu nitelikler BTM'nin temel ögesi olan ve bir test maddesinin bileşenlerini gösteren Q-matrisin oluşturulmasında kullanılır. Vurgulandığı gibi, Q-matris, bilişsel teorilerin ve ölçme modellerinin entegrasyonunun gerçekleştiği BTM'nin temel taşıdır (Chiu, 2013; de la Torre, 2016; Shu-Liang, 2019; Tatsuoka, 2009).

Q-matrisin oluşturulmasının önemi, sınıflandırmanın temel unsurları olan bilgi yapılarına odaklanmayı sağlamasıdır. Q-matris, belirlenen madde nitelikleri doğrultusunda oluşturulur. Q-matris, maddenin cevaplanabilmesine ait özel niteliklerin düzenli olarak gösterimini temsil eder. Kısacası, her maddenin başarılı bir şekilde tamamlanması için gereken nitelikleri sıralar. Detaylı olarak ise, her bir sınav katılımcısı "e" ve nitelik sayısını "K" ile sembolize edilsin. Buna göre, α_e yetkilik durumunu (*mastery profile*) gösteren vektör ve $\alpha_{ek} = 1$ ya da $\alpha_{ek} = 0$ olmak üzere $\alpha_e = (\alpha_{e1}, \alpha_{e2}, \alpha_{e3}, \dots \dots \alpha_{eK})$ vektöründe her bir niteliğin yetkinlik düzeyine ulaşma durumu gösterilir. Bu durumda K tane niteliğin ölçüldüğü bir bilişsel tanılayıcı testte 2^K sayısınınca farklı yetkinlik durumlarını gösteren örtük sınıflar ortaya çıkar. Ayrıca, her maddenin ölçmeyi hedeflediği nitelikler ise bir q vektörü ile temsil edilir, örneğin "madde j" için her bir niteliğin ölçülme durumuna göre $q_{iK} = 1$ veya $q_{iK} = 0$ olmak üzere $q_i = (q_{i1}, q_{i2}, q_{i3}, \dots \dots q_{iK})$ vektörü oluşturulur. Her bir maddeye ait q vektörlerinin bir araya getirilmesi ile oluşan matris ise BTM'nin istatistiksel süreçleri için anahtar görevi gören ve testteki madde sayısı "J" olmak üzere $J \times K$ boyutlu bir Q-matristir. Örneğin, dört nitelik ölçmeyi hedefleyen çoktan seçmeli bir bilişsel tanılayıcı testin, onuncu maddesi dört nitelikten ilk ikisini ve sonuncusunu ölçüyorsa Q-matristeki onuncu vektör şu şekilde gösterilir: $q_{10} = (1, 1, 0, 1)$ Belirtilen vektörün en az bir değerinin 1 olması önemlidir, çünkü BTM'de her madde en az bir niteliği ölçmek üzere tasarlanmıştır. Tüm değerleri 0 olan bir vektör olamaz. Aşağıda Tablo 1'de dört niteliği ölçen tanılayıcı bir testin ilk altı maddesine ait madde-nitelik dizilimi örneklendirilmiştir. Q-matris için örnekteki benzer biçimde oluşturulan madde-nitelik dizilimi kullanılır.

Tablo 1*Madde-Nitelik Dizilimi Örneği*

Maddeler	Nitelikler			
	α_1	α_2	α_3	α_4
1	0	0	0	1
2	1	0	1	0
3	0	0	1	1
4	0	0	1	1
5	1	1	0	0
6	1	0	0	0

Tablo 1’de görüldüğü gibi, örnek verilirse birinci madde sadece dördüncü niteliği, ikinci madde birinci ve üçüncü nitelikleri ölçmektedir. Daha önce belirtildiği gibi, tanılayıcı bir testte dört nitelik ölçülüyorsa ve niteliklerin birbirinden bağımsız olduğu varsayılırsa, dört elemanı bulunan bir kümenin alt kümeleri olarak, $2^4=16$ tane $c_1=(0,0,0,0)$, $c_2=(1,0,0,0)$, $c_3=(0,0,1,1)$, $c_4=(1,0,1,1)$, $c_{16}=(1,1,1,1)$ biçiminde sınıflar oluşur. Her sınıf, o sınıfa dahil edilen bireylerin yetkinlik durumunu göstererek her bir nitelikte yetkin olup olmadıkları hakkında bilgi verir. Örneğin, c_1 sınıfı bireylerin tüm niteliklerde yetkinlik seviyesine ulaşamadıklarını gösterir. Diğer taraftan, c_{16} sınıfındaki bireyler tüm niteliklerde yetkinlik seviyesindedir. Dört numaralı sınıftaki bireyler sadece ikinci nitelikte yetkin olmayıp, diğer niteliklerde yetkindirler. Bir bireyin her bir nitelikteki yetkinlik seviyesi, o niteliğin yetkinlik seviyesini belirleyen olasılığın 0.50'nin altında veya üstünde olup olmamasına göre belirlenir. Bir niteliğin yetkin olma olasılığı 0.50'nin altında ise, katılımcı o nitelikte “*yetkin değil*” biçiminde sınıflandırılır ve 0 olarak atanır; diğer durumda ise 1 olarak atanır. Ayrıca kesme puanları çalışmanın doğasına göre ayarlanabilir. Örneğin, 0.40 ile 0.60 arasındaki yetkinlik seviyesi olasılıkları “*belirsiz*” olarak ayarlama yapılabilir.

Sonuç olarak BTM, tanılayıcı testlere yanıt verenleri farklı nitelik yetkinlik durumları ile ilişkilendirilmiş bir dizi belirgin örtük sınıflardan birine sınıflandırmak için kullanılan ayrık gizil değişkenlere sahip istatistiksel modellerdir (Rupp ve diğerleri, 2010). Bu modeller literatürde farklı isimlerle ve istatistiksel çerçevelerle kullanılsa da (*cognitive psychometric models* (Gorin & Svetina, 2012; Rupp, 2007; Svetina ve diğerleri, 2011), *diagnostic*

classification models (Bradshaw, 2016; Rupp ve Templin, 2008; von Davier & Lee, 2019), *latent response models* (Embretson, 2013; Maris, 1995; Ulitzsch ve diğerleri, 2020), *restricted latent class models* (Haertel, 1989; Vermunt, 2001; Xu & Shang, 2018), *multiple classification latent class models* (Maris, 1999; Nussbeck & Eid, 2015; Xu & von Davier, 2019), *structured item response theory models* (Rupp & Mislevy, 2007), *cognitive diagnosis models* (de la Torre, 2009; Liu ve diğerleri, 2023; George ve diğerleri, 2016; Templin & Henson, 2006), *cognitive diagnostic models* (DiBello & Stout, 2007; Heller ve diğerleri, 2015; Ravand & Robitzsch, 2015) bilişsel tanılamada kullanılmak üzere modellenmişlerdir. Kullanılan isim çeşitliliği bile fazlaca olan BTM'nin gelişim sürecinde kullanılan istatistiksel modelleri de az değildir ve bunların sınıflandırılması Rupp ve diğerleri (2010) tarafından aşağıda verilen Tablo 2'de yapılmıştır.

Tablo 2

BTM'nin Sınıflandırılması (Rupp ve diğerleri, 2010, s.98)

Gözlenen Cevap Değişkenleri	Gizil Yordayıcı Değişkenler		Model Türü	
	<i>İki kategorili (dichotomous)</i>	<i>Çok kategorili (polytomous)</i>		
<i>İki kategorili (dichotomous)</i>	RSM		Tamamlayıcı Olmayan (Noncompensatory)	
	AHM			
	DINA			
	HO-DINA			
	MS-DINA			
	NIDA			
	RERUM			
	BIN	BIN		
	MCLCM	MCLCM		
	Full NC-RUM	Full NC-RUM		
	Reduced NC-RUM	Reduced NC-RUM		
	DINO			Tamamlayıcı Olan (Compensatory)
	NIDO			
	BIN	BIN		
MCLCM	MCLCM			
C-RUM	C-RUM			
GDM	GDM			
H-GDM	H-GDM			

Tablo 2 (Devamı)

	LCDM	LCDM	
	G-DINA	G-DINA	
<i>Çok kategorili (polytomous)</i>	RSM		Tamamlayıcı Olmayan (Noncompensatory)
	AHM		
	BIN	BIN	
	MCLCM	MCLCM	
	Full NC-RUM	Full NC-RUM	
	Reduced NC-RUM	Reduced NC-RUM	
	BIN	BIN	Tamamlayıcı Olan (Compensatory)
	MCLCM	MCLCM	
	C-RUM	C-RUM	
	GDM	GDM	
	H-GDM	H-GDM	
	LCDM	LCDM	
	G-DINA	G-DINA	

Not: RSM:Rule-space method, AHM:Attribute hierarchy method, BIN: Bayesian inference network, DINA: Deterministic inputs, noisy “and” gate, HO-DINA: Higher-order DINA, MS-DINA: Multistrategy DINA, LCDM: Loglinear Cognitive Diagnosis Model, DINO: Deterministic inputs, noisy “or” gate, NIDA: Noisy inputs, deterministic “and”gate, NIDO: Noisy inputs, deterministic “or” gate, RUM: Reparametrized unified model / Fusion model, C-RUM: Compensatory RUM, NC-RUM: Non-compensatory RUM, GDM: General diagnostic model, LCDM: Loglinear cognitive diagnosis model, MCLCM: Multiple classification latent class model, GDINA: Generalized DINA.

Tablo 2’de görüldüğü üzere modeller gözlenen ve gizil yordayıcı değişkenlerin iki kategorili veya çok kategorili olma durumları ile modellerin tamamlayıcı olup olmama durumlarına göre sınıflandırılmıştır. Tamamlayıcı modeller, bilişsel süreçlerdeki eksikliklerin diğer beceriler tarafından telafi edilebileceğini öne sürer. Yani, ölçülmek istenen bir beceride yetkin olunmaması durumunda, yetkin olunan beceriler tarafından dengeleme potansiyeline sahiptir. Tamamlayıcı modellerde, zayıf olunan beceri ya da beceriler yetkin olunan beceri ya da beceriler tarafından tamamlanabilir. Tamamlayıcı olmayan modellerde ise her bir beceri, doğru bir yanıt verebilmek için gereklidir. Yani, zayıf olunan beceri ya da beceriler diğer beceri ya da beceriler tarafından dengeleme şansına sahip değildir. Herhangi bir zayıflık, doğru bir yanıt alabilmek için kritiktir. Eğer bir beceride zayıflık varsa bu diğer becerilerle telafi edilemez ve bu beceri tamamlanmadıkça doğru bir yanıt mümkün değildir. Bu iki model türü, bilişsel süreçlerin nasıl işlediğine dair farklı varsayımları temsil eder. Tamamlayıcı modellerde, bilişsel esneklik ve dengeleme

vurgulanırken; tamamlayıcı olmayan modellerde ise her bir becerinin maddeye verilecek cevapta kritik bir öneme sahip olduğu vurgulanmaktadır.

GDINA Modeli

G-DINA (*generalized deterministic inputs, noisy "and" gate*) modeli (de la Torre, 2011), daha esnek varsayımlara sahip olarak DINA modelinin geliştirilmiş biçimidir. Saturated G-DINA modeli (*doyurulmuş GDINA*), alternatif bağlantı fonksiyonlarına dayalı bilişsel tanı için diğer genel modellere eşdeğerdir. Saturated terimi, bir modelin içerdiği tüm parametreleri veya serbestlik derecelerini kullanma durumunu ifade eder. Yaygın olarak kullanılan bilişsel tanı modelleri uygun kısıtlamalar altında genel modellerin özel durumları olduğu bilinmektedir. Genel bir BTM çerçevesi olan G-DINA modeli, tasarım (*design*) ve ağırlık (*weight*) matrislerine dayalı olarak her bir madde için hem model tahmini bileşenini hem de Wald testine dayalı olarak model karşılaştırma bileşenini içerir. GDINA modelinin çerçevesi, tekli olarak puanlanan test maddeleri için mevcut olan en basit ve dolayısıyla en kısıtlayıcı, yorumlanabilir bilişsel tanı modellerinden biri olan DINA modeline dayanmaktadır. Maddenin ölçtüğü nitelik sayısından bağımsız olarak doğru veya yanlış olarak puanlanan maddeler bağlamında DINA modeli, her bir madde için yalnızca iki parametreyi gerektiren tutumlu bir modeldir. DINA modeline göre, bir maddenin doğru cevaplanabilmesi için eşit derecede önemli birkaç niteliğin bir arada olması gerekir. Bir maddenin gereken bir niteliği eksik olması durumunda, maddenin tamamı için gereken tüm niteliklerin eksik olduğu durumla aynı olarak görülür. DINA modelinde, K sayıda nitelik ölçen bir testteki j . madde, $j = 1, \dots, J$, 2^K gizil nitelik vektörünü veya sınıflarını iki gizil gruba böler. Madde j 'nin nitelik gereksinimleri verildiğinde, her bir 2^K nitelik vektörü, deterministik olarak, o maddenin tüm gerekli niteliklerine sahip olan nitelik vektörleri grubu η_{j1} veya aynı maddenin en az bir gerekli niteliği eksik olan nitelik vektörleri grubu η_{j0} 'a ait olarak sınıflandırılır. DINA modelinde, aynı gruptaki nitelik vektörleri, bir maddeye doğru cevap verme olasılığının aynı olduğu şeklinde varsayılır. Yani, bir madde için gereken

aynı niteliklere sahip olan öğrencilerin, o maddeye cevap verme olasılıkları birbirine eşittir. Niteliklerin doğru bir şekilde tanımlandığı ve belirtildiği durumlarda bile, bu varsayım her zaman η_{j0} grubu için geçerli olmayabilir. Başka bir deyişle, bazen bu varsayım eksik olabilir veya gerçek hayattaki verilerle tam olarak uyuşmayabilir. Bu gruptaki nitelik vektörlerinin gereken niteliklere olan eksiklik derecelerinin farklılık göstermesi nedeniyle, başarı olasılıkları aynı olmayabilir. Her bir nitelik vektörünün, eksik olan niteliklerine bağlı olarak başarı olasılığı farklılık gösterebilir. İşte GDINA, DINA modelinin geliştirilmiş bir biçimidir. Özellikle, G-DINA modeli, η_{j0} grubundaki tüm nitelik vektörlerinin başarı olasılığının eşit olduğu varsayımını yumuşatır. DINA modeli, η_{j0} grubundaki tüm nitelik vektörlerinin eşit başarı olasılıklarına sahip olduğunu belirtirken G-DINA modeli, bu gruptaki nitelik vektörlerinin farklı yeterlilik derecelerine sahip olabileceğini kabul eder ve bu durumun farklı başarı olasılıklarına yol açabileceğini belirtir (de la Torre, 2011).

G-DINA modeli kullanımını genişleten ve BTM alanında genel bir çerçeve sağlayan olanaklar sunmaktadır. G-DINA modeli, uygun tasarım matrisleri ve dönüşümleri kullanarak diğer bilişsel tanı modelleri için bir temel oluşturmaktadır. Bu, farklı bilişsel tanı modellerinin geliştirilmesine ve bazı yaygın tanılayıcı modellerin tahmin edilmesine olanak tanır. Ayrıca, GDINA'nın sahip olduğu genel çerçeve sınırlı (constrained, reduced) modellerin, tam belirtilmiş (doyurulmuş, unconstrained, saturated) modellerle karşılaştırılarak etkililiğinin değerlendirilmesine yardımcı olur. Bu sebeple, G-DINA modeli, bilişsel tanı modellerinin geliştirilmesi ve uygulanması sürecinde önemli bir araç olarak görülmektedir.

GDINA modelde her örtük grubun bir nitelik vektörü olarak gösterildiği belirtilmiştir. Her bir nitelik vektörü α_{ij}^* olmak üzere, her bir örtük grubun ilgili maddeyi doğru cevaplama olasılığı vardır ve bu olasılık $P(\alpha_{ij}^*)$ ile gösterilmek üzere;

$$P(\alpha_{ij}^*) = \delta_{j0} + \sum_{k=1}^{K_j^*} \delta_{jk} \alpha_{lk} + \sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'} \alpha_{lk} \alpha_{lk'} + \dots + \delta_{j12\dots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{lk}$$

olarak verilir. Aşağıda bu olasılık değerine ait açıklamalar verilmiştir.

δ_{j0} : j maddesine ait kesişim değeri

$\sum_{k=1}^{K_j^*} \delta_{jk} \alpha_{lk}$: j maddesine ait niteliklerin ana etkileri

$\sum_{k'=k+1}^{K_j^*} \sum_{k=1}^{K_j^*-1} \delta_{jkk'} \alpha_{lk} \alpha_{lk'} + \dots + \delta_{j12\dots K_j^*} \prod_{k=1}^{K_j^*} \alpha_{lk}$: j maddesine ait niteliklerin etkileşim etkisi

δ_{jk} : α_k 'den kaynaklanan ana etki

$\delta_{jkk'}$: α_k ve $\alpha_{lk'}$ 'den kaynaklanan etkileşim etkisi

$\delta_{j12\dots K_j^*}$: $\alpha_1, \alpha_2, \dots, \alpha_{K_j^*}$ 'den kaynaklanan etkileşim etkisi

Bu parametreler şu şekilde yorumlanabilir: δ_0 , gerekli hiçbir nitelikte yetkin olunmadığı durumda doğru cevap verme olasılığını yani temel olasılığı temsil eder. δ_k , tek bir nitelikte yetkin olma durumunda doğru cevap verme olasılığındaki değişiklik; $\delta_{kk'}$, birinci dereceden bir etkileşimin etkisini ve α_k ve $\alpha_{lk'}$ niteliklerinde yetkin olma durumunda doğru cevap verme olasılığına etki eden toplam etkiyi ve $\delta_{12\dots K_j^*}$, tüm gerekli niteliklerde yetkin olma durumunda, ana ve düşük dereceli etkileşim etkilerinin toplam etkisi üzerine ek olarak doğru cevap verme olasılığındaki değişiklik anlamındadır (de la Torre, 2011).

Burada şunu belirtmek gerekir ki GDINA modelde üç tür bağlama fonksiyonu (*link function*) ile tanımlama yapılmaktadır ve bunun seçimi kullanılacak olan bilişsel tanı modeli ile ilişkilidir. Bu bağlama fonksiyonları identity link, logit link ve log link fonksiyonlarıdır. Örneğin identity link ile DINA, DINO, A-CDM türü modellerin cevap olasılıkları GDINA çerçevesi ile hesaplanabilirken, logit link ile LLM (C-RUM) modeline ait cevaplama olasılığının log-odds değeri elde edilir ve log link fonksiyonu ile R-RUM modelinde

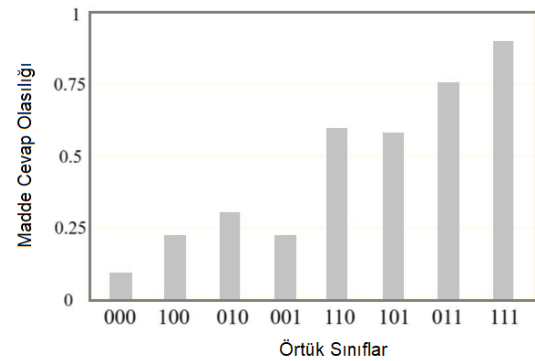
cevaplama olasılığının logaritma değeri verilir. Yukarıda paylaşılan fonksiyon ise GDINA çerçevesinde identity link fonksiyonudur.

Yukarıda bahsedilen sınıf profili ve buna göre hesaplanan başarı olasılığının üç niteliği ölçen bir j maddesi için aşağıda Şekil 1'de örneklendirmesi yapılmıştır.

Şekil 1

GDINA Modeli Örtük Sınıflama ve Cevaplama Olasılıkları Örneği

Sınıf	Sınıf Profili	Cevaplama Olasılığı
1	000	δ_0
2	100	$\delta_0 + \delta_1$
3	010	$\delta_0 + \delta_2$
4	001	$\delta_0 + \delta_3$
5	110	$\delta_0 + \delta_1 + \delta_2 + \delta_{12}$
6	101	$\delta_0 + \delta_1 + \delta_3 + \delta_{13}$
7	011	$\delta_0 + \delta_2 + \delta_3 + \delta_{23}$
8	111	$\delta_0 + \delta_1 + \delta_2 + \delta_3 + \delta_{12} + \delta_{13} + \delta_{23} + \delta_{123}$



Örneğin şeklin sol tarafındaki tabloda 000 sınıf profiline ait doğru cevaplama olasılığı yani kesişim değeri δ_0 kadarken bu olasılık sadece birinci nitelikte yetkin olunan durumda kesişim ve ana etkinin toplamı kadar $\delta_0 + \delta_1$, birinci ve ikinci niteliklerde yetkin olunan durumlarda kesişim, birinci ve ikinci niteliklerin ana etkisi ile bu ikisinin birbiriyle olan etkileşimlerinin toplamı kadar $\delta_0 + \delta_1 + \delta_2 + \delta_{12}$, her üç nitelikte de yetkin olunan durumda kesişim, ana etkiler, ikili etkileşimler ve üçlü etkileşimin toplamı kadardır. Şekildeki grafikte ise bu örneklendirme sayısal olarak yapılmıştır. DINA modelde ise yukarıda verilen aynı koşulda (üç nitelik ölçen bir j maddesi için), ilk yedi sınıf için doğru cevaplama olasılık değeri δ_0 olarak belirlenir ve sadece ilgili maddede ölçülen niteliklerin üçünde de yetkin durumda olduğu gözlenen sekizinci sınıf profilinde doğru cevaplama olasılığı sunulur.

Bilişsel Tanı Modelleri Uygulamasına Yönelik Yapılan Araştırmalar

Bilişsel Tanı Modelleri (BTM) literatürü, çok geniş bir yelpazede çalışma alanlarını kapsamaktadır. Bu modellerin etkin bir biçimde kullanılması ve gelişimi için farklı araştırmalar üzerine yoğunlaşarak ilerlemelerin kaydedildiği görülmektedir. Literatürde BTM'nin tanıtılması, istatistiksel olarak dayandığı temeller, kullanım alanları ve günümüze kadar olan gelişiminin kaleme alındığı kitaplar ve genel incelemeler mevcuttur (Bolt, 2007; de la Torre & Minchen, 2014; DiBello & Stout, 2007; Rupp & Leighton, 2016; Rupp ve diğerleri, 2010; von Davier & Lee, 2019). Aynı zamanda araştırmacılar, BTM'yi ele alırken literatürde bir dizi alt konuya odaklanıldığı görülmektedir. İkili veri tipleri için BTM'ler üzerine yapılan çalışmalar, öğrencilerin sınav performansını daha iyi anlamak için gelişmiş modeller ve yöntemler sunmaktadır (Bradshaw & Templin, 2014; de la Torre & Douglas, 2004; Hong ve diğerleri, 2015). Buna ek olarak, bilişsel tanı çerçevesinde, örneğin yapay sinir ağları, küme analizi, bulanık modelin entegrasyonu, yapay zekanın bilişsel teşhise katkısı gibi konu ve yaklaşımlar da incelenmektedir. Bu yaklaşımlar, öğrencilerin bilişsel süreçlerini daha kapsamlı bir şekilde analiz etmeyi amaçlamaktadır (Briggs & Cinci, 2017; Chiu ve diğerleri, 2009; Liu ve diğerleri, 2018).

Çoktan seçmeli veri tipleri için BTM'ler, test performansının daha karmaşık yapılarını ele almak için önemli bir araştırma alanıdır (Chen & de la Torre, 2018; Culpepper, 2019; Tu ve diğerleri, 2018). Aynı şekilde, çoktan seçmeli nitelikler için BTM'ler de öğrencilerin farklı yeteneklerini değerlendirmede kullanılmaktadır (Chen & de la Torre, 2013; Zhan ve diğerleri, 2019). Bazı araştırmalar, bilişsel tanı modellerini birden fazla stratejiyi içeren durumlar için genişletmeyi amaçlamaktadır (de la Torre & Douglas, 2008; Ma, 2019a). Öte yandan, bazı çalışmalar, öğrencilerin seçenekleri nasıl kullandığını anlamak için seçenekler için özel BTM'ler geliştirmektedir (DiBello ve diğerleri, 2015). Son olarak, boylamsal veriler üzerine yapılan çalışmalar, öğrenci gelişimini ve performansını zaman içinde izlemeyi amaçlamaktadır (Madison & Bradshaw, 2018; Wang ve diğerleri,

2018). Bu arařtırmalar, biliřsel tanı modellerinin çeřitli uygulama alanlarını ve ölçme gereksinimlerini ele alırken sunduđu fırsatları vurgulamaktadır.

BTM'nin istatistiksel gelişimine katkı sunan arařtırmalar model kestirimi ve tanımlanabilirlik gibi önemli konuları içermektedir (Fang ve diđerleri, 2019; Huebner & Wang, 2011; Xu, 2017). Ayrıca, güvenilirlik konusu da BTM'nin etkinliđi açısından literatürde önemle vurgulanmaktadır (Johnson & Sinharay, 2018; Templin & Bradshaw, 2013). Model uyumu ve kiři uyumu deđerlendirmesi de BTM arařtırmalarının merkezinde yer almaktadır (Chen ve diđerleri, 2018; Santos ve diđerleri, 2020; řen & Bradshaw, 2017). Ayrıca, çeřitli biliřsel tanı modelleri arasında çıkarımsal model karşılařtırmaları da yapılmaktadır (de la Torre & Lee, 2013; Sorrel ve diđerleri, 2017). Q-matrix için yöntemlerin incelenmesi de BTM arařtırmalarının odak noktalarından biridir (Chiu, 2013; Lim & Drasgow, 2017; Ma & de La Torre, 2020). Bununla birlikte, BTM'nin uygulama alanları da incelenmektedir. Örneđin, CD-CAT gibi özel uygulamalar üzerine yapılan çalışmalar, biliřsel tanı modellerinin pratikte nasıl kullanılabileceđini göstermektedir (Cheng, 2009; Kaplan ve diđerleri, 2015; Yiđit ve diđerleri, 2019). Ayrıca, BTM'nin çeřitli alanlardaki uygulamaları da incelenmiřtir, bu da bu modellerin geniř bir yelpazede nasıl kullanılabileceđini göstermektedir (Ma & Meng, 2014; Sessoms & Henson, 2018; Tu ve diđerleri, 2017).

BTM çalışmalarının geniř konu yelpazesi içerisinde bu çalışma uygulama ve betimleme üzerine odaklanmıřtır. İlk adım olarak öğrencilerin cebirsel düşünme becerilerine ait yetkinlik düzeylerini gösterir betimsel bir tablo elde edebilmek amacıyla hazırlanan tanılayıcı testlere verdikleri cevaplar incelenmiřtir ve buna göre öğrencilerin belirlenen becerilerdeki yetkinlik durumları belirlenmiřtir. Bu çerçevede de literatürde var olan uygulama ve betimlemeye dönük çalışmaların paylaşılması uygun görülmüřtür.

Bu kapsamda Jurich ve Bradshaw (2014) bir arařtırmasında bir yükseköğretim kurumunun müfredatının bir bölümüne iliřkin öğrenme sonuçlarına yönelik detaylı tanı sağlamak için didaktik olarak yönlendirilmiř bir log-lineer biliřsel tanı modeli (LCDM;

Henson ve diğ erleri, 2009) uygulaması yapılmıřtır. Kullanılan log-lineer biliřsel tanı modeli, lisans öğrencilerinin psikososyal arařtırmaya iliřkin dört öğrenme çıktılarıyla ilgili olarak iki farklı zaman noktasında deęerlendirildięi bir bağlamda kullanılmıřtır. Çalışma, toplam 1710 öğrenciyi kapsamıř olup, öğrenci gelişimini incelemek amacıyla ön test ve son test verilerini kullanmıřtır. Ön test verileri, LCDM kullanılarak madde parametreleri, öğrenci sınıflandırmaları ve psikososyal arařtırma öğrenme sonuçları arasındaki iliřkilerin tahmin edilmesi için analiz edilmiřtir. Ayrıca, en uygun yapısal modelin belirlenmesi hedefiyle farklı log-lineer yapısal modeller istatistiksel olarak karşılařtırılmıř ve nitelikler arasındaki en uygun iliřkiyi temsil eden ikili log-lineer yapısal model seçilmiřtir. Model deęerlendirmeleri ve karşılařtırmaları ön test verileri kullanılarak tamamlanmıřtır. Post test için öğrenci sınıflandırmaları, tüm model parametrelerinin kabul edilen ön test modelinde bulunan deęerlere eşitlenmesiyle tahmin edilmiřtir. Analizlerde kullanılan ikili log-lineer yapısal modele göre ön test ve son test marjinal nitelik yetkinlik ve sınıflandırma oranları sunulmuřtur. Ayrıca, grup düzeyinde ve öğrenci düzeyinde geri bildirim saęlayan tanılama puanı raporu sunulmuřtur. Son olarak, öğrenci gelişimini gözlemek amacıyla ön test ve son testten elde edilen yetkinlik düzeyi sonuçları karşılařtırılmıřtır.

Hu vd. (2021) çalışmalarında bilimsel açıklama kavramına ait sekiz nitelięi belirlemiřler ve bu nitelikler, TIMSS (Uluslararası Matematik ve Fen Eğilimleri Arařtırması) test maddeleriyle kodlanmıřtır. Çin'in Hangzhou şehrindeki 574 dördüncü sınıf öğrencisinden toplanan veriler, ABD, Singapur, Avustralya, Birleřik Krallık ve Rusya'nın verileriyle birleřtirilerek çalışmada kullanılmıřtır. Altı ülkedeki öğrenciler tarafından ortak olarak test edilen 30 TIMSS test maddesi seçilmiřtir. Çalışmada DINA modeli kullanılmıř ve bu yapılan analizlerden elde edilen sonuçlar ışığında uluslararası karşılařtırmalı analiz, bilgi durumu sıralaması ve bilimsel açıklamanın öğrenme yolu ile ilgili olarak üç yönden karşılařtırmalar yapılmıřtır. Bu çalışma, BTM'nin özellikle nitel bir yaklařıma dayanan geleneksel deęerlendirmelere kıyasla öğrencilerin bilimsel açıklama niteliklerindeki yetkinliklerine zamanında tanılayarak geribildirim saęlamak için umut verici

bir araç olduğunu göstermektedir. BTM'nin sağladığı bu sonuçlar, eğitimcilerin eğitim sistemleri, standartları, müfredatları ve önceki değerlendirmeleri gözden geçirmelerine ve kültürel etkileri göz önünde bulundurmalarına olanak tanımaktadır. Bu çalışmanın vurgusu, belirli bir grup öğrencinin belirli bir bağlamda izlediği bilişsel sıralamayı temsil eden öğrenme yolunun, bir müfredat şeklinde sunulan bir disiplinin mantıksal sıralaması ile örtüşmediğidir. Bu çalışmada saptanan sonuçlara göre gözlenen öğrenme yolunun, öğrencilerin öğrenme sürecini yansıttığı ve gelişimlerini teşvik ettiği, disipline dayalı bir öğrenme yolunu yansıtmadığı vurgulanmıştır. Öğrencilerin bireysel sonuçları ve öğrenme yollarından elde edilen sonuçlar bir araya getirildiğinde, BTM yaklaşımıyla oluşturulan öğrenme yolunun, fen öğretmenlerine öğretim materyalleri, stratejiler, etkinlikler ve değerlendirme yöntemleri seçmeleri konusunda rehberlik sağlama potansiyeline sahip olduğu belirtilmiştir.

Jia ve Zhu (2021) yaptıkları araştırmada Uluslararası Öğrenci Değerlendirme Programı (PISA) 2012'de uygulanan sınavdan istatistik konusu temelli dokuz madde seçilmiş ve bu maddeler dört niteliğe dayalı olarak 14 ülkeden 30.092 öğrencinin cevapları temel alınarak, belirsizlik, ortalama, istatistiksel grafik ve veri işleme nitelikleri çerçevesinde Lineer Logistik Model (LLM) ile analiz edilmiştir. Analizlerde işlemler iki adımdan oluşmaktadır. İlk adım, mantıksal durumlarına dayanarak bir bilgi hiyerarşisi oluşturmaktır, ikinci adım ise, bu hiyerarşi boyunca daha sık görülen durumların öğrenme yolları olarak belirlenmesidir. Her ülkedeki öğrencilerin yetkinlik durumuna göre çeşitli nitelik sınıfları bulunmaktadır. Bu nitelik sınıflarına dayalı olarak çıkarılan bilgi durumu hiyerarşik yapısı, öğrencilerin istatistiksel öğrenme yolunu ifade eder. Çalışmada her ülkedeki öğrencilerin her bir nitelik için ortalama yetkinlik düzeyleri sunulmaktadır ve ülkelerin öğrenme yollarına ait görseller paylaşılmaktadır. Araştırma, bazı ülkelerin sadece tek bir öğrenme yoluna sahipken bazılarının ise ikinci bir öğrenme yoluna da sahip olduğunu göstermektedir. 14 ülke arasından 13 farklı ülkede, dört istatistiksel nitelik için nispeten tek tip bir öğrenme yolu mevcut olduğu belirtilmiştir: belirsizlik→veri

işleme→istatistiksel grafik→ortalama. Ayrıca, bu öğrenme yoluyla öğretim, öğrencilerin öğrenmesi için uygun bir yol olduğu tavsiye edilmektedir. Bu öğrenme yolu ile bazı ülkelerdeki ders kitaplarında öğretim sırasının da aynı olduğu belirtilmektedir.

Bradshaw vd. (2014) çalışmasında, psikometri uzmanları ve matematik eğitimi araştırmacılarının iş birliğiyle gerçekleştirilen Öğretmenlerin Çarpma Akıl Yürütme Becerilerini Tanılama (DTMR) Projesi çerçevesinde, öğretmenlerin çarpma akıl yürütme becerileri ve rasyonel sayıları anlama düzeylerini belirlemek ve onlara geri bildirim sağlamak amacıyla bir tanılayıcı test geliştirilmiş ve uygulanmıştır. Bu çalışma, özellikle kesir aritmetiği konusunda akıl yürütme üzerine odaklanmıştır ve test, bu sınırlandırılmış konuya uygun olarak geliştirilmiştir. 990 hizmet içi ortaokul matematik öğretmeninden yanıt veri toplanmıştır. Kesir testi, 28 maddeden oluşmuş ve bunlardan 20'si tek bir niteliği ölçerken, kalan maddeler ise iki niteliği ölçmüştür. Çalışmadaki nitelikler Referans Birimler (RB), Bölme ve Ardışıklama (BA), Uygunluk (UY), Çarpma Karşılaştırma (ÇK) olarak belirlenmiş ve yanıtlar log-lineer bilişsel tanı modeli (LCDM; Henson ve diğerleri, 2009) kullanılarak analiz edilmiştir. Çalışmada hem test maddeleri hem de modele ait uyum indeksleri detaylıca incelenerek testin hedeflenen amaç doğrultusunda ölçme uygulaması yaptığının incelemeleri detaylıca sunulmuştur. Çalışma, hem gruba hem de bireylere ait yetkinlik durumlarını paylaşarak geri bildirim sağlamıştır. Bu araştırma, bir CDM uygulaması olarak geri bildirim sağlama amacını başarıyla yerine getirmiş ve LCDM ile elde edilen tanılanmış sonuçları matematik eğitiminde kesir öğretimi çerçevesinde tartışmıştır.

BTM çerçevesinde test geliştirmeden analiz safhalarına kadar güzel bir örnek proje uygulaması olarak literatürde yer alan "Ulusal Bilim Vakfı CAREER: Bilişsel Tanılayıcı Değerlendirme Projesi için bir Bilişsel Modelleme Yaklaşımı" kapsamında, sekizinci sınıf öğrencileri için orantısal akıl yürütme becerisinin tanılanması amacıyla matematikçiler, matematik eğitimi araştırmacıları ve deneyimli ortaokul öğretmenleri tarafından test formu oluşturulmuştur. Ayrıca, maddelerin oluşturulması, gözden geçirilmesi ve geliştirilmesi

süreci de ayrıntılı bir şekilde açıklanmıştır (Tjoe ve de la Torre, 2013). Öğretmenlerle birlikte maddeler üzerinde çalışarak ve öğrencilerle sesli düşünme görüşmeleri yaparak, orantısal akıl yürütme becerisinin değerlendirilmesi için niteliklerin belirlenmesi ve doğrulama süreçleri derinlemesine tartışılmıştır (Tjoe & de la Torre, 2014). Tanılayıcı değerlendirme kapsamında altı niteliğin ölçülmesi hedeflenmiştir. Bu nitelikler; orantısal akıl yürütme için gereken önkoşul beceriler ve kavramlar, kesirleri karşılaştırma ve sıralama, oranlar ve orantılar oluşturma, değer setleri arasındaki çarpansal ilişkiyi belirleme, orantılı bir ilişkiyi orantısız bir ilişkiden ayırt etme, orantısal akıl yürütme problemlerini çözmede algoritmaları uygulama olarak belirlenmiştir (Tjoe ve de la Torre; 2014). Bu temel çalışmalar neticesinde sekizinci sınıf öğrencilerine tanılayıcı test uygulanmış, elde edilen cevap örüntüleri üzerinden BTM ile MTK model karşılaştırmaları yapılmıştır (Ma ve diğerleri, 2020). Çalışma sonucunda, MTK bulgularına dayanarak aynı seviyedeki iki öğrencinin yetenek düzeyleri, sonuçlarına dayanarak farklı nitelik yetkinlik seviyelerini göstermiştir. Ayrıca, tek boyutlu IRT ile karşılaştırıldığında, DCM daha ayrıntılı bilgi sağlamıştır. Bireysel düzeyde elde edilen bu bilgilere ek olarak, matematik öğretim programına yardımcı olmak için okul veya sınıf düzeyinde bilgiler de sunulmuştur.

BTM'nin gelişimiyle birlikte literatürde geri bildirim sunmak amaçlı daha pek çok araştırmalar mevcuttur. Bu araştırmalar özellikle *Uluslararası Matematik ve Fen Bilimleri Eğilimleri Araştırması (TIMSS, Trends in International Mathematics and Science Study)* ve *Uluslararası Öğrenci Değerlendirme Programı (PISA, Programme for International Student Assessment)* gibi sınavlardan elde edilen cevap örüntüleri tanısal modellemeye uyumlandırılarak (*retrofitting*) hem öğrencilerin ölçülen nitelikler açısından yetkinlik düzeyleri hakkında genel bir betimleme sunabilmek, hem bireysel bilgiler paylaşmak hem de ülkeler arası karşılaştırmalar yapmak amaçlı çalışmalar fazlaca mevcuttur (Birenbaum ve diğerleri, 2005; Dogan & Tatsuoka, 2008; Şen & Arıcan, 2015; Tatsuoka ve diğerleri, 2004; Toker, 2010). Bu çalışmalar, uluslararası düzeyde önemli karşılaştırmalı veriler

sunarak eğitim politikalarının ve öğretim yaklaşımlarının geliştirilmesine katkıda bulunmaktadır.

Test Eşitleme

Farklı alanlarda, genellikle ortak bir olgu üzerinden gelen ölçümleri karşılaştırmak ilgi çekicidir: "*Farklı yöntemler ve araştırmacılar tarafından farklı koşullarda yapılan ölçümlerin karşılaştırılabilirliği, tüm bilim için temel bir ön koşuldur*" (Dorans ve Holland 2000). Test puanlarının çeşitli amaçlar çerçevesinde önemli kararlar almak için kullanıldığı göz önünde bulundurulursa test puanlarının karşılaştırılabilirliği de önemli bir konu olarak karşımıza çıkar. Kararlar bireylere ait çeşitli özelliklerin karşılaştırılması ile alınır ve bu özelliklerin bir puanlama sistemi ile ölçülmeye çalışılması ve bireyler arasındaki karşılaştırmanın puanlar üzerinden yapıyor olması puanların karşılaştırılabilirliğin önemini vurgular. Bu sebeple bir değerlendirme ve karar almada kullanılan puanların adil bir şekilde rapor edilmesi ön plana çıkar. Diğer bir taraftan da ölçme programları kapsamında güvenlik nedenleriyle aynı özelliği ölçmeyi hedefleyen farklı test formları da uygulamaya konulmaktadır. Bu test formlarının mümkün olduğunca benzer özelliklerde olmasına dikkat edilse de test ve madde zorluk derecelerindeki farklar kaçınılmazdır. Bu durum, aynı özelliği ölçmeyi amaçlayan farklı test formlarının bazı katılımcılara daha kolay uygulanmasına yol açabilir. Sonuç olarak, puanların adil bir şekilde değerlendirilebilmesi için test formlarının eşitlenmesi ihtiyacı doğar.

Test eşitleme yöntemi, farklı testlerden elde edilen puanların birbirinin yerine kullanılmasına olanak tanıyan bir süreçtir. Kolen ve Brennan'a (2014) göre, bu süreç, benzer içerik ve zorluk seviyesine sahip test formları arasındaki farklılıkları düzenleyerek gerçekleşir ve test formlarından elde edilen puanların karşılaştırılabilir hale getirilmesini sağlar. Bu süreç yedi aşama ile tarif edilir (Kolen ve Brennan, 2014). İlk aşama eşitlemenin amacının belirlenmesidir. Hangi genel amaç ile eşitleme yapılacağı saptanmalıdır. Belirlenen amaç çerçevesinde ikinci aşamaya geçilerek, benzer içerik ve

istatistiksel özelliklere sahip uygun test formlarının oluşturulmasıdır. Üçüncü aşamada eşitleme deseni belirlenerek eşitlemede kullanılacak verilerin nasıl bir uygulama ile elde edileceği belirlenir ve dördüncü aşamada belirlenen eşitleme deseni doğrultusunda uygun veri toplama işlemi gerçekleştirilir. Beşinci aşama olarak operasyonel tanımların belirlenmesi ve bu çerçevede verilerin uygun biçimde hazırlanmasının sağlanmasıdır. Altıncı aşamada ise geçerli ve güvenilir sonuçlar elde etmek amacıyla eşitlemede kullanılacak istatistiksel kestirim yöntemlerinin seçimi yapılır. Son aşamada olarak elde edilen eşitleme sonuçları değerlendirilmekle birlikte eşitleme sürecinin de etkililiği değerlendirilmeye alınır.

Puan farklarının sadece test formlarının zorluk seviyelerindeki farklılıklardan kaynaklanmadığına dikkat çekilerek, bir gruptaki sınav katılımcısının diğer gruptakinden daha yetenekli olma ihtimalinden de söz edilmektedir. Bu durumda, gruplar arasında test puanları kullanılarak yapılan bir karşılaştırmanın iki gruptaki bireylerin yetenekleriyle karıştırılabileceği belirtilir (González & Wiberg, 2017). Bu durumda Livingston (2004) eğer iki test formundaki puanlar grup içinde aynı göreceli konumu temsil ediyorsa eşdeğeri olarak dikkate alınması gerektiğini söyler. Sonuç olarak test eşitleme sürecinin testlerin zorluk derecesindeki farklılıkları telafi etme amacının yanında eşitleme hedefi çerçevesinde testleri dikkatle oluşturmak, içerik uyumu, test maddelerinin yazımına dair detayların takip edildiği, paralel test formlarının oluşturulmasını kapsayan detaylı süreç olduğu açıktır.

Eşitlemeye ihtiyaç duyulmayan bir ortamda test formları birebir her yönden paralelliği sağlamış ve aynı psikometrik özelliktedir. Fakat böyle bir ortamın imkansızlığının kabulüyle test oluşturma basamaklarından itibaren hassas davranarak testleri mümkün olduğunca paralel formlar haline getirmeye çalışmak ve test puanlarını eşitlemeye hazırlamak önemlidir. Eşitleme literatüründe bir eşitlemenin eşitleme olarak kabul edilebilmesi için gerekli koşullar belirtilmiştir (Dorans & Holland, 2000, Kolen & Brennan, 2004). Bu koşullardan ilki testlerin eşit yapı geçerliliğine sahip olması ve ikincisi

testlerin eşit güvenilirlikte olmasıdır. Aynı yapı ve güvenilirliğe ek olarak Kelecioğlu (1994) testlerin aynı ortalama güçlüğe, test puanlarının aynı ortalama ve varyansa sahip olmalarını belirtir. Bu koşullar testlerin paralelliğine dikkat çeken koşullar olarak karşımıza çıkmaktadır. Dolayısı ile testteki madde sayısından bu maddelerin formatına ve ölçme hedefine kadar her unsurunda bütünlük olması eşitleme açısından önemlidir. Diğer bir koşul, eşitlemede simetrikliktir. Buna göre örneğin birinci test formuna ait puanlar ikinci test formuna ait puanlara dönüştürüldüğünde bunun tersi de doğru olmalıdır, yani eşitleme fonksiyonunun tersi kullanılarak ikinci test formuna ait puanlar birinci test formuna ait puanlara dönüştüğünde eşit olması beklenmektedir. İlk testten 13 puan alan bir öğrencinin eşitlenen ikinci teste göre eşitlenmiş puanı 15 ise ikinci testten 15 alan başka bir öğrencinin eşitlenen birinci teste göre eşitlenmiş puanı 13 olmalıdır. Dördüncü bir koşul ise eşitliktir, bir adayın eşitlenen testlerden herhangi birini alması adaylar açısından bir kaygı ya da şüphe oluşturmamalıdır. Beşinci koşul olarak gruptan bağımsızlık kavramı öne çıkar. Buna göre iki testi birbirine bağlamak için kullanılan eşitleme fonksiyonu grupta bulunan her alt grup için aynı biçimde işlem görmelidir.

Burada literatüre bağlı olarak belirtilen eşitleme koşulları aynı zamanda eleştirilerle de karşılaşmıştır. Von Davier (2007) bu koşulların eleştirilebilir olduğunu belirtmiş, Dorans ve diğerleri (2010) bu koşulların pratik olarak görülmediği, katı ve belirsizliğe neden olduğu görüşleri aktarmıştır. Örneğin eşitlik koşulu Lord (1980) için en temel koşullardan biri olarak belirtilirken bu koşulun Livingston (2004) tarafından pratikte uygulanamaz olduğu aktarılmıştır. Lord (1980) aynı zamanda eşitlik koşulunun aynı yapının ölçülmesi ve aynı güvenilirlik koşullarını kapsadığını belirtmiştir. Ayrıca simetriklik koşulunun regresyon yöntemlerinin eşitleme için kullanımını engellediğine dair eleştirilerde de bulunulmuştur.

Eşitlemede Veri Toplama Desenleri

Tek Grup Deseni

Tek grup deseninde, farklı test formlarının aynı gruba uygulanması oldukça yaygın bir uygulamadır. Bu düzenlemede, her iki testin de aynı evrende uygulandığı varsayılır ve her bir öğrenciye hem X formu hem de Y formu verilir. Bu şekilde, her bir öğrencinin cevapları $(x_i, y_i), i = 1, 2, 3, \dots, n$ biçiminde birer ikili eşleme oluşturur. Ancak, uygulama sırasının test puanlarını etkileyebileceği ve bu durumun sıra etkisi olarak adlandırıldığına dikkat edilmelidir. Öğrencilerin birinci sırada X formunu ve daha sonra Y formunu yanıtlaması veya tam tersi durumunda, ikinci formun birinci formdan etkilenmesi muhtemeldir. Bu etki, ikinci formda yorgunluk veya aşinalık gibi faktörlerle kendini gösterebilir ve bu durumu dengelemek için test formlarının dengelenmesi gerekebilir. Ayrıca, tüm öğrencilerin her iki test formuna da yanıt vermesi istendiği için toplam test süresi oldukça uzun olabilir. Bu düzen, test puanlarının eşitlenmesinde kullanılan basit bir yöntem olmasına rağmen, sıra etkisi gibi faktörlerin dikkate alınması gerekmektedir.

Denk Gruplar Deseni (Eşdeğer gruplar deseni)

Denk gruplar deseni, test eşitleme çalışmalarında sıkça kullanılan bir yöntemdir. Bu desen, farklı test formlarının her birinin temsil edildiği büyük ve heterojen bir grup öğrencinin rastgele iki alt gruba ayrılmasıyla oluşturulur. Bu alt gruplardan birine X formu, diğerine ise Y formu verilir. Bu şekilde elde edilen veriler, her bir test formunun performans düzeyleri arasındaki farkı yansıtır. Ancak, eşit gruplar deseni ile karşılaştırıldığında, her iki grubun da yeteneklerinin eşit olduğu varsayımı geçersizdir. Bu durumda, gruplar arasındaki yetenek farklılıklarını kontrol etmek için özellik ve yeteneklerine göre gruplar oluşturulur. Livingston (2004), bu farklılığı dengelemek için grupların yetenek düzeylerinin birbirine paralel olacak şekilde belirlenebileceğini öne sürmüştür. Bu yaklaşım, gruplar arasındaki farklılıkların kontrol altına alınmasına yardımcı olur.

Denk gruplar deseninde, iki bağımsız grup (G1 ve G2) aynı hedef evrenden örneklenir ve her bir gruba sadece bir test formu uygulanır. Bu şekilde elde edilen veriler iki bağımsız değişken olarak ifade edilir. Ancak, testler farklı zaman noktalarında gruplara uygulandığında ve gruplardaki test katılımcılarının bileşimi zamanla değiştiğinde, denk gruplar deseninin kullanılması bazen problemlere yol açabilir çünkü eşdeğerlik varsayımı geçerli olmayabilir.

Dengelenmiş Grup Deseni

Dengelenmiş grup deseni, test eşitleme çalışmalarında kullanılan önemli bir yöntemdir. Bu desende, tek bir popülasyondan seçilen örneklem, bağımsız iki gruba bölünür ve her bir gruba eşitlenmesi planlanan test formlarından biri uygulanır, ancak sıraları farklıdır. Birinci gruba ilk form, ikinci gruba ise ikinci form önce verilir. Bu şekilde, her iki test formunun performansı incelenir ve sıra etkisi göz önünde bulundurulur. Örneğin, n_{G1} ve n_{G2} test katılımcılarının sırasıyla G1 ve G2 gruplarından seçildiğini varsayalım, elde edilen veriler iki bağımsız bivariate vektörden oluşur: $(x_{1i}, y_{2i}) (i = 1, 2, 3, \dots, n_{G1})$ ve $(x_{2j}, y_{1j}) (j = 1, 2, 3, \dots, n_{G2})$. Bu desen, özellikle küçük örneklerde kullanışlıdır ancak her iki formun da katılımcılar tarafından yanıtlanması gerektiği için uygulama süresi uzun olabilir (González & Wiberg, 2017).

Tek grup deseninin dezavantajlarından biri, testlerin veriliş sırasının bir hata kaynağı oluşturmasıdır. Bu nedenle, dengelenmiş tek grup deseni geliştirilmiş ve bu yöntemle testlerin veriliş sırasından kaynaklanacak hatalar giderilmeye çalışılmıştır. Bu desende, test formlarının sıralarını değiştirerek her iki formun performansı incelenir ve böylelikle sıra etkisi azaltılmış olur ve testlerin güvenilirliği artar (González & Wiberg, 2017; Kolen & Brennan, 2014). Ancak, iki formun da uygulanması zaman açısından bir problem olabilir.

Denk Olmayan Gruplarda Ortak Madde Deseni (DOG-OM)

Denk olmayan gruplarda ortak madde deseni, tek bir oturumda birden fazla test formunun uygulanmadığı durumlarda kullanılır. farklı P ve Q popülasyonlarından iki grup bağımsız olarak örneklenir. Her bir gruba X ve Y test formlarından biri uygulanır ve her iki gruba da ortak bir anchor test formu A uygulanır. Gözlemlenen anchor test puanı a ise, elde edilen veriler $(x_i, a_i)(i = 1,2,3, \dots, n_x)$ ve $(y_j, a_j)(j = 1,2,3, \dots, n_y)$ olmak üzere iki bağımsız vektördür. Eğer ortak maddeler bireyin puanına katkı sağlıyorsa iç ortak test, sağlamazsa dış ortak test olarak adlandırılır (Crocker & Algina, 1986; Kolen & Brennan, 2004). Ortak test, gruplar arasındaki farklılıkları düzenlemek için bütün test formunun içerikte ve istatistiksel olarak temsilcisi olmalıdır. Bu, testin küçük bir versiyonunun kullanılması anlamına gelir (Kolen & Brennan, 2004). Angoff (1984), ortak testin en az 20 maddeden oluşması veya her bir X ve Y formundaki madde sayısının %20'sinden az olmamasını önermiştir. Ortak test kullanıldığında, grupların yetenek düzeyleri farklı olsa bile testler eşitlenebilir. DOG-OM desenine potansiyel bir tehdit, test katılımcılarının anchor maddelerini tanıyabilecekleri ve bu maddelere geri kalan test kadar ciddi yaklaşmayabilecekleri gerçeğidir. Ayrıca, standartlaştırılmış testler kullanan tüm ölçme programlarının anchor testlere sahip olmadığını ve bu nedenle DOM-OM tasarımının bazı durumlarda kullanılamayacağını belirtmek gerekir (Gonzalez & Wiberg, 2017).

Denk Olmayan Gruplarda Ortak Değişken Deseni (DOG-OD)

Bu desen türü, denk olmayan gruplarda ortak madde desenine (DOG-OM) bir alternatif olarak kabul edilir. DOG-OM desenindeki bir sorun, her testte ortak maddelerin bulunmamasıdır. Branberg ve Wiberg (2011), iki test formunu eşitlemek için ortak maddeler dışında ortak değişkenlerin de kullanılabileceğini belirtir. Gruplar arasındaki sistematik farklılıkları en iyi açıklayan ve test ile yüksek ilişkili olan ortak değişkenlerin seçilmesi önerilmektedir. DOG-OM deseninde, farklı P ve Q popülasyonlarından bağımsız olarak iki grup örneklenir ve her birine X ve Y test formlarından biri uygulanır. Bir referans test formunun olmaması durumunda, denk olmayan gruplarda ortak madde (DOM-OD)

deseni, test alan gruplar arasındaki farklılıkları hesaba katabilecek ilgili ortak değişkenleri kullanır. Böylelikle grupların yetenek farkı kontrol altına alınabilir. Ortak değişkenler doğru seçildiğine eşitleme hatası azalarak eşitlemenin doğruluğu artmaktadır. Eşitlenecek gruplara göre yanlılık içeren bir ortak değişken seçildiği durumda ise değerlendirmenin yanlı olması sebebiyle test adaleti sağlamak adına yapılan eşitlemede eşitlenmiş puanların geçerliği düşerek eşitlenmeye zarar verecektir. (Wiberg & Bränberg, 2015).

Klasik Test Kuramına Dayalı Eşitleme Yöntemleri

X ve Y test formları, bu testlerden elde edilen rastgele test puan verileri X ve Y , bu verilerdeki olası test puanları ise x ve y olmak üzere;

X ve Y formları, T hedef popülasyonunda $F_Y(y) = F_{\varphi(x)}(y)$ durumunda $\varphi(x)$ tarafından eşitlenmiş olarak adlandırılır. Buna göre, Y dağılımı ile eşitlenmiş puan dağılımı $\varphi(X)$ 'in aynı olduğu durumlarda x ve y puanlarının eşitlendiği belirtilmiş olur. Böylece, T hedef popülasyonunda X 'i Y 'ye eşitlemede kullanılan φ için açık bir form elde edilmiş olur. Aslında, x ve y 'nin X ve Y test puanlarının dağılımlarında herhangi bir ortak yığılmalı oran p için dağılım dilimleri olduğunu varsayıldığında, yani, $x(p) = F_X^{-1}(p)$ ve $y(p) = F_Y^{-1}(p)$ ise her $p \in [0,1]$ için $F_Y(y) = F_{\varphi(x)}(y)$ gerekliliği ile Y 'deki bir eşdeğer y puanı, X 'teki bir x puanı için şu şekilde elde edilebilir:

$$y = \varphi(x) = F_Y^{-1}(F_X(x))$$

Elde edilen bu eşitlik, eşitliğin genel tanımının matematiksel gösterimidir. Eşitlemede kullanılan farklı yöntemlerin temeli bu tanımdır. Aşağıda KTK'ya göre yapılan eşitleme yöntem ve özellikleri paylaşılmıştır.

Ortalama Eşitleme

Ortalama eşitleme yöntemi, X ve Y puan dağılımlarının yalnızca konum aileleri söz konusu olduğunda veya her iki konum-ölçek dağılımının varyansının eşit olduğu varsayıldığında ortaya çıkar (Gonzalez & Wiberg, 2017). Ortalama eşitleme, X ve Y test formlarının yalnızca puan ölçeği boyunca ortalamalarına (μ_x, μ_y) göre zorluk açısından farklılık gösterdiği varsayımına dayanır. Ortalama eşitleme süreci, X ve Y test formları arasındaki ortalama puan farkını kullanarak eşitleme yapmayı ifade eder (Livingston, 2004). Ortalama eşitleme aşağıdaki eşitlikteki gibi ifade edilmektedir.

$$\varphi(x; \mu_x, \mu_y) = x + \mu_x - \mu_y$$

Örneğin, eğer X formu Y formundan 3 puan daha zor ise, her iki formda yer alan katılımcılar için X formunun, Y formundan 3 puan daha zor olduğu belirtilir (Kolen & Brennan, 2014). Ortalama eşitleme yöntemi, her iki formun tüm katılımcı yanıtlarını göz önünde bulundurduğu için performans düzeyinden bağımsız olarak kullanılabilir. Bu yöntemde, X formundan elde edilen puanların Y formundaki karşılıklarını bulmak için, her iki formun ortalama puanları arasındaki farktan elde edilen sabit, X formundan elde edilen ham puanlara eklenir (Kolen & Brennan, 2014).

Gonzalez ve Wiberg (2017) ortalama eşitlemenin genellikle sadece açıklayıcı amaçlar için kullanıldığını çünkü birçok gerçek test durumunda aşırı basit olabileceğini vurgulamaktadır. Yani, ortalama eşitleme yöntemi, bazı test koşullarını anlamak veya açıklamak için kullanılabilir, ancak gerçek test durumlarında kullanımı sınırlı olabilir çünkü karmaşıklığı yeterince ele almayabilir veya gereksinimleri karşılamayabilir. Bu nedenle, daha karmaşık veya uygun yöntemler genellikle gerçek test durumlarında tercih edilir.

Doğrusal Eşitleme

Doğrusal eşitleme, test formları arasındaki zorluk farklarının tamamen puan rastgele değişkenlerin (veya dağılımlarının) ilk iki momentleriyle, yani test formları X ve Y'den gelen ortalama (μ_x, μ_y) ve standart sapmalar (σ_x, σ_y) ile açıklanabileceğini

varsayar. Doğrusal eşitleme fonksiyonunun genel tanımı aşağıdaki gibi belirtilmiştir (Gonzalez & Wiberg, 2017).

$$\varphi(x, \mu_x, \mu_y, \sigma_x, \sigma_y) = \frac{\sigma_y}{\sigma_x} [x - \mu_x] + \mu_y$$

Doğrusal eşitleme, pratik ve kullanışlı bir yöntem olmasına rağmen bazı problemlerden de bahsedilmektedir. Livingston (2004) özellikle puan aralığı ve sonuçların gruba bağlı olması konularına dikkat çeker. X formunun Y formuna doğrusal eşitleme yöntemi kullanılarak eşitlenmesi sürecinde, X formundaki çok yüksek veya çok düşük puanların eşitlenmiş puanları, Y formunun olası puan aralığının dışına çıkabildiğini belirtir. Örneğin, her biri 80 sorudan oluşan iki test, doğrusal eşitleme yöntemi kullanılarak eşitlendiğinde, eğer X formu Y formundan daha zor ise X formundan 75 puan alan bir bireyin eşitlenmiş puanı, Y formunda 83 puana denk gelebilir. Üst puan sınırı 80 puan olan test formlarından eşitleme sonrası 83 puan alınması, açıklanması zor bir durumdur. Sonuçların gruba bağlı olması durumunda ise farklı güçlük düzeylerine sahip iki formun doğrusal eşitliği, başarılı bir grup ile daha az başarılı bir grupta farklılık gösterebilir. Başarılı gruba ait doğrunun eğimi daha yatay konumda iken başarısız gruba ait doğrunun eğimi daha dik konumda olacaktır. Bu sorunun üstesinden gelmek için heterojen büyük bir çalışma grubu seçilmesi önerilir.

Eşit Yüzdellikli Eşitleme

Eşit yüzdellikli eşitleme işlemi iki aşamadan oluşmaktadır. İlk aşamada, eşitlenmek istenen iki formun puan dağılımlarından yığılmalı frekans dağılımları oluşturulur veya bu dağılımların grafikleri çizilir. İkinci aşamada ise, elde edilen yığılmalı frekanslarda aynı yüzdellik sıraya denk gelen puanlar eşitlenir. Örneğin, belirli bir yüzdellik sıraya denk gelen bir kişinin X formundan aldığı puan, aynı yüzdellik sıraya denk gelen puanın Y formundaki değeri ile eşleştirilir. Eşit yüzdellikli eşitleme dönüşümü, X ve Y'nin sırasıyla F_x ve F_y kümülatif dağılım fonksiyonlarına sahip sürekli rastgele değişkenler olduğunu

varsayarsak, kapalı bir formda yazılabilir. Bu işlem, test formlarının farklı yüzdellik sıralarındaki puanlarının eşitlenmesini sağlar.

$$y = \varphi(x) = F_Y^{-1}(F_X(x))$$

Bu eşitlik en başta verilen eşitliğin genel tanımının matematiksel gösterimi ile aynıdır ve şunu belirtmekte fayda vardır ki bütün eşitleme dönüşümleri bu genel tanımın matematiksel gösteriminin özel durumlarıdır.

Eşit yüzdellikli eşitleme iki aşamalı bir süreçtir. İlk aşamada, eşitlenecek iki formun puan dağılımları incelenir ve yığılmış frekans dağılımları oluşturulur. İkinci aşamada ise, elde edilen bu yığılmış frekanslarda, aynı yüzdellik sıradaki puanlar eşitlenir. Ancak, uygulamada, iki formun ham puanları genellikle aynı yüzdellik sıraya gelmez. Bu sorunu çözmeye dağılımı sürekli hale getirmeyi sağlayan öteleme ve düzgünleştirme yöntemleri kullanılır. Bu yöntemler, dağılımın daha doğru ve güvenilir bir şekilde eşitlenmesini sağlar. Düzgünleştirme yöntemleri, ön-düzgünleştirme (*pre-smoothing*) ve son-düzgünleştirme (*post-smoothing*) olarak ikiye ayrılmaktadır. Ön-düzgünleştirme yöntemleri, eşitleme işlemi gerçekleştirilmeden önce ham puanlara ait frekans dağılımına uygulanır. Bu yöntemler, puan dağılımındaki belirgin dalgalanmaları düzeltmek ve dağılımı daha düzgün hale getirmek için kullanılır. Öte yandan, son-düzgünleştirme yöntemleri, eşitleme işlemi tamamlandıktan sonra elde edilen dönüşümlere uygulanır. Bu yöntemler, eşitleme sonrası elde edilen puan dağılımındaki herhangi bir belirgin dalgalanmayı azaltmak ve sonuçları daha stabilize etmek için kullanılır. Ön-düzgünleştirme ve son-düzgünleştirme yöntemleri, eşitleme sürecinin daha güvenilir ve istikrarlı olmasına yardımcı olur (Kolen ve Brennan, 2014).

Düzgünleştirme yöntemlerinin yanı sıra, puan dağılımlarını düzenlemede kullanılan başka bir yöntem de ötelemedir. Livingston (2004), öteleme yönteminin eşit yüzdellikli eşitlemede puanların her zaman aynı yüzdellik sıraya denk gelmeme problemine tam olarak çözüm sağlamayabileceğini, ancak eşit yüzdellikli eşitleme işlemi için kolaylık sağlayabileceğini belirtmiştir. Öteleme, puanların bir formdan diğerine dönüştürülmesi

sırasında belirli bir miktar artırılması veya azaltılması işlemidir. Bu yöntem, puanların dağılımını daha dengeli hale getirmek ve eşitleme sürecini optimize etmek için kullanılır.

Test Eşitleme Desenlerine Göre Varsayımlar

Eşitleme tasarımlarında ya tek bir ortak T popülasyonundan ya da P ve Q gibi farklı iki popülasyondan seçilen örneklemelerden elde edilen veriler kullanılmaktadır. Tek grup, denk grup ve dengelenmiş grup desenlerinde tek bir örneklem, iki bağımsız örneklem ve ortak bir hedef popülasyonundan iki örneklemde test formlarının sıralarının karıştırılarak kullanılmasıyla elde edilen veriler φ 'nin tahmin edilmesinde kullanılır. Diğer taraftan denk olmayan gruplarda ortak madde veya ortak değişken desenlerinde örneklemeler farklı iki popülasyondan seçildiklerinden dolayı bu iki popülasyondan bir hedef popülasyon tanımlanır. Eşitlemenin genel tanımı için kullanılan $y = \varphi(x) = F_Y^{-1}(F_X(x))$ eşitliğinin tanımı ortak bir T popülasyonu için oluşturulduğundan özellikle denk olmayan gruplarda ortak madde tasarımından elde edilen verilerin eşitlenmesinde bazı varsayımlara ihtiyaç duyulmuştur (Gonzalez & Wiberg, 2017). Bu varsayımlar aşağıda desenlere göre sırasıyla ele alınmıştır.

Tek grup, denk grup ve dengelenmiş grup desenlerinin ortak özelliği, birbirinden farklı olmayan bir popülasyondan örneklemeler alınmasıdır. Bu durum, denk gruplarda tek değişkenli veya tek grup ve dengelenmiş grupta çift değişkenli puan dağılımlarının doğrudan elde edilmesi ve bu dağılımların parametrelerinin doğrudan tahmin edilebilir olmasını sağlar. Dolayısıyla, φ 'nin doğrudan tahmini için kullanılabilir. Denk grup, tek grup ve dengelenmiş grup desenlerinde, eşitleme dönüşümü üzerine çıkarımda bulunmak için özel bir varsayıma gerek yoktur çünkü tüm veriler aynı popülasyondan gelmektedir ve bu nedenle her şey doğrudan gözlemlenebilir ve tahmin edilebilir.

Denk grup, tek grup ve dengelenmiş grup tasarımlarda örneklemeler ortak T popülasyonundan gelirken denk olmayan gruplarda ortak madde deseninde örneklemeler

farklı P ve Q popülasyonlarından gelmektedir. Denk olmayan gruplarda ortak madde deseninde, T hedef popülasyonu, P ve Q 'nun ağırlıklı bir birleşimi olarak oluşturulur ve bu popülasyona sentetik popülasyon da denilmektedir. Bu çerçevede, sentetik popülasyon $w_P + w_Q = 1$ ve $w_P, w_Q \geq 0$ olmak üzere $w_P P + w_Q Q = T$ biçiminde tanımlanmıştır. Bu tanıma dayanarak, X ve Y 'nin ortak T hedef popülasyonu üzerinde $f_{XT}(x)$ ve $f_{YT}(y)$ olarak belirtilen puan dağılımları, ağırlıklandırılmış bir şekilde birleştirilmiş olarak $f_{XT}(x) = w_P f_{XP}(x) + w_Q f_{XQ}(x)$ ve $f_{YT}(y) = w_P f_{YP}(y) + w_Q f_{YQ}(y)$ ile ifade edilir. Burada, $f_{XP}(x)$ ve $f_{XQ}(x)$ sırasıyla P ve Q popülasyonlarında X puanlarının dağılımlarını temsil ederken, benzer şekilde, $f_{YP}(y)$ ve $f_{YQ}(y)$ P ve Q popülasyonlarında Y puanlarının dağılımlarını belirtir. Ancak, denk olmayan gruplarda ortak madde deseninde X test formu yalnızca P popülasyonuna uygulanırken, Y test formu yalnızca Q popülasyonuna uygulandığı için $f_{XQ}(x)$ ve $f_{YP}(y)$ puan dağılımları toplanan verilerden tahmin edilemez. Bu problem için yaygın bir çözüm, hem P hem de Q popülasyonlarına bir bağlama testi, A (*ancor*), uygulamaktır. Gözlemlenmeyen dağılımlar, belirli varsayımlar getirilerek ve sonuç olarak elde edilen koşullu puan dağılımları $f_{XP}(x|a)$ ve $f_{YQ}(y|a)$ 'nin T üzerindeki (marjinal) bağlama puanlarının dağılımı olan $f_{AT}(a)$ üzerinden marjinalize edilmesiyle tahmin edilebilir.

Test Eşitleme Desenlerine Göre KTK'ya Dayalı Eşitleme Yöntemleri

Eşitleme yapılırken tek grup, denk grup veya dengelenmiş grup desenleri kullanıldığında belirli varsayımlara ihtiyaç olmadığı belirtilmiştir. Bu desenlerde elde edilen verilerden direk olarak eşitleme tanımındaki φ tahmin edilebilir. Dolayısıyla, lineer eşitleme denkleminde ihtiyaç duyulan ortalama ve standart sapmalar, ortalama eşitleme denkleminde ihtiyaç duyulan ortalamalar mevcut veriler üzerinden doğrudan belirlenerek ilgili tanımlarda φ tahmininde kullanılabilirler. Yine aynı şekilde eşit yüzdelli eşitleme

dönüşümünde de belirli varsayımlara başvurmadan verilerden eşitleme puanı tahmin edilebilir.

Denk olmayan gruplarda ortak madde deseni için eşitleme yöntemleri, ya eşit yüzdelikli dönüşümlerde gözlemlenmeyen $f_{XQ}(x)$ ve $f_{YP}(y)$ dağılımlarını tahmin etmek için ya da doğrusal dönüşümlerde bunlara karşılık gelen μ_{XQ} , σ_{XQ}^2 , μ_{YP} ve σ_{YP}^2 konum-ölçek parametrelerinin tahmin edilmesi için farklı varsayımlar kullanılır. Farklı varsayımlara dayanan bu eşitleme teknikleri doğrusal ve eşit yüzdelikli eşitleme yöntemlerine göre aşağıda sunulmuştur (Gonzalez & Wiberg, 2017).

Denk Olmayan Gruplarda Ortak Madde Deseninde KTK'ya Dayalı Eşitleme

Yöntemleri

Denk olmayan gruplarda ortak madde (DOG-OM) deseninde doğrusal veya eşit yüzdelikli eşitleme dönüşümleri ile kullanılabilen iki yaygın yaklaşım zincirleme eşitleme (*chained equating*) ve frekans tahmini (*frequency estimation*) eşitlemedir. Zincirleme eşitleme, X'i Y'ye ortak maddeler (*Anchor*, A) aracılığıyla eşitler ve T üzerinde tanımlanan puan dağılımlarını elde etmek için sentetik bir popülasyon kavramını kullanmaz. X testindeki puanları A'dakilerle ilişkilendiren $\varphi_A(x)$ dönüşümü ve A'daki puanları Y testindeki puanlarla ilişkilendiren $\varphi_Y(a)$ dönüşümü olmak üzere, X'ten Y'ye puan dönüşümünü zincirleme eşitleme $\varphi(x) = \varphi_Y(x) = \varphi_Y(\varphi_A(x))$ eşitliğini kullanarak yapar. $\varphi_A(x)$ dönüşümü P üzerinde ve $\varphi_Y(a)$ dönüşümü Q üzerinde tahmin edildiği için $\varphi_Y(x)$ dönüşümünün hangi popülasyonda geçerli olduğu belirsizdir. $\varphi_Y(x)$ 'nin ortak bir popülasyon üzerinde tanımlı geçerli bir eşitleme dönüşümü olması için bazı varsayımlara ihtiyaç duyulmaktadır. Diğer yandan, frekans tahmini yaklaşımı, eşitleme dönüşümünün oluşturulduğu T'deki X ve Y'nin puan dağılımlarını elde etmek için sentetik popülasyondan yararlanır. Bu yaklaşımda kullanılan tüm puan dağılımları eldeki verilerden tahmin edilemez ve T üzerinde tanımlanan eşitleme dönüşümünü elde etmek için bazı varsayımlar gerekir.

DOG-OM deseninde doğrusal eşitleme yapmak için yukarıda da verilen $\varphi(x, \mu_x, \mu_y, \sigma_x, \sigma_y) = \frac{\sigma_y}{\sigma_x} [x - \mu_x] + \mu_y$ denkleminin kullanılabilmesi için önce T hedef popülasyonu üzerinde tanımlanan X ve Y puan dağılımlarının ortalama ve standart sapmalarının belirlenmesi gerekir. Sentetik popülasyondaki parametrelerin belirlenmesi için aşağıdaki eşitlikler kullanılır.

$$\mu_{XT} = w_P \mu_{XP} + w_Q \mu_{XQ}$$

$$\mu_{YT} = w_P \mu_{YP} + w_Q \mu_{YQ}$$

$$\sigma_{XT}^2 = w_P \sigma_{XP}^2 + w_Q \sigma_{XQ}^2 + w_P w_Q [\mu_{XP} - \mu_{XQ}]^2$$

$$\sigma_{YT}^2 = w_P \sigma_{YP}^2 + w_Q \sigma_{YQ}^2 + w_P w_Q [\mu_{YP} - \mu_{YQ}]^2$$

Daha önce belirtildiği gibi, DOG-OM deseninde iki gruptan elde edilen cevap örüntülerinden tüm parametreler doğrudan tahmin edilemez. Örneğin, μ_{XQ} doğrudan tahmin edilemez çünkü Q popülasyonuna yalnızca Y formu uygulanmıştır. Benzer şekilde, σ_{YP}^2 de tahmin edilemez çünkü P popülasyonunda yalnızca X formu uygulanmıştır. Aşağıda açıklanan doğrusal ve eşit yüzdellikli eşitleme yöntemleri doğrudan tahmin edilemeyen parametrelerin elde edilmesi için farklı varsayımlar kullanarak eşitlemede kullanılır.

Tucker Eşitleme Yöntemi. Tucker eşitleme iki varsayıma dayanır:

- (i) $E_P(X|A) = E_Q(X|A)$ ve $E_P(Y|A) = E_Q(Y|A)$; yani her iki popülasyonda da A'ya bağlı olarak X'in ve Y'nin koşullu beklentileri eşittir.
- (ii) $Var_P(X|A) = Var_Q(X|A)$ ve $Var_P(Y|A) = Var_Q(Y|A)$; yani her iki popülasyonda da A'ya bağlı olarak X'in ve Y'nin koşullu varyansları aynıdır.

Geleneksel olarak, doğrusal regresyon modellerinde yapılan varsayımlara benzer biçimde koşullu beklentilerin doğrusal ve koşullu varyansların sabit olduğu varsayılır. Bu nedenle, yukarıdaki iki varsayım eğimlerin, kesme noktalarının ve varyansların her iki

doğrusal regresyon için (X'in A'ya göre ve Y'nin A'ya göre regresyonu) aynı olması sağlanarak eşdeğer şekilde yeniden formüle edilebilir. Buna göre, P ve Q'daki regresyonlarda δ kesişim ve γ eğim parametreleri olmak üzere;

$$\delta_P = \mu_{XP} - \gamma_P \mu_{AP}, \quad \gamma_P = \frac{\sigma_{XP,AP}}{\sigma_{AP}^2}$$

$$\delta_Q = \mu_{XQ} - \gamma_Q \mu_{AQ}, \quad \gamma_Q = \frac{\sigma_{YQ,AQ}}{\sigma_{AQ}^2}$$

verilir.

Herhangi iki rastgele değişken Z ve W arasındaki kovaryansı temsil eden $\sigma_{Z,W}$ ile $\gamma_P = \gamma_Q$ ve $\delta_P = \delta_Q$ yapılarak sentetik hedef popülasyonunun ortalamaları ve varyansları

$$\mu_{XT} = \mu_{XP} - w_Q \gamma_P [\mu_{AP} - \mu_{AQ}]$$

$$\mu_{YT} = \mu_{YQ} - w_P \gamma_Q [\mu_{AP} - \mu_{AQ}]$$

$$\sigma_{XT}^2 = \sigma_{XP}^2 - w_Q \gamma_P^2 [\sigma_{AP}^2 - \sigma_{AQ}^2] + w_P w_Q \gamma_P^2 [\mu_{AP} - \mu_{AQ}]^2$$

$$\sigma_{YT}^2 = \sigma_{YQ}^2 - w_P \gamma_Q^2 [\sigma_{AP}^2 - \sigma_{AQ}^2] + w_P w_Q \gamma_Q^2 [\mu_{AP} - \mu_{AQ}]^2$$

biçiminde tanımlanabilir. Burada popülasyon P'de X'in A'ya regresyonu için eğim γ_P ve popülasyon Q'da Y'nin A'ya regresyonu için γ_Q eğimdir. Böylece bu eğim değerleri yukarıdaki eşitlikte kullanılarak ortalama ve varyanslar belirlenir ve sonrasında bu değerler eşitleme dönüşümü hedefiyle genel doğrusal eşitleme denkleminde kullanılır.

Levine Gözlenen Puan Eşitleme Yöntemi. Bu yöntemin temelleri KTK'nın gerçek puan ve hata puanları toplamının gözlenen puana eşitliği ilkesine dayanmaktadır. Levine gözlenen-puan eşitleme yönteminde, varsayımlar gerçek puanlar üzerinden yapılırken, pratikte eşitleme için gözlenen test puanları kullanılmaktadır. Teorik açıdan gerçek puanların incelendiği varsayımların, pratikte gözlenen puanlar üzerinden çalışılmaktadır. Birinci varsayıma göre, test formları X ve Y'deki gerçek puanlar ile ortak test formu A'daki gerçek puanlar arasındaki korelasyon 1'dir, yani mükemmel bir ilişki vardır. İkinci

varsayıma göre, test formları X ve Y'deki gerçek puanların, ortak test formu A üzerindeki gerçek puanlar üzerine lineer regresyonun katsayıları aynıdır. Üçüncü varsayıma göre ise farklı popülasyonlarda ölçme hatası varyansı, test formları X, Y ve A için aynıdır. Belirtilen varsayımlar kullanılarak, γ eğimleri gerçek-puan standart sapmaları ile tanımlanmaktadır ancak pratikte bu gerçek-puan standart sapmaları doğrudan gözlemlenmemektedir. KTK varsayımlarına dayalı olarak da belirlenebilen γ terimleri klasik kongenerik modelden türetilmiş olarak aşağıda paylaşılmıştır (Gonzalez & Wiberg, 2017).

$$\gamma_P = \frac{\sigma_{XP}^2}{\sigma_{XP,AP}}, \quad \gamma_Q = \frac{\sigma_{YQ}^2}{\sigma_{YQ,AQ}}$$

Burada bulunan γ değerleri yukarıda Tucker eşitleme yönteminde verilen eşitliklerde yerine koyularak ortalama ve varyans değerleri belirlenerek genel doğrusal eşitleme formülünde kullanılır.

Levine Gerçek Puan Eşitleme Yöntemi. Bu eşitleme yöntemi varsayımları yukarıda Levine gözlenen puan eşitleme yönteminde belirtilen varsayımlar ile aynıdır. Kavramsal olarak tek fark gözlenen puan yerine gerçek puanların kullanılmasıdır. Levine gerçek puan eşitlemesi için aşağıdaki eşitlik kullanılmaktadır.

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

Zincirleme Doğrusal Eşitleme. Zincirleme eşitleme yöntemi, P ve Q popülasyonlarında test puanları ve ortak madde puanları için ortalama ve standart sapmalar hesaplanır ve sonrasında X'i A'ya bağlayan ve A'yı Y'ye bağlayan tahmini doğrusal dönüşümler birleştirilerek gerçekleştirilir. Bu dönüşümde kullanılan eşitlik

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

olarak verilmiştir.

DOG-OM desenlerinde eşit yüzdellikli eşitlemede kullanılan iki yöntem daha önce bu bölümün başında da belirtildiği gibi zincirleme (*chained*) ve frekans tahmini (*frequency estimation*) eşitleme yöntemleridir. DOG-OM desenlerinde eşit yüzdellikli eşitlemede kullanılan bu yöntemler kısaca aşağıda belirtilmiştir.

Frekans Tahmini Eşitleme. Frekans tahmini eşitleme yöntemi, T hedef popülasyonu içindeki X ve Y puan dağılımlarını tahmin etmek için bir ortak madde testi kullanır. Yapılan tahminlerden elde edilen puan dağılımlarında yüzdellik sıraları belirlenir ve test formları, eşit yüzdellikli eşitleme kullanılarak birbirine eşitlenir. Ortak madde test puanları (*anchor*) A iken X ve Y puanlarının P ve Q popülasyonlarındaki koşullu dağılımları sırasıyla $f_{XP}(x|a)$, $f_{YP}(y|a)$, $f_{XQ}(x|a)$ ve $f_{YQ}(y|a)$ ve ortak madde test puanlarına ait marjinal dağılımlar $f_{AP}(a)$ ve $f_{AQ}(a)$ olarak verilsin. Buna göre T hedef popülasyonda marjinal puan dağılımlarının tanımlanabilmesi için gerekli iki varsayıma göre $f_{XP}(x|a) = f_{XQ}(x|a)$ ve $f_{YP}(y|a) = f_{YQ}(y|a)$ koşullu puan dağılımlarına ait eşitlikler verilir. Bu varsayımlara dayanarak

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

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

eşitliklere ulaşılır.

Zincirleme Eşit Yüzdellikli Eşitleme. Zincirleme eşit yüzdellikli eşitlemede X ve Y test puanları ortak maddeler aracılığı ile bağlanır veya zincirlenir. X testi öncelikle A ortak madde testine (*anchor*) sonrasında ise A ortak madde testi Y testine eşitlenir. Bu eşitleme iki varsayıma i) $\varphi_{AP}(x) = \varphi_{AQ}(x) = \varphi_A(x)$ ve ii) $\varphi_{YP}(a) = \varphi_{YQ}(a) = \varphi_Y(a)$ dayandırılarak

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

dönüşümü ile gerçekleşir.

DOG-OM desenlerinde ayrıca doğrusal ve eşit yüzdellikli eşitleme yöntemlerinin her ikisine de dayanan hibrit bir yöntem olan Braun-Holland Eşitleme yöntemi de bulunmaktadır. Bu yöntem frekans tahmini yönteminin doğrusal bir versiyonudur. Denklemi eşit yüzdellikli eşitlemede verilen $f_{XT}(x)$ ve $f_{YT}(y)$ sentetik hedef dağılımlarının ortalama ve standart sapma tahminlerine dayanır.

$$\mu_{XT} = \sum_x x f_{XT}(x), \quad \sigma_{XT}^2 = \sum_x [x - \mu_{XT}]^2 f_{XT}(x)$$

$$\mu_{YT} = \sum_y y f_{YT}(y), \quad \sigma_{YT}^2 = \sum_y [y - \mu_{YT}]^2 f_{YT}(y)$$

Bu eşitlikler aracılığı ile bulunan ortalama ve standart sapma değerleri ise genel doğrusal eşitleme fonksiyonunda yerine koyularak Braun-Holland eşitleme yöntemi gerçekleştirilir.

Klasik Yöntemlerde Eşitleme Hataları

Eşitleme dönüşümleri, farklı testler arasında puanları eşitlemek için kullanılan istatistiksel yöntemlerdir. Ancak, bu yöntemlerin kullanımı sırasında bazı hatalar meydana gelebilir. Bu hatalar iki türde incelenir: sistematik hatalar ve rastgele hatalar. Sistematik hatalar tahminlerin gerçek değerden ne kadar sapma gösterdiğini belirler. Eşitleme dönüşümleri kullanılırken, tahminlerin gerçek puanlardan sürekli bir şekilde sapma göstermesi durumunda sistematik bir hata meydana gelir. Örneğin, bir eşitleme yöntemi, genellikle belirli bir grup için diğerlerinden daha yüksek veya düşük puanlar tahmin ediyorsa, bu sistematik bir hataya işaret edebilir. Rastgele hatalar tahminler arasındaki değişkenliği ölçer. Eşitleme dönüşümleri uygulandığında, aynı yöntemi farklı veri setlerine veya örneklere uyguladığınızda elde edilen sonuçlar arasında değişkenlik olabilir. Bu değişkenlik, rastgele hata olarak adlandırılır ve tahminlerin güvenilirliğini etkiler. Bu nedenle, eşitleme dönüşümleri kullanılırken, sistematik hataların belirlenmesi için yanlılık (*bias*) gözlemlenmeli ve rastgele hataların belirlenmesi için bootstrap standart hatalar ile kök ortalama kare hatası (*Root Mean Squared Error, RMSE*) hesaplanmalıdır. Bu adımlar, eşitleme yöntemlerinin doğruluğunu ve güvenilirliğini değerlendirmede yardımcı olur.

Eşitlemenin Bootstrap Hatası. Bootstrap hata kavramı bir popülasyondan seçilen çok sayıda farklı örneklerde yapılan eşitlemelerden her bir ham puan için elde edilen eşitlenmiş puanların oluşturduğu dağılımın standart sapması, o puan değeri için eşitlemenin standart hatasını temsil eder. Başka bir deyişle, farklı tekrarlı eşitlemelerde eşitleme sonuçlarının birbirinden ne kadar farklı olduğunu gösteren bir ölçüdür. Ne kadar düşükse, eşitleme işleminin o puan için o kadar tutarlı olduğu anlamına gelir. Ancak, gerçek hayatta genellikle sadece bir örnek veri kümesine erişim bulunmaktadır ve bu nedenle eşitleme standart hatasını tahmin etmek için kullanılacak yöntemlere ihtiyaç duyulmuştur. İşte burada bootstrap yöntemi devreye girer. Bu yöntem, gözlemlenen örnek veri kümesinden tekrarlı çekimlerle birçok rastgele örneklemin kullanılmasıyla eşitlemenin standart hatasının tahmin edilmesini sağlar. Özellikle eşitleme bağlamında son derece değerli olan bu yöntem, standart hatalar için açık formüllerin türetilmesinin zor veya imkansız olduğu durumlarda boşluğu doldurur. L , tekrarlı alınan toplam örneklem sayısı ve $\hat{\varphi}_l(x_i)$, l . örneklemdeki x_i test puanına ait eşitlenmiş puan olmak üzere, bootstrap eşitleme hatası

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

olarak verilmiştir.

Yanlılık (Bias) ve RMSE. Yukarıda verildiği gibi L , tekrarlı alınan toplam örneklem sayısı, $\hat{\varphi}_l(x_i)$, l . örneklemdeki x_i test puanına ait eşitlenmiş puan ve $\varphi(x_i)$ gerçek puan olmak üzere,

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

verilmiştir. Diğer taraftan yanlılığın karesi ve standart hatanın karesinin toplamının karekökü RMSE değerini vermektedir. Buna göre RMSE değerlerini belirlemek için

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

$$RMSE(x_i) = \sqrt{\frac{1}{L} \sum_{l=1}^L [\hat{\varphi}_l(x_i) - \varphi(x_i)]^2}$$

eşitlikleri kullanılmaktadır.

Kernel Eşitleme

Kernel eşitleme, von Davier vd. (2004) belirttiği gibi, doğrusal eşitlemeyi özel bir durum olarak içeren esnek bir eşit yüzdelikli eşitleme benzeri fonksiyonlarını temel alan test eşitleme için birleştirici bir yaklaşımdır. Bu yaklaşım, farklı testler arasında eşitleme yaparken kullanılan çeşitli işlevleri birleştirir. Bu yöntem, yarı-parametrik eşitleme tahmincileri kategorisine girer. Tahmin için hem sonlu parametreler hem de dağılım fonksiyonları kullanılır (Gonzalez & Wiberg, 2017). Kernel eşitlemenin parametrik yönü, x ve y ile temsil edilen iki test formunun puan dağılımlarındaki puan olasılığı parametrelerini içerir, bu parametreler genellikle çoklu nominal dağılımın parametreleri olarak modellenir. Öte yandan, kernel eşitlemenin nonparametrik yönü, kernel düzleştirme tekniklerini kullanarak puan dağılım fonksiyonlarını tahmin etmeyi içerir. Kernel düzgünleştirme, bir rastgele değişkenin gözlemlenen veri noktalarına dayanarak olasılık yoğunluk fonksiyonunu tahmin etme yöntemidir. Bu yöntem, yakındaki veri noktalarına ağırlıklar atayarak temel dağılımı tahmin etmeyi içerir. Kernel eşitlemede kullanılan eşitleme dönüşümü, r ve s olarak temsil edilen puan olasılığı parametrelerini ve h_x ve h_y olarak adlandırılan bant genişliği yani süreklileştirmede düzgünleştirme derecesini kontrol eden parametrelerini içeren aşağıdaki eşitleme dönüşümünü temel alır.

$$\hat{\varphi}(x; r, s) = F_{h_y}^{-1}(F_{h_x}(x; \hat{r}); \hat{s}) = \hat{F}_{h_y}^{-1}(\hat{F}_{h_x}(x))$$

Özetlemek gerekirse, kernel eşitleme, farklı test formları arasında puanları eşitlemek için parametrik ve nonparametrik yaklaşımları birleştirir, böylece bir testin farklı versiyonları arasında puanları eşitleme konusunda esneklik ve doğruluk sağlar.

Kernel eşitleme von Davier vd. (2004) tarafından beş aşamada özetlenmiştir: ön-düzgünleştirme, puan olasılıklarının tahmini, süreklileştirme, eşitleme dönüşümünün hesaplanması, eşitleme doğruluğunun hesaplanması. İlk aşamada öncelikle bir eşitleme tasarımından elde edilen ham verilere uygun istatistiksel modeller uygulayarak, genellikle log-lineer olan uygun modellerle tek değişkenli veya iki değişkenli puan olasılıklarını önceden düzgünleştirir. Eşitlemede kullanılacak verilerinin toplanması aşamasında ilgili popülasyonlardan belirli bir desene göre örneklemeler belirlenir. Elde edilen veriler ya denk grup desenlerinde olduğu gibi bağımsız puan rastgele değişken çiftleri ya da tek grup, dengelenmiş grup, denk olmayan gruplarda ortak madde desenlerindeki gibi iki değişkenli rastgele vektörler şeklindedir. Örneklem büyüdükçe, puan dağılımının popülasyonu daha iyi temsil edeceği düşünülür. Ancak, örneklem hatası nedeniyle, bu dağılımlarda genellikle düzensizlikler görülür ve kesikli dağılımların uygun bir şekilde tahmin edilmesi gerekir. Bu düzensizlikleri azaltmanın yaygın bir yolu, belirli bir veri toplama tasarımından elde edilen puan dağılımlarını ön-düzgünleştirmektir. Matematiksel olarak, ön-düzgünleştirme, puan dağılımlarını modelleyerek puan olasılıklarını model parametreleri ile ilişkilendirme anlamına gelir. İkinci adım, eşitleme desenine göre bir desen fonksiyonu (*design function*, DF) ile hedef popülasyonda puan olasılıklarını tahmin etmektir. Kernel eşitlemede ham veriler ve log-lineer modellerle önceden düzgünleştirilen veriler bir matriste saklanır. Her sütun ve satır, tek grup deseni, denk grup ve denk olmayan gruplarda ortak madde deseni için iki test formunda olası bir puanı temsil eder. Ancak, daha sonraki işlemde giriş bir olasılık vektörüdür. Bu nedenle, DF, iki test formunun birleşik puan dağılımını marjinal bir haline dönüştürmek için kullanılan bir matristir. Özellikle, eğer veriler rastgele denk grup deseninde toplanmışsa, tek değişkenli log-lineer bir modelle daha fazla dönüşüme gerek yoktur ve DF burada bir birim matris görevindedir. Ancak, veriler başka desenlerde toplanırsa, daha sofistike bivaryant modeller kullanılır. Bu nedenle, bir olasılık vektörü elde etmek için, yalnızca 1 ve 0 içeren karmaşık matrislere ihtiyaç vardır. Üçüncü adım, test puanları için kesikli kümülatif dağılım fonksiyonlarının kernel düzgünleştirme teknikleri ile sürekli hale getirildiği bir süreçtir. Bu süreç, kesikli puan değişkeni, kernel

düzgünleştirmeyi karakterize eden sürekli bir rastgele değişken ve düzgünleştirme derecesini kontrol eden bir parametre olmak üzere üç bileşenin birleşimi olan bir sürekli rastgele değişken aracılığıyla gerçekleştirilir. Dördüncü adım, kernel eşitleme çerçevesi altında tanımlanan genel eşit yüzdelikli eşitleme fonksiyonu ile test formlarını eşitlemektir. Bu adımda, test formları arasında eşitleme yapmak için belirlenen genel eşit yüzdelikli eşitleme fonksiyonu kullanılır, bu da kernel eşitleme çerçevesi altında tanımlanmıştır. Son olarak, eşitleme fonksiyonları arasındaki eşitlemenin standart hatası (*SEE*) ve eşitleme farkının standart hatası (*SEED*), kernel eşitleme performansını değerlendirmek için kriter olarak hesaplanır (von Davier vd., 2004). Diğer eşitleme yöntemlerini değerlendirirken olduğu gibi, *SEE*, örneklem verileriyle popülasyon parametrelerini yorumlamanın neden olduğu rastgele bir hatanın göstergesidir. *SEED*, kernel eşitlemede belirgin bir kriterdir ve iki kernel eşitleme fonksiyonu arasındaki farkların standart sapmasını gösterir. Von Davier ve diğerleri (2004) tarafından yapılan bir çalışmaya göre, $-2SEED$ ve $+2SEED$ arasındaki kernel eşitleme farkları, daha çok fonksiyonlar yerine örneklem belirsizliğinden kaynaklanmaktadır. Puan dağılımlarının ön düzleştirilmesi ve devamlı hale getirilmesinin avantajlarına dayanarak, kernel eşitlemenin özellikle geleneksel olanlar gibi diğer eşitleme yöntemlerine kıyasla eşitleme doğruluğu açısından eşdeğer veya daha iyi olduğu kanıtlanmıştır. (Arıkan ve Gelbal, 2018; Wang ve diğerleri, 2020).

X formunun Y formuna eşitlenmesinde kullanılan eşitleme dönüşümünün tahminindeki belirsizliği ölçen *SEE* şu şekilde gösterilir.

$$SEE_Y(x) = \hat{\sigma}_Y(x) = \sqrt{Var(\hat{\varphi}(x))} = \sqrt{Var(\hat{\varphi}(x; \hat{\mathbf{r}}, \hat{\mathbf{s}})}$$

Farklı kernel eşitleme dönüşümlerini karşılaştırmak için *SEED* kullanılabilir. $\hat{\varphi}_1$ ve $\hat{\varphi}_2$, iki kernel eşitleme dönüşümünü temsil etmek üzere *SEED*, iki *SEE* vektörü arasındaki farkın Öklid normu olarak aşağıdaki gibi tanımlanır:

$$SEED_Y(x) = \sqrt{Var(\hat{\varphi}_1(x) - \hat{\varphi}_2(x))}$$

Farklı eşitleme dönüşümlerini karşılaştırmanın başka bir yolu, puan dağılımlarının momentlerini kullanmaktır. Yüzde göreceli hata (*Percent Relative Error, PRE*), eşitlenen değerlerin dağılımının momentlerini, puanların eşitlendiği dağılımın momentleri ile karşılaştırır. *PRE*'nin amacı, birkaç momenti karşılaştırarak eşitlenmiş puanların gözlemlenen dağılıma ne kadar iyi uyduğunu ölçmektir. Yüksek *PRE* değerleri, daha az etkili bir puan eşitlemeyi işaret eder. $PRE(p)$, eşitlenmiş puanların gerçek dağılıma ne kadar yakın olduğunu ölçmek için kullanılan bir ölçüdür. *PRE*'nin genel olarak p . momenti, $PRE(p)$, $\mu_p(\varphi(X)) = \sum_j (\varphi(x_j))^p r_j$ ve $\mu_p(Y) = \sum_k (y_k)^p s_k$ olmak üzere;

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

olarak tanımlanır (von Davier ve diğerleri, 2004).

Test Eşitleme ile İlgili Araştırmalar

Test eşitleme, eğitimde ve psikometride önemli bir konudur ve farklı testler arasında geçerliliğin, güvenilirliğin ve karşılaştırılabilirliğin sağlanması için kritik bir süreçtir. Bu süreç, farklı zamanlarda veya farklı gruplar arasında uygulanan testlerin sonuçlarının doğru bir şekilde yorumlanmasını sağlamak amacıyla kullanılır. Test eşitleme araştırmaları, farklı eşitleme yöntemlerinin karşılaştırılması, yeni eşitleme tekniklerinin geliştirilmesi, eşitleme sürecinin zorluklarının ele alınması ve eğitimde kullanılan ölçme araçlarının daha adil ve güvenilir bir şekilde kullanılmasını sağlamak için önemlidir. Bu sebeple, burada eğitimde test eşitleme ile ilgili yapılan araştırmalara kısa bir bakış sunulmuştur.

Livingston ve Kim (2010) tarafından yapılan bir çalışmada, rastgele gruplar deseninde eşitleme için beş yöntem incelenmiştir. Bu çalışmada, 400, 200, 100 ve 50 test katılımcısından oluşan örneklemelerle yeniden örnekleme çalışmaları yapılmıştır. Dokuz bin veya daha fazla test katılımcısına uygulanan altı test formu madde havuzu olarak kullanılmış ve buradan eşitlenmiş form çiftleri oluşturmak için kullanılmıştır. Kriter

eşitleme olarak tüm test katılımcılarının grubunda doğrudan eşit yüzdelli eşitleme seçilmiştir. İncelemede düzgünleştirilmiş dağılımların eşit yüzdelli eşitlemesi, doğrusal eşitleme, ortalama eşitleme, simetrik daire yayı eşitleme ve basitleştirilmiş daire yayı eşitleme yöntemleri karşılaştırılmıştır ve daire yayı yöntemleri en doğru sonuçları üretmiştir. Daire yayı yöntemleri örneklem ne kadar küçükse, diğer yöntemlere göre doğrulukta daha büyük bir avantaja sahiptirler. Özellikle yüksek puan seviyelerinde eşitleme için son derece doğru sonuçlar vermişlerdir. Ortalama eşitleme, ortalama ve alt ortalama puanlar için iyi performans göstermiştir, ancak yüksek puanlar için gösterememiştir. Doğrusal eşitleme ve eşit yüzdelli eşitleme, 25. yüzdeliğin altındaki veya 90. yüzdeliğin üstündeki puanlar için kötü performans göstermiştir. Birim (identity) kullanımı, bu çalışmalar için oluşturulan test formlarını eşitlemek için iyi bir alternatif olmamıştır. Küçük örneklem verileri ile eşitleme yapmak gereken test programlarında, madde zorluk derecelerinin doğru tahminleri için gerekli büyük örneklem öntest verileri genellikle bulunmamaktadır. Bu nedenle, birim (identity) bu test programlarında puan eşitlemek için yeterli bir alternatif olacağı pek olası görünmemektedir.

von Davier vd. (2004) tarafından gözlenen puan eşitleme yöntemleri olan CE ve PSE yöntemlerinin NEAT tasarımında ilişkisi incelenmiştir. CE ve PSE yöntemlerinin aynı sonuçları verdiği ideal koşullarını belirterek ortak madde testinin operasyonel testlerle olan korelasyonları ve alt popülasyonların ortak madde test puanlarında ne kadar farklı olduğu üzerinde vurgu yapmışlardır. İlk bulguları, ortak madde testinin hem P hem de Q popülasyonlarında aynı dağılıma sahip olduğu durumu kapsar (X ve Y 'nin ne kadar benzer olduğu veya ortak madde testinin "testin minyatür versiyonu" olup olmadığına dair ek varsayımlar olmaksızın). Bu durumda, hem CE hem de PSE'nin tam olarak aynı eşitlemeyle sonuçlanacağı belirtilmiştir. Ortak madde testi tarafından ölçülen yeteneklerin dağılımı açısından iki popülasyon birbirine çok benzer olduğunda CE ve PSE'nin neredeyse aynı sonuçları vermesi gerektiğini vurgularlar. NEAT tasarımında ortak madde testinin diğer iki testle mükemmel bir şekilde korelasyonu olduğunda, CE ve PSE'nin her

iki popülasyon arasındaki fark ne kadar büyük olursa olsun popülasyon değişmezliği ve özdeşliği olduğunu gösterir ve hem CE hem de PSE aynı eşitleme fonksiyonunu verir. Ayrıca, Dorans ve Holland'ın (2000) basit eşdeğer gruplar ve tek grup eşitleme desenlerinde P popülasyonu için kullanılan eşitleme fonksiyonlarının popülasyon değişmezliği için *RMSD* ölçümlerini NEAT tasarımında P ve Q popülasyonları için CE ve PSE yöntemlerine genelleştirilmiştir.

Bränberg ve Wiberg (2011) hem denk gruplar deseni hem de denk olmayan gruplar deseni için Kovaryantlar ile Doğrusal Eşitleme Modelini önermiştir. Bir simülasyon çalışmasına dayalı olarak çok terimli dağılımdan elde edilen gözlemler ve gerçek verilerden oluşan bir veri seti kullanılmıştır. Gerçek veriler olarak, İsveç'te üniversite ve kolejlere kabul için kullanılan ve her seferinde yeni bir test formuyla iki kez verilen SweSAT kullanılmış ve bir log-lineer modelle düzgünleştirilmiştir. SweSAT, 122 maddenin beş alt teste ayrıldığı bir testtir. Bu makalede sunulan model iki bölümden oluşmaktadır. İlk olarak, gözlemlenen puanları kovaryantlarla ilişkilendiren bir doğrusal regresyon bölümü bulunmaktadır. İkinci olarak, iki test formundaki gözlemlenen puanları birbirine bağlayan doğrusal bir eşitleme bölümü bulunmaktadır. Modelin parametrelerini tahmin etmek için normal bir dağılım varsayılmış ve maksimum olabilirlik tahminini kullanılmıştır. Test eşitlemede ise Tucker yöntemi kullanılmıştır. Simülasyon çalışmasında denk gruplar desenine göre bir eşitlemenin doğruluğunun kovaryantların (çalışmada cinsiyet ve eğitim olarak belirlenmiştir.) eşitleme sürecinde kullanılmasıyla artırılabilirliğini göstermektedir. Denk olmayan gruplar deseninde ise eşitleme parametrelerinin tahmin edilmesinin etkili yolunun iyi bir ortak madde testi kullanmak olduğunu göstermektedir. Tahminlerin doğruluğunun kovaryant gözlemleri ile ortak madde testi puanlarının desteklenmesiyle iyileştirilebilir, ancak bu iyileşme küçüktür. Bir ortak madde testinin olmaması durumunda, eşitlemeyi iyileştirmek için kovaryantların gruplar arasındaki farkları düzeltmek için kullanılabilirliği, bu farkları açıklayabilecek kovaryantlar bulunması şartıyla geçerlidir. Bu

özellikle, yasal, idari veya ekonomik kısıtlamaların bir ortak madde testinin kullanımını zorlaştırdığı SweSAT gibi test programları için önemlidir.

Liu and Low (2008) çalışmalarında kernel eşitleme ile geleneksel yöntemleri karşılaştırmışlardır. Çalışmada kernel eşitleme birisi çok benzer bir popülasyonda eşitleme ve diğeri oldukça farklı popülasyonda eşitleme olmak üzere iki farklı eşitleme senaryosunda uygulanmıştır. Denk olmayan gruplarda ortak madde deseni kullanılmış ve ortak maddeler dış test olarak belirlenmiştir. Çalışma bulgularına göre doğru sayısına göre puanlanan testler gibi rastgele veri düzensizliklerine sahip puan dağılımları üreten test programları için, Kernel süreklileştirme ve geleneksel eşit yüzdelli süreklileştirme arasındaki farkların dağılımlardaki düzensizlikler ön-düzleştirildikten sonra küçük olması beklendiği belirtilmiştir. Dolayısı ile Kernel eşitleme fonksiyonları ile geleneksel eşitleme fonksiyonları arasındaki farkların da küçük olacağı vurgulanmıştır. Formül puanlarının kullanımından kaynaklananlar gibi yapılandırılmış düzensizlikler içeren puan dağılımlarına sahip test programları için, Kernel süreklileştirme ile geleneksel eşit yüzdelli süreklileştirme arasındaki farklar göz ardı edilemeyeceği belirtilmiştir. Ayrıca sonuçlar, iki testi alan iki popülasyonun ortak madde test puan dağılımları benzer olduğunda, farklı eşitleme yöntemlerinin, farklı varsayımlara sahip olsalar bile, aynı veya çok benzer sonuçları verdiğini göstermişlerdir. Kernel eşitleme fonksiyonları, geleneksel eşit yüzdelli eşitleme yöntemine göre daha az doğru ve daha fazla yanlı olabileceği işaret edilmiştir. Kernel eşitlemede kullanılan *PRE* ve *SEED* gibi eşitleme doğruluğuna dair kullanılan değerler geleneksel eşitleme yöntemleri tarafından da uygulanabilir olması tavsiye edilmiştir. Bu nedenle, kernel eşitleme, süreklileştirme ve standart hatanın geliştirilmesi açısından geleneksel gözlenen puan eşitleme yöntemlerine göre heyecan verici bir alternatif olmasına rağmen, pratik bir bakış açısından geleneksel eşitleme yöntemlerinin kernel eşitleme ile değiştirilmesini önermemektedirler.

İki test arasında eşitleme yapılırken geleneksel yaklaşımda genellikle ortak test katılımcıları ve/veya ortak maddeler kullanılmaktadır. Ancak Wiberg ve Bränberg (2015)

tarafından yapılan bir çalışmada, test puanları ile ilişkili değişkenlerin (örneğin, okul notları ve diğer test puanları) ortak maddelerin yerine kullanılması amaçlanmıştır. Bu amaç çerçevesinde çalışma kernel eşitleme ile denk olmayan gruplarda ortak değişken deseni, denk gruplar deseni ve denk olmayan gruplarda ortak madde deseninde son-tabakalama yönteminin geliştirilmesi çerçevesinde uygulanmış ve karşılaştırılmıştır. Gerçek veriler, bir kolej giriş testinden alınmıştır. Çalışmada denk olmayan gruplarda ortak madde deseni, denk gruplar desenine göre daha düşük standart hatalar vermiştir. Ortak madde ile birlikte ortak değişkenler de kullanıldığında geniş bir test puanı aralığında standart hataların çok düşük olduğu gözlenmiştir. Standart testlerin hepsinde ortak madde bulunmadığı için denk gruplar deseninde ortak değişkenlerin dağılımındaki farklılıklardan kaynaklanan test puanı dağılımlarındaki farklılıklara göre ayarlama yapılarak eşitlemeler geliştirilebileceği belirtilir.

Ortak madde testlerinin, eşitlenen testlerin içeriği ve istatistiksel özellikleri açısından mini versiyonları (minitestler) olması gerektiği geniş bir kabul görür. Sinharay ve Holland (2007), bu inancın istatistiksel özelliklerle ilgili temellerini incelemektedir. Çalışmalarında içerik açısından temsilci olan ortak madde testlerinin istatistiksel temsil ediciliği gereksinimini incelerler. Çalışmada içerik açısından temsilci olan, toplam testlerle aynı ortalama zorlukta olan ve madde zorluklarının yayılımı toplam testlerinkinden daha az olan ortak madde testlerine odaklanılmıştır. Minitestler, miditestler ve yarı-miditestler gibi eşitlenen testlerden farklı istatistiksel özelliklere sahip olan çeşitli türdeki ortak madde testlerinin eşitleme performansı, bir dizi simülasyon çalışması ve gerçek olmayan veri örneği aracılığıyla incelenmiştir. Çalışmada tek boyutlu ve çok boyutlu madde tepki kuramı modellerinden üretilen veriler kullanılarak denk olmayan gruplarda ortak madde deseni bağlamında minitest ve diğer ortak madde testlerinin eşitleme performansı karşılaştırılmıştır. Simülasyonlarda test uzunluğu, örneklem büyüklüğü, iki sınav grubunun ortalama yetenekleri farkı, iki testin ortalama zorlukları arasındaki fark, ortak madde testi, eşitleme yöntemi gibi faktörler kontrol edilmiştir. Çalışmanın sonuçlarına göre, madde zorluklarının yayılımı toplam testten daha az olan ortak madde testlerinin yanlılığı ve

standart hatası açısından minitestlerle aynı performansı sergilediği belirtilmiştir. Bu nedenle, bir ortak madde testinin toplam testin istatistiksel özelliklerini yansıtması gereğinin çok kısıtlayıcı olabileceği ve en iyi seçenek olmayabileceği sonucuna varılmıştır.

Kim vd. (2010) farklı formattaki test maddelerine dayalı test formları üzerinde eşitleme incelemeleri yapmışlardır. Çalışmalarında, çoktan seçmeli ve yapılandırılmış yanıt maddeleri içeren testler için denk olmayan gruplarda ortak madde deseninde farklı durumlar incelemişler ve hangi tasarımın eşitlenen iki test arasında denk puanlar üretmede daha etkili olduğunu belirlemeye çalışmışlardır. Büyük ölçekli bir sınavdan elde edilen veriler kullanılarak, klasik eşitleme yöntemleri bağlamında doğrusal eşitleme yöntemleriyle (chained linear, Tucker, Levine-observed score) eşitleme çalışmaları yapılmıştır. Çalışma kapsamında dört eşitleme tasarımı incelenmiştir, bunlar; sadece çoktan seçmeli maddeleri içeren bir ortak madde testi, hem çoktan seçmeli hem de yapılandırılmış yanıt maddelerini içeren karma formatlı ortak madde testi, yine hem çoktan seçmeli hem de yeniden puanlanmış (trend scoring) yapılandırılmış yanıt maddelerini içeren karma formatlı ortak madde testi ve bir ortak madde testi olmaksızın yeniden puanlama içeren yapılandırılmış yanıt maddeleri içeren testlerin denk gruplar deseninde eşitleme incelemeleridir. Burada şuna açıklık getirmekte fayda var, yeniden puanlama (trend scoring) denilen durumda eski formda puanlanan yapılandırılmış yanıt içeren maddeler yeni test formundaki yapılandırılmış yanıt içeren maddeleri değerlendirecek olan grup tarafından yeniden değerlendirilmiştir. Böylelikle farklı gruplara bağlı değerlendirilerek verilen puanlardan kaynaklanan hatalardan eşitlenmenin arındırılması hedeflenmiştir. Eşitleme sonuçlarına göre; yalnızca çoktan seçmeli maddeleri ortak madde testi olarak kullanılan veya yeniden puanlama olmadan yapılandırılmış yanıt maddelerinden oluşan ortak maddeli test desenleri, diğer iki desene göre çok daha büyük yanlılığa yol açmıştır. Yeniden puanlamaya dayalı yapılandırılmış yanıt maddeli denk grup deseni ile yapılan eşitleme en küçük yanlılığa ve en küçük RMSE hatasına sahip olmuştur.

Moses vd. (2007) ise alternatif formlardaki madde sırasının eşitlemeye olan etkisini incelemiştir. Kernel yöntemi içinde farklı madde sıralamalarının puanlara etkisini değerlendirmek için önerilen bazı prosedürleri yeniden gözden geçirmişlerdir. Bu değerlendirmede biri genel puan dağılımlarına odaklanırken (Hanson, 1996) diğeri eşitlenmiş puanlara odaklanan (Dorans & Lawrence, 1990) iki strateji kullanılmıştır. Bu amaçla, kernel yöntemi Hanson (1996) ve Dorans ve Lawrence (1990) prosedürlerini bütünleştirmek için kullanışlı bir çerçeve sağlayarak madde sıra etkileri hakkında tamamlayıcı ve kapsamlı bilgi sunmuş ve böylece madde sıra etkileri için uygun eşitleme fonksiyonu (identity, curvilinear, linear) tavsiye edilmiştir. Puan eşdeğerliği, iki farklı ileri düzey yerleştirme sınavının (AP) iki farklı türü için değişen madde sıralamaları iki biçimde değerlendirilmiştir. Sınav hacimlerinin büyük veya küçük oluşuna ve madde sırası etkilerinin yüksek veya düşük oluşuna göre prosedürler madde sırası etkilerinin hem istatistiksel hem de pratik olarak önemli veya önemsiz olma durumuna göre eşitlemeye dair önerilerde bulunulmuştur. Örneğin, sınav hacimleri ve madde sırası etkileri küçük olduğunda, prosedürler madde sırası etkilerinin hem istatistiksel hem de pratik olarak önemsiz olduğunu ve dolayısıyla eşitlemenin garanti edilmediğini öne sürmüştür.

Kan (2011) farklı yıllarda uygulanan OKS Türkçe alt testleri puanları açısından bir eşitleme çalışması yapmıştır. Çalışmada 2003 ve 2005 yıllarında düzenlenen 25 çoktan seçmeli sorudan oluşan OKS Türkçe alt testleri birleştirilerek 50 soruluk bir test oluşturuldu ve 2005–2006 eğitim-öğretim yılında 1030 8. sınıf öğrencisine uygulandı. Öğrencilere test, tek oturumda ve 50 dakika süre verilerek uygulandı. 2003 OKS Türkçe alt test puanları, 2005 OKS Türkçe alt test puanlarına, klasik eşitleme metotlarından doğrusal eşitleme ve tek grup düzeneği kullanılarak eşitlendi. Eşitleme işlemi sırasında tek grup düzeneği için eşitleme hatası (SHE) hesaplandı. İki test arasındaki denkliliği değerlendirmek için eşitlemenin standart hatasına dayalı güven aralıkları kullanıldı. Sonuç olarak, elde edilen bulgular, özdeşlik fonksiyonunun tanımlanan ± 2 SHE güven aralığı

içinde olmaması nedeniyle 2003 ve 2005 OKS Türkçe alt testlerinin birbirine eşdeğer olduğunu veya birbirinin yerine kullanılabileceğini belirtmenin zor olduğunu belirtmiştir.

Ozdemir (2017) yaptığı araştırmada 2011 ve 2007 yıllarında uygulanan TIMSS matematik testlerinin eşitlenmesini amaçlamaktadır. TIMSS 2011'den elde edilen puanlar, denk olmayan gruplarda ortak madde deseni altında doğrusal olmayan gözlenen puan eşitleme yöntemleri kullanılarak TIMSS 2007 formundan elde edilen puanlarla eşitlenmiştir. Çalışma aynı zamanda, kernel eşitleme yöntemlerine dayalı farklı eşitleme yöntemlerinin, daire-yay eşitleme yöntemi adı verilen yeni bir doğrusal olmayan eşitleme yöntemiyle karşılaştırılmasını içermektedir. Türk öğrencilere uygulanan sekizinci sınıf matematik alt testlerinden elde edilen veriler kullanılarak gerçekleştirilen araştırmada, TIMSS 2011 matematik alt test kitapçığı 12'den elde edilen ham puanlar, TIMSS 2007 matematik alt test kitapçığı 14'e eşitlenmiştir. Bu eşitleme sürecinde, eşit yüzdelli eşitleme ve daire-yay eşitleme yöntemleri kullanılmıştır. Çalışmanın sonuçları, daire-yay eşitleme yönteminin eşit yüzdelli eşitleme yöntemine göre daha doğru sonuçlar verdiğini ve farklı yıllarda uygulanan testler arasında eşitsizliği gösterdiğini ortaya koymaktadır. Ayrıca, yeni daire-yay eşitleme yöntemlerinin, diğer doğrusal olmayan eşitleme yöntemlerine kıyasla daha başarılı olduğu ve daha tutarlı sonuçlar verdiği bulunmuştur.

Ankara Üniversitesi tarafından geliştirilen Yabancı Öğrenciler Sınavı (AYOS) kapsamında Temel Öğrenme Becerileri Testi (BLST) verilerini kullanarak Altıntaş ve Wallin (2021) tarafından yapılan bir araştırmada 2017 ve 2018 yıllarında uygulanan toplam 5,223 bireye ait test puanlarının eşitlenmesi amaçlanmıştır. Bu test, sözlü olmayan bir yetenek ölçümü olan AYOS BLST'nin iki bölümden oluşmakta olup, toplamda 100 çoktan seçmeli maddeden oluşmaktadır. Araştırma kapsamında, farklı test formlarının istatistiksel eşitliğini belirlemek için kernel eşitleme yöntemi ve denk olmayan gruplarda ortak değişken (DOG-OD) deseni kullanılmıştır. Bu bağlamda, test formlarını eşitlemek için cinsiyet ve yaş gibi kovaryatlar kullanılmış ve eşitleme sonuçları SEE ve SEED gibi hatalar kullanılarak değerlendirilmiştir. Elde edilen sonuçlara göre, ortak değişken

kullanılarak test skorları arasındaki eşitlik sağlanmış ve testlerin eşdeğer olduğu sonucuna varılmıştır.

Benzer biçimde Başaran vd (2023), TIMSS 2019'da dördüncü sınıf seviyesinde 1699 adayın fen testi puanlarını kullanarak klasik test teorisi (KTT) ve kernel eşitleme temelli yöntemlerin, denk olmayan gruplarda ortak madde (DOG-OM) ve denk olmayan gruplarda ortak değişken (DOG-OD) desenlerinde eşitlenen puanları karşılaştırmışlardır. Bu karşılaştırmalarda KTT'de Tucker, Levine gerçek puan, Levine gözlenen puan, öndüzdünleştirme ve son-düzdünleştirme eşit yüzdelikli eşitleme yöntemleri ile kernel eşitlemede doğrusal ve eşit yüzdelikli yöntemler kullanılmıştır. DOG-OM deseni eşitlemeleri kapsamında "evde öğrenme kaynakları, matematik ve fen derslerinde güven, fen öğrenmeyi sevme, fen derslerinde öğretimsel açıklık, matematik dersi başarısı, cinsiyet ve evde test dilini konuşma" değişkenler olarak kullanılmıştır. Karşılaştırmalar eşitleme hataları ve eşitlenmiş puanlar arasındaki farklar açısından değerlendirilerek desenlere göre daha iyi olan eşitleme ve yöntemler belirtilmiştir. Araştırmada elde edilen sonuçlar arasında, DOG-OM deseninde eşitleme yapılamadığı durumlarda matematik başarısı ve evde öğrenme kaynaklarının denk olmayan gruplarda ortak değişken deseninde kovaryant olarak kullanılabilmesi belirtilmiştir. Yine ortak değişken kapsamında kullanılan diğer değişkenlerin denk gruplar ile yapılan eşitlemeler ile benzer sonuçlar verdiği görülmüştür.

Ülkemizde yapılan eşitleme çalışmaları için yukarıda verilen araştırma örneklerinin yanında yine test eşitleme ile ilgili olarak farklı eşitleme yöntemlerinin karşılaştırılması (Akin Arıkan & Gelbal, 2021; Mutluer & Çakan, 2023; Özsoy & Kilmen, 2023), şans başarısı (Bozdağ & Kan, 2010; Koçak, 2020), ortak maddelerde değişen madde fonksiyonu (Demirus & Gelbal, 2016; Gübes & Uyar, 2020), MTK yöntemlerinin test eşitlemede karşılaştırılması (Aksekioğlu, 2017; Öztürk Gübeş & Kelecioğlu, 2015; Pektaş & Kılınç, 2016), çok kategorili puanlanan maddelerden oluşan testlerin eşitlenmesi (Çörtük & Sinan, 2023), eşitlemede yerel bağımsızlık (Doğuyurt & Tan, 2023), MTK'da eşitleme

hataları (Salmaner Dođan & Tan, 2022), grup deđişmezliđi (Suna & Tan, 2017; İnal & Arıkan, 2017), küçük örneklem (Caglak, 2016, Tan, 2015) ve çok boyutşu eşitleme (Burcu & Yeşiltaş, 2017) gibi konularda da araştırmalar mevcuttur. Bu tez çalışmasının konusu ise daha önceden de belirtildiđi üzere bilişsel tanı modelleri ve test eşitleme ile ilgilidir. Türkiye'deki ve dünyadaki literatür incelendiđinde ise bu iki alanı bađlar nitelikte sadece iki çalışma ile karşılaşılmıştır. Çalışmalardan biri yüksek lisans tezi olarak Türkiye'de diđeri ise yurt dıřında yapılmıř ve bilişsel tanı modellerinde yerel eşitlemeye dayalı olarak yayımlanmıř bir çalışmadır.

Canpolat (2019) BTM ve test eşitleme çalışmaları kapsamında yaptıđı araştırmada, BTM standartlarına uygun olarak geliřtirilen iki farklı test formunun psikometrik özellikleri incelenmiştir. Her iki test formu, içerdikleri ortak ve farklı maddelerle birlikte, açık uçlu ve çoktan seçmeli maddelerin yer aldıđı 15 soruluk bir yapıya sahiptir. Aynı zamanda, her iki test formunun farklı cevaplayıcı grupları tarafından tamamlanmasıyla elde edilen veriler, DINA modeli kullanılarak detaylı bir şekilde analiz edilmiştir. Araştırmada, BTM parametrelerine göre belirlenen örtük yetenek sınıflarına ait cevaplayıcıların diđer testi almıř olması durumunda atanacakları örtük yetenek sınıflarının ankor maddeler üzerinden belirlenmesine yönelik bir yöntem önerilmiştir. Veri analizi üç aşamada gerçekleştirilmiştir. İlk olarak, Test 1'in uygulandıđı cevaplayıcıların cevap örüntüleri ve örtük yetenek sınıfları belirlenmiştir. Daha sonra, Test 1 ve Test 2'den elde edilen madde parametreleri kullanılarak simülasyon verileri üretilmiş ve DINA modeliyle analiz edilmiştir. Son olarak, aynı örüntüye sahip olan dataların örtük yetenek sınıfları eşleştirilmiştir. Bu sayede, Test 1'i tamamlayan cevaplayıcıların diđer testi alması durumunda hangi örtük yetenek sınıflarına atanacakları belirlenmiştir.

Xin ve Zhang (2015) tarafından bilişsel tanı modeli gözlenen puanlarının yerel eşitlemesine dayalı bir çalışma gerçekleştirilmiştir. Bu çalışmada eşitleme kriterlerinin karşılanmasının yanı sıra ortak madde testi olmadan eşitleme sorununu çözmek için bilişsel tanı modeli gözlenen puanın yerel bir eşitleme yöntemi tanıtılmıştır. Test yapı

kriteri nitelik düzeyinde yeniden tanımlanmış ve ardından yerel eşitlemenin ana kaygısı olan eşitlik kriterinden, nitelik yetkinlik örüntüsüne göre verilen bir eşitlik dönüşüm ailesi türetilmiştir. Simülasyon çalışmalarının sonuçlarına göre, denk gruplar deseni altında geleneksel eşit yüzdelikli eşitleme için önemli bir yanlılık olduğunu, ancak yerel yöntem için çok daha küçük bir yanlılık gösterdiği belirtilmiştir. Araştırmacılar tarafından önerilen yöntemin aynı zamanda denk olmayan grup tasarımı altında ortak madde testi olmadan bile tatmin edici performanslar gösterdiği belirtilmiş, yerel yöntemin örneklem büyüklüğündeki değişikliklere göre oldukça sağlam olduğu ancak model uyumsuzluğunun artmasıyla birlikte yanlılığın da arttığı vurgulanmıştır.

Özet olarak, bilişsel tanı modelleri ve test eşitleme yöntemleri, kendi içlerinde derin ve geniş bir gelişme göstermiş psikometrik alanın farklı iki alt dalını oluşturmaktadır. Bu alt dalların kendilerine özgü araştırmaları, literatürde geniş bir yelpazede yer aldığı bilinmektedir. Ancak, bu iki alanın kesişimini inceleyen bu tez, bu iki alan arasındaki etkileşimi ve potansiyel bağlantıları anlamak için bir köprü görevi görmektedir. Bu çalışma, bilişsel tanı modellerinin test eşitleme yöntemlerine entegrasyonu üzerine odaklanarak, psikometrik araştırmaya yeni bir bakış açısı sunmayı amaçlamaktadır.

Bölüm 3

Yöntem

Bu bölümde araştırmanın türü, çalışma grubu, araştırmada kullanılan veri toplama araçları ve süreci ile verilerin analizine dair detaylar paylaşılmıştır.

Araştırmanın Türü

Bu çalışmada, tarama modelleri çerçevesinde genel tarama model türünden kesit alma yaklaşımı ile test uygulaması yapılan bir çalışmadan (Erdoğan,2023) elde edilmiş yanıt örüntülerinin bu çalışmanın hedefleri doğrultusunda yeniden analiz edilerek, denk gruplar (DG) vedenk olmayan gruplarda ortak madde (DOG-OM) desenlerinde KTK'ya dayalı geleneksel eşitleme (ortalama, doğrusal, eşit yüzdelikli) ve Kernel eşitleme yöntemleri ile denk olmayan gruplarda ortak değişken (DOG-OD) deseninde KTK'ya dayalı Kernel eşitleme yöntemi uygulanarak beceri puanlarının eşitlenme durumu eşitleme puanları ve hataları açısından incelenmiştir.Bu araştırma, tanılayıcı test uygulamasında gözlenen yanıt örüntüleri üzerinden bilişsel tanı modelleri uygulanarak elde elden analiz sonuçları, beceri eşitleme sürecinde veri olarak kullanıldığı ve mevcut eşitleme yöntem ve tekniklerin gerçek verilerle test edilerek beceri puanlarının eşitlenebilme durumunun incelendiği betimsel bir araştırmadır (Karasar, 2009).

Çalışma Grubu

Erdoğan (2023) tarafından uygulanan tanılayıcı testler (Test A, Test B, Test C) ortaokulu tamamlayan ve liseye yeni geçen öğrencileri kapsadığı için Ankara'nın Çankaya İlçesi'nde bulunan 27 lisenin dokuzuncu sınıf veya hazırlık öğrencilerinden alınan yanıt örüntüleri bu çalışmada kullanılmıştır. Bahsedilen araştırmaya öğrenci alımını Liselere Giriş Sınavı (LGS) puanlarına dayalı olarak gerçekleştiren 15 okul ile sınav puanları olmaksızın öğrenci kabul eden 12 okul katılmıştır. Okul türleri göz önüne alındığında, Fen Lisesi (FL), Anadolu Lisesi (AL), Sosyal Bilimler Lisesi (SBL), Anadolu İmam Hatip Lisesi

(AİHL) ve Mesleki ve Teknik Anadolu Lisesi (MTAL) öğrencileri bu çalışmaya katılmıştır. Aşağıda Tablo 3'te, okul türlerine göre öğrenci katılımcı bilgileri sunulmuştur.

Tablo 3

Okul Türlerine ve Kabul Koşullarına Göre Katılımcı Sayısı

Okul Türleri	LGS puanı ile kabul		Sınavsız kabul	
	<i>f</i>	<i>n</i>	<i>f</i>	<i>n</i>
Fen Lisesi	2	315		
Anadolu Lisesi	6	931	9	705
Sosyal Bilimler Lisesi	2	238		
Anadolu İmam Hatip Lisesi	3	236		
Mesleki ve Teknik Anadolu Lisesi	2	78	3	113
Toplam	15	1798	12	818

f: katılımcı okul sayısı, *n*: katılımcı kişi sayısı

Tablo 3'te öğrenci kabul koşullarına ve okul türlerine göre çalışmaya katılan okul ve öğrenci sayıları paylaşılmıştır. Araştırmaya katılan toplam 27 okulun 15'i LGS puanlarına dayalı olarak öğrenci kabul etmekteyken, geriye kalan 12 okul ise sınav puanları olmaksızın öğrenci alımı gerçekleştirmiştir. Okul türlerine bakıldığında, Fen Lisesi (FL), Anadolu Lisesi (AL), Sosyal Bilimler Lisesi (SBL), Anadolu İmam Hatip Lisesi (AİHL) ve Mesleki ve Teknik Anadolu Lisesi (MTAL) öğrencileri bu çalışmada yer almıştır. LGS puanına dayalı kabul koşuluna sahip olan okullardan toplamda 1798 katılımcı bulunurken, sınav puanı koşulu olmayan okullardan ise 818 katılımcı vardır. AL grubu, en fazla okul ve öğrenci sayısına sahip olarak dikkat çekmektedir. Ayrıca, sadece sınav kriterine göre öğrenci kabul eden FL, SBL ve AİHL gibi okullar öğrenci sayısında AL grubundan daha düşük bir katılıma sahiptir.

Ayrıca, Tablo 4'te uygulanan her bir test için okul türlerine göre katılan öğrenci sayısı paylaşılmıştır.

Tablo 4*Testlerdeki Okul Türlerine Göre Katılımcı Sayısı*

	<i>AİHL</i>	<i>AL</i>	<i>FL</i>	<i>MTAL</i>	<i>SBL</i>	<i>Toplam</i>
<i>TEST A</i>	81	540	107	61	78	867
<i>TEST B</i>	79	550	102	63	77	871
<i>TEST C</i>	73	540	106	64	82	865

AİHL: Anadolu İmam Hatip Lisesi, *AL*: Anadolu Lisesi, *FL*: Fen Lisesi, *MTAL*: Mesleki ve Teknik Anadolu Lisesi, *SBL*: Sosyal Bilimler Lisesi

Tablo 3 ve Tablo 4'te görüldüğü üzere AL okul sayısı diğer okul türü sayılarından fazla olduğu için örnekleme katılım yapan öğrenci grubu da diğer okul türlerine göre fazladır. Ancak okulların dağılımının çalışmaya ait evrende de bu biçimde olduğunu vurgulamakta fayda bulunmaktadır. Sınavla öğrenci alan FL, SBL, MTAL ve AİHL okul sayıları sınavla öğrenci alan AL okul sayısından çok daha azdır, bunun yanında sınav koşulu olmadan öğrenci kabul eden okulların büyük bir çoğunluğu AL grubundadır. Çankaya bölgesindeki sınav koşulu ile öğrenci kabul eden tüm okullara ulaşılmış sadece 3 okuldaki gönüllü katılım desteği alınamamıştır. Bu tablodaki verilerin toplama koşullarında örneklem dahilindeki ulaşılabilecek yüksek değerlerdir.

Tablo 5*Okulların Kabul Koşullarına Göre Testlerdeki Katılımcı Sayısı*

<i>Kabul Koşulu</i>	<i>TEST A</i>	<i>TEST B</i>	<i>TEST C</i>
Sınav Puanı Koşulu	595	591	605
Sınav Puanı Koşulsuz	272	280	260

Tablo 5'te görüldüğü üzere, sınav puanı ile kabul koşulu olan okullardaki katılım, sınav puanı koşulu olmayan okullara kıyasla yaklaşık olarak iki katı kadar yüksektir. Sınav puanı koşulu olmayan okullardaki düşük katılımın nedeni, bu okulların uygulanacak testlere gönüllü olarak katılım yapmak istememeleri olmuştur.

Son olarak, cinsiyet alt gruplarına göre testlere katılan öğrenci sayısını sunmak da örnekleme tanıtmaya katkı sağlayacaktır. Tablo 6 her testte yer alan kız ve erkek öğrenci sayılarını sunmaktadır.

Tablo 6

Her Teste Göre Cinsiyet Bilgisi

	<i>n</i>	<i>kız</i>	<i>erkek</i>	<i>belirtmeyen</i>
TEST-A	867	435	396	36
TEST-B	871	432	414	25
TEST-C	865	444	386	35

n: katılımcı sayısı

Veri Toplama Süreci

Bu çalışmada Erdoğan (2023) tarafından daha önce tamamlanan bir çalışmada elde edilen veriler kullanılmıştır. Veriler örnekleme uygulanan ve üç form biçiminde hazırlanan tanılayıcı testlerden elde edilen yanıt örüntüleri, katılımcılara ait okul ve sınıf bilgileri, LGS puanları ve cinsiyet bilgileri ile sınırlıdır. Veriler Mart 2021’de liseye geçiş yapmış dokuzuncu sınıf veya hazırlık öğrencilerinin tanılayıcı test formlarına verdiği cevaplardan elde edilmiştir. Her sınıfta Test A, B ve C formları sırasıyla her bir öğrenciye dağıtılmış ve böylelikle her sınıf üç gruba ayrılarak test uygulaması yapılmıştır. Öğrencilere testi tamamlamaları için 30 dakika süre verilmiştir. Testler her okulda tüm sınıflarda aynı anda uygulanmıştır.

Veri Toplama Araçları

Çalışmada kullanılan tanılayıcı testler üç form olarak Erdoğan (2023) tarafından hazırlanmıştır. Testler cebirsel düşünme becerilerini ölçme hedefiyle oluşturulmuştur. Cebirsel düşünme ile ilgili olarak geniş ve derin bir alanyazın çalışması yapan Stephens ve diğerlerinin (2017) bu alanyazını sunarken kullandıkları çerçeve, ilgili çalışmadaki

tanılayıcı testlerin nitelikleri olarak belirlenmiştir. Bu niteliklerin ölçülmesi hedefiyle hazırlanan maddeler bir araya getirilerek tanılayıcı test formları oluşturulmuştur. BTM çerçevesinde ölçülmesi hedeflenen nitelikler, “genelleştirilmiş aritmetik, fonksiyonel düşünme, niceliksel mantık yürütme, genelleştirme ve mantık yürütme, görselleştirme ve gerekçelendirme” olarak altı cebirsel düşünme becerisinden oluşmuştur. Bu niteliklerin esasında yine alanyazın temelinde iki faktöre bağlı olduğunu belirlemek önemlidir. İlk üç nitelik cebirsel düşünmenin temel yönleri ve bileşenleri (Kaput, 2008) ve diğer üç nitelik ise cebirsel düşünmenin temel uygulamaları (Blanton ve diğerleri) kapsamında olduğu belirtilmiştir. Erdoğan (2023) oluşturulan test formuna ait süreçleri madde yazımından, test oluşturma, öğrenci, öğretmen, uzman ve akademisyenlerden alınan geri bildirimlerle yapılan düzenlemeleri ve testin son haline gelişi ile ilgili bilgileri ilgili çalışmasında detaylı bir biçimde açıklamıştır.

İlk adımda test maddelerinin geliştirilmesi ile ilgili olarak, ortaokulu tamamlayan ve liseye geçiş yapan matematik öğrencilerinin sahip olduğu cebir bilgilerini değerlendirmek amacıyla bir çerçeve oluşturulması hedeflenmiştir. Bu amaçla, ortaokul ders kitapları, lise giriş sınavları, ulusal ve uluslararası büyük ölçekli sınavlar ile akademik makalelerdeki maddeler incelenmiş, cebir ile ilgili maddeler seçilmiş ve bu maddelerin sınıflandırılması yapılmıştır. Seçilen bu maddeler doğrudan çalışmada kullanılmadan öğrencilere ne tür sorular sorulduğıyla ilgili genel durumu görmek ve araştırmacıya madde yazımında rehberlik etmek amacıyla seçilmiştir. Sınıflandırmalar ise maddelerin ilgili olduğu cebir alt konuları, araştırma kapsamında belirlenen cebirsel düşünme becerileri ve maddelerin öngörülen zorluk seviyelerine göre yapılmıştır. Bu sınıflandırmalar test geliştirme aşamasında belirtke tablosuna (Crocker & Algina, 1986) göre soru seçimini kolaylaştırmak amacıyla kullanılmıştır. Çalışmanın amacı çerçevesinde matematik öğretim programı ortaokul cebir programına dayanarak geliştirilen belirtke tablosu, tanılayıcı değerlendirme için hazırlanan testlerde faydalı bir araç olarak kullanılmıştır. Cebir konuları, altıncı sınıfta kısa bir tanıtım ile başlayarak yedinci ve sekizinci sınıflarda kapsamlı öğretimle birlikte

öğretim programında yer almıştır. Ortaokuldaki matematik müfredatındaki cebir programına dair genel bilgiler Tablo 7'de özetlenmiştir.

Tablo 7

Ortaokul Cebir Programı Konu Kazanım ve Süre Bilgisi

Konu	Sınıf Seviyesi			6			7			8		
	6	7	8	Kazanım Sayısı	Süre (saat)	%	Kazanım Sayısı	Süre (saat)	%	Kazanım Sayısı	Süre (saat)	%
Cebirsel İfadeler	√	√		3 (1)	10 (5)	6 (7)	3 (1)	10 (6)	6 (8)			
Eşitlik ve Denklem		√					4 (1)	20 (7)	11 (10)			
Cebirsel ifadeler ve Özdeşlikler			√							4 (1)	15 (7)	8 (10)
Doğrusal Denklemler			√							6 (1)	30 (8)	17 (11)
Eşitsizlikler			√							3 (1)	10 (6)	6 (8)

Not: Parantez içi verilen değerler ortaokul matematik öğretim programının uygulamalı matematik dersine aittir (MEB, 2018)

Tablo 7, okul konuları ve sınıf düzeyinde öğretim programının bir parçası olarak belirlenen hedef sayısını, konuya ayrılan toplam süreyi ve toplam ders saati süresi içerisindeki yüzdeleri göstermektedir. Parantez içinde verilen değerler, uygulamalı matematik derslerine aittir. Önemli bir not olarak, beşinci sınıf düzeyi ortaokul kapsamında yer almasına rağmen öğretim programında cebir konuları altıncı sınıftan itibaren başladığı için bu tabloda yer almamaktadır. Test formları için hazırlık yapılırken bu tablo belirteci tablosunun oluşturulmasında kullanılmıştır. 20 soruluk testin konu dağılımı, ders saatlerine göre oranı göz önüne alınarak belirlenmiştir. Altıncı sınıfta verilen cebir dersi, konuyu derinleştirmek amacıyla yedinci ve sekizinci sınıflarda tekrarlandığı için bu oran hesaplanırken, sadece yedinci ve sekizinci sınıfların ders saatleri dikkate alınarak hesaplanmıştır. Bu orana göre, tanılayıcı testlerde her alt konuya ayrılacak soru sayıları cebirsel ifade ve özdeşlikler için altı ya da yedi soru, eşitlik ve denklemler için beş soru, doğrusal denklemler için altı soru ve eşitsizlikler için üç soru olabilecek biçimdedir.

Ayrıca şunu da vurgulamak gerekir ki içerik belirleme işlemi ile birlikte çalışmada nitelikleri maddelerle uyumlu bir biçimde ayarlamak önemli bir aşama olarak yer almıştır. Çalışmaya başlamadan test formlarının oluşturulmasına kılavuzluk etmesi için belirtke tablosu hazırlanmış, bu tablo ışığında madde havuzu oluşturulmuş ve oluşturulan maddelerden belirtke tablosu ışığında uygun olan maddeler seçilerek ilk test formu oluşturulmuştur. Aşağıdaki tabloda ilk test formu için madde havuzundan belirtke tablosu ışığında seçilen soru ve nitelik dağılımları paylaşılmıştır.

Tablo 8

Konu ve Niteliklere Göre Madde Dağılımları

Konu	Madde Sayısı	Nitelikler					
		Temel Yön & Bileşenler			Temel Uygulamalar		
		GA	FD	NMY	G & MY	T	G
Cebirsel ifadeler ve Özdeşlikler	7	4	1	2	2	4	2
Eşitlik & Denklemler	7	4	1	3	3	4	2
Doğrusal Denklemler	7	1	6	2	4	4	1
Eşitsizlikler	3	2		3	3	3	2
		11	8	10	12	15	7

Not: GA: Genelleştirilmiş Aritmetik, FD: Fonksiyonel Düşünme, NMY: Niceliksel Mantık Yürütme, G&MY: Genelleme ve Mantık Yürütme, T: Temsilleştirme, G: Gereçlendirme

Test geliştirme süreci, Crocker ve Algina (1986) ile Özçelik (1989) tarafından belirtilen test hazırlama ve geliştirme kılavuzları ışığında devam ettirilmiştir. Ana form olarak adlandırılan Test A'nın ilk taslağı için 20 soruluk bir test hazırlanmış ve test düzenleme süreci aşağıdaki adımlarla takip edilmiştir:

- i. İlk olarak, test bir haftalık aralıklarla üç kez incelenmiş ve araştırmacı tarafından değişiklikler yapılmıştır.
- ii. Ardından, dört ortaokul ve lise öğretmeni testleri çözerek incelemiş ve görüşlerini araştırmacıyla paylaşmış, bu doğrultuda yenilemeler yapılmıştır.

- iii. Üç dokuzuncu sınıf öğrencisi testleri araştırmacının gözetiminde yüz yüze sesli olarak çözmüş ve çözüm süreçleri araştırmacı tarafından incelenmiştir. Buna ek olarak üç dokuzuncu sınıf öğrencisi ise testleri uzaktan kendi çalışma ortamlarında sesli olarak çözmüş, videolarını göndermiş ve araştırmacı tarafından videolar incelenmiştir.
- iv. Araştırmacı, yapılan incelemelere göre öğrenci cevapları göz önüne alınarak test maddelerinde düzenlemeler yapılmıştır.
- v. Düzenleme yapılan test bir grup on beş kişilik dokuzuncu sınıf seviyesindeki öğrencilere uygulanmıştır. Her öğrencinin test üzerinde yapmış olduğu yazılı çalışmalar detaylı bir biçimde incelenmiştir.
- vi. Öğrenci cevaplarının incelenmesi ve matematik öğretmenleriyle yapılan görüşmeler doğrultusunda Test-A'da yeniden düzenlemeler yapılmıştır.
- vii. Test, altıncı maddede belirtilen son haliyle iki test uzmanının incelemesine tabi tutulmuştur. Onların test üzerine özel çalışmaları ve araştırmacı ile yapılan görüşmeler ışığında test düzenlemeleri devam etmiştir.
- viii. Son aşamada ise matematik eğitimi alanında bir fakülte üyesi tarafından Test-A incelenmiş ve önerilere dayanarak bazı düzenlemeler yapılmıştır.

Bu aşamadan sonra ana form olarak belirlenen cebirsel düşünme yeteneği testi Test-A'dan yeni türetilen diğer iki test Test B ve Test C'nin geliştirme süreci de başlatılmıştır. Bu test geliştirme sürecinde araştırmacının yanı sıra, farklı okullarda aktif olarak görev yapan beş orta ve lise matematik öğretmeni, matematik alanında uzmanlaşmış dört test geliştirme uzmanı, matematik eğitimi alanında dört fakülte üyesi, eğitimde ölçme ve değerlendirme alanında olan iki fakülte üyesi, sekizinci ve dokuzuncu sınıf öğrencileri, Türkçe öğretmeni ve testin biçimsel düzeni için bir mizanpaj uzmanı bu sürece katılmıştır. Araştırmacı, süreçte koordinasyon sağlamış, yapılan geri bildirimleri değerlendirmiş ve testlerde düzenleme işlemlerini uygulamıştır. Araştırmacı, matematik

eđitimi ve öğretilmi, madde yazımı ve test geliştirme konularında uzmanlık ve tecrübelerini bu test geliştirme sürecinde çalışmalara aktarmıştır. Üç testin geliştirme süreci yaklaşık altı ay sürmüştür. Toplam 20 maddeden oluşan her bir test altı ortak maddeye sahiptir. Angoff (1971) tarafından öne sürüldüğü gibi, ortak soruların testin genel yapısı, kapsamı, soru türleri ve zorluk seviyeleri gibi özellikler açısından tam bir temsilci olması gerektiği düşünölmektedir. Benzer şekilde, Kolen ve Brennan (2004), ortak soruların özellikle zorluk derecesi ve içeriđi açısından tüm testi doğru bir şekilde yansıtmalı gerektiđini vurgulamışlardır. Ayrıca, her iki test formunda da ortak soruların herhangi bir kelime veya ifade deđişikliği olmadan ve aynı soru numaralarına karşılık gelecek şekilde yerleştirilmesinin önemli olduđu belirtilmiştir. Ortak maddelerin diđer önemli bir özelliđi, bu maddelerin tüm test formları içinde yeterli miktarda olması gerekliliđidir. Angoff (1971), ortak testin uzunluđunun, tüm testteki toplam madde sayısının %20'si kadar olması gerektiđini öne sürmüştür. Hambleton vd. (1991) ile Woldbeck (1998), ortak maddelerin, testin toplam madde sayısının yaklaşık %20 ila %25'i arasında olması gerektiđini belirtmişlerdir. Kolen ve Brennan (2004) ise, ortak test deseni kullanılarak yapılan eşitleme çalışmalarında ortak madde sayısının artmasının eşitleme hatalarını azalttığını bildirmişlerdir. Çok boyutlu ölçüm yapılan bu çalışmada konu ve ölçölen altı nitelik birlikte göz önüne alınarak 20 maddeden altı madde ortak madde olarak seçilerek eşitleme hatalarının azaltılması hedeflenmiştir. Her testte altı niteliđin ölçöldüğü ve her teste ait Q-matrisin aynı olduđu, yani aynı özellikleri ölçen her maddenin her test formunda aynı soru numarası sırası ile sorulduđu bilgisi önemlidir.

Testlere ait maddelerin yazımında her testte aynı beceriyi ölçen maddeler aynı sırada testlere yerleştirilmiştir. Testlerdeki maddeler aynı becerileri ölçmesinin yanında aynı kazanımı ölçmelerine dikkat edilmiş, paralel formlarda bazı maddelerde sayı deđişimi, bazılarında cevap şıkkı deđişimi, bazılarında ise ufak senaryo deđişimi yapılmıştır. Bazı durumlarda ise aynı beceri çerçevesinde ölçölüp benzer kazanımlar kapsamında maddeler yazılmıştır. Maddelerin aynı beceri ve kazanım çerçevesinde farklı

yönleri bakımından örnek olması amacıyla aşağıda her teste ait birkaç madde örneği verilmiştir.

Şekil 2

Testlere Ait Madde 8

TEST A

8. Dik koordinat düzleminde

$$\frac{3x}{5} - \frac{5y}{3} = 1$$

doğrusu veriliyor.

Bu doğrunun eğimi kaçtır?

- A) $\frac{3}{5}$ B) $\frac{9}{25}$ C) $-\frac{5}{3}$ D) $-\frac{1}{3}$

TEST B

8. Dik koordinat düzleminde

$$\frac{5y}{4} - \frac{4x}{5} = 1$$

doğrusu veriliyor.

Bu doğrunun eğimi kaçtır?

- A) $\frac{4}{5}$ B) $\frac{16}{25}$ C) $-\frac{1}{16}$ D) $-\frac{20}{25}$

TEST C

8. Dik koordinat düzleminde

$$\frac{2x}{3} - \frac{3y}{2} = 2$$

doğrusu veriliyor.

Bu doğrunun eğimi kaçtır?

- A) $\frac{2}{3}$ B) $\frac{4}{9}$ C) $-\frac{4}{3}$ D) $-\frac{2}{9}$

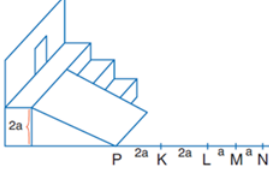
Madde 8'de genelleştirilmiş aritmetik, fonksiyonel düşünme ve genelleme ve mantık yürütme becerileri çerçevesinde denklemi verilmiş olan bir doğrunun eğiminin bulunması istenmiştir. Her testte kullanılan ifade, çözüm yolu, denklemlerde kullanılan sayıların türü açısından ortaklık söz konusudur. Maddelerde sadece rakamsal olarak değişiklik yapılmıştır.

Şekil 3

Testlere Ait Madde 10

TEST A

10. Kapısı yerden $2a$ birim yükseklikte olan bir binanın girişine bir rampa yapılacaktır.



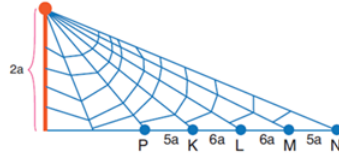
Şekildeki gibi P noktasından başlayacak biçimde yapılan rampanın eğimi % 10'dur.

Buna göre, eğimi % 8 olması için rampa yerdeki doğrusal olarak verilen noktaların hangisinden başlamalıdır?

- A) K B) L C) M D) N

TEST B

- 10.



Yukarıda bir oyun parkında halatlarla yapılan örümcek ağı biçimindeki oyuncağın bir kesiti verilmiştir.

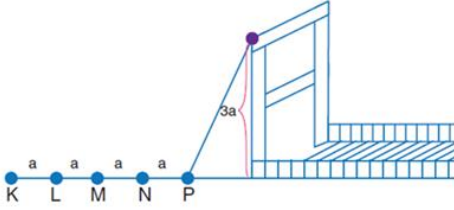
Şekildeki $2a$ uzunluğundaki direğin tepe noktasından yerde doğrusal olarak verilen noktaların her birine halatlar gerilerek bağlanmıştır. P noktasına bağlanan halatın eğimi % 25'tir.

Buna göre, verilen noktaların hangisine bağlanan halatın eğimi % 8'dir?

- A) K B) L C) M D) N

TEST C

- 10.



Bir köprü yapımında direklerin üzerinde belirlenen bazı noktalardan zeminde belirlenen doğrusal noktalara çelik halatlar sabitlenecektir.

$3a$ yüksekliğindeki köprü direğinin tepe noktasından P noktasına sabitlenecek bir çelik halatın eğimi % 75 olacaktır.

Buna göre, % 50 eğime sahip olacak çelik halat zemindeki hangi noktaya sabitlenmelidir?

- A) K B) L C) M D) N

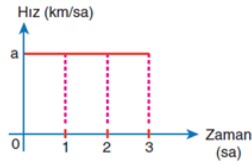
Madde 10 her testte genelleştirilmiş aritmetik, niceliksel mantık yürütme ile genelleme ve mantık yürütme becerileri çerçevesinde eğim kavramı kapsamında hazırlanmış olup y eksenindeki değişimin x eksenindeki değişime olan oranı üzerine kurgulanmış bir maddedir. Görünüşte farklı senaryolar verilmiş olsa da her testte çözüm için gerekli bilgi ve yöntem aynı temeldedir.

Şekil 4

Testlere Ait Madde 11

TEST A

11.

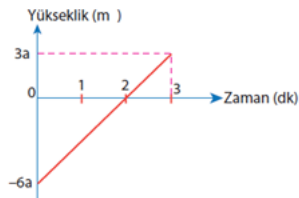


Yukarıda bir otomobilin hızının zamana bağlı değişimini gösteren grafik verilmiştir.

Bu otomobilin 3 saat boyunca her 1 saatlik zaman dilimi içerisinde aldığı yolun uzunluğu aşağıdakilerin hangisinde doğru olarak temsil edilmiştir?

1. saat dilimi 2. saat dilimi 3. saat dilimi
- A) $\frac{a}{3}$ $\frac{a}{3}$ $\frac{a}{3}$
- B) a $2a$ $3a$
- C) $3a$ $3a$ $3a$
- D) $\frac{a}{3}$ $\frac{a}{3}$ $\frac{a}{3}$

11. Deniz seviyesinden 6a metre derinlikte bulunan bir sandık, halata bağlanarak güvertesi deniz seviyesinden 3a metre yükseklikte olan bir gemiye 3 dakikada çıkarılıyor. Bu sandığın güverteye çıkarılırken yüksekliğinin zamana bağlı değişimini gösteren grafik aşağıda verilmiştir.

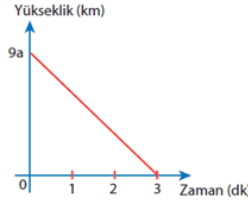


Bu sandığın bulunduğu derinlikten gemi güvertesine çıkartılırken her 1 dakikalık zaman dilimi içerisinde yüksekliğindeki değişim miktarı aşağıdakilerin hangisinde doğru olarak temsil edilmiştir?

1. dakika dilimi 2. dakika dilimi 3. dakika dilimi
- A) $\frac{a}{3}$ $\frac{a}{3}$ $\frac{a}{3}$
- B) $3a$ $3a$ $3a$
- C) $9a$ $6a$ $3a$
- D) $3a$ $6a$ $9a$

TEST B

11.



Yukarıda 9a km yükseklikten 3 dakikada iniş yapan bir uçağın yerden yüksekliğinin zamana bağlı değişimini gösteren grafik verilmiştir.

Bu uçağın bulunduğu yükseklikten iniş yaparken her 1 dakikalık zaman dilimi içerisinde yüksekliğindeki değişim miktarı aşağıdakilerin hangisinde doğru olarak temsil edilmiştir?

1. dakika dilimi 2. dakika dilimi 3. dakika dilimi
- A) $\frac{a}{3}$ $\frac{a}{3}$ $\frac{a}{3}$
- B) $3a$ $3a$ $3a$
- C) $9a$ $6a$ $3a$
- D) $3a$ $6a$ $9a$

Madde 11'de öğrencilerin eğitim bilgisi ve grafik okumalarıyla ilgili kazanımlarına bağlı olarak fonksiyonel düşünme, niceliksel mantık yürütme, genelleme ve mantık yürütme ile temsilleştirme becerileri ölçülmektedir. Üç testte aynı beceri çerçevesinde

eğim bilgisi temelinde sabit, azalan ve artan grafikler kullanılmıştır. Testteki maddeler arasında testler arası ölçülen beceriler aynı olmasına rağmen maddelerin bağlı olduğu kazanım bakımından test kapsamında en çok farklılık gösteren maddenin madde 10 olduğu söylenebilir. Burada bunun vurgulanma sebebi testlerde genel olarak her maddenin doğru cevaplanması için gerekli olan çözüm yolunun ve ihtiyaç duyulan düşünme becerilerinin her bir test için aynı olmasına özen gösterilmiştir ve amaç testlerin paralellik göstermesidir. Madde 11 ise aynı beceride aynı konuya ait alt konuya bağlı farklı kazanımlarda maddelerin durumunu izlemek adına teste yerleştirilen sorulardan biridir.

Şekil 5

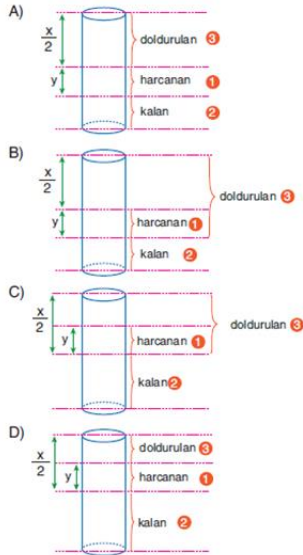
Testlere Ait Madde 12

TEST A

12. "Toplam kapasitesi x litre olan bir su deposunun yarısı doludur. Depodaki suyun y litresi harlandıktan sonra su deposunun tamamı dolduruluyor."

Yukarıda ifade edilen durumdaki aşamalar "harcanan (1), kalan (2) ve doldurulan (3)" su miktarı olarak modelleniyor.

Buna göre, aşağıdaki seçeneklerden hangisinde bu aşamalar doğru biçimde gösterilmiştir?

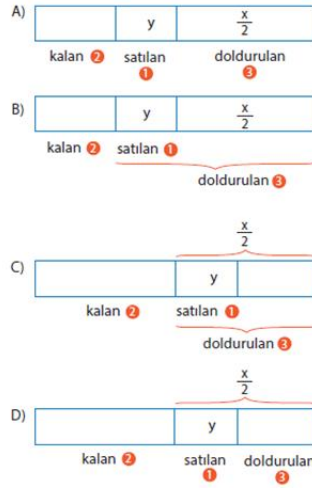


TEST B

12. "Toplam kapasitesi x ton olan bir buğday deposunun yarısı doludur. Depodaki buğdayın y tonu satıldıktan sonra deponun tamamı dolduruluyor."

Yukarıda ifade edilen durumdaki aşamalar "satılan (1), kalan (2) ve doldurulan (3)" buğday miktarı olarak modelleniyor.

Buna göre, aşağıdaki seçeneklerden hangisinde bu aşamalar doğru biçimde gösterilmiştir?



Şekil 5 (Devamı)

TEST C

12. "Bir kolonya üreticisi toplam kapasitesi x litre olan bir bidondan kolonya satışı yapıyor. Bu üretici bidonun yarısı dolu iken gelen müşterilere toplamda y litre kolonya satıyor. Daha sonra üretici bidonu kolonya ile tamamen dolduruyor."

Yukarıda ifade edilen durumdaki aşamalar "satılan (1), kalan (2) ve doldurulan (3)" kolonya miktar olarak modelleniyor.

Buna göre, aşağıdaki seçeneklerden hangisinde bu aşamalar doğru biçimde gösterilmiştir?

A)

	y	$\frac{x}{2}$
kalan 2	satılan 1	doldurulan 3

B)

	y	$\frac{x}{2}$
kalan 2	satılan 1	doldurulan 3

C)

	y	$\frac{x}{2}$
kalan 2	satılan 1	doldurulan 3

D)

	y	$\frac{x}{2}$
kalan 2	satılan 1	doldurulan 3

Madde 12'de farklı senaryoların aynı durumu tarif ettiği ve birbirinin çok benzeri olduğu görülmektedir. Madde 12 niceliksel mantık yürütme ve temsilleştirme becerileri çerçevesinde sözel ifadelerin matematiksel gösterimini şekil üzerinde beklemektedir. Soruda istenilen durum şıklarda farklı olarak silindir ve dikdörtgen üzerinde görselleştirilmiştir.

Maddelerin yazımı açısından testlerin geneli için bir yorum yapılacak olunursa, maddeler genel olarak birbirine çok benzer ve sayı değeri değiştirilmiş senaryolar üzerinden oluşturulmuştur. Hesaplamalarda her test için sayısal olarak aynı tür sayıların kullanılmasına özen gösterilmiş ve testte de genel olarak tam sayılarla işlem yapılabilecek değerler ayarlanmıştır. Testteki beceri ve ilgili konulardaki ilgili kazanımların ölçümleri arasında farklı sayı gösterimlerine (köklü, kesirli, ondalıklı ifadeler gibi) ait değerlerden uzak durulmuştur ki sayıların sahip olduğu işlemsel özelliklere ait bilgiler ölçülmesi hedeflenen beceri ve kazanımların önüne geçmesin. Örneğin bir testte ilgili maddedeki

işlemlerde tam sayılar kullanılarak bir çözüm yapılmış ise diğer testteki paralel olarak hazırlanan maddedeki çözüme ait işlemlerde ondalıklı sayılara ait işlemsel özelliklerin kullanımından uzak durulmuştur. Test maddelerinde farklı senaryolarda çözüm için aynı yöntem ve yolların takibine dikkat edilmiştir. En farklı maddeler olarak adlandırılabilir yukarıda örneği verilen madde 10, 11 ve 12 denilebilir ki bu maddelerin de bağlı olduğu konu ve alt konu, düşünme becerisi, çözüme giden yoldaki yöntem ve stratejileri çok benzer veya aynı temeldedir. Testler kapsamında birebir aynı (ortak maddeler), benzer (sayısal değişimler yapılan), benzer senaryolar ile aynı kazanım ve becerilerin ölçüldüğü veya benzer senaryolarda aynı alt konunun farklı durumlarına bağlı kazanımların aynı beceri çerçevesinde ölçülmesi hedeflenmiştir. Mümkün olduğunca paralel testler hazırlanması hedeflenen bu test geliştirme sürecinde cebirsel düşünme becerilerinin eşitlenebilme durumunun incelenmesinde geri planda madde yazımından düzenlemelerine, test maddelerinin çözümü için ihtiyaç duyulan düşünme becerilerinden ait olduğu konu ve kazanımlarına ve çözüme götüren matematiksel yöntemlerin paralelliğine dikkat edilmiştir. Burada vurgulanan geri plandaki düzenlemeler testlerin ölçmeyi hedeflediği becerileri ölçebilmesi ve bu çalışma kapsamında becerilerin farklı testlerde eşitlenebilme durumlarının incelenmesi açısından önemlidir.

Verilerin Analizi

Bu çalışmada verilerin analizi iki kısımdan oluşmaktadır. İlk aşaması verilerin BTM çerçevesindeki analizidir. İkinci aşaması ise BTM analizlerinden elde edilen bireylere ait beceri yetkinlik düzeylerinin bu çalışmanın amacı çerçevesinde üç test formunun birbiriyle çaprazlanarak (AxB, AxC, BxC) uygulanan beceri eşitlemelerinin incelenmesidir. Çalışmada BTM analizleri sonrasında bireylere ait beceri yetkinlik düzeyleri kullanılarak beceri puanları elde edilmiş ve bu puanların farklı grup desenlerinde, denk gruplar (DG) ve denk olmayan gruplarda ortak madde (DOG-OM) desenlerinde KTK'ya dayalı geleneksel eşitleme (ortalama, doğrusal, eşit yüzdelikli) ve Kernel eşitleme yöntemleri ile

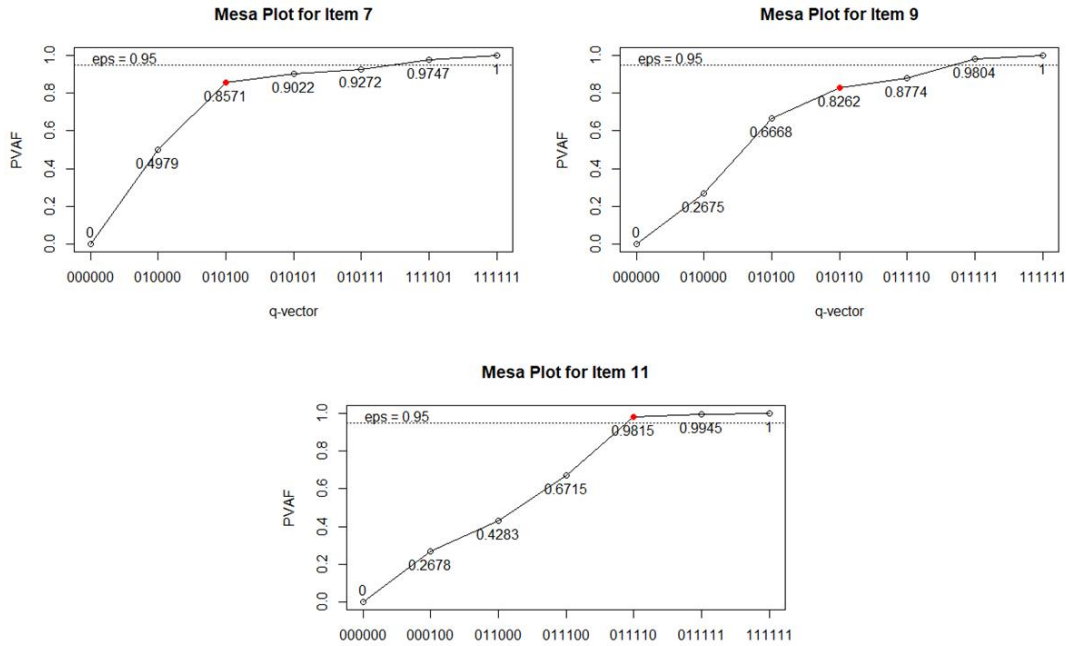
denk olmayan gruplarda ortak deęişken (DOG-OD) deseninde KTK'ya dayalı Kernel eşitleme yöntemi uygulanarak beceri puanlarının test eşitleme durumları incelenmiştir. Bu analizleri gerçekleştirmek için R (R Core Team 2016) programlama dilinde *psyc* (Revelle & Revelle, 2015), *GDINA*(Ma & de la Torre, 2016), *dplyr* (Wickham ve dięerleri, 2023), *equate* (Albano, 2022), *kequate* (Andersson ve dięerleri, 2022) paketleri kullanılmıştır.

1. Aşama: BTM Modeli Uygulaması

Çalışmanın analiz safhasında ilk olarak BTM çerçevesinde GDINA model uygulamaları gerçekleştirilmiştir. Madde parametreleri ve katılımcıların yetkinlik düzeylerinin belirlenmesinde *expected a posterior (EAP)* metodu kullanılmıştır. Her bir analiz her test formu için ayrı olarak uygulanmıştır. Her testte veriye en uygun modeli seçme aşamasına geçmeden önce Q-matris geçerlilięi *the proportion of variance accounted for (PVAF)* (de la Torre & Chiu, 2016) metodu ile kontrol edilmiştir. Bu metot ile bir kesme deęer belirlenir (çalışmada bu deęer 0.95 alınmıştır) ve bu kesme deęerinin altında kalan *q*-vektörler için madde uygun niteliklere sahip *q*-vektörleri tavsiye edilmektedir. Ayrıca de la Torre and Ma (2016) GDINA analizlerinde Q-matris geçerlilięi kontrolünde *q*-vektörlerin *mesa plot* ile görselleştirilmesini önermişlerdir. Burada önemli olan nokta, *PVAF* deęerleri ve *mesa plot* görsellerinin yardımıyla sunulan tavsiyelerin kabulü araştırmacılara baęlıdır. Duruma göre madde testten çıkarılabilir, analizlerle önerilen *q*-vektör kabul edilerek Q-matris yenilenir ve yeni analizlere geçilir veya tavsiye uygun bulunmayarak analizlere devam edilir. Örnek olması açısından *PVAF* deęerleri, kesme puan çizgisi ve tavsiye edilen *q*-vektörlerin *mesa plot* görselleri aşağıda Test A formuna ait üç farklı durum Şekil 6'da paylaşılmıştır.

Şekil 6

Test A'dan PVAF Değerlerine Göre Mesa Plot Örnekleri



Örneğin bu grafiklerde PVAF değeri madde 7 için 0,86, madde 9 için 0,83 ve madde 11 için 0,98 olarak belirtilmiştir. Madde 7 ve 9 için kesme puanının geçen ilk q -vektör önerilmiştir. Bu örnekte de belirtildiği gibi üç test formunun her birinde Q-matris geçerliliği Erdoğan (2023) tarafından kontrol edilmiş, maddelerin durumu incelenmiş, ilk oluşturulan Q-matris ile analizlere devam edildiği belirtilmiştir. Çalışmada GDINA analizlerinde kullanılan Q-matris aşağıda Tablo 9'da sunulmuştur ve şimdiki çalışmada da aynı Q-matris GDINA model analizlerinde kullanılmıştır.

Tablo 9

Test Formlarına ait Q-matris

Maddeler	Temel Yönler & Bileşenler			Temel Uygulamalar		
	GA	FD	NMY	G&MY	T	G
1	1	0	0	0	1	0
2	1	0	0	1	1	0
3	1	0	0	0	1	0
4	1	0	0	1	0	1

5	0	1	1	1	1	1
6	0	1	0	1	1	1
7	0	1	0	1	0	0
8	1	1	0	1	0	0
9	0	1	0	1	1	0
10	1	0	1	1	0	0
11	0	1	1	1	1	0
12	0	0	1	0	1	0
13	1	0	1	0	1	0
14	1	0	0	1	0	1
15	1	0	1	0	1	0
16	1	0	0	0	1	0
17	0	0	1	0	1	1
18	0	1	1	1	1	0
19	1	0	1	0	1	1
20	0	1	1	1	1	1

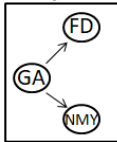
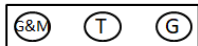
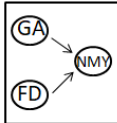



Not: GA: Genelleştirilmiş Aritmetik, FD: Fonksiyonel Düşünme, NMY: Niceliksel Mantık Yürütme, G&R: Genelleme ve Mantık Yürütme, T: Temsilleştirme, G: Gerekçelendirme

Erdoğan (2023) model veri uyumu için GDINA modelinde hem genel saturated model hem de nitelikler arasında olabilecek tüm hiyerarşik yapıları test ederek karşılaştırmıştır. Hiyerarşik yapıları test ederken önce ilk faktördeki nitelikler arasındaki hiyerarşik yapıları test etmiş, ikinci faktördeki nitelikleri serbest bırakmıştır. Bu durumda her test için oniki hiyerarşik yapı test edilmiştir ve model-veri uyumu en uygun olan hiyerarşi belirlenmiştir. Daha sonra ikinci adım olarak, ilk faktördeki nitelikleri serbest bırakmış, ikinci faktördeki nitelikler arasındaki hiyerarşik yapıları test etmiştir. Bu durumda da yine on iki farklı hiyerarşik yapı test edilmiş ve model-veri uyumu en uygun olan hiyerarşik model belirlenmiştir. Son olarak ise birinci durumdaki ve ikinci durumdaki en uygun hiyerarşik yapılar seçilerek birinci ve ikinci faktördeki hiyerarşik yapıları nitelikleri aynı anda karma hiyerarşik yapı modeli (Tu ve diğerleri, 2019) çerçevesinde incelemiştir. Erdoğan (2023) her test için bir kez saturated GDINA modeli (özellikler arasında yapısal

bir ilişki olmadığı varsayılan, ancak özellikler arasındaki etkileşime izin verilen model), 24 kez faktörlerden birine ait niteliklerin hiyerarşik yapıları diğer faktöre ait niteliklerin serbest bırakıldığı GDINA modeli ve bir kez her iki faktöre ait niteliklerin karma hiyerarşik yapıları olarak uygulandığı GDINA model analizleri yaparak toplamda 26 model arasından en uygun model-veri uyumu olan modeli belirlemiştir. Bu işlemleri her test için tekrarlamış ve her testteki model-veri uyumunu kontrol ederek en uygun modeli belirlemiştir. Üç test için toplamda 78 model karşılaştırması yapmıştır. Aşağıda Şekil 7 ve Şekil 8'de Erdoğan (2023, s.85-86) tarafından test edilen ve paylaşılan hiyerarşik yapıları model uygulamalarının görselleri sunulmuştur. Görsellerde nitelikler arasında iraksak (*divergent*), yakınsak (*convergent*) ve doğrusal (*linear*) yapıları niteliklere göre çeşitlendirilerek test edildiği gösterilmektedir.

Şekil 7

Sadece Birinci Faktörün Hiyerarşik Yapılı Olma Durumunun Test Edilmesi

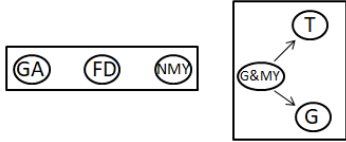
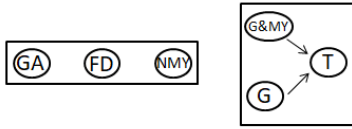
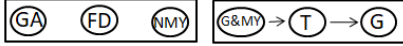
Sadece Birinci Faktör Hiyerarşik Yapılı ise	Test Edilen Yapılar
<p>1. Faktör: Temel Yön ve Bileşenler</p>  <p>2. Faktör: Temel Uygulamalar</p>  <p>Sembol: $(GA \rightarrow FD) \& (GA \rightarrow NMY)$</p>	<p>i. $(GA \rightarrow FD) \& (GA \rightarrow NMY)$ ii. $(FD \rightarrow GA) \& (FD \rightarrow NMY)$ iii. $(NMY \rightarrow GA) \& (NMY \rightarrow FD)$</p>
<p>Yakınsak</p>   <p>Sembol: $(GA \& FD) \rightarrow NMY$</p>	<p>i. $(GA \& FD) \rightarrow NMY$ ii. $(GA \& NMY) \rightarrow FD$ iii. $(FD \& NMY) \rightarrow GA$</p>
<p>Doğrusal</p>   <p>Sembol: $GA \rightarrow FD \rightarrow NMY$</p>	<p>i. $GA \rightarrow FD \rightarrow NMY$ ii. $GA \rightarrow NMY \rightarrow FD$ iii. $FD \rightarrow GA \rightarrow NMY$ iv. $FD \rightarrow NMY \rightarrow GA$ v. $NMY \rightarrow GA \rightarrow FD$ vi. $NMY \rightarrow FD \rightarrow GA$</p>

Not: GA: Genelleştirilmiş Aritmetik, FD: Fonksiyonel Düşünme, NMY: Niceliksel Mantık Yürütme, G&MY: Genelleme ve Mantık Yürütme, T: Temsilleştirme, G: Gerekçeleme

Araştırmacı yapmış olduğu modellere ait test uygulamalarında Şekil 9'de sadece birinci faktörün hiyerarşik yapıları olduğu durumda test edilen yapılardan doğrusal yapı hiyerarşide belirtilen beşinci sıradaki yapının (NMY → GA → FD) modele uyumlu olduğunu belirtmiştir. Bunun yanında aşağıda Şekil 10'da sadece ikinci faktör hiyerarşik yapıları olduğu durumda test edilen yapılardan doğrusal yapı hiyerarşide belirtilen üçüncü sıradaki hiyerarşik yapının (T → G&MY → G) modele uyumlu olduğunu açıklamıştır.

Şekil 8

Sadece İkinci Faktörün Hiyerarşik Yapılı Olma Durumunun Test Edilmesi

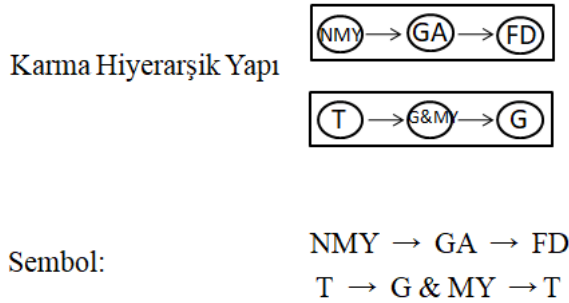
	Sadece İkinci Faktör Hiyerarşik Yapılı ise	Test Edilen Yapılar
	<p>1. Faktör: Temel Yön ve Bileşenler</p> <p>2. Faktör: Temel Uygulamalar</p>	
Iraksak	 <p>Sembol: (G & MY → T) & (G & MY → G)</p>	<p>i. (G&MY → T)&(G&MY → G) ii. (T → G&MY) & (T → G) iii. (G → T)&(G → G&MY)</p>
Yakınsak	 <p>Sembol: (G&MY & G) → T</p>	<p>i. (G&MY & T) → G ii. (G&MY & G) → T iii. (T & G) → G&MY</p>
Doğrusal	 <p>Sembol: G & MY → T → G</p>	<p>i. G&MY → T → G ii. G&MY → G → T iii. T → G&MY → G iv. T → G → G&MY v. G → G&MY → T vi. G → T → G&MY</p>

Not: GA: Genelleştirilmiş Aritmetik, FD: Fonksiyonel Düşünme, NMY: Niceliksel Mantık Yürütme, G&MY: Genelleme ve Mantık Yürütme, T: Temsilleştirme, G: Gerekçeleme

Bu aşamadan sonra araştırmacı Şekil 9'daki ve Şekil 10'daki öne çıkan uyumlu modellere göre her iki faktöre ait nitelikleri hiyerarşik yapılandırarak karma hiyerarşik yapı GDINA model uygulamıştır. Aşağıda Şekil 9'da araştırmacının çalışmasında uyguladığı karma hiyerarşik model paylaşılmıştır.

Şekil 9

Her İki Faktörün Hiyerarşik Yapılı Olma Durumunun Test Edilmesi



Not: GA: Genelleştirilmiş Aritmetik, FD: Fonksiyonel Düşünme, NMY: Niceliksel Mantık Yürütme, G&R: Genelleme ve Mantık Yürütme, T: Temsilleştirme, G: Gerekçeleştirme

Erdoğan (2023) Test A, Test B ve Test C için uyguladığı bu model testlerinden en uygun olan modeli modellere ait *Relative Fit Indeks (RFI)* ve *Likelihood Ratio Test (LRT)* değerleri ile karşılaştırmış ve sonuç olarak her testte karma hiyerarşik yapılı GDINA model veriye en uygun olan model olarak seçilmiştir. Model seçimi aşamasında, *relative fit* indeksleri olarak bilinen -2 Log-likelihood ($-2LL$), Akaike Information Criterion (AIC; Akaike, 1974) and Bayesian Information Criterion (BIC, Schwarz, 1978) değerlerine göre değerlendirme yapılmıştır. Bu indekslerin küçük değerleri daha iyi bir model-veri uyumu olduğu anlamındadır. Ki-kare değerlerine ait p değeri Likelihood Ratio testleri ile incelenmiştir. Tüm karşılaştırmalarda, sıfır hipotezi (*null hypothesis*) kapsamında hiyerarşik yapılandırılmış GDINA modelinin veriye saturated GDINA modeli kadar iyi uyduğu kabul edilmiştir. Öte yandan, alternatif hipotez (*alternative hypothesis*) kapsamında, iki modelin eşdeğer olarak kabul edilmediği yani, hiyerarşik yapının kabul görmediği durum incelenmiştir. Eğer p değeri 0,01'den büyükse, bu hiyerarşik yapılı GDINA modelinin veriye saturated GDINA modelden daha iyi uyduğu anlamına gelmektedir. Aşağıda saturated model, her iki faktöre ait seçilen en uygun hiyerarşik yapılı modeller ile bu hiyerarşik yapıların birlikte test edildiği karma hiyerarşik yapılara ait modellerin relative fit indeks değerleri paylaşılmıştır.

Tablo 10*Test Formlarında Model Uyumu*

	Model	parm. sayısı	RFI			LRT		
			Sapma (Deviance)	AIC	BIC	χ^2	df	p
TEST A	Saturated	283	15679,06	16245,06	17593,56			
	NMY-GA-FD yapı	215	15699,31	16129,31	17153,80	20,25	68	1
	T-G&MY-Gyapı	189	15725,73	16103,73	17004,33	46,68	94	1
	Karma Hiyerarşik Yapı.	148	15794,84	16090,84	16796,07	115,78	135	0,88
TEST B	Saturated	283	14988,30	15554,30	16904,11			
	NMY-GA-FD yapı	215	14966,48	15396,48	16421,95	-21,82		
	T-G&MY-G yapı	189	14999,89	15377,89	16279,35	11,59	94	1
	Karma Hiyerarşik Yapı.	148	15099,09	15395,09	16101,00	110,79	135	0,94
TEST C	Saturated	283	15225,95	15791,95	17139,80			
	NMY-GA-FD yapı	215	15237,86	15667,86	16691,85	11,91	68	1
	T-G&MY-G yapı	189	15265,79	15643,79	16543,94	39,84	94	1
	Karma Hiyerarşik Yapı.	148	15346,37	15642,37	16347,25	120,42	135	0,81

Not 1: GA: Genelleştirilmiş Aritmetik, FD: Fonksiyonel Düşünme, NMY: Niceliksel Mantık Yürütme, G&MY: Genelleme ve Mantık Yürütme, T: Temsilleştirme, G: Gerekçeleştirme

Not 2: parm. sayısı: Parametre Sayısı, RFI: Relative Fit İndeksi, LRT: Likelihood Ratio Test, AIC: Akaike Information Criterion, BIC: Bayesian Information Criterion, χ^2 : Ki-kare değeri, df: degree of freedom (serbestlik derecesi), $p > 0,01$ anlamlıdır.

Tablo 10'da, saturated GDINA modeli ile karşılaştırıldığında her bir testteki karma hiyerarşik yapılı GDINA modellerinin daha düşük AIC ve BIC değerlerine sahip olduğu ve LR testlerinde $p > 0.01$ ile en iyi uyum sağlayan modeller olduğu görülmüştür. Bu nedenle, üç testin her birinde en iyi uyum sağlayan modelin, nitelikleri arasında NMY \rightarrow GA \rightarrow FD ve T \rightarrow G&MY \rightarrow G hiyerarşik yapısal ilişkisi bulunan karma hiyerarşik yapılı GDINA modeli olduğu sonucuna varılmıştır. Seçilen bu model mutlak uyum (*absolute fit*), madde uyumu (*item fit*), kişi uyumu (*person fit*) ve güvenilirlik (GDINA R-paketinde test ve nitelik seviyesinde sınıflama doğruluğu (*accuracy*) değerleri veriliyor.) indeksleri ile desteklemiştir. Aşağıda bu aşamalar sırasıyla paylaşılmıştır.

Aşağıda Tablo 11'de model-veri uyumunda *absolute fit* (mutlak uyumluluk) ve *item fit* (madde uyumluluğu) değerleri paylaşılmıştır.

Tablo 11

Testlere Formlarında Absolute Fit ve Item Fit İndeksleri

Karma Hiyerarşik Yapılı GDINA Model	<i>Absolute Fit</i>				<i>Item Fit</i>	
	<i>Max()</i>	<i>ABS(fcor)</i>	<i>RMSEA2</i>	<i>SRMSR</i>	<i>RMSD</i>	<i>MAD</i>
TEST A	5,22*	0,11*	0,034	0,036	0,026	0,02
TEST B	8,72*	0,12*	0,049	0,036	0,024	0,015
TEST C	11,52*	0,15*	0,05	0,042	0,023	0,015

* $p > 0.05$

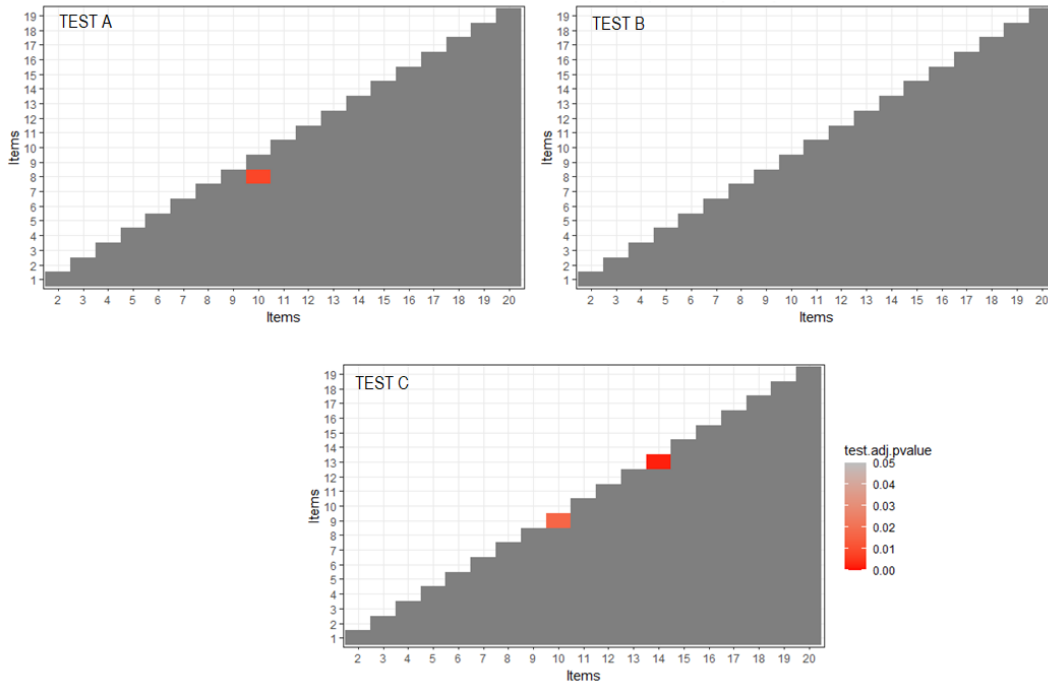
Absolute fit (mutlak uyumluluk), mevcut modelin veriyle ne kadar iyi bir biçimde uyumlu olduğunun değerlendirilmesi sürecidir. Testteki tüm madde çiftleri için bağımsızlık testi istatistiklerinin maksimum değeri modelin uyumsuzluğunu belirlemek amacıyla kullanılmaktadır. Anlamli olmayan bir p değeri, madde çiftinin istatistiksel bağımsızlığının ihlal edilmediği ve modelin başarıyla veriye uyduğu anlamına gelir. Tablo 11'de karma hiyerarşik yapıli GDINA modelin abosulte fit değerlerindirmesinde *Max()* değerlerinin $p > 0,05$ olduğu belirtilmiştir. *ABS(f cor)* istatistiği, Fisher-dönüştürülmüş korelasyonların mutlak sapmalarını ölçer ve anlamli olmayan bir p değeri ile birlikte sifıra yakın değerlerle kabul gören bir model uyumunu bildirir (Chen et al., 2013; Hu et al., 2016). Tablo 11'de karma hiyerarşik yapıli GDINA modelinde *ABS(f cor)* istatistiğine göre $p > 0,05$ olduğu ve anlamli olmadığı belirtilmiştir. Yine *absolute fit* indeksi olarak kullanılan *root mean squared error approximation (RMSEA)* ve *standardized root mean square residual (SRMSR)* değerleri Tablo 11'de paylaşılmıştır. *RMSEA*, serbestlik derecesi başına gözlemlenen kovaryans matrisi ile modelin öngördüğü kovaryans matrisi arasındaki uyuşmazlık miktarının bir ölçüsüdür. *SRMSR* ise gözlemlenen ve modelin öngördüğü kovaryans matrisleri arasındaki standartlaştırılmış artıkların ortalamasının bir ölçüsüdür (Chen, 2007). *RMSEA* değerleri sifirdan bire kadar değişir ve 0,06'dan küçük olan değerler iyi bir uyumu gösterir. *SRMSR* indeksi için kabul edilebilir değerler sifirdan 0,08'e kadar uzanır

(Shafipoor ve diğeri, 2021). Hu et al. (2016) *Limited Information RMSEA (RMSEA2)*'nin 0,05'ten küçük bir değer olmasının iyi bir uyum olduğunu belirtirken, Maydeu-Olivares ve Joe (2014) bu değer 0.089'dan küçük olmasının da kabul edilebilir bir uyum olduğunu da belirtir. Ayrıca *SRMSR* değerlerinin de 0,05'ten küçük olmasının yeterli uyumu ifade ettiğini vurgularlar. Tabloda *RMSEA2* ve *SRMSR* değerlerinin her bir test için 0,05'ten düşük olduğu görülmektedir. Dolayısıyla, Tablo 11'de belirtilen tüm *absolute fit (mutlak uyumluluk)* indekslerinin karma hiyerarşik yapıları GDINA modeline uygun olduğu görülmektedir.

Yine Tablo 11'de model-veri uyumunu destekleyen item fit (madde uyumluluk) indeks değerleri paylaşılmıştır. Bu kapsamda *Root Mean Square Deviation (RMSD)*; eski yayınlarda *RMSEA* olarak adlandırılır ve *Mean Absolute Deviation (MAD)* değerleri kullanılmıştır. *RMSEA*, gözlemlenen bir madde özelliği ile model tarafından öngörülen bir özellik arasındaki farkın standartlaştırılmış bir ölçüsüdür. Eğer *RMSEA*=0 ise, bu model ile veri arasında hiçbir uyumsuzluk olmadığını gösterir ve *RMSEA* değeri 0.1'den büyük olan modellerin iyi bir uyuma sahip olmadığı ve göz ardı edilmesi gerektiğini belirtilir (Oliveri & von Davier; 2011). *MAD*, her bir veri noktasının ortalamaya olan ortalama mesafesini verir ve bir veri setindeki değişkenliğe ilişkin bilgi sunar. Tablo 11'de *RMSD* ve *MAD* değerlerinin sıfıra yakın olması, maddelerin karma hiyerarşik yapıları GDINA modeline uyumlu olduğunu göstermektedir. Madde uyumunda *RMSD* ve *MAD* değerleri ile her bir maddenin seçilen modele tek tek ne kadar iyi uyum sağladığının değerlendirilmesinin yanında R GDINA paketinde maddelerin ikili karşılaştırmaları ile bu uyum incelenebilmektedir ve bu karşılaştırma ısı haritaları ile sunulabilmektedir. Aşağıda Şekil 10'da her test formuna ait maddelerin ikili karşılaştırılmasıyla model uyumlarının incelenmesi ısı haritaları ile sunulmuştur.

Şekil 10

Her Test Formundaki Madde Çiftleri İçin Isı Haritası



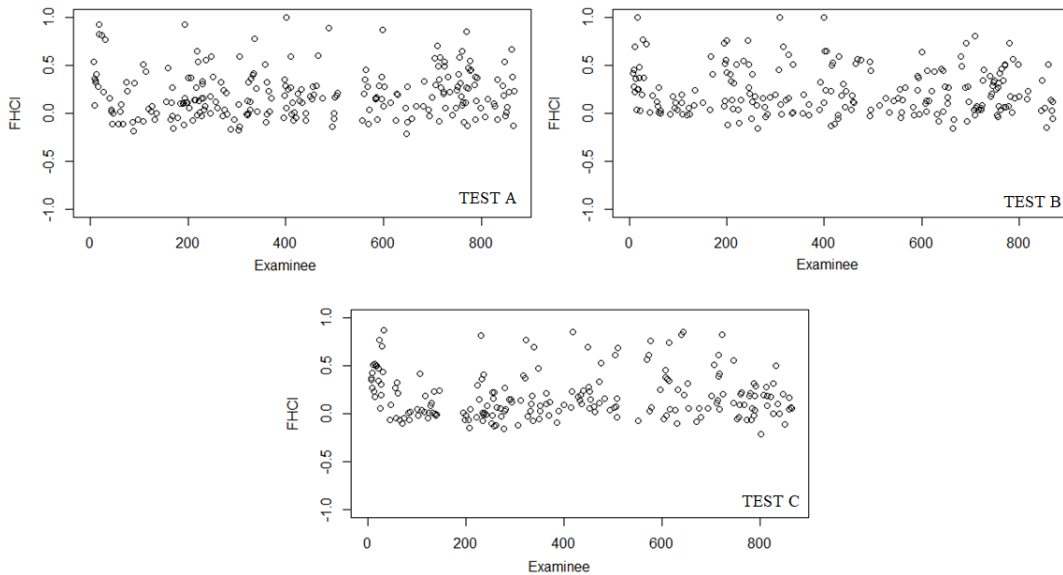
Eğer gözlemlenen hatalı madde çifti sayısı, şansa bağlı olarak beklenen hatalı madde çifti sayısından daha düşükse, modelin veriye tatmin edici bir uyum sağladığı belirtilmiştir (Carragher et al., 2019). Şansa bağlı olarak beklenen hatalı madde çifti sayısının hesaplanması toplam madde çifti sayısının 0,05 katına eşittir. Bu çalışmada her testte 20 madde olduğu için 190 madde çifti ($20 \times 19 / 2 = 190$) bulunmakta ve test başına yaklaşık 10 madde çiftinin ($0,05 \times 190 = 9,5$) tesadüfen uyumsuz olması beklenebilmektedir. Şekil 10'deki görsellerde Test B'de hiçbir uyumsuz madde çiftinin bulunmadığını ve Test A ve C'de ihmal edilebilir düzeyde sırasıyla yalnızca bir ve iki uyumsuz madde çiftinin bulunduğu görülmektedir. Bu durum karma hiyerarşik yapılı GDINA modelinin verilere iyi uyum gösterdiği anlamına gelmektedir.

Person fit (kişi uyumu), sınava girenlerin yanıt örüntülerinin modele uyumunu analiz etmek için kullanılır. Çalışmada seçilen modelle yanıt tutarlılığını değerlendirmek için *Full Hierarchy Consistency Index (FHCI, Tam Hiyerarşi Tutarlılık İndeksi)* (Akbay ve Kılınç, 2018) kullanılmıştır. Madde türü, niteliklerin yapısı ve sınava giren kişilerin ölçülen

niteliklere göre ait oldukları örtük sınıflar gibi faktörlerden bağımsız olarak hesaplanan *FHCI* değerlerinin kişi uyumu için 0,00 ile 1,00 arasında değerler alması beklenmektedir. Aşağıda Şekil 11’de testlere göre *FHCI* değerleri saçılım grafiklerinde paylaşılmıştır. Bu grafiklere göre *FHCI* değerlerinin 0,00 ile 1,00 arasında dağıldığı gözlenmektedir, özellikle 0,00 ile 0,5 arasında yoğunlaşmış olduğu söylenebilir. Sıfırın altındaki değerlerin sifıra çok yakın ve ihmal edilebilir düzeyde olduğu belirtilebilir. Sonuç olarak, yanıt örüntülerinin karma hiyerarşik yapıları GDINA model ile uyumlu olduğu söylenebilmektedir.

Şekil 11

Kişi Uyumunu Gösteren FHCI Değerlerinin Saçılım Grafikleri



GDINA paketi, test ve nitelikler düzeyinde sınıflandırma doğruluğu sunar ki bu güvenilir bir gösterge olarak kabul edilir (Mei & Chen, 2022). Test katılımcılarının uygun örtük sınıflarına doğru bir şekilde kategorize edildikleri derece, test düzeyindeki sınıflandırma doğruluğu olarak bilinir ve 0.80’den yüksek değerler tatmin edici olarak kabul edilir (Ravand ve Robitzsch, 2018). Testte ölçülen her bir nitelik için test katılımcılarının ilgili niteliklerde “yetkin” veya “yetkin olmayan” gruplara doğru bir şekilde kategorize edildikleri derece, nitelik düzeyindeki sınıflandırma doğruluğu olarak bilinir. Güvenirlilik olarak 0,20’den küçük değerler yetersiz; 0,25-0,50 arası zayıf; 0,50-0,65 arası orta; 0,65-0,80 arası iyi; 0,80-0,90 arası çok iyi ve 0,90’dan büyük değerler mükemmel güvenilirliği

temsil eder (Shafipoor et al., 2021). Aşağıda Tablo 12’de çalışmada kullanılan test formlarına ait test ve nitelik seviyesindeki sınıflandırma doğruluk düzeyleri (*accuracy*) paylaşılmıştır.

Tablo 12

Test Formlarının ve Niteliklerin Sınıflandırma Doğruluk Düzeyleri

Karma Hiyerarşik Yapılı GDINA Model	Test Güvenirliği	Nitelik Güvenirlikleri					
		GA	FD	NMY	G&MY	T	G
TEST A	0,66	0,90	0,85	0,88	0,89	0,88	0,93
TEST B	0,66	0,87	0,85	0,87	0,89	0,92	0,86
TEST C	0,66	0,85	0,86	0,85	0,87	0,89	0,87

Not: GA: Genelleştirilmiş Aritmetik, FD: Fonksiyonel Düşünme, NMY: Niceliksel Mantık Yürütme, G&MY: Genelleme ve Mantık Yürütme, T: Temsilleştirme, G: Gerekçelendirme

Tablo 12’de görüldüğü üzere, her bir test için test düzeyindeki doğruluk indeksi 0.66 olarak belirlenmiştir, bu da makul güvenilirlik aralığının üst sınır değeri ve iyi güvenilirlik aralığının alt sınır değeri ile uyumludur. Nitelik düzeyinde ise doğruluk değerlerinin 0.85 ile 0.93 arasında değiştiği gözlenmektedir, bu da çok iyi ve mükemmel güvenilirlik aralıklarına denk gelmektedir.

Sonuç olarak, burada bahsedilen *relative fit* (göreceli uyum), *absolute fit* (mutlak uyum), *item fit* (madde uyumu) ve *person fit* (kişi uyumu) bilgileri ışığında karma hiyerarşik yapıları GDINA modelinin katılımcılardan toplanan yanıt örüntülerini ne derece iyi bir biçimde temsil ettiği anlaşılmaktadır. Bu durum ise yapılan istatistiksel analizlerin sonuçlarının geçerlik ve güvenilirliğine sağlam desteklerdir. Bütün bu bilgilere ait açıklamalar araştırmacıya ait Erdoğan (2023)’teki çalışmasında da detaylarıyla paylaşılmıştır.

Karma hiyerarşik yapıları GDINA model analizi ile testlere katılan bireylere ait her bir nitelik için yetkinlik düzeyinin belirlenebildiği vurgulanmıştır. Aşağıda Şekil 14’te R programında GDINA pakette yapılan analizler sonucu oluşturulan bir tabloya ait ekran kesiti paylaşılmış olup bu kesitte her bir bireye ait genel test puanının yanı sıra karma

hijerarşik yapılı GDINA model analizi sonunda bireysel beceri yetkinlik düzeyleri örnek olarak paylaşılmıştır.

Şekil 12

Bireylere Ait Test Puanı ve Beceri Yetkinlik Düzeyleri Örneği

Test_Pn	GA	FD	NMY	G&MY	T	G
14	1.00	0.93	1.00	0.98	0.99	0.81
14	1.00	0.89	1.00	0.98	1.00	0.67
19	1.00	0.96	1.00	1.00	1.00	0.86
17	1.00	0.92	1.00	0.99	1.00	0.85
12	0.99	0.81	1.00	0.77	0.99	0.40
11	1.00	0.85	1.00	0.91	0.97	0.69
7	0.97	0.52	0.98	0.47	0.78	0.37
9	0.66	0.57	0.70	0.88	1.00	0.49
13	0.96	0.90	1.00	0.98	0.98	0.83
11	0.99	0.74	1.00	0.67	1.00	0.32
12	0.98	0.83	0.99	0.67	1.00	0.41

Not: Test_Pn: Genel Test Puanı (20 üzerinden), GA: Genelleştirilmiş Aritmetik Yetkinlik Düzeyi, FD: Fonksiyonel Düşünme Yetkinlik Düzeyi, NMY: Niceliksel Mantık Yürütme Yetkinlik Düzeyi, G&MY: Genelleme ve Mantık Yürütme Yetkinlik Düzeyi, T: Temsilleştirme Yetkinlik Düzeyi, G: Gereçlendirme Yetkinlik Düzeyi

Şekil 12'ye göre örneğin maddelere verilen doğru cevaplar üzerinden genel test puanı 20 üzerinden 9 olan bir öğrencinin NMY becerisi yetkinlik düzeyi 0,70 iken T becerisi yetkinlik düzeyi 1.00 olarak belirlenmiştir. Yine örneğin maddelere verilen doğru cevaplar üzerinden genel test puanı 20 üzerinden 7 olan bir katılımcının FD becerisi yetkinlik düzeyi 0,52 iken GA becerisi yetkinlik düzeyi 0,97 olarak sonuçlanmıştır. Burada önemle vurgulanması gereken bir nokta ise eşit genel test puanına sahip olan öğrencilerin beceri yetkinlik düzeylerinin farklılık gösterebilmesidir. Bilişsel tanı modellerini diğer modellere göre eğitimde etkin kılan yönü de burasıdır. Bu araştırma kapsamında ise çalışmanın 1. aşamasında elde edilen ve bu ekran görüntüsünde örneği paylaşılan beceri yetkinlik düzeyleri çalışmanın 2. aşamasındaki analizlerde girdi olarak kullanılmıştır.

2. Aşama: Beceri Puanları Eşitleme Uygulamaları

Çalışmanın analiz safhasının ikinci aşamasında ise karma hiyerarşik yapılı GDINA analizlerinden elde edilen bireylere ait beceri yetkinlik düzeyleri eşitleme uygulamaları için veri olarak kullanılmıştır. BTM analizlerinden elde edilen bireylere ait yetkinlik düzeyleri 0,00-1,00 arasında ihtimal değerleri olarak verildiğinden bu değerler beceri puanlarına çevrilmiştir.

Test eşitlemesi yapılmadan önce eşitlenecek puanların dağılımlarının belirlenmesi gereklidir. Eşitleme fonksiyonları bu dağılımı kullanmaktadır. Eşitleme yapmadan önce, verilerin kalitesini artırmak için birkaç adım bulunmaktadır. Bunlardan birisi de puan dağılımlarının sürdürülmesi ve düzgünleştirilmesidir (Dorans ve diğerleri, 2010). Gözlenen puanların eşitlenmesi için eşitlenecek olan her iki forma ait gözlenen puan dağılımlarının bulunması gerekir (Kolen & Brennan, 2013). Örnekleme büyüklüğü ne kadar büyük olursa puan dağılımı popülasyonu o kadar iyi temsil eder, ancak örneklem hatası nedeniyle bu dağılımlarda düzensizlikler görülmekte ve kesikli dağılımın iyi tahmin edilmesine ihtiyaç duyulmaktadır (González & Wiberg, 2017). Eğer bir eşitleme çalışmasında çok büyük örneklem veya tüm popülasyon mevcut olsa puan dağılımları ve eşit yüzdellik ilişkileri oldukça düzgün olurdu (Kolen & Brennan, 2013). Burada belirtilenlerden anlaşıldığı üzere test eşitlemesi yapılmadan önce eşitlenecek puanların dağılımları önemlidir. Eşitleme sürecinde eşitlenecek olan puanlar değişken konumuna geçmektedir. Bu puanların her birinin örnekleme yeterli sayıda temsil edilmesi eşitleme yapılabilmesi için gereklidir. R programında kullanılan eşitleme paketlerinde de eşitleme sürecine hazırlık olarak puanlara ait frekans tabloları oluşturulmakta ve bu tablolar üzerinden eşitleme süreci başlatılmaktadır. Çalışma kapsamında becerilere ait yetkinlik düzeyleri kullanılarak beceri puanları elde edilme aşamasında bu puanlara ait olan ölçeğin de belirlenmesi işlemi beceri puanlarının eşitleme süreci açısından önemlidir. Her bir testi alan öğrenci sayısı 860 civarında olduğu için burada belirtilenler ışığında beceri puanı ölçeğindeki her bir kesikli puanın eşitleme sürecinde yeterli sayıda temsil edilebilmesi için puan ölçeğinin dar bir

ranja sahip olmasının uygun olacağı çıkarımı yapılmıştır. Eşitleme literatüründe bu puan ölçeğinin değişimi ve örneklem ilişkisine dair net bir bilgi sunulmamakla birlikte her puanın yeteri kadar temsil edilmesi yorumu test eşitlemenin teori ve uygulamaları çerçevesinde bilinmektedir. Örneğin, eşit yüzdeliği eşitleme için puan dağılımının gergin ve sıkışık (puan aralıklarının daha küçük) olması ile bir formun puan dağılımının diğer formun puan dağılımına denk geleceği belirtilir (Angoff, 1971; Woldbeck, 1998). Bu sebeple bu çalışmada eşitleme incelemelerinde karma hiyerarşik yapılı GDINA modelinden elde edilen beceri yetkinlik düzeylerinin beceri puanlarına 25 puan üzerinden dönüştürülmesi uygun görülmüştür. Beceri yetkinlik düzeylerinin beceri puanlarına dönüştürülürken farklı puan ölçekleri (10, 20, 50, 100 puan gibi) ve eşitleme durumlarının incelenmesi ile ilgili çalışmalar farklı araştırmalar için konu olabileceği de burada vurgulanabilir.

Bu çalışma kapsamında her bir beceri puanı 25 puan üzerinden hesaplanmıştır. Beceriye ait yetkinlik düzeyi ile belirlenen 25 puan çarpıldığında her bireye ait olan beceri puanları elde edilmiştir. Örneğin bir bireye ait ilgili beceride yetkinlik düzeyi olarak 0,85 ihtimal belirtilmiş ise bu bireyin beceri puanı $0,85 \times 25 = 21,25$ 'tir. Eşitlemenin yapılabilmesi için kesikli puanlar elde etmek amacıyla birer basamağına göre yuvarlama işlemi uygulanarak ilgili beceri 21 olarak belirlenmiştir.

Şekil 13

Bireysel Beceri Yetkinlik Düzeylerine Karşılık Gelen Beceri Puanları Örneği

Test_Pn	GA	FD	NMY	G&MY	T	G	GA_BPn	FD_BPn	NMY_BPn	G&MY_BPn	T_BPn	G_BPn
14	1.00	0.93	1.00	0.98	0.99	0.81	25	23	25	25	25	20
14	1.00	0.89	1.00	0.98	1.00	0.67	25	22	25	24	25	17
19	1.00	0.96	1.00	1.00	1.00	0.86	25	24	25	25	25	22
17	1.00	0.92	1.00	0.99	1.00	0.85	25	23	25	25	25	21
12	0.99	0.81	1.00	0.77	0.99	0.40	25	20	25	19	25	10
11	1.00	0.85	1.00	0.91	0.97	0.69	25	21	25	23	24	17
7	0.97	0.52	0.98	0.47	0.78	0.37	24	13	25	12	20	9
9	0.66	0.57	0.70	0.88	1.00	0.49	16	14	17	22	25	12
13	0.96	0.90	1.00	0.98	0.98	0.83	24	23	25	25	25	21
11	0.99	0.74	1.00	0.67	1.00	0.32	25	18	25	17	25	8
12	0.98	0.83	0.99	0.67	1.00	0.41	24	21	25	17	25	10

Not: Test_Pn: Genel Test Puanı (20 üzerinden), GA: Genelleştirilmiş Aritmetik Yetkinlik Düzeyi, FD: Fonksiyonel Düşünme Yetkinlik Düzeyi, NMY: Niceliksel Mantık Yürütme Yetkinlik Düzeyi, G&MY: Genelleme ve Mantık Yürütme Yetkinlik Düzeyi, T: Temsilleştirme Yetkinlik Düzeyi, G: Gerekleştirme Yetkinlik Düzeyi, GA_BPn: Genelleştirilmiş Aritmetik Beceri Puanı, FD_BPn: Fonksiyonel Düşünme Beceri Puanı, NMY_BPn: Niceliksel Mantık Yürütme Beceri Puanı, G&MY_BPn: Genelleme ve Mantık Yürütme Beceri Puanı, T_BPn: Temsilleştirme Beceri Puanı, G_BPn: Gerekleştirme Beceri Puanı

Şekil 13'e göre örneğin maddelere verilen doğru cevaplar üzerinden genel test puanı 20 üzerinden 9 olan bir öğrencinin NMY becerisi 0,70 yetkinlik düzeyinde ve beceri puanı olarak 25 üzerinden 17 puana denk gelmektedir. Yine örneğin maddelere verilen doğru cevaplar üzerinden genel test puanı 20 üzerinden 7 olan bir katılımcının FD becerisi 0,52 yetkinlik düzeyinde ve beceri puanı olarak 25 üzerinden 13 puana denk gelmektedir.

Görüldüğü üzere, nitelikler açısından bireysel yetkinlik düzeyleri yine nitelikler açısından bireysel beceri puanlarına dönüştürülmüştür. Klasik test kuramına göre eşitleme yöntemlerinde gözlenen test puanların eşitlenmesi durumu yani testlerden alınan toplam test puanlarının eşitlenmesi sık kullanılan bir eşitlemedir. Bu çalışmada ise klasik eşitleme yöntemlerinin beceri yetkinlik düzeylerinden elde edilen beceri puanları için incelenmesi öne çıkmaktadır. Eşitlemede gözlenen bir test puanı yerine örtük bir beceri puanı kullanılması hedeflenmiştir.

Bu çalışma kapsamında altı niteliğe ait yetkinlik düzeyinden altı beceri puanı elde edilebilmiştir ancak eşitleme çalışmaları tekrarlardan sakınmak ve çalışmayı altı nitelik için tekrarlı analizlerle karmaşıkleştirmamak adına sadece birinci faktöre bağlı üç niteliğe ait beceri puanları üzerinden incelenmiştir. Bu üç nitelik ile yapılan eşitleme çalışmalarının bu tez kapsamındaki hedefler için yeterli olduğu düşünülmüştür. Çalışmada üç test formu birbiriyle çaprazlanarak (AxB, AxC, BxC) üç ayrı eşitleme uygulaması farklı desenlerde incelenmiştir. Her bir eşitleme uygulaması KTK temelinde ve üç eşitleme deseni üzerindedir. İzlenen adımlar aşağıda sırasıyla paylaşılmıştır.

i. Denk gruplar (DG) deseni. DG deseninde ortalama, doğrusal, eşit yüzdellikli ve Kernel türü eşitlemeler yapılmıştır. Kernel eşitlemede *uniform* ve *logistic kernel* yöntemleri uygulanmış olup DG desenlerinde *gauss* kernel yönteminin uygun olmadığı belirtilmiştir (González & Wiberg, 2017). DG deseninde ortalama, doğrusal, eşit yüzdellikli ve Kernel türü eşitlemelerinin uygunluğu standart hatalar açısından kontrol edilmiş olup, karşılaştırmalarla en uygun eşitleme türü de belirtilmiştir.

ii. **Denk olmayan gruplarda ortak madde (DOG-OM) deseni.** DOG-OM desenlerinde de doğrusal, eşit yüzdelikli ve Kernel türü eşitlemeler uygulanmıştır. Doğrusal eşitleme türünde *Tucker, Levine Observed-Score, Levine True-Score, Braun-Holland* ve *chained linear* eşitleme yöntemleri uygulanmıştır. Eşit yüzdelikli eşitleme türünde *chained equipercentile equating* ve *frequency estimation equating* eşitleme yöntemleri kullanılmıştır. DOG-OM desenine göre ortalama, doğrusal ve eşit yüzdelikli eşitleme türlerinin uygunluğu *Bootstrap Standard errors of equating, SE, Bias and RMSE* açısından kontrol edilmiş olup, karşılaştırmalarla en uygun eşitleme türü de belirtilmiştir. Kernel eşitleme türünde ise *poststratification equating (PSE) (frequency estimation kernel* eşitlemede PSE olarak bilinmektedir (González & Wiberg, 2017)), *chained equating (CE), Gaus, uniform, logistic kernels* yöntemleri kullanılmıştır. Yapılan kernel eşitlemenin uygunluğu *standard error of equating (SEE), the percent relative error (PRE), standart error of equating differences (SEED)* değerleri ile kontrol edilmiştir.

iii. **denk olmayan gruplarda ortak değişken (DOG-OD) deseni.** Ortaokulu tamamlayan dokuzuncu sınıf veya hazırlık sınıfı öğrencilerinden cebirsel düşünme becerisi tanılayıcı testlerinden elde edilen yanıt örüntülerinin yanında okulların kabul koşulu (LGS puan kriteri ile öğrenci kabul eden ve etmeyen), bu okulların türleri (AL, AİHL, FL, MTAL, SBL) ve cinsiyet bilgileri de toplanmıştır. İlk olarak okulların kabul koşulunun bulunup bulunmama durumu ve okul türleri ortak değişken olarak belirlenmiş ve bu değişkenlere göre test eşitleme durumu incelenmiştir. İkinci olarak, ortak değişken değiştirilerek okulların kabul koşulunun bulunup bulunmama durumuna ve cinsiyet değişkenlerine göre ikinci bir eşitleme durumu incelenmiştir. Burada yapılan eşitlemeler de kernel eşitleme türüyle yapılmış olup PSE ve CE metotları kullanılmıştır. Eşitlemelerin uygunluğu *standard error of equating (SEE), the percent relative error (PRE), standart error of equating differences (SEED)* değerleri ile kontrol edilmiştir.

Tablo 13*Eşitleme Bulguları Tablosu*

<i>Eşitleme Desenler</i>	<i>Eşitleme Türü</i>	<i>Eşitlenen Testler</i>		
		<i>AxB</i>	<i>AxC</i>	<i>BxC</i>
<i>Denk Gruplar</i>	Ortalama Doğrusal Eşit Yüzdelikli Kernel			
<i>Denk Olmayan Gruplarda Ortak Madde</i>	Doğrusal Eşit Yüzdelikli Kernel			
<i>Denk Olmayan Gruplarda Ortak Değişken</i>	Kernel (değişkenler: okul kabul koşulu & okul türü Kernel (değişkenler: okul kabul koşulu & cinsiyet			

Bulgular kısmında her bir beceri puanına, eşitleme deseni ve türüne göre analiz sonuçları ayrı ayrı paylaşılması ardından Tablo13'te eşitleme yapılıp yapılamadığına göre \sqrt veya – ile doldurularak eşitleme yapılabilir ise hangi yöntemin tercih edilebilir olduğu belirtilecektir. Böylece tabloda gözlenen genel duruma göre sonuçlar yorumlanacaktır.

Çalışma Grubuna Ait Betimleyici Bilgiler

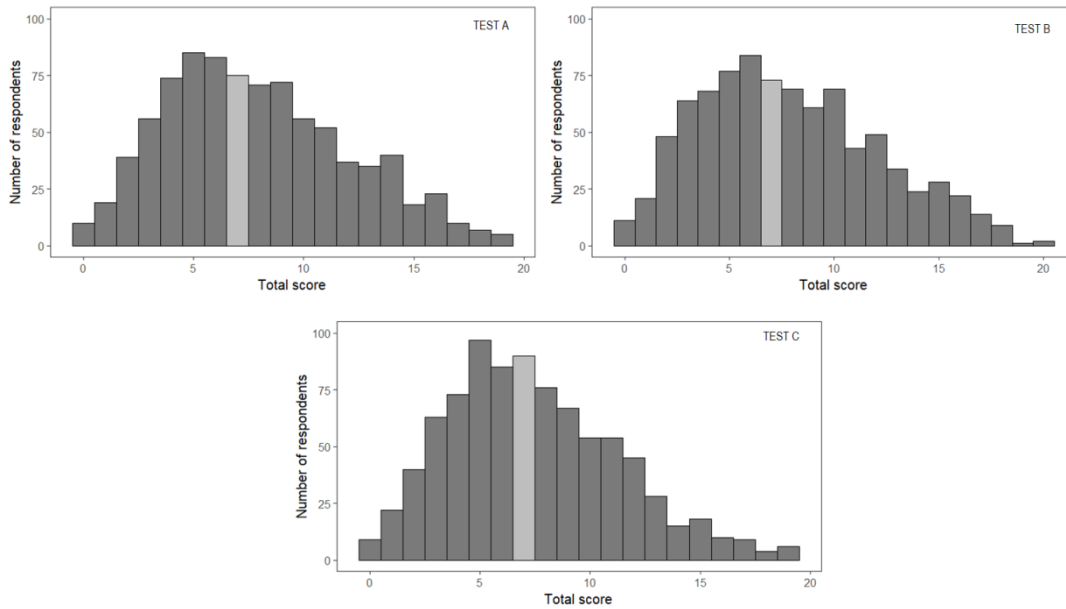
Burada örnekleme ait genel bilgiler ve test istatistiklerinin paylaşılmasındaki amaç paralel formlar oluşturmak hedefiyle hazırlanan testlerin yapı bakımından paralelliğinin yanında bu testlerden elde edilen genel test puanlarının ve madde istatistiklerinin de gruplar arasındaki durumunun karşılaştırılarak kontrol edilmesi bu formların paralelliğine yine destek sunan kanıtlardır. Bu kanıtlar bu testlerin analizleriyle elde edilen beceri puanlarının klasik yöntemlerle eşitleme çalışmalarında kullanılmasında her biri dayanak olmaktadır. Bu sebeple maddelere verilen doğru-yanlış yanıtlardan elde edilen genel test puanlarına ve testlerin genel istatistiklerine ait bilgilerin paylaşılması önemlidir. Böylelikle paralel formların karma hiyerarşik yapıları GDINA çerçevesindeki analizlerden elde edilen

beceri yetkinlik düzeylerinin beceri puanlarına dönüştürülerek eşitleme çalışmalarının temelinde paralel test uygulamalarına bağlı olduğu vurgulanmıştır. Bu bağlamda maddelere verilen yanıtlardan elde edilen genel test puanlarına ve ortak maddelere ait betimleyici istatistiklere, testlerin KTK'ya göre ortalama güçlüğü ve güvenilirlik indeksleri, maddelerin KTK'ya göre madde güçlük ve ayırtedicilik indeksleri ve yine maddelerin GDINA'ya göre yetkin ve yetkin olmayan gruplar tarafından cevaplama olasılıkları, her testte bulunan ortak maddelerin betimsel istatistikleri ve bu ortak maddelerden alınan puanların genel test puanları ile korelasyonları yukarıda bahsedildiği gibi bu testlerden elde edilen beceri puanlarının eşitleme incelemelerine dayanak olması amacıyla aşağıda sırasıyla paylaşılmıştır.

Her testteki örnekleme ait test puanları dağılımları Şekil 14'de ve bu dağılımlara ait bilgiler Tablo 14'te paylaşılmıştır.

Şekil 14

Test Puanları Dağılımı



Tablo 14*Genel Test Puanlarına Ait Betimsel İstatistikler*

	<i>n</i>	<i>Ortalama</i>	<i>Medyan</i>	<i>Mod</i>	<i>Standart Sapma</i>	<i>Çarpıklık K.</i>	<i>Basıklık K.</i>
TEST-A	867	7,90	7	5	4,14	0,42	-0,48
TEST-B	871	7,86	7	6	4,25	0,41	-0,52
TEST-C	865	7,44	7	5	3,88	0,53	-0,08

Tablo 14'teki bilgiler ve Şekil 14'teki grafiklerden görüleceği üzere grupların ortalama, medyan ve mod değerlerinin çok yakın ve hafif sağa çarpık bir dağılım sergiledikleri görülmektedir. Dağılımlarda normallik açısından genel bir kılavuz olarak -1 ile $+1$ arasındaki çarpıklık değeri mükemmel olarak kabul edilirken -2 ile $+2$ arasındaki bir değer ise kabul edilebilir olarak görülür. -2 'den küçük ve $+2$ 'den büyük değerler de önemli düzeyde normal olmayan bir durumu gösterdiği belirtilir (Hair ve diğerleri, 2022). Ayrıca eğer basıklık ($+2$)'den büyükse, dağılım çok sivridir ve -2 'den küçük bir basıklık ise çok düz bir dağılımı göstermektedir. Hem çarpıklık hem de basıklık sifıra yakınsa, yanıt deseni normal bir dağılım olarak kabul edilir (George & Mallery, 2019). Tablo 14'te görüldüğü üzere dağılımların çarpıklık ve basıklık katsayıları 0'a yakın düzeydedir ve öğrencilerin genel test puanlarının dağılımı normal dağılım olarak varsayılabilir.

Normal dağılıma sahip olduğu varsayılan bu üç grup arasında varyans farklılıklarını değerlendirmek amacıyla Levene'in varyans homojenliği testi tercih edilmiştir. Levene testi, iki veya daha fazla gruba ait bir değişkenin varyanslarının eşitliğini değerlendirmek için kullanılan bir istatistiktir. Yapılan analizlerde Levene testine göre bu üç grup arasında varyanslar açısından anlamlı bir fark olduğu anlaşılmıştır ($F(2, 2600)=5,46; p>0,05$). Her üç test için F -testi testi ($\alpha =0,05$ ile gruplar arasında varyans homojenliği değerlendirilmiştir. Test sonuçlarına göre A ve B testleri için gruplar arasında istatistiksel olarak anlamlı bir varyans farklılığı bulunmamaktadır ($F(866, 870)=0,95, p>0,05$). Bunun yanında B ve C Testleri için gruplar arasında istatistiksel olarak anlamlı bir varyans farklılığı bulunmaktadır ($F(866, 864)=1,14;p<0.05$) ve bu

durum B ve C Testleri için de aynıdır ($F(870, 864) = 1,20; p < 0.05$). Sonuç olarak, Test C'nin varyansı diğer iki testten anlamlı olarak farklılık göstermektedir.

Ayrıca bu üç testten elde edilen genel test puanlarının ortalamaları arasında anlamlı bir fark olup olmadığı da incelenmiştir. Bu gruplarda varyansların homojenliği sağlanmadığı için ortalamaları arasındaki farkın incelenmesinde ANOVA testi yerine Kruskal-Wallis testinin tercih edilmesi daha güvenilir sonuçlar vermektedir. Kruskal-Wallis testi, non-parametrik bir ANOVA türüdür ve üç veya daha fazla bağımsız grup arasındaki medyan farkını test etmek için kullanılır. Bu üç grup için Kruskal-Wallis testi uygulandığında gruplar arasında medyan değerlerinde anlamlı bir fark bulunmadığı görülmüştür ($H(2) = 5,05, p > 0,05$).

Ortak maddelere ait betimsel istatistikler ise genel çalışma grubuna göre, sınav koşuluyla ve sınav koşulu olmadan öğrenci kabul eden okullara göre Tablo 15'te paylaşılmıştır.

Tablo 15

Ortak Maddelere Ait Betimsel İstatistikler

	<i>Genel</i>			<i>Sınav Koşulu</i>			<i>Sınav Koşulsuz</i>			<i>Test ile Korelasyonu</i>
	<i>n</i>	<i>M</i>	<i>ss</i>	<i>n</i>	<i>M</i>	<i>ss</i>	<i>n</i>	<i>M</i>	<i>ss</i>	
TEST-A	867	2,49	1,46	595	2,86	1,44	272	1,68	1,13	0,82
TEST-B	871	2,43	1,50	591	2,84	1,46	280	1,55	1,17	0,85
TEST-C	865	2,44	1,46	605	2,80	1,43	260	1,62	1,17	0,82

6 ortak maddenin genel ortalamasının en yüksek 2,5 civarında olduğu, bunun sınav koşuluyla öğrenci kabul eden okullarda 3,00 civarında ve sınav koşulu olmayan okullarda ise 1,7 civarında olduğu görülmektedir. Ortak maddeden alınan toplam puanlar ile genel test puanların korelasyonu ise 0,80 civarındadır. Genel olarak sınav koşulu ile öğrenci kabulü yapmayan okullarda öğrencilerin ortak maddelerde düşük ortalamaya sahip olduğu görülmektedir. Sınav koşuluyla öğrenci kabul eden okulların ise ortalaması orta düzeydedir.

Test formlarına ait test ortalama güçlükleri ve güvenilirlikleri aşağıda Tablo 16'da paylaşılmıştır.

Tablo 16

KTK'ya göre Testlerin Ortalama Güçlük ve Güvenirlik İndeksleri

	Güçlük	α
TEST-A	0,50	0,78
TEST-B	0,51	0,81
TEST-C	0,50	0,76

Test puanlarının aritmetik ortalamasının testten alınacak genel tam puana oranı test güçlüğüdür (Özçelik, 2010). Yani her maddenin 1 puan olarak kabul edildiği bir puanlamada testteki doğru sayısı ortalamasının madde sayısına oranı da test güçlüğüdür. Bu oran 0 ile 1 arasındadır ve 1'e ne kadar yakınsa test o kadar kolay, 0'a ne kadar yakınsa test o kadar zordur anlamına gelmektedir. Tablo 16'da üç testin her birinin ortalama güçlüğü 0,50 civarında olduğu görülmektedir ve bu testlerin orta güçlükte oldukları söylenebilir. Buna ek olarak Özçelik (2010), eğitimde kullanılan testlerin güvenilirliğinin, soru sayısı çok az olan testler dışında, genellikle 0,80'in altına düşmemesi gerektiğini ifade eder. Aynı zamanda, öğrencilere yönelik alınacak ciddi kararlara temel olacak ölçümler elde etmek için kullanılacak testlerin güvenilirliğinin 0,90'ın üzerinde olması gerektiğini vurgular. Tablo 16'da bu yönden bakıldığında 20 maddelik test formlarına ait güvenilirliklerin 0,80 civarında olduğu ve yeterli görüldüğü söylenebilir.

Testlerin ortalama güçlüklerinin yanında testlere ait madde güçlüklerinin ve ayırt ediciliklerinin belirtilmesi test formlarının paralelliğinin maddesel boyuta kadar incelenmesine ve bu paralelliğe etki edebilecek durumların saptanmasına katkı sağlayacaktır. Fakat burada kısaca genel bir bakış sunmak amacıyla aşağıda Tablo 17'de KTK'ya göre test maddelerinin güçlük ve ayırt edicilik indeksleri paylaşılmıştır.

Tablo 17*KTK'ya Göre Testlerin Madde Güçlük ve Ayırt Edicilik İndeksleri*

<i>Madde</i>	<i>Madde Güçlükleri</i>			<i>Madde Ayırt Edicilikleri</i>		
	<i>TEST A</i>	<i>TEST B</i>	<i>TEST C</i>	<i>TEST A</i>	<i>TEST B</i>	<i>TEST C</i>
1	0,64	0,70	0,64	0,76	0,70	0,60
2	0,48	0,50	0,50	0,67	0,72	0,56
3	0,69	0,70	0,67	0,61	0,74	0,75
4	0,44	0,40	0,43	0,46	0,43	0,51
5	0,56	0,54	0,58	0,67	0,69	0,67
6	0,31	0,27	0,30	0,39	0,50	0,27
7	0,72	0,69	0,69	0,36	0,47	0,56
8	0,48	0,49	0,45	0,32	0,45	0,33
9	0,21	0,28	0,26	0,09	0,21	0,07
10	0,35	0,55	0,59	0,62	0,58	0,61
11	0,57	0,42	0,45	0,64	0,43	0,39
12	0,56	0,47	0,46	0,57	0,56	0,50
13	0,63	0,57	0,42	0,58	0,72	0,51
14	0,37	0,47	0,35	0,33	0,41	0,41
15	0,29	0,28	0,23	0,23	0,37	0,05
16	0,53	0,61	0,61	0,59	0,57	0,64
17	0,47	0,71	0,62	0,42	0,73	0,48
18	0,66	0,66	0,64	0,70	0,65	0,65
19	0,67	0,59	0,59	0,63	0,72	0,67
20	0,42	0,36	0,46	0,59	0,38	0,48

Not: Koyu renk ile belirtilen maddeler her testte ortak olan maddelerdir.

Tablo 17'de verilenlere göre koyu renkte belirtilen 6 madde (2, 4, 5, 7, 9, 19) her testte ortak olan maddelerdir. Bu maddelerin güçlük dereceleri açısından her test için birbirine yakın değerler aldığı söylenebilir. Ortak maddelerden sadece 19. maddede test A diğer iki teste göre farklı güçlük derecesine sahiptir ve bu madde Test A'yı alan grup tarafından daha yüksek oranda doğru cevaplanmıştır. Farklı test formlarındaki aynı test maddeleri benzer güçlük derecesine sahip olsalar bile ayırt edicilik değerlerinin test gruplarına göre değiştiği görülmektedir. Ortak olmayan maddelerin güçlük ve ayırt edicilik değerleri incelendiğinde ise genel olarak madde güçlük dereceleri yakın olan maddelerin ayırt ediciliklerinin test formlarına göre değiştiği görülmektedir. Öğrencilerin en çok zorlandıkları maddelerin ise eğim kavramı ile ilişkili olan 6. ve 9. maddeler ile verilen bir problemde bilinmeyenleri kullanarak denklem oluşturma ve çözmeye bağlı olan 15.

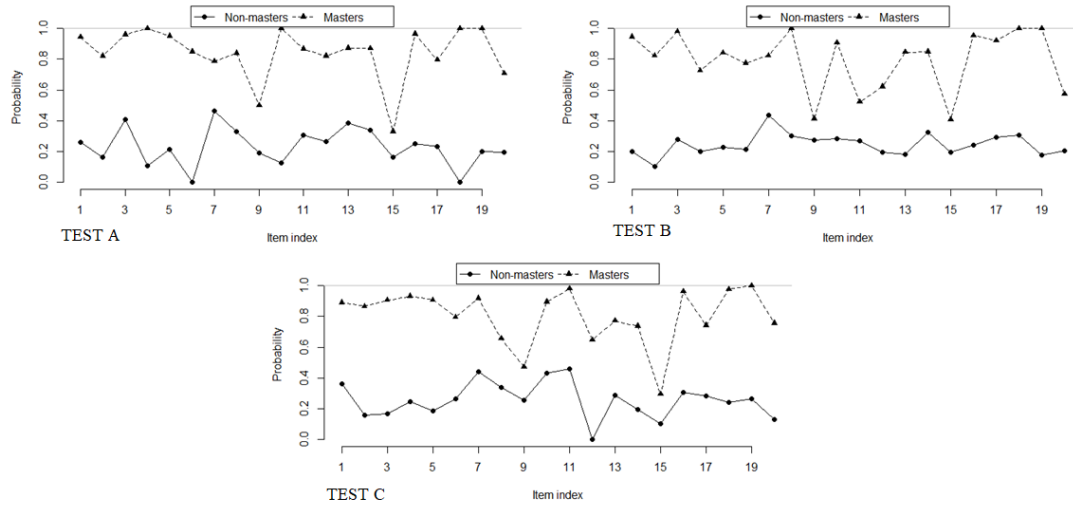
maddedir. Bu maddelerin senaryo deęişikliğinden ziyade maddelerde ölçölmek istenen ilgili kavram, kazanım ve becerilerle ilgili olarak özellikle matematik eğitimi çerçevesinde deęerlendirilmesi uygundur. Bunun yanında farklı senaryolarda aynı kazanımı ölçen, aynı yöntem ve stratejilere baęlı çözümleri olan maddelerde madde güçlük deęerlerinin deęiştii de gözlenmiştir. Örneğin madde 17 buna güzel bir örnektir. İlgili maddede niceliksel mantık yürütme, temsilleştirme ve gerekçelendirme nitelikleri eşitlik kavramı ile ilgili olarak üç farklı senaryoda ölçölmüştür. Senaryo gereęi Test A formunda bu nitelikler terazi-kütle ilişkisi kullanılarak yapılan eşitleme, Test B’de kazandıkları parayı paylaşan iki çocuğun paralarının eşitlenmesi ve Test C’de farklı kütlelerde verilen çuvalların deęişimi ile yapılan eşitleme söz konusudur. Burada çözüm için kullanılan yöntem aynı iken maddelerin doęru cevaplanma oranları farklıdır ve bu farklılığın temelinde yatan sebepler matematik eğitimi çerçevesinde deęerlendirilmelidir. Mesela böyle bir durumda Test B’de günlük hayatta dięer formlardaki senaryolardan daha çok kullanılan paraya dayalı bir senaryo verilmiş ve bu maddenin daha çok aday tarafından doęru cevaplandığı görölmüştür. Yine aynı durum için öğrencilerin sahip olduęu “kütle” kavramına bakılarak da durum analiz edilebilir. Belirtildięi gibi maddelerdeki güçlük ve ayırt edicilik deęerlerinin incelenmesinde maddelere has özellikler hem matematik eğitimi açısından hem de ölçme ve deęerlendirme açısından yorumlanmalıdır. Burada bu çalışma kapsamında uzaklaşmamak amacıyla maddeler açısından tek tek bir deęerlendirmeye girilmeyecek olup genel olarak maddelerin hazırlanma safhasında ölçme hedefi doęrultusunda gösterilen hassasiyetlere rağmen maddelerin cevaplanma oranlarında farklılıklar ortaya çıkabilmektedir. İşte bu sebeple ortaya çıkan bu farklılıklardan kaynaklı farklı test puanlarının veya beceri puanlarının eşitlenmesi önem kazanmaktadır.

Aynı test maddelerinin yukarıda belirtilen ve KTK’ya dayalı olarak yapılan analiz sonuçlarının yanında daha detaylı bilgiler sunan karma hiyerarşik yapılı GDINA modeli ile analizlerinde ilgili maddelerin ölçölmek istenen becerilerde yetkin olan ve olmayan öğrenci

grupları tarafından doğru cevaplanma durumunu gösteren grafikler aşağıda Şekil 15'te paylaşılmıştır.

Şekil 15

GDINA Modele Göre Yetkin ve Yetkin Olmayan Gruplarda Maddelerin Cevaplanma Olasılıkları



Şekil 15'te görüldüğü gibi her testte yetkin ve yetkin olmayan öğrenci gruplarının maddeleri cevaplama eğilimlerinin genel olarak benzer olduğu görülmektedir. Burada da ilgili becerilerde yetkin ve yetkin olmayan öğrencilerin özellikle 9. ve 15. maddelerde güçlükleri olduğu görülmektedir. Testlere göre bazı maddelerde de yetkin gruplar arasında doğru cevaplama olasılıklarında farklılıkların olduğu durumlar mevcuttur. Örneğin madde 11'de ölçülmesi hedeflenen becerilerde yetkin olan öğrenciler Test A'da % 87 ve Test C'de % 98 olasılığında bu maddeyi doğru cevaplayabilirken Test B'de bu olasılık % 52'dir. Klasik yöntemlerde bu detay hiç fark edilmezken bu çalışmada karma hiyerarşik yapıllı GDINA model analizleriyle saptanabilmiştir. Fakat testlerin geneline bakıldığında yetkin grubun maddeleri cevaplama olasılıklarının ortalaması Test A için 0,85, Test B için 0,80 ve Test C için 0,81'dir. Yetkin olmayan grubun ise maddeleri cevaplama olasılıklarının ortalaması Test A ve Test B için 0,23, Test C için 0,26'dır. Genel anlamda testler benzer olasılıklar sunmaktadır.

Genel olarak testler benzer analiz sonuçları verse de test maddelerinin özelinde bakıldığında ve özellikle karma hiyerarşik yapıları GDINA model analizleri çerçevesinden incelendiğinde maddelerde ölçülmek istenen becerilerde yetkin ve yetkin olmayan grupların eğilimine ve cevaplama oranlarına kadar bilgiler elde edilebilmektedir. Bu bilgiler ışığında farklı formlarda aynı hedeflerle hazırlanan maddelerde benzer sonuçlar alındığı gibi benzer olaylar sonuçlarının da alındığı görülmektedir. Bu farklılıklar da bireysel olarak hem test genel puanına hem de bireysel beceri yetkinlik düzeylerine ve doğal olarak bu yetkinlik düzeylerinden elde edilen beceri puanlarının test formlarına göre farklılık göstermesine sebep olmaktadır. Farklı formları alan öğrencilerin arasında yapılacak olan bir puan karşılaştırmasında ise bu farklılıkların öğrenciler arasında eşit olmayan koşullar oluşturmasının önüne geçilmesi amacıyla beceri puanlarının eşitlenmesi önemi kendini göstermektedir.

Dorans ve Holland (2000) literatürde eşitleme koşulu olarak aynı yapıyı ölçme, güvenilirliklerin eşit olması, simetriklik, eşitlik ve gruptan bağımsızlık vurgusunun öne çıkarıldığını belirtir. Buraya kadar verilen bilgiler bir araya getirilecek olunursa testler için madde hazırlığından itibaren paralel testler oluşturulmasına önem verilmiş, her testte aynı becerileri ölçen maddeler aynı sırada yer almış, madde yazımında aynı becerileri ölçen testler farklı testler için kullanıma göre hazırlanmış ancak mümkün olduğunca paralel maddeler olmasına dikkat edilmiş, maddelerin bağlı olduğu kazanımlar özenle seçilmiş ve bu açıdan da maddeler arası paralellik göz önünde bulundurulmuştur. KTK'ya göre yapılan analizler ve karma hiyerarşik yapıları GDINA'ya göre yapılan analizler birlikte değerlendirilirse, KTK açısından her testin ortalama güçlüğü'nün birbirine çok yakın olması, test ve madde güvenilirliklerinin testler arasındaki yakın değerleri, yine karma yapıları hiyerarşik GDINA model sonuçlarına göre test seviyesinde güvenilirliklerinin aynı olması, her bir nitelikte yetkinlik açısından sınıflandırılma doğruluğunun yüksek güvenilirlikte ve yakın değerlerde olması, KTK'ya göre madde analizlerinde farklı güçlük ve ayırteçicilik indeksler olmakla birlikte testlerin genelinin bu açıdan da benzerlik sergilediği, karma

hijerarşik yapılı GDINA model sonuçlarına göre ise testlerdeki maddelere yetkin ve yetkin olmayan grupların her bir testte benzer eğilimler gösterdiği görülmektedir. Genel test puanları açısından test formlarının ortalamalarının birbirine çok yakın olduğu, meydanlarının aynı olduğu ve mod değerlerinin yine çok yakın olduğu görülmektedir. Test A ve Test B'de varyanslarının homojen olmakla birlikte Test C'de bu durumun anlamlı olarak farklı olduğu, buna bağlı olarak yapılan non-parametrik testlerde bu gruplar arasında medyan farkının bulunmadığı, belirlenmiştir. Bütün bu bilgiler bir araya getirildiğinde araştırma karma hijerarşik yapılı GDINA model ile analiz edilen testlerden elde edilen beceri puanlarının eşitlenmesine dair bir inceleme çalışmasının yapılması ve sonuçların değerlendirilmesi uygun görülmüştür.

Eşitleme çalışmalarına geçmeden önce karma hijerarşik yapılı GDINA model uygulamasında elde edilen bireysel beceri yetkinlik düzeylerinin 25 puan üzerinden beceri puanlarına dönüştürüldüğünde örnekleme ait beceri puanlarının betimsel istatistikleri aşağıda Tablo 18'de paylaşılmıştır. Daha önce de belirtildiği gibi bu çalışmada kullanılan test formlarında altı bilişsel beceri üzerinden tanılama yapılmıştır. Karma hijerarşik yapılı GDINA modeli analizleri sonucunda bu altı nitelik için bireysel yetkinlik düzeyleri elde edilmiş ve her biri için beceri puanları belirlenmiştir. Fakat bu çalışma çerçevesinde, altı beceri puanı kullanarak çalışmayı altı niteliğin her biri için tekrarlı analizlerle karmaşıktırmamak adına sadece birinci faktöre bağlı üç niteliğe ait beceri puanları üzerinden eşitleme incelemeleri yapılmasının çalışma hedefleri çerçevesinde yeterli olacağı düşünülmüştür.

Tablo 18

Beceri Puanlarının Testlere Göre Betimsel İstatistikleri

<i>Test</i>	<i>GA</i>			<i>FD</i>			<i>NMY</i>		
	<i>M</i>	<i>ss</i>	<i>medyan</i>	<i>M</i>	<i>ss</i>	<i>medyan</i>	<i>M</i>	<i>ss</i>	<i>medyan</i>
TEST A	12,05	10,69	11	9,41	9,22	6	19,32	7,69	23
TEST B	11,73	10,09	11	9,31	9,22	6	14,70	9,80	18
TEST C	12,35	9,67	12	4,85	6,49	2	15,12	9,30	18

Beceri puanlarına bakıldığında çalışma grubunun her bir becerideki genel durumu yansımaktadır. 25 puan üzerinden belirlenen beceri puanlarında öğrencilerin genel olarak niceliksel mantık yürütme becerisinde diğer becerilere göre çok daha iyi durumda olduğu görülmektedir. Bunun yanında öğrencilerin fonksiyonel düşünme becerilerinde yüksek düzeyde zorlandıkları anlaşılmaktadır. Testler açısından bakıldığında ise Test C'de diğer test formlarına göre fonksiyonel düşünme becerisinde öğrencilerin daha çok güçlük çektikleri ve düşük puana sahip oldukları anlaşılmaktadır. Gerek KTK'ya göre yapılan analizlerde ve gerek karma yapılı hiyerarşik GDINA modeli analizlerinde testlerin genelinde sergilenen özellikler yakın olmalarına rağmen daha detay özellik olarak beceriler açısından karşılaştırıldığında betimsel özelliklerde test formları arasında farklılıkların ortaya çıktığı gözlenmektedir. İşte eğitimde tanılama amaçlı kullanılan testlerde nitelikler açısından elde edilen sonuçların karşılaştırma amaçlı kullanılacağı durumlarda farklı test formları ile yapılan ölçmelerden kaynaklanan puan farklılıklarının eşitlenmesi ile daha adil bir karşılaştırma, daha geçerli ve güvenilir değerlendirmeler yapılabilmesi adına önem kazanmaktadır.

Bölüm 4

Bulgular, Yorum ve Tartışma

Bu bölümde ilk olarak denk gruplar deseni, denk olmayan gruplarda ortak madde deseni ve denk olmayan gruplarda ortak değişken deseni çerçevesinde genelleştirilmiş aritmetik beceri puanlarının AxB, AxC ve BxC testlerinde eşitlenebilme durumu incelenmiştir. Yine aynı çerçeve kapsamında devam edilerek fonksiyonel düşünme ve niceliksel mantık yürütme beceri puanları için eşitleme durumu incelenmiştir. Her bir desen sonunda ilgili becerinin AxB, AxC ve BxC testlerinde eşitlenebilme durumu tablolaştırılarak sunulmuş ve son aşamada bulgular bir araya getirilerek genel bir yorum yapılabilmesi sağlanmıştır. Yapılan incelemeler sonucunda karma hiyerarşik yapılı GDINA modelden elde edilen yetkinlik düzeyi değerlerinin beceri puanlarına çevrilerek eşitleme yapılabilme durumu ile ilgili bu çalışmanın başında belirtilen araştırma sorularının yanıtları aranmış ve bulgulardan yola çıkılarak yorumlar paylaşılmıştır.

Birinci Araştırma Sorusu Kapsamındaki Bulgular

Genelleştirilmiş Aritmetik (GA) Becerisinin Denk Grup (DG) Desenine Göre Eşitleme İncelemeleri

DG deseninde GA becerileri için her üç testte çaprazlama olarak eşitleme uygulamaları yapılmıştır. Bu desen çerçevesinde ortalama, doğrusal, eşit yüzdelikli ve kernel eşitleme uygulamaları yapılmıştır. Aşağıda AxB testine ait eşitlemeler ayrıntılarıyla paylaşılmış olup, AxC ve BxC eşitlemelerine ait genel bulgular desen ile ilgili genel tabloda sunulmuştur.

GA Becerisinin Test A ve Test B için DG Deseninde Eşitleme Uygulamaları

DG deseni kapsamında GA becerisinin AxB testlerinde eşitlenebilme durumları incelenmiştir. Bu desen çerçevesinde ortalama, doğrusal, eşit yüzdelikli ve kernel eşitleme

uygulamaları yapılmış, farklı yöntemler denenmiş ve bu yöntemlere dair eşitlenmiş puanlar, bu puanların ham puanlardan farkı ve eşitleme hataları paylaşılmıştır.

i. Ortalama Eşitleme (AxB / DG / Ortalama)

GA becerisine ait ham puanlar ve AxB test eşitlemesinde ortalama eşitleme ile elde edilen eşitlenen puanlar Tablo 19'da verilmiştir.

Tablo 19

DG Deseni GA Becerisi Ortalama Eşitleme Puanı

<i>Ham puan</i>	<i>Eşitlenmiş Puan</i>	<i>Ham puan</i>	<i>Eşitlenmiş Puan</i>
0	-0,32	13	12,68
1	0,68	14	13,68
2	1,68	15	14,68
3	2,68	16	15,68
4	3,68	17	16,68
5	4,68	18	17,68
6	5,68	19	18,68
7	6,68	20	19,68
8	7,68	21	20,68
9	8,68	22	21,68
10	9,68	23	22,68
11	10,68	24	23,68
12	11,68	25	24,68

Tablo 19'a göre yapılan ortalama eşitlemede eşitlenen puanların ham puanlardan 0.32 puan daha düşük olduğu görülmektedir. Eşitlenen puanların üst sınır ham puanı (25 puan) aşmadığı fakat alt sınır ham puandan (0 puan) daha düşük olduğu görülmektedir.

ii. Doğrusal Eşitleme (AxB / DG / Doğrusal)

GA becerisine ait ham puanlar, AxB test eşitlemesinde doğrusal eşitleme ile elde edilen eşitlenmiş puanlar ve eşitlemeye ait standart hatalar Tablo 20'de paylaşılmıştır.

Tablo 20*DG Deseni GA Becerisi Doğrusal Eşitleme Puanı*

<i>Ham puan</i>	<i>Eşitlenmiş Puan</i>	<i>SE</i>	<i>Ham puan</i>	<i>Eşitlenmiş Puan</i>	<i>SE</i>
0	0,357	0,234	13	12,628	0,236
1	1,301	0,233	14	13,572	0,238
2	2,245	0,231	15	14,516	0,240
3	3,189	0,231	16	15,460	0,243
4	4,133	0,230	17	16,404	0,246
5	5,077	0,230	18	17,348	0,249
6	6,021	0,229	19	18,292	0,252
7	6,965	0,230	20	19,236	0,255
8	7,909	0,230	21	20,180	0,259
9	8,853	0,231	22	21,124	0,263
10	9,797	0,232	23	22,068	0,268
11	10,741	0,233	24	23,012	0,272
12	11,684	0,234	25	23,956	0,277

Tablo 20'da görüldüğü üzere eşitlenen puanlar ham puana ait sınırlar içerisinde olup, eşitlenen puanlar ilk 6 puana kadar ham puan değerlerinden yüksek ve sonrasında ham puanlardan daha düşük olduğu görülmektedir. Bu farklar 0,36 ile -1,04 değerleri arasında değişmektedir. Eşitlenmiş puanlara ait standart hatalar ise 0,23 ile 0,28 arasında değişmektedir.

iii. Eşit Yüzdellikli Eşitleme (AxB / DG / EYE)

GA becerisine ait ham puanlar, AxB test eşitlemesinde eşit yüzdellikli eşitleme ile elde edilen eşitlenmiş puanlar ve eşitlemeye ait standart hatalar Tablo 21'de paylaşılmıştır.

Tablo 21*DG Deseni GA Becerisi EYE Puanı*

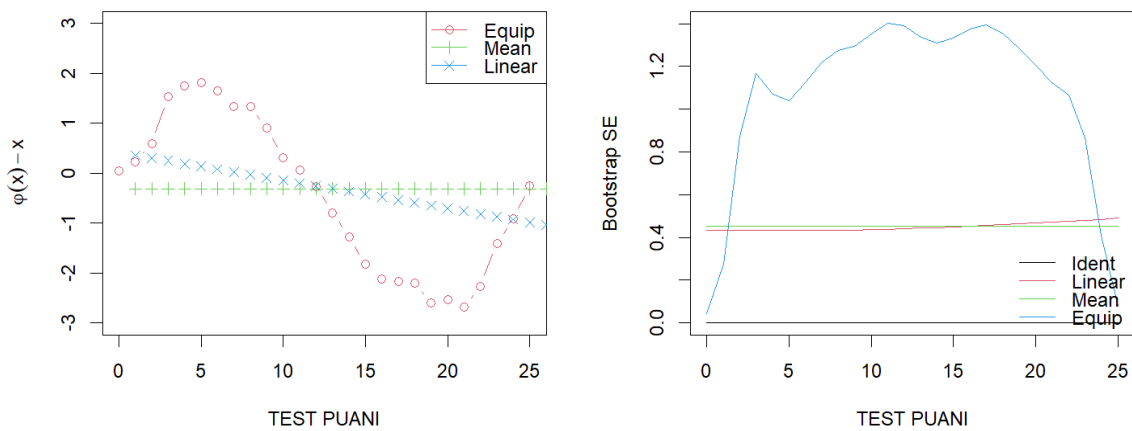
<i>Ham puan</i>	<i>Eşitlenmiş Puan</i>	<i>SE</i>	<i>Ham puan</i>	<i>Eşitlenmiş Puan</i>	<i>SE</i>
0	0,05	0,06	13	11,74	1,39
1	1,23	0,24	14	12,20	1,15
2	2,59	0,94	15	12,73	1,15
3	4,53	1,07	16	13,17	1,48

Tablo 21 (Devamı)

4	5,74	1,20	17	13,87	2,07
5	6,81	1,03	18	14,83	1,03
6	7,64	1,29	19	15,79	1,02
7	8,33	1,30	20	16,39	1,57
8	9,34	2,08	21	17,46	0,91
9	9,90	0,99	22	18,32	0,90
10	10,31	0,99	23	19,73	0,52
11	11,06	2,32	24	21,59	0,46
12	0,05	1,39	25	23,08	0,10

Tablo 21'de görüldüğü üzere eşitlenen puanlar ham puana ait sınırlar içerisinde olup, eşitlenen puanlar ilk 11 puana kadar ham puan değerlerinden yüksek ve sonrasında ham puanlardan daha düşük olduğu görülmektedir. Bu farklar 1,74 ile -2,61 değerleri arasında değişmektedir. Ham puanlara göre 16 ile 23 puan aralığında eşitlenmiş puan ile ham puan farkının yükseldiği görülmektedir. Eşitlenmiş puanlara ait standart hatalar ise 0,1 ile 2,08 arasında değişmektedir. Eşitlenmiş puan aralığına göre uçlarda daha düşük hata puanları gözlenirken ortalarda hatalar uçlara göre daha yüksek durumdadır.

Aşağıda ortalama, doğrusal ve eşit yüzdelikli eşitlemeye ait eşitlenmiş puanların ham puanlardan farkı ve eşitlenen puanların bootstrap standart hataları ile ilgili görseller Şekil 16 sunulmuştur.

Şekil 16*DG Deseni GA Becerisi Eşitlenmiş Puan Fark ve Bootstrap Standart Hataları*

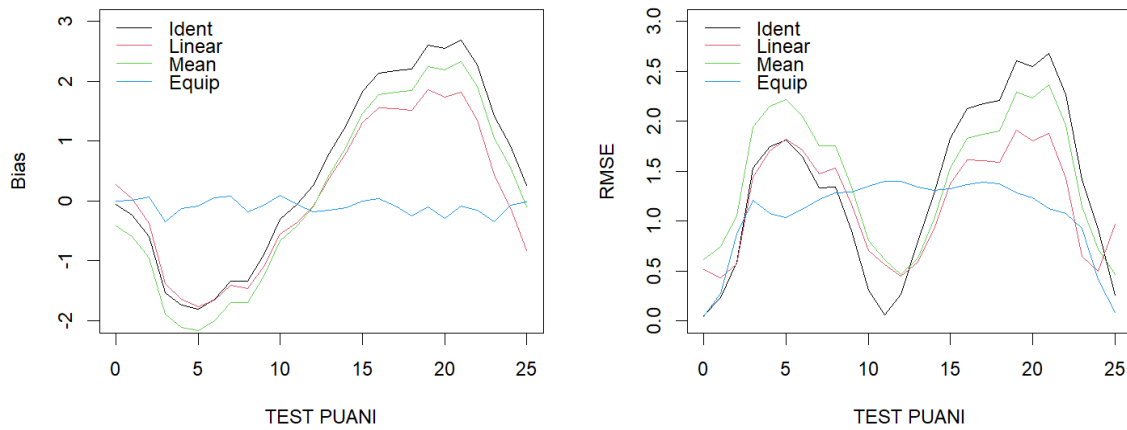
Şekil 16'da verilenlere göre eşitlenmiş puanların ham puanlardan farkının eşit yüzdelikli eşitlemede daha yüksek olduğu, uç noktalarda ve 11-13 puan aralığında

birbirine yakın değerler aldığı söylenebilir. Bootstrap standart hatalar incelendiğinde ise ortalama ve doğrusal eşitleme hatalarının 0,4 ile 0,5 arasında yakın değerler aldıkları, eşit yüzdellikli eşitlemenin ise 1,4 civarına kadar değişen değerler olduğu görülmektedir.

Eşitlenmiş puanlara ait yanlılık ve RMSE değerleri Şekil 17’de paylaşılmıştır.

Şekil 17

DG Deseni GA Becerisi Eşitlenmiş Puan Yanlılık ve RMSE Değerleri



Şekil 17’de yanlılık değerleri incelendiğinde eşit yüzdellikli eşitlemede bu değerlerin 0 civarında değiştiği görülmektedir. Diğer eşitlemelerde ise bu değerlerin 10 puana kadar -2 ile 0 aralığında değiştiği, 10 puan sonrasında ise 0 ile 2,5 civarı değiştiği görülmektedir. Ortalama eşitlemenin yanlılık değeri doğrusal eşitlemeye göre daha yüksektir. RMSE değerleri incelendiğinde ise eşit yüzdellikli eşitlemenin diğer eşitlemelere göre genel olarak düşük değerlerde olduğu, 10 ile 15 puan aralığında ortalama ve doğrusal eşitlemelerden daha yüksek değerdedir. Diğer puan aralıklarında ortalama eşitleme değerlerinin ise en yüksek RMSE değerlerine sahip olduğu görülmektedir. Genel olarak bakıldığında ise yanlılık değerlerinin -2 ile 2,5 ve RMSE değerleri 0 ile 2,6 arasında değişmektedir.

Tablo 22 ortalama, doğrusal ve eşit yüzdellikli eşitlenmiş puanlara ve birim puanlara ait bootstrap hata, yanlılık ve RMSE değerleri ile bunların ağırlıklandırılmış değerlerine ait ortalamaları sunmaktadır.

Tablo 22*DG Deseni GA Becerisi Eşitlenmiş Puan Bootstrap Hata, Yanlılık ve RMSE Değerleri*

	<i>SE</i>	<i>SE_w</i>	<i>BIAS</i>	<i>BIAS_w</i>	<i>RMSE</i>	<i>RMSE_w</i>
birim	0,00	0,00	1,58	1,14	1,58	1,14
doğrusal	0,45	0,46	1,22	0,91	1,30	1,02
ortalama	0,45	0,45	1,51	1,10	1,58	1,19
eye	1,14	0,74	0,15	0,12	1,14	0,74

Not: eye: eşit yüzdellikli eşitleme

Tablo 22'ye göre doğrusal ve ortalama eşitlemede eşitlenen puanların standart hataları eşit, 0,45 olmakla birlikte eşit yüzdellikli eşitlemede bu değer 1,14'tür. Yanlılık değerleri 0,15 ile 1,58 arasında değişmekte, RMSE değerleri ise 1,14 ile 1,58 değerleri arasında değişmektedir. Tabloya genel olarak bakıldığında değerler düşük olmakla birlikte standart hata değeri eşit yüzdellikli eşitlemede diğer eşitlemelerden biraz daha yüksek olsa da yanlılık ve RMSE değerlerinin diğer eşitlemelerden daha düşük olduğu görülmektedir.

iv. Kernel Eşitleme (AxB / DG / Kernel)

GA becerisine ait ham puanlar, AxB test eşitlemesinde logistic kernel eşit yüzdellikli eşitleme ile doğrusal eşitlemeden elde edilen eşitlenmiş puanlar ve eşitlemeye ait standart hatalar Tablo 23'te paylaşılmıştır.

Tablo 23*DG Deseni GA Becerisi Logistic Kernel EYE ve Doğrusal Eşitleme Puan ve Hataları*

<i>Ham puan</i>	<i>EYE</i>		<i>Doğrusal</i>	
	<i>Eşitlenmiş Puan</i>	<i>SE</i>	<i>Eşitlenmiş Puan</i>	<i>SE</i>
0	0,04	0,046	0,36	0,484
1	1,23	0,266	1,30	0,482
2	2,71	0,833	2,24	0,481
3	4,22	1,174	3,19	0,480
4	5,62	1,496	4,13	0,479
5	6,78	1,351	5,08	0,479
6	7,75	1,343	6,02	0,479
7	8,59	1,459	6,96	0,479
8	9,30	1,355	7,91	0,479

Tablo 23 (Devamı)

9	9,98	1,160	8,85	0,480
10	10,58	1,435	9,80	0,481
11	11,11	1,188	10,74	0,482
12	11,65	1,399	11,68	0,484
13	12,13	1,219	12,63	0,486
14	12,68	1,400	13,57	0,488
15	13,20	1,304	14,52	0,490
16	13,85	1,285	15,46	0,492
17	14,54	1,558	16,40	0,495
18	15,33	1,478	17,35	0,498
19	16,27	1,414	18,29	0,502
20	17,37	1,451	19,24	0,505
21	18,64	1,274	20,18	0,509
22	19,99	0,872	21,12	0,513
23	21,47	0,747	22,07	0,517
24	23,05	0,322	23,01	0,522
25	24,76	0,084	23,96	0,526

Tablo 23'te görüldüğü üzere logistic kernel eşit yüzdelikli ve doğrusal eşitleme ile eşitlenen puanlar ham puana ait sınırlar içerisindedir. Logistic kernel eşit yüzdelikli eşitleme ile eşitlenen puanlar ilk 11 puana kadar ham puan değerlerinden yüksek ve sonrasında ham puanlardan daha düşük olduğu görülmektedir. Ham puanlara göre 16 ile 20 puan aralığında eşitlenmiş puan ile ham puan farkının diğer puan aralıklarına göre daha yüksek olduğu görülmektedir. Eşitlenmiş puanlara ait standart hatalar ise 0,05 ile 1,56 arasında değişmektedir. Logistic kernel doğrusal eşitleme ile eşitlenen puanlar ise ilk 6 puana kadar ham puan değerlerinden yüksek ve sonrasında ise ham puan değerlerinden düşüktür. Farklar 0-6 ham puan aralığında azalırken, sonrasında artmaktadır. Genel olarak farkların düşük olduğu görülmektedir. Eşitlenmiş puanlara ait standart hatalar ise 0,48 ile 1,53 arasında değişmektedir, bu değişim düşük puandan yüksek puana artarak ilerlemiştir.

Ayrıca uniform kernel eşit yüzdelikli ve doğrusal eşitlemeler de incelenmiş olup eşitlenmiş puanlar ve bu puanlara ait standart hatalar Tablo 24'te paylaşılmıştır.

Tablo 24*DG Deseni GA Becerisi Uniform Kernel EYE ve Doğrusal Eşitleme Puan ve Hataları*

<i>Ham puan</i>	<i>EYE</i>		<i>Doğrusal</i>	
	<i>Eşitlenmiş Puan</i>	<i>SE</i>	<i>Eşitlenmiş Puan</i>	<i>SE</i>
0	0,04	0,045	0,36	0,484
1	1,26	0,271	1,30	0,482
2	2,71	0,919	2,24	0,481
3	4,26	1,199	3,19	0,480
4	5,62	1,383	4,13	0,479
5	6,77	1,370	5,08	0,479
6	7,74	1,336	6,02	0,479
7	8,59	1,306	6,96	0,479
8	9,33	1,310	7,91	0,479
9	9,98	1,293	8,85	0,480
10	10,57	1,295	9,80	0,481
11	11,12	1,294	10,74	0,482
12	11,63	1,312	11,68	0,484
13	12,14	1,311	12,63	0,486
14	12,66	1,343	13,57	0,488
15	13,22	1,342	14,52	0,490
16	13,83	1,377	15,46	0,492
17	14,52	1,404	16,40	0,495
18	15,33	1,396	17,35	0,498
19	16,28	1,396	18,29	0,502
20	17,38	1,359	19,24	0,505
21	18,62	1,160	20,18	0,509
22	19,97	0,982	21,12	0,513
23	21,46	0,769	22,07	0,517
24	23,01	0,376	23,01	0,522
25	24,76	0,078	23,96	0,526

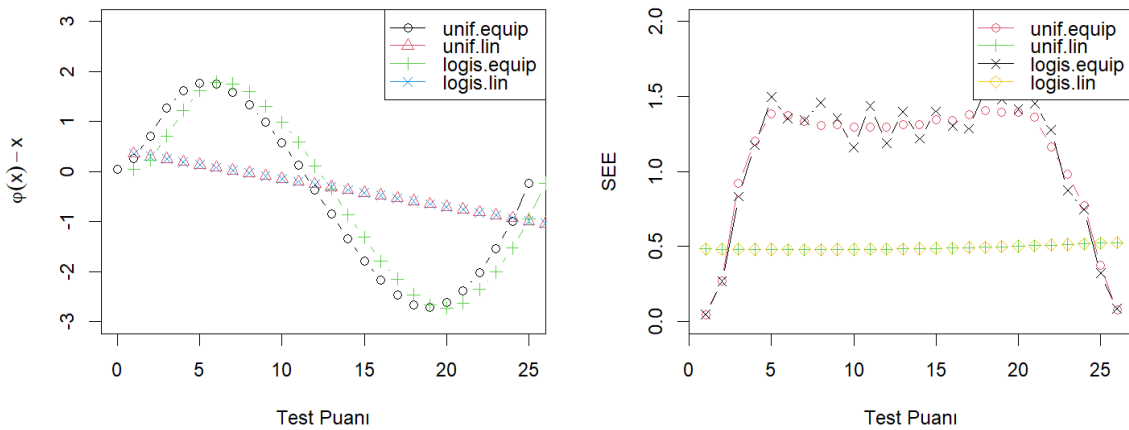
Tablo 24'de uniform kernel eşit yüzdelikli ve doğrusal eşitleme ile eşitlenen puanlar ham puana ait sınırlar içerisinde. Uniform kernel eşit yüzdelikli eşitleme ile eşitlenen puanlar ilk 11 puana kadar ham puan değerlerinden yüksek ve sonrasında ham puanlardan daha düşüktür. Ham puanlara göre 16 ile 22 puan aralığındaki eşitlenmiş puan ile ham puan farkının diğer puan aralıklarına göre daha yüksek olduğu görülmektedir. Eşitlenmiş puanlara ait standart hatalar ise 0,05 ile 1,40 arasında değişmektedir. Standart hataların bazı aralıklarda artış gösterdiği bazı aralıklarda ise düşük değerler aldığı fakat belli bir örüntüyü takip etmediği gözlenmiştir. Uniform kernel

doğrusal eşitleme ile eşitlenen puanlar ise ilk 6 puana kadar ham puan değerlerinden yüksek ve sonrasında ise ham puan değerlerinden düşüktür. Farklar 0-6 ham puan aralığında azalırken, sonrasında artmaktadır. Genel olarak farkların düşük olduğu görülmektedir. Eşitlenmiş puanlara ait standart hatalar ise 0,48 ile 0,53 arasında değişmektedir, bu değişim ilk 8 puana kadar düşmüş, sonrasında ise artarak ilerlemiştir. Fakat genel olarak standart hata değerleri birbirine çok yakındır.

Şekil 18'de yukarıda verilen logistic ve uniform kernel eşitlemelerde eşit yüzdelikli ve doğrusal eşitlemelerde eşitlenmiş puan ile ham puan farklarının ve standart hatalarının yer aldığı grafikler birlikte verilmiştir.

Şekil 18

DG Deseni GA Becerisi Kernel Eşitleme Puan Farkları ve Standart Hataları



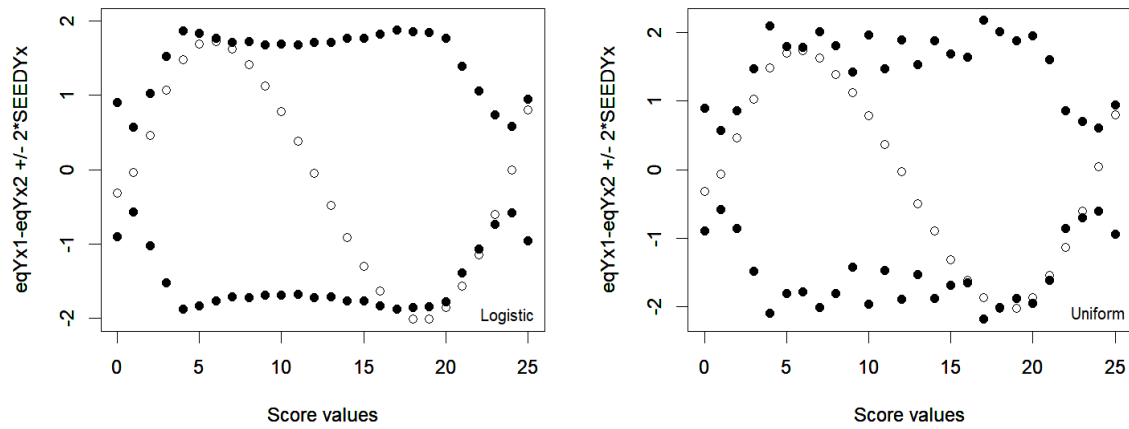
Şekil 18'de logistic ve uniform doğrusal eşitlemelere ait eşitlenmiş puan farklarının çakıştığı eşit yüzdelikli eşitlemelerin ise çok yakın değerler aldığı görülmektedir. Logistic kernelde eşit yüzdelikli eşitleme ile eşitlenmiş puanların farkı 0,04 ile -2,73 ve doğrusal eşitleme ile eşitlenmiş puanların ham puanlardan farkı ise -1,04 ile 0,36 değerleri arasında değişmektedir. Uniform kernel eşit yüzdelikli eşitleme ile eşitlenmiş puanların ham puanlardan farkı 0,04 ile -2,72 değerleri arasında ve uniform kernel doğrusal eşitleme ile eşitlenmiş puanların ham puanlardan farkı -1,04 ile 0,02 değerleri arasında değişmektedir. Eşitleme hatalarına bakıldığında ise uniform kernel doğrusal eşitleme ile logistic kernel

doğrusal eşitlemeye ait değerlerin çakıştığı, eşit yüzdellikli eşitlemeye ait değerlerin ise yakın olduğu görülmektedir.

Logistic kernel ve uniform kernel eşitlemelerde eşit yüzdellikli eşitleme ile doğrusal eşitleme arasındaki eşitleme puanları farkına ait standart hatalar (*standard error of equating differences, SEED*) değerleri Şekil 19'da paylaşılmıştır. Bu grafiklerde hata değerlerinin hata bantları arasında yer alması beklenir.

Şekil 19

DG Deseni GA Becerisi Logistic ve Uniform Kernel SEED Değerleri



Şekil 19 incelendiğinde logistic ve uniform kernel eşitlemelerde eşit yüzdellikli eşitleme puanları ile doğrusal eşitleme puanları arasındaki farkların standart hata değerlerinin hata bantları içerisinde olduğu görülmektedir.

Tablo 25 ise logistic ve uniform kernel eşitlemelere ait yüzde göreceli hata (percent relative error, PRE) değerlerini sunmaktadır.

Tablo 25

DG Deseni GA Becerisi Uniform ve Logistic Kernel Yüzde Göreceli Hata Değerleri

<i>Ham Puan</i>	<i>Uniform</i>	<i>Logistic</i>
1	0,07	0,08
2	0,28	0,34
3	0,44	0,53
4	0,57	0,69

Tablo 25 (Devamı)

5	0,67	0,82
6	0,76	0,91
7	0,81	0,98
8	0,85	1,02
9	0,86	1,03
10	0,84	1,01

Tablo 25'te görüldüğü üzere ilk 10 moment değeri için elde edilen yüzde göreceli değerleri (*PRE*) uniform kernel eşit yüzdelliği eşitleme için 0,07 ile 0,84 ve logistic kernel eşit yüzdelliği eşitleme için 0,08 ile 1,01 arasında değerler almaktadır. İyi bir eşitleme için *PRE* değerlerinin düşük değerde olması beklenir, tabloda bu değerlerin genel olarak düşük değerler olduğu görülmektedir.

Denk grup desenlerinde geliştirilmiş aritmetik beceri puanlarının *AxB* testleri için eşitlenme durumu ortalama, doğrusal, eşit yüzdelliği eşitleme ve kernel eşitleme ile incelendiğinde eşitleme uygulamalarının kabul edilebilir sonuçlar verdiği görülmektedir. *AxB* test eşitleme uygulamaları benzer biçimde *AxC* ve *BxC* test eşitlemeleri için de uygulanmıştır. Bu uygulamalardan elde edilen bulgular Tablo 26'da sunulmuştur.

Tablo 26

DG Deseninde *GA* Becerisinin Testlere Göre Eşitlenebilme Durumları

<i>Desen</i>	<i>Eşitleme Türü</i>	<i>Eşitlenen Testler</i>		
		<i>AxB</i>	<i>AxC</i>	<i>BxC</i>
<i>Denk Gruplar</i>	Ortalama	√	√	√
	Doğrusal	√	√	√
	Eşit Yüzdelliği	√	√	√
	Kernel	√	√	√

Tablo 26'da paylaşılan bulgulara göre denk gruplar deseninde geliştirilmiş aritmetik becerisi puanlarında *AxB*, *AxC* ve *BxC* eşitleme uygulamalarında ortalama, doğrusal, eşit yüzdelliği ve kernel eşitlemelerin uygun sonuçlar verdiği belirtilmektedir.

GA Becerisinin Denk Olmayan Gruplarda Ortak Madde (DOG-OM) Desenine Göre Eşitleme İncelemeleri

DOG-OM deseninde GA becerileri için eşitleme incelemelerine devam edilmiştir. Bu desen çerçevesinde doğrusal, eşit yüzdelikli ve kernel eşitleme uygulamaları yapılmıştır. Aşağıda AxB testine ait eşitlemeler verilmiş olup, AxC ve BxC eşitlemelerine ait bulgular desen ile ilgili genel tabloda paylaşılmıştır.

GA Becerisinin Test A ve Test B için DOG-OM Deseninde Eşitleme Uygulamaları

GA becerisinin DOG-OM deseninde AxB testlerinde eşitleme incelenmeleri yapılmıştır. Desende uygulanan doğrusal, eşit yüzdelikli ve kernel eşitleme uygulamaları çerçevesinde farklı yöntemler kullanılarak ve bu yöntemlerden elde edilen eşitlenmiş puanlar, bu puanların ham puanlardan farkı, eşitlemeye ait standart hatalar, yanlılık, RMSE ve eşitlemenin uygunluğu ile ilgili bilgi sunacak diğer bilgiler paylaşılmıştır.

i. Doğrusal Eşitleme (AxB / DOG-OM/ Doğrusal)

DOG-OM deseninde doğrusal eşitleme kapsamında GA becerisine ait puanlar Tucker, Chain, Levine gözlenen puan, Levine gerçek puan ve Braun-Holland yöntemleri ile eşitlemeler yapılmıştır. Elde edilen eşitlenmiş puanlar bir arada Tablo 26'da paylaşılmıştır.

Tablo 27

DOG-OM Deseni GA Becerisi Doğrusal Eşitleme Puanları

<i>Ham Puan</i>	<i>Tucker</i>	<i>Chain</i>	<i>Levine OS</i>	<i>Levine TS</i>	<i>Braun/Holland</i>
0	0,56	1,38	5,82	3,32	0,57
1	1,50	2,27	6,42	4,12	1,49
2	2,44	3,17	7,01	4,92	2,41
3	3,37	4,06	7,60	5,72	3,33
4	4,31	4,96	8,20	6,51	4,25
5	5,25	5,86	8,79	7,31	5,17
6	6,19	6,75	9,38	8,11	6,09
7	7,12	7,65	9,97	8,91	7,01

Tablo 27 (Devamı)

8	8,06	8,54	10,57	9,71	7,93
9	9,00	9,44	11,16	10,50	8,85
10	9,94	10,34	11,75	11,30	9,77
11	10,87	11,23	12,35	12,10	10,69
12	11,81	12,13	12,94	12,90	11,61
13	12,75	13,02	13,53	13,70	12,53
14	13,69	13,92	14,13	14,49	13,45
15	14,62	14,82	14,72	15,29	14,37
16	15,56	15,71	15,31	16,09	15,29
17	16,50	16,61	15,91	16,89	16,21
18	17,44	17,50	16,50	17,69	17,13
19	18,37	18,40	17,09	18,48	18,05
20	19,31	19,30	17,69	19,28	18,97
21	20,25	20,19	18,28	20,08	19,89
22	21,19	21,09	18,87	20,88	20,81
23	22,12	21,98	19,47	21,68	21,73
24	23,06	22,88	20,06	22,47	22,65
25	24,00	23,78	20,65	23,27	23,57

Tablo 27'da eşitlenen puanlar incelendiğinde *Tucker*, *Chain* ve *Braun/Holland* yöntemlerinden elde edilen eşitleme puanlarının ham puanlara yakın değerlerde eşitlendiği, ancak *Levine gözlenen puan* ve *Levine gerçek puan* yöntemleri ile yapılan eşitlemelerde eşitlenen puanlar uç noktalarda ham puanlara göre yüksek farka sahip olduğu görülmektedir. *Levine gözlenen puan* 5,82 ile *Levine gerçek puan* 3,32 ile başlangıç yaparak bu açık fark ham puanlarda 11 puana kadar devam etmiş, 12-16 puana kadar azalmış ve 17 puandan itibaren fark yeniden açılmıştır. Burada gözlenen durum uç noktalardaki eşitlenen puanların kullanıma uygun olmadığıdır. Ancak eşitleme için uygun görünen diğer yöntemlerden elde edilen eşitlenen puanların hata değerleri açısından incelemeleri de yapılmalıdır.

ii. Eşit Yüzdellikli Eşitleme (AxB / DOG-OM / Eşit Yüzdellikli)

DOG-OM deseninde GA becerisinin eşit yüzdellikli eşitleme çalışmalarında frekans tahmini (*frequency estimation, FE*) ve zincir eşitleme (*chain equating, CE*) yöntemleri kullanılmıştır. AxB test eşitlemesinde bu yöntemlerden elde edilen eşitlenmiş puanlar

Tablo 28'te ve bu puanların ham puanlardan farklarının görselleştirildiği ve FE ile CE yöntemlerinin karşılaştırılma imkanının sunulduğu Şekil 20 aşağıda paylaşılmıştır.

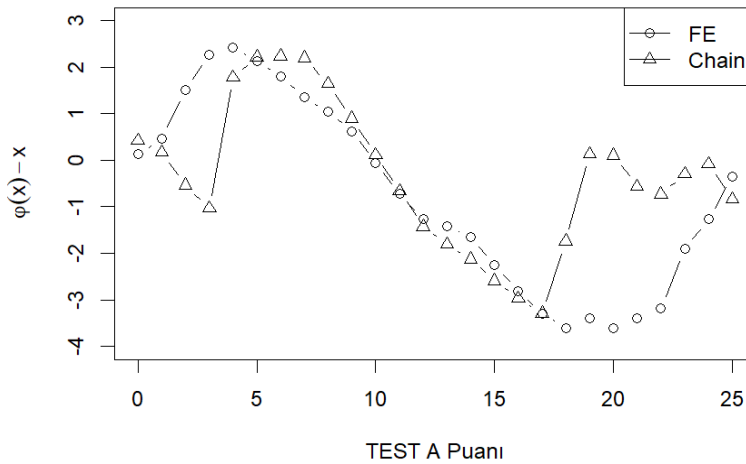
Tablo 28

DOG-OM Deseni GA Becerisi EYE Puanları

<i>Ham Puanlar</i>	<i>FE</i>	<i>CE</i>	<i>Ham Puanlar</i>	<i>FE</i>	<i>CE</i>
0	0,12	0,42	13	11,58	11,19
1	1,45	1,17	14	12,34	11,87
2	3,50	1,46	15	12,75	12,41
3	5,27	1,97	16	13,19	13,03
4	6,43	5,78	17	13,71	13,71
5	7,13	7,22	18	14,40	16,25
6	7,79	8,23	19	15,60	19,13
7	8,36	9,19	20	16,39	20,11
8	9,04	9,64	21	17,60	20,43
9	9,63	9,89	22	18,82	21,27
10	9,94	10,11	23	21,09	22,70
11	10,28	10,33	24	22,73	23,91
12	10,73	10,57	25	24,66	24,16

Şekil 20

DOG-OM Deseni GA Becerisi FE ve Chain Eşitleme Puan Farkları



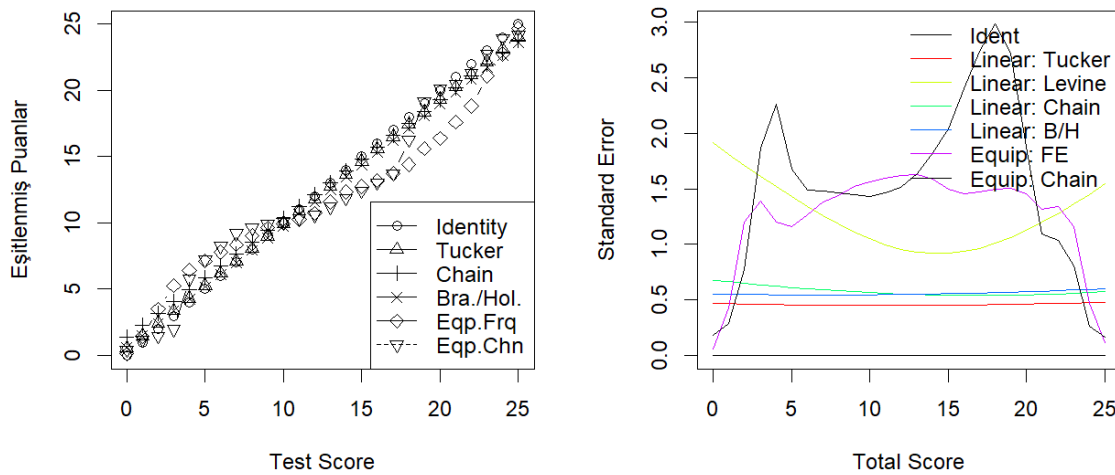
Tablo 28'de görüldüğü üzere FE ve CE yöntemleriyle eşitlenen puanlar ham puana ait sınırlar içerisinde olup, FE'de eşitlenen puanlar ilk 10 puana kadar ham puan değerlerinden yüksek ve sonrasında ham puanlardan daha düşük olduğu görülmektedir. CE'de eşitlenen puanların ham puanlara göre durumunda herhangi bir örüntü

görülmemektedir. Bazı puan aralıklarında ham puandan yüksek, bazı puan aralıklarında ham puanlardan düşüktür. FE eşitlemede bu farklar -3,61 ile 2,43 değerleri arasında ve CE eşitlemede -3,29 ile 2,23 arasında değişmektedir. FE ve CE eşitlenmiş puanların uç puanlara doğru farklılaştığı, orta puanlara doğru yakınlaştığı görülmektedir. CE puanlarının FE puanlarına göre ham puana daha yakın değerler aldığı söylenebilir ancak hata ve yanlılık incelemesine göre karar verilmelidir.

Doğrusal ve eşit yüzdelikli eşitlemelerde elde edilen eşitlenmiş puanların ham puanlara göre durumunun paylaşıldığı Şekil 21 'deki görselde bu eşitlemelere ait standart hataların yönelimi de sunulmuştur.

Şekil 21

DOG-OM Deseni GA Becerisi Eşitlenmiş Puanve Standart Hatalar

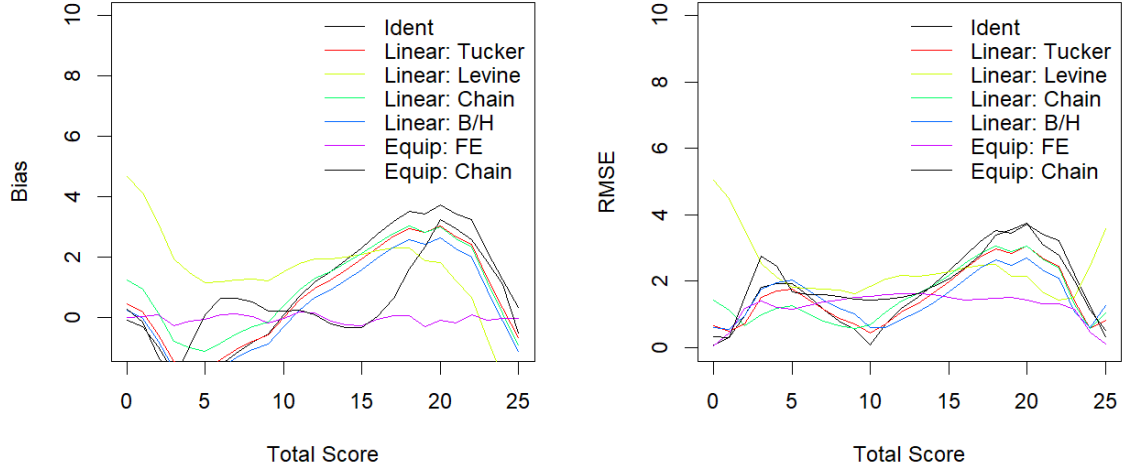


Şekil 21'de doğrusal eşitleme yapılan yöntemlerden elde edilen eşitlenmiş puanların ham puanlara çok yakın olduğu, eşit yüzdelikli eşitleme puanlarının ise ham puanlara göre sapma gösterdiği görülmektedir. Standart hatalar incelendiğinde ise Tucker, Chain ve Braun/Holland yöntemlerine ait hataların 0,5 civarındadır. Eşit yüzdelikli eşitlemede ise CE yöntemine ait standart hataların FE'ye göre çok daha yüksek değerler almaktadır. Standart hatalar bakımından doğrusal eşitleme yöntemleri ile FE yöntemi daha uygun görünmektedir.

Aynı biçimde doğrusal ve eşit yüzdelikli eşitlemelere ait yanlılık ve RMSE değerleri Şekil 22 aracılığı ile incelenmiştir.

Şekil 22

DOG-OM Deseni GA Becerisi Eşitlenmiş Puan Yanlılık ve RMSE Değerleri



Şekil 22’de görüldüğü üzere FE eşit yüzdelikli eşitleme ait yanlılık değerlerinin 0 civarında değerler almaktadır. Doğrusal eşitlemeye ait yanlılık değerleri ise -2 civarı ile 3,8 civarı değerler arasında değişmektedir. RMSE değerleri ise FE eşit yüzdelikli eşitlemede 0 ile 2 arasında değerler alarak diğerlerine göre daha düşüktür. Diğer eşitlemelerde ise 0 ile 4 arasında birbirine yakın ve puanlara göre farklılaşan RMSE değerleri göze çarpmaktadır. Bu değerler 0-5 ile 10-20 puan aralılarında artarken diğer puan aralıklarında azalmaktadır.

Tablo 29 DOG-OM deseni çerçevesinde uygulanmış olan doğrusal ve eşit yüzdelikli eşitlemelere ait bootstrap standart hata, yanlılık, RMSE ortalama değerleri ile bunların ağırlıklandırılmış değerlerine ait ortalamaları Tablo 28’de verilmiştir.

Tablo 29*DOG-OM Deseni GA Becerisi Eşitlenmiş Puan Bootstrap Hata, Yanlılık ve RMSE**Değerleri*

	<i>SE</i>	<i>SE_w</i>	<i>BIAS</i>	<i>BIAS_w</i>	<i>RMSE</i>	<i>RMSE_w</i>
Birim	0,00	0,00	2,10	1,54	2,10	1,54
Tucker	0,46	0,47	1,70	1,24	1,76	1,32
Chain	0,58	0,61	1,68	1,36	1,78	1,49
B/H	0,56	0,57	1,56	1,19	1,65	1,32
FE	1,32	0,87	0,14	0,09	1,32	0,87
CE	1,68	1,08	1,35	1,14	2,15	1,57

Tablo 29'daki değerler incelendiğinde doğrusal eşitleme yöntemlerinin genel olarak yakın hata değerlerine sahip olduğu görülmektedir. Doğrusal eşitleme yöntemleri içerisinde standart hatası en düşük olan yöntem Tucker, yanlılıkta ve RMSE değerlerinde en düşük olan yöntem ise Braun/Holland'tır. Eşit yüzdelikli eşitlemelerde FE yöntemi CE yöntemine göre daha düşük değerlere sahiptir. Tabloya genel olarak bakıldığında yapılan eşitlemelerin düşük hatalara sahip olduğu belirtilebilir.

iii. Kernel Eşitleme (AxB / DOG-OM/ Kernel)

DOG-OM deseninde kernel eşitleme uygulaması Son Tabakalama (*poststratification equating, PSE*) ve Zincir Eşitleme (*chain equating, CE*) yöntemleri ile yapılmıştır. PSE ve CE yöntemleri ile elde edilen eşitlenmiş puan değerlerinin yakın olması sebebiyle burada Tablo 29'da sadece PSE kernel eşit yüzdelikli ve doğrusal eşitlemelere ait eşitlenen puanlar ve standart hataları paylaşılmıştır.

Tablo 30*DOG-OM Deseni GA Becerisi Kernel PSE EYE ve Doğrusal Eşitleme Puan ve Hataları*

<i>Ham Puan</i>	<i>EYE</i>		<i>Doğrusal</i>	
	<i>Eşitlenmiş Puan</i>	<i>SE</i>	<i>Eşitlenmiş Puan</i>	<i>SE</i>
0	0,12	0,04	0,47	0,441
1	1,43	0,18	1,42	0,440
2	2,75	0,36	2,36	0,439
3	4,02	0,53	3,31	0,439

Tablo 30 (Devamı)

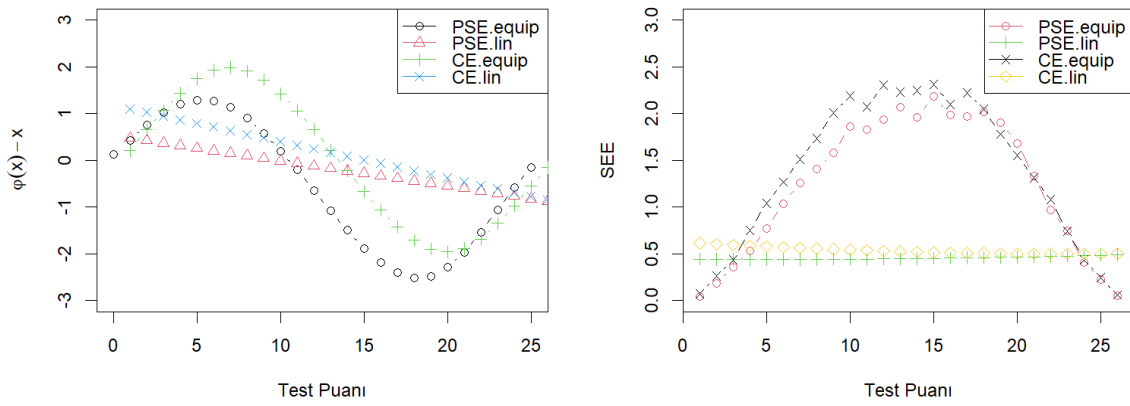
4	5,20	0,77	4,26	0,439
5	6,29	1,04	5,20	0,439
6	7,26	1,26	6,15	0,439
7	8,13	1,41	7,10	0,439
8	8,90	1,58	8,04	0,440
9	9,58	1,86	8,99	0,441
10	10,19	1,83	9,93	0,442
11	10,79	1,93	10,88	0,444
12	11,34	2,07	11,83	0,445
13	11,92	1,96	12,77	0,447
14	12,50	2,18	13,72	0,450
15	13,12	1,98	14,67	0,452
16	13,82	1,97	15,61	0,455
17	14,59	2,02	16,56	0,457
18	15,48	1,90	17,50	0,461
19	16,52	1,68	18,45	0,464
20	17,71	1,33	19,39	0,467
21	19,03	0,97	20,34	0,471
22	20,46	0,74	21,29	0,475
23	21,93	0,40	22,23	0,479
24	23,41	0,22	23,18	0,483
25	24,84	0,05	24,12	0,488

Tablo 30'da eşitlenen puanlar ham puana ait sınırlar içerisinde olup, eşitlenen puanlar PSE kernel eşit yüzdelikli eşitlemede ilk 10 puana kadar ham puan değerlerinden yüksek ve sonrasında ham puanlardan daha düşüktür. Eşitlenmiş puanlar ile ham puanlar arasındaki fark eşit yüzdelikli eşitlemede 1,29 ile -2,52 değerleri arasında değişmektedir. Eşitlenmiş puan ile ham puan farkının 14 ile 19 puan aralığında diğer puan aralıklarına göre daha fazla yükseldiği görülmektedir. Eşitlenmiş puanlara ait standart hatalar ise 0,04 ile 2,18 arasında değişmektedir. Eşitlenmiş puan aralığına göre uçlarda daha düşük hata değerleri gözlenirken ortalarda hata değerleri uçlara göre biraz daha yüksek durumdadır. PSE kernel doğrusal eşitlemede ise ilk 8 puana kadar eşitlenmiş puanlar ham puan değerlerinden yüksek ve sonrasında ham puanlardan daha düşüktür. Eşitlenmiş puanlar ile ham puanlar arasındaki fark 0,47 ile -0,88 değerleri arasında değişmektedir ve bu eşitlenmiş puanlara ait standart hata değerleri de 0,44 ile 0,49 değerleri arasındadır. Genel olarak PSE kernel eşit yüzdelikli ve doğrusal eşitlemelerde eşitlenen puanlar ve

standart hatalar bakımından uygun bir eşitleme görülmektedir. Burada bahsedilen eşitlenen puanlar ile ham puanların farkına ve eşitlemelerin standart hatalarına ait verilerle CE kernel eşit yüzdelikli ve doğrusal eşitlemelere ait bilgiler aşağıda Şekil 23'te görselleştirilmiştir.

Şekil 23

DOG-OM Deseni GA Becerisi Kernel PSE ve CE Eşitleme Puan Fark ve Standart Hataları

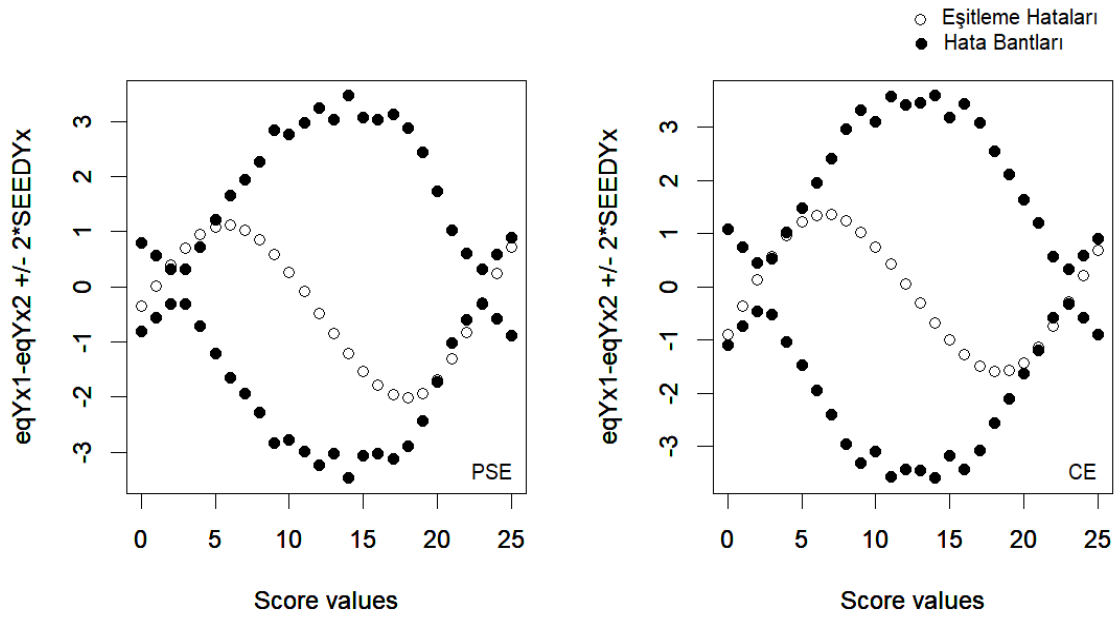


Şekil 23'te PSE ve CE kernel eşitlemelerinin ham puan farklarına ait doğru ve eğri grafiklerinin benzer yönelimde oldukları görülmektedir. Aynı durum bu eşitlemelere ait standart hata değerleri için de gözlenmektedir. Doğrusal eşitlemeler 0 ve civarında farka sahipken, eşit yüzdelikli eşitlemelerde bu aralık daha geniştir. Doğrusal eşitlemelerde hata değerleri test puanları boyunca daha durağan bir düzeydeyken, eşit yüzdelikli eşitlemelerde ise bu değerler 15 puan civarına kadar artan ve sonrasında azalan eğri yönelimindedir.

Eşitlemenin uygunluğunun incelenebileceği diğer bir hata türü de *SEED* değerleridir. PSE kernel ve CE kernel eşitlemelerinde eşit yüzdelikli eşitleme ile doğrusal eşitleme arasındaki eşitleme puanları farkına ait standart hata (*SEED*) değerleri Şekil 24'te paylaşılmıştır.

Şekil 24

DOG-OM Deseni GA Becerisi Kernel PSE ve CE SEED Değerleri



Şekil 24 incelendiğinde PSE ve CE kernel eşitlemelerde eşit yüzdelli eşitleme puanları ile doğrusal eşitleme puanları arasındaki farkların standart hata değerlerinin hata bantları içerisinde olduğu görülmektedir. Kernel eşitlemelerin uygun görüldüğü beirtilebilir.

Ayrıca, Tablo 31 PSE ve CE kernel eşitlemelere ait yüzde göreceli hata (PRE) değerlerini sunmaktadır.

Tablo 31

DOG-OM Deseni GA Becerisi Kernel PSE ve CE Yüzde Göreceli Hata Değerleri

	PSE	CE	
		B to OM	OM to A
1	0,02	0,04	0,07
2	0,15	-0,38	0,08
3	0,23	-0,29	0,06
4	0,28	-0,24	0,06
5	0,31	-0,30	0,07
6	0,32	-0,44	0,10
7	0,31	-0,65	0,14
8	0,29	-0,94	0,19
9	0,25	-1,30	0,26
10	0,20	-1,72	0,34

Tablo 31’de ilk 10 moment değeri için elde edilen yüzde göreceli değerleri (*PRE*) PSE kernel eşit yüzdeliği eşitleme için 0,02 ile 0,32 değerleri arasında değişmektedir. Değerler sıfıra çok yakın ve 10. değer ise 0,20’dir. CE kernel eşit yüzdeliği eşitlemenin iki aşamasındaki *PRE* değerlerigenel olarak sıfır civarında ve bu değerler -1,72 ile 0,04 arasında değişmektedir. *PRE* değerlerinin tabloda genel olarak düşük değerlerde olduğu görülmektedir.

DOG-OM desenlerinde genelleştirilmiş aritmetik beceri puanlarının AxB testleri için eşitleme durumu doğrusal, eşit yüzdeliği ve kernel eşitleme ile incelendiğinde eşitleme uygulamalarının uygunluğu doğrultusunda bulgulara ulaşılmıştır. AxB test eşitleme uygulamaları benzer biçimde AxC ve BxC test eşitlemeleri için de uygulanmıştır. Bu uygulamalardan elde edilen bulgular Tablo 32’de sunulmuştur.

Tablo 32

DOG-OM Deseninde GA Becerisinin Testlere Göre Eşitlenebilme Durumları

<i>Desen</i>	<i>Eşitleme Türü</i>	<i>Eşitlenen Testler</i>		
		<i>AxB</i>	<i>AxC</i>	<i>BxC</i>
<i>Denk Olmayan Gruplarda Ortak Madde</i>	Doğrusal	Tucker, B/H	Tucker	Tucker
	Eşit Yüzdeliği	FE	-	-
	Kernel	PSE, CE	-	-

Tablo 32’de verilenlere göre denk olmayan gruplarda ortak madde deseninde uygulanan üç farklı eşitlemeden sadece AxB eşitlemesinde doğrusal, eşit yüzdeliği ve kernel eşitlemeler yapılabilmiş, diğer eşitlemelerde ise sadece doğrusal eşitleme yapılabilirken eşit yüzdeliği ve kernel eşitlemelerde uygun bulgulara ulaşılmamıştır.

GA Becerisinin Denk Olmayan Gruplarda Ortak Değişken (DOG-OD) Desenine Göre Eşitleme İncelemeleri

DOG-OD deseninde GA becerileri için eşitleme incelemelerinde okulların kabul koşulu, okul türleri ve cinsiyet değişkenleri kullanılmıştır. Bu desen çerçevesindeki incelemelerde kernel eşitleme uygulamaları kullanılmıştır. Aşağıda AxB testine ait okulların kabul koşulu-okul türleri ile okulların kabul koşulu-cinsiyet değişkenleri çerçevesinde eşitlemeler verilmiş olup, AxC ve BxC eşitlemelerine ait bulgular desen ile ilgili genel tabloda sunulmuştur.

GA Becerisinin Test A ve Test B için DOG-OD Deseninde Eşitleme Uygulamaları

i. Okulların Kabul Koşulu ve Okul Türleri Değişkenine Göre

DOG-OD deseninde kernel eşitleme uygulaması Son Tabakalama (poststratification equating, PSE) yöntemi ile yapılmıştır. PSE kernel eşit yüzdelikli ve doğrusal eşitlemeye ait eşitlenen puanlar ve standart hataları Tablo 33'te verilmiştir.

Tablo 33

DOG-OD Deseni GA Becerisi Kernel EYE ve Doğrusal Eşitleme Puan ve Hataları

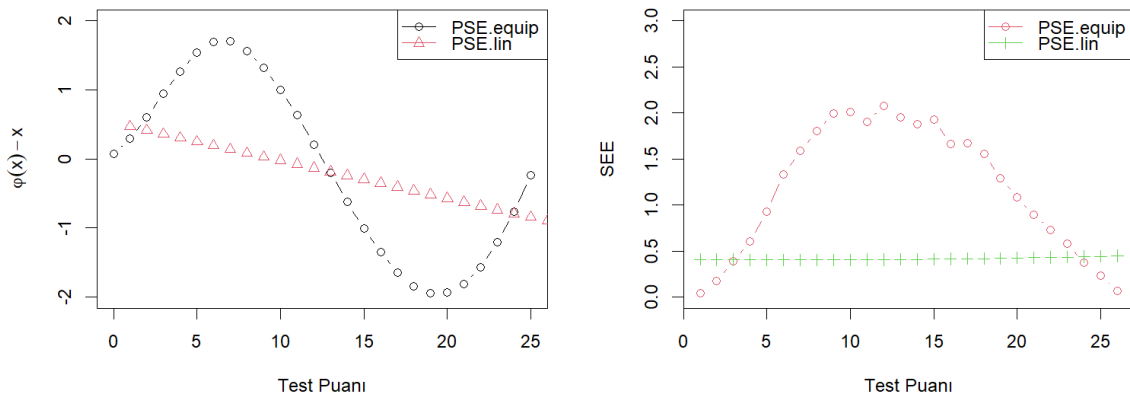
<i>Ham Puan</i>	<i>EYE</i>	<i>SEE</i>	<i>Doğrusal</i>	<i>SEE</i>
0	0,08	0,044	0,47	0,408
1	1,29	0,178	1,41	0,406
2	2,60	0,392	2,36	0,405
3	3,95	0,603	3,30	0,404
4	5,27	0,924	4,25	0,403
5	6,54	1,333	5,19	0,403
6	7,69	1,589	6,14	0,403
7	8,70	1,799	7,08	0,403
8	9,56	1,996	8,03	0,403
9	10,32	2,009	8,97	0,404
10	11,00	1,898	9,92	0,405
11	11,63	2,075	10,87	0,406
12	12,21	1,948	11,81	0,408
13	12,80	1,878	12,76	0,409
14	13,37	1,927	13,70	0,411
15	13,99	1,658	14,65	0,413

Tablo 33 (Devamı)

16	14,65	1,669	15,59	0,416
17	15,36	1,551	16,54	0,418
18	16,15	1,292	17,48	0,421
19	17,05	1,081	18,43	0,424
20	18,06	0,893	19,37	0,427
21	19,18	0,730	20,32	0,431
22	20,44	0,581	21,26	0,435
23	21,80	0,369	22,21	0,438
24	23,23	0,231	23,16	0,443
25	24,77	0,069	24,10	0,447

Şekil 25

DOG-OD Deseni GA Becerisi Kernel PSE EYE ve Doğrusal Eşitleme Puan Fark ve Standart Hataları

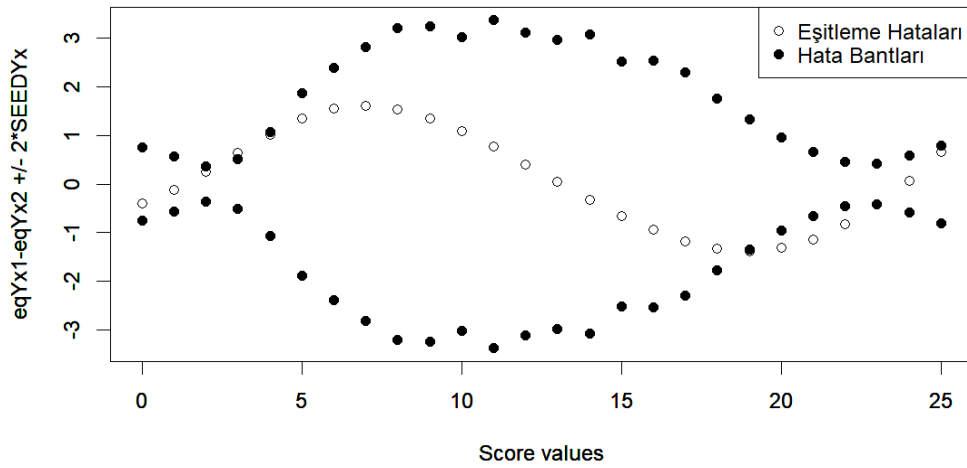


Tablo 33 ve Şekil 25 birlikte değerlendirilecek olursa eşitlenen puanlar ham puana ait sınırlar içerisinde olup, eşitlenen puanlar PSE kernel eşit yüzdelikli eşitlemede ilk 12 puana kadar ham puan değerlerinden yüksek ve sonrasında ham puanlardan daha düşüktür. Eşitlenmiş puanlar ile ham puanlar arasındaki fark eşit yüzdelikli eşitlemede 1,70 ile -1,95 değerleri arasında değişmektedir. Eşitlenmiş puan ile ham puan farkının uç puan aralıklarında diğer puan aralıklarına göre daha düşük olduğu görülmektedir. Eşitlenmiş puanlara ait standart hatalar ise 0,04 ile 2,07 arasında değişmektedir. Eşitlenmiş puan aralığına göre uçlarda daha düşük hata değerleri gözlenirken ortalarda hata değerleri uçlara göre biraz daha yüksek durumdadır. PSE kernel doğrusal eşitlemede

ise ilk 8 puana kadar eşitlenmiş puanlar ham puan değerlerinden yüksek ve sonrasında ham puanlardan daha düşüktür. Eşitlenmiş puanlar ile ham puanlar arasındaki fark -0,90 ile 0,47 değerleri arasında değişmektedir ve bu eşitlenmiş puanlara ait standart hata değerleri de 0,40 ile 0,45 değerleri arasındadır. Genel olarak PSE kernel eşit yüzdelikli eşitlemede eşitlenen puanlar ve standart hatalar bakımından uygun bir eşitleme görülmektedir. Burada bahsedilen eşitlenen puanlar ile ham puanların farkına ve eşitlemelerin standart hatalarına ait bilgiler yanında eşit yüzdelikli eşitleme ile doğrusal eşitleme arasındaki eşitleme puanları farkına ait standart hata (*SEED*) değerleri Şekil 26'da paylaşılmıştır.

Şekil 26

DOG-OD Deseni GA Becerisi Kernel PSE EYE ve Doğrusal Eşitleme SEED Değerleri



Şekil 26 incelendiğinde hata bantlarının -3 ile 3 seviyelerinde olduğu görülmektedir. PSE kernel eşitlemelerde eşit yüzdelikli eşitleme puanları ile doğrusal eşitleme puanları arasındaki farkların standart hata değerlerinin genel olarak bu hata bantları içerisinde olduğu görülmektedir. Kernel eşitlemenin uygun görüldüğü belirtilebilir. Ayrıca, Tablo 34 PSE kernel eşit yüzdelikli eşitlemeye ait yüzde göreceli hata (PRE) değerlerini sunmaktadır.

Tablo 34*DOG-OD Deseni GA Becerisi Kernel PSE EYE Yüzde Göreceli Hata Değerleri*

	PRE
1	0,00
2	0,11
3	0,18
4	0,22
5	0,21
6	0,23
7	0,22
8	0,20
9	0,15
10	0,08

Tablo 34'de ilk 10 moment değeri için elde edilen yüzde göreceli hata değerleri (PRE) PSE kernel eşit yüzdellikli eşitleme için 0,00 ile 0,23 değerleri arasında değişmektedir. Değerler sıfıra çok yakın ve 10. değer ise 0,08'dir. PRE değerlerinin tabloda düşük değerlerde olduğu görülmektedir.

Okulların kabul koşulu ve okul türleri değişkenlerine göre yapılan PSE kernel eşit yüzdellikli eşitlemede eşitlenmiş puanlar, eşitleme hataları, eşitlenmiş puan ve ham puan farkları, SEED ve PRE değerlerinden elde edilen verilere göre eşitlemenin uygun olduğu söylenebilir.

ii. Okulların Kabul Koşulu ve Cinsiyete Göre

DOG-OD deseninde okulların kabul koşulu ve cinsiyet değişkenlerinin kullanılarak uygulanan kernel PSE eşitlemede elde edilen eşit yüzdellikli ve doğrusal eşitlemeye ait eşitlenen puanlar ve standart hataları Tablo 35'da paylaşılmıştır. Bu puanlara ait fark ve standart hatalar Şekil 27'de görselleştirilmiştir.

Tablo 35*DOG-OD Deseni GA Becerisi Kernel EYE ve Doğrusal Eşitleme Puan ve Hataları*

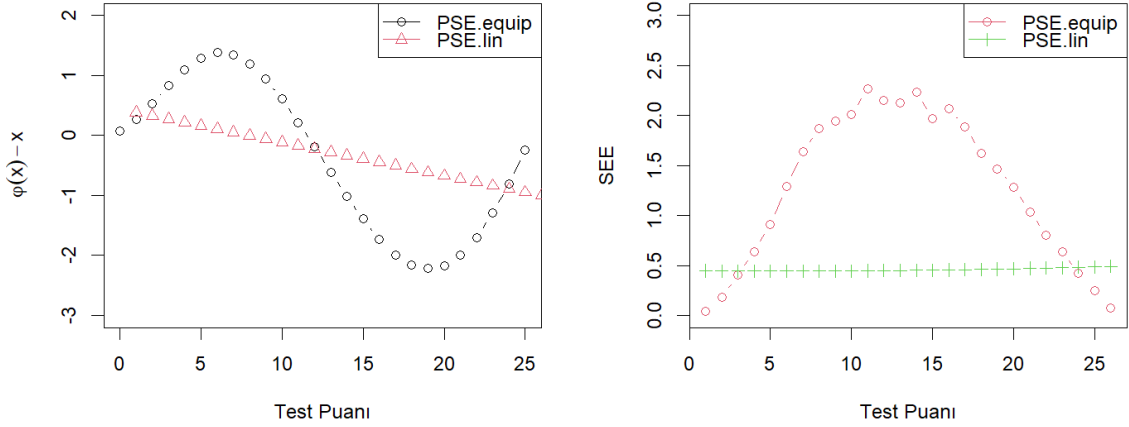
Ham Puan	EYE	SEE	Doğrusal	SEE
0	0,07	0,046	0,38	0,447
1	1,26	0,183	1,33	0,446
2	2,53	0,403	2,27	0,445
3	3,83	0,641	3,22	0,444
4	5,09	0,910	4,16	0,444
5	6,29	1,287	5,11	0,444
6	7,38	1,640	6,05	0,444
7	8,34	1,868	7,00	0,444
8	9,18	1,941	7,94	0,445
9	9,93	2,007	8,88	0,446
10	10,60	2,265	9,83	0,447
11	11,21	2,153	10,77	0,448
12	11,81	2,128	11,72	0,450
13	12,38	2,232	12,66	0,452
14	12,98	1,968	13,61	0,454
15	13,61	2,065	14,55	0,456
16	14,26	1,886	15,50	0,459
17	15,01	1,624	16,44	0,462
18	15,85	1,468	17,38	0,464
19	16,78	1,281	18,33	0,468
20	17,83	1,037	19,27	0,471
21	19,00	0,800	20,22	0,475
22	20,29	0,641	21,16	0,479
23	21,71	0,421	22,11	0,483
24	23,19	0,247	23,05	0,487
25	24,76	0,074	24,00	0,491

Tablo 35'te verilenlere göre eşit yüzdelikli eşitlemede eşitlenmiş puanlar 11 puana kadar ham puan değerlerinden yüksek, sonrasında ise düşüktür. Genel olarak bakıldığında 15 puan sonrasında eşitlenen puanlarda ham puanlara göre daha açık fark bulunurken uç puanlarda bu durum gözlenmemiştir. Doğrusal eşitlemede ise 7 puana kadar eşitlenmiş puanlar ham puanlardan yüksek sonrasında ise düşüktür. Eşitlenen puanların ham puanlardan farkı 7 puana kadar azalırken sonrasında bu fark 25 puana kadar artış göstermiştir. Genel olarak değerlendirildiğinde ise eşitlenen puanlar ile ham puanlar arasındaki fark uygun seviyede görünmektedir. Aşağıda Şekil 27 tablodaki

eşitlenen puanlar ile ham puanlar arasındaki farkı ve eşitlemeye ait standart hatayı görsel olarak sunmaktadır.

Şekil 27

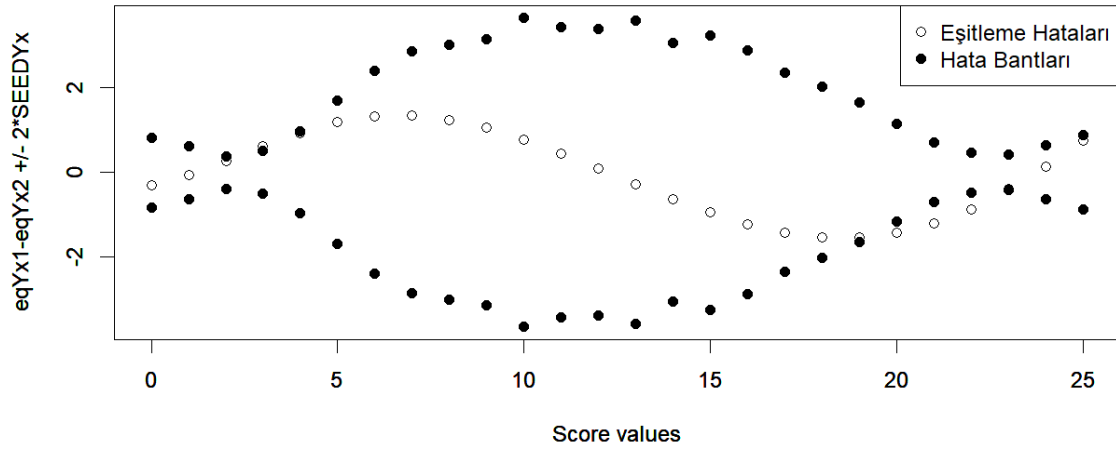
DOG-OD Deseni GA Becerisi PSE Kernel EYE ve Doğrusal Eşitleme Puan Fark ve Standart Hataları



Yukarıda bahsedilen fark puanları Şekil 27'de de görselleştirildiği üzere eşit yüzdelikli eşitleme için -2,22 ile 1,38 arasında ve doğrusal eşitleme için -1 ile 0,38 arasında değişmektedir. Standart hatalar açısından doğrusal eşitlemede 0,45 ile 0,49 arasında değerler alan hemen hemen sabit görünümde bir doğru görülmektedir. Eşit yüzdelikli eşitlemede ise uçlarda hata değerlerinin 0 civarına yaklaşırken orta puanlarda bu değer 2,27 seviyelerine çıkmıştır. Burada bahsedilen eşitlenen puanlar ile ham puanların farkına ve eşitlemelerin standart hatalarına ilişkin bilgiler yanında eşit yüzdelikli eşitleme ile doğrusal eşitleme arasındaki eşitleme puanları farkına ait standart hata (SEED) değerleri Şekil 30'da paylaşılmıştır.

Şekil 28

DOG-OD Deseni GA Becerisi PSE Kernel EYE ve Doğrusal Eşitleme SEED Değerleri



Şekil 28 incelendiğinde hata bantlarının -3 ile 3 seviyelerinde olduğu görülmektedir. PSE kernel eşitlemelerde eşit yüzdelli eşitleme puanları ile doğrusal eşitleme puanları arasındaki farkların standart hata değerlerinin genel olarak bu hata bantları içerisinde olduğu görülmektedir. Kernel eşitlemenin uygun görüldüğü belirtilebilir. Ayrıca, Tablo 35 PSE kernel eşit yüzdelli eşitlemeye ait yüzde göreceli hata (PRE) değerlerini sunmaktadır.

Tablo 36

DOG-OD Deseni GA Becerisi Kernel PSE EYE Yüzde Göreceli Hata Değerleri

	PRE
1	0,01
2	0,13
3	0,20
4	0,24
5	0,26
6	0,25
7	0,21
8	0,16
9	0,08
10	-0,01

Tablo 36'de ilk 10 moment değeri için elde edilen yüzde göreceli hata değerleri (PRE) PSE kernel eşit yüzdelikli eşitleme için -0,01 ile 0,26 değerleri arasında değişmektedir. PRE değerlerinin tabloda düşük değerlerde olduğu görülmektedir.

Okulların kabul koşulu ve cinsiyet değişkenlerine göre yapılan PSE kernel eşit yüzdelikli eşitlemede eşitlenmiş puanlar, eşitleme hataları, eşitlenmiş puan ve ham puan farkları, SEED ve PRE değerlerinden elde edilen verilere göre eşitlemenin uygun olduğu söylenebilir.

Denk olmayan gruplarda ortak değişken desenlerinde genelleştirilmiş aritmetik beceri puanlarının AxB testlerinde okulların kabul koşulu-okul türü ve okulların kabul koşulu-cinsiyet değişkenleri çerçevesinde eşitlenme durumu PSE kernel eşit yüzdelikli eşitleme ve kernel eşitleme ile incelendiğinde eşitleme uygulamalarının kabul edilebilir sonuçlar verdiği görülmektedir. AxB test eşitleme uygulamaları benzer biçimde AxC ve BxC test eşitlemeleri için de uygulanmış ve bu uygulamalardan elde edilen bulgular Tablo 37'da sunulmuştur.

Tablo 37

DOG-OD Deseninde GA Becerisinin Testlere Göre Eşitlenebilme Durumları

<i>Desen</i>	<i>Eşitleme Türü</i>	<i>Eşitlenen Testler</i>		
		<i>AxB</i>	<i>AxC</i>	<i>BxC</i>
<i>Denk Olmayan Gruplarda Ortak Değişken</i>	Kernel (değişkenler: okul kabul koşulu & okul türü	PSE-EYE Doğrusal	–	–
	Kernel (değişkenler: okul kabul koşulu & cinsiyet	PSE-EYE Doğrusal	–	–

Tablo 37'da verilenlere göre denk olmayan gruplarda ortak değişken deseninde uygulanan üç farklı eşitlemeden sadece AxB eşitlemesi uygun bulgular vermiş, diğer eşitlemelerde AxC ve BxC için uygun bulgulara ulaşılmamıştır.

İkinci Araştırma Sorusu Kapsamındaki Bulgular

Fonksiyonel Düşünme (FD) Becerisinin Denk Grup (DG) Desenine Göre Eşitleme İncelemeleri

DG deseninde FD beceri puanları için de GA beceri puanlarının eşitleme incelemelerindeki aşamalar takip edilerek her üç testte çaprazlama olarak eşitleme uygulamaları yapılmıştır. Bu desen çerçevesinde ortalama, doğrusal, eşit yüzdelikli ve kernel eşitleme uygulamaları yapılmıştır. Aşağıda FD becerisinin AxB testine ait eşitlemeleri ayrıntılarıyla paylaşılmış olup, AxC ve BxC eşitlemelerine ait genel bulgular desen ile ilgili genel tabloda sunulmuştur.

FD Becerisinin Test A ve Test B için DG Deseninde Eşitleme Uygulamaları

DG deseni kapsamında FD beceri puanlarının AxB testlerinde eşitlenebilme durumları incelenmiştir. Bu desen çerçevesinde ortalama, doğrusal, eşit yüzdelikli ve kernel eşitleme uygulamalarında farklı yöntemler denenmiş ve bu yöntemlere dair eşitlenmiş puanlar, bu puanların ham puanlardan farkı ve eşitleme hataları paylaşılmıştır.

i. Ortalama Eşitleme (AxB / DG / Ortalama)

FD becerisine ait ham puanlar ve AxB test eşitlemesinde ortalama eşitleme ile elde edilen eşitlenen puanlar Tablo 38'de sunulmuştur.

Tablo 38

DG Deseni FD Becerisi Ortalama Eşitleme Puanı

<i>Ham puan</i>	<i>Eşitlenmiş Puan</i>	<i>Ham puan</i>	<i>Eşitlenmiş Puan</i>
0	-0,09	13	12,91
1	0,91	14	13,91
2	1,91	15	14,91
3	2,91	16	15,91
4	3,91	17	16,91
5	4,91	18	17,91
6	5,91	19	18,91
7	6,91	20	19,91

Tablo 38 (Devamı)

8	7,91	21	20,91
9	8,91	22	21,91
10	9,91	23	22,91
11	10,91	24	23,91
12	11,91	25	24,91

Tablo 38'deki verilere göre yapılan ortalama eşitlemede eşitlenen puanların ham puanlardan 0,09 puan daha düşük olduğu görülmektedir. Eşitlenen puanların üst sınır ham puanı aşmadığı fakat alt sınır ham puandan daha düşük olduğu görülmektedir.

ii. Doğrusal Eşitleme (AxB / DG / Doğrusal)

FD becerisinin AxB test eşitlemesinde doğrusal eşitleme ile elde edilen eşitlenmiş puanları ve eşitlemeye ait standart hataları Tablo 39'da verilmiştir.

Tablo 39*DG Deseni FD Becerisi Doğrusal Eşitleme Puanı*

<i>Ham puan</i>	<i>Eşitlenmiş Puan</i>	<i>SE</i>	<i>Ham puan</i>	<i>Eşitlenmiş Puan</i>	<i>SE</i>
0	-0,093	0,139	13	12,904	0,232
1	0,907	0,142	14	13,904	0,244
2	1,907	0,146	15	14,904	0,257
3	2,907	0,151	16	15,904	0,270
4	3,906	0,156	17	16,903	0,284
5	4,906	0,162	18	17,903	0,298
6	5,906	0,168	19	18,903	0,313
7	6,906	0,175	20	19,903	0,329
8	7,905	0,183	21	20,902	0,345
9	8,905	0,192	22	21,902	0,362
10	9,905	0,201	23	22,902	0,380
11	10,905	0,211	24	23,902	0,398
12	11,905	0,221	25	24,902	0,417

Tablo 39 incelendiğinde eşitlenmiş puanların ortalama eşitleme puanları ile hemen hemen birebir olduğu, binde birler basamağına göre sayısal değişiklik içerdiği gözlenmektedir. Eşitlenen puanların ham puanlardan düşük olduğu ve eşitlenen puan ile

ham puan farklarının -0,093 ile -0,098 arasında deęiřtięi gözlenmektedir. Eřitlenmiř puanlara ait standart hatalar ise 0,14 ile 0,42 arasında olduęu görölmektedir.

iii. Eřit Yüzdelikli Eřitleme (AxB / DG / Eřit Yüzdelikli)

FD becerisine ait ham puanlar, AxB test eřitlemesinde eřit yüzdelikli eřitleme ile elde edilen eřitlenmiř puanlar ve eřitlemeye ait standart hatalar Tablo 40'ta paylařılmıřtır.

Tablo 40

DG Deseni FD Beceri Puanları EYE Puanı

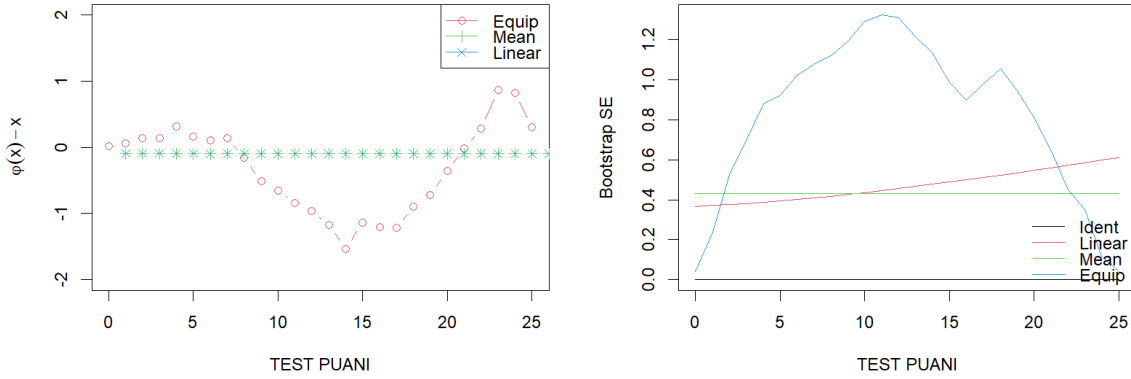
<i>Ham puan</i>	<i>Eřitlenmiř Puan</i>	<i>SE</i>	<i>Ham puan</i>	<i>Eřitlenmiř Puan</i>	<i>SE</i>
0	0,02	0,05	13	11,82	0,78
1	1,07	0,24	14	12,46	0,78
2	2,14	0,57	15	13,86	0,82
3	3,14	0,55	16	14,80	0,92
4	4,31	1,22	17	15,79	0,86
5	5,16	0,74	18	17,10	2,06
6	6,11	1,09	19	18,28	0,78
7	7,14	1,15	20	19,64	0,91
8	7,84	0,99	21	20,99	0,73
9	8,49	0,99	22	22,28	0,46
10	9,35	1,15	23	23,86	0,37
11	10,15	1,71	24	24,82	0,13
12	11,04	2,05	25	25,31	0,06

Eřit yüzdelikli eřitleme ile eřitlenen FD beceri puanları Tablo 40'ta göröldüęü üzere eřitlenen puanlar ilk 8 puana kadar ham puan deęerlerinden yüksek, sonrasında 21 puana kadar ham puanlardan daha düşük ve 25 puana kadar tekrar ham puanlardan daha yüksek olduęu görölmektedir. Bu farklar -1,54 ile 0,86 deęerleri arasında deęiřmektedir. Genel olarak 10 puan ile 20 puan aralıęında eřitlenmiř puan ile ham puan farkının yükseldięi görölmektedir. Eřitlenmiř puanlara ait standart hatalar ise 0,05 ile 2,06deęerleri arasında deęiřmektedir. Eřitlenmiř puan aralıęına göre uçlarda daha düşük hata deęerleri gözlenirken ortalarda hatalar uçlara göre daha yüksek durumdadır.

Aşağıda ortalama, doğrusal ve eşit yüzdellikli eşitlemeye ait eşitlenmiş puanların ham puanlardan farkı ve eşitlenen puanların bootstrap standart hataları ile ilgili görseller Şekil 29'da sunulmuştur.

Şekil 29

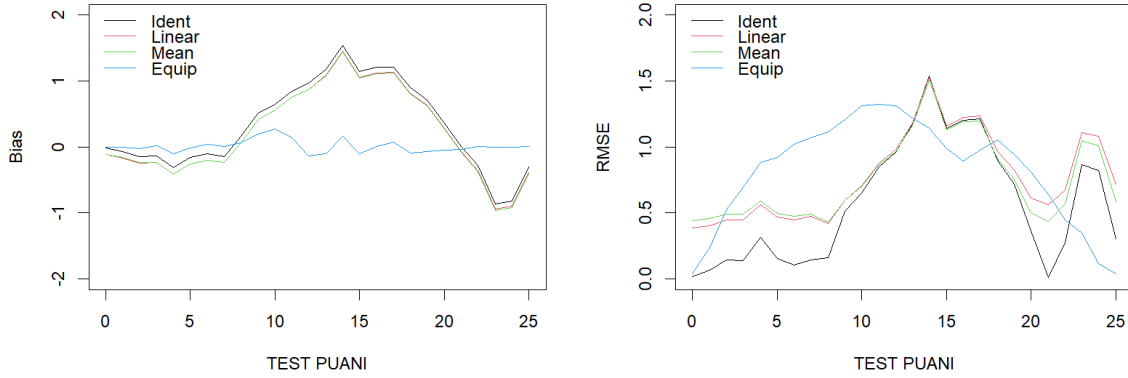
DG Deseni FD Becerisi Eşitlenmiş Puan Fark ve Bootstrap Standart Hataları



Şekil 29'da verilenlere göre ortalama ve doğrusal eşitlemede eşitlenmiş puanların yakınlığı ve bu puanların ham puanlardan farkının 0 dolaylarında olduğu görülmektedir. Eşit yüzdellikli eşitleme ile elde edilen eşitlenmiş puanların düşük beceri puanlarında ham puanlardan farkının 0 dolaylarında olduğu ancak daha sonra bu farkın -1,5 civarı ile 1 civarında farklılaştığı görülmektedir. Genel anlamda bu farkların düşük seviyede olduğu söylenebilir. Bootstrap standart hatalar incelendiğinde ise ortalama eşitlemede 0,43 seviyesinin gözlemlendiği, doğrusal eşitlemede bu hatalarının 0,4 ile 0,6 dolaylarında değerler aldıkları, eşit yüzdellikli eşitlemenin ise 0 ile 1,4 civarına kadar değişen değerler aldığı görülmektedir. Eşitlenmiş puanlara ait yanlılık ve RMSE değerleri Şekil 30'da paylaşılmıştır.

Şekil 30

DG Deseni FD Becerisi Eşitlenmiş Puan Yanlılık ve RMSE Değerleri



Şekil 30'da yanlılık değerleri incelendiğinde eşit yüzdelikli eşitlemede bu değerlerin 0 civarında farklı değerler aldığı görülmektedir. Diğer eşitlemelerde ise bu değerlerin 8 puana kadar yine 0 civarında olduğu sonrasında ise -1 ile 1,5 civarı arasında değiştiği görülmektedir. Ortalama ve doğrusal eşitlemenin yanlılık değeri açısından da çok yakın olduğu anlaşılmaktadır. RMSE değerlerinin ise 0 ile 1,5 değerleri arasında değiştiği görülmektedir. Ortalama ve doğrusal eşitlemeye ait değerlerin yine birbirine yakın olduğu, eşit yüzdelikli eşitlemenin ise bunlardan farklı eğilimde olduğu görülmektedir. Yanlılık açısından düşük değerlere sahip olan eşit yüzdelikli eşitleme RMSE değerleri açısından ortalama ve doğrusal eşitlemeye göre genel olarak daha yüksek değerler almıştır. Ancak genel olarak yanlılık ve RMSE değerleri her üç eşitleme türünde de düşük olduğu söylenebilir.

Denk grup desenlerinde ortalama, doğrusal ve eşit yüzdelikli eşitleme ile eşitlenmiş puanlara ait bootstrap hata, yanlılık ve RMSE değerleri ile bunların ağırlıklandırılmış değerlerine ait ortalamaları aşağıda Tablo 41'de verilmiştir.

Tablo 41*DG Deseni FD Becerisi Eşitlenmiş Puan Bootstrap Hata, Yanlılık ve RMSE Değerleri*

	<i>SE</i>	<i>SE_w</i>	<i>BIAS</i>	<i>BIAS_w</i>	<i>RMSE</i>	<i>RMSE_w</i>
birim	0,00	0,00	0,72	0,57	0,72	0,57
doğrusal	0,48	0,46	0,69	0,56	0,84	0,72
ortalama	0,43	0,43	0,69	0,56	0,81	0,71
eye	0,91	0,67	0,09	0,07	0,91	0,67

Not: eye: eşit yüzdellikli eşitleme

Tablo 41'e göre eşitleme hata, yanlılık ve RMSE değerlerine genel olarak bakıldığında 0 ile 0,91 arasında değişen değerler olduğu görülmektedir. Bu tabloda ortalama ve doğrusal eşitlemelerin standart hataları ve RMSE değerleri eşit yüzdellikli eşitlemeden düşüktür. Yanlılık açısından bakıldığında en düşük değere eşit yüzdellikli eşitleme sahiptir. Tablodaki verilerin genel olarak eşitlemelerin uygunluğu açısından destek verdiği söylenebilir.

iv. Kernel Eşitleme (AxB / DG / Kernel)

FD becerisine ait ham puanlar, AxB test eşitlemesinde logistic kernel eşit yüzdellikli eşitleme ile doğrusal eşitlemeden elde edilen eşitlenmiş puanlar ve eşitlemeye ait standart hatalar Tablo 42'de verilmiştir.

Tablo 42*DG Deseni FD Becerisi Logistic Kernel EYE ve Doğrusal Eşitleme Puan ve Hataları*

<i>Ham Puan</i>	<i>EYE</i>		<i>Doğrusal</i>	
	<i>Eşitlenmiş Puan</i>	<i>SE</i>	<i>Eşitlenmiş Puan</i>	<i>SE</i>
0	0,01	0,040	-0,09	0,373
1	1,06	0,187	0,91	0,377
2	2,16	0,430	1,91	0,382
3	3,24	0,695	2,91	0,388
4	4,24	0,859	3,91	0,395
5	5,16	0,908	4,91	0,402
6	6,05	0,919	5,91	0,410
7	6,91	0,966	6,91	0,419
8	7,76	1,086	7,91	0,428
9	8,59	1,237	8,91	0,438

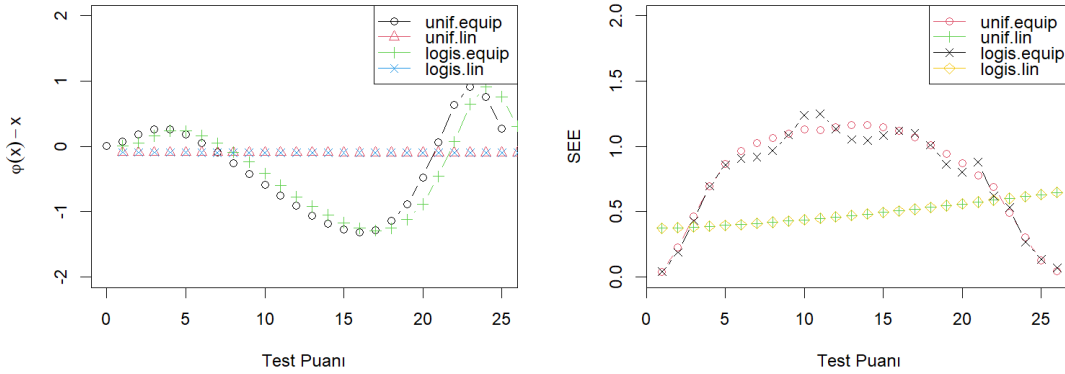
Tablo 42 (Devamı)

10	9,40	1,250	9,90	0,448
11	10,23	1,134	10,90	0,459
12	11,08	1,053	11,90	0,470
13	11,95	1,044	12,90	0,482
14	12,83	1,082	13,90	0,494
15	13,75	1,119	14,90	0,506
16	14,71	1,102	15,90	0,519
17	15,75	1,009	16,90	0,532
18	16,88	0,864	17,90	0,546
19	18,11	0,801	18,90	0,559
20	19,54	0,877	19,90	0,573
21	21,07	0,621	20,90	0,587
22	22,65	0,532	21,90	0,602
23	23,91	0,270	22,90	0,616
24	24,76	0,138	23,90	0,631
25	25,31	0,069	24,90	0,646

Tablo 42’de logistic kernel eşit yüzdelikli eşitleme ile eşitlenen puanlar ilk 6 puana kadar ve 21’den itibaren ham puan değerlerinden yüksektir. 6 ile 21 puan aralığında ise eşitlenmiş puanların ham puanlardan daha düşük olduğu görülmektedir. Ayrıca bu aralıkta eşitlenmiş puan ile ham puan farkının diğer puan aralıklarına göre daha yüksek olduğu görülmektedir. Fakat genel olarak eşitlenmiş puan ile ham puan farkının düşük olduğu söylenebilir. Eşitlenmiş puanlara ait standart hatalar ise 0,04 ile 1,25 arasında değişmektedir. Logistic kernel doğrusal eşitleme ile eşitlenen puanlar ham puan değerlerinden genel olarak 0,10 puan düşüktür. Eşitlenmiş puanlara ait standart hatalar ise 0,37 ile 0,65 arasında değişmektedir, bu değişim düşük puandan yüksek puana artarak ilerlemiştir. Aşağıda Şekil 31’da FD becerisine ait logistic ve uniform kernel eşitlemelerde eşit yüzdelikli ve doğrusal eşitlemelerden elde edilen eşitlenmiş puan ile ham puan farklarının ve standart hatalarının yer aldığı grafikler paylaşılmıştır.

Şekil 31

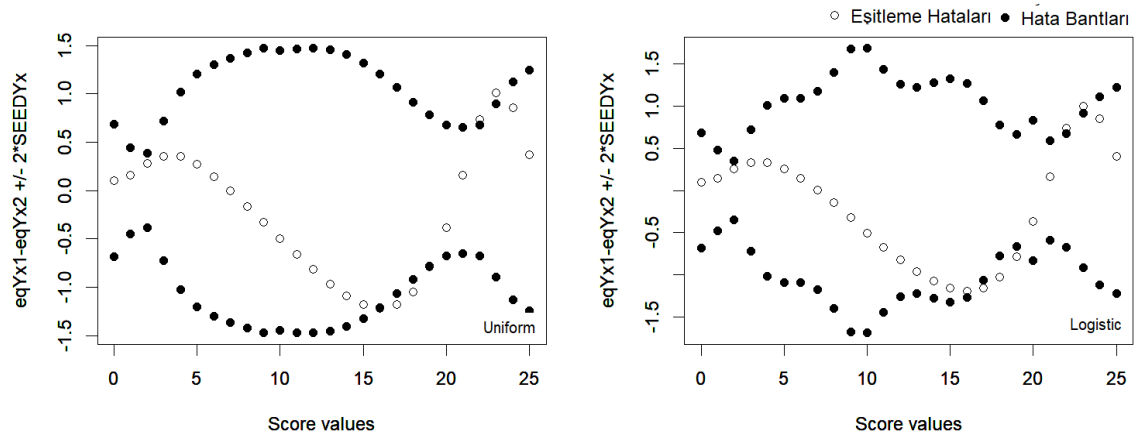
DG Deseni FD Becerisi Kernel Eşitleme Puan Fark ve Standart Hataları



FD becerisinin logistic ve uniform doğrusal eşitlemelere ait eşitlenmiş puan farklarının çakıştığı eşit yüzdelikli eşitlemelerin ise çok yakın değerler aldığı Şekil 31'de verilmiştir. Logistic ve uniform kernel eşit yüzdelikli eşitlemede eşitlenmiş puanların ham puanlardan farkı -1,29 ile 0,91 değerleri arasında değişmektedir. Diğer taraftan logistic ve uniform kernel doğrusal eşitlemede eşitlenmiş puanların ham puanlardan farkının sabit denilebilecek kadar ufak sayısal fark ile -0,10 ile -0,09 arasındadır. Eşitleme hatalarına bakıldığında ise uniform kernel doğrusal eşitleme ile logistic kernel doğrusal eşitlemeye ait değerlerin çakıştığı ve 0,4 ile 0,7 dolaylarında değişen değerler aldığı görülmektedir. Yine logistic ve uniform kernel eşit yüzdelikli eşitlemeye ait hata değerlerinin ise birbirine yakın olduğu, 0 ile 1,25 arasında değerler aldığı görülmektedir. Ayrıca logistic kernel ve uniform kernel eşitlemelerde eşit yüzdelikli eşitleme ile doğrusal eşitleme arasındaki eşitleme puanları farkına ait standart hata (*standard error of equating differences, SEED*) değerleri Şekil 32'de paylaşılmıştır.

Şekil 32

DG Deseni FD Becerisi Uniform ve Logistic Kernel SEED Değerleri



Şekil 32 incelendiğinde logistic ve uniform kernel eşitlemelerde eşit yüzdelikli eşitleme puanları ile doğrusal eşitleme puanları arasındaki farkların standart hata değerlerinin genel olarak hata bantları içerisinde olduğu görülmektedir. Aşağıda Tablo 43 ise logistic ve uniform kernel eşitlemelere ait yüzde göreceli hata (percent relative error, PRE) değerlerini sunmaktadır.

Tablo 43

DG Deseni FD Becerisi Uniform ve Logistic Kernel Yüzde Göreceli Hata Değerleri

	uni.PRE	logis.PRE
1	0,003	0,009
2	0,027	0,111
3	0,037	0,168
4	0,037	0,203
5	0,033	0,232
6	0,030	0,261
7	0,031	0,295
8	0,037	0,338
9	0,050	0,390
10	0,072	0,454

Tablo 43'de görüldüğü üzere ilk 10 moment değeri için elde edilen yüzde göreceli değerleri (*PRE*) uniform kernel eşit yüzdelikli eşitleme için 0,003 ile 0,072 değerleri arasında ve logistic kernel eşit yüzdelikli eşitleme için 0,009 ile 0,45 değerleri arasında değişmiştir. İyi bir eşitleme için *PRE* değerlerinin düşük değerde olması beklenir, tabloda bu değerlerin uniform kernel için çok daha düşük olduğu görülmekle birlikte logistic kernel için de görülen değerlerin düşük olduğu söylenebilir.

Denk grup desenlerinde fonksiyonel düşünme beceri puanlarının *AxB* testleri için eşitlenme durumu ortalama, doğrusal, eşit yüzdelikli eşitleme ve kernel eşitleme ile incelendiğinde eşitleme uygulamalarından elde edilen bulguların kabul edilebilir bilgiler sunduğu görülmektedir. *AxB* test eşitleme uygulamaları benzer biçimde *AxC* ve *BxC* test eşitlemeleri için de uygulanmıştır. Bu uygulamalardan elde edilen bulgular Tablo 44'de aşağıda sunulmuştur.

Tablo 44

DG Deseninde FD Becerisinin Testlere Göre Eşitlenme Durumları

<i>Desen</i>	<i>Eşitleme Türü</i>	<i>Eşitlenen Testler</i>		
		<i>AxB</i>	<i>AxC</i>	<i>BxC</i>
<i>Denk Gruplar</i>	Ortalama	√	-	-
	Doğrusal	√	-	-
	Eşit Yüzdelikli	√	-	-
	Kernel	√	-	-

Tablo 44'de paylaşılan bulgulara göre denk gruplar deseninde fonksiyonel düşünme becerisi puanlarında *AxB* eşitleme uygulamalarında ortalama, doğrusal, eşit yüzdelikli ve kernel eşitlemelerin uygun sonuçlar verdiği ancak *AxC* ve *BxC* eşitleme uygulamalarında ise bu eşitlemelerin uygun sonuçlar vermediği belirtilmektedir.

FD Becerisinin Denk Olmayan Gruplarda Ortak Madde (DOG-OM) Desenine Göre Eşitleme İncelemeleri

DOG-OM deseninde FD becerileri için eşitleme incelemeleri çerçevesinde doğrusal, eşit yüzdelikli ve kernel eşitleme uygulamaları yapılmıştır. Aşağıda AxB testine ait eşitlemeler verilmiş olup, AxC ve BxC eşitlemelerine ait bulgular desen ile ilgili genel tabloda paylaşılmıştır.

FD Becerisinin Test A ve Test B için DOG-OM Deseninde Eşitleme Uygulamaları

FD becerisinin DOG-OM deseninde AxB testlerinde uygulanan doğrusal, eşit yüzdelikli ve kernel eşitleme uygulamaları çerçevesinde farklı yöntemler kullanılarak ve bu yöntemlerden elde edilen eşitlenmiş puanlar, bu puanların ham puanlardan farkı, eşitlemeye ait standart hatalar, yanlılık, RMSE ve eşitlemenin uygunluğu ile ilgili bilgi sunacak diğer bilgiler paylaşılmıştır.

i. Doğrusal Eşitleme (AxB / DOG-OM / Doğrusal)

DOG-OM deseninde doğrusal eşitleme kapsamında FD becerisine ait puanlar Tucker, Chain, Levine gözlenen puan (Levine Observed Score, *Levine-OS*), Levine gerçek puan (Levine True Score, *Levine-TS*) ve Braun-Holland (*BH*) yöntemleri ile eşitlemeler yapılmıştır. Elde edilen eşitlenmiş puanlar bir arada Tablo 45'te paylaşılmıştır.

Tablo 45

DOG-OM Deseni FD Becerisi Doğrusal Eşitleme Puanları

Ham Puanlar	Tucker	Chain	LevineOS	LevineTS	BH
0	-0,11	0,33	3,37	6,74	0,32
1	0,89	1,28	4,02	7,04	1,34
2	1,88	2,24	4,68	7,35	2,36
3	2,87	3,19	5,34	7,65	3,38
4	3,87	4,15	5,99	7,95	4,4
5	4,86	5,11	6,65	8,26	5,43
6	5,85	6,06	7,3	8,56	6,45
7	6,85	7,02	7,96	8,86	7,47

Tablo 45 (Devamı)

8	7,84	7,97	8,62	9,17	8,49
9	8,84	8,93	9,27	9,47	9,51
10	9,83	9,88	9,93	9,77	10,54
11	10,82	10,84	10,59	10,08	11,56
12	11,82	11,79	11,24	10,38	12,58
13	12,81	12,75	11,9	10,68	13,6
14	13,81	13,71	12,55	10,99	14,62
15	14,8	14,66	13,21	11,29	15,65
16	15,79	15,62	13,87	11,59	16,67
17	16,79	16,57	14,52	11,9	17,69
18	17,78	17,53	15,18	12,2	18,71
19	18,77	18,48	15,84	12,5	19,73
20	19,77	19,44	16,49	12,81	20,76
21	20,76	20,39	17,15	13,11	21,78
22	21,76	21,35	17,8	13,41	22,8
23	22,75	22,31	18,46	13,72	23,82
24	23,74	23,26	19,12	14,02	24,84
25	24,74	24,22	19,77	14,32	25,87

Tablo 45'te eşitlenen puanlar incelendiğinde *Tucker*, *Chain* ve *Braun/Holland* yöntemlerinden elde edilen eşitleme puanlarının ham puanlara yakın değerlerde eşitlendiği, ham puanlar ile eşitlenen puanları arası farkın mutlak değer olarak 0,1 ile 0,9 arasında değiştiği görülmektedir. Ancak *Levine gözlenen puan* ve *Levine gerçek puan* yöntemleri ile yapılan eşitlemelerde eşitlenen puanların ham puanlara göre genel olarak yüksek fark ile eşitlendiği ve özellikle uç noktalarda ham puanlara göre yüksek farka sahip olduğu görülmektedir. *Levine gözlenen puan* 0 ham puana karşılık olarak 3,37 ile *Levine gerçek puan* 6,74 ile başlangıç yaparak 25 ham puana karşılık 19,77 ve 14,32 puan ile tamamlamıştır. Burada gözlenen durum bu iki yönteme ait eşitlenen puanların kullanıma uygun olmadığıdır. Bunun yanında eşitleme için uygun görünen diğer yöntemlerden elde edilen eşitlenen puanların hata değerleri açısından incelemeleri de grafiklerde paylaşılmıştır.

ii. Eşit Yüzdellikli Eşitleme (AxB / DOG-OM / Eşit Yüzdellikli)

DOG-OM deseninde FD becerisinin eşit yüzdellikli eşitleme inceleme çalışmalarında frekans tahmini (*frequency estimation, FE*) ve zincir eşitleme (*chain equating, CE*) yöntemleri kullanılmıştır. AxB test eşitlemesinde bu yöntemlerden elde edilen eşitlenmiş puanlar Tablo 47'da ve bu puanların ham puanlardan farklarının görselleştirildiği Şekil 33 aşağıda paylaşılmıştır.

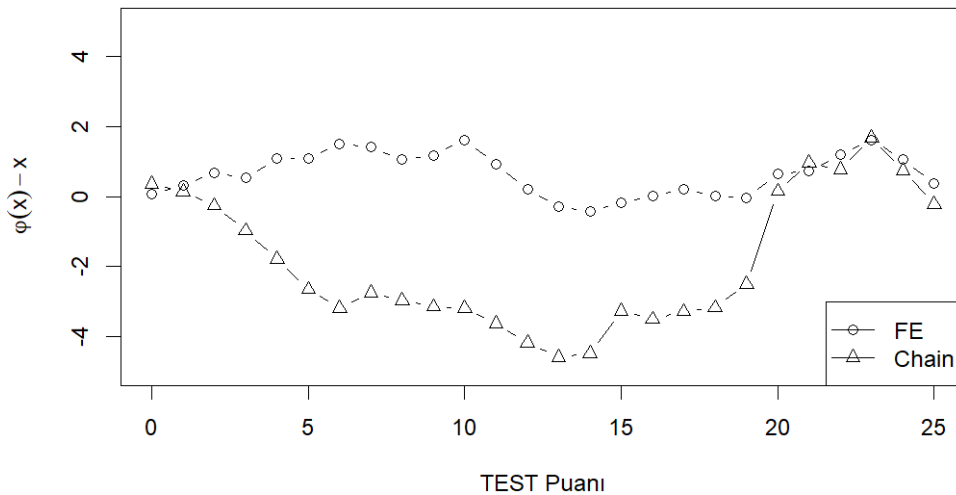
Tablo 47

DOG-OM Deseni FD Becerisi EYE Puanları

Ham Puan	freq	chain	Ham Puan	freq	chain
0	0,07	0,36	13	12,72	8,40
1	1,33	1,14	14	13,59	9,51
2	2,67	1,74	15	14,82	11,72
3	3,54	2,02	16	16,02	12,49
4	5,08	2,20	17	17,20	13,71
5	6,09	2,35	18	18,02	14,82
6	7,51	2,81	19	18,95	16,50
7	8,41	4,24	20	20,66	20,15
8	9,06	5,02	21	21,73	21,96
9	10,18	5,85	22	23,20	22,77
10	11,61	6,80	23	24,63	24,68
11	11,92	7,36	24	25,05	24,74
12	12,22	7,82	25	25,37	24,78

Şekil 33

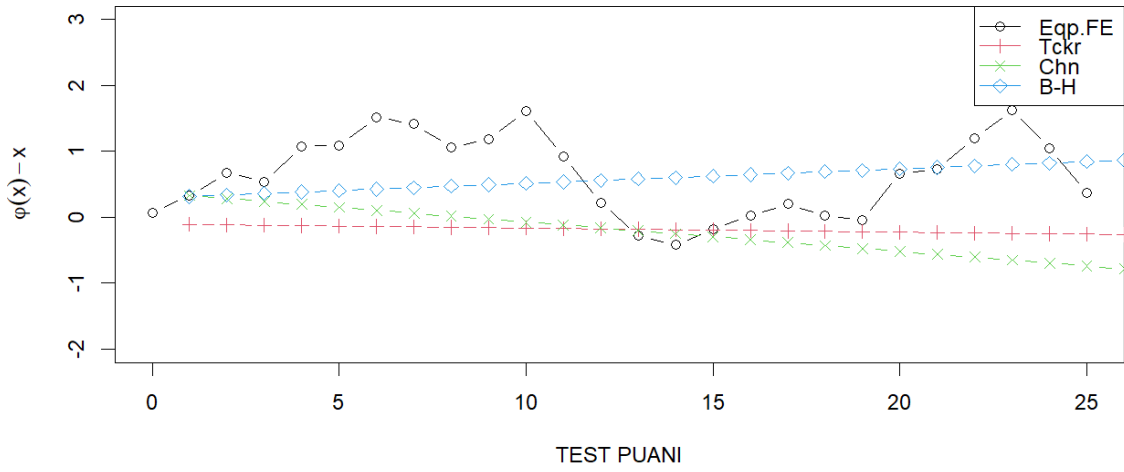
DOG-OM Deseni FD Becerisi FE ve Chain Eşitleme Puan Farkları



Tablo 47 ve Şekil 33'te verilenlere göre FE eşitleme ile elde edilen eşitlenmiş puanlar ham puanlardan -0,41 ile 1,63 arasında değişen farklara sahipken CE eşitleme ile elde edilen eşitlenmiş puanlar ise ham puanlardan -4,60 ile 1,68 arasında değişen farklara sahiptir. Eşit yüzdelikli eşitlemede FE eşitlemenin CE eşitlemeye göre daha uygun eşitlenmiş puanlar sunduğu görülmektedir. Şekil 34'da ayrıca doğrusal eşitleme ve eşit yüzdelikli eşitlemede daha uygun bulgular sunan yöntemlerden eşitlenmiş puan ile ham puan farklarını gösteren grafik paylaşılmıştır.

Şekil 34

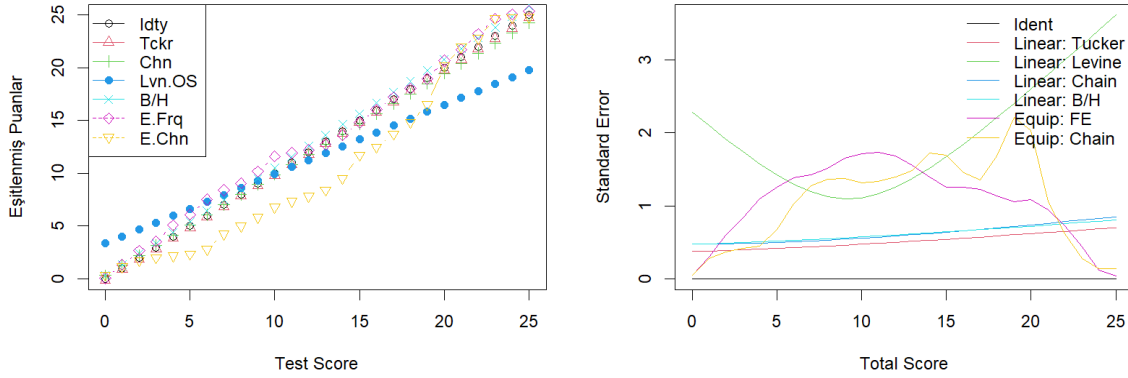
DOG-OM Deseni FD Becerisi Eşitlenmiş Puan Farkları



Şekil 34'te eşitlenmiş puanlar ile ham puanların farkının -0,5 ile 1,6 civarı değerlerde değiştiği görülmektedir. Eşit yüzdelikli eşitlemede bu farkın doğrusal eşitlemelere göre daha fazla olduğu görülse de genel olarak farklar düşüktür. Eşit yüzdelikli eşitlenmiş puanların 11 ile 21 puanlar arasında doğrusal eşitlenmiş puanlara yakın olduğu ancak diğer aralıklarda daha çok farka sahip olduğu görülmektedir. Şekil 35'te ise bu eşitlenmiş puanlar birlikte grafik üzerinde verilerek bunlara ait standart hatalar yine görsel üzerinde paylaşılmıştır.

Şekil 35

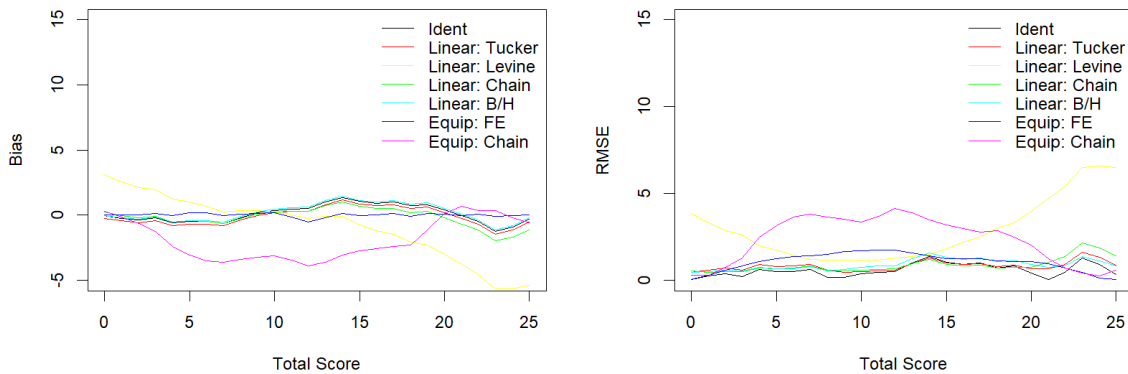
DOG-OM Deseni FD Becerisi Eşitlenmiş Puanlar ve Standart Hataları



Şekil 35'te doğrusal ve eşit yüzdelli eşitleme ile elde edilen eşitlenmiş puanların genelinde gözlenen durum Levine gözlenen puan ve CE eşit yüzdelli eşitleme puanlarının diğer yöntemlerle elde edilen puanlara göre ham puanlardan dikkat çeken biçimde sapma göstermesi ve diğer doğrusal eşitlemeler ile FE eşit yüzdelli eşitlemenin ham puanlar civarında eşitlenmiş puan sunmasıdır. Bu puanlara ait standart hataların ise bu durumu yansıtarak Levine doğrusal eşitleme ile CE eşit yüzdelli eşitleme haricindeki diğer eşitlemelerin uygun aralıklarda olduğu görülmektedir. Doğrusal eşitlemelerde Tucker, Chain ve B/H eşitlemelerin 0,5 ile 0,9 arasında değişen standart hatalara sahip olduğu görülürken FE eşit yüzdelli eşitlemede ise 0 ile 1,8 arasında değişen standart hatalar görülmektedir. Genel olarak bu hataların düşük düzeyde olduğu söylenebilir. Aşağıda Şekil 36 ise bu eşitlemelere ait yanlılık ve RMSE değerlerini sunmaktadır.

Şekil 36

DOG-OM Deseni FD Becerisi Eşitlenmiş Puan Yanlılık ve RMSE Değerleri



Yanlılık değerleri açısından Levine doğrusal eşitleme ve CE eşit yüzdelikli eşitlemelerin -5 ile 4 dolaylarında değiştiği, diğer eşitlemelerin de birbirine yakın değerler alarak -1 ile 1 arasında değiştiği, özellikle FE eşit yüzdelikli eşitlemenin 0 civarında değerler aldığı görülmektedir. RMSE değerler açısından Tucker, Chain ve B/H doğrusal eşitlemeler ile FE eşit yüzdelikli eşitlemelerin daha uygun bulgular sunduğu görülürken Levine doğrusal eşitleme ve FE eşit yüzdelikli eşitlemelerin bu değerlerin çok üzerinde değerler aldığı görülmektedir. Ayrıca, DOG-OM deseni çerçevesinde uygulanmış olan doğrusal ve eşit yüzdelikli eşitlemelere ait bootstrap standart hata, yanlılık, RMSE ortalama değerleri ile bunların ağırlıklandırılmış değerlerine ait ortalamaları Tablo 48'de verilmiştir.

Tablo 48

DOG-OM FD Becerisi Eşitlenmiş Puan Bootstrap Hata, Yanlılık ve RMSE Değerleri

	<i>SE</i>	<i>SE_w</i>	<i>BIAS</i>	<i>BIAS_w</i>	<i>RMSE</i>	<i>RMSE_w</i>
identity	0,00	0,00	0,68	0,60	0,68	0,60
Tucker	0,53	0,50	0,69	0,65	0,86	0,82
Levine	2,09	2,29	2,62	3,09	3,35	3,84
Chain	0,63	0,61	0,74	0,74	0,97	0,96
B/H	0,63	0,60	0,71	0,60	0,94	0,84
EF	1,17	0,87	0,15	0,10	1,18	0,87
CE	1,21	0,94	2,41	1,71	2,69	1,95

Tablo 48'deki değerler incelendiğinde doğrusal eşitleme yöntemlerinin genel olarak yakın hata değerlerine sahip olduğu görülmekle birlikte standart hata, yanlılık ve RMSE değerlerinde en düşük olan yöntem Tucker doğrusal eşitlemedir. Eşit yüzdelikli eşitlemelerde FE yöntemi CE yöntemine göre daha düşük ve uygun değerlere sahiptir. Tabloya genel olarak bakıldığında Levine doğrusal eşitleme ve CE eşit yüzdelikli eşitlemeler dışındaki eşitlemelerin düşük hatalara sahip olduğu belirtilebilir.

iii. Kernel Eşitleme (AxB / DOG-OM / Kernel)

DOG-OM deseninde FD beceri puanlarının kernel ile eşitleme incelemesinde Son Tabakalama (*poststratification equating, PSE*) ve Zincir Eşitleme (*chain equating, CE*) yöntemleri kullanılmıştır. PSE ve CE kernel eşit yüzdelikli ve doğrusal eşitlemelere ait eşitlenen puanlar ve standart hataları Tablo 49 ve Tablo 50'de paylaşılmıştır.

Tablo 49

DOG-OM Deseni FD Becerisi PSE EYE ve Doğrusal Eşitleme Puan ve Hataları

Ham Puan	EYE		Doğrusal	
	Eşitlenmiş Puan	SE	Eşitlenmiş Puan	SE
0	-0,02	0,03	-0,202	0,340
1	0,93	0,11	0,799	0,346
2	1,87	0,20	1,800	0,353
3	2,81	0,30	2,801	0,361
4	3,74	0,42	3,801	0,370
5	4,67	0,56	4,802	0,380
6	5,59	0,70	5,803	0,390
7	6,52	0,85	6,804	0,401
8	7,46	1,00	7,805	0,413
9	8,40	1,14	8,805	0,425
10	9,36	1,28	9,806	0,438
11	10,34	1,40	10,807	0,451
12	11,34	1,51	11,808	0,465
13	12,36	1,59	12,809	0,479
14	13,40	1,64	13,809	0,494
15	14,46	1,63	14,810	0,508
16	15,54	1,56	15,811	0,523
17	16,62	1,42	16,812	0,539
18	17,70	1,23	17,812	0,555
19	18,78	1,03	18,813	0,570
20	19,85	0,83	19,814	0,586
21	20,90	0,65	20,815	0,603
22	21,94	0,48	21,816	0,619
23	22,97	0,32	22,816	0,636
24	23,99	0,18	23,817	0,653
25	25,00	0,06	24,818	0,670

Tablo 49'da eşitlenen puanlar PSE kernel eşit yüzdelikli eşitlemede eşitlenen puanların ham puan değerlerinden düşük olduğu ve 25 ham puan değerine eşit bir biçimde tamamladığı görülmektedir. Eşitlenmiş puanlar ile ham puanlar arasındaki fark eşit yüzdelikli eşitlemede -0,66 ile 0 değerleri arasında değişmektedir. Eşitlenmiş puanlara ait standart hatalar ise 0,03 ile 1,64 arasında değişmektedir. Eşitlenmiş puan aralığına göre uçlarda daha düşük hata değerleri gözlenirken ortalarda hata değerleri uçlara göre biraz daha yüksek durumdadır. PSE kernel doğrusal eşitlemede ise eşitlenmiş puanlar ham puan değerlerinden daha düşüktür. Eşitlenmiş puanlar ile ham puanlar arasındaki fark -0,20 ile -0,18 değerleri arasında değişmektedir ve bu eşitlenmiş puanlara ait standart hata değerleri de 0,34 ile 0,67 değerleri arasındadır. Genel olarak PSE kernel eşit yüzdelikli ve doğrusal eşitlemelerde eşitlenen puanlar ve standart hatalar bakımından uygun bir eşitleme görülmektedir.

Tablo 50 ise FD beceri puanlarının kernel CE eşit yüzdelikli ve doğrusal eşitlemelerden elde edilen eşitlenmiş puan ve hataları göstermektedir.

Tablo 50

DOG-OM Deseni FD Becerisi Kernel CE EYE ve Doğrusal Eşitleme Puan ve Hataları

Ham Puan	EYE		Doğrusal	
	Eşitlenmiş Puan	SE	Ham Puan	Eşitlenmiş Puan
0	0,04	0,05	0,175	0,462
1	1,08	0,15	1,147	0,467
2	2,08	0,26	2,118	0,472
3	3,05	0,38	3,090	0,478
4	3,99	0,51	4,062	0,485
5	4,92	0,65	5,033	0,493
6	5,84	0,82	6,005	0,501
7	6,74	1,00	6,977	0,510
8	7,63	1,19	7,949	0,519
9	8,52	1,37	8,920	0,529
10	9,41	1,50	9,892	0,540
11	10,31	1,60	10,864	0,551
12	11,23	1,67	11,836	0,562
13	12,16	1,71	12,807	0,573
14	13,12	1,74	13,779	0,586
15	14,11	1,72	14,751	0,598

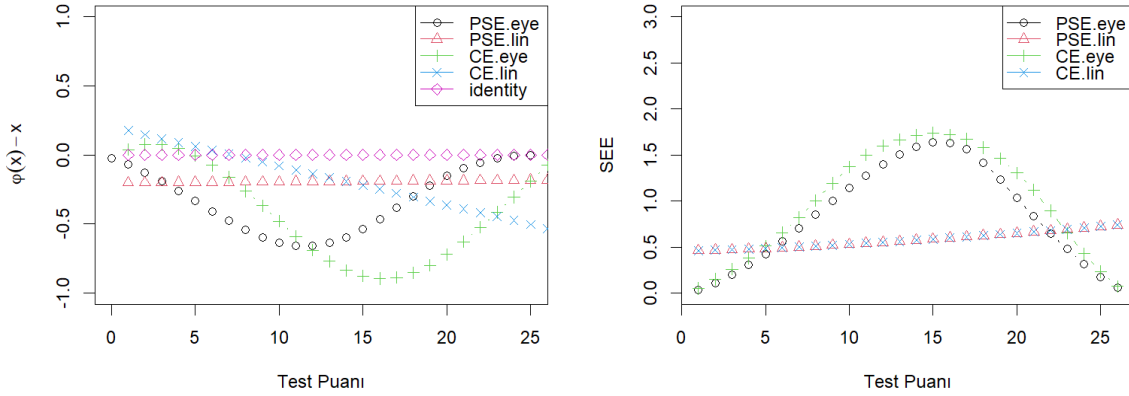
Tablo 47 (Devamı)

16	15,11	1,67	15,722	0,611
17	16,15	1,58	16,694	0,624
18	17,20	1,46	17,666	0,637
19	18,28	1,30	18,638	0,651
20	19,37	1,12	19,609	0,665
21	20,48	0,89	20,581	0,679
22	21,58	0,65	21,553	0,693
23	22,69	0,43	22,525	0,708
24	23,81	0,23	23,496	0,723
25	24,93	0,07	24,468	0,738

Tablo 50'de eşitlenen puanlar kernel CE eşit yüzdelikli eşitlemede eşitlenen puanların 3 puana kadar ham puanlardan yüksek fakat genel olarak ham puan değerlerinden düşük olduğu görülmektedir. Eşitlenmiş puanlar ile ham puanlar arasındaki fark eşit yüzdelikli eşitlemede -0,89 ile 0,08 değerleri arasında değişmektedir. Eşitlenmiş puanlara ait standart hatalar ise 0,05 ile 1,74 arasında değerler almaktadır. Eşitlenmiş puanlarda uçlarda daha düşük hata değerleri gözlenirken bu durum ortalarda tersidir. Kernel CE doğrusal eşitlemede eşitlenmiş puanlar ile ham puanlar arasındaki fark -0,53 ile 0,18 değerleri arasında değişmektedir ve bu eşitlenmiş puanlara ait standart hata değerleri ise 0,46 ile 0,74 arasında farklı değerler almaktadır. Genel olarak kernel CE eşit yüzdelikli ve doğrusal eşitlemelerde eşitlenen puanlar ve standart hatalar bakımından uygun bir eşitleme görülmektedir. Tablo 49 ve 50 için sunulan açıklamalar ayrıca aşağıda Şekil 37'de görselleştirilmiştir.

Şekil 37

DOG-OM Deseni FD Becerisi Kernel PSE ve CE Eşitleme Puan Fark ve Standart Hataları

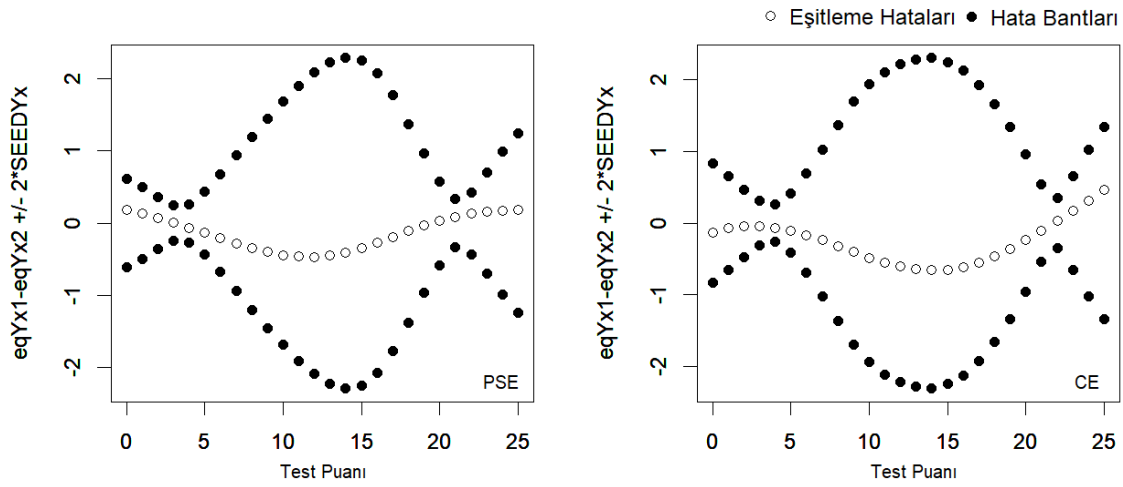


Şekil 37'de kernel PSE ve CE eşitlemelerinde eşitlenen puanların ham puanlardan farkının genelde düşük değerlerde olduğu en fazla -0,9 civarına ulaştığı gözlenmektedir. Kernel PSE doğrusal eşitlemede elde edilen eşitlenen puanların ham puanlardan farkı sabit bir doğruyu gösterirken CE doğrusal eşitlemenin azalan doğru grafiğini göstermektedir. PSE ve CE eşit yüzdelikli eşitlemelere ait eşitlenen puan ile ham puan farkını gösteren eğrilerdeki fark da grafiklerden gözlenmektedir, PSE eşit yüzdelikli eşitlemenin daha düşük farka sahiptir, ancak hatalar açısından bakıldığında her iki yöntemde doğrusal ve eşit yüzdelikli eşitlemeler çok yakın hata değerlerine sahiptir.

Kernel PSE ve CE eşitlemelerinde eşit yüzdelikli eşitleme ile doğrusal eşitleme arasındaki eşitleme puanları farkına ait standart hata (*SEED*) değerleri Şekil 38'te paylaşılmıştır.

Şekil 38

DOG-OM Deseni GA Becerisi Kernel PSE ve CE SEED Değerleri



Şekil 38 incelendiğinde kernel PSE ve CE eşitlemelerinde eşit yüzdelikli eşitleme puanları ile doğrusal eşitleme puanları arasındaki farkların standart hata değerlerinin hata bantları içerisinde olduğu görülmektedir. Bu grafiklere göre kernel eşitlemelerin uygun olduğu belirtilebilir. Ayrıca, Tablo 51 kernel PSE ve CE eşitlemelere ait yüzde göreceli hata (PRE) değerlerini sunmaktadır.

Tablo 51

DOG-OM Deseni FD Becerisi Kernel PSE ve CE Yüzde Göreceli Hata Değerleri

	PSE	CE	
		B to OM	OM to A
1	0,030	-0,067	0,053
2	0,029	-0,177	0,066
3	0,033	-0,139	0,039
4	0,037	-0,203	0,049
5	0,040	-0,355	0,078
6	0,040	-0,589	0,124
7	0,039	-0,902	0,186
8	0,036	-1,292	0,263
9	0,033	-1,758	0,356
10	0,029	-2,296	0,465

Tablo 51’de ilk 10 moment değeri için elde edilen yüzde göreceli değerleri (*PRE*) kernel PSE eşit yüzdellikli eşitleme için 0,03 ile 0,04 değerleri arasında değişmektedir. Değerler sıfıra çok yakın ve 10. değer ise 0,029’dur. Kernel CE kernel eşit yüzdellikli eşitlemenin ise PSE eşitlemeye göre mutlak değer olarak daha yüksek değerler aldığı görülmektedir ancak CE eşitlemesinin ikinci aşamasında ortak maddeden Test A’ya yapılan eşitlemede *PRE* değerlerinin 0 civarında düşük değerlerde olduğu görülmektedir.

DOG-OM desenlerinde fonksiyonel düşünme becerisi puanlarının doğrusal, eşit yüzdellikli ve kernel eşitlemeleri AxB testlerinde incelendiğinde eşitlemelerin uygunluğu doğrultusunda bulgulara ulaşılmıştır. AxB test eşitleme uygulamaları benzer biçimde AxC ve BxC test eşitlemeleri için de uygulanmıştır. Bu uygulamalardan elde edilen bulgular Tablo 52’de sunulmuştur. Tablo 52’de verilenlere göre denk olmayan gruplarda ortak madde deseninde uygulanan üç farklı eşitlemeden sadece AxB eşitlemesinde doğrusal, eşit yüzdellikli ve kernel eşitlemeler yapılabilmiş, diğer eşitlemelerde ise uygun bulgulara ulaşılmamıştır.

Tablo 52

DOG-OM Deseninde FD Becerisinin Testlere Göre Eşitleme Durumları

<i>Desen</i>	<i>Eşitleme Türü</i>	<i>Eşitlenen Testler</i>		
		<i>AxB</i>	<i>AxC</i>	<i>BxC</i>
<i>Denk Olmayan Gruplarda Ortak Madde</i>	Doğrusal	Tucker, Chain, Braun/Holland	-	-
	Eşit Yüzdellikli	FE	-	-
	Kernel	PSE, CE	-	-

FD Becerisinin Denk Olmayan Gruplarda Ortak Değişken (DOG-OD) Desenine Göre Eşitleme İncelemeleri

DOG-OD deseninde okulların kabul koşulu, okul türleri ve cinsiyet değişkenleri FD becerileri için eşitleme incelemelerinde kullanılmıştır. Bu desen çerçevesinde kernel eşitleme uygulamaları yapılmıştır. Aşağıda AxB testine ait okulların kabul koşulu-okul türleri ile okulların kabul koşulu-cinsiyet değişkenleri çerçevesinde eşitlemeler verilmiş olup, AxC ve BxC eşitlemelerine ait bulgular desen ile ilgili genel tabloda paylaşılmıştır.

FD Becerisinin Test A ve Test B için DOG-OD Deseninde Eşitleme Uygulamaları

i. Okulların Kabul Koşulu ve Okul Türleri Değişkenine Göre

DOG-OD deseninde FD beceri puanı eşitlemelerinde kernel eşitleme uygulaması Son Tabakalama (poststratification equating, PSE) yöntemi kullanılmıştır. Tablo 53'te kernel PSE eşit yüzdelikli ve doğrusal eşitlemeye ait eşitlenen puanlar ve standart hataları paylaşılmıştır. Eşitlenen bu puanlar ile ham puanların farkı ve eşitlenen puanlara ait standart hatalar ise Şekil 39'da görsel ile sunulmuştur.

Tablo 53

DOG-OD Deseni FD Becerisi Kernel EYE ve Doğrusal Eşitleme Puan ve Hataları

Ham Puan	EYE		Doğrusal	
	Eşitlenmiş Puan	SE	Ham Puan	Eşitlenmiş Puan
0	0,07	0,04	-0,003	0,335
1	1,21	0,15	1,001	0,339
2	2,30	0,30	2,005	0,343
3	3,29	0,45	3,009	0,349
4	4,19	0,59	4,014	0,355
5	5,02	0,73	5,018	0,362
6	5,79	0,93	6,022	0,369
7	6,53	1,13	7,026	0,377
8	7,26	1,19	8,030	0,385
9	8,04	1,25	9,034	0,395
10	8,87	1,37	10,038	0,404
11	9,76	1,47	11,043	0,414
12	10,74	1,50	12,047	0,425

Tablo 53 (Devamı)

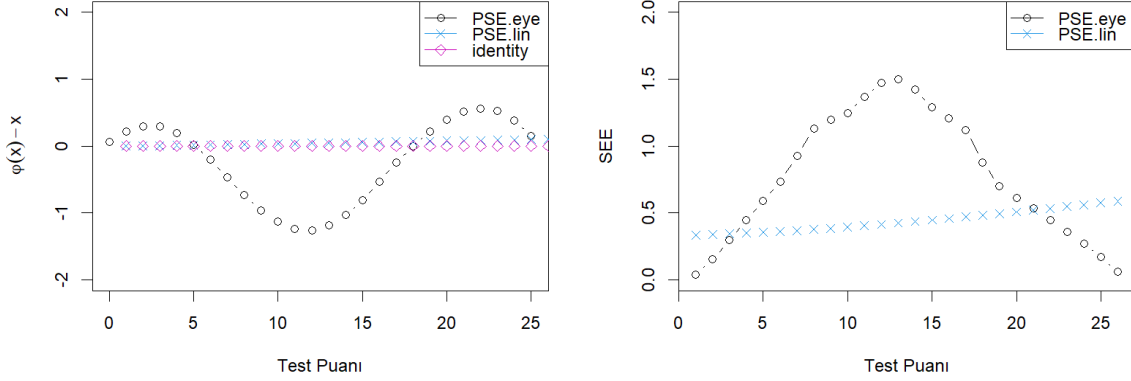
13	11,81	1,42	13,051	0,436
14	12,97	1,29	14,055	0,447
15	14,19	1,21	15,059	0,458
16	15,47	1,12	16,063	0,470
17	16,75	0,88	17,068	0,482
18	18,00	0,70	18,072	0,495
19	19,22	0,61	19,076	0,508
20	20,40	0,54	20,080	0,521
21	21,52	0,45	21,084	0,534
22	22,56	0,36	22,088	0,547
23	23,52	0,27	23,093	0,561
24	24,38	0,17	24,097	0,575
25	25,15	0,06	25,101	0,589

Tablo 53'deki kernel PSE eşit yüzdelliği eşitleme ile eşitlenen puanlar uçlarda ham puanlardan yüksek ve orta puanlarda ham puanlardan düşüktür. Kernel PSE doğrusal eşitlemede ise eşitlenen puanlar ham puanlardan yüksek ve aradaki fark 0 civarındadır. Ancak hem eşit yüzdelliği eşitlemede hem de doğrusal eşitlemede eşitlenen puanların üst puan limitini 0,15 ve 0,10 puan kadar aştıkları görülmektedir. Eşitleme hataları açısından eşit yüzdelliği eşitlemenin uç puan değerlerinde daha düşük olduğu görülmekte ve genel olarak bu eşitleme hataları 0,04 ile 1,50 arasında farklı değerler almaktadır. PSE doğrusal eşitlemede ise eşitlenen puanlar ile ham puan değerlerinin birine çok yakın değerlerde olduğu görülmektedir. Bunun yanında eşitleme hatalarının düşük puandan yüksek puana doğru artış gösterdiği görülmekte ve bu hatalar 0 dolaylarındadır. Eşitlenen puanların ham puanlardan farkı ve bu puanların standart hataları aşağıda Şekil 39 ile görselleştirilmiştir.

Şekil 39

DOG-OD Deseni FD Becerisi Kernel PSE EYE ve Doğrusal Eşitleme Puan Fark ve

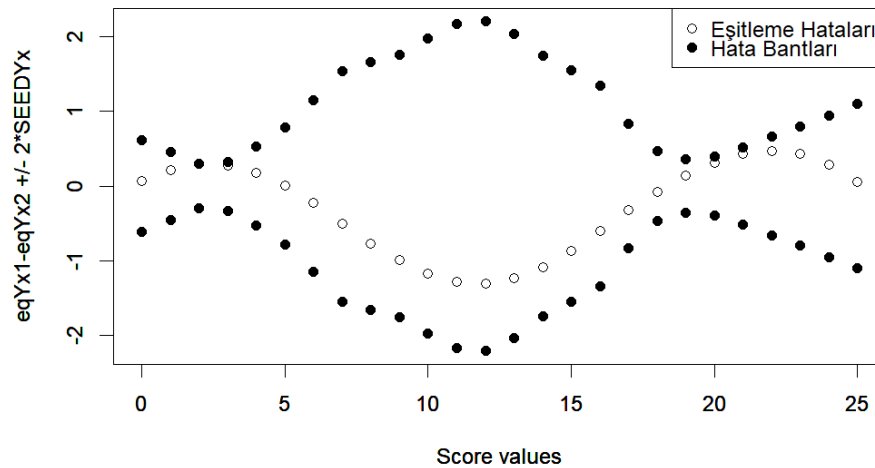
Standart Hataları



Şekil 39'te görüldüğü üzere kernel PSE eşit yüzdelikli eşitlemeye ait eşitlenmiş puan ve ham puan farkı -1,26 ile 0,56 arasında değişmektedir. PSE doğrusal eşitlemeye ait eşitlenmiş puan ile ham puan farkı ise 0 ile 0,10 arasındadır. Doğrusal eşitlemeye ait standart hatalar daha düşük olmakla birlikte eşit yüzdelikli eşitlemeye ait standart hatalar ise 1,50 civarına kadar çıkmaktadır. Genel olarak bakıldığında ise eşitlemelere ait standart hataların düşük olduğu görülmektedir. Burada bahsedilenler yanında eşit yüzdelikli eşitleme ile doğrusal eşitleme arasındaki eşitleme puanları farkına ait standart hata (*SEED*) değerleri Şekil 40'ta paylaşılmıştır.

Şekil 40

DOG-OD Deseni FD Becerisi Kernel PSE EYE ve Doğrusal Eşitleme SEED Değerleri



Şekil 40 incelendiğinde hata bantlarının -2 ile 2 seviyelerinde olduğu görülmektedir. Kernel PSE eşitlemelerde eşit yüzdeliği eşitleme puanları ile doğrusal eşitleme puanları arasındaki farkların standart hata değerlerinin genel olarak bu hata bantları içerisinde olduğu görülmektedir. Buna göre, kernel eşitlemenin uygun görüldüğü belirtilebilir. Ayrıca, PSE kernel eşit yüzdeliği eşitlemeye ait yüzde göreceli hata (PRE) değerleri Tablo 54'te paylaşılmıştır.

Tablo 54

DOG-OD Deseni FD Becerisi Kernel PSE EYE Yüzde Göreceli Hata Değerleri

	PRE
1	-0,044
2	0,012
3	0,022
4	0,021
5	0,022
6	0,027
7	0,039
8	0,057
9	0,082
10	0,114

Tablo 54'te ilk 10 moment değeri için elde edilen yüzde göreceli hata değerleri (PRE) kernel PSE eşit yüzdeliği eşitleme için 0,01 ile 0,11 değerleri arasında değişmektedir. Tabloda PRE değerlerinin sıfıra çok yakın değerlerde olduğu görülmektedir ve kernel eşitlemenin uygunluğu açısından destek veriler olduğu belirtilebilir.

Okulların kabul koşulu ve okul türleri değişkenlerine göre yapılan kernel PSE eşit yüzdeliği eşitlemede eşitlenmiş puanlar, eşitleme hataları, eşitlenmiş puan ve ham puan farkları, SEED ve PRE değerlerinden elde edilen bulgulara göre eşitlemenin uygun olduğu söylenebilir.

ii. Okulların Kabul Koşulu ve Cinsiyete Göre

DOG-OD deseninde okulların kabul koşulu ve cinsiyet değişkenlerine bağlı olarak FD beceri puanı eşitlemelerinde kernel eşitleme uygulaması Son Tabakalama (poststratification equating, PSE) yöntemi kullanılmıştır. Tablo 55'de kernel PSE eşit yüzdelikli ve doğrusal eşitleme sonucu elde edilen eşitlenen puanlar ve standart hataları verilmiştir.

Tablo 55

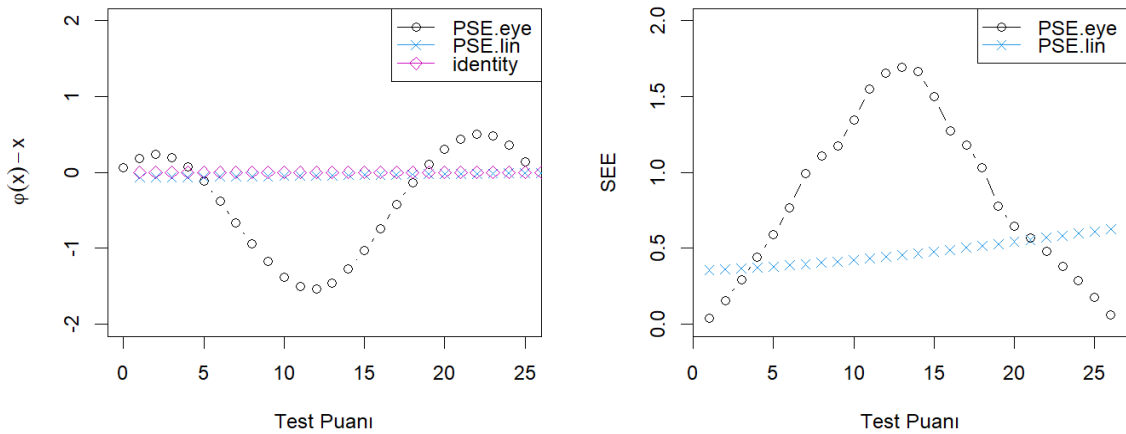
DOG-OD Deseni FD Becerisi Kernel EYE ve Doğrusal Eşitleme Puan ve Hataları

Ham Puan	EYE		Doğrusal	
	Eşitlenmiş Puan	SE	Ham Puan	Eşitlenmiş Puan
0	0,06	0,04	-0,071	0,357
1	1,18	0,16	0,931	0,361
2	2,24	0,29	1,934	0,367
3	3,20	0,44	2,937	0,373
4	4,08	0,59	3,939	0,379
5	4,88	0,77	4,942	0,387
6	5,62	0,99	5,944	0,395
7	6,33	1,11	6,947	0,404
8	7,06	1,17	7,950	0,413
9	7,82	1,35	8,952	0,423
10	8,62	1,55	9,955	0,433
11	9,49	1,65	10,957	0,443
12	10,47	1,69	11,960	0,455
13	11,55	1,67	12,962	0,466
14	12,72	1,50	13,965	0,478
15	13,97	1,28	14,968	0,490
16	15,25	1,18	15,970	0,503
17	16,58	1,03	16,973	0,515
18	17,86	0,78	17,975	0,528
19	19,10	0,65	18,978	0,542
20	20,30	0,57	19,980	0,555
21	21,44	0,48	20,983	0,569
22	22,51	0,38	21,986	0,583
23	23,48	0,29	22,988	0,597
24	24,36	0,18	23,991	0,611
25	25,14	0,06	24,993	0,626

Tablo 55’de kernel PSE eşit yüzdelikli eşitleme ile eşitlenen puanlar uç noktalarda ham puanlardan yüksek değerde iken orta noktalarda ham puanlardan düşük düzeydedir. Eşitlenen puan ile ham puanlar farkı -1,51 ile 0,51 arasında değişmektedir. Bu fark görsel olarak Şekil 41’de aşağıda paylaşılmıştır. Ayrıca uygulanan kernel PSE doğrusal eşitlemede eşitlenen puanlar ham puanlardan düşük olup ham puanlara çok yakın değerdedir. Doğrusal eşitlenmiş puanların ham puanlardan farkı -0,071 ile -0,007 arasındadır ve bu fark oldukça düşüktür. Eşitlenmiş puanların standart hataları ise görselde de verildiği üzere eşit yüzdelikli için 0,04 ile 1,67 arasındayken doğrusal eşitlemede bu değerler 0,36 ile 0,63 arasındadır.

Şekil 41

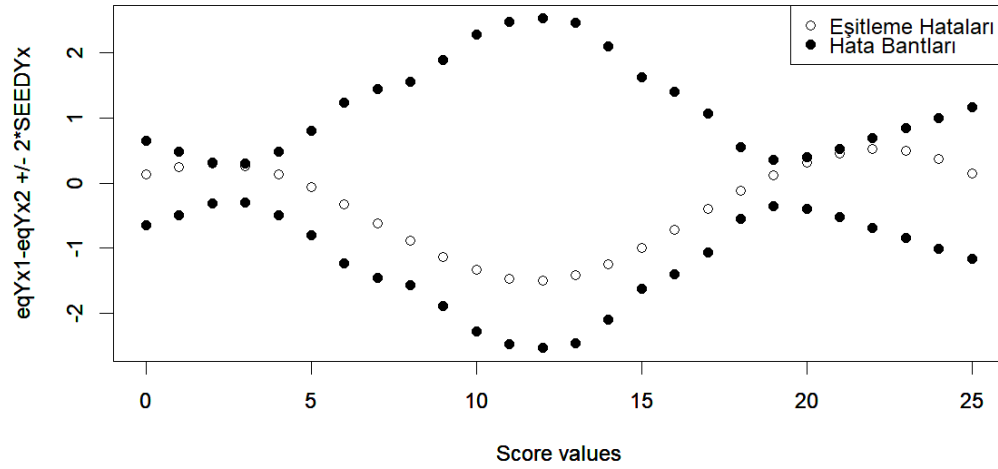
DOG-OD Deseni FD Becerisi Kernel PSE EYE ve Doğrusal Eşitleme Puan Fark ve Standart Hataları



Ayrıca eşit yüzdelikli eşitleme ile doğrusal eşitleme arasındaki eşitleme puanları farkına ait standart hata (*SEED*) değerleri Şekil 42’de paylaşılmıştır.

Şekil 42

DOG-OD Deseni FD Becerisi Kernel PSE EYE ve Doğrusal Eşitleme SEED Değerleri



Şekil 42'de hata bantları -2 ile 2 seviyelerinde olduğu görülmektedir. PSE kernel eşitlemelerde eşit yüzdelikli eşitleme puanları ile doğrusal eşitleme puanları arasındaki farkların standart hata değerlerinin genel olarak bu hata bantları arasında olduğu görülmektedir. Bunun yanında, Tablo 56 kernel PSE eşit yüzdelikli eşitlemeye ait yüzde göreceli hata (PRE) değerlerini sunmaktadır.

Tablo 56

DOG-OD Deseni FD Becerisi Kernel PSE EYE Yüzde Göreceli Hata Değerleri

	PRE
1	-0,033
2	0,007
3	0,006
4	-0,004
5	-0,012
6	-0,016
7	-0,013
8	-0,003
9	0,013
10	0,037

Tablo 56'de ilk 10 moment değeri için elde edilen yüzde göreceli hata değerleri (PRE) kernel PSE eşit yüzdelikli eşitleme için -0,02 ile 0,04 değerleri arasında değişmektedir. PRE değerlerinin tabloda düşük değerlerde olduğu görülmektedir.

Okulların kabul koşulu ve cinsiyet değişkenlerine göre yapılan kernel PSE eşit yüzdelikli eşitlemede eşitlenmiş puanlar, eşitleme hataları, eşitlenmiş puan ve ham puan farkları, SEED ve PRE değerlerinden elde edilen verilere göre eşitlemenin uygun olduğu söylenebilir.

DOG-OD desenlerinde okul kabul koşulu, okul türü ve cinsiyet değişkenlerine bağlı olarak FD beceri puanlarının AxB testleri için kernel eşitleme incelemelerinin uygunluğu doğrultusunda bulgulara ulaşılmıştır. AxB test eşitleme uygulamaları benzer biçimde AxC ve BxC test eşitlemeleri için de uygulanmıştır. Bu uygulamalardan elde edilen bulgular Tablo 57'de sunulmuştur.

Tablo 57

DOG-OD Deseninde FD Becerisinin Testlere Göre Eşitleme Durumları

<i>Desen</i>	<i>Eşitleme Türü</i>	<i>Eşitlenen Testler</i>		
		<i>AxB</i>	<i>AxC</i>	<i>BxC</i>
<i>Denk Olmayan Gruplarda Ortak Değişken</i>	Kernel (değişkenler: okul kabul koşulu & okul türü	√	-	-
	Kernel (değişkenler: okul kabul koşulu & cinsiyet	√	-	-

Tablo 57'de görüldüğü göre denk olmayan gruplarda ortak değişken deseninde okul kabul koşulu-okul türü ile okul kabul koşulu-cinsiyet değişkenleri çerçevesinde uygulanan eşitlemelerden sadece AxB eşitlemesinde eşitlemeler yapılabilmiş, AxC ve BxC eşitlemelerinde uygun bulgulara ulaşılmamıştır.

Üçüncü Araştırma Sorusu Kapsamındaki Bulgular

Niceliksel Mantık Yürütme (NMY) Becerisinin DG, DOG-OM ve DOG-OD Desenlerine Göre Eşitleme İncelemeleri

Bu tez kapsamında yürütülen çalışmalar beceri puanları çerçevesinde AxB, AxC ve BxC testlerinin eşitleme incelemeleri olup buraya kadar GA ve FD beceri puanlarının AxB eşitleme örnek incelemeleri kapsamında tablo ve grafik detaylarıyla paylaşılmış olup AxB, AxC ve BxC eşitlemelerine dair bulgular bir arada genel tablo ile sunulmuştur. NMY beceri puanları da AxB eşitleme incelemeleri yapılmış olup inceleme örneğine dair bulgular detaylı tablo ve grafiklerle EK-A, EK-B ve EK-C'de verilmiştir. NMY beceri puanlarının AxB, AxC ve BxC eşitlemelerine ait bulgular ise desenlere göre tablolar halinde aşağıda paylaşılmıştır.

DG deseni çerçevesinde NMY becerisi eşitleme incelemelerine dair bulgular Tablo 58'te paylaşılmıştır. Tabloda görüldüğü üzere DG deseninde AxB ve AxC test eşitlemeleri yapılamamışken BxC test eşitlemelerinin klasik yöntemlerde ve kernel eşitleme yöntemlerinde eşitlemelerin gerçekleştiği belirtilmiştir.

Tablo 58

DG Deseninde NMY Becerisinin Testlere Göre Eşitlenme Durumları

<i>Eşitleme Desenler</i>	<i>Eşitleme Türü</i>	<i>Eşitlenen Testler</i>		
		<i>AxB</i>	<i>AxC</i>	<i>BxC</i>
<i>Denk Gruplar</i>	Ortalama	–	–	√
	Doğrusal	–	–	√
	Eşit Yüzdelikli	–	–	√
	Kernel	–	–	√

DOG-OM deseni çerçevesinde NMY becerisi eşitleme incelemelerine dair bulgular ise Tablo 59'da sunulmuştur. Tabloda görüldüğü üzere DOG-OM deseninde AxB ve AxC test eşitlemeleri yapılamamışken BxC test eşitlemelerinin klasik yöntemlerde ve kernel eşitleme yöntemlerinde eşitlemelerin gerçekleştiği belirtilmiştir.

Tablo 59

DOG-OM Deseninde NMY Becerisinin Testlere Göre Eşitleme Durumları

<i>Eşitleme Desenler</i>	<i>Eşitleme Türü</i>	<i>Eşitlenen Testler</i>		
		<i>AxB</i>	<i>AxC</i>	<i>BxC</i>
<i>Denk Olmayan Gruplarda Ortak Madde</i>	Doğrusal	-	-	Tucker
	Eşit Yüzdelikli	-	-	FE
	Kernel	-	-	PSE

DOG-OD deseni çerçevesinde ise NMY becerisi eşitleme incelemelerine dair bulgular Tablo 60'ta sunulmuştur. Tabloda verilenlere göre DOG-OD deseninde AxB ve AxC test eşitlemeleri yapılamamışken BxC test eşitlemelerinin kernel eşitleme yöntemiyle her iki değişken grubuna göre eşitlemelerin gerçekleştiği belirtilmiştir.

Tablo 60

DOG-OD Deseninde NMY Becerisinin Testlere Göre Eşitleme Durumları

<i>Eşitleme Deseni</i>	<i>Eşitleme Türü</i>	<i>Eşitlenen Testler</i>		
		<i>AxB</i>	<i>AxC</i>	<i>BxC</i>
<i>Denk Olmayan Gruplarda Ortak Değişken</i>	Kernel (değişkenler: okul kabul koşulu & okul türü)	-	-	√
	Kernel (değişkenler: okul kabul koşulu & cinsiyet)	-	-	√

Yorum ve Tartışma

Bu araştırma bugüne kadar test eşitlemede kullanılan klasik yöntem ve bakış açılarının bilişsel tanı modelleri bakış açısıyla birleştirilerek karma hiyerarşik yapılı GDINA model analizleri sonucu elde edilen beceri yetkinlik düzeyi olasılıkları kullanılarak hesaplanan beceri puanlarının testler arasında eşitlenebilme durumunun incelenmesi çalışmasıdır. Çalışmada her beceri için farklı grup desenlerinde eşitleme çalışmaları yapılmıştır. Bu kapsamda bazı becerilerin eşitlenebildiği ve bazılarının ise eşitlenemediği görülmüştür. Gözlenen bu durumun ardındaki genel özelliğin anlaşılabilmesi amacıyla her beceri için farklı grup desenlerinde yapılan eşitlemeler bir araya getirilerek aşağıda tablolaştırılmıştır. Böylece genel durum değerlendirilmesinin kolaylıkla yapılması ve sonuçların belirlenmesi hedeflenmiştir.

Genelleştirilmiş aritmetik beceri puanlarının eşitlenebilme durumu ile ilgili genel bir sonuca ulaşabilmek için testlere göre eşitleme durumuna ait bulgular Tablo 61’de bir bütün olarak sunulmuştur.

Tablo 61

Genelleştirilmiş Aritmetik Beceri Puanlarının Testlere Göre Genel Eşitleme Durumu

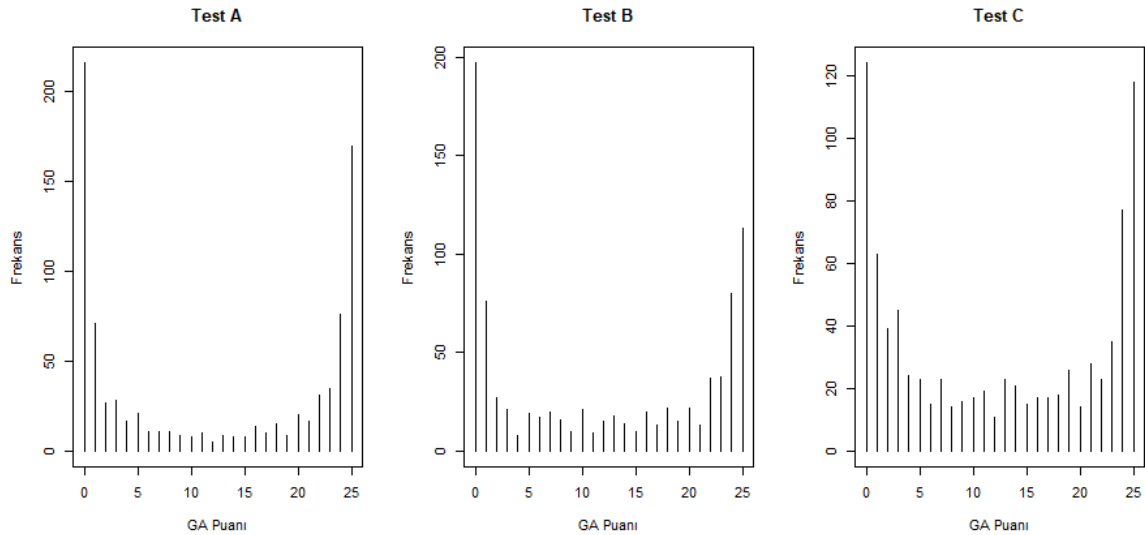
<i>Eşitleme Desenler</i>	<i>Eşitleme Türü</i>	<i>Eşitlenen Testler</i>		
		<i>AxB</i>	<i>AxC</i>	<i>BxC</i>
<i>Denk Gruplar</i>	Ortalama	√	√	√
	Doğrusal	√	√	√
	Eşit Yüzdelikli	√	√	√
	Kernel	√	√	√
<i>Denk Olmayan Gruplarda Ortak Madde</i>	Doğrusal	Tucker	Tucker	Tucker
	Eşit Yüzdelikli	FE	-	-
	Kernel	CE	-	-
<i>Denk Olmayan Gruplarda Ortak Değişken</i>	Kernel (değişkenler: okul kabul koşulu & okul türü)	PSE-EYE Doğrusal	-	-
	Kernel (değişkenler: okul kabul koşulu & cinsiyet)	PSE-EYE Doğrusal	-	-

Not: PSE: Post Stratification Equating (Son Tabakalama Eşitleme), EYE: Eşit Yüzdelikli Eşitleme

Bu tabloya göre genelleştirilmiş aritmetik becerileri için AxB için her üç grup deseninde (DG, DOG-OM, DOG-OD) eşitlemeler yapılabildiği gözlenirken AxC ve BxC için sadece DG deseninde eşitlemelerin yapılabildiği görülmüştür. Yine aynı testlerde DOG-OM deseninde sadece doğrusal eşitleme olarak Tucker eşitlemenin yapılabildiği gözlenmiştir. AxB test eşitlemesinin her üç grup deseni için uygun olması tabloda dikkat çekmektedir. Bu durumun puan dağılımları ile ilgisi incelenmiş ve aşağıda her teste ait genelleştirilmiş aritmetik beceri puanları dağılımı Şekil 43'te verilmiştir.

Şekil 43

Genelleştirilmiş Aritmetik Beceri Puanları Dağılımları



Not: Test A ($M=12.05$, medyan=11, $ss=10,69$),
 Test B ($M=11,73$, medyan=11, $ss=10,09$),
 Test C ($M=12.35$, medyan=12, $ss=9,67$).

Genelleştirilmiş aritmetik puan dağılımları görsel olarak incelendiğinde A ve B testlerine ait dağılımların çok benzer olduğu görülmektedir. Beceri puanlarına ait betimsel istatistikler incelendiğinde ise bu iki teste ait genelleştirilmiş aritmetik puanlarına ait ortalama ve standart sapma değerlerinin çok yakın olduğu ve meydanlarının eşit olduğu görülmektedir. Test C'nin dağılım grafiği diğer iki testten farklı görünmekle birlikte ortalama, medyan ve standart sapma değerleri diğer testlerle benzer özellikler göstermektedir. Genelleştirilmiş aritmetikte AxB testlerinde gözlenen bu eşitleme durumu

A ve B testlerine ait puan dağılımlarının benzerlik göstermesine bağlanmıştır. AxC ve BxC testlerinde DG deseninde gerçekleştirilen eşitlemenin ise Test C'nin moment değerlerinin diğer iki testin moment değerlerine yakın olmasına bağlamakla birlikte puan dağılımdaki farklılığının yanında ortak madde ve ortak değişkenlerinin bu puan dağılımına olan etksi sebebiyle DOG-OM ve DOG-OD desenlerinde puanların eşitlenmesine etkisinin olabileceği sonucuna varılmıştır.

Benzer biçimde fonksiyonel düşünme beceri puanlarının testlere göre eşitlenme durumuna ait bir yorum yapabilmek amacıyla bulgular Tablo 62'da bir bütün olarak sunulmuştur.

Tablo 62

Fonksiyonel Düşünme Beceri Puanlarının Testlere Göre Genel Eşitleme Durumu

<i>Eşitleme Desenleri</i>	<i>Eşitleme Türü</i>	<i>Eşitlenen Testler</i>		
		<i>AxB</i>	<i>AxC</i>	<i>BxC</i>
<i>Denk Gruplar</i>	Ortalama	√	-	-
	Doğrusal	√	-	-
	Eşit Yüzdelikli	√	-	-
	Kernel	√	-	-
<i>Denk Olmayan Gruplarda Ortak Madde</i>	Doğrusal	Tucker, Chain, Braun/Holland	-	-
	Eşit Yüzdelikli	FE	-	-
	Kernel	PSE, CE	-	-
<i>Denk Olmayan Gruplarda Ortak Değişken</i>	Kernel (değişkenler: okul kabul koşulu & okul türü)	√	-	-
	Kernel (değişkenler: okul kabul koşulu & cinsiyet)	√	-	-

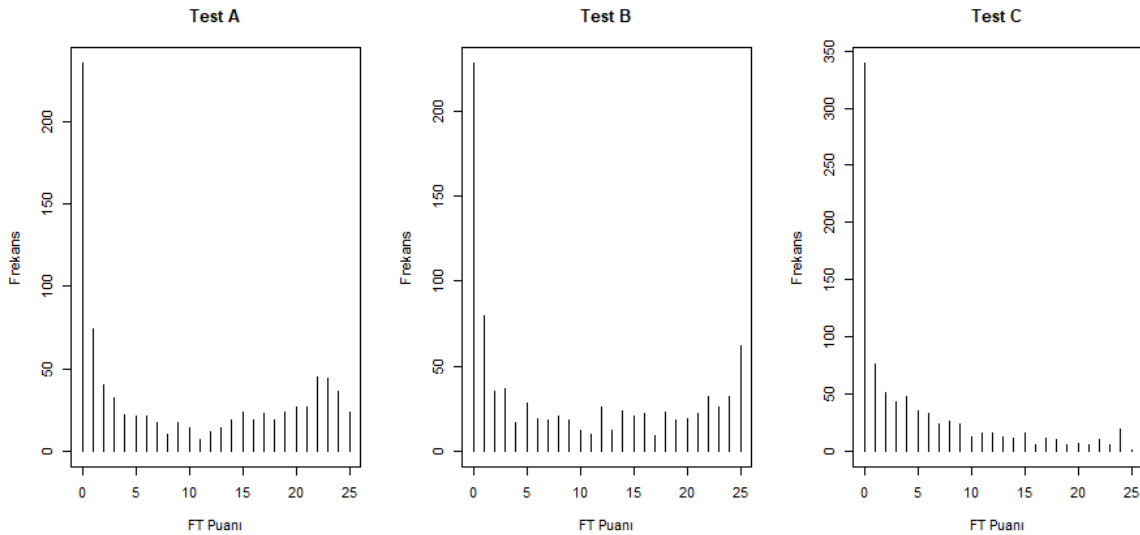
Not: FE (Frequency Estimation), PSE (Poststratification Equating), CE (Chain Equating)

Bu tabloya göre fonksiyonel düşünme beceri puanları için A ve B testlerinde her üç grup deseni için (DG, DOG-OM, DOG-OD) eşitlemeler yapılabildiği gözlenirken AxC ve BxC testlerinde ise eşitleme yapılamadığı görülmektedir. Bu durumun puan dağılımları ile

ilgisi incelenmiş ve aşağıda Şekil 44'te fonksiyonel düşünme becerisi için her teste ait beceri puanları dağılımı verilmiştir.

Şekil 44

Fonksiyonel Düşünme Beceri Puanlarının Dağılımları



Not: Test A ($M=9,41$, medyan=6, $ss=9,22$),
 Test B ($M=9,31$, medyan=6, $ss=9,22$),
 Test C ($M=4,85$, medyan=2, $ss=6,49$).

Fonksiyonel düşünme becerisi puan dağılımları görsel olarak incelendiğinde A ve B testlerine ait dağılımların benzer ve Test C dağılımının ise bu iki testten çok farklı olduğu görülmektedir. Beceri puanlarına ait betimsel istatistikler incelendiğinde ise Test C'nin ortalama, standart sapma ve medyan değerlerinin Test A ve Test B'den oldukça farklı ve düşük olduğu görülmektedir. Eşitleme yapılabilen Test A ve Test B dağılımlarında ise dağılımın benzerliğinin yanında bu iki teste ait fonksiyonel düşünme becerisi puan ortalaması ve standart sapma değerlerinin çok yakın olduğu ve meydanlarının eşit olduğu görülmektedir. Fonksiyonel düşünme becerisi puanları için AxB testlerinde gözlenen eşitleme durumu her iki testin dağılımlarının benzerlik göstermesine, ortalama, medyan ve standart sapma değerlerinin yakınlığına bağlanmıştır. AxC ve BxC testlerinde DG, DOG-OM ve DOG-OD desenlerinde puan eşitlemesinin gözlenememe durumu ise eşitlenen

puanlara ait dağılımların, ortalama, medyan ve standart sapma değerlerinin birbirinden farklı olmasına bağlanmıştır.

Niceliksel mantık yürütme beceri puanlarının desenlere ve testlere göre genel eşitleme durumlarını gösteren Tablo 63 aşağıda paylaşılmıştır.

Tablo 63

Niceliksel Mantık Yürütme Beceri Puanlarının Testlere Göre Genel Eşitleme Durumu

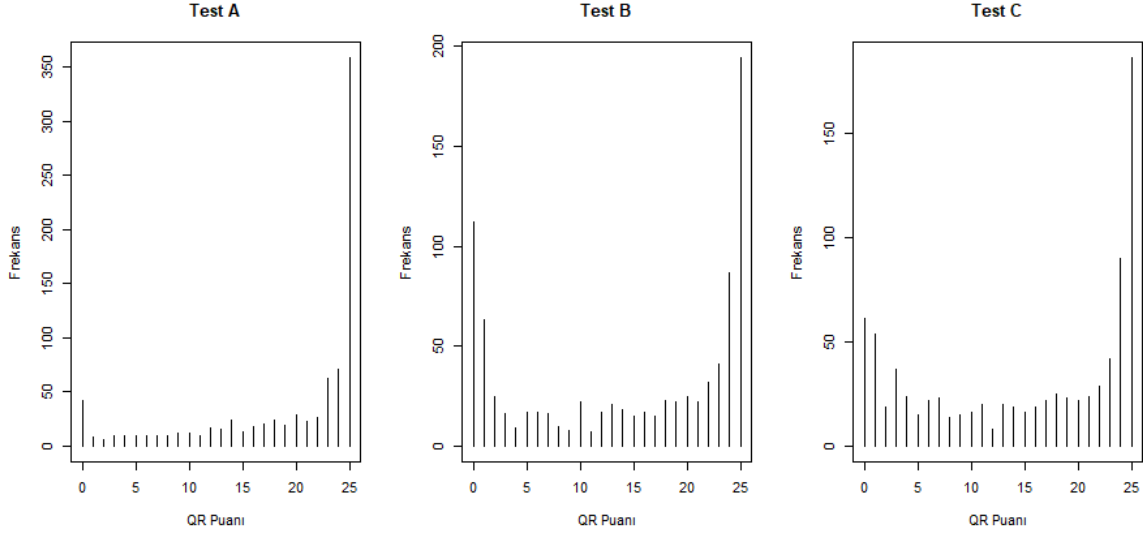
<i>Eşitleme Desenleri</i>	<i>Eşitleme Türü</i>	<i>Eşitlenen Testler</i>		
		<i>AxB</i>	<i>AxC</i>	<i>BxC</i>
<i>Denk Gruplar</i>	Ortalama	-	-	√
	Doğrusal	-	-	√
	Eşit Yüzdelikli	-	-	√
	Kernel	-	-	√
<i>Denk Olmayan Gruplarda Ortak Madde</i>	Doğrusal	-	-	Tucker
	Eşit Yüzdelikli	-	-	FE
	Kernel	-	-	PSE
<i>Denk Olmayan Gruplarda Ortak Değişken</i>	Kernel (değişkenler: okul kabul koşulu & okul türü)	-	-	√
	Kernel (değişkenler: okul kabul koşulu & cinsiyet)	-	-	√

Not: FE (Frequency Estimation), PSE (Poststratification Equating)

Bu tabloya göre niceliksel mantık yürütme beceri puanları için BxC testlerinde her üç grup deseni için (DG, DOG-OM, DOG-OD) eşitlemeler yapılabildiği gözlenirken AxB ve AxC testlerinde ise eşitleme yapılamadığı görülmektedir. Diğer becerilerde olduğu gibi niceliksel mantık yürütme beceri puanlarında da eşitleme durumunun puan dağılımları ile ilgisi incelenmiş ve aşağıda her teste ait niceliksel mantık yürütme beceri puanları dağılımı verilmiştir.

Şekil 45

Niceliksel Mantık Yürütme Beceri Puanları Dağılımları



Not: Test A ($M=19,32$, medyan=23, $ss=7,7$),

Test B ($M=14,7$, medyan=18, $ss=9,8$),

Test C ($M=15,12$, medyan=18, $ss=9,3$).

Verilen grafiklerde niceliksel mantık yürütme becerisine ait puan dağılımları incelendiğinde B ve C testlerine ait dağılımların benzer ve Test A dağılımının ise bu iki testten çok farklı olduğu görülmektedir. Beceri puanlarına ait betimsel istatistikler incelendiğinde ise Test A'nın ortalama, medyan ve standart sapma değerlerinin Test B ve Test C'den oldukça farklı olduğu görülmektedir. Eşitleme yapılabilen Test B ve Test C dağılımlarında ise dağılımın benzerliğinin yanında bu iki teste ait niceliksel mantık yürütme becerisi puan ortalaması ve standart sapma değerlerinin çok yakın olduğu ve meydanlarının eşit olduğu görülmektedir. Niceliksel mantık yürütme beceri puanları için BxC testlerinde gözlenen eşitleme durumu her iki testin dağılımlarının benzerlik göstermesine, ortalama, medyan ve standart sapma değerlerinin yakınlığına bağlanmıştır. AxB ve AxC testlerinde DG, DOG-OM VE DOG-OD desenlerinde puan eşitlemesinin gözlenememe durumu ise eşitlenen puanlara ait dağılımların, ortalama, medyan ve standart sapma değerlerinin birbirinden farklı olmasına bağlanmıştır.

Sonuç olarak, üç farklı test uygulamasından elde edilen üç farklı beceri puanı üzerinden yapılan eşitleme incelemesinde karma hiyerarşik yapılı GDINA analizlerinden çıktı olarak elde edilen veri, test eşitleme analizlerinde girdi, yani beceri puanı olarak kullanılmıştır. Bu beceri puanlarının bazılarında eşitleme yapılabilmesi ve bazılarında yapılamaması durumunun özellikle eşitleme yapılabilen beceri puanlarının benzer dağılıma sahip olması, sağa çarpık ya da sola çarpık biçimde olmaması, ortalama, medyan, standart sapma gibi moment değerlerinin yakın olmasının etkisi olduğu düşünülmüştür. Bu beceri puanları paralel formların uygulanmasından elde edilmiştir ancak her birinin kendi içerisinde beceriye ait bir dağılımı mevcuttur. Yapı, konu, güçlük derecesi, maddelerin sırası, maddelerin senaryosu, kazanımı, BTM çerçevesindeki nitelikleri bakımından paralellik gösteren bu testler ölçülen becerilerde test maddelerinin özel durumlarına göre farklılıklar sergilemiş ve bu testi alan ilgili grupların beceri puanları dağılımına yansımıştır.

Literatürde bu çalışmadaki gibi dolaylı olarak elde edilen puanlar üzerinde eşitleme çalışması mevcut değildir ancak literatürdeki var olan yorumların ışığında bu çalışmanın bulgularına bakmak da faydalı olacaktır. Tong ve Kolen (2005) çalışmalarında testlerin zorluk derecesinin ve test puan dağılımlarının eşitlemeye olan etkilerini rastgele gruplar deseninde incelemişlerdir. Araştırmacılar alternatif formlardaki zorluk derecelerindeki farklılığın ve büyüklüğün kullanılacak eşitleme yöntemlerinin özelliğinin geçerlilik derecesini etkilediğini belirtmişlerdir. BTM çerçevesinde yapılan analizler açısından bakılırsa test formları genel anlamda hem KTK hem de BTM temelinde paralel formlar olarak istatistiksel sonuçlar vermiş olmasalar bile madde bazında farklı güçlük derecesine sahip maddeler nitelik düzeylerinde yani beceri düzey ve puanlarında farklılıklara sebep olan temel unsurdur. Dolayısı ile beceri puanları açısından farklı dağılımların elde edildiği görülmektedir. İşte bu beceri puanlarında yapılan eşitlemelerde ise dağılım özellikleri ön plan çıktığı görülmektedir.

Yine Tong ve Kolen (2005) çalışmalarının devamında test puanı dağılımları benzer özellik gösteren gruplarda eşitleme incelemelerinde bulunmuşlar, dağılımları benzer olan gruplarda kullandıkları yöntemlerin hepsinde eşitleme yapılabilirken, farklı olan gruplarda eşitleme durumu yöntemlere göre farklılıklar göstermiş veya hiç yapılamamıştır. Buradaki test puanları dağılımları açısından bir benzerlik olarak bu çalışma için düşünülürse dağılımlardaki benzerlik ve eşitleme durumu her üç desen için eşitlemenin gözlemediği ve gözlenemediği durumlarla karşılaşılmıştır. Çalışmadaki denk gruplarda yapılan eşitlemelere genel anlamda bakıldığında dağılımları benzer özellik gösteren gruplarda eşitlemelerin yapılabildiğidir. Eşitlemenin yapılamadığı durumlarda dağılımlardan biri sağa veya sola çarpık, yani dağılım özellikleri birbirinden farklıdır. Denk olmayan gruplarda ortak değişken desenlerinde bazı testlerin eşitlenmesinde ortak değişkenlerin kullanılabilirdiği gözlemediği gibi bazı testlerin eşitlenmesinde bu durum gözlenememiştir. Ortak değişken ile eşitleme gözlenemeyen durumlarda yine testlerden birinin sağa veya sola çarpık oluşu ve farklı dağılımlar dikkat çekmektedir. Denk olmayan gruplarda ortak madde desenlerinde de yine bazı testlerde eşitlemenin yapıldığı bazılarında ise yapılamadığı gözlenmiştir. Bu durumda da diğer iki desendeki aynı durum yani farklı puan dağılımlarının yansımaları olarak düşünülmüştür.

Eşitlemelerin gerçekleşemediği durumların ardında eşitleme yapılan iki grubun farklı puan dağılımları öne çıkmakla birlikte, özellikle DOG-OM deseninde benzer dağılım varsayımı olmadığı göz önüne alınırsa gözlenen bu duruma farklı açılardan da yaklaşmak uygundur. Örneklem büyüklüğünün test eşitleme sonuçları üzerinde etkili olduğu bilinmektedir (Eid, 2005; Fitzpatrick & Yen, 2001; Kilmen & Demirtasli, 2012). Çalışmada gözlenen duruma sebep olabilecek bir neden ise sağa veya sola çarpık dağılımlardaki puanların örneklemin küçük olması sebebi ile yeteri kadar temsil edilememesi olabilir. Bu durumda örneklem büyüklüğü ve puan dağılımlarının sağa veya sola çarpık olması bir arada etki etmiş olabilir. Bu çalışmada gözlenen durumu daha büyük örneklerde çarpık dağılım olsa bile puanların temsil edilebilirliği ile farklı durumların gözlenme ihtimali de vardır. Bu

sebeple, örneklem büyüklüğü, puan dağılımları, puanların örnekleme temsil edilebilirliği faktörleri de burada gözlenen sonuçların ardındaki nedenlerden olma olasılığı da düşünülmektedir. Ayrıca ortak maddeler ile ilgili yapılan test eşitleme çalışmaları (Ricker & von Davier, 2007; Sinharay & Holland, 2006; Yang & Houang, 1996) ışığında bu çalışmaya bakılırsa, ortak maddelerin sayısı, ortak maddelerden elde edilen beceri puanlarının örneklemlerdeki dağılımı, ortak maddelerin zorluk derecesinin dağılımı öne çıkar. BTM açısından da bakılırsa ortak maddelerin ölçülen gizil beceriler açısından dağılımı, BTM çerçevesinde ortak madde sayısının yeterliliği gibi konular da puan dağılımları ve puanların temsil edilebilirliği ile ilişkili olarak düşünülmüştür. Denk olmayan grupta ortak madde deseninde hem genel beceri puanları açısından hem de ortak maddelerden elde edilen beceri puanlarının özellikle dağılım, momentler ve yeterli örneklem ile puanların temsil edilebilirliği gibi durumlar inceleme kapsamına alınması faydalıdır.

Bu bulgu ve yorumlar ışığında beceri puanlarının eşitlenmesi hedefi taşıyan test uygulamalarında puan dağılımları ve puanların temsil edilebilirliği ile ilgili olarak test geliştirme aşaması ile ilişkili hususların bulunduğu da anlaşılmaktadır. Bunlardan biri oluşturulan Q-matrisin en iyi biçimde tanımlanması ve maddelerin ölçmeyi hedeflediği niteliği gerçekten ölçüyor olmasının yetkinlik düzeyinin belirlenmesi açısından önemi (de la Torre & Chiu, 2016; & Suen, 2013; Qin & Guo, 2023) beceri puanlarının eşitlenmesi açısından da kendisini göstermektedir. Niteliklerin Q-matristeki dağılımının beceri puanlarının dağılımına ve eşitlemedeki rolüne olan etkisi düşünülmelidir. Testte ölçülecek becerilerin ilgili literatür kaynaklarından öğrencilerin yetkinlik durumları, öğrenmedeki zorlukları gibi konular tespit edilerek Q-matristeki yetenek dağılımı ayarlanması beceri puanlarının güvenilir bir biçimde tespit edilebilmesine katkı sağlayacaktır. Örneğin fonksiyonel düşünme becerisinde literatür kaynaklarında ve öğretmen deneyimleri ışığında öğrencilerin güçlük çektiği bir beceri olarak bilinmektedir. Bu çalışmada fonksiyonel becerilerin puan dağılımları sağa çarpık biçimde olduğu için düşük puanlara

denk gelen öğrenci sayısı yüksek iken diğer uçtaki öğrenci sayısı eşitlemeye imkan verebilecek kadar yeterli değildir.

Literatür ışığında çeşitli çalışmalarda ölçülmesi hedeflenen bir beceri ile ilgili olarak öğrencilerin zorluk yaşadığı veya düşük yetkinlik düzeylerinde olduğu belirtilmiş ise o beceri ile ilgili daha çok veri elde edebilmek adına Q-matriste ilgili becerinin dağılımının yanında testi oluşturan maddelerin zorluk derecelerinin çeşitliliğine (Haladyna & Rodriguez, 2013; Henson & Douglas, 2005; Kline, 2015) de dikkat edilmesi geçerli ve güvenilir ve geçerli beceri puanlarının elde edilmesinde etkili olacaktır. Yine fonksiyonel düşünme becerisi için örnek verilecek olursa daha fazla maddede farklı zorluk derecelerinde testlerde yer alması uygun olabilirdi. Buradaki testler bu açıdan yeniden revize edilebilir. Sonuç olarak, çalışmada görüldüğü üzere ilgili becerilerde madde zorluk derecesi çeşitliliği azlığı sebebiyle yeterli veri toplanamayan beceriler sağa ya da sola çarpık olma durumu sergileyebileceği için puanların temsil edilememesinden kaynaklı olarak eşitleme için yeterli olmayacaktır. O halde madde yazımında ilgili becerilere ait puan dağılımı da göz önünde tutulması ve madde zorluk derecesi çeşitlendirilmiş maddelerin kullanımına önem verilmesi BTM çerçevesinde elde edilen puanlara ve bu puanların eşitlenmesi çalışmalarına katkı sağlayacağı düşünülmektedir. Burada amaç ilgili becerilerde yeteri kadar veriye ulaşarak test eşitleme yapabilmek için dağılımların her puan seviyesinde yeteri kadar bilgi sunabilmesini sağlanmaktır. Dağılımların sağa veya sola çarpık olmasının eşitlemeye olumsuz etkisi bu çalışmada gözlemlendiği üzere ilgili puanlarda yeterli kişi sayısına ulaşılamadığı ve eşitlemenin olumsuz etkilendiği anlaşılmaktadır. Bu durumdan yola çıkarak, uç noktalarda yoğunluğun olduğu ve ara puanlarda yoğunluğun çok az olduğu dağılımlarda da eşitlemenin problemlili olacağı düşünülmelidir. Testlerden birinin dahi bu durumda olması eşitlemenin yapılamayacağı anlamına geldiği bu çalışma ile de görülmüştür. Dağılımlarda her puan diliminde eşitleme için yeterli kişi sayısının olmasının önemi anlaşılmıştır.

Bu çalışma BTM çerçevesinde beceri puanları elde ederek bu puanların eşitlenmesi hedefini test maddelerinin yazımından başlayarak test geliştirme ve uygulama aşamalarından itibaren kontrol etmiştir. Dolayısı ile ortaya çıkan bulgularda araştırmanın kapsamında yer alan testlerdeki madde yazımının da çalışmadaki beceri puanlarının dağılım özelliklerine ve eşitleme incelemelerine ait bulgularda etkisi olduğu düşünülerek buraya da değinmek önemlidir. Genel test puanlarına ait istatistiklerde testler arası farklılıklar gözlenmesine rağmen bu fark dikkate değer olarak tarif edilemezken, özel olarak beceri puanları çerçevesinden bakıldığında, örneğin fonksiyonel düşünme beceri puanlarında Test C'deki durum için diğer testlere göre farkının çok daha fazla olduğu ile ilgili tarif yapılabilir. İşte bu öngörü ile yazılan test maddelerindeki farklılıkların beceri puanlarına olan etkisi de göz önünde tutularak her becerinin yeterli biçimde bilgi sunabilmesi sağlanmalıdır. Test maddelerinin senaryosunun, ilgili çözüm yolları ve çözüme giden düşünme biçimlerinin, aynı beceriyi ölçen maddelerin konu, kazanım ve zorluk derecesi bakımından da benzer olmasına önem verilmelidir. Çalışmada görüldüğü üzere örneğin testlerdeki 12. madde örneği gibi (araştırmanın yöntem bölümünde paylaşılmıştır) aynı çözüm yoluna sahip olan maddelerin sadece senaryoda kullanılan durumdan dahi cevaplanma oranının etkilendiği görülmüştür. Sadece sayısı değişen bir matematiksel gösterimin (testlerdeki madde 8) farklı oranlarda cevaplandığı anlaşılmıştır. Hatta aynı kazanıma ait farklı görsellerin (madde 11) farklı yorum ile çözüme ulaşılması sebebiyle testlerde hem bu maddelerin yapılabilme oranları hem de ilgili becerilerin gruplardaki yetkinlik düzeylerinde açık farklara sebep olduğuna rastlanılmıştır. Test maddelerindeki aynı becerinin aynı kazanım çerçevesinde aynı çözüm yolları olsa bile farklı senaryolarla sorulmasından kaynaklı olarak farklı cevaplanma oranlarıyla ve dolayısıyla farklı beceri yetkinlik düzeylerinin ortaya çıkmasına sebep olduğu uygulama aracılığı ile gözlenmiştir. Bu sebeple ilgili beceri ve kazanımların hem zorluk derecesi çeşitliliği hem de niteliklerin dağılımlarının literatür temelli olarak öğrencilerin genel durumunun farkındalığı ile hazırlanmasının katkısı beceri puanlarının eşitlemesinde ortaya çıkacaktır.

Tanılayıcı testlerden elde edilen beceri puanlarında eşitleme incelemelerindeki bu çalışma çerçevesindeki bulgularından yola çıkılarak beceri puanlarının dağılımına etki edebilecek nedenlerden olabileceği belirtilen nitelik dağılımı ve madde zorluk derecesinin çeşitliliğinin dolaylı olarak testteki madde sayısının beceri puanları açısından önemine de işaret ettiği düşünülebilir. BTM çerçevesinde yapılan çalışmalarda nitelik sayısının arttıkça test uzunluğunun artmasının niteliklerdeki yetkinlik düzeyinin daha doğru biçimde belirlenebileceği bildirilmektedir (Henson & Douglas, 2005; Huebner ve diğerleri, 2018; Kuo ve diğerleri, 2016; Wang, 2013). Tanılayıcı testlerde ilgili becerilerde madde zorluk derecesi çeşitliliğinin sağlanabilmesi ve niteliklerin Q-matristeki dağılımının yeterli düzeyde olabilmesi için test madde sayısının da yeterli düzeyde olması, böylece yetkinlik düzeylerinin daha doğru bir biçimde belirlenebilmesi, dolayısı ile beceri puanlarının saptanabilmesi ve puan dağılımları bakımından önem kazanmaktadır.

BTM çerçevesinde yapılan araştırmalarda sınıflandırma doğruluğu sonuçlarının nitelik sayısının fazla olması durumlarında negatif yönde etkilenebileceği, bu sebeple nitelik sayısının fazla tutulmaması tavsiye edilmektedir (Henson ve diğerleri, 2008; Rupp ve diğerleri, 2010; Wang ve diğerleri, 2015; Xu, 2019). Eşitleme amacı olan tanılayıcı testlerde de nitelik sayısının fazla olması ilgili nitelikler hakkında yeterli bilgi sağlamada etkisiz kalabileceği düşünülmelidir. Bu sebeple tanılayıcı testlerde eşitleme hedefi var ise ölçülecek nitelik sayısına da dikkat edilmelidir, çünkü nitelik sayısı ilgili niteliğin farklı madde zorluk derecesi ile tekrarlı ölçme yapılabilmesi nitelik sayısından etkilenecektir. Nitelik sayısı fazla ise madde zorluk derecesi çeşitliliği ile ölçülmesi önerilen nitelikler daha az sayıda ölçülmüş olacaktır.

Test geliştirme aşamalarından itibaren maddelerin denenmeye başlanması, ona göre revize edilirken beceri puanlarının dağılım biçimine dair ihtimalleri de kendisini göstermeye başlayacaktır. Maddelerin gerçek uygulamadan önce denenmesi, beceri puanlarının ait olduğu testteki dağılımın sağa ya da sola çarpık olmasına etkisi olup olmayacağı test oluştururken dikkate alınmalıdır. Eğer maddenin daha az ya da daha çok

kiři tarafından cevaplanacağı öngörölüyorsa benzer yönelimi olan maddelerin dağılımların çarpık olma ihtimali göz önüne alınarak aynı testte bir arada kullanılmamasına özen gösterilmelidir.

Bu çalışma BTM çerçevesinde geliştirilen ve analiz edilen testlerden elde edilen yetkinlik düzeylerinin beceri puanlarına transferi ile beceri puanlarının eşitlenmesi üzerine başlangıç yapılmış bir çalışmadır. Çalışmanın bulgularına dayanarak varolan durumun sebepleri üzerinde düşünöldüğünde BTM ve test eşitleme hedefinin beceri puanları açısından test geliştirme aşaması, nitelik sayısı, niteliklerin testlere dağılımı, test ve ortak madde sayısı, madde zorluk derecelerinin yayılımı, öğrencilerin güçlük çektiği konu ve becerilerin daha alt birimlerine göre ve kolaydan zora daha çok madde sayısı ile ölçülmesi, bütün bunların ortak madde testleri açısından da düşünölmesi gibi faktörlerin göz önüne alınarak arařtırmalara devam edilmesi önemlidir.

Bölüm 5

Sonuç ve Öneriler

Bu çalışma kapsamında elde edilen en genel sonuç bilişsel tanılama ölçme hedefleriyle hazırlanan testlerin bilişsel tanı modelleri çerçevesinde analiz edilerek bireylere ait beceri yetkinlik düzeyine ilişkin olasılık değerleri beceri puanlarına çevrilerek becerilerde eşitleme yapılabildiğidir. Yapılan uygulamada becerilerin eşitlenebildiği gözlenmiştir ancak eşitlemelerin uygun olmadığı bulgular da elde edilmiştir. Bu iki durumun gözlenme sebebinin eşitlenecek testlere ait beceri puanlarının dağılım biçimi, ortalama, medyan ve standart sapmaları ile ilgili olduğu sonucuna ulaşılmıştır.

Bulgulardan yola çıkılarak varılan sonuç tanılayıcı testlerden elde edilen beceri puanları dağılımı sağa ya da sola çarpık biçimde olan testlerde puan dağılımları yeteri kadar temsil edilemediği için eşitlemenin yapılamadığıdır. Eşitlenecek iki testten en az birinin dağılımı sağa ya da sola çarpık ise eşitleme gözlenmemiştir.

Bulgulardan elde edilen bu sonuçlar eşitleme desenlerine göre herhangi bir farklılık sergilememiştir. Dağılım, ortalama ve standart sapmadaki farklılık her desende eşitlemeye izin vermemiştir. Dolayısı ile eşitleme desenlerinde dağılım ile ilgili herhangi bir farklı eğilim gözlenmemiştir. Dağılım çarpık ise her desende farklı eşitlemeler uygulanmasına rağmen eşitlemenin yapılamadığına dair bulgular elde edilmiştir.

Özel bir sonuç olarak, genelleştirilmiş aritmetik becerisinde DOG-OM ve DOG-OD desenlerinde eşitleme yapılamayan AxC ve BxC testlerinde DG deseninde eşitleme gözlenmiştir. Burada GA becerilerinin A ve B testlerindeki dağılımları benzer, ortalama ve standart sapma değerleri yakın olduğu için eşitlemenin her desende yapılabilmesinin gözlenmesi sonucunu doğrulamaktadır. Ancak Test C'ye ait dağılım biçimi A ve B testlerine benzemekle birlikte özellikle standart sapması bu iki testten farklıdır. Dolayısı ile bu farklılık sebebiyle DOG-OM ve DOG-OD desenlerinde eşitleme gerçekleşmemiş ancak DG desenlerinde eşitleme yapılabilmektedir. Eşitlenecek testlerin dağılımları benzerlik

gösterse de standart sapmalarının farklı olması DOG-OM ve DOG-OD desenlerinde eşitlemeyi uygun kılmaz.

Araştırmacılara Yönelik Öneriler

- Uygulanan bu çalışma ve paylaşılan sonuçları ışığında yeni araştırmalar yapılarak burada elde edilen sonuçların daha çok dayanak bulmasına katkı sağlayacak biçimde yeni çalışma önerileri kapsamında öncelikli olarak böyle bir çalışmanın daha çok örneklem ve daha fazla madde sayısı ile tekrarlmasıdır. Madde sayısının fazla olmasının altında yatan sebep ise öğrencilerin fonksiyonel düşünme becerisinde genel anlamda zorluk yaşadıkları ve bu beceriyi ölçen basit, orta ve zor güçlükte maddelerin yanında basit-orta arası ile orta-zor arası maddelere yer verilerek bu beceriden elde edilen puanların daha geniş bir biçimde yansıtılması hedeflenmelidir. Böylelikle farklı becerilerde farklı sayıda örneklem ile yapılan bilişsel tanı uygulamalarından elde edilen beceri puanlarının eşitlenmesi çalışmalarına da katkı sağlanacaktır. Literatürde bu tür bir çalışma mevcut olmamakla birlikte bu çalışmanın benzerleri uygulanarak eşitleme durum incelemeleri yapılarak buradaki sonuçları destekleyen ve geliştiren çalışmalara yer verilebilir.

- Bu çalışmada bilişsel tanı modelleri çerçevesinde elde edilen yetkinlik düzeyi olasılık değerlerinden beceri puanları elde edildiği belirtilmişti. Bu çalışmada beceri puanları 25 puan üzerinden değerlendirilerek eşitleme uygulamaları yapılmıştır. Ancak benzer bir çalışmanın farklı puanlarla (100, 50, 25,15, 10 gibi) eşitleme çalışmaları yapılarak puan aralığının eşitlemeye olan etkisini inceleyen çalışmaların da destek olabileceği düşünülmektedir.

- Çalışmanın sonucu olarak eşitlemede kullanılan dağılımların benzer ve yeterli kişi sayısında olmasının önemi vurgulanmıştır. Aynı çalışmada her testte dağılımların biçimleri aynı tutularak ve örneklemin artırıldığı simülatif çalışmalarla denenmelidir. Ayrıca sağa çarpık veya sola çarpık iki dağılımda büyük örneklemelerde eşitleme çalışmaları

incelemeleri yapılmalıdır. Buradaki çalışmada hem örneklem sayısının yetersiz olması hem de eşitlenecek testin ikisinin de aynı anda sağa veya sola çarpık olan bir durum incelemesi yapılamamıştır. Paralellik ve diğer koşulları sağlayan iki testte büyük örneklerle aynı anda sağa veya sola çarpık dağılımlarda eşitleme durumu incelenmelidir.

- Niteliklerin Q-matristeki dağılımının beceri puanları elde edilmesine olan etkisi incelenerek bunun eşitlemeye olan etkisinin araştırılması önerilebilir. Ayrıca ölçülen niteliklerin maddelerin zorluk derecesine göre dağılımlarının beceri puanlarının saptanmasına olan etkisinin incelenmesi de eşitleme durumu ile ilgili olarak yeni bakış açıları kazandırabilir.

- Her test için mühim derecede önemli olan madde yazımı tanılayıcı testler açısından da çok önemlidir. Maddelerin ölçtüğü beceriler yanında eğer eşitleme hedefi var ise en az üç test ile testin biri referans seçilerek yeni bir araştırma yapılabilir. Bu çalışmada diğer ikinci bir test mümkün olduğunca referans testin özellikleri ile bir tutularak, maddelere ait bazı faktörler (konu, kazanım, senaryo gibi) sabit tutulmaya çalışılarak çok ufak değişimlerle hazırlanabilir. Diğer üçüncü bir test ise bazı maddelerinde aynı konu ve benzer kazanımlar olmasına rağmen senaryoda, görsellerde veya aynı konunun farklı bir kazanımında olduğu biçimde ayarlamalar yapılarak test uygulaması denenebilir. Bu üç testin farklılaşan maddelere göre beceri puanları incelenebilir ve eşitlemeler karşılaştırılabilir. Böylece tanılayıcı testlerde madde yazımının testlerde ölçülmek istenen niteliklerin yetkinlik düzeyine ve beceri puanlarının eşitlenmesine olan etkisi de incelenmiş olur.

- Bu çalışmadan çok daha farklı bir yaklaşımla, bilişsel tanı modellerinde cevap örüntüsünün ve Q-matrisin direk olarak etkin olduğu, model niteliklerinin ve modelde saptanan parametre ve diğer değerlerin kullanılarak kendi içinde bilişsel tanı modellerine has çerçevede eşitleme analizleri yapılabilmesi tanılayıcı testler açısından büyük bir

ihtiyaçtır. Bu yöndeki araştırma ve geliştirme çalışmalarına da önem verilmesi ve literatürdeki bu açığın kapatılması beklenmektedir.

Uygulayıcılara Yönelik Öneriler

Günümüzde ders programlarında belirlenen kazanımların edinilmesinin önemi yanında bilişsel becerilerin edinilmesi ve geliştirilmesi önemi sürekli vurgulanmaktadır. Bilgi her çağda önemlidir ve bizim bulunduğumuz bu çağda bilginin etkin kullanılmasını sağlayan temellerin düşünme becerilerinin etkin kullanılmasına bağlı olduğu eğitim-öğretim topluluklarında dile getirilmektedir. Bu çalışma eğitim ve öğretimde hem öğrencilerin bilişsel becerilerinin belirlenerek geri bildirim almalarında hem de elde edilen beceri puanlarının farklı gruplar için eşitleme yapılabilmesi eğitim öğretimde eşitlik ve anlamlı karşılaştırmalar yapabilme açısından katkı sağlayan bir uygulamadır.

- Bu uygulamanın büyük kitlesel sınavlarda kullanılabilmesi için eğitimde karar alıcıların aracılığı ile bu çalışmaya benzer uygulamaların daha geniş örneklerde pilot uygulamalar yapılarak geliştirilmesi ve ileriki aşamalarda kitlesel sınavlarda kullanılması sağlanabilir. Ülkemizde geniş ölçekli sınav yapan MEB ve ÖSYM gibi kurumların bu gibi çalışmaların sonuçları üzerinden kurum bünyesinde yeni çalışmalar yaparak uygulamalar yapması ülkemizde ölçme uygulamalarının gelişmesine katkı sağlayacaktır.

- Bu çalışmadan da görüldüğü üzere bilginin veya ders kazanımlarının ölçülmesine odaklanan testlerin yanında bilgilerin kullanımını sağlayan beceri ölçmesinin de günümüzde yapılabildiği görülmektedir. Becerileri nitelik olarak belirleyen ve ölçen testlerin kullanımı örneğin ALES gibi akademik personel seçiminde becerilerin ön plana çıktığı durumlarda kullanılabilir ve tekrarlı olarak yapılan bu sınavların beceri eşitlemelerinin yapılarak eşitliğinin sağlanmasına katkı sağlayacaktır. Örneğin ÖSYM'nin uyguladığı herhangi bir matematik testi lise öğretim programları kapsamındaki kazanımlar çerçevesinde ölçme yapıyor olsun. Klasik test kuramı bağlamında bu test ile öğrencilere ait puanlara göre sıralama yapılabilirken diğer taraftan belirlenmiş olan birkaç beceri

çerçevesinde de ölçmeler yapılabilir. Mesela çözümleme, analiz etme, yorumlama, mantık yürütme gibi beceriler belirlenerek her bir maddenin ölçmeyi hedeflediği becerilerle oluşturulan bir Q-matris ve uygun bir bilişsel tanı modeli ile sınav uygulanan adayların klasik test kuramı ile belirlenen puanlarının yanında ayrıca bu belirlenen becerilere ait yetkinlik düzeyleri veya beceri puanları da belirtilebilir. Böylelikle adayların kendileri hakkında bir geribildirim sağlanmasının yanında puan esasına dayanarak yapılan iş görüşme mülakatlarında verilen bu geri bildirimler işverenler tarafından da adayların seçiminde kullanılabilir. İşte bu çalışma çerçevesinde gerçekleştirilen beceri puanlarının eşitlenmesi de farklı yıllarda aynı sınavlara giren kişiler arasında yanlılık olmaması adına beceri puanlarının eşitlenmesi yapılabilir. ÖSYM ve MEB gibi ölçme yapan kurumlar beceri sınavlarının hazırlanmasında ve uygulanmasında, testlerin eşitlenmesinde bu test sonuçlarını kullanan kurumların daha etkin bir biçimde seçme yapmasına katkı sağlayacaktır.

Kaynaklar

- Akın Arıkan, C., & Gelbal, S. (2021). A Comparison of Kernel Equating and Item Response Theory Equating Methods. *Eurasian Journal of Educational Research*, 93, 179-198.
- Akın Arıkan, Ç., and Gelbal, S. (2018). A comparison of traditional and kernel equating methods. *Int. J. Assess. Tools Educ.* 5, 417–427. <https://doi:10.21449/ijate.409826>
- Aksekiöğlü, B. (2017). Madde tepki kuramına dayalı test eşitleme yöntemlerinin karşılaştırılması: PISA 2012 fen testi örneği.
- Albano A (2022). *equate: Observed-Score Linking and Equating*. R package version 2.0-8, <https://cran.r-project.org/web/packages/equate/equate.pdf>.
- Albano, A. D. (2016). equate: An R package for observed-score linking and equating. *Journal of Statistical Software*, 74(8), 1–36.
- Altıntaş, Ö., & Wallin, G. (2021). Equality of admission tests using kernel equating under the non-equivalent groups with covariates design. *International Journal of Assessment Tools in Education*, 8(4), 729-743.
- Andersson, B., Branberg, K., Wiberg, M., & Andersson, M. B. (2022). *kequate: The Kernel Method of Test Equating*. R package version 1.6.4, <https://CRAN.R-project.org/package=kequate>
- Atar, B., & Yeşiltaş, G. (2017). Çok Boyutlu Eşitleme Yöntemlerinin Eşdeğer Olmayan Gruplarda Ortak Madde Deseni için Performanslarının İncelenmesi. *Journal of Measurement and Evaluation in Education and Psychology*, 8(4), 421-434.
- Başaran, E. S., Mutluer, C., & Çakan, M. (2023). A Comparison of Covariates, Equating Designs, and Methods in Equating TIMSS 2019 Science Tests. *Participatory Educational Research*, 10(5), 41-63.

- Bejar, I. I. (1984). Educational diagnostic assessment. *Journal of educational measurement, 21*(2), 175-189.
- Birenbaum, M., Tatsuoka, C., & Xin, T. (2005). Large-scale diagnostic assessment: Comparison of eighth graders' mathematics performance in the United States, Singapore and Israel. *Assessment in Education: Principles, Policy & Practice, 12*(2), 167-181.
- Blanton, M., Levi, L., Crites, T., Dougherty, B., & Zbiek, R. M. (2011). *Developing essential understanding of algebraic thinking for teaching mathematics in grades 3-5*. National Council of Teachers of Mathematics.
- Bolt, D. (2007). The present and future of IRT-based cognitive diagnostic models (ICDM) and related methods. *Journal of Educational Measurement, 44*(4), 377–383.
- Bozdağ, S., & Kan A. (2010). Şans başarısının test eşitlemeye etkisi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 39*(39), 91-108.
- Bradshaw, L. (2016). Diagnostic classification models. *The Wiley handbook of cognition and assessment: Frameworks, methodologies, and applications, 297-327*.
- Bradshaw, L., & Templin, J. (2014). Combining item response theory and diagnostic classification models: a psychometric model for scaling ability and diagnosing misconceptions. *Psychometrika, 79*(3), 403–425. <https://doi:10.1007/s11336-013-9350-4>
- Bradshaw, L., Izsak, A., Templin, J., & Jacobson, E. (2014). Diagnosing teachers' understandings of rational numbers: Building a multidimensional test within the diagnostic classification framework. *Educational measurement: Issues and practice, 33*(1), 2-14.
- Bränberg, K., & Wiberg, M. (2011). Observed score linear equating with covariates. *Journal of Educational Measurement, 48*(4), 419-440.

- Briggs, D. C., & Circi, R. (2017). Challenges to the use of artificial neural networks for diagnostic classifications with student test data. *International Journal of Testing*, 17(4), 302–321. <https://doi:10.1080/15305058.2017.1297816>
- Canpolat, A. (2019). *Bilişsel tanı modellerinde madde parametreleri üzerinden örtük yetenek sınıflarının eşitlenmesi* (Master's thesis, Eğitim Bilimleri Enstitüsü).
- Chen, F., Liu, Y., Xin, T., & Cui, Y. (2018). Applying the m_2 statistic to evaluate the fit of diagnostic classification models in the presence of attribute hierarchies. *Frontiers in psychology*, 9, 1875. <https://doi:10.3389/fpsyg.2018.01875>
- Chen, H., & Chen, J. (2016). Retrofitting non-cognitive-diagnostic reading assessment under the generalized DINA model framework. *Language Assessment Quarterly*, 13(3), 218-230.
- Chen, J., & de la Torre, J. (2013). A general cognitive diagnosis model for expert-defined polytomous attributes. *Applied Psychological Measurement*, 37(6), 419–437. <https://doi:10.1177/0146621613479818>
- Chen, J., & de la Torre, J. (2018). Introducing the general polytomous diagnosis modeling framework. *Frontiers in psychology*, 9, 1474. <https://doi:10.3389/fpsyg.2018.01474>
- Cheng, Y. (2009). When cognitive diagnosis meets computerized adaptive testing: CD-CAT. *Psychometrika*, 74(4), 619–632. <https://doi:10.1007/s11336-009-9123-2>
- Chiu, C. Y. (2013). Statistical refinement of the Q-matrix in cognitive diagnosis. *Applied Psychological Measurement*, 37(8), 598-618.
- Chiu, C.Y. (2013). Statistical refinement of the Q-matrix in cognitive diagnosis. *Applied Psychological Measurement*, 37(8), 598–618.
- Chiu, C.-Y., Douglas, J. A., & Li, X. (2009). Cluster analysis for cognitive diagnosis: Theory and applications. *Psychometrika*, 74(4), 633–665.

- Culpepper, S. A. (2019). An exploratory diagnostic model for ordinal responses with binary attributes: Identifiability and estimation. *Psychometrika*, 84(4), 921–940.
- Çağlak, S. (2016). Comparison of several small sample equating methods under the NEAT design. *Turkish Journal of Education*, 5(3), 96-118.
- Çörtük, M., & Sinan, A. (2023). Çok Kategorili Puanlanan Maddelerden Oluşan Testlerde Klasik Test Kuramı ve Madde Tepki Kuramı'na Dayalı Test Eşitleme Yöntemlerinin Karşılaştırılması. *Turkish Academic Research Review*, 8(4), 1429-1439.
- de la Torre, J. (2009). A cognitive diagnosis model for cognitively based multiple-choice options. *Applied Psychological Measurement*, 33(3), 163-183.
- de la Torre, J. (2011). The generalized DINA model framework. *Psychometrika*, 76, 179-199.
- de la Torre, J., & Chiu, C. Y. (2016). A general method of empirical Q-matrix validation. *Psychometrika*, 81, 253-273.
- de la Torre, J., & Douglas, J. A. (2004). Higher-order latent trait models for cognitive diagnosis. *Psychometrika*, 69(3), 333–353. <https://doi:10.1007/bf02295640>
- de la Torre, J., & Douglas, J. A. (2008). Model evaluation and multiple strategies in cognitive diagnosis: An analysis of fraction subtraction data. *Psychometrika*, 73(4), 595–624. <https://doi:10.1007/s11336-008-9063-2>
- de la Torre, J., & Karelitz, T. M. (2009). Impact of diagnosticity on the adequacy of models for cognitive diagnosis under a linear attribute structure: A simulation study. *Journal of Educational Measurement*, 46(4), 450-469.
- de la Torre, J., & Lee, Y.-S. (2013). Evaluating the wald test for item-level comparison of saturated and reduced models in cognitive diagnosis. *Journal of Educational Measurement*, 50(4), 355–373. <https://doi:10.1111/jedm.12022>

- de la Torre, J., & Minchen, N. (2014). Cognitively diagnostic assessments and the cognitive diagnosis model framework. *Psicología Educativa*, 20(2), 89-97.
- Demirus, K. B., & Gelbal, S. (2016). Ortak maddelerin deęişen madde fonksiyonu gösterip göstermemesi durumunda test eşitlemeye etkisinin farklı yöntemlerle incelenmesi. *Journal of Measurement and Evaluation in Education and Psychology*, 7(1), 182-201.
- DiBello, L. V., & Stout, W. (2007). Guest editors' introduction and overview: IRT-based cognitive diagnostic models and related methods. *Journal of Educational Measurement*, 285-291.
- DiBello, L. V., Henson, R. A., & Stout, W. F. (2015). A family of generalized diagnostic classification models for multiple choice option-based scoring. *Applied Psychological Measurement*, 39(1), 62–79. <https://doi:10.1177/0146621614561315>
- Dogan, E., & Tatsuoka, K. (2008). An international comparison using a diagnostic testing model: Turkish students' profile of mathematical skills on TIMSS-R. *Educational Studies in Mathematics*, 68, 263-272.
- Doğuyurt, M. F., & Şeref, T. A. N. (2023). Test eşitlemede yerel bağımsızlık varsayımının ihlalinin delta ve bootstrap eşitleme hatalarına etkisinin çeşitli deęişkenlere göre incelenmesi. *Uludağ Üniversitesi Eğitim Fakültesi Dergisi*, 36(1), 23-50.
- Dorans, N. J., & Holland, P. W. (2000). Population invariance and the equatability of tests: Basic theory and the linear case. *Journal of educational measurement*, 37(4), 281-306.
- Dorans, N. J., Moses, T. P., & Eignor, D. R. (2010). Principles and practices of test score equating. *ETS Research Report Series*, 2010(2), i-41.
- Eid, G. K. (2005). The effects of sample size on the equating of test items. *Education*, 126(1).

- Embretson, S. E. (2013). Multidimensional latent trait models in measuring fundamental aspects of intelligence. In *Human Abilities* (pp. 117-132). Psychology Press.
- Erdoğan, M. (2023). *Diagnosing students' mastery profile of algebraic thinking ability: An application of mixed hierarchically structured cognitive diagnostic models* (Doctoral dissertation). Middle East Technical University, Ankara
- Fang, G., Liu, J., & Ying, Z. (2019). On the identifiability of diagnostic classification models. *Psychometrika*, *84*(1), 19–40. <https://doi:10.1007/s11336-018-09658-x>
- Fitzpatrick, A. R., & Yen, W. M. (2001). The effects of test length and sample size on the reliability and equating of tests composed of constructed-response items. *Applied Measurement in Education*, *14*(1), 31-57.
- George, A. C., Robitzsch, A., Kiefer, T., Groß, J., & Ünlü, A. (2016). The R package CDM for cognitive diagnosis models. *Journal of Statistical Software*, *74*, 1-24.
- George, D. & Mallery, M. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference*, 17.0 update (10a ed.) Boston: Pearson.
- Gierl, M. J., & Cui, Y. (2008). Defining characteristics of diagnostic classification models and the problem of retrofitting in cognitive diagnostic assessment.
- Gierl, M. J., & Leighton, J. P. (2007). Directions for future research in cognitive diagnostic assessment. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 341–351). Cambridge University Press.
- Gierl, M. J., & Leighton, J. P. (2007). Directions for future research in cognitive diagnostic assessment. In J. P. Leighton & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 341–351). Cambridge University Press.
- Gierl, M. J., Leighton, J. P., & Hunka, S. (2007). Using the attribute hierarchy method to make diagnostic inferences about examinees' cognitive skills. In J. P. Leighton &

- M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education: Theory and applications* (pp. 242–274). Cambridge University Press.
- González, J., & Wiberg, M. (2017). Applying test equating methods. *Cham: Springer International Publishing*.
- González, J., & Wiberg, M. (2017). Applying test equating methods. *Cham: Springer International Publishing*.
- Gorin, J. S., & Svetina, D. (2012). Cognitive psychometric models as a tool for reading assessment engineering. *Reaching an understanding: Innovations in how we view reading assessment*, 169-183.
- Gübeş, N., & Uyar, S. (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.
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2022). *A Primer on Partial Least Squares Structural Equation Modeling (PLS-SEM)* (3 ed.). Thousand Oaks, CA: Sage.
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. Routledge.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.
- Heller, J., Stefanutti, L., Anselmi, P., & Robusto, E. (2015). On the link between cognitive diagnostic models and knowledge space theory. *Psychometrika*, 80, 995-1019.
- Henson, R. A., Templin, J. L., & Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika*, 74, 191-210.

- Henson, R., & Douglas, J. (2005). Test construction for cognitive diagnosis. *Applied Psychological Measurement, 29*(4), 262-277.
- Henson, R., Templin, J., & Douglas, J. (2007). Using efficient model-based sum-scores for conducting skills diagnoses. *Journal of Educational Measurement, 44*(4), 361-376.
- Holland, P. W., Dorans, N. J., & Petersen, N. S. (2006). 6 Equating Test Scores. *Handbook of statistics, 26*, 169-203.
- Hong, H., Wang, C., Lim, Y. S., & Douglas, J. (2015). Efficient models for cognitive diagnosis with continuous and mixed-type latent variables. *Applied Psychological Measurement, 39*(1), 31–43. <https://doi:10.1177/0146621614524981>
- Hu, T., Yang, J., Wu, R., & Wu, X. (2021). An international comparative study of students' scientific explanation based on cognitive diagnostic assessment. *Frontiers in Psychology, 12*, 795497.
- Huebner, A., & Wang, C. (2011). A note on comparing examinee classification methods for cognitive diagnosis models. *Educational and Psychological Measurement, 71*(2), 407–419. <https://doi:10.1177/0013164410388832>
- Huebner, A., Finkelman, M. D., & Weissman, A. (2018). Factors affecting the classification accuracy and average length of a variable-length cognitive diagnostic computerized test. *Journal of Computerized Adaptive Testing, 6*(1).
- Huff, K., & Goodman, D. P. (2007). The demand for cognitive diagnostic assessment. In J. P. Leighton, & M. J. Gierl (Eds.), *Cognitive diagnostic assessment for education* (pp. 19–60). Cambridge University Press.
- Huff, K., & Goodman, D. P. (2007). The demand for cognitive diagnostic assessment. *Cognitive diagnostic assessment for education: Theory and applications*, 19-60.
- Izsák, A., & Templin, J. (2016). Coordinating conceptualizations of mathematical knowledge with psychometric models. *Journal for Research in Mathematics Education, Monograph, 15*, 5-30.

- İnal, H., & Akin Arıkan, Ç. (2017). An investigation of group invariance in test equating according to gender. *Journal of Measurement and Evaluation in Education and Psychology*, 8(1), 128-145.
- Jia, B., & Zhu, Z. (2021). International comparative study of statistics learning trajectories based on PISA data on cognitive diagnostic models. *Frontiers in Psychology*, 12, 657858.
- Johnson, M. S., & Sinharay, S. (2018). Measures of agreement to assess attribute-level classification accuracy and consistency for cognitive diagnostic assessments. *Journal of Educational Measurement*, 55(4), 635–664.
- Jurich, D. P., & Bradshaw, L. P. (2014). An illustration of diagnostic classification modeling in student learning outcomes assessment. *International Journal of Testing*, 14(1), 49-72.
- Kan, A. (2011). Test eşitleme: OKS testlerinin istatistiksel eşitliğinin sınanması. *Eğitim ve Bilim*, 36(160).
- Kaplan, M., de La Torre, J., & Barrada, J. R. (2015). New item selection methods for cognitive diagnosis computerized adaptive testing. *Applied Psychological Measurement*, 39(3), 167–188.
- Kaput, J. (2008). What is algebra? What is algebraic reasoning? In J. Kaput, D. W. Carraher, & M. L. Blanton (Eds.), *Algebra in the early grades* (pp. 5-18). Routledge.
- Ketabi, S., Alavi, S. M., & Ravand, H. (2021). Diagnostic Test Construction: Insights from Cognitive Diagnostic Modeling. *International Journal of Language Testing*, 11(1), 22-35.
- Kilmen, S., & Demirtasli, N. (2012). Comparison of test equating methods based on item response theory according to the sample size and ability distribution. *Procedia-Social and Behavioral Sciences*, 46, 130-134.

- Kim, S., Walker, M. E., & McHale, F. (2010). Comparisons among designs for equating mixed-format tests in large-scale assessments. *Journal of Educational Measurement, 47*(1), 36-53.
- Kline, P. (2015). *A handbook of test construction (psychology revivals): introduction to psychometric design*. Routledge.
- Koçak, D. (2020). The Effect of Chance Success on Equalization Error in Test Equation Based on Classical Test Theory. *International Journal of Progressive Education, 16*(2).
- Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling, and linking* (3rd ed). Springer
- Kolen, M. J., & Brennan, R. L. (2013). *Test equating: Methods and practices*. Springer Science & Business Media.
- Kuo, B. C., Chen, C. H., Yang, C. W., & Mok, M. M. C. (2016). Cognitive diagnostic models for tests with multiple-choice and constructed-response items. *Educational Psychology, 36*(6), 1115-1133.
- Lee, Y. S., & Luna-Bazaldua, D. A. (2019). How to conduct a study with diagnostic models. *Handbook of Diagnostic Classification Models: Models and Model Extensions, Applications, Software Packages, 525-545*.
- Leighton, J. P., Gierl, M. J., & Hunka, S. (2004). The attribute hierarchy model: An approach for integrating cognitive theory with assessment practice. *Journal of Educational Measurement, 41*, 205–236.
- Leighton, J., & Gierl, M. (Eds.). (2007). *Cognitive diagnostic assessment for education: Theory and applications*. Cambridge University Press.
- Li, H., & Suen, H. K. (2013). Constructing and validating a Q-matrix for cognitive diagnostic analyses of a reading test. *Educational Assessment, 18*(1), 1-25.

- Lim, Y. S., & Drasgow, F. (2017). Nonparametric calibration of item-by-attribute matrix in cognitive diagnosis. *Multivariate behavioral research*, 52(5), 562–575. <https://doi:10.1080/00273171.2017.1341829>
- Liu, J., & Low, A. C. (2008). A comparison of the kernel equating method with traditional equating methods using SAT® data. *Journal of Educational Measurement*, 45(4), 309-323.
- Liu, Q., Wu, R., Chen, E., Xu, G., Su, Y., Chen, Z., & Hu, G. (2018). Fuzzy cognitive diagnosis for modelling examinee performance. *ACM Transactions on Intelligent Systems and Technology*, 9(4), 1–26. <https://doi:10.1145/3168361>
- Liu, Y., Zhang, T., Wang, X., Yu, G., & Li, T. (2023). New development of cognitive diagnosis models. *Frontiers of Computer Science*, 17(1), 171604.
- Livingston, S. A. (2004). *Equating test scores (without IRT)*. Princeton, NJ:ETS.
- 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.
- Lohman, D. F. (2000). Complex information processing and intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 285–340). New York: Cambridge University Press.
- Lord, F. M. (2012). *Applications of item response theory to practical testing problems*. Routledge.
- Ma, W. (2019a). A diagnostic tree model for polytomous responses with multiple strategies. *The British Journal of Mathematical and Statistical Psychology*, 72(1), 61–82. <https://doi:10.1111/bmsp.12137>
- Ma, W., & de la Torre, J. (2019). Digital Module 05: Diagnostic Measurement—The G-DINA Framework <https://ncme.elevate.compartners.com>. *Educational Measurement: Issues and Practice*, 38(2), 114-115.

- Ma, W., & de La Torre, J. (2020). An empirical Q-matrix validation method for the sequential generalized DINA model. *The British Journal of Mathematical and Statistical Psychology*, 73(1), 142–163. <https://doi:10.1111/bmsp.12156>
- Ma, X., & Meng, Y. (2014). Towards personalized English learning diagnosis cognitive diagnostic modelling for efl listening. *Asian Journal of Education and e-Learning*, 2(5), 336–348.
- Madison, M. J., & Bradshaw, L. P. (2018). Assessing growth in a diagnostic classification model framework. *Psychometrika*, 83(4), 963–990. <https://doi:10.1007/s11336-018-9638-5>
- Maris, E. (1995). Psychometric latent response models. *Psychometrika*, 60, 523-547.
- Moses, T., Yang, W. L., & Wilson, C. (2007). Using Kernel Equating to Assess Item Order Effects on Test Scores. *Journal of Educational Measurement*, 44(2), 157-178.
- Mutluer, C., & Çakan, M. (2023). Comparison of Test Equating Methods Based on Classical Test Theory and Item Response Theory. *Uludağ Üniversitesi Eğitim Fakültesi Dergisi*, 36(3), 866-906.
- Nichols, P. D., Chipman, S. F., & Brennan, R. L. (Eds.). (1995). *Cognitively diagnostic assessment*. Hills-dale, NJ:Erlbaum.
- Nichols, P. D., Chipman, S. F., & Brennan, R. L. (Eds.). (2012). *Cognitively diagnostic assessment*. Routledge.
- Norris, S. P., Macnab, J. S., & Phillips, L. M. (2007). Cognitive modeling of performance on diagnostic achievement tests. *Cognitive diagnostic assessment for education: Theory and applications*, 61-84.
- Nussbeck, F. W., & Eid, M. (2015). Multimethod latent class analysis. *Frontiers in Psychology*, 6, 1332.

- Özdemir, B. (2017). Equating TIMSS Mathematics Subtests with Nonlinear Equating Methods Using NEAT Design: Circle-Arc Equating Approaches. *International Journal of Progressive Education*, 13(2), 116-132.
- Özsoy, Ş. N., & Kilmen, S. (2023). Comparison of Kernel equating methods under NEAT and NEC designs. *International Journal of Assessment Tools in Education*, 10(1), 56-75.
- Öztürk Gübeş, N., & Kelecioğlu, H. (2015). Farklı Test Eşitleme Yöntemlerinin Eşitlik Özelliği Ölçütüne Göre Karşılaştırılması. *Journal of Faculty of Educational Sciences*, 48(1).
- Pektaş, S., & Kılınç, M. (2016). PISA 2012 matematik testlerinden iki kitapçığın gözlenen puan eşitleme yöntemleri ile eşitlenmesi. *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi*, 1(40).
- Qin, H., & Guo, L. (2023). Using machine learning to improve Q-matrix validation. *Behavior Research Methods*, 1-20.
- R Core Team (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Ravand, H. (2016). Application of a cognitive diagnostic model to a high-stakes reading comprehension test. *Journal of Psychoeducational Assessment*, 34(8), 782-799.
- Ravand, H., & Robitzsch, A. (2015). Cognitive diagnostic modeling using R. *Practical Assessment, Research, and Evaluation*, 20(1), 11.
- Revelle, W., & Revelle, M. W. (2015). Package 'psych'. *The comprehensive R archive network*, 337(338).
- Ricker, K. L., & von Davier, A. A. (2007). The impact of anchor test length on equating results in a nonequivalent groups design. *ETS Research Report Series*, 2007(2), i-19.

- Rupp, A. A. (2007). The answer is in the question: A guide for describing and investigating the conceptual foundations and statistical properties of cognitive psychometric models. *International Journal of Testing*, 7(2), 95-125.
- Rupp, A. A., & Leighton, J. P. (Eds.). (2016). *The handbook of cognition and assessment: frameworks, methodologies, and applications*. Chichester UK and Hoboken NJ: Wiley Blackwell.
- Rupp, A. A., & Mislevy, R. J. (2007). Cognitive foundations of structured item response models. *Cognitive diagnostic assessment for education: Theory and applications*, 205-241.
- Rupp, A. A., & Templin, J. (2008). Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. *Measurement: Interdisciplinary Research and Perspectives*, 6(4), 219-262.
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford Press.
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement: Theory, methods, and applications*. Guilford press.
- Ryan, J., & Brockmann, F. (2009). A Practitioner's Introduction to Equating with Primers on Classical Test Theory and Item Response Theory. *Council of Chief State School Officers*.
- Salmaner Doğan, R., & Tan, Ş. (2022). Madde Tepki Kuramında Eşitleme Hatalarının Belirlenmesinde Kullanılan Delta ve Bootstrap Yöntemlerinin Çeşitli Değişkenlere Göre İncelenmesi. *Gazi University Journal of Gazi Educational Faculty (GUJGEF)*, 42(2).
- Santos, K. C. P., de la Torre, J., & von Davier, M. (2020). Adjusting person fit index for skewness in cognitive diagnosis modeling. *Journal of Classification*, 37, 399-420.

- Sen, S., & Arican, M. (2015). A diagnostic comparison of Turkish and Korean students' mathematics performances on the TIMSS 2011 assessment. *Journal of Measurement and Evaluation in Education and Psychology*, 6(2).
- Sen, S., & Bradshaw, L. (2017). Comparison of relative fit indices for diagnostic model selection. *Applied Psychological Measurement*, 41(6), 422–438. <https://doi:10.1177/0146621617695521>
- Sessoms, J., & Henson, R. A. (2018). Applications of diagnostic classification models: A literature review and critical commentary. *Measurement: Interdisciplinary Research and Perspectives*, 16(1), 1–17. <https://doi:10.1080/15366367.2018.1435104>
- Shu-Liang, D. I. N. G. (2019). Q matrix and its applications in cognitive diagnosis. *Journal of Psychological Science*, (3), 739.
- Sinharay, S., & Holland, P. (2006). Choice of anchor test in equating. *ETS Research Report Series*, 2006(2), i-43.
- Sinharay, S., & Holland, P. W. (2007). Is it necessary to make anchor tests mini-versions of the tests being equated or can some restrictions be relaxed?. *Journal of Educational Measurement*, 44(3), 249-275.
- Sorrel, M. A., de La Torre, J., Abad, F. J., & Olea, J. (2017). Two-step likelihood ratio test for item-level model comparison in cognitive diagnosis models. *Methodology*, 13(Supplement 1), 39–47. <https://doi:10.1027/1614-2241/a000131>
- Stephens, A. C., Ellis, A. B., Blanton, M., & Brizuela, B. M. (2017). In J. Cai (Ed.), Algebraic thinking in the elementary and middle grades. *Compendium for Research in Mathematics Education* (pp. 299-328). National Council of Teachers of Mathematics.
- Sternberg, R. J. (1977). *Intelligence, information processing and analogical reasoning: The componential analysis of humanabilities*. New York: Wiley.

- Sun, Y., & Suzuki, M. (2013). Diagnostic assessment for improving teaching practice. *International Journal of Information and Education Technology*, 3(6), 607.
- Suna, H. T., & Tan, Ş. (2017). Grup Değişmezliği Özelliğinin Farklı Eşitleme Yöntemlerinde İncelenmesi. *Karaelmas Eğitim Bilimleri Dergisi*, 5(2), 316-327.
- Svetina, D., Gorin, J. S., & Tatsuoka, K. K. (2011). Defining and comparing the reading comprehension construct: A cognitive-psychometric modeling approach. *International journal of testing*, 11(1), 1-23.
- Tan, Ş. (2015). Küçük örneklerde beta4 ve polynomial loglineer öndüzgünleştirme ve kübik eğri sondüzgünleştirme metotlarının uygunluğu. *Gazi Üniversitesi Gazi Eğitim Fakültesi Dergisi*, 35(1), 123-151.
- Tatsuoka, C., Clements, D. H., Sarama, J., Izsák, A., Orrill, C. H., de la Torre, J., Khasanova, E., & Tatsuoka, K. K. (2016). Chapter 4: Developing Workable Attributes for Psychometric Models Based on the Q-Matrix. *Journal for Research in Mathematics Education, Monograph*, 15, 73–96.
- Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction to the rule space method*. Routledge.
- Tatsuoka, K. K. (2009). *Cognitive assessment: An introduction to the rule space method*. Routledge.
- Tatsuoka, K. K., Corter, J. E., & Tatsuoka, C. (2004). Patterns of diagnosed mathematical content and process skills in TIMSS-R across a sample of 20 countries. *American educational research journal*, 41(4), 901-926.
- Templin, J. L., & Henson, R. A. (2006). Measurement of psychological disorders using cognitive diagnosis models. *Psychological methods*, 11(3), 287.
- Templin, J., & Bradshaw, L. (2013). Measuring the reliability of diagnostic classification model examinee estimates. *Journal of Classification*, 30(2), 251–275.
<https://doi.org/10.1007/s00357-013-9129-4>

- Tjoe, H., & de la Torre, J. (2013). Designing cognitively-based proportional reasoning problems as an application of modern psychological measurement models. *Journal of Mathematics Education*, 6(2), 17-26.
- Tjoe, H., & de la Torre, J. (2014). The identification and validation process of proportional reasoning attributes: An application of a cognitive diagnosis modeling framework. *Mathematics Education Research Journal*, 26, 237-255.
- Tu, D., Gao, X., Wang, D., & CAI, Y. (2017). A new measurement of internet addiction using diagnostic classification models. *Frontiers in psychology*, 8, 1768. <https://doi:10.3389/fpsyg.2017.01768>
- Tu, D., Wang, S., Cai, Y., Douglas, J., & Chang, H. H. (2019). Cognitive diagnostic models with attribute hierarchies: Model estimation with a restricted Q-matrix design. *Applied Psychological Measurement*, 43(4), 255-271.
- Tu, D., Zheng, C., CAI, Y., Gao, X., & Wang, D. (2018). A polytomous model of cognitive diagnostic assessment for graded data. *International Journal of Testing*, 18(3), 231–252. <https://doi:10.1080/15305058.2017.1396465>
- Ulitzsch, E., vonDavier, M., & Pohl, S. (2020). A hierarchical latent response model for inferences about examinee engagement in terms of guessing and item-level non-response. *British Journal of Mathematical and Statistical Psychology*, 73, 83-112.
- Vermunt, J. K. (2001). The use of restricted latent class models for defining and testing nonparametric and parametric item response theory models. *Applied Psychological Measurement*, 25(3), 283-294.
- 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.

- von Davier, M., & Lee, Y. S. (2019). Handbook of diagnostic classification models. Cham: Springer International Publishing.
- vonDavier, M., & Lee, Y. S. (2019). Handbook of diagnostic classification models. Cham: Springer International Publishing.
- Wang, C. (2013). Mutual information item selection method in cognitive diagnostic computerized adaptive testing with short test length. *Educational and psychological measurement, 73*(6), 1017-1035.
- Wang, S., Yang, Y., Culpepper, S. A., & Douglas, J. A. (2018). Tracking skill acquisition with cognitive diagnosis models: A higher-order, hidden Markov model with covariates. *Journal of Educational and Behavioral Statistics, 43*(1), 57-87.
- Wang, S., Zhang, M., & You, S. (2020). A comparison of IRT observed score kernel equating and several equating methods. *Frontiers in psychology, 11*, 512026.
- Wang, W., Song, L., Chen, P., Meng, Y., & Ding, S. (2015). Attribute-level and pattern-level classification consistency and accuracy indices for cognitive diagnostic assessment. *Journal of Educational Measurement, 52*(4), 457-476.
- Wiberg, M., & Bränberg, K. (2015). Kernel equating under the non-equivalent groups with covariates design. *Applied Psychological Measurement, 39*(5), 349-361.
- Wickham, H., François, R., Henry, L., Müller, K., & Wickham, M. H. (2023). *dplyr: A Grammar of Data Manipulation. R package version, 1.1.4*, <https://cran.r-project.org/web/packages/dplyr/dplyr.pdf>
- Wooldbeck, T. (1998). "Basic concept in modern methods of test equating." *Paper presented at the annual meeting of the Southwest Psychological Association New Orleans, L. A.*
- Xin, T., & Zhang, J. (2015). Local equating of cognitively diagnostic modeled observed scores. *Applied Psychological Measurement, 39*(1), 44-61.

- Xu, G. (2017). Identifiability of restricted latent class models with binary responses. *The Annals of Statistics*, 45(2), 675–707. <https://doi:10.1214/16-AOS1464>
- Xu, G. (2019). Identifiability and cognitive diagnosis models. *Handbook of Diagnostic Classification Models: Models and Model Extensions, Applications, Software Packages*, 333-357.
- Xu, G., & Shang, Z. (2018). Identifying latent structures in restricted latent classmodels. *Journal of the American Statistical Association*, 113(523), 1284-1295.
- Xu, X., & vonDavier, M. (2019). Modeling structured multiple classification latent classes in multiple populations. *Advances in Latent Class Analysis: A Festschrift in Honor of C. Mitchell Dayton*, 239.
- Yang, W. L., & Houang, R. T. (1996). The Effect of Anchor Length and Equating Method on the Accuracy of Test Equating: Comparisons of Linear and IRT-Based Equating Using an Anchor-Item Design.
- Yigit, H. D., Sorrel, M. A., & de La Torre, J. (2019). Computerized adaptive testing for cognitively based multiple-choice data. *Applied Psychological Measurement*, 43(5), 388–401. <https://doi:10.1177/0146621618798665>
- Zhan, P., Wang, W.-C., & Li, X. (2019). A partial mastery, higher-order latent structural model for polytomous attributes in cognitive diagnostic assessments. *Journal of Classification*, 30(2), 195. <https://doi:10.1007/s00357-019-09323-7>

EK-A: Niceliksel Mantık Yürütme (NMY) Becerisinin Denk Grup (DG) Desenine Göre Eşitleme İncelemeleri

NMY Becerisinin Test A ve Test B için DG Deseninde Eşitleme Uygulamaları

i. Ortalama, Doğrusal ve Eşit Yüzdelikli Eşitleme

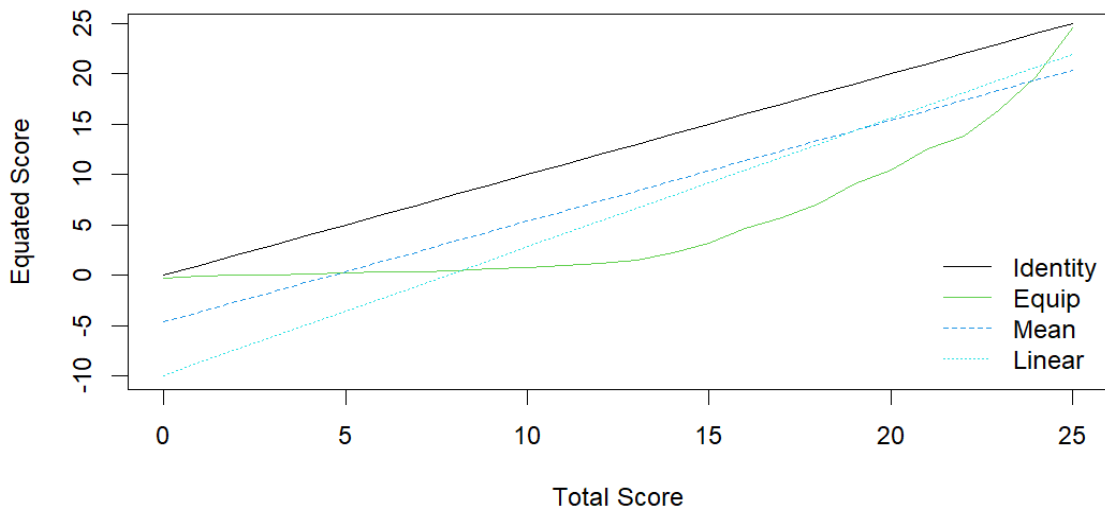
Tablo 64

DG Deseni NMY Becerisi Ortalama Eşitleme Puanı

<i>Ham Puan</i>	<i>Ortalama Eşitleme</i>	<i>Doğrusal Eşitleme</i>	<i>Eşit Yüzdelikli Eşitleme</i>
0	-4,62	-9,91	-0,31
1	-3,62	-8,64	-0,09
2	-2,62	-7,37	-0,02
3	-1,62	-6,09	0,05
4	-0,62	-4,82	0,13
5	0,38	-3,54	0,21
.	.	.	.
.	.	.	.
.	.	.	.
20	15,38	15,57	10,45
21	16,38	16,84	12,55
22	17,38	18,12	13,76
23	18,38	19,39	16,43
24	19,38	20,66	19,73
25	20,38	21,94	24,57

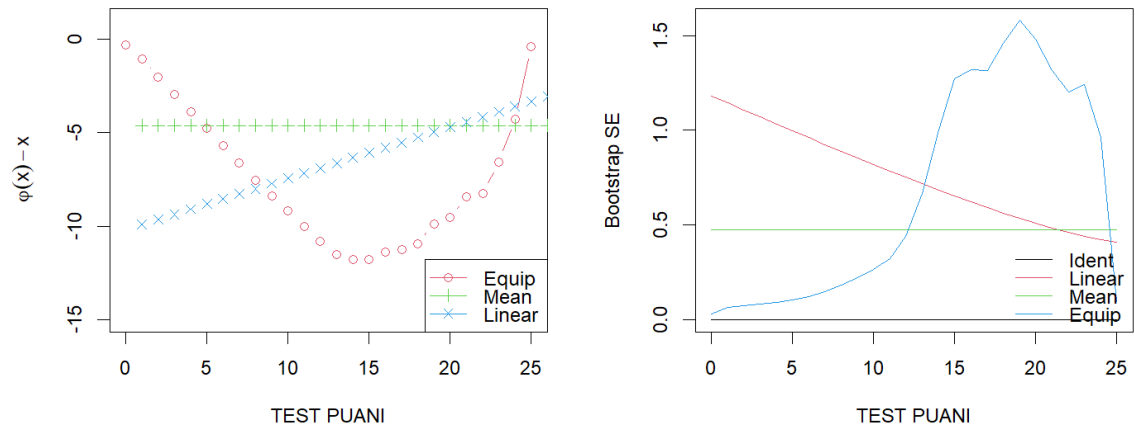
Şekil 46

DG Deseni NMY Becerisi Eşitlenmiş Puanlar

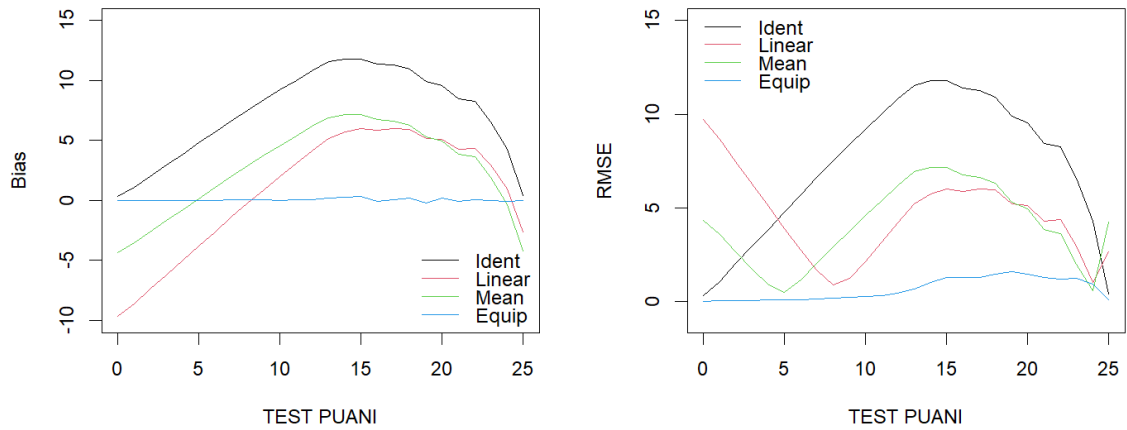


Şekil 47

DG Deseni NMY Becerisi Eşitlenmiş Puan Fark ve Bootstrap Standart Hataları

**Şekil 48**

DG Deseni NMY Becerisi Eşitlenmiş Puan Yanlılık ve RMSE Değerleri

**Tablo 65**

DG Deseni NMY Becerisi Eşitlenmiş Puan Bootstrap Hata, Yanlılık ve RMSE Değerleri

	se	se_w	bias	bias_w	rmse	rmse_w
i	0,00	0,00	8,16	6,22	8,16	6,22
lin	0,79	0,61	4,98	4,19	5,04	4,23
mean	0,47	0,47	4,54	4,32	4,57	4,35
equip	0,87	0,78	0,13	0,10	0,87	0,78

ii. Kernel Eşitleme (AxB / DG / Kernel)

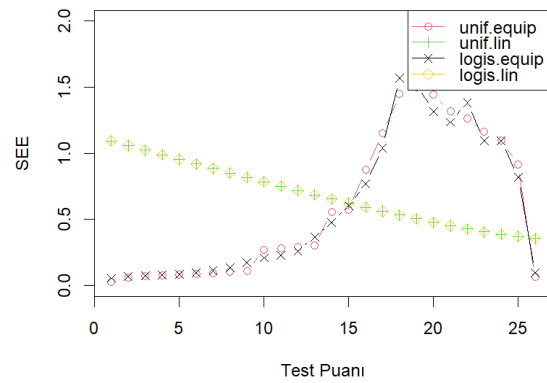
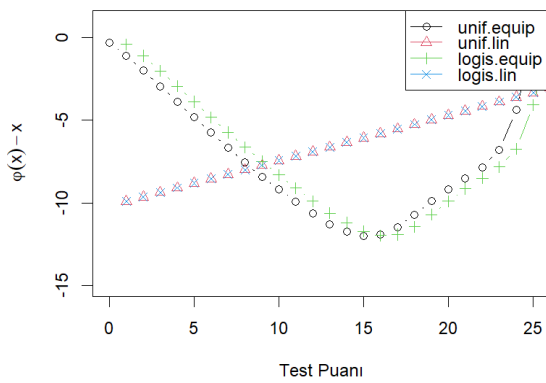
Tablo 66

DG Deseni NMY Becerisi Logistic Kernel EYE ve Doğrusal Eşitleme Puan ve Hataları

	eqYx	SEEX	eqYxLIN	SEEXLIN
0	-0,43	0,055	-9,91	1,094
1	-0,11	0,068	-8,64	1,059
2	-0,02	0,074	-7,37	1,024
3	0,05	0,079	-6,09	0,989
4	0,11	0,086	-4,82	0,954
5	0,18	0,096	-3,54	0,920
.
.
.
20	10,86	1,235	15,57	0,453
21	12,50	1,381	16,84	0,429
22	14,17	1,096	18,12	0,407
23	16,26	1,095	19,39	0,388
24	19,93	0,817	20,66	0,371
25	24,53	0,098	21,94	0,357

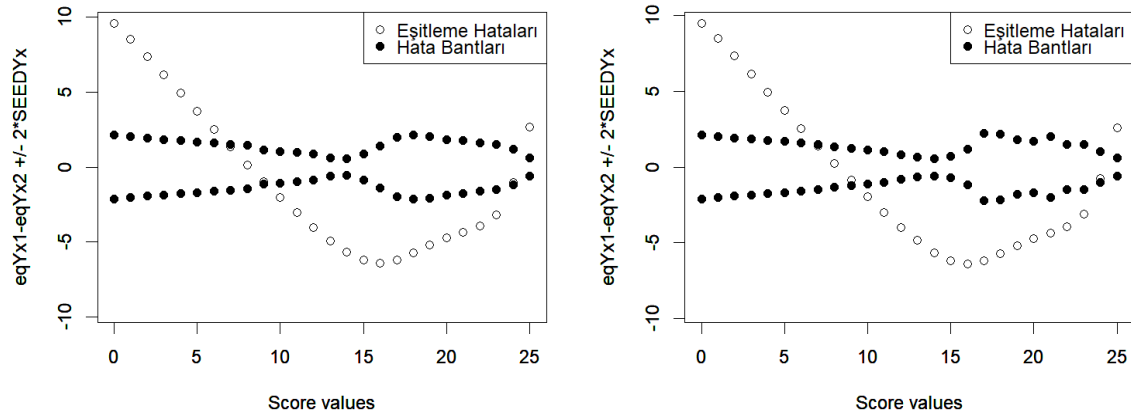
Şekil 49

DG Deseni NMY Becerisi Kernel Eşitleme Puan Fark ve Standart Hataları



Şekil 50

DG Deseni NMY Becerisi Logistic ve Uniform Kernel SEED Değerleri

**Tablo 67**

DG Deseni NMY Becerisi Uniform ve Logistic Kernel Yüzde Göreceli Hata Değerleri

	Unf.PRE	Logis.PRE
1	1,03	1,12
2	2,15	2,19
3	3,20	3,10
4	4,15	3,84
5	5,00	4,41
6	5,72	4,83
7	6,33	5,11
8	6,82	5,26
9	7,21	5,29
10	7,48	5,22

**EK-B: NMY Becerisinin Denk Olmayan Gruplarda Ortak Madde (DOG-OM) Desenine
Göre Eşitleme İncelemeleri**

NMY Becerisinin Test A ve Test B için DOG-OM Deseninde Eşitleme Uygulamaları

i. Doğrusal Eşitleme (AxB / DOG-OM / Doğrusal)

Tablo 68

DOG-OM Deseni NMY Becerisi Doğrusal Eşitleme Puanları

xscale	tucker	chain	levineTS	BH
0	-8,42	-2,78	9,92	-8,63
1	-7,18	-1,75	10,47	-7,38
2	-5,94	-0,73	11,01	-6,13
3	-4,69	0,30	11,55	-4,88
.
.
.
23	20,19	20,77	22,43	20,10
24	21,43	21,79	22,97	21,35
25	22,67	22,82	23,51	22,59

Tablo 69

DOG-OM Deseni NMY Becerisi Levine OS Doğrusal Eşitleme Puanı

	levineOS		levineOS
0	1,06	13	14,06
1	2,06	14	15,06
2	3,06	15	16,06
3	4,06	16	17,06
4	5,06	17	18,06
5	6,06	18	19,06
6	7,06	19	20,06
7	8,06	20	21,06
8	9,06	21	22,06
9	10,06	22	23,06
10	11,06	23	24,06
11	12,06	24	25,06
12	13,06	25	26,06

ii. Eşit Yüzdellikli Eşitleme (AxB / DOG-OM / Eşit Yüzdellikli)

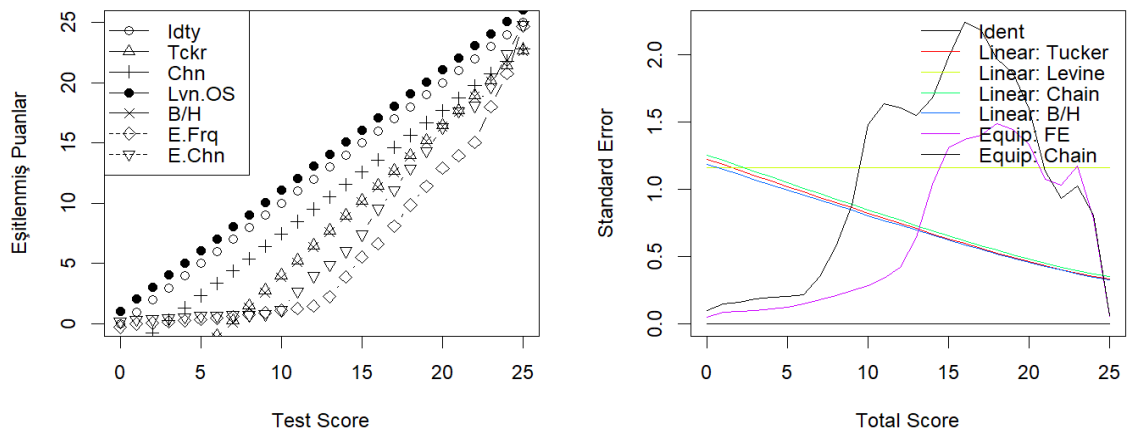
Tablo 70

DOG-OM Deseni NMY Becerisi EYE ve CE Puanları

Ham Puanlar	Freq. Est.	Chain	Ham Puanlar	Freq. Est.	Chain
0	-0,27	0,20	13	2,27	4,92
1	0,01	0,35	14	3,88	6,05
2	0,08	0,40	15	5,53	7,42
3	0,17	0,47	16	6,65	9,54
.
.
.
10	1,11	1,17	23	17,99	19,65
11	1,27	2,70	24	20,77	22,38
12	1,48	3,97	25	24,68	24,77

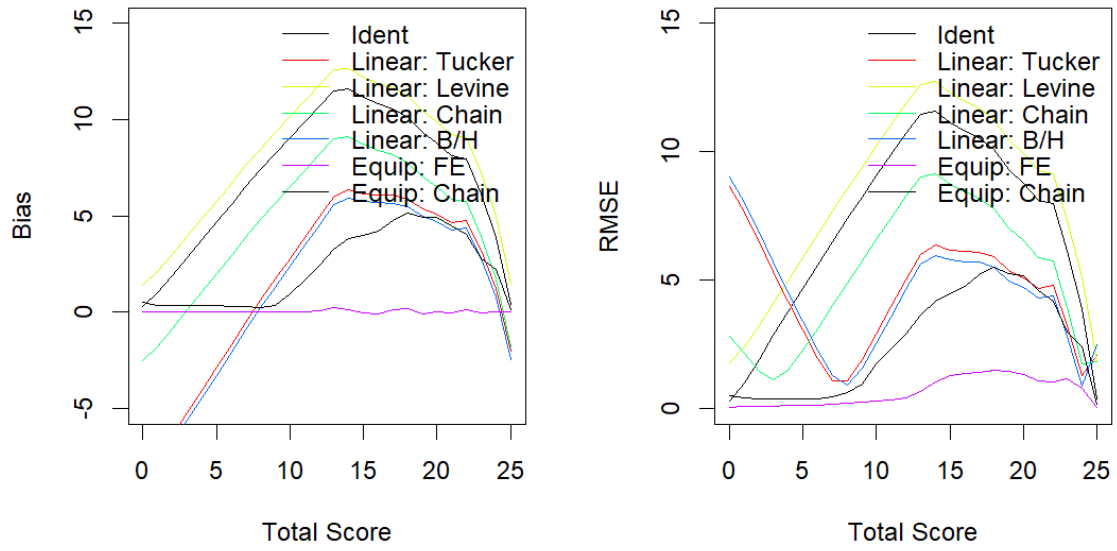
Şekil 51

DOG-OM Deseni NMY Becerisi Eşitlenmiş Puanlar ve Standart Hataları



Şekil 52

DOG-OM Deseni NMY Becerisi Eşitlenmiş Puan Yanlılık ve RMSE Değerleri

**Tablo 71**

DOG-OM Deseni GA Becerisi Eşitlenmiş Puan Bootstrap Hata, Yanlılık ve RMSE

Değerleri

	se	se_w	bias	bias_w	rmse	rmse_w
i	0,00	0,00	7,85	5,91	7,85	5,91
lt	0,67	0,49	4,87	3,98	4,91	4,01
ll	1,00	1,00	8,74	6,65	8,80	6,72
lc	0,73	0,52	5,76	4,48	5,80	4,51
lbh	0,62	0,44	4,76	3,96	4,80	3,98
ef	0,71	0,63	0,09	0,07	0,72	0,63
ec	0,99	0,78	2,80	2,38	2,97	2,50

iii. Kernel Eşitleme (AxB / DOG-OM / Kernel)

Tablo 72

DOG-OM Deseni NMY Becerisi Logistic Kernel EYE ve Doğrusal Eşitleme Puan ve

Hataları

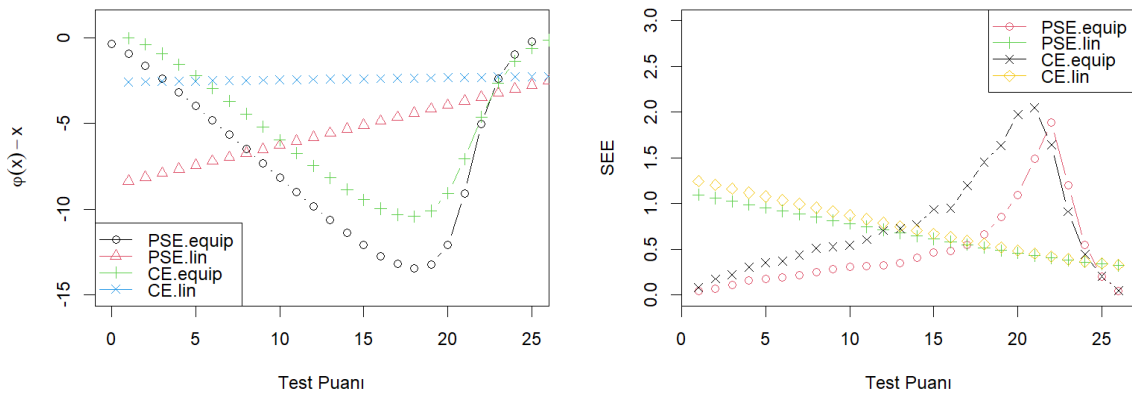
	NEAT_PSE	see.pse	NEAT_CE	see.ce
0	-0,35	0,04	-0,02	0,08
1	0,06	0,07	0,60	0,17
2	0,34	0,11	1,06	0,22
3	0,60	0,15	1,43	0,30

.
.
.
12	2,15	0,35	3,83	0,72
13	2,35	0,40	4,14	0,76
.
.
.
23	20,61	0,55	21,63	0,45
24	23,01	0,20	23,39	0,21
25	24,77	0,05	24,85	0,05

Şekil 53

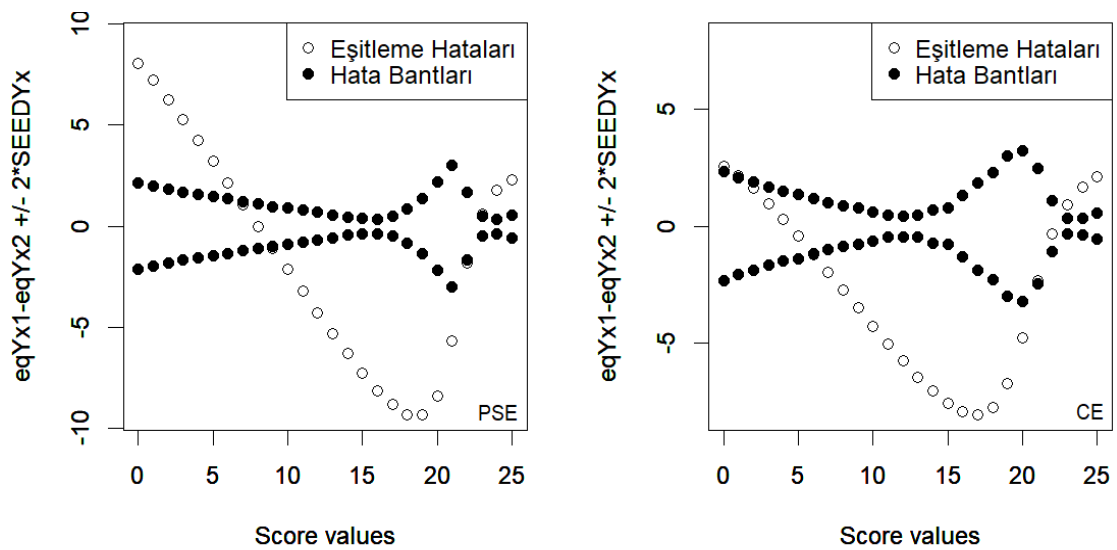
DOG-OM Deseni NMY Becerisi Kernel PSE ve CE Eşitleme Puan Fark ve Standart

Hataları



Şekil 54

DOG-OM Deseni NMY Becerisi Kernel PSE ve CE SEED Değerleri



Tablo 73*DOG-OM Deseni NMY Becerisi Kernel PSE ve CE Yüzde Göreceli Hata Değerleri*

	PSE	CE/B-ancor	CE/ancor-A
1	0,34	0,08	0,15
2	0,49	0,00	0,18
3	0,54	0,00	0,17
4	0,56	0,00	0,15
5	0,55	-0,02	0,14
6	0,51	-0,05	0,12
7	0,45	-0,09	0,11
8	0,36	-0,16	0,10
9	0,26	-0,24	0,09
10	0,13	-0,33	0,09

EK-C: NMY Becerisinin Denk Olmayan Gruplarda Ortak Değişken (DOG-OD) Desenine Göre Eşitleme İncelemeleri

NMY Becerisinin Test A ve Test B için DOG-OD Deseninde Eşitleme Uygulamaları

i. Okulların Kabul Koşulu ve Okul Türleri Değişkenine Göre

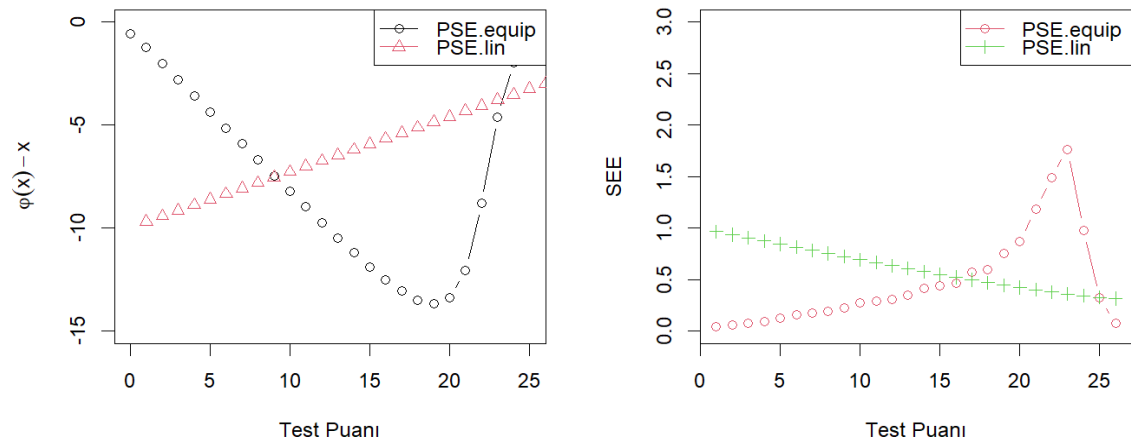
Tablo 74

DOG-OD Deseni NMY Becerisi Kernel EYE ve Doğrusal Eşitleme Puan ve Hataları

	eqYx	SEEYx	eqYxLIN	SEEYxLIN
0	-0,56	0,04	-9,69	0,97
1	-0,24	0,06	-8,42	0,94
2	-0,02	0,07	-7,15	0,90
3	0,18	0,09	-5,89	0,87
.
.
.
12	2,26	0,35	5,53	0,61
13	2,53	0,42	6,80	0,58
.
.
.
23	18,36	0,98	19,48	0,34
24	22,03	0,32	20,75	0,33
25	24,56	0,07	22,01	0,31

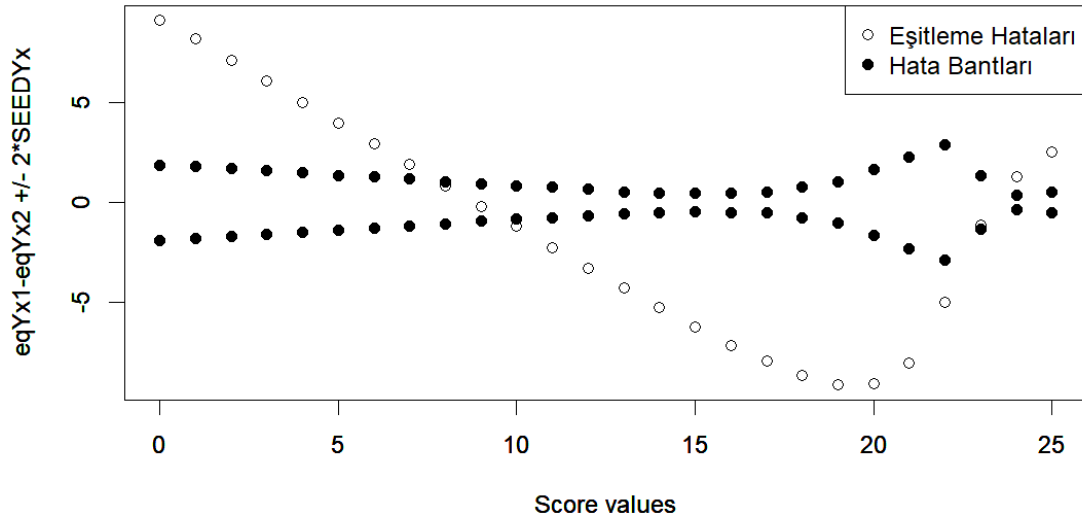
Şekil 55

DOG-OD Deseni NMY Becerisi Kernel PSE EYE ve Doğrusal Eşitleme Puan Fark ve Standart Hataları



Şekil 56

DOG-OD Deseni NMY Becerisi Kernel PSE EYE ve Doğrusal Eşitleme SEED Değerleri

**Tablo 75**

DOG-OD Deseni NMY Becerisi Kernel PSE EYE Yüzde Göreceli Hata Değerleri

	PRE
1	0,50
2	0,80
3	0,96
4	1,03
5	1,04
6	0,99
7	0,88
8	0,72
9	0,52
10	0,27

ii. Okulların Kabul Koşulu ve Cinsiyete Göre

Tablo 76

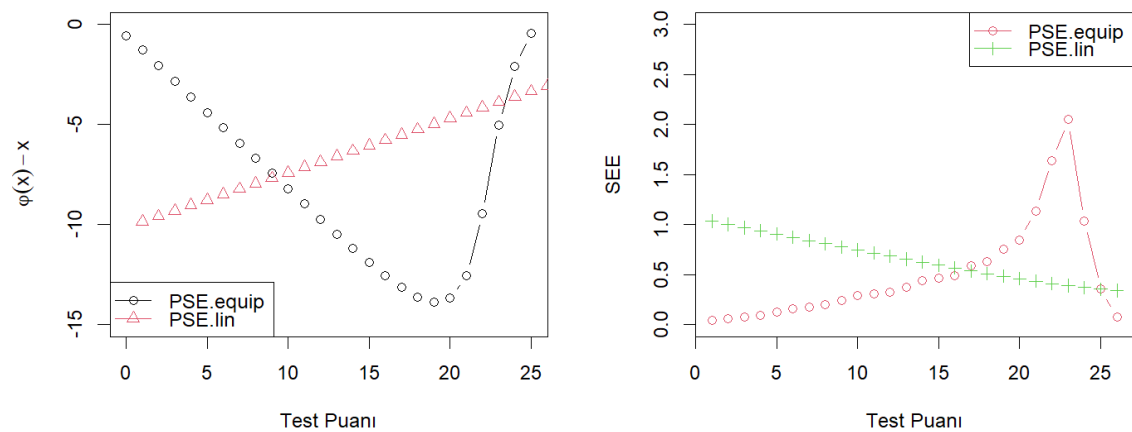
DOG-OD Deseni NMY Becerisi Kernel EYE ve Doğrusal Eşitleme Puan ve Hataları

	eqYx	SEYx	eqYxLIN	SEYxLIN
0	-0,58	0,04	-9,855	1,034
1	-0,27	0,06	-8,583	1,001
2	-0,05	0,07	-7,312	0,968
3	0,15	0,09	-6,040	0,936
.
.
.
12	2,26	0,37	5,403	0,653
13	2,52	0,44	6,674	0,623
.
.
.
22	12,54	2,05	18,117	0,389
23	17,97	1,03	19,389	0,370
24	21,90	0,35	20,660	0,353
25	24,54	0,08	21,931	0,339

Şekil 57

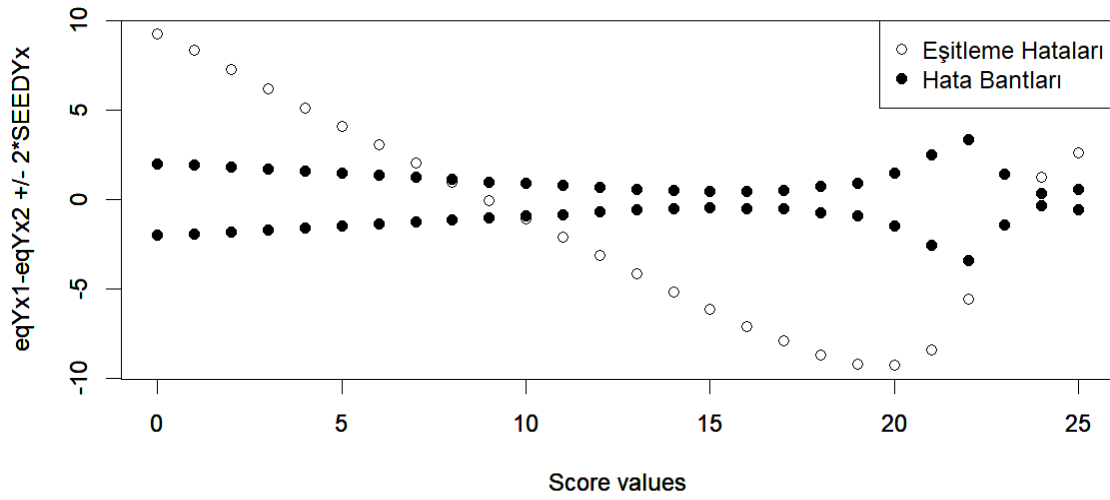
DOG-OD Deseni NMY Becerisi Kernel PSE EYE ve Doğrusal Eşitleme Puan Fark ve

Standart Hataları



Şekil 58

DOG-OD Deseni NMY Becerisi Kernel PSE EYE ve Doğrusal Eşitleme SEED Değerleri

**Tablo 77**

DOG-OD Deseni NMY Becerisi Kernel PSE EYE Yüzde Göreceli Hata Değerleri

	PRE
1	0,54
2	0,87
3	1,04
4	1,12
5	1,13
6	1,07
7	0,94
8	0,77
9	0,54
10	0,26

EK-D:Araştırma Etik Komisyon İzin Muafiyeti Formu/Araştırma Etik Komisyonu**Onay Bildirimi**

UYGULAMALI ETİK ARAŞTIRMA MERKEZİ
APPLIED ETHICS RESEARCH CENTER



ORTA DOĞU TEKNİK ÜNİVERSİTESİ
MIDDLE EAST TECHNICAL UNIVERSITY

DUMLUPINAR BULVARI 06800
ÇANKAYA ANKARA/TURKEY
T: +90 312 210 22 91
F: +90 312 210 79 59
uram@meb.odu.tr
www.egitim.meb.gov.tr

Sayı: 28620816 / 126

16 MART 2020

Konu: Değerlendirme Sonucu

Gönderen: ODTÜ İnsan Araştırmaları Etik Kurulu (İAEK)

İlgi: İnsan Araştırmaları Etik Kurulu Başvurusu

Sayın Prof.Dr. Kürşat ERBAŞ

Danışmanlığını yaptığınız Melek ERDOĞAN'ın "Öğrencilerin Cebirsel Düşünme Becerilerinin Bilişsel Tanı Modelleri ile Tanınması" başlıklı araştırması İnsan Araştırmaları Etik Kurulu tarafından uygun görülmüş ve 126 ODTU 2020 protokol numarası ile onaylanmıştır.

Saygılarımızla bilgilerinize sunarız.

Prof.Dr. Mine MISIRLISOY

Başkan

Prof. Dr. Tolga CAN

Üye

Doç.Dr. Pınar KAYGAN

Üye

Dr. Öğr. Üyesi Ali Emre TURGUT

Üye

Dr. Öğr. Üyesi Şerife SEVİNÇ

Üye

Dr. Öğr. Üyesi Müge GÜNDÜZ

Üye

Dr. Öğr. Üyesi Süreyya Özcan KABASAKAL

Üye

EK-E: MEB Araştırma İzni Onay Bildirimi

T.C.
ANKARA VALİLİĞİ
Milli Eğitim Müdürlüğü

Sayı : 14588481-605.99-E.17642271
Konu : Araştırma izni

03.12.2020

ORTA DOĞU TEKNİK ÜNİVERSİTESİNE

İlgi: a) 19.11.2020 tarihli ve 249 sayılı yazımız.
b) MEB Yenilik ve Eğitim Teknolojileri Genel Müdürlüğünün 2020/2 nolu Genelgesi.

Üniversiteniz Fen Bilimleri Enstitüsü Doktora Öğrencisi Melek ERDOĞAN'ın "Öğrencilerin Cebirsel Düşünme Becerilerinin Bilişsel Tanı Modelleri İle Tanımlanması" konulu tezi kapsamında ilçenize bağlı liselerde uygulanacak olan veri toplama araçları ilgi (b) Genelge çerçevesinde incelenmiştir.

Yapılan inceleme sonucunda, söz konusu araştırmanın Müdürlüğümüzde muhafaza edilen ölçme araçlarının; Türkiye Cumhuriyeti Anayasası, Milli Eğitim Temel Kanunu ile Türk Milli Eğitiminin genel amaçlarına uygun olarak, ilgili yasal düzenlemelerde belirtilen ilke, esas ve amaçlara aykırılık teşkil etmeyecek, eğitim-öğretim faaliyetlerini aksatmayacak şekilde okul ve kurum yöneticilerinin sorumluluğunda gönüllülük esasına göre uygulanması Müdürlüğümüzce uygun görülmüştür.

Bilgilerinizi ve gereğini rica ederim.

Turan AKPINAR
Vali a.
Milli Eğitim Müdürü

Ek:
Uygulama Araçları (18 sayfa)
Dağıtım:
Gereği:
ODTÜ
Bilgi:
Çankaya İlçe MEM

Adres: Emniyet Mah. Alparslan Türkeş Cad. 4/A Yenimahalle

Bilgi için: Emine Kocuk

Elektronik Ağ: ankara.meb.gov.tr
e-posta: isticat06@meb.gov.tr

Tel: 0 (312) 306 89 30
Faks: 0 () _____

Bu evrak güvenli elektronik imza ile imzalanmıştır. <https://evraksorgu.meb.gov.tr> adresinden fe87-0f9d-36a7-b41b-540e kodu ile teyit edilebilir.

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.

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

(İmza)

Melek ERDOĞAN

EK-G: Doktora Tez Çalışması Orijinallik Raporu

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

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

Tez Başlığı: BİLİŞSEL TANI MODELLERİ İLE ELDE EDİLEN BİREYSEL BECERİ PUANLARININ BENZERLİĞİNİN PARALEL FORMLARLA İNCELENMESİ

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
25/03/2024	170	228072	29/01/2024	%6	2266957916

Uygulanan filtreler:

- Kaynaklar hariç
- Alıntılar dâhil
- 5 kelimedenden daha az örtüşme içeren metin kısımları hariç

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

Ad Soyadı: MELEK ERDOĞAN

Öğrenci No.: N14245784

Ana Bilim Dalı: Eğitim Bilimleri

Programı: Eğitimde Ölçme ve Değerlendirme

Statüsü: Y.Lisans Doktora Bütünleşik Dr.

İmza

DANIŞMAN ONAYI

UYGUNDUR.

Prof. Dr. Burcu ATAR

EK-H: Thesis/Dissertation Originality Report

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

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

Thesis Title: EXAMINING THE COMPARABILITY OF INDIVIDUAL SKILL SCORES OBTAINED THROUGH COGNITIVE DIAGNOSTIC MODELS WITH PARALLEL FORMS

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
25/03/2024	170	228072	29/01/2024	6%	2266957916

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: MELEK ERDOĞAN

Student No.: N14245784

Department: Educational Sciences

Program: Educational Measurement and Evaluation

Status: Masters Ph.D. Integrated Ph.D.

Signature

ADVISOR APPROVAL

APPROVED
Prof. Dr. Burcu Atar

EK-I: Yayınlanma 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 yada 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 tarihinden itibaren 6 ay ertelenmiştir.⁽²⁾
- O Tezimle ilgili gizlilik kararı verilmiştir.⁽³⁾

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

(imza)

Melek ERDOĞAN

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

- (1) Madde 6.1. Lisansüstü teze ilgili patent başvurusu yapılması veya patent alma sürecinin devam etmesi durumunda, tez danışmanının önerisi ve enstitü ana bilim 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 danışmanın önerisi ve enstitü ana bilim 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 ana bilim dalının uygun görüşü üzerine enstitü veya fakülte yönetim kurulu tarafından karar verilir.

