

**TESTLERİN BOYUTLULUĞUNUN, ORTAK MADDE
FORMATININ, YETENEK DAĞILIMININ VE ÖLÇEK
DÖNÜŞTÜRME YÖNTEMLERİNİN KARMA TESTLERİN
EŞİTLENMESİNE ETKİSİ**

**THE EFFECTS OF TEST DIMENSIONALITY, COMMON
ITEM FORMAT, ABILITY DISTRIBUTION AND SCALE
TRANSFORMATION METHODS ON MIXED-FORMAT TEST
EQUATING**

Neşe ÖZTÜRK GÜBEŞ

Hacettepe Üniversitesi

Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliğinin

Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı İçin

Öngördüğü

Doktora Tezi


olarak hazırlanmıştır.

2014

Eđitim Bilimleri Enstitüsü M¼d¼rl¼đ¼'ne,

NeŒe ÖZT¼RK G¼BEŒ' in hazırladıđı "Testlerin Boyutluluđunun, Ortak Madde Formatının, Yetenek Dađılıminin ve Ölçek DönüŒt¼rme Yöntemlerinin Karma Testlerin EŒitlenmesine Etkisi" baŒlıklı bu alıŒma j¼rimiz tarafından **Eđitim Bilimleri Anabilim Dalı, Eđitimde Ölme ve Deđerlendirme Bilim Dalı'nda Doktora Tezi** olarak kabul edilmiŒtir.

BaŒkan


Prof. Dr., Œener B¼Y¼KÖZT¼RK


¼ye (DanıŒman)


Prof. Dr., H¼lya KELECIOđLU

¼ye


Do. Dr., Nuri DOđAN

¼ye


Yrd. Do. Dr., Tahsin Ođuz BAŒOKU

¼ye


Yrd. Do. Dr., Derya OBANOđLU AKTAN

ONAY

Bu tez Hacettepe ¼niversitesi Lisans¼st¼ Eđitim-Öđretim ve Sınav Yönetmeliđi'nin ilgili maddeleri uyarınca yukarıdaki j¼ri ¼yeleri tarafından/...../..... tarihinde uygun gör¼lm¼Œ ve Enstit¼ Yönetim Kurulunca/...../..... tarihinde kabul edilmiŒtir.

Prof. Dr. Berrin AKMAN
Eđitim Bilimleri Enstitüsü M¼d¼r¼

TESTLERİN BOYUTLULUĞUNUN, ORTAK MADDE FORMATININ, YETENEK DAĞILIMININ VE ÖLÇEK DÖNÜŞTÜRME YÖNTEMLERİNİN KARMA TESTLERİN EŞİTLENMESİNE ETKİSİ

Neşe ÖZTÜRK GÜBEŞ

ÖZ

Bu araştırma kapsamında boyutluluğun, ortak madde seti formatının, grupların yetenek dağılımının ve farklı ölçek dönüştürme yöntemlerinin karma testlerin eşitlenmesinde eşitlik özelliğinin korunumu üzerine etkisi incelenmiştir. Araştırma, eşdeğer olmayan gruplar ortak test deseni altında Madde Tepki Kuramı gerçek puan eşitleme (GR_PE) ve gözlenen puan eşitleme (GZ_PE) yöntemleri kullanılarak yürütülmüştür. Eşitlik özelliği, birinci-sıra eşitlik (BSE) ve ikinci-sıra eşitlik özelliği (İSE) ölçütlerine göre D_1 ve D_2 indeksleri hesaplanarak değerlendirilmiştir.

Araştırma, simülasyon verileri kullanılarak yürütülmüştür. Çalışma kapsamında, verilerinin türetilmesinde SimuMIRT, madde parametrelerinin kestiriminde PARSCALE 4.1, ölçek dönüştürme işleminde STUIRT, test puanlarının eşitlenmesinde POLYEQUATE, D_1 ve D_2 indekslerini hesaplamak için gerekli olan koşullu ortalama ve ölçmenin standart hatasını elde etmek için POLYCSEM programı kullanılmıştır. Araştırma kapsamında incelenen faktörlerin GR_PE ve GZ_PE sonuçlarının BSE ve İSE özelliğinin korunumuna etkisi 64 farklı koşul ($4 \times 2 \times 2 \times 4$) için 100 tekrar sonucu elde edilen D_1 ve D_2 indekslerinin ortalaması alınarak değerlendirilmiştir. Ayrıca tüm faktörlerin BSE ve İSE özelliklerinin korunumu ölçütüne göre eşitleme sonuçları üzerinde istatistiksel olarak manidar etkiye sahip olup olmadığı SPSS 20 programında 4-yönlü ANOVA yapılarak incelenmiştir.

Araştırmanın sonucunda BSE özelliği ölçütüne göre tüm faktörlerin, İSE özelliği ölçütüne göre ise ortak madde seti formatı dışında kalan tüm faktörlerin GR_PE ve GZ_PE sonuçları üzerinde istatistiksel olarak anlamlı etkisinin olduğu görülmüştür. Araştırmadan elde edilen bulgular şöyle özetlenebilir (1) BSE ve İSE eşitlik özellikleri en iyi tek boyutlu en kötü ise çok boyutluluğun derecesinin en yüksek olduğu koşul altında korunmuştur. (2) Karma ortak madde seti kullanılarak elde

edilen GR_PE ve GZ_PE sonuçları BSE özelliğini çoktan seçmeli madde seti kullanılan koşullara göre daha iyi korumuştur. (3) Grupların eşdeğer olduğu koşullar altında elde edilen eşitleme sonuçları eşdeğer olmadığı koşullarla karşılaştırıldığında her iki eşitlik özelliğini daha iyi korumuştur. (4) Tek boyutluluğun sağlandığı koşullarda her ne kadar moment ölçek dönüştürme yöntemleri BSE ve İSE özelliklerini korumada karakteristik eğri yöntemlerine yakın performans göstermiş olsa da özellikle çok boyutluluğun derecesinin arttığı koşullarda karakteristik eğri ölçek dönüştürme yöntemleri her iki eşitlik özelliğini korumada üstün performans göstermiştir.

Anahtar sözcükler: Test eşitleme, karma test, eşitlik özelliği, birinci-sıra eşitlik, ikinci-sıra eşitlik.

Danışman: Prof. Dr. Hülya KELEÇİOĞLU, Hacettepe Üniversitesi, Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı

THE EFFECTS OF TEST DIMENSIONALITY, COMMON ITEM FORMAT, ABILITY DISTRIBUTION AND SCALE TRANSFORMATION METHODS ON MIXED-FORMAT TEST EQUATING

Neşe ÖZTÜRK GÜBEŞ

ABSTRACT

The purpose of this study was to examine the impact of test dimensionality (unidimensional versus multidimensional), characteristics of common-item set (format representative versus non-representative), differences in group ability distributions (equivalent versus nonequivalent), and different scale transformation methods on preserving equity property with mixed-format test equating. The study was conducted by using item response theory (IRT) true score equating (TSE) and IRT observed score equating (OSE) methods under non-equivalent groups anchor test design. The equity property was evaluated based on first-order equity (FOE) and second-order equity (SOE) by calculating D_1 and D_2 indexes.

Simulated data was used in this study. The computer program SimuMIRT was used for data generation, PARSCALE 4.1 was used to estimate the parameters, STUIRT was used to do scale transformation, POLYEQUATE was used to equate test scores and lastly POLYCSEM was used to provide conditional means and standard errors of measurement which were necessary for calculating D_1 and D_2 indexes. The impact of factors on IRT true score and observed score equating results in terms of preserving FOE and SOE properties were evaluated by taking average of D_1 and D_2 indexes over 100 replications for the 64 different conditions (4x2x2x4). Also, 4-way ANOVA was performed to determine if the investigated factors had significant effects on equating results with regarding to FOE and SOE properties. ANOVA was conducted by using computer program SPSS 20.

The results showed that in terms of FOE property criteria, all investigated factors had significant effect on TSE and OSE results. Similarly, based on SOE property except for common-item characteristics all factors had significant effect on both TSE and OSE results. The major findings can be summarized as follows: (1) The FOE and SOE properties best preserved under unidimensional condition, worst preserved when the degree of multidimensionality was severe. (2) The true score

and observed score equating results which were provided by using mixed-format common item set, preserved FOE better compared to equating results which were provided only multiple-choice common item set. (3) The equivalent groups condition outperformed the nonequivalent group condition on preserving both equity properties. (4) Under unidimensional test structure, moment scale transformation methods performed as well as characteristic curve methods. However, under multidimensional test structure characteristic curve scale transformation methods performed significantly superior than moment scale transformation methods in terms of preserving FOE and SOE properties.

Keywords: Test equating, mixed-format test, equity property, first-order equity, second-order equity.

Advisor: Prof. Dr. Hülya KELEÇİOĞLU, Hacettepe University, Department of Educational Sciences, Division of Measurement and Evaluation in Education

ETİK BEYANNAMESİ

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 tüm bilgi ve sonuçları bilimsel ahlak kurallarına uygun olarak sunduğumu,
- başkalarının eserlerinden yararlanılması durumunda ilgili eserlere bilimsel normlara uygun olarak atıfta bulunduğumu,
- atıfta bulunduğum eserlerin tümünü kaynak olarak gösterdiğimi,
- kullanılan verilerde herhangi bir tahrifat yapmadığımı,
- ve bu tezin herhangi bir bölümünü bu üniversitede veya başka bir üniversitede başka bir tez çalışması olarak sunmadığımı

beyan ederim.


Neşe ÖZTÜRK GÜBEŞ

TEŞEKKÜR

Doktora tez çalışmamın başından sonuna kadar her aşamasında yanımda olan, problemlerime çözümler üreten, benimle birlikte sevinen ve heyecanlanan, yanında kendimi rahatça ifade edebildiğim tez danışmanım canım hocam Prof. Dr. Hülya KELECİOĞLU'na;

Lisansüstü eğitimime başladığım ilk günden itibaren babacan tavırlarıyla desteğini hep yanımda hissettiğim çok değerli hocam Prof. Dr. Selahattin GELBAL'a; güler yüzünü ve mütevazî kişiliğini akademisyen olarak örnek aldığım sevgili hocam Doç. Dr. Nuri DOĞAN'a; görüş ve önerileriyle tez çalışmama katkı sağlayan kıymetli hocalarım Prof. Dr. Şener BÜYÜKÖZTÜRK'e, Yrd. Doç. Dr. Derya ÇOBANOĞLU-AKTAN'a ve Yrd. Doç. Dr. Tahsin Oğuz BAŞOKÇU'ya,

Tezimin her aşamasında sorduğum her soruyu büyük bir içtenlikle yanıtlayan ve bana çözüm yolları gösteren Dr. Andrews'e; Florida State Üniversitesi'nde bulunduğum süre boyunca bana akademik danışmanlık yapan, engin alan bilgisine hayran kaldığım ve kendisinden çok şey öğrendiğim çok kıymetli hocam Dr. Paek'e;

Amerika'da bulunduğum süre boyunca hayatımı kolaylaştıran ve varlıklarıyla bana güç veren başta Şeyma ve Seyfullah TINGİR, Erkan SAYILIR, Hülya YÜREKLİ, Tingxuan LI ve Yi-jhen WU olmak üzere burada isimlerini saymadığım tüm arkadaşlarım ve hocalarıma;

Hacettepe Üniversitesi'nde görev yaptığım süre boyunca destek ve katkılarıyla yanımda olan başta sevgili arkadaşım Dr. Kubra ATALAY-KABASAKAL olmak üzere Eğitim Bilimleri Bölümü'nde görev yapan tüm meslektaşlarıma;

Hayatımın her anında yanımda olan, beni koşulsuz seven ve bugünlere gelmemi sağlayan sevgili anneme, babama ve kardeşime;

Her daim bana destek olan, nazımı ve stresimi çeken, varlığı ile hayatımı anlamlandıran, mutluluğumun kaynağı canım eşim ve meslektaşım Murat GÜBEŞ'e;

Doktora eğitimim boyunca yurt içi doktora burs olanağı sağlayan TÜBİTAK'a;

gönülden teşekkür ederim.

İÇİNDEKİLER

ÖZ	iii
ABSTRACT	v
ETİK BEYANNAMESİ	vii
TEŞEKKÜR.....	viii
İÇİNDEKİLER.....	ix
TABLolar DİZİNİ	xi
ŞEKİLLER DİZİNİ.....	xii
SİMGELER VE KISALTMALAR DİZİNİ	xiii
1. GİRİŞ.....	1
1.1. Problem Durumu	1
1.2. Araştırmanın Amacı ve Önemi.....	4
1.3. Problem Cümlesi.....	5
1.3.1. Alt Problemler.....	5
1.4. Sınırlılıklar	6
1.5. Tanımlar.....	6
1.6. Araştırmanın Kuramsal Temeli	6
1.6.1. Test Eşitleme Desenleri	7
1.6.2. Test Eşitleme Yöntemleri	9
1.6.2.1. Madde Tepki Kuramı Modelleri	9
1.6.2.2. Madde Kalibrasyonu	14
1.6.2.3. Ölçek Dönüştürme.....	14
1.6.2.4. Madde Tepki Kuramına Dayalı Test Eşitleme	17
1.6.3. Eşitleme Sonuçlarının Değerlendirilmesi.....	18
2. İLGİLİ ARAŞTIRMALAR.....	21
2.1. Karma Testlerin Eşitlenmesini Etkileyen Faktörlerin İncelendiği Araştırmalar	21
2.2. Eşitlik Özelliğinin Değerlendirme Ölçütü Olarak Alındığı Araştırmalar	25
2.3. İlgili Araştırmalar Özet	29
3. YÖNTEM	30
3.1. Araştırmanın Türü	30
3.2. Araştırma Deseni	30
3.3. Araştırma Verileri	31
3.4. Simülasyon Faktörleri ve Koşulları	35

3.5. Verilerin Türetilmesi	38
3.6. Simülasyon Verilerinin Kontrolü	40
3.7. Verilerin Analizi	41
4. BULGULAR VE TARTIŞMA	45
4.1. Alt Problem 1'e İlişkin Bulgular	45
4.1.1. Testlerin boyutluluğu eşitlik özelliği korunumunu nasıl etkilemektedir?	47
4.1.2. Ortak madde seti formatı eşitlik özelliğinin korunumunu nasıl etkilemektedir?	49
4.1.3. Grupların yetenek dağılımları eşitlik özelliği korunumunu nasıl etkilemektedir?	50
4.1.4. Ölçek dönüştürme yöntemleri eşitlik özelliğinin korunumunu nasıl etkilemektedir?	51
4.1.5. Faktörlerinin etkileşimi eşitlik özelliğinin korunumunu nasıl etkilemektedir?	53
4.2. Alt Problem 2'ye İlişkin Bulgular	58
4.2.1. Testlerin boyutluluğu eşitlik özelliği korunumunu nasıl etkilemektedir?	60
4.2.2. Ortak madde seti formatı eşitlik özelliğinin korunumunu nasıl etkilemektedir?	62
4.2.3. Grupların yetenek dağılımı eşitlik özelliğinin korunumunu nasıl etkilemektedir?	63
4.2.4. Ölçek dönüştürme yöntemleri eşitlik özelliğinin korunumunu nasıl etkilemektedir?	64
4.2.5. Faktörlerin etkileşimi eşitlik özelliği korunumunu nasıl etkilemektedir?	66
4.3. Tartışma ve Yorum	73
5. SONUÇ VE ÖNERİLER	78
5.1. Sonuçlar	78
5.2. Öneriler	80
5.2.1. Araştırmanın Bulgularına Dayalı Olarak Öneriler	80
5.2.2. İleride Yapılacak Araştırmalara Dönük Öneriler	81
KAYNAKÇA	83
EKLER DİZİNİ	89
EK 1. VERİ TÜRETMEDE KULLANILAN MADDELERİN EVREN PARAMETRE DEĞERLERİ	90
EK 2. FORM X'İN KALİBRASYONU İÇİN PARSCALE KODU	91
EK 3. ETİK KURUL İZİN MUAFİYETİ FORMU	92
EK 4. ORJİNALLİK RAPORU	93
ÖZGEÇMİŞ	94

TABLolar DİZİNİ

Tablo 3.1. Madde Havuzunu Oluşturan Maddelerin Konu Alanına Göre Dağılımı.....	33
Tablo 3.2. Testlerde Yer Alan Maddelerin Madde Formatı ve Ölçtüğü Konu Alanına Göre Dağılımı.....	33
Tablo 3.3. Form X, Form Y ve Ortak Testlere Seçilen Maddelerin Ortalama Parametre Değerleri.....	35
Tablo 3.4. Araştırmada İncelenen Faktörler ve Koşullar.....	38
Tablo 3.5. Veri Türetiminde Kullanılan Koşullar.....	38
Tablo 3.6. Form X ve Form Y'nin Madde Parametreleri için Hesaplanan RMSE ve Bias Değeri Ortalamaları.....	41
Tablo 4.1. MTK Gerçek Puan Eşitleme Yöntemine Göre Elde Edilen D_1 İndeksi Ortalamaları.....	46
Tablo 4.2. MTK Gerçek Puan Eşitleme Yöntemine Göre Elde Edilen D_2 İndeksi Ortalamaları.....	46
Tablo 4.3. GR_PE Sonuçlarının BSE ve İSE Özelliği Ölçütlerine Göre 4-yönlü ANOVA Testi Sonuçları.....	47
Tablo 4.4. Boyutluluğa İlişkin İkili Karşılaştırma Testi Sonuçları.....	48
Tablo 4.5. Marjinal Ortalamalar.....	48
Tablo 4.6. Ölçek Dönüştürme Yöntemlerine İlişkin İkili Karşılaştırma Testi Sonuçları.....	52
Tablo 4.7. Marjinal Ortalamalar.....	52
Tablo 4.8. MTK Gözlenen Puan Eşitleme Yöntemine Göre Elde Edilen D_1 İndeksi Ortalamaları.....	59
Tablo 4.9. MTK Gözlenen Puan Eşitleme Yöntemine Göre Elde edilen D_2 İndeksi Ortalamaları.....	59
Tablo 4.10. MTK Gözlenen Puan Eşitleme Sonuçlarının BSE ve İSE Korunumu Ölçütlerine Göre Dört Yönlü ANOVA Testi Sonuçları.....	60
Tablo 4.11. Boyutluluk Faktörünün İkili Karşılaştırma Testi Sonuçları.....	62
Tablo 4.12. Marjinal Ortalamalar.....	62
Tablo 4.13. Ölçek Dönüştürme Yöntemleri Faktörünün İkili Karşılaştırma Testi Sonuçları.....	65
Tablo 4.14. Marjinal Ortalamalar.....	66

ŞEKİLLER DİZİNİ

Şekil 3.1. Eşdeğer olmayan gruplar ortak test deseni	30
Şekil 4.1. GR_PE yöntemi ile D_1 ölçütüne göre büyük ve orta etkiye sahip ikili etkileşimlerin grafikleri	54
Şekil 4.2. GR_PE yöntemi ile D_2 ölçütüne göre büyük ve orta etkiye sahip ikili etkileşimlerin grafikleri	55
Şekil 4.3. GZ_PE yöntemi ile D_1 ölçütüne göre büyük ve orta etkiye sahip ikili etkileşimlerin grafikleri	68
Şekil 4.4. GZ_PE yöntemi ile D_1 ölçütüne göre büyük ve orta etkiye sahip ikili etkileşimlerin grafiklerinin devamı.....	69
Şekil 4.5. GZ_PE yöntemi ile D_2 ölçütüne göre büyük ve orta etkiye sahip ikili etkileşimlerin grafikleri	70

SİMGELER VE KISALTMALAR DİZİNİ

MTK: Madde Tepki Kuramı

NEAT: Non-equivalent groups anchor test design (Eşdeğer olmayan gruplar ortak test deseni)

TIMSS: Uluslararası Fen ve Matematik Eğilimleri Araştırması

3PL: Üç parametrelili lojistik model

M-3PL: Multidimensional three-parameter logistic model (Çok boyutlu üç parametrelili lojistik model)

GPC: Generalized partial credit model (Genelleştirilmiş kısmi puan modeli)

M-2PPC: Multidimensional two-parameter partial credit model (Çok boyutlu iki-parametrelili kısmi puan modeli)

BSE: Birinci-sıra eşitlik

İSE: İkinci-sıra eşitlik

RMSE: Root mean square errors (Hata kareleri ortalamasının karekökü)

O-O: Ortalama/ortalama

O-S: Ortalama/Sigma

SL: Stocking-Lord

HA: Haebara

GR_PE: Gerçek Puan Eşitleme

GZ_PE: Gözlenen Puan Eşitleme

1. GİRİŞ

Bu bölümde, araştırmanın temelini oluşturan problem durumu, araştırmanın amacı ve önemi, problem cümlesi, alt problemler, sınırlılıklar, tanımlar ve araştırmanın kuramsal temeline ilişkin bilgi sunulmuştur.

1.1. Problem Durumu

Günümüzde eğitim kurumları için önemli olan, bireylerin hızla değişen bilgileri elde etmeleri ve elde ettikleri bu bilgileri gerçek yaşam durumlarında kullanabilmeleridir. Dolayısıyla, öğretim programları öğrencilerde alt düzey düşünmeyi gerektiren bir anlayıştan daha çok üst düzey düşünmeyi gerektiren bir anlayışa; öğretim yöntem ve teknikleri de öğretmen merkezli bir yapıdan öğrenci merkezli bir yapıya dönüşmüştür. Bu durum öğretim programının en önemli öğelerinden biri olan ölçme ve değerlendirmenin de bilginin ne derece kazanıldığını ölçen bir yaklaşımdan, bilginin yeni durumlarda ya da gerçek yaşam durumlarında nasıl kullanılabildiğini ölçen bir yapıya dönüşmesini gerektirmiştir.

Çoktan seçmeli testler, eğitim kurum ve kuruluşlarının birçoğunda en çok kullanılan ölçme araçlarından biridir. Çoktan seçmeli testlerin, öğretmenlerin sınıf uygulamalarından ulusal ve uluslararası sınavlara kadar birçok düzeyde ve hemen hemen her öğretim alanında kullanıldığı görülmektedir. Özellikle ülkemizde çok sayıda öğrencinin katıldığı öğrenci seçme ve yerleştirme sınavlarında, açıköğretim veya uzaktan eğitim kuruluşlarının sınavlarında çoktan seçmeli testlerin kullanılan tek sınav türü olduğu söylenebilir (Doğan, 2009).

Çoktan seçmeli test formatının özellikle geniş ölçekli sınavlarda tercih edilmesinin en önemli sebepleri arasında kısa bir zaman diliminde çok sayıda soru sorularak geniş bir kapsam alanının ölçülebilmesi ve objektif puanlanarak güvenilir puanlar elde edilebilmesi gösterilebilir (Kim ve Walker, 2012; Livingston, 2009). Ancak, çoktan seçmeli test formatı bir takım sınırlılıklarından dolayı eleştirilmektedir. Haladyna (1997), çoktan seçmeli test formatının kullanımını çoktan seçmeli öğretime yol açtığı, yazma ve yaratıcı düşünme gibi becerilerin ölçülmesi için yetersiz olduğu gerekçeleri ile eleştirmektedir. Benzer şekilde Livingston (2009) okulda öğretilen çoğu becerilerin çoktan seçmeli maddelerle etkili bir şekilde ölçülemeyecek kadar karmaşık olduğunu belirtmiş ve şu örneği vermiştir: “Çoktan

seçmeli bir tarih testi ile öğrencilerin olgusal bilgileri ölçülebilir ancak tarihsel bir sorun üzerine mantıklı bir kompozisyon yazıp yazamayacakları belirlenemez.” (s.2).

Üst düzey düşünme becerilerini ölçmek için açık uçlu maddelerin kullanılması önerilmektedir. Açık uçlu maddeler, öğrencinin verilen seçeneklerden doğru cevabı seçmek yerine kendi cevabını oluşturmasını gerektirir. Böylece çoktan seçmeli maddelerde var olan şansa doğru cevap verme olasılığı açık uçlu maddelerde oluşmaz. Açık uçlu maddeler, öğrencinin sadece bir iki cümle yazarak yanıtlayacağı kadar basit olabileceği gibi bir makaleyi okuyup ana fikrini belirleyip bu ana fikre uygun anlamlı bir kompozisyon yazmasını gerektirecek kadar karmaşık da olabilir (Tankersley, 2007). Kompozisyonlar, öğrencinin bir kelime, bir sayı ya da bir formül yazmasını gerektiren kısa cevaplı maddeler ve belirli işlem basamaklarını çözmesini gerektiren matematik problemleri en çok kullanılan açık uçlu soru tipleridir. Genellikle bu maddeler, doğru-yanlıştan ziyade çok kategorili puanlanırlar ve bu sayede öğrenciler eksik ya da kısmen doğru olan cevaplarından da puan alabilirler. Açık uçlu maddelerle üst düzey düşünme becerileri etkili bir biçimde ölçülebilir fakat bu maddeleri objektif ve güvenilir olarak puanlamak çoktan seçmeli maddelerle kıyaslanınca oldukça zordur (Kirkpatrick, 2005).

Çoktan seçmeli ve açık uçlu madde formatlarının yukarıda saydığımız bir takım sınırlılıklarından dolayı Programme for International Student Assessment (PISA), Trends in International Mathematics and Science Study (TIMSS), International Assessment of Education Progress (IAEP) gibi uluslararası düzeyde uygulanan geniş ölçekli sınavlar karma test formatında hazırlanmakta, çoktan seçmeli ve açık uçlu maddeler birlikte kullanılmaktadır. Bu sayede farklı madde türlerinin birbirine olan üstünlüğü bir avantaja dönüştürülmektedir.

PISA ve TIMSS gibi uluslararası geniş ölçekli sınavlarda test güvenliğini sağlamak, aynı soruları sormadan ölçmedeki değişimi takip etmek ya da her bir öğrencinin bütün sorulara cevap vermesini gerektirmeden daha fazla sayıda soruyu örneklemek için farklı test kitapçıkları kullanılmaktadır (Xu, 2009). Bu sınavlarda, test maddeleri bloklara bölünmektedir. Bu bloklar madde sayısı, kapsam ve madde formatı açısından her kitapçık birbirine eşdeğer olacak şekilde farklı kitapçıklara dağıtılmaktadır. Öğrenciler, herhangi bir kitapçığa tesadüfî olarak atanmaktadır.

Her ne kadar kitapçıklar aynı yapıyı ölçmek için benzer kapsam ve istatistiksel özelliklere sahip olacak şekilde hazırlansa da test puanlarının aynı ölçekte ifade edilmesi bir başka deyişle eşitlemesi gerekir. Eşitleme yapıldıktan sonra öğrencilerin hangi kitapçığı aldığına bir öneminin olmaması; geniş ölçekli sınavlar için bütün sonuçların öğrencilerin tümünün her bir uygulamada bütün soruları almış gibi oluşması gerekir (Xu, 2009). Dolayısıyla, her uygulamada birden fazla test kitapçığının kullanıldığı uluslararası sınavlar için test eşitleme önemli bir süreçtir.

Bir testte birden fazla madde formatının kullanılması eşitleme sürecine birtakım güçlükleri de beraberinde getirmektedir. Tek bir madde formatının kullanıldığı testlerde olduğu gibi karma formatta hazırlanan testlerin çok boyutlu yapıya sahip olduğu durumlarda eşitleme sürecinde bir takım zorluklarla karşılaşmaktadır. Çok boyutluluğunun oluşmasının birçok nedeni olabilir. Karma formatta hazırlanan testlerde eğer çoktan seçmeli ve açık uçlu maddeler, bireylerin farklı bilişsel süreçlerini ölçüyor ise çok boyutluluk madde formatına dayalı olarak oluşabilir (Traub, 1993). Bir diğer deyişle çoktan seçmeli ve açık uçlu maddeler aynı kapsamı fakat farklı bilişsel süreçleri ölçüyor olabilirler. Bu durumu Traub (1993) format etkisi olarak açıklamış ve format etkisinin karma testlerde formata dayalı çok boyutluluğa neden olabileceğini belirtmiştir.

Çok boyutluluğun Madde Tepki Kuramına (MTK) dayalı eşitleme sürecini etkileyen bir faktör olduğu tanımlanmıştır (Kolen ve Harris, 1990). Ancak günümüzde birçok test programında tek boyutlu MTK modellerinin kullanımı daha yaygındır. Dolayısıyla çok boyutluluğun olduğu test durumlarında tek boyutlu MTK modelleri kullanılarak yapılan eşitlemenin ne kadar dayanıklı olduğunu araştırmak önemlidir. Bu araştırmanın temel amaçlarından biri çok boyutluluğun olduğu test durumunda tek boyutlu MTK modelleri kullanılarak yapılan eşitlemenin eşitlik özelliğinin korunumuna ne kadar dayanıklı olduğunu incelemektir.

Test eşitleme çalışmalarında ortak testin önemli bir rolü vardır. En çok kullanılan ve en popüler desen olan eşdeğer olmayan gruplar ortak test deseni ile kaliteli bir eşitleme çalışması yürütebilmek ortak testin kalitesine bağlıdır (Sinharay ve Holland, 2006). Çoktan seçmeli testler için ortak madde setinin, testin bütünü minyatürü olması gerektiği konusunda yaygın bir görüş bulunmaktadır. Birçok

uzman (von Davier, Holland ve Thayer, 2004; Kolen ve Brennan, 2004) ortak testin, testin bütününün kapsam ve istatistiksel özellikler bakımından bir aynası olması gerektiğini belirtmişlerdir. Ancak, sadece çoktan seçmeli maddelerden oluşan testlere ait ortak madde setinin oluşturulmasına yönelik olan bu görüş, ne ölçüde karma testlere genellenebilir? Dolayısıyla bu araştırmanın diğer bir amacı da ortak madde seti formatının karma testlerin eşitlenmesine etkisini incelemektir.

Literatürde karma testlerde ölçek dönüştürme yöntemlerinin performansının incelendiği çok az çalışma bulunmakla birlikte yapılan araştırmalar (Kim 2004; Kim ve Lee, 2006; Kim ve Kolen, 2006) karakteristik eğri ölçek dönüştürme yöntemlerinin moment ölçek dönüştürme yöntemlerine kıyasla daha kararlı sonuçlar verdiğini göstermiştir. Eşitlik özelliği ölçüt olarak alındığında ölçek dönüştürme yöntemlerinin performansı nasıldır? Bu araştırmanın bir diğer amacı ölçek dönüştürme yöntemlerinin eşitlik özelliği korunumu ölçütüne göre performansını incelemektir.

1.2. Araştırmanın Amacı ve Önemi:

Bu araştırmanın amacı testlerin boyutluluğunun, ortak madde seti formatının, grupların yetenek dağılımlarının ve farklı ölçek dönüştürme yöntemlerinin karma testlerin eşitlenmesinde eşitlik özelliğinin korunumuna etkisini incelemektir.

Literatür incelendiğinde, karma testlerin eşdeğer olmayan gruplar ortak test deseni kullanılarak eşitlenmesinde ortak madde setinin nasıl oluşturulması gerektiğine dair bir uzlaşma olmadığı görülmüştür. Her ne kadar Kolen ve Brennan (2004) ortak madde setinin kapsam ve istatistiksel özellikler bakımından testin bütününü yansıtmaması gerektiğini savunsa da uygulamada genellikle açık uçlu maddeler birtakım sınırlılıklarından dolayı ortak madde setine dâhil edilmemektedir (He, 2011). Bu araştırma ile simülasyon verileri kullanılarak ortak madde setine açık uçlu maddelerin dahil edilmesinin ya da edilmemesinin eşitlemenin eşitlik özelliğini nasıl etkilediği incelenerek, bu konudaki tartışmalara katkıda bulunmak hedeflenmektedir.

Karma testlerin eşitlenmesi üzerine yapılan araştırmalar (Andrews, 2011; Bastari, 2000; Cao, 2008; Hagge, 2010; He, 2011; Kim ve Walker, 2012; Kirkpatrick, 2005; Tian, 2011; Wang, 2013; Wolf, 2013) incelendiğinde testlerin boyutluluğun ve ortak

madde seti özelliklerinin bir arada ele alınıp dört farklı ölçek dönüştürme yönteminin performansının eşitlik özelliği korunumu ölçütüne dayalı olarak karşılaştırıldığı herhangi bir çalışmaya rastlanmamıştır. Araştırma bu yönüyle yapılan diğer araştırmalardan farklılık göstermektedir.

Ülkemizde test eşitlemeye ilişkin yapılan çalışmalar (Atalay-Kabasakal, 2014; Bozdağ ve Kan, 2010; Gök, 2012; Kelecioğlu, 1994; Kilmen, 2010; Öztürk, 2010; Öztürk ve Anıl, 2012) incelendiğinde hepsinin 1-0 puanlanan çoktan seçmeli testlerden elde edilen puanların eşitlenmesi üzerine olduğu görülmüş, karma testlerin eşitlenmesine ilişkin herhangi bir araştırmaya rastlanmamıştır. Bu araştırma, ülkemizde karma testlerin eşitlenmesi üzerine yapılan ilk çalışma olma özelliğini de taşımaktadır.

Ülkemizde test güvenliğini ve gizliliğini sağlamak için yılın belirli dönemlerinde bir sınavın farklı formları kullanılmakta ve bu farklı sınav formlarının eşdeğer olduğu varsayılmaktadır. Ancak, test geliştirme uzmanları her ne kadar içerik ve istatistiksel özellikler bakımından olabildiğince birbirine benzer formlar hazırlamaya çalışsa da formların tam olarak birbirine eşdeğer olması beklenemez. Dolayısıyla bu araştırma, ülkemizde aynı amaç doğrultusunda farklı dönemlerde uygulanan sınavlardan elde edilen puanları karşılaştırabilmek için gerekli olan test eşitleme sürecine dikkat çekmesi ve bu sürece rehber olması açısından da önemlidir.

1.3. Problem Cümlesi:

Testlerin boyutluluğu, ortak madde seti formatı, grupların yetenek dağılımları ve farklı ölçek dönüştürme yöntemleri karma testlerin eşitlenmesinde eşitlik özelliğinin korunumunu nasıl etkilemektedir?

1.3.1. Alt Problemler:

1) Madde Tepki Kuramı gerçek puan eşitleme yöntemi ile yapılan eşitlemelerde, eşitlik özelliğinin korunumunu;

- a) Testlerin boyutluluğu,
- b) Ortak madde seti formatı,
- c) Grupların yetenek dağılımları,
- d) Ölçek dönüştürme yöntemleri ve
- e) Bu faktörlerin etkileşimleri nasıl etkilemektedir?

2) Madde Tepki Kuramı gözlenen puan eşitleme yöntemi ile yapılan eşitlemelerde, eşitlik özelliğinin korunumunu;

- a) Testlerin boyutluluğu,
- b) Ortak madde seti formatı,
- c) Grupların yetenek dağılımları,
- d) Ölçek dönüştürme yöntemleri ve
- e) Bu faktörlerin etkileşimleri nasıl etkilemektedir?

1.4. Sınırlılıklar:

Bu araştırma,

1. Simülasyon verileri,
2. Boyutluluğun madde formatına dayalı oluşturulduğu iki boyutlu test yapısı ve
3. Çok boyutlu telafi edici MTK modelleri ile sınırlıdır.

1.5. Tanımlar:

Karma Test: Farklı oranlarda çoktan seçmeli ve açık uçlu madde içeren testtir.

Eşitlik Özelliği: Eşitleme yapıldıktan sonra alternatif formaların koşullu dağılımlarının aynı olmasıdır.

Birinci-Sıra Eşitlik: Eşitleme yapıldıktan sonra alternatif formların koşullu ortalamalarının eşit olmasıdır.

İkinci-Sıra Eşitlik: Eşitleme yapıldıktan sonra alternatif formlara ait ölçmenin koşullu standart hatasının (SEM) eşit olmasıdır.

1.6. Araştırmanın Kuramsal Temeli

Angoff (1984), test formlarının eşitlenmesini bir formun birim sistemini diğer formun birim sistemine dönüştürmek olarak tanımlamış ve dönüştürmeden sonra formların eşdeğer olacağını belirtmiştir. Kolen ve Brennan (2004) eşitlemeyi; benzer içerik ve güçlük düzeyine sahip test formları arasındaki farklılıkları düzenleyerek, bu formlardan elde edilen puanların birbiri yerine kullanılmasını sağlayan istatistiksel bir süreç olarak tanımlamıştır. En genel tanımıyla eşitleme, test formları arasındaki güçlük farklılıklarını düzenleyerek elde edilen puanların birbiri yerine kullanılmasını sağlayan sürece verilen addır (Öztürk ve Anıl, 2012).

1.6.1. Test Eşitleme Desenleri

Bir eşitleme sürecine başlamadan önce veri toplamak için gerekli olan eşitleme desenine karar vermek gerekir. Kolen ve Brennan (2004) uygulamada veri toplamada kullanılan üç temel deseni şöyle sıralamışlardır; tek grup deseni (single group design), eşdeğer gruplar deseni (random groups design) ve eşdeğer olmayan gruplar ortak test (non-equivalent groups anchor test (NEAT)) deseni.

Tek grup deseninde iki veya daha fazla test formu aynı gruba uygulanır. Tek grup desenin dezavantajlı yönü formların uygulanma sırasının birbirinden etkilenmesidir. Bu duruma "sıra etkisi" adı verilmektedir. Sıra etkisi ikinci formun uygulanmasında yorgunluk ve sorulara aşinalık olarak ortaya çıkabilir. Kolen ve Brennan (2004) sıra etkisini kontrol altına almak için dengelenmiş tek grup desenini önermiştir. Bu yöntemde grup ikiye ayrılır ve birinci grup ilk olarak X formunu daha sonra Y formunu alırken ikinci grup ilk olarak Y formunu sonra X formunu alır.

Eşdeğer gruplar deseninde, X ve Y formlarındaki bütün puanları temsil edecek yeterlikte büyük bir heterojen grup seçilir. Grup tesadüfi olarak ikiye ayrılır. Gruplara bireyler atanırken sarmal bir süreç izlenebilir. Bunun için kitapçıklar önce X formu ardından Y formu gelecek şekilde paketlenir ve ilk kişi X formunu ikinci kişi Y formunu, üçüncü ise X formunu alacak şekilde dağıtılır. Bu sarmal süreç sayesinde X ve Y formunu alan eşdeğer iki grup oluşur. Tek grup deseninden farklı olarak bu desende bireyler sadece bir formu alırlar fakat bütün test formlarının aynı oturumda uygulanması gerekmektedir. Bir oturumda, bütün test formlarının uygulanma zorunluluğu test güvenliğini sağlamada sorunlara neden olabilir. Ayrıca, her bir formu örneklemin belirli bir oranı aldığı için büyük örneklemelere ihtiyaç vardır (Kolen ve Brennan, 2004).

Eşdeğer olmayan gruplar ortak test deseninde (non-equivalent groups anchor test; NEAT) denk olmayan iki grubun her birine ayrı bir test uygulanır. Test formlarını alan grupların eşdeğer yetenek dağılımına sahip olması varsayımı yoktur. Puan farklılıklarına formlar arasındaki farklılıklar sebep olduğu kadar grup farklılıklarının etkileri de karışır. Grupların farklılığı her iki formda yer alan ortak madde seti ile giderilmeye çalışılarak ortak maddeler aracılığı ile formlar birbirine eşitlenir (He, 2011).

NEAT deseninde grup ve form farklılıklarını doğru olarak ayırt edebilmek için eşitlemede kullanılan ortak maddelerin titizlikle oluşturulması gerekir. Kolen ve Brennan (2004), özellikle ortak madde setinin testin bütününe "mini bir versiyonu" nu yansıtabilecek şekilde oluşturulması gerektiğini belirtmiştir. Ortak madde seti, testin bütünü ile aynı içerik ve istatistiksel özelliklere sahip olmalı, her iki test formunda da benzer konumda ya da soru sırasında yer almalıdır. Ayrıca, ortak madde setindeki madde sayısı doğru eşitleme sonuçları verebilecek sayısal çoklukta olmalıdır.

Karma formattaki testlerin NEAT deseni ile eşitlenmesinde, ortak madde setinin hem çoktan seçmeli hem açık uçlu maddeleri içererek testin bütününe oransal olarak temsil etmesi gerekir. Ancak, uygulamada genellikle ortak madde seti sadece çoktan seçmeli maddelerden oluşturulmaktadır. Ortak madde setine açık uçlu maddelerin alınmamasının birçok sebebi vardır (He, 2011). Örneğin, kompozisyon oluşturma gibi açık uçlu maddeleri hatırlamak çoktan seçmeli maddeleri hatırlamaktan daha kolaydır. Dolayısıyla açık uçlu maddeler, ortak madde olarak kullanılırsa öğrenciler tarafından açıklanabilir bu da sınavın güvenliğinde sorunlara neden olabilir (Muraki, Hombo ve Lee, 2000). İkincisi, açık uçlu maddelerin cevaplanması ve puanlanması daha uzun zaman aldığı için çoktan seçmeli maddelerle kıyaslanınca çok daha az sayıda açık uçlu maddeye karma testlerde yer verilmektedir. Dolayısıyla, sadece birkaç tane açık uçlu maddenin var olduğu bir karma testten, testin bütününe temsil edecek ortak açık uçlu maddeyi seçmek çok zordur. Üçüncü olarak, açık uçlu maddeler puanlayıcılar tarafından puanlandığı için, elde edilen puanlara puanlayıcıdan kaynaklı hata karışabilir (He, 2011; Kirkpatrick, 2005).

Bir grup araştırmacı (Baghi Bent, DeLain ve Hennings, 1995; Ercikan vd., 1998) açık uçlu maddeler de içeren karma test formları arasındaki farklılıkların sadece çoktan seçmeli maddelerden oluşan ortak test ile kontrol edilebileceğini savunurken bir diğer grup araştırmacı ise (Kim ve Kolen, 2006; Kim, Walker ve McHale, 2010; Li, Lissitz ve Yang, 1999) çoktan seçmeli ve açık uçlu maddelerin farklı yapıları ölçtüğünü savunarak ortak madde setinin tamamen çoktan seçmeli maddelerden oluşmasının eşitlemede yanlış sonuçlara neden olacağına dair kanıtlar sunmuştur. Karma testlerin eşitlenmesinde kullanılan ortak madde setinin

nasıl oluşturulması gerektiğine ilişkin literatürde bir uzlaşma olmadığı görülmektedir. Dolayısıyla bu araştırmada ortak madde setini oluşturan madde formatı bir faktör olarak ele alınmıştır.

1.6.2. Test Eşitleme Yöntemleri

Eşdeğer olmayan gruplar ortak test deseni için klasik test kuramı ve madde tepki kuramına (MTK) dayalı birçok test eşitleme yöntemi geliştirilmiştir. Tucker yöntemi (Gulliksen, 1950), Levine gerçek ve gözlenen puan yöntemi (Levine, 1955) ve Braun-Holland doğrusal yöntemi (Braun ve Holland, 1982) NEAT deseni altındaki geleneksel doğrusal eşitleme yöntemleridir. Frekans kestirimi (Angoff, 1971) ve zincirleme (chained) eşit yüzdelikli eşitleme yöntemleri (Angoff, 1971) ise NEAT deseni altındaki geleneksel eşit yüzdelikli eşitleme yöntemleridir.

MTK'na dayalı eşitleme yöntemleri, genel olarak MTK gerçek puan eşitleme ve MTK gözlenen puan eşitleme olmak üzere iki başlık altında toplanabilir. MTK'na dayalı eşitleme genellikle madde kalibrasyonu, ölçek dönüştürmesi ve eşitleme olmak üzere üç aşamada tamamlanır. Birinci aşama olan madde kalibrasyonunu yapabilmek için belirli bir MTK modeli ya da karma testler için MTK modellerini seçmek gerekir. Madde parametrelerini kestirmek için birçok tek boyutlu ve çok boyutlu model geliştirilmiştir.

1.6.2.1. Madde Tepki Kuramı Modelleri

Tek boyutlu MTK modellerinde bireylerin cevaplarını sadece tek bir örtük değişkenin etkilediği varsayılır. İki kategorili diğer bir deyişle 1-0 şeklinde puanlanan maddelere ilişkin madde parametreleri genellikle bir-parametrelili lojistik model (1PL), iki-parametrelili lojistik (2PL) model ve üç-parametrelili lojistik (3PL) modellerinden biri seçilerek kestirilir. Tek boyutlu ikili puanlanan maddeler için geliştirilen MTK modelleri içerisinde geniş ölçekli sınavlarda en çok kullanılanı 3PL modeldir (Huggins, 2012). 3PL model altında θ yeteneğine sahip bir i bireyinin j maddesine doğru cevap verme olasılığı matematiksel olarak 1.1 eşitliğinde olduğu gibi ifade edilebilir:

$$P_{ij} = P(\theta_i; a_j, b_j, c_j) = c_j + (1 - c_j) \frac{\exp[Da_i(\theta_i - b_j)]}{1 + \exp[Da_j(\theta_i - b_j)]} \quad (1.1)$$

Eşitlik 1.1’de yer alan a_j j maddesinin ayırt edicilik parametresi; b_j güçlük parametresi ve c_j psödo şans parametresini bir diğer deyişle düşük yeteneğe sahip bir bireyin (θ) maddeye doğru cevap verme olasılığını gösterir. Eşitlikteki, D ölçekleme sabitidir ve değeri 1.7’dir.

Psödo şans parametresi sıfıra eşitlenirse ya da şansla başarının olmadığı varsayılır ise iki parametrelili lojistik model (2PL) elde edilir. Eğer c parametresinin sıfıra eşitlenmesine ek olarak a parametresinin tüm maddeler için aynı olduğu kabul edilip sabitlenir ise 1PL model elde edilir. TIMSS uygulamalarında çoktan seçmeli maddeler 3PL model kullanılarak kalibre edildiği için bu araştırma kapsamında da çoktan seçmeli maddelere ait madde cevaplarının türetilmesinde ve parametrelerin kestiriminde 3PL model kullanılmıştır.

Her ne kadar MTK modelleri ile yapılan çalışmaların büyük bir çoğunluğu ikili puanlanan maddelerden elde edilen veri analizine yoğunlaşmış olsa da ikiden fazla puan kategorisine sahip maddeler için de önemli modeller geliştirilmiştir. Aşamalı tepki modeli (Graded Response Model (GRM); Samejima, 1969), kısmi puan modeli (Partial Credit Model (PCM); Masters, 1982) ve genelleştirilmiş kısmi puan modeli (Generalized Partial Credit (GPCM); Muraki, 1992) üzerinde en çok çalışılan ve çok boyutlu MTK çerçevesinde genellemesi yapılan üç modeldir (akt. Reckase, 2009).

Aşamalı tepki modeli (GRM; Samejima, 1969), likert ölçeği gibi sıralı kategorilere sahip çoklu puanlanan maddelerin modellenmesi için uygun bir model olup 2PL modelin genelleştirilmiş bir versiyonu olarak düşünülebilir. GRM’e göre bir bireyin x_j kategorisinde ya da daha üst kategoride cevap verme olasılığı matematiksel olarak (De Ayala, 2009):

$$P_{x_j}^*(\theta) = \frac{e^{a_j(\theta - \delta_{x_j})}}{1 + e^{a_j(\theta - \delta_{x_j})}} \quad (1.2)$$

1.2 eşitliğindeki θ örtük yeteneği, a_j j maddesinin ayırt edicilik parametresini, δ_{x_j} ise x_j kategorisi ($x_j = 0, 1, \dots, m_j$) için kategori sınır güçlüğü (category boundary location) parametresini temsil etmektedir. Kategori güçlük parametresi k ve $k-1$

kategorileri arasındaki sınır (eşik) olarak da adlandırılabilir. GRM' de δ_{x_j} parametreleri her zaman artan sırada olup bir j maddesi için m_j tane kategori sınır gücülüğü parametresi vardır.

Çok kategorili puanlanan maddeler için geliştirilen bir diğer model kısmi puan modelidir (PCM; Masters,1982). PCM, çözümünde birden fazla işlem basamağı gerektiren ve çözümü için bu işlem basamaklarının tamamlanmasının önemli olduğu maddelerin modellenmesi için geliştirilmiştir. Dolayısıyla, PCM matematik problemleri gibi kısmi puanlamanın mümkün olduğu başarı testlerine verilen madde cevaplarını modellemek için uygundur. Örneğin, öğrencilere şöyle bir matematik maddesi sorduğumuzu varsayalım (De Ayala, 2009, s.163):

$$(6/3)+2=?$$

Bu maddeye doğru cevabı verebilmek için belirli işlem basamaklarının doğru olarak başarılması gerekir. İlk işlem basamağı öğrencinin 6/3 oranını bilmesi, ikinci işlem basamağı ilk işlem basamağından elde ettiği sonucu 2 ile toplayabilmesidir. Madde iki işlem basamağından oluştuğu için $x_i = \{0, 1, 2\}$ şeklinde kısmi olarak puanlanabilir. Bu madde için şöyle bir puanlama anahtarı hazırlayabiliriz: öğrenci ilk işlemi yanlış yaparsa sıfır, ilk işlemi doğru yapıp ikinciyi yanlış yaparsa 1, her iki işlemi de doğru yaparsa 2 puanını alacaktır. Bu puanlara kategori puanı adı verilir ve kategori puanı aynı zamanda başarıyla tamamlanan işlem basamağı sayısını gösterir. Daha yüksek kategori puanı, daha düşük kategori puanı ile karşılaştırılınca daha yüksek yeteneği gerektirir. Bu yaklaşımda, bireylerin cevapları m_j+1 kategori puanına sahip olur. Örneğin, $x_j=\{0,1,..m_j\}$ eşitliğindeki m_j j maddesini doğru olarak cevaplamak için gerekli olan işlem basamağı sayısını gösterir. Yukarıdaki örnekte $m_j=2$ 'dir. PCM, 1PL modelin bir uzantısı olarak düşünülebilir. θ yeteneğine sahip bir bireyin bir x_j kategori puanını elde etmesinin koşullu olasılığını $P(x_j|\theta, \delta_{jh})$ temsil etmek üzere PCM matematiksel olarak (De Ayala, 2009):

$$P(x_j|\theta, \delta_{jh}) = \frac{\exp\left[\sum_{h=0}^{x_j} (\theta - \delta_{jh})\right]}{\sum_{k=0}^{m_j} \exp\left[\sum_{h=0}^k (\theta - \delta_{jh})\right]} \quad (1.3)$$

1.3 eşitliğindeki δ_{jh} , j maddesinin adım (işlem) güçlük parametresini (step difficulty parameter); m_j kategori sayısı temsil etmektedir ve $k = \{1, \dots, m_j\}$. PCM' de cevap kategorilerinin mutlaka sıralı olması gerekir fakat bu koşul adım güçlük parametresinin sıralı olmasını gerekli kılmamaktadır.

Genelleştirilmiş kısmi puan (GPC) modeli ise ayırt edicilik parametresi olan a 'nın kısmi puan modeline (PCM) eklenmesi ile onun bir uzantısı olarak geliştirilmiştir. PCM'de olduğu gibi GPC modeli de başarı ile tamamlanması için bir dizi işlem basamağının yerine getirilmesini gerektiren madde cevaplarının modellenmesi için uygundur. Benzer şekilde, bir maddeden maksimum puan almak için bütün işlem basamaklarının doğru olarak tamamlanması gerekir. GPC modeli, çok sayıda farklı bileşenin puanlandığı açık uçlu madde cevaplarının modellenmesi için uygun bir modeldir. Her bir bileşen doğru/yanlış ya da başarılı/başarısız şeklinde puanlanır ve bir maddeden alınacak toplam puan başarı ile tamamlanan bileşen sayısı kadardır. GPC modeli, matematiksel olarak 1.4 eşitliğinde olduğu gibi ifade edilebilir (Embretson ve Reise, 2000; De Ayala, 2009):

$$P(x_{jk} | \theta, a_j, \delta_{jk}) = \frac{\exp\left[\sum_{h=1}^{k_j} a_j (\theta - \delta_{jh})\right]}{\sum_{c=0}^{m_j} \exp\left[\sum_{h=1}^c (\theta - \delta_{jh})\right]} \quad (1.4)$$

TIMSS uygulamalarında açık uçlu maddeler GPC modeli kullanılarak kestirildiği için bu araştırmada da açık uçlu maddelere ait madde cevaplarının türetilmesinde ve madde parametrelerinin kestirilmesinde GPC modeli kullanılmıştır.

Yukarıda açıklanan MTK modelleri testlerin tek boyutlu olduğunu varsaymaktadır. Ancak, eğitimde ve psikolojide kullanılan çoğu test bir ölçüde çok boyutlu yapıya sahiptir. Dolayısıyla MTK modellerini uygulamadan önce testlerin boyutluluğunun incelenmesi gerekir. Tek boyutlu modellerde yer alan örtük özellik ya da yetenek parametrelerine yenilerinin eklenmesi ile tek boyutlu MTK modellerinin bir uzantısı olarak çok boyutlu MTK modelleri geliştirilmiştir (Ackerman, 1994; akt. Kang, 2006).

Çok boyutlu MTK modellerini Ackerman (1989), telafi edici (compensatory) ve telafi edici olmayan (non-compensatory) modeller olmak üzere iki kategoride sınıflandırmıştır. Telafi edici model, bir yetenek düzeyindeki yetkinliğin diğer boyut

veya boyutlardaki eksik yetkinliği tamamlaması durumunda kullanılmaktadır. Telafi edici model boyutların etkileşmesine, bir boyuttaki yüksek yeteneğin ikinci boyuttaki daha düşük yeteneği tamamlamasına izin vermektedir. Telafi edici olmayan modellerde ise yüksek puan alabilmek için bir bireyin her iki boyutta da yetkin olması gerekmektedir (Yao ve Schwarz, 2006). Bu araştırma kapsamında, çok boyutlu ikili puanlanan madde cevapları telafi edici çok boyutlu üç parametrelili lojistik modele (multidimensional three-parameter logistic model; M-3PL), çok kategorili puanlanan madde cevapları çok boyutlu iki parametrelili kısmi puan modeline (multidimensional two-parameter partial credit model; M-2PPC) dayalı olarak türetildiği için aşağıda sadece bu modellere ilişkin bilgi sunulmuştur.

Çok boyutlu üç parametrelili lojistik model (M-3PL), tek boyutlu 3PL modelin geliştirilmiş halidir. Çok boyutlu modellerde tıpkı tek boyutlu modellerde olduğu gibi bir maddeye verilen doğru cevap $X_{ij} = 1$ ve yanlış cevap $X_{ij} = 0$ olarak kodlanır. Madde karakteristik eğrisi yerini doğru cevap verme yüzeyine bırakır. M-3PL modeli 1.5 eşitliğinde olduğu gibi ifade edilebilir (Reckase, 2009):

$$P(X_{ij} = 1 | \theta_j, a_i, d_i, c_i) = c_i + (1 - c_i) \frac{\exp[a_i' \theta_j + d_i]}{1 + \exp[a_i' \theta_j + d_i]} \quad (1.5)$$

Eşitlik 1.5' te yer alan θ_j , M boyutluluktaki örtük yeteneklerin bir vektörü olmak üzere; a_i madde eğimlerinin bir vektörü, d_i madde güçlüğü ile ilişkili skaler bir parametredir.

Maddeler ikiden fazla puan kategorisine sahip ise X_{ij} cevabı, $k = \{1, \dots, K_i\}$ puan kategorilerinden biriyle kodlanır. Çok boyutlu iki-parametrelili kısmi puan modeli tek boyutlu geliştirilmiş kısmi puan modelinin bir uzantısı olarak geliştirilmiştir ve matematiksel olarak M-2PPC model 1.6 eşitliğinde olduğu gibi ifade edilebilir (Yao ve Schwarz, 2006):

$$P(X_{ij} = k | \theta_j, a_i, d_{ik}) = \frac{\exp[\sum_{v=1}^k a_i' \theta_j + d_{iv}]}{\sum_{h=1}^{K_i} \exp[\sum_{v=1}^k a_i' \theta_j + d_{iv}]} \quad (1.6)$$

Eşitlik 1.6'da yer alan θ_j , örtük yeteneklerin bir vektörü olmak üzere; a_i madde eğimlerinin bir vektörü ve d_{ik} eşik parametreleridir.

1.6.2.2. Madde Kalibrasyonu

Uygun MTK modelleri seçildikten sonra test eşitleme sürecinin ilk basamağı olan ayrı (separate) ya da eş zamanlı (concurrent) kalibrasyon yapılarak madde ve yetenek parametreleri kestirilebilir. Ayrı kalibrasyonda her iki test formuna ait madde ve yetenek parametreleri bilgisayar programı her bir form için ayrı çalıştırılarak kestirilir. Eş zamanlı kalibrasyonda ise her iki test formuna ait madde ve yetenek parametreleri eş zamanlı olarak tek bir seferde kestirilir. Her iki test formuna ait bütün parametrelerin eş zamanlı olarak kestirilmesi aynı ölçekte elde edilmelerini garantilerken ayrı kalibrasyon yapılarak kestirilen madde parametreleri aynı ölçekte olmayacaktır. Ayrı kalibrasyon yapıldığında, örtük değişkene ait ölçeğin ortalama ve standart sapması örtük değişkenin dağılımına sabitlenmektedir. Eğer madde parametrelerini kestirmede kullanılan örneklemeler farklı evrenlerden geliyor ise kestirilen parametreler farklı ölçeklerde elde edilir (Hanson ve Beguin, 2002).

Her ne kadar eş zamanlı kalibrasyonu uygulamak ayrı kalibrasyondan daha kolay ve pratik gibi görünse de literatürde yapılan araştırmalar tek boyutluluk varsayımının sağlanmadığı durumlarda ayrı kalibrasyonun daha doğru sonuçlar verdiğini göstermiştir (Beguin, Hanson ve Glas, 2000; Kim ve Kolen, 2006). Kolen ve Brennan (2004), ayrı kalibrasyonun daha güvenilir olduğunu belirtirken Cao (2008) eş zamanlı kalibrasyonda çoğu zaman yakınsaklık (convergence) problemleri ile karşılaşıldığının altını çizmiştir. Bu araştırma kapsamında çok boyutluluk bir faktör olarak alındığı için madde ve yetenek parametre kestiriminde ayrı kalibrasyon yöntemi tercih edilmiştir.

1.6.2.3. Ölçek Dönüştürme

Test eşitlemenin yapılabilmesi için farklı formlara ait madde ve yetenek parametrelerinin aynı ölçekte olması gerekir (Kolen ve Brennan, 2004). Dolayısıyla ayrı kalibrasyon sonucu elde edilen madde ve yetenek parametreleri ancak ölçek dönüştürme işlemi uygulanarak aynı ölçekte ifade edilebilir.

Ölçek dönüştürme işleminin temel amacı yeni formun madde ve yetenek parametrelerini eski formun madde ve yetenek parametrelerinin ölçeğine dönüştürecek iki bağlanma (linking) katsayısını bir diğer deyişle eğim (A) ve keşişim (B) sabitlerini bulmaktır. Eski formu " E ", yeni formu " Y " harfleri temsil

etmek üzere yeni formun yetenek parametreleri 1.7' de verilen eşitlik ile madde parametreleri (ayrıt edicilik ve güçlük) 1.8 ve 1.9 numaralı eşitlik ile eski formun ölçeğine dönüştürülebilir. Şans parametresi olasılık ölçeğinde olduğu için herhangi bir dönüştürme işleminin uygulanmasına gerek yoktur (Kolen ve Brennan, 2004) :

$$\theta_E = A\theta_Y + B \quad (1.7)$$

$$a_{jE} = a_{jY}/A \quad (1.8)$$

$$b_{jE} = Ab_{jY} + B \quad (1.9)$$

Kolen ve Brennan (2004) yaygın olarak kullanılan dört ölçek dönüştürme yöntemini ortalama/sigma (Marco, 1977), ortalama/ortalama (Loyd ve Hoover, 1980), Haebara (Haebara, 1980) ve Stocking-Lord (Stocking ve Lord, 1983) yöntemi olarak sıralamıştır. Ortalama/ortalama ve ortalama/sigma yöntemleri moment, Haebara ve Stocking-Lord yöntemleri ise karakteristik eğri yöntemleri olarak da adlandırılmaktadır.

Ortalama/sigma yönteminde, 1.8 ve 1.9 numaralı eşitliklerdeki A (eğim) ve B (kesişim) sabitlerini elde etmek için ortak maddelerden elde edilen b -parametre kestirimlerinin ortalama ve standart sapmaları alınır. Ortalama/ortalama yönteminde ise ortak maddelerden elde edilen a -parametre kestirimlerinin ortalaması alınarak A -sabitini elde edilir. Eşitliklerdeki B -sabitini elde etmek için ise ortak maddelerden elde edilen b -parametre kestirimlerinin ortalaması alınır. MTK modellerinin varsayımları tam olarak karşılanmaz ise genellikle ortalama/sigma ve ortalama/ortalama yöntemleri farklı sonuçlar verir. Bazen ortalama/ortalama yöntemi yerine ortalama/sigma yöntemi tercih edilir. Bunun sebebi, b -parametre kestirimlerinin a -parametre kestirimlerine göre daha kararlı olmasıdır. Ancak Baker ve Al-Karni (1991) ortalamaların standart sapmalardan daha kararlı olmasından dolayı ortalama/ortalama yönteminin tercih edilebileceğini belirtmiştir (Kolen ve Brennan, 2004).

Eğim A ve kesişim B sabitlerinin kestirimleri, ortak maddelerin madde parametrelerinden ziyade madde karakteristik eğrileri arasındaki farka dayalı olarak tanımlanan ölçüt fonksiyonu minimize edilerek de elde edilebilir. Bundan dolayı Stocking ve Lord (1983) bu tür ölçek dönüştürme yöntemlerine karakteristik eğri yöntemleri adını vermiştir. Karakteristik eğri yöntemlerinde, bir test

karakteristik fonksiyonu oluşturmak üzere toplanan yeni forma ait madde tepki eğrileri, eski formun ölçeğine dönüştürülür. Haebara ve Stocking-Lord literatürde yaygın olarak kullanılan iki karakteristik eğri ölçek dönüştürme yöntemidir (Kim, 2004).

Haebara yönteminde, karakteristik eğriler arasındaki farkın karelerinin toplamı minimize edilerek ölçüt fonksiyon elde edilir. Madde karakteristik eğrileri arasındaki fark, belirli bir yetenek düzeyindeki bireylerin her bir madde için elde edilen madde karakteristik eğrileri arasındaki farkın karesi alınıp toplanarak bulunur. Belirli bir θ_i için maddeler üzerinden fark karelerinin toplamı matematiksel olarak şu şekilde elde edilebilir (Kolen ve Brennan, 2004):

$$Hdiff(\theta_i) = \sum_{j=v} \left[p_{ij}(\theta_{ji}; \hat{a}_{jj}, \hat{b}_{jj}, \hat{c}_{jj}) - p_{ij}(\theta_{ji}; \frac{\hat{a}_{jj}}{A}, A\hat{b}_{jj} + B, \hat{c}_{jj}) \right]^2 \quad (1.10)$$

1.10 numaralı eşitlikteki ($j:v$) toplamın ortak maddeler üzerinden alındığını göstermekte olup eşitlikte, iki ölçekte yer alan karakteristik eğriler arasındaki farkın karesi alınıp toplanmaktadır. Daha sonra $Hdiff$ bütün bireyler üzerinden toplanıp 1.11 numaralı eşitlik minimize edilerek A ve B sabitleri elde edilmektedir (Kolen ve Brennan, 2004).

$$Hcrit = \sum_i Hdiff(\theta_i) \quad (1.11)$$

Stocking-Lord yönteminde ise Haebara yönteminin tam tersi olarak maddeler üzerinden toplamın farkının karesi alınmaktadır. Stocking-Lord yönteminde, karesi alınmadan önce her bir parametre seti kestirimi üzerinden toplam alınır. Bu yöntem de süreç test karakteristik eğrileri üzerinden işlemektedir (Kolen ve Brennan, 2004).

$$SLdiff(\theta_i) = \left[\sum_{j:v} p_{ij}(\theta_{ji}; \hat{a}_{jj}, \hat{b}_{jj}, \hat{c}_{jj}) - \sum_{j:v} p_{ij}(\theta_{ji}; \frac{\hat{a}_{jj}}{A}, A\hat{b}_{jj} + B, \hat{c}_{jj}) \right]^2 \quad (1.12)$$

A ve B sabitleri aşağıdaki ölçüt minimize edilerek bulunmaktadır:

$$SLcrit = \sum_i SLdiff(\theta_i) \quad (1.13)$$

1.13 numaralı eşitlikteki toplam bireyler üzerinden alınmaktadır. A ve B sabitleri için 1.11 ve 1.12 numaralı eşitlikleri çözmek iteratif bir yaklaşımı gerektirmektedir.

Moment yöntemleri A ve B sabitlerinin kestirmedeki kolaylığından dolayı tercih edilirken Stocking – Lord yöntemi daha doğru sonuçlar verdiği gerekçesi ile genellikle ölçüt yöntem olarak kullanılmaktadır (Kim ve Lee, 2006).

1.6.2.4. Madde Tepki Kuramına Dayalı Test Eşitleme

Madde ve yetenek parametreleri kestirilip aynı ölçekte elde edildikten sonra MTK gerçek puan eşitleme ya da MTK gözlenen puan eşitleme yapılarak testler birbirine eşitlenebilir. MTK gerçek puan eşitleme yönteminde, bir test formuna ait θ yetenek düzeyi ile ilişkili gerçek puanın diğer formun θ yetenek düzeyi ile ilişkili gerçek puanına eşdeğer olduğu kabul edilir:

$$\tau_X(\theta_i) = \tau_Y(\theta_i) \quad (1.14)$$

MTK gerçek puan eşitleme üç aşamalı süreçte tamamlanır:

1. X formundan bir gerçek puan (τ_X) seçilir.
2. τ_X 'e karşılık gelen θ_i değeri belirlenir.
3. Y formunda θ_i değerine karşılık gelen gerçek puan (τ_Y) bulunur.

Bu süreç, X formunda yer alan tüm gerçek puan değerleri için tekrar edilir (Kolen ve Brennan, 2004).

MTK'ya dayalı yürütülen bir diğer eşitleme yöntemi MTK gözlenen puan eşitlemedir. MTK gözlenen puan eşitleme yönteminde, MTK modelleri kullanılarak her bir formun gözlenen puan dağılımları kestirildikten sonra eşit yüzdelikli eşitleme yöntemi ile puanlar birbirine eşitlenir. MTK gözlenen puan eşitleme yöntemi, dört basamaktan oluşan bir süreçte tamamlanır (Kolen ve Brennan, 2004):

1. Her bir formun koşullu dağılımı $f(x | \theta_i)$ ve $g(y | \theta_i)$, Lord-Wingersky yineleme formülü ile kestirilir.
2. İki grupta yer alan bireylerin yetenek dağılımları kullanılarak her bir formun ve her bir popülasyonun gözlenen puan dağılımları 1.15'ten 1.18'e kadar verilen eşitlikleri yardımıyla elde edilir:

$$f_1(x) = \sum_i f(x | \theta_i) \phi_1(\theta_i) \quad (1.15)$$

$$f_2(x) = \sum_i f(x | \theta_i) \phi_2(\theta_i) \quad (1.16)$$

$$g_1(y) = \sum_i g(y | \theta_i) \phi_1(\theta_i) \quad (1.17)$$

$$g_2(y) = \sum_i g(y | \theta_i) \phi_2(\theta_i) \quad (1.18)$$

Eşitliklerde yer alan φ_1 ve φ_2 Popülasyon 1 ve Popülasyon 2'nin yetenek dağılımlarını göstermektedir.

3. NEAT deseni altında iki farklı popülasyon bulunmasına rağmen genellikle Popülasyon 1 ve Popülasyon 2'nin bileşimi olan tek bir popülasyon için eşitleme ilişkisi tanımlanır. İki popülasyonun bileşimine *sintetik* popülasyon adı verilir. X ve Y formunun sintetik popülasyonları olan $f_s(x)$ ve $g_s(y)$ için gözlenen puan dağılımları, 1.15' ten 1.18'e kadar olan denklemler aşağıda verilen $f_s(x)$ ve $g_s(y)$ eşitliklerinde yerine konularak kestirilir.

$$f_s(x) = w_1 f_1(x) + w_2 f_2(x), \quad (1.19)$$

$$g_s(y) = w_1 g_1(y) + w_2 g_2(y), \quad (1.20)$$

Popülasyon 1 ve Popülasyon 2'nin ağırlıkları w_1 ve w_2 olmak üzere;

$$w_1 + w_2 = 1 \text{ ve } w_1, w_2 \geq 0 \text{ dir.}$$

4. Son olarak $f_s(x)$ ve $g_s(y)$ arasında eşit yüzdellik ilişkisi bulunarak puanlar birbirine eşitlenir.

Kolen ve Brennan (2004) MTK gerçek puan eşitlemenin MTK gözlenen puan eşitlemeye üstünlüğünün hesaplama kolaylığı ve elde edilen dönüşümün grubun yetenek dağılımından bağımsız elde edilebilmesi, sınırlılığının ise uygulamada var olmayan gerçek puanları eşitlemesi olarak açıklamıştır. MTK gözlenen puan eşitlemenin üstünlüğü ise gözlenen puanlar arasındaki eşitleme ilişkisini tanımlamasıdır.

1.6.3. Eşitleme Sonuçlarının Değerlendirilmesi

Eşitleme çalışmalarından elde edilen sonuçları değerlendirmek için birçok ölçüt geliştirilmiştir. Harris ve Crouse (1993) eşitleme çalışmalarında kullanılan değerlendirme ölçütlerine ilişkin yaptıkları kapsamlı çalışmalarında bu ölçütlerden bazılarını şöyle sıralamışlardır: RMSD (root mean square differences), MSE (mean square error), yanlılık (bias), eşitlemenin standart hatası ve RMSE (root mean squared error).

Eşitleme sonuçlarının yeterliğini değerlendirmede kullanılan bir diğer ölçüt eşitleme sonuçlarının eşitlik özelliğini sağlayıp sağlamadığının belirlenmesidir.

Eşitlemenin eşitlik özelliği ilk kez Lord (1980) tarafından tanımlanmıştır. Lord' un eşitlik özelliği ancak bir bireyin X ya da Y formunu almasının fark oluşturmadığı durumda sağlanır. Lord' un eşitlik özelliği iki test formu birebir birbirinin aynısı olmadığı sürece sağlanamayacaktır. Böyle bir durumda da zaten eşitleme yapmak gereksizdir (Kolen ve Brennan, 2004). Bu sebeple Divgi (1981) ve Morris (1982) zayıf eşitlik özelliği ya da birinci-sıra eşitlik (first-order equity) özelliği de denilen bir kavramı öne sürmüştür. Lord'un eşitlik özelliğine göre daha esnek olan bu kavrama Yen (1983) tarafından tau-eşdeğerliği de denilmiştir. Birinci-sıra eşitlik (BSE) özelliği, eşitleme yapıldıktan sonra alternatif formların koşullu ortalamalarının eşit olmasını gerektirmektedir. Morris (1982), ayrıca eşitleme yapıldıktan sonra alternatif formlara ait ölçmenin koşullu standart hatasının (SEM) eşit olması durumunda sağlanan ikinci-sıra eşitlik (second-order equity; İSE) özelliğini önermiştir (akt. Harris ve Crouse, 1993). Matematiksel olarak BSE ve İSE özellikleri sırayla 1.21 ve 1.22 eşitliklerinde olduğu gibi ifade edilebilir:

$$E[eq_Y(X) | \tau] = E(Y | \tau) \quad (1.21)$$

$$SEM_{eq_Y(x) | \tau} = SEM_{Y | \tau} \quad (1.22)$$

1.21 ve 1.22 eşitliklerinde yer alan τ örtük yeteneği, $eq_Y(X)$ X formundaki bir puanı Y formunun ölçeğine dönüştüren eşitleme fonksiyonunu göstermektedir. BSE özelliği Tong ve Kolen (2005) tarafından geliştirilen D_1 indeksi hesaplanarak değerlendirilebilir:

$$D_1 = \frac{\sqrt{\sum_i q_i \{E[Y | \theta_i] - E[e\hat{q}_Y(x) | \theta_i]\}^2}}{SD_Y} \quad (1.23)$$

D_1 eşitliğindeki; $E[Y | \theta_i]$ bir θ_i yetenek düzeyi için eski formun koşullu ortalamasıdır. $E[e\hat{q}_Y(x) | \theta_i]$ ise bir θ_i düzeyi için eşitlenmiş puanların koşullu ortalaması; $q_i \theta_i$ 'deki quadrature ağırlığıdır. Son olarak, SD_Y Y formunun standart sapmasıdır. D_1 indeksi değeri küçüldükçe BSE özelliği daha iyi korunmaktadır.

İSE özelliğinin değerlendirilmesine ilişkin D_2 indeksi yine Tong ve Kolen (2005) tarafından geliştirilmiştir:

$$D_2 = \frac{\sqrt{\sum_i q_i (SEM_{Y | \theta_i} - SEM_{e\hat{q}_Y(x) | \theta_i})^2}}{SD_Y} \quad (1.24)$$

D_2 indeksinde yer alan $SEM_Y | \theta_i$ eski formu alan θ_i yetenek düzeyindeki bireylere ait koşullu ölçmenin standart hatası (SEM); $SEM_{\hat{q}_Y(x)} | \theta_i$ eşitlenen yeni formu alan θ_i yetenek düzeyindeki bireylere ait koşullu ölçmenin standart hatasıdır. Benzer şekilde D_2 indeksi değeri küçüldükçe İSE özelliği daha iyi korunmaktadır.

Bireylerin X ya da Y formlarından herhangi birini almaktan dolayı oluşacak avantaj ya da dezavantajın önüne geçmek için BSE özelliğinin; ölçmenin kesinliğini (precision) arttırmak ve ölçme hatalarının azalmasını sağlamak için ise İSE özelliğinin korunması önemlidir (He ve Kolen, 2011). Eğer BSE ve İSE özellikleri korunmaz ise alternatif formlardan elde edilen puanların birbiri yerine kullanılabilmesi durumu oluşmaz. Örneğin, eğer BSE özelliği sağlanmaz ise test formlarından birinde belirli bir yetenek düzeyinde beklenen ölçek puanı daha yüksek olacak, o test formunu alan öğrenciler diğer formu alan aynı yetenek düzeyindeki öğrencilerden daha avantajlı olacaklardır. Benzer bir durum BSE özelliği korunup İSE özelliği korunmadığında da oluşur. Böyle bir durumda, iki test formunun koşullu ortalamaları geniş bir yetenek aralığı için eşit olmasına rağmen alt düzey yeteneğe sahip öğrenciler varyansı daha büyük olan test formunu almak isteyeceklerdir. Çünkü alt düzey yeteneğe sahip öğrenciler, ölçme hatasından daha çok yarar sağlar durumda olacaklardır. Üst düzey yeteneğe sahip öğrenciler ise üst düzey yeteneğin daha doğru ölçüldüğü varyansı küçük olan formu almak isteyeceklerdir (Andrews, 2011).

Kolen ve Lee (2011), BSE ve İSE özelliklerinin alternatif formlardan elde edilen puanların değerlendirilmesi için kullanışlı bir ölçüt olduğunu belirtmiş; bu ölçütlerin yeterince sağlanmadığı durumlarda test geliştirme sürecinin yeniden gözden geçirilip formların birbirine daha benzer olacak şekilde düzenlenmesi gerektiğini belirtmiştir. Test eşitleme çalışmalarında eşitlik özelliklerinin değerlendirilmesi, bir eşitleme yönteminin formlar arasındaki güçlük farklılıklarını ne kadar iyilikte düzenlediğinin önemli bir kanıtı olarak kullanılmaktadır. Dolayısıyla bu araştırmada değerlendirme ölçütü olarak eşitlik özelliği seçilmiş ve çeşitli simülasyon koşulları altında elde edilen eşitleme sonuçlarının BSE ve İSE özelliklerini korumadaki performansları karşılaştırılmıştır.

2. İLGİLİ ARAŞTIRMALAR

2.1. Karma Testlerin Eşitlenmesini Etkileyen Faktörlerin İncelendiği Araştırmalar

Bastari (2000), yürüttüğü simülasyon çalışmasında karma testlerin eşitlenmesini etkileyen çeşitli faktörleri (test uzunluğu, ÇS maddelerin AU maddelere oranı, ortak madde seti uzunluğu, grupların yetenek dağılımı, madde kalibrasyon yöntemleri ve örneklem büyüklüğü) NEAT deseni altında incelemiştir. Madde ve yetenek parametrelerinin türetilmesi ve kestirilmesinde 3PL/GPC model kompozisyonunu kullanmıştır. Araştırmanın sonucunda: testin ve ortak madde setinin uzunluğu, testte yer alan çoktan seçmeli madde oranı ve örneklem büyüklüğü arttıkça, gruplar arasındaki yetenek dağılımı farkı azaldıkça daha doğru eşitleme sonuçları elde edildiği bulunmuştur. Ayrıca, eş zamanlı kalibrasyon sonucu daha doğru eşitleme sonuçları elde edildiği görülmüştür.

Kirkpatrick (2005), çalışmasında karma testlerin eşitlenmesine ortak madde seti kompozisyonunun etkisini incelemiştir. Araştırma, hem gerçek veri hem de simülatif veri üzerinden yürütülmüştür. Gerçek veri ile farklı madde formatlarının eşitlemeye etkisi incelenirken simülasyon çalışması ile bu etkilerin nedenleri incelenmiştir. MTK gerçek puan eşitleme, gözlenen puan eşitleme ve yetenek kestirimi eşitlemesi (estimated ability equating) olmak üzere üç eşitleme yöntemi kullanılmıştır. Simülasyon çalışmasında madde formatına bağlı olarak oluşan boyutluluk (tek boyutlu, $\rho=0.50$ ve $\rho=0.80$ olan ilişkili iki boyut) ve grupların yetenek dağılımı olmak üzere iki faktör incelenmiştir. Elde edilen eşitleme sonuçları eşdeğer puanların ortalamasına ait delta grafikleri, kesme puanlar ve dönüştürme sabitlerine ait standart hatalar ölçüt alınarak değerlendirilmiştir. Kirkpatrick (2005) araştırmasının sonucunda şu temel bulgulara ulaşmıştır: Gerçek veri setinde madde formatı etkilerinin eşitleme yöntemi ve veri setine göre farklılık göstermektedir. Madde formatına dayalı oluşturulan boyutlar arasındaki korelasyon 0.50'ye düştüğünde eşitleme sonuçlarında önemli farklılıklar olduğu görülmüştür. Simülasyon çalışması, eşitleme sonuçlarının evren yetenek dağılımının bir madde formatından diğerine farklılaşmasından etkilendiğini göstermiştir. Bu araştırma, farklı madde formatlarındaki performans

farklılaşmasının eşitleme sonuçlarını madde formatları arasındaki korelasyondan daha çok etkilediğini göstermiştir.

Kim ve Kolen (2006) yürüttükleri simülasyon çalışmasında, karma testlerde dört farklı MTK ölçek dönüştürme yönteminin ve eş zamanlı kalibrasyonun madde formatına dayalı oluşan çok boyutluluğa olan dayanıklılığını incelemiştir. Araştırmada üç düzey format etkisi, iki düzey karma-test türü ve üç düzey grupların yetenek dağılımı farklılığı olmak üzere üç temel faktör incelenmiştir. Değerlendirme ölçütü olarak MSE (mean square error) ve OSD (observed score distribution) indisleri kullanılmıştır. Araştırmanın sonucunda, ölçek dönüştürme yöntemleri içinde karakteristik eğri yöntemlerinin moment yöntemlerine göre daha iyi performans gösterdiği fakat genel olarak eş zamanlı kalibrasyonun format etkisine ölçek dönüştürme yöntemlerinden daha dayanıklı olduğu görülmüştür.

Kim ve Lee (2006), dört ölçek dönüştürme yönteminin (ortalama-ortalama, ortalama-sigma, Haebara, and Stocking-Lord) uzantısını karma testler için geliştirmiş ve dört yöntemin farklı koşullar altındaki performansını incelemiştir. Simülatif verinin kullanıldığı çalışmada 3PL/GPC modelleri kullanılarak karma testler oluşturulmuştur. Araştırmada grupların yetenek dağılımı, örneklem büyüklüğü, karma formatta oluşturulan test kompozisyonu ve farklı ortak madde seti kompozisyonu (karma, sadece çoktan seçmeli, sadece açık uçlu) olmak üzere 4 temel faktör incelenmiştir. SB (square bias) ve MSE (mean squared errors) indekslerinin değerlendirme ölçütü olarak kullanıldığı araştırmanın sonucunda: karakteristik eğri yöntemlerinin moment yöntemlerinden daha az eşitleme hatası verdiği ve Haebara yönteminin dört yöntem içinde en az hata veren yöntem olduğu bulunmuştur. Ayrıca ortak madde setinin karma oluşturulduğu koşulların ortak madde setinde tek bir madde türünün kullanıldığı durumlardan daha iyi performans gösterdiği ve eşdeğer gruplarla yürütülen eşitlemenin eşdeğer olmayan grupla yürütülenden daha az hataya sahip olduğu görülmüştür.

Cao (2008), yaptığı çalışmasında ortak madde setinin istatistiksel, kapsam ve format özellikleri bakımından temsil edebilirliğinin karma testlerin eşitlenmesine etkisini incelemiştir. Simülasyon çalışması ile testin boyutluluğu, grupların yetenek dağılımları ve ortak madde setinin istatistiksel, kapsam ve format bakımından temsil edebilirliği olmak üzere beş faktörü manipüle etmiştir. Araştırma, NEAT

kullanılarak yürütülmüştür. Cao (2008), araştırmasında ayrıca tek boyutlu MTK eşitleme yöntemlerini ve bu yöntemlerin çeşitli çok boyutluluk test yapılarındaki dayanıklılığını incelemiştir. Araştırmada değerlendirme ölçütü olarak BIAS, RMSE ve sınıflama tutarlılığı indisleri kullanılmıştır. Araştırma sonucunda şu temel bulgulara ulaşılmıştır: Tüm simülasyon koşulları dikkate alındığında grupların yetenek dağılımı faktörünün eşitleme sonuçları üzerinde önemli etkisinin olduğu gözlenmiştir. Eşdeğer grup koşulu her zaman eşdeğer olmayan grup koşulundan daha iyi sonuçlar vermiştir. İkinci olarak; tek boyutlu yapı altında kapsam ve format temsil edebilirliği faktörlerinin eşitleme sonuçları üzerinde küçük bir etkisinin olduğu gözlenmiştir. Ancak, istatistiksel temsil edebilirlik faktörü eşzamanlı kalibrasyonu anlamlı şekilde etkilemiştir. Madde formatına bağlı çok boyutluluk arttıkça format temsil edebilirliğinin özellikle eşdeğer olmayan gruplarda önemli farklılıklar oluşturduğu tespit edilmiştir. Ancak, farklı kapsam alanlarına dayalı olarak çok boyutluluğun artması eşitleme sonuçları üzerinde etkili olmamıştır. Son olarak, tek boyutluluk varsayımının ihlaline eş zamanlı kalibrasyonun çok dayanıklı olmadığı çok boyutluluğun derecesi arttıkça tek boyutlu MTK modellerinin anlamlı ölçüde performanslarının azaldığı görülmüştür.

Hagge (2010), çalışmasında karma testlerin özelliklerinin ve ortak madde setinin kompozisyonunun NEAT deseninde eşitleme sonuçlarının doğruluğuna etkisini incelemiştir. Analizler operasyonel ve psödo-test formlarına ait madde cevapları kullanılarak yapılmıştır. Operasyonel test formları ile gerçekleştirilen analizlerde iki faktör incelenmiştir; (1) eski ve yeni test formunu alan grupların yeterliklerindeki farklılıklar ve (2) çoktan seçmeli maddeler ile açık uçlu maddelerin göreceli güçlükleri. Psödo-test formları ile ise bu faktörlere ek olarak ortak madde setinin format ve istatistiksel olarak temsil edebilirliği incelenmiştir. Her bir koşul için frekans kestirimi ve zincirleme eşit yüzdelli eşitleme olmak üzere iki geleneksel; MTK gerçek ve MTK gözlenen puan eşitleme olmak üzere iki MTK eşitleme yöntemi dikkate alınmıştır. Araştırmanın sonucunda; eski ve yeni formları alan grupların yeterlikleri arasındaki farklılık arttıkça yanlılığın arttığı; yanlılıktaki artışın en fazla frekans kestirimi yönteminde, en az MTK eşitleme yöntemlerinde olduğu gözlenmiştir. Eşitlemenin standart hatasının MTK gözlenen puan eşitlemede en küçük olma eğilimindeyken zincirleme eşit yüzdelli eşitlemede en büyük olma

eğiliminde olduğu bulunmuştur. Psödo-test formları operasyonel test formlarına benzer olarak oluşturulduğunda elde edilen sonuçların benzer olduğu görülmüştür.

Tian (2011), yaptığı simülasyon çalışmasında karma testlerde MTK eşitleme yöntemlerinin performansını değerlendirmiştir. Çalışmasını 55 çoktan seçmeli 15 açık uçlu madde ile NEAT deseni kullanarak yürütmüştür. Madde cevaplarını 2003 TIMSS 8. Sınıf matematik test maddelerine dayalı olarak türetmiştir. Araştırmada iki farklı ölçek dönüştürme yöntemi (Stocking Lord ve eş zamanlı kalibrasyon), üç farklı ortak madde seti uzunluğu ve üç çeşit ortak madde seti (karma, sadece ÇS ve sadece AU) olmak üzere toplam 18 koşulu incelemiştir. Değerlendirme ölçütü olarak MSE (mean square error), SD (standart sapma) ve RMSD (root mean square difference) indekslerini kullanmıştır. Araştırmanın sonucunda, eş zamanlı kalibrasyonun madde parametrelerinin telafisinde genellikle daha iyi performans gösterdiğini ve daha doğru sonuçlar verdiğini bulmuştur.

Kim ve Walker (2012), hem çoktan seçmeli hem açık uçlu madde içeren karma testlerde ortak madde kompozisyonunun uygunluğunu alt grup değişmezliği indislerini kullanarak değerlendirmişlerdir. Araştırma, NEAT deseni ile a) sadece çoktan seçmeli ve b) çoktan seçmeli-açık uçlu ortak madde seti kullanılarak yürütülmüştür. Her bir ortak madde seti koşulunda bağlanma fonksiyonları hem kızlar hem de erkekler için elde edilmiş ve toplam gruptan elde edilen fonksiyonla karşılaştırılmıştır. Eşitleme, zincirleme eşit yüzdelikli eşitleme yöntemi kullanılarak yapılmıştır. Ortak madde setinin sadece çoktan seçmeli maddelerden oluştuğu durumda, alt gruplardan elde edilen fonksiyon ile grubun tamamından elde edilen fonksiyon arasındaki farkın kesme puanları da içeren puan bölgesinde önemli olduğu, özellikle düşük performanslı bireyler için verilecek geçme/kalma kararlarında tutarsızlığa neden olduğu görülmüştür. Araştırmanın sonucunda, erkekler ve kızlar arasındaki alt grup değişmezliğini sağlamada çoktan seçmeli ve açık uçlu maddelerden oluşan karma ortak madde setinin daha iyi olduğu görülmüştür.

Wang (2013), simülatif veri kullanarak yaptığı çalışmasında NEAT deseni altında test ve birey özelliklerinin karma testlerin eşitlenmesine etkisini incelemiştir. Araştırma kapsamında madde türüne bağlı boyutluluk, grupların yetenek dağılımı, ortak madde setinin uzunluğu, kompozisyonu ve ortak madde setinin format

açısından temsil edilebilirliği olmak üzere beş ana faktör incelenmiştir. Öndüzgünleştirilmiş frekans kestirim eşitleme yöntemi ve öndüzgünleştirilmiş zincirleme eşit yüzdelikli eşitleme yöntemlerini farklı koşullar altında mutlak yanlılık (absolute bias), eşitlemenin standart hatası ve RMSD indekslerine dayalı olarak karşılaştırmıştır. Araştırmanın sonucunda, madde türüne bağlı olarak oluşan çok boyutluluğun tesadüfi hata üzerinde önemli etkisinin olmadığı fakat ortak madde setinin oluşturulma şekline bağlı olarak yanlılık üzerinde etkisinin olduğu bulunmuştur. Benzer şekilde grupların yetenek dağılımı farklılığının tesadüfi hata üzerinde önemli etkisinin olmadığını, gruplar farklılaştıkça daha uzun ortak madde setine sahip koşulların daha az eşitleme hatasına sahip olduğu görülmüştür. Grupların yetenek dağılımları arasında fark oluştuğunda karma formatta oluşturulan ortak madde seti daha doğru eşitleme sonuçları vermiştir. Grupların yetenek dağılımlarının eşdeğer olduğu ve tek boyutluluğun sağlandığı koşullarda sadece çoktan seçmeli maddelerden oluşan ortak madde seti, karma formatta oluşturulan ortak madde seti ile benzer performans göstermiştir.

2.2. Eşitlik Özelliğinin Değerlendirme Ölçütü Olarak Alındığı Araştırmalar

Bolt (1999), MTK gerçek puan eşitlemenin performansını çok boyutluluğun olduğu çeşitli koşullar altında geleneksel eşitleme yöntemleri ile karşılaştırarak incelemiştir. Değerlendirme ölçütü olarak BSE ve İSE özelliğini kullandığı çalışmasını simülatif veri kullanarak yürütmüştür. Araştırmanın simülasyon faktörlerini madde güçlüğü ile test yapısı arasındaki etkileşim ile boyutluluk düzeyi (0.3, 0.5, 0.7 ve 1.0) oluşturmuştur. Araştırmanın sonucunda, iki boyut arasındaki korelasyonun yüksek (≥ 0.7) olduğu durumlarda MTK gerçek puan eşitlemenin eşit yüzdelikli eşitlemeden daha iyi performans gösterdiği; boyutlar arasındaki korelasyonun düşük olduğu koşullar altında ise MTK eşitleme yöntemlerinin en az eşit yüzdelikli eşitleme kadar iyi olduğu görülmüştür.

Kim, Brennan ve Kolen (2005), çalışmalarında dört eşitleme yöntemini (3PL gerçek puan eşitleme, 3PL gözlenen puan eşitleme, beta 4 gerçek puan eşitleme ve beta 4 gözlenen puan eşitleme) yöntemlerini BSE, İSE, CMSED (conditional-mean-square-error-difference) ve eşit yüzdelikli eşitleme özelliği ölçütlerine göre incelemiştir. Araştırma sadece çoktan seçmeli maddelerden oluşan gerçek veri kullanılarak eşdeğer gruplar deseni altında yürütülmüştür. Araştırmanın

sonucunda, gerçek puan dağılımının model alındığı durumda gerçek puan eşitleme yönteminin BSE özelliğini korumada gözlenen puan eşitlemede daha iyi olduğu; İSE özelliği, CMSED ve eşit yüzdellikli eşitleme özelliği ölçüt olarak alındığında ise gözlenen puan eşitlemenin eşit yüzdellikli ve gerçek puan eşitlemeden daha iyi performans gösterdiği bulunmuştur.

Tong ve Kolen (2005), üç farklı eşitleme yönteminin (ön-düzenleştirilmiş eşit yüzdellikli eşitleme, MTK gerçek ve gözlenen puan eşitleme) performansını eşitlik özelliğinin korunumunu eşdeğer gruplar deseni altında incelemiştir. Araştırma gerçek ve simülatif veri kullanılarak yürütülmüştür. Araştırmanın sonucunda, eğer eşitlenecek test formları benzer ise eşitlik özelliklerini korumada tüm yöntemlerin benzer performansa sahip olduğu görülmüştür. Ancak, test formları birbirinden çok farklılık gösteriyorsa eşit yüzdellikli eşitleme ve MTK gözlenen puan eşitlemenin aynı dağılım özelliğini korumada; MTK gerçek-puan eşitlemenin BSE özelliğini korumada ve MTK gözlenen-puan eşitlemenin İSE özelliğini korumada en iyi performans gösterdiği sonucuna ulaşılmıştır.

He (2011), araştırmasında, sadece çoktan seçmeli maddelerin ortak madde setinde yer aldığı durumda karma testlerin eşitlenmesini etkileyen faktörleri araştırmıştır. Eşitleme sonuçlarını aynı dağılıma sahip olma, BSE ve İSE özelliği olmak üzere üç farklı ölçüte göre değerlendirmiştir. Araştırma gerçek veri kullanılarak yürütülmüş ve araştırmada şu faktörler incelenmiştir: ÇS ve AU maddeler arasındaki korelasyon, ortak madde oranı, ÇS madde puanı oranı ve iki formun benzerliği. Frekans kestirimi (FE), zincirleme eşit yüzdellikli eşitleme (CE), MTK gerçek puan eşitleme ve MTK gözlenen puan eşitleme olmak üzere dört eşitleme yöntemi kullanılmıştır. Araştırmanın sonucunda şu temel bulgulara ulaşılmıştır: (1) MTK yöntemleri arasında MTK gerçek puan eşitlemenin gözlenen puan eşitlemeden BSE özelliğini daha iyi koruduğu; MTK gözlenen puan eşitlemenin İSE özelliğini ve aynı dağılıma sahip olma özelliklerini MTK gerçek puana göre daha iyi koruduğu görülmüştür. (2) Geleneksel eşitleme yöntemleri içerisinde CE yöntemi BSE eşitlik özelliğini daha iyi korurken hem FE hem CE yöntemi BSE ve aynı dağılıma sahip olma özelliklerini korumada benzer sonuçlar vermiştir. (3) Eşit yüzdellikli eşitlemede kullanılan düzenleştirmenin İSE ve aynı dağılıma sahip olma özelliğinin korunmasına yardımcı olduğu tespit edilmiştir. (4)

Her iki MTK eşitleme yönteminde ÇS ve AU maddeler arasındaki korelasyonun yüksek olmasının BSE özelliğinin korunması ile ilişkili olduğu görülmüştür. (5) MTK gerçek puan eşitleme için ÇS ve AU maddeler arasındaki korelasyonun yüksek olması ile İSE özelliğinin korunması arasında ilişki olduğu bulunmuştur. (6) MTK gözlenen puan eşitleme için ÇS ve AU maddeler arasındaki korelasyonun yüksek olması ile aynı dağılıma sahip olma özelliğinin korunması arasında ilişki olduğu görülmüştür. (7) Ortak madde oranı, ÇS madde oranı ve iki formun benzerlikle eşitlik özelliklerinin korunması arasında bir ilişki bulunmamıştır.

Andrews (2011), karma testlerin eşitlenmesinde eşitlik özelliğinin korunumunu tek boyutlu ve çok boyutlu MTK altında incelenmiştir. Araştırma gerçek ve simülatif veri ile beş farklı eşitleme (frekans kestirimi, zincirleme eşit yüzdelli eşitleme, tek boyutlu MTK gerçek puan ve gözlenen puan eşitleme, çok boyutlu MTK gözlenen puan eşitleme) yöntemi ile NEAT deseni kullanılarak yürütülmüştür. Yetenek dağılımları arasındaki korelasyon (0.50, 0.65, 0.80, 0.95), boyutluluk (tek boyutlu-çok boyutlu), değerlendirme ölçütü (beklenen puan-CSEM), yetenek kestirimi (MLE, EAP, TCC), puan türü (ham puan-ölçek puanı) olmak üzere beş faktörü incelemiştir. Gerçek verilerden elde edile analizler, test eşitleme yöntemlerinin performanslarının eşitlik özelliğini değerlendirmek için kullanılan çerçeveye (tek boyutlu-çok boyutlu) göre değiştiğini göstermiştir. Simülasyon çalışmasının sonucunda, yetenekler arasında korelasyona bağlı olmaksızın BSE özelliğinin korunmasında tek boyutlu MTK gerçek puan eşitleme, gözlenen puan eşitlemeden daha iyi performans göstermiştir. Tam tersi olarak İSE özelliğinin korunmasında ise MTK gözlenen puan eşitleme daha iyi performans göstermiştir. Çok boyutluluğun derecesi arttıkça, çok boyutlu MTK gözlenen puan eşitleme yöntemi eşitlik özelliğini en iyi koruyan yöntem olmuştur. Yetenekler arasına korelasyon azaldıkça ve gruplar arasındaki farklılıklar arttıkça tek boyutlu MTK eşitleme yöntemleri geleneksel eşitleme yöntemleri ile aynı performansı göstermiştir.

Duong (2011), simülasyon verisi kullanarak yürüttüğü çalışmasında çeşitli geleneksel (FE, CE) ve MTK'na dayalı test eşitleme yöntemlerini (GR_PE ve GZ_PE) eşit yüzdelli eşitleme, tam eşitlik, BSE ve İSE ölçütlerine göre karşılaştırmıştır. Araştırmada NEAT deseni kullanılmış, madde cevapları 3PL model ile türetilmiştir. Araştırma kapsamında test formları arasındaki farklılık ve

grupların yetenek dağılımları simülasyon faktörü olarak ele alınmıştır. Araştırmanın sonucunda, grupların yetenek dağılımları eşit olduğunda dört eşitleme yönteminin kullanılan değerlendirme ölçütüne bağlı olmaksızın benzer sonuçlar verdiği görülmüştür. Formlar arasında büyük farklılıkların olduğu koşullarda tüm eşitleme yöntemlerinin kötü sonuçlar verdiği görülmüştür. Araştırmanın bulgularına göre GR_PE yöntemi, BSE özelliğini korumada en iyi performans gösteren yöntem iken GZ_PE yöntemi ise İSE özelliğini korumada hiçbir eşitlemenin yapılmadığı (identity equating) yöntemden sonra en iyi performans gösteren yöntem olmuştur.

Lee, Lee ve Brennan (2012), çeşitli eşitleme yöntemlerinin karma testlerin eşitlenmesinde eşitlik özelliğini korumadaki performansını, psikometrik modellerin etkisini, test güvenilirliği ve eşitlik özelliği arasındaki ilişkiyi incelemiştir. Araştırma gerçek veri kullanılarak yürütülmüştür. Araştırmanın sonucunda, MTK gerçek-puan eşitleme yönteminin BSE özelliğini korumada en iyi performans gösterdiği; MTK gözlenen-puan eşitlemenin ise İSE özelliğini korumada en iyi olduğu görülmüştür.

Wolf (2013), çalışmasında dört farklı eşitleme yöntemi kullanarak (MTK gerçek puan, MTK gözlenen puan, frekans kestirim ve zincirleme eşit yüzdelikli eşitleme) NEAT deseni altında karma testlerin eşitlenmesinde eşitlik özelliğinin korunumunu etkileyen faktörleri incelemiştir. Araştırma simülatif veri kullanılarak yürütülmüştür. Test yapısı (tek boyutlu-çok boyutlu), eşdeğer olmayan gruplar ve ortak madde seti özellikleri (format açısından temsil eden-etmeyen) olmak üzere üç temel faktör incelenmiştir. Araştırmanın sonucunda, MTK eşitleme yöntemlerinin eşitlik özelliğini korumada tüm koşullar altında geleneksel eşitleme yöntemlerinden daha iyi performans gösterdiği; gruplar eşdeğer olduğunda ise geleneksel eşitleme yöntemlerinin MTK yöntemleri ile benzer performans gösterdiği görülmüştür. BSE özelliğinin korunmasında MTK gerçek-puan eşitleme yönteminin test yapısına bağlı olmadan en iyi performans gösterdiği, tek boyutluluk koşulu sağlandığında ise İSE özelliğini korumada tüm yöntemlerin benzer performans gösterdiği bulunmuştur. Çok boyutluluk olduğu durumlarda, MTK gerçek puan eşitleme yöntemi İSE özelliğini korumada en iyi performansı göstermiştir. Grupların yetenek dağılımları arasındaki fark arttıkça karma formatta ortak madde setine sahip koşullar daha düşük eşitlik indekslerine sahip olmuştur.

2.3. İlgili Arařtırmalar Özet

Karma testlerin eşitlenmesi üzerine yapılan arařtırmaları genel olarak özetleyecek olursak; eşitlenecek testlerin karakteristik özelliklerinin (örneğin boyutluluğunun, uzunluğunun), ortak madde seti özelliklerinin (örneğin uzunluğunun, testin genelini temsil edip etmemesinin), örneklem büyüklüğünün ve grupların yetenek dağılımının (eşdeğer gruplar, eşdeğer olmayan gruplar) karma testlerin eşitlenmesinde etkili faktörler olduğunu söyleyebiliriz.

Eşitlik özelliğinin değerlendirme ölçütü olarak alındığı arařtırmalarda çoğunlukla eşitleme yöntemlerinin karşılaştırıldığını söyleyebiliriz. Yapılan arařtırmalar incelendiğinde genel olarak MTK gerçek puan eşitlemenin BSE özelliğini korumada daha iyi performans gösterme eğiliminde olduğu, MTK gözlenen puan eşitlemenin ise İSE özelliğini korumada daha iyi performans gösterme eğiliminde olduğunu söyleyebiliriz. Ayrıca yapılan arařtırmalara dayalı olarak MTK eşitleme yöntemlerinin her iki eşitlik özelliğini korumada geleneksel eşitleme yöntemlerine göre daha iyi performans gösterdiğini söyleyebiliriz.

3. YÖNTEM

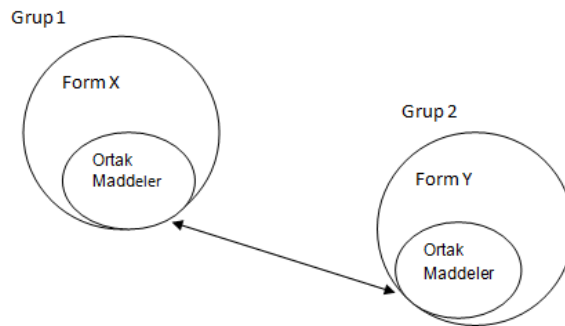
3.1. Araştırmanın Türü

Bu araştırmada; testlerin boyutluluğunun, ortak madde seti formatının, grupların yetenek dağılımının ve ölçek dönüştürme yöntemlerinin karma testlerin eşitlenmesinde eşitlik özelliğinin korunumu üzerine etkisi simülasyon verileri kullanılarak incelenmiştir. Araştırmadan elde edilen sonuçlar, mevcut teorik bilgiye doğrudan katkı getirecek nitelikte olduğu için araştırma simülasyon verileri ile yürütülen bir temel araştırmadır (Karasar, 2009).

3.2. Araştırma Deseni

Bu araştırmada testler, eşdeğer olmayan gruplar ortak test (nonequivalent groups anchor test (NEAT)) deseni kullanılarak eşitlenmiştir. Bu eşitleme deseninde, ortak maddelere sahip iki test formu farklı iki gruba uygulanır. Farklı formları alan grupların yetenek dağılımlarının eşdeğer olması beklenmemektedir. Puan farklılıklarına formlar arasındaki farklılıklar sebep olduğu kadar grup farklılıklarının etkileri de karışır. Grupların farklılığı her iki formda yer alan ortak madde seti ile giderilmeye çalışılarak ortak maddeler aracılığı ile formlar birbirine eşitlenir. NEAT deseni iki şekilde oluşturulabilir. Eğer ortak madde puanı bireyin test puanına ekleniyorsa bu tür ortak testlere iç ortak (internal anchor) test eğer bireyin puanına eklenmiyorsa dış ortak (external anchor) test denir (Kolen ve Brennan, 2004).

Bu araştırma, iç ortak test kullanılarak yürütülmüştür. X formu yeni form, Y formu eski form olarak adlandırılmış ve X formu Y formuna eşitlenmiştir. Araştırma deseni Şekil 3.1.'de görülmektedir.



Şekil 3.1. Eşdeğer olmayan gruplar ortak test deseni

Kaynak: Kolen, M. J. & Brennan, R. L. (2004). Test equating, scaling, and linking: Methods and practices (2nd ed.). New York, NY: Springer-Verlang.

3.3. Araştırma Verileri

Bu araştırma simülasyon verileri kullanılarak yürütülmüştür. Simülasyon verileri test eşitleme çalışmaları için araştırmacıya çok sayıda faktörü inceleme imkânı vermesi yönüyle değerlidir. Fakat simülasyon verileri, olabildiğince gerçek test durumlarını yansıtacak şekilde türetilmelidir (Harris ve Crouse, 1993). Gerçek test durumlarına benzer olması açısından bu araştırmada kullanılan madde cevapları, 2011 yılında uygulanan Uluslararası Fen ve Matematik Eğilimleri Araştırması (Trends in International Mathematics and Science Study–TIMSS) 8. Sınıf matematik testlerine ait yayınlanan madde parametreleri kullanılarak türetilmiştir.

TIMSS, Uluslararası Eğitim Başarısı Değerlendirme Kuruluşu (International Association for Evaluation of Educational Achievement-IEA) tarafından, katılımcı ülkelerin dört yılda bir 4. ve 8. sınıf düzeylerinde uygulanan, öğrencilerin matematik ve fen başarılarını ölçmeyi amaçlayan uluslararası bir karşılaştırma sınavıdır. TIMSS en son, 2011 yılında uygulanmıştır. TIMSS’ te çoktan seçmeli ve açık uçlu maddeler birlikte değerlendirilmektedir. Çoktan seçmeli maddeler doğru cevap için “1” yanlış cevap için “0” verilerek puanlanırken açık uçlu maddeler her bir maddeye özgü hazırlanan puanlama rehberleri kullanılarak iki farklı şekilde puanlanmaktadır (Mullis ve Martin, 2011):

- 1-puana sahip açık uçlu maddeler; doğru cevap için “1” yanlış cevap için “0” verilerek,
- 2-puana sahip açık uçlu maddeler; tamamen doğru cevap için “2”, kısmen doğru cevap için “1” ve yanlış cevap için “0” verilerek puanlanmaktadır.

TIMSS’in 2011 yılında gerçekleşen uygulamasında matris örnekleme yaklaşımı kullanılarak matematik ve fen maddelerini içeren madde havuzunun bütününden her bir sınıf düzeyi için 14 öğrenci başarı kitapçığı oluşturulmuş ve her öğrenci sadece bir kitapçıktaki soruları cevaplamıştır. TIMSS maddeleri, 8. sınıf düzeyinde her biri 12-18 madde içeren bloklar halinde gruplandırılmıştır. Her bir bloktaki maddelerin dağılımının kapsam ve bilişsel alanlar açısından madde havuzunun tümüyle eşleşmesine çalışılmıştır. 8. sınıf düzeyinde her bir kitapçıkta ortalama 31 tane matematik sorusu bulunmaktadır. Kitapçıklarda bulunan çoktan seçmeli/açık uçlu madde oranları kitapçıktan kitapçığa değişkenlik göstermekle birlikte bütün

sorulardan alınacak toplam puanın en az yarısının çoktan seçmeli sorulardan gelecek şekilde kitapçıklar düzenlenmiştir (Mullis vd., 2009).

TIMSS uygulamasındaki yer alan matematik testlerine benzerlik göstermesi açısından bu araştırma kapsamında eşitlenecek testlerin 30 maddeden oluşmasına karar verilmiştir. TIMSS' te kullanılan, çoktan seçmeli maddelerden elde edilecek puanın testten elde edilecek toplam puanın yarısından fazla olması ölçütü dikkate alınmış dolayısıyla her bir test 24 çoktan seçmeli ve 6 açık uçlu maddeden oluşturulmuştur. Bir testten elde edilecek toplam madde puanı 36'dır ($24*1+6*2=36$) ve çoktan seçmeli madde puanının açık uçlu madde puanına oranı 2'dir.

NEAT deseninde, doğru eşitleme sonuçları elde etmek için testlerin yeterli ortak madde sayısına sahip olması gerekir (Cao, 2008). Kolen ve Brennan (2004) 40 ve üzerinde madde sayısına sahip testler için ortak madde sayısının toplam testin uzunluğunun en az %20'si kadar olmasını önermiştir. Bu araştırma kapsamında, ortak madde sayısının testte yer alan toplam madde sayısının 1/3'ü kadar olmasına karar verilmiştir. Bu oran, Kim ve Kolen (2006) ile Cao (2008)'nin çalışmalarında kullandıkları oran ile aynıdır. Her bir test 30 maddeden oluştuğu için bu çalışmada kullanılan ortak madde sayısı 10'dur. Araştırmanın amacı doğrultusunda ortak test-1 ve ortak test-2 olmak üzere iki farklı ortak test oluşturulmuştur. Ortak test-1 karma test formatına sahip olup çoktan seçmeli madde sayısının açık uçlu madde sayısına oranı testin tamamındaki orana ($24/6$) eşittir. Ortak test-1, 8 çoktan seçmeli ve 2 açık uçlu maddeden ($8/2$) oluşmaktadır. Ortak test-2 ise ortak test-1'deki 2 açık uçlu madde her iki testte de ortak olmasına karar verilen 2 çoktan seçmeli madde ile değiştirilerek oluşturulmuştur. Ortak test-2 sadece çoktan seçmeli 10 madden oluşmaktadır.

Araştırma kapsamında, TIMSS 2011'in yayınlanan 8. sınıf matematik madde parametrelerinden bir madde parametre havuzu oluşturulmuştur. X ve Y testi ile ortak testi oluşturacak maddelere ilişkin parametreler, oluşturulan madde parametre havuzundan seçilmiştir. Madde parametre havuzunu oluşturan maddelerin format ve konu alanına göre dağılımı Tablo 3.1.' de verilmiştir.

Tablo 3.1. Madde Havuzunu Oluşturan Maddelerin Konu Alanına Göre Dağılımı

<i>Konu Alanı</i>	<i>Madde Formatı</i>	
	<i>Çoktan Seçmeli</i>	<i>Açık Uçlu</i>
<i>Sayılar</i>	50	10
<i>Cebir</i>	48	7
<i>Geometri</i>	39	3
<i>Veri ve Olasılık</i>	31	6

*Bu araştırma kapsamında sadece 0,1,2 şeklinde kısmi puanlanan açık uçlu maddeler ele alınmış, 0-1 şeklinde puanlanan açık uçlu maddeler madde parametre havuzuna alınmamıştır.

Madde parametre havuzundan X ve Y testlerinde yer alan maddelerin parametreleri elde edilirken, TIMSS' in matematik testlerini hazırlarken kullandığı konu alanı örneklemedeki madde sayıları ve parametreleri dikkate alınmıştır. TIMSS 2011'de 8. sınıflarda matematik testinde yer alan maddelerin %30'u sayılar, %30'u cebir, %20'si geometri ve %20'si veri ve olasılık konu alanlarını ölçmektedir (Mullis vd., 2009). Dolayısıyla X ve Y testlerine sayılar ve cebir konu alanlarını ölçen 6'şar maddenin, geometri ile veri ve şans konu alanlarını ölçen 4'er maddenin parametreleri seçilmiştir. Ortak testte yer alan maddelerin parametreleri seçilirken aynı konu alanı oranları kullanılmıştır. X ve Y testleri ile ortak testlere ait maddelerin format ve konu alanına göre dağılımları Tablo 3.2.' de özetlenmiştir.

Tablo 3.2. Testlerde Yer Alan Maddelerin Madde Formatı ve Ölçtüğü Konu Alanına Göre Dağılımı

<i>Konu Alanı</i>	<i>X testi</i>	<i>Ortak test-1</i>	<i>Ortak test-2</i>	<i>Y testi</i>
<i>Sayılar (%30)</i>	6 (5ÇS+1AU)	3 (2ÇS+1AU)	3 (ÇS)	6 (5ÇS+1AU)
<i>Cebir (%30)</i>	6 (5ÇS+1AU)	3 (2ÇS+1AU)	3 (ÇS)	6 (5ÇS+1AU)
<i>Geometri (%20)</i>	4 (3ÇS+1AU)	2 (ÇS)	2 (ÇS)	4 (3ÇS+1AU)
<i>Veri ve Olasılık (%20)</i>	4 (3ÇS+1AU)	2 (ÇS)	2 (ÇS)	4 (3ÇS+1AU)
<i>Toplam</i>	20	10	10	20

ÇS: Çoktan Seçmeli, AU: Açık Uçlu

Madde parametre havuzundan ilk olarak eski forma (Y testine), 2011 TIMSS'in yayınlanan 5. kitapçığındaki matematik testini oluşturan maddelerin ortalama parametre değerleri ($\bar{a} = 1.19$, $\bar{b} = 0.51$ ve $\bar{c} = 0.22$) temel alınarak 16 çoktan seçmeli ve 4 açık uçlu olmak üzere toplam 20 madde seçilmiştir. Y formunu oluşturan maddelerin ayırt edicilik, güçlük ve şans parametre ortalamaları sırayla 1.19, 0.50 ve 0.19'dur.

İkinci olarak, madde parametre havuzundan ortak test-1 ve ortak test-2 için madde parametreleri seçilmiştir. NEAT deseninde ortak testin, grup farklılıklarını yeterince yansıtabilmesi için hem kapsam hem de istatistiksel özellikler bakımından testin bütününe yansıtması gerekir (Kolen ve Brennan, 2004). İstatistiksel temsil edebilirlikle kastedilen eşitlenecek forma ait (unique) maddelerin ortalama güçlüğü'nün ortak madde setinin ortalama güçlüğü ile aynı olmasıdır. Ortak testin ortalama güçlüğü belirlenirken eski formun (Y testinin) ortalama güçlüğü dikkate alınmıştır (Cao, 2008; Sinharay ve Holland, 2007). Dolayısıyla madde formatı, kapsam ve istatistiksel özellikler bakımından Y testini temsil edecek şekilde 8 çoktan seçmeli ve 2 açık uçlu madde ortak test-1'e seçilmiştir. Ortak test-1'i oluşturan maddelerin ayırt edicilik, güçlük ve şans parametre ortalamaları sırayla 1.19, 0.50 ve 0.20'dir.

Ortak test-2'nin ortak test-1'den farkı ise tamamen çoktan seçmeli maddelerden oluşmasıdır. Ortak test-1'deki iki açık uçlu madde, her iki testte de ortak olarak belirlenen ve aynı kapsamı ölçen iki çoktan seçmeli maddenin parametreleri ile değiştirilerek ortak test-2 oluşturulmuştur. Ortak test-2, çoktan seçmeli 10 maddeden oluşmaktadır. Bu araştırmada, iç ortak test kullanıldığı ve ortak test puanları X ve Y test puanlarına dahil edildiği için her iki koşulda (ortak-1 ve ortak-2 testinin kullanıldığı) ortak testlerden elde edilecek puanlar değişse bile toplam test puanının korunması önemlidir (Andrews, kişisel iletişim). Bu nedenle açık uçlu 2 madde her iki testte ortak olan fakat ortak madde setinde yer almayan 2 çoktan seçmeli madde ile değiştirilmiştir. Bu sayede, her iki koşulda testlerden elde edilebilecek toplam puanın eşit olması ($24*1+6*2=36$) sağlanmıştır. Ortak test-2'yi oluşturan maddelerin ayırt edicilik, güçlük ve şans parametre ortalamaları sırayla 1.20, 0.50 ve 0.20'dir.

Son olarak, yeni forma (X testine) madde parametre havuzundan madde parametreleri seçilmiştir. Test uygulamalarında genellikle yeni formun güçlük düzeyinin olabildiğince eski form ile aynı olmasına çalışılmaktadır. Ancak, öngörülmeven sebeplerden dolayı yeni formun güçlük düzeyi eski formdan daha kolay ya da daha zor olabilmektedir (Sinharay ve Holland, 2007). Test eşitlemenin amacı, formlar arasında oluşan bu güçlük farkını düzenlemek olduğu için bu araştırma kapsamında, X testine ortalama madde güçlük değeri Y testinden

yüksek olacak şekilde 16 çoktan seçmeli ve 4 açık uçlu madde seçilmiştir. X testinin ayırt edicilik ve şans parametre değeri Y testi ile aynı, sadece güçlük parametre değeri Y testinden 0.44 birim fazladır ($\bar{a}_X = 1.19, \bar{b}_X = 0.94, \bar{c}_X = 0.19$), bir diğer deyişle X testi Y testinden daha zor olacak şekilde oluşturulmuştur.

Form X, Form Y ile ortak testlere seçilen maddelerin ortalama parametre değerleri Tablo 3.3.' te özetlenmiştir. Bu formları oluştururken kullanılan madde parametre değerleri Ek-1'de sunulmuştur.

Tablo 3.3. Form X, Form Y ve Ortak Testlere Seçilen Maddelerin Ortalama Parametre Değerleri

<i>Madde Parametre Ortalamaları</i>	<i>Y testi</i>	<i>Ortak test-1</i>	<i>Ortak test-2</i>	<i>X testi</i>
<i>a</i>	1.19	1.19	1.20	1.19
<i>b</i>	0.50	0.50	0.50	0.94
<i>c</i>	0.19	0.20	0.20	0.19

3.4. Simülasyon Faktörleri ve Koşulları

Literatürde yer alan karma testlerin eşitlenmesi üzerine yapılan araştırmalar (Andrews, 2011; Bastari, 2000; Cao, 2008; Hagge, 2010; He, 2011; Kim ve Lee, 2004; Kirkpatrick, 2005; Tian, 2011) incelenerek bu araştırma kapsamında karma testlerin eşitlenmesini etkilediği düşünülen boyutluluk, ortak madde seti formatı, ölçek dönüştürme yöntemleri ve grupların yetenek dağılımları olmak üzere dört faktörün ele alınmasına karar verilmiştir.

Bu araştırmada ele alınan faktörlerden ilki *testlerin boyutluluğudur*. Karma testlerle çalışıldığında genellikle akıllarda oluşan “Çoktan seçmeli ve açık uçlu maddeler aynı yapıyı ölçüyor mu?” sorusu kritik bir sorudur (Kim ve Kolen, 2006, s.358). Tıpkı çoktan seçmeli testlerde olduğu gibi karma testlerde de çok boyutluluğun oluşmasının birçok sebebi olabilir. Madde formatına bağlı olarak oluşan çok boyutluluğun sebebini Traub (1993), çoktan seçmeli ve açık uçlu maddelerin bireylerin farklı bilişsel süreçlerini (yeteneklerini) ölçmesi olarak açıklamıştır. Bu araştırma kapsamında, çok boyutluluğun madde formatına dayalı olarak oluştuğu kabul edilmiştir. Çok boyutlu testleri, tek boyutlu MTK modelleri kullanarak eşitlemek uygun olmayabilir. Ancak gerçek test durumlarında uygulama kolaylığı açısından tek boyutlu modellerin kullanılarak testlerin eşitlenmesinin daha yaygın olduğu görülmektedir. Dolayısıyla tek boyutlu MTK modellerine dayalı olan test

eşitleme yöntemlerinin, madde formatına dayalı olarak oluşan çok boyutlu testlerin eşitlenmesine ne kadar dayanıklı olduğunu araştırmak önemlidir (Cao, 2008). Dolayısıyla bu araştırma kapsamında madde formatına dayalı oluşan çok boyutluluk bir faktör olarak alınmıştır.

Bu araştırma kapsamında boyutluluk ile kastedilen iki boyutlu test yapısıdır. Çoktan seçmeli maddelerin θ_1 , açık uçlu maddelerin θ_2 yeteneğini ölçtüğü varsayımı altında θ_1 ve θ_2 yetenekleri arasındaki korelasyon 1.00, 0.80, 0.61 ve 0.50 olarak belirlenerek boyutluluk koşulu altında dört faktör oluşturulmuştur. Bu korelasyonlardan 1.00, 0.80 ve 0.50 literatürde yapılan araştırma sonuçlarına (Kim ve Kolen, 2006; Andrews, 2011) dayanılarak belirlenmiştir. 0.61 korelasyon değeri ise TIMSS 2011 8. sınıflarda uygulanan 5. kitapçıktaki matematik testinden elde edilmiştir. TIMSS 2011'in Türkiye örnekleminde yer alan 490 öğrencinin matematik testine verdiği cevaplar üzerinden polikorik korelasyon matrisi kullanılarak temel bileşenler analizi yapılmıştır. Yapılan temel bileşenler analizi sonucu, maddelerin iki bileşen altında toplandığı ve bileşenlerin aralarındaki korelasyonun 0.61 olduğu görülmüştür. Tek boyutlu test durumunu θ_1 ve θ_2 yetenekleri arasındaki korelasyonun 1.00 olduğu durum temsil etmektedir. Korelasyon değeri azaldıkça veri setinde oluşan çok boyutluluğun düzeyi artmaktadır.

Bu araştırma kapsamında ele alınan ikinci faktör *ortak madde seti formatıdır*. NEAT deseni kullanıldığında bir diğer soru ortak madde setinin nasıl oluşturulması gerektiğidir. İdeal olan ortak madde setinin testin bütününe temsil edecek şekilde çoktan seçmeli ve açık uçlu maddeler birlikte kullanılarak oluşturulmasıdır. Bu sayede ortak madde seti kapsam ve istatistiksel özellikler bakımından testin bütününe daha iyi temsil edecektir. Fakat uygulamada karşılaşılan bir takım zorluklardan (güvenirlilik, güvenlik, puanlayıcı tutarsızlığı vb.) dolayı ortak madde seti genellikle sadece çoktan seçmeli maddelerden oluşturulmaktadır (Kim ve Lee, 2004). Bu araştırma kapsamında ortak madde seti formatı bir faktör olarak alınarak, test eşitlemenin eşitlik özelliğinin korunumunu ortak madde setine açık uçlu maddelerin dâhil edilip edilmemesinin etkisi incelenmiştir. Bu amaç doğrultusunda ortak madde seti için iki farklı koşul oluşturulmuştur: ortak test -1 eşitlenecek testleri temsil edecek şekilde karma formatta oluşturulurken ortak test-2 ise sadece çoktan seçmeli maddelerle oluşturulmuştur.

Bu araştırma kapsamında ele alınan üçüncü faktör farklı *ölçek dönüştürme yöntemleridir*. Yapılan araştırmalar, karakteristik eğri ölçek dönüştürme yöntemlerinin moment yöntemleri ile karşılaştırıldığında daha kararlı sonuçlar verdiğini göstermiştir (Hanson ve Beguin, 2002; Kim ve Cohen, 1992, Kim ve Lee, 2004). Literatür incelendiğinde BSE ve İSE özelliğini korunumunun ölçüt olarak alındığı test durumlarında, ölçek dönüştürme yöntemlerinin performansının değerlendirildiği herhangi bir çalışmaya rastlanmamıştır. Dolayısıyla bu araştırma kapsamında ortalama/ortalama (O-O), ortalama/sigma (O-S), Haebara ve Stocking-Lord (S-L) olmak üzere dört farklı ölçek dönüştürme yöntemi bir faktör olarak ele alınmıştır.

Bu araştırma kapsamında ele alınan son faktör *grupların yetenek dağılımlarıdır*. Bu faktör altında eşdeğer olan gruplar ve eşdeğer olmayan gruplar olmak üzere iki koşul ele alınmıştır. NEAT deseninde her ne kadar grupların yetenek dağılımlarının eşit olması beklenmese de grupların eşdeğer olmama derecesini karşılaştırma açısından grupların eşdeğer olduğu durum bir koşul olarak alınmıştır. Bu bağlamda testlerin tek boyutlu ve çok boyutlu olduğu durumda yetenek dağılımı için iki koşul oluşturulmuştur. Grupların eşdeğer olduğu tek boyutlu koşul altında, her iki grup ortalaması 0 ve standart sapması 1 olan standart normal dağılıma sahiptir. Çok boyutlu koşul altında ise her iki grup her iki yetenekte ortalaması 0 ve standart sapması 1 olan çok değişkenli normal dağılıma sahiptir. Matematiksel olarak şöyle ifade edilebilir:

$$(\theta_1, \theta_2)_Y \sim BN(\mu_1 = 0, \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 1)$$

$$(\theta_1, \theta_2)_X \sim BN(\mu_1 = 0, \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 1)$$

Grupların eşdeğer olmadığı tek boyutlu koşul altında, Grup 1'in yetenek ortalaması 0 iken Grup 2'nin yetenek dağılımının ortalaması 1 birim arttırılmıştır (Hanson ve Beguin, 2002; Kim ve Lee, 2006). Grup 1 ve Grup 2'nin standart sapması aynı ve 1'dir. Grupların eşdeğer olmadığı çok boyutlu koşul altında, Grup 1'in (Y formunu alan grup) her iki yetenekteki ortalaması sıfır ve standart sapması 1'dir. Grup 2'nin (X formunu alan grup) ise her iki yetenekteki ortalaması 1 birim Grup 1'den daha yüksek olup, standart sapması Grup 1 ile aynı ve 1'dir. Bu durum matematiksel olarak şöyle ifade edilebilir:

$$(\theta_1, \theta_2)_Y \sim BN(\mu_1 = 0, \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 1)$$

$$(\theta_1, \theta_2)_X \sim BN(\mu_1 = 1, \mu_2 = 1, \sigma_1 = 1, \sigma_2 = 1)$$

Bu araştırma kapsamında dört faktör altında incelenen koşullar Tablo 3.4.'te özetlenmiştir.

Tablo 3.4. Araştırmada İncelenen Faktörler ve Koşullar

<i>Faktörler</i>	<i>Düzeyleri</i>
Boyutluluk	1.00, 0.80, 0.61, 0.50
Ortak Madde Seti Formatı	Karma, sadece çoktan seçmeli
Grupların Yetenek Dağılımları	Eşdeğer gruplar, eşdeğer olmayan gruplar
Ölçek Dönüştürme Yöntemi	Ortalama/ortalama, ortalama/sigma, Haebara, Stocking-Lord

3.5. Verilerin Türetilmesi

Araştırmada 64 [4 (boyutluluk düzeyi) x 2 (ortak madde seti formatı) x 2 (grupların yetenek dağılımı) x 4 (ölçek dönüştürme yöntemi)] farklı koşul incelenmiştir. Ortak madde seti formatı ve ölçek dönüştürme yöntemi faktörleri araştırma kapsamında incelenen fakat veri türetiminde bir değişken olarak yer almayan faktörlerdir. Dolayısıyla araştırma, her bir form ve 8 [4 (boyutluluk düzeyi) x 2 (grupların yetenek dağılımı)] koşul için 800 (8x100) farklı veri seti türetilerek yürütülmüştür. Araştırma kapsamında veri türetiminde kullanılan sekiz koşul Tablo 3.5.'te özetlenmiştir.

Tablo 3.5. Veri Türetiminde Kullanılan Koşullar

<i>Koşul</i>	<i>Boyutlar Arası Korelasyon</i>	<i>Yetenek Dağılımı</i>
1	1.00	$(\theta_1)_Y \sim N(\mu_1 = 0, \sigma_1 = 1)$ $(\theta_1)_X \sim N(\mu_1 = 0, \sigma_1 = 1)$
2	0.80	$(\theta_1, \theta_2)_Y \sim BN(\mu_1 = 0, \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 1)$
3	0.61	$(\theta_1, \theta_2)_X \sim BN(\mu_1 = 0, \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 1)$
4	0.50	
5	1.00	$(\theta_1)_Y \sim N(\mu_1 = 0, \sigma_1 = 1)$ $(\theta_1)_X \sim N(\mu_1 = 1, \sigma_1 = 1)$
6	0.80	$(\theta_1, \theta_2)_Y \sim BN(\mu_1 = 0, \mu_2 = 0, \sigma_1 = 1, \sigma_2 = 1)$
7	0.61	$(\theta_1, \theta_2)_X \sim BN(\mu_1 = 1, \mu_2 = 1, \sigma_1 = 1, \sigma_2 = 1)$
8	0.50	

Y formunu alan Grup 1 ve X formunu alan Grup 2 için madde cevapları Ek-1'de verilen madde parametreleri dikkate alınarak türetilmiştir. Her bir koşul altında her bir form için 3000 madde cevabı türetilmiştir. Yapılan araştırmalar (Hanson ve

Beguin, 2002; Kim ve Lee, 2004) 3000 örneklem büyüklüğünün doğru eşitleme sonuçları elde etmek için yeterli olduğunu göstermiştir.

Veri üretiminde SimuMIRT (Yao, 2003) bilgisayar programında yer alan “SimulateRWO.bat” dosyası kullanılmıştır. “SimulateRWO” hem iki hem çok kategorili puanlanan maddeler için çok boyutlu telafi edici MTK modellerine dayalı olarak madde cevabı türetebilen bir programdır. Çok boyutlu güçlük parametresini temsil eden “d” parametresi ile tek boyutlu güçlük parametresini temsil eden “b” parametresi arasında tutarlılığı sağlamak amacıyla güçlük parametresine Reckase (2009) tarafından önerilen $d = (-)b * \sqrt{\sum_{v=1}^m a_v^2}$ dönüşümü uygulanmıştır.

Tek boyutlu simülasyon koşulları altında, çoktan seçmeli maddeler için madde cevapları 3PL model, açık uçlu maddeler için ise GPC modeli kullanılarak türetilmiştir. Veri üretiminde, bu modellerin seçiminin sebebi TIMSS uygulamalarındaki başarı testlerinin, çoktan seçmeli maddelerin 3PL model; kısmi puanlanan açık uçlu maddelerin GPC (Muraki, 1992) modelinin kullanılarak kalibre edilmesidir (Mullis vd., 2009). Çok boyutlu koşullar altında ise bu modellerin uzantısı olarak geliştirilen çok boyutlu 3PL model (M-3PL, Reckase 1997) ve çok boyutlu genelleştirilmiş iki-parametrelili kısmi puan modeli (M-2PPC; Yao ve Shwarz, 2006) kullanılmıştır.

Araştırmada boyutluluk, çoktan seçmeli ve açık uçlu maddelerin farklı iki yeteneği ölçtüğü kabul edilerek oluşturulmuştur. Çoktan seçmeli maddelerin sadece θ_1 yeteneğini ölçtüğü varsayılar, bu maddeler için birinci boyutu ölçen ayırt edicilik parametresi (a_1) değer alırken, ikinci boyutu ölçen ayırt edicilik parametresi (a_2) sıfır değerini almıştır. Açık uçlu maddelerin ise sadece θ_2 yeteneğini ölçtüğü varsayılmış ve bu maddeler için ikinci boyutu ölçen a_2 parametresi değer alırken, birinci boyutu ölçen a_1 parametresi sıfır değerini almıştır. Çok boyutluluğun derecesi programın girdi dosyasında yer alan varyans-kovaryans matrisi aracılığı ile ayarlanmıştır. Örneğin, iki yetenek arasındaki korelasyonun 0.80 olduğu iki boyutlu koşul altında oluşturulan varyans-kovaryans matrisi şöyle ifade edilebilir:

$$\begin{bmatrix} 1.00 & 0.80 \\ 0.80 & 1.00 \end{bmatrix}$$

Tek boyutlu koşul altında ise varyans-kovaryans matrisi yerine sadece tek bir rakam “1” yazılmıştır (Yao, kişisel iletişim). X ve Y testlerine ortak maddeler dahil edilerek her bir test formu için madde cevapları ayrı ayrı türetilmiştir.

3.6. Simülasyon Verilerinin Kontrolü

Bu araştırmada, sekiz farklı koşul altında türetilen verinin kontrolü, her bir forma ait kestirilen madde parametreleri için RMSE (Root Mean Squared Error; Hata kareleri ortalamasının karekökü) ve bias (yanlılık) değerleri hesaplanarak yapılmıştır. RMSE, parametre kestiriminde mutlak doğruluğun (absolute accuracy) bir ölçüsüdür. Kestirilen parametre değeri ile gerçek parametre değeri arasındaki farkın karesinin ortalamasıdır. Bias gerçek ve kestirilen parametre değerleri arasındaki farkın ortalaması olarak ifade edilebilir. Bias parametre kestirimindeki sistematik hataların bir ölçüsünü verir. Parametre kestiriminin varyansı büyüdükçe ya da bias arttıkça RMSE değeri de artabilir. Matematiksel olarak RMSE ve bias eşitlik 3.1 ve 3.2 de olduğu gibi ifade edilebilir (Demars, 2003; Jurich ve Goodman, 2009).

$$RMSE = \sqrt{\frac{\sum_{k=1}^{100} (\hat{\tau}_{jk} - \tau_j)^2}{100}} \quad (3.1)$$

$$Bias = \frac{\sum_{k=1}^{100} \hat{\tau}_{jk} - \tau_j}{100} \quad (3.2)$$

τ_j : j parametresinin gerçek değeri

$\hat{\tau}_{jk}$: tekrar eden veri seti (k=1...100) için j parametresinin kestirilen değeri

Tablo 3.6.'da her bir formun kestirilen madde parametrelerine dair hesaplanan RMSE ve bias ortalamaları verilmiştir. Bias ortalamaları, bias değerlerinin mutlak değeri alınarak hesaplanmıştır. Bunun sebebi pozitif ve negatif bias değerlerinin birbirini nötrlemesini önlemek, aradaki gerçek farkı görebilmektir (Stone, 1992). Koşul-1,2,3 ve 4'te gruplar eşdeğer, her iki yetenek arasındaki korelasyon sırayla 1.00, 0.80, 0.61 ve 0.50'dir. Koşul-5, 6, 7 ve 8'de ise X ve Y formunu alan gruplar eşdeğer olmamakla birlikte benzer şekilde iki yetenek arasındaki korelasyon sırayla 1.00, 0.80, 0.61 ve 0.50'dir (Bkz. Tablo 3.5).

Tablo 3.6.'da yer alan X ve Y formlarını alan gruplar için tek boyutlu veri setinin türetilip tek boyutlu madde kalibrasyonunun yapıldığı Koşul-1 ve Koşul-5 simülasyon sürecinin kontrolü için ölçüt olarak alınabilir. Koşul-1 ve Koşul-5'te her iki formun

kestirilen madde parametrelerine ilişkin RMSE değerleri incelendiğinde en küçük RMSE değerine güçlük parametresinin (*b*) sahip olduğu ve onu ayırt edicilik parametresinin (*a*) takip ettiği görülmektedir.

Tablo 3.6. Form X ve Form Y'nin Madde Parametreleri için Hesaplanan RMSE ve Bias Değeri Ortalamaları

Koşul	Form X						Form Y					
	RMSE			Bias			RMSE			Bias		
	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>c</i>
1	0.13	0.10	0.40	0.03	0.03	0.01	0.12	0.11	0.41	0.04	0.04	0.02
2	0.18	0.13	0.43	0.10	0.09	0.03	0.16	0.15	0.43	0.08	0.10	0.03
3	0.25	0.18	0.45	0.18	0.15	0.05	0.21	0.19	0.45	0.13	0.15	0.04
4	0.30	0.21	0.46	0.24	0.18	0.06	0.24	0.21	0.45	0.17	0.17	0.05
5	0.11	0.13	0.42	0.04	0.06	0.02	0.12	0.11	0.41	0.04	0.04	0.02
6	0.18	0.19	0.46	0.13	0.14	0.05	0.16	0.15	0.43	0.08	0.10	0.03
7	0.44	0.34	0.51	0.39	0.32	0.11	0.22	0.20	0.45	0.14	0.16	0.05
8	0.89	0.47	0.55	0.83	0.45	0.15	0.25	0.22	0.45	0.18	0.18	0.05

En yüksek RMSE değerine ise şans parametresi (*c*) sahiptir. Literatürde yapılan araştırmalar da parametre doğrulanmasının (recovery) en iyi güçlük parametre kestiriminde olduğunu göstermiştir (Jurich ve Goodman, 2003; Paek ve Cai, 2013; Stone, 1992).

Tablo 3.6.'da eşdeğer ve eşdeğer olmayan gruplar koşulu altında çok boyutluluğun derecesi arttıkça X ve Y formlarına ait madde parametrelerinin RMSE ve Bias değerlerinin de arttığı görülmektedir. Bu durumun oluşması beklendiktir. Çünkü madde cevapları çok boyutlu MTK modelleri kullanılarak türetilmiş, madde parametreleri ise tek boyutlu MTK modelleri kullanılarak kestirilmiştir. Tek boyutlu parametre kestiriminin çok boyutluluktan olumsuz yönde etkilendiğini söyleyebiliriz.

3.7. Verilerin Analizi

Veri türetimi tamamlanıp kontrolü yapıldıktan sonra verilerin analizine geçilmiştir. Verilerin analizi madde kalibrasyonu, ölçek dönüştürme, test eşitleme ve eşitleme sonuçlarının değerlendirilmesi olmak üzere dört aşamada tamamlanmıştır.

1. Madde Kalibrasyonu: MTK eşitleme yöntemleri kullanıldığında, madde parametreleri ayrı ya da eş zamanlı kalibrasyon yöntemlerinden biri kullanılarak kestirilir. Yapılan araştırmalar (Beguin, Hanson ve Glass, 2000; Kim ve Kolen, 2006), MTK'nın tek boyutluluk varsayımı sağlandığında eş zamanlı kalibrasyonun ayrı kalibrasyondan daha iyi performans gösterdiğini

fakat çok boyutluluğun olduğu durumlarda ise ayrı kalibrasyonun MTK varsayımlarının ihlaline daha dayanıklı olduğunu ve daha doğru sonuçlar verdiğini göstermiştir. Bu araştırmada boyutluluk bir koşul olarak ele alındığından tek boyutluluk varsayımının ihlaline daha dayanıklı olan ayrı kalibrasyon yöntemi seçilerek X ve Y formlarına ait madde parametreleri kestirilmiştir. Çoktan seçmeli maddelerin parametre kestiriminde 3PL model, açık uçlu maddelerin parametre kestiriminde GPC modeli kullanılmıştır. Madde parametreleri Marginal Maksimum Likelihood (MML) yöntemi ile PARSCALE 4.1 (Muraki ve Bock, 2003) programı kullanılarak kestirilmiştir. PARSCALE programında madde parametrelerini lojistik ölçekte elde edebilmek için ölçek sabiti "D=1" olarak tanımlanmıştır (Childs ve Chen, 1999). Araştırma kapsamında kullanılan PARSCALE kodlarının bir örneği Ek-2'de sunulmuştur.

2. Ölçek dönüştürme: X formunun madde parametreleri ve yetenek dağılımı (quadrature noktaları) ortalama/ortalama, ortalama/sigma, Haebara ve Stocking-Lord olmak üzere dört farklı ölçek dönüştürme yöntemi kullanılarak Y formunun ölçeğine dönüştürülmüştür. Ölçek dönüştürme işlemi STUIRT (Kim ve Kolen, 2004) bilgisayar programı kullanılarak yapılmıştır.
3. Testlerin eşitlenmesi: Veri analizinin üçüncü aşamasında MTK gerçek puan eşitleme ve MTK gözlenen puan eşitleme yöntemleri kullanılarak X testi Y testine eşitlenmiştir. Testler birbirine karma formattaki testlerin eşitlenmesi için geliştirilen POLYEQUATE (Kolen, 2004) bilgisayar programı kullanılarak eşitlenmiştir. MTK kalibrasyonu sonucu kestirilen Y formunun madde parametreleri ve ölçek dönüştürme sonucu elde edilen X formunun madde parametreleri programın girdi dosyasını oluşturmuştur.
4. Eşitleme sonuçlarının değerlendirilmesi: Araştırmada incelenen 64 farklı koşul altında MTK gerçek puan ve MTK gözlenen puan eşitleme yöntemleri kullanılarak elde edilen eşitleme sonuçları birinci-sıra ve ikinci-sıra eşitlik özelliğini koruma ölçütlerine göre değerlendirilmiştir. Birinci-sıra (BSE) ve ikinci-sıra eşitlik (İSE) özelliklerini değerlendirebilmek için belirli bir test teori modelinin kabul edilmesi gerekir. Bu araştırmada, çoktan seçmeli maddeler için 3PL model, açık uçlu maddeler için GPC modeli dikkate alınmıştır.

BSE özelliği Tong ve Kolen (2005) tarafından geliştirilen D_1 indeksi kullanılarak hesaplanmıştır:

$$D_1 = \frac{\sqrt{\sum_i q_i \{E[Y | \theta_i] - E[e\hat{q}_Y(x) | \theta_i]\}^2}}{SD_Y} \quad (3.3)$$

D_1 eşitliğindeki; $E[Y | \theta_i]$ bir θ_i yetenek düzeyi için eski formun koşullu ortalamasıdır. $E[e\hat{q}_Y(x) | \theta_i]$ ise bir θ_i düzeyi için eşitlenmiş puanların koşullu ortalaması; $q_i\theta_i$ 'deki quadrature ağırlığıdır. Son olarak, SD_Y Y formunun standart sapmasıdır. D_1 değeri küçüldükçe BSE özelliği daha iyi korunmaktadır.

İSE özelliğini değerlendirmek için yine Tong ve Kolen (2005) tarafından geliştirilen D_2 indeksi hesaplanmıştır:

$$D_2 = \frac{\sqrt{\sum_i q_i (SEM_Y | \theta_i - SEM_{\hat{q}_Y(x)} | \theta_i)^2}}{SD_Y} \quad (3.4)$$

D_2 indeksinde yer alan $SEM_Y | \theta_i$ eski formu alan θ_i yetenek düzeyindeki bireylere ait ölçmenin koşullu standart hatası (SEM); $SEM_{\hat{q}_Y(x)} | \theta_i$ ise eşitlenen yeni formu alan θ_i yetenek düzeyindeki bireylere ait ölçmenin koşullu standart hatasıdır. D_2 indeksi değerinin artması İSE özelliğinin korunumu azaldığının göstergesidir.

Veri türetmede kullanılan evren parametre değerleri bilindiğinden dolayı BSE ve İSE özelliği bilinen evren parametre değerleri kullanılarak değerlendirilmiştir. D_1 ve D_2 indekslerinde yer alan eski formun koşullu ortalaması ve ölçmenin koşullu standart hatası X ve Y formlarının bilinen evren parametre değerleri kullanılarak hesaplanmıştır. Her bir tekrar için D_1 ve D_2 indekslerinde yer alan eski forma ait koşullu ortalama ve standart hatalar aynı kalmış, sadece eşitlenmiş puanlara ait koşullu ortalama ve standart hatalar değişmiştir (Bolt, 1999; Andrews, 2011).

D_1 ve D_2 indekslerini hesaplamak için POLYCSEM (Kolen, 2004) bilgisayar programında koşullu ortalama ve koşullu SEM'ler hesaplanmıştır. Son olarak, elde edilen koşullu ortalama ve koşullu SEM'ler kullanılarak D_1 ve D_2 indeksleri hesaplanmıştır.

Verilerin analizi için yukarıda açıklanan 1'den 4'e kadar olan basamakların hepsi 64 koşul için 100 kez tekrarlanmıştır. Literatürde yapılan araştırmalar, kararlı eşitleme sonuçları elde etmek için 100 tekrarın yeterli olduğunu gösterdiğinden dolayı tekrar sayısı 100 seçilmiştir (Andrews, 2011; Cao, 2008; Kim ve Lee, 2004; Kim ve Kolen, 2006). 100 tekrarın sonunda elde edilen D_1 ve D_2 indekslerinin ortalaması alınarak eşitleme sonuçları değerlendirilmiştir. Ayrıca incelenen faktörlerin tek başına ve etkileşimlerinin eşitlik özelliğinin korunumu üzerinde istatistiksel olarak anlamlı etkisinin olup olmadığı 4-yönlü varyans analizi (ANOVA) yapılarak değerlendirilmiştir. Araştırmada kullanılan PARSCALE, STUIRT, POLYEQUATE ve POLYCSEM bilgisayar programları R (R Development Core Team, 2005) tabanlı çalıştırılarak tüm analizler seri şekilde yapılmıştır. Ayrıca, tüm programların girdi dosyalarının hazırlanmasında, çıktı dosyalarının düzenlenmesinde ve D_1 ile D_2 indekslerinin hesaplanmasında R programlama dili kullanılmıştır.

4. BULGULAR VE TARTIŞMA

Bu bölümde araştırmanın bulguları, araştırma problemleri çerçevesinde ele alınan birinci-sıra (BSE) ve ikinci-sıra eşitlik (İSE) özelliği korunumu ölçütlerine göre incelenmiştir. Gerçek puan eşitleme (GR_PE) ve gözlenen puan eşitleme (GZ_PE) yöntemine göre elde edilen eşitleme sonuçlarını BSE özelliği korunumu ölçütüne göre değerlendirmek için D_1 indeksi, İSE özelliği korunumu ölçütüne göre değerlendirmek için D_2 indeksi hesaplanmıştır. Araştırma kapsamında ele alınan faktörler, 64 farklı koşul için 100 tekrar sonucu hesaplanan D_1 ve D_2 indeksi ortalamalarına (\bar{D}_1 ve \bar{D}_2) göre incelenmiştir. Göreceli olarak düşük \bar{D}_1 ve \bar{D}_2 değerleri eşitlik özelliğinin daha iyi korunduğunun bir göstergesi olarak ele alınmıştır. Faktörlerin tek başına ve etkileşimlerinin eşitlik özelliğinin korunumu üzerinde istatistiksel olarak anlamlı etkisinin olup olmadığı 4-yönlü ANOVA yapılarak değerlendirilmiştir. ANOVA testi sonucu elde edilen etki büyüklüklerini değerlendirmede Cohen (1988) tarafından önerilen ölçüt (*kısmi* η^2 : 0.01 küçük etki; 0.06 orta etki; 0.14 büyük etki) baz alınmıştır. Elde edilen bulgular, alt problemlerin sırasına göre sunulmuştur.

4.1. Alt Problem 1'e İlişkin Bulgular

Madde Tepki Kuramı gerçek puan eşitleme yöntemi ile yapılan eşitlemelerde, eşitlik özelliğinin korunumunu;

- a) Testlerin boyutluluğu,
- b) Ortak madde seti formatı,
- c) Grupların yetenek dağılımları,
- d) Ölçek dönüştürme yöntemleri ve
- e) Bu faktörlerin etkileşimleri nasıl etkilemektedir?

GR_PE eşitleme yöntemi sonucu elde edilen eşitleme sonuçlarının birinci-sıra eşitlik (BSE) ve ikinci-sıra eşitlik (İSE) özelliklerine dayalı olarak değerlendirmek üzere hesaplanan D_1 ve D_2 indeksi ortalamaları (\bar{D}_1 ve \bar{D}_2) sırayla Tablo 4.1. ve Tablo 4.2.'de sunulmuştur. Boyutluluk, ortak madde seti formatı, yetenek dağılımı ve ölçek dönüştürme yöntemi faktörlerinin ve bu faktörlerin etkileşimlerinin BSE ve İSE özelliklerinin korunması üzerinde istatistiksel olarak anlamlı etkisinin olup olmadığı 4-yönlü ANOVA yapılarak test edilmiş ve elde edilen bulgular Tablo 4.3.'te sunulmuştur.

Tablo 4.1. MTK Gerçek Puan Eşitleme Yöntemine Göre Elde Edilen D_1 İndeksi Ortalamaları

Yetenek Dağılımı	Ortak Madde Seti	Boyutluluk	Ölçek Dönüştürme Yöntemleri			
			O-O	O-S	SL	HA
Eşdeğer Gruplar (EG)	KOM	1.00	0.068	0.072	0.067	0.068
		0.80	0.099	0.093	0.078	0.084
		0.61	0.135	0.121	0.092	0.105
		0.50	0.163	0.143	0.104	0.121
	ÇSOM	1.00	0.073	0.119	0.067	0.067
		0.80	0.111	0.143	0.081	0.081
		0.61	0.158	0.165	0.101	0.102
		0.50	0.193	0.179	0.115	0.116
Eşdeğer Olmayan Gruplar (EOG)	KOM	1.00	0.115	0.103	0.076	0.087
		0.80	0.177	0.142	0.105	0.125
		0.61	0.336	0.248	0.183	0.227
		0.50	0.462	0.335	0.250	0.307
	ÇSOM	1.00	0.125	0.128	0.076	0.082
		0.80	0.227	0.213	0.134	0.141
		0.61	0.451	0.419	0.253	0.260
		0.50	0.617	0.578	0.338	0.347

*KOM: Karma ortak madde, ÇSOM: Çoktan seçmeli ortak madde, O-O: Ortalama-Ortalama, O-S: Ortalama-Sigma, SL: Stocking-Lord, HA: Haebara

Tablo 4.2. MTK Gerçek Puan Eşitleme Yöntemine Göre Elde Edilen D_2 İndeksi Ortalamaları

Yetenek Dağılımı	Ortak Madde Seti	Boyutluluk	Ölçek Dönüştürme Yöntemleri			
			O-O	O-S	SL	HA
Eşdeğer Gruplar (EG)	KOM	1.00	0.038	0.038	0.037	0.037
		0.80	0.037	0.036	0.038	0.038
		0.61	0.038	0.035	0.040	0.038
		0.50	0.042	0.036	0.044	0.041
	ÇSOM	1.00	0.039	0.056	0.037	0.037
		0.80	0.035	0.055	0.038	0.038
		0.61	0.033	0.048	0.037	0.037
		0.50	0.035	0.048	0.040	0.040
Eşdeğer Olmayan Gruplar (EOG)	KOM	1.00	0.038	0.038	0.044	0.042
		0.80	0.046	0.038	0.048	0.044
		0.61	0.100	0.053	0.070	0.062
		0.50	0.176	0.077	0.097	0.085
	ÇSOM	1.00	0.037	0.059	0.044	0.043
		0.80	0.040	0.053	0.041	0.040
		0.61	0.086	0.077	0.054	0.054
		0.50	0.128	0.111	0.078	0.076

*KOM: Karma ortak madde, ÇSOM: Çoktan seçmeli ortak madde, O-O: Ortalama-Ortalama, O-S: Ortalama-Sigma, SL: Stocking-Lord, HA: Haebara

Tablo 4.3. GR_PE Sonuçlarının BSE ve İSE Özelliği Ölçütlerine Göre 4-yönlü ANOVA Testi Sonuçları

<i>Etkiler</i>	<i>sd</i>	<i>D₁</i>		<i>D₂</i>	
		<i>F</i>	η^2	<i>F</i>	η^2
<i>Boyutluluk (B)</i>	3	4379.949*	0.675	1577.399*	0.428
<i>Format (F)</i>	1	1146.193*	0.153	0.000	0.000
<i>Yetenek Dağılımı (YD)</i>	1	10752.619*	0.629	4836.287*	0.433
<i>Ölçek Dönüştürme Yöntemi (ÖDY)</i>	3	1112.716*	0.345	225.274*	0.096
<i>B*F</i>	3	131.931*	0.059	34.652*	0.016
<i>B*YD</i>	3	1859.424*	0.468	1451.394*	0.407
<i>B*ÖDY</i>	9	115.187*	0.141	128.390*	0.154
<i>F*YD</i>	1	443.838*	0.065	46.008*	0.007
<i>F*ÖDY</i>	3	172.468*	0.075	333.495*	0.136
<i>YD*ÖDY</i>	3	312.274*	0.129	300.074*	0.124
<i>B*F*YD</i>	3	111.465*	0.050	9.481*	0.004
<i>B*F*ÖDY</i>	9	10.940*	0.015	18.161*	0.025
<i>B*YD*ÖDY</i>	9	36.484*	0.049	132.421*	0.158
<i>F*YD*ÖDY</i>	3	26.930*	0.013	41.728*	0.019
<i>B*F*YD*ÖDY</i>	9	13.087*	0.018	17.947*	0.025
<i>Hata</i>	6336				
<i>Toplam</i>	6399				

*p<0.0001

4.1.1. Testlerin boyutluluğu eşitlik özelliği korunumunu nasıl etkilemektedir?

Boyutluluğun MTK GR_PE sonuçlarının BSE özelliğinin korunumu üzerindeki etkisini değerlendirmek üzere Tablo 4.1.' de verilen \bar{D}_1 değerleri incelenmiştir. Tablo 4.1.'de verilen \bar{D}_1 değerlerine göre tüm koşullar altında, çok boyutluluğun derecesi arttıkça \bar{D}_1 değerlerinin de arttığı görülmektedir. En küçük \bar{D}_1 değeri, iki yetenek arasındaki korelasyonun 1.00 olduğu tek boyutlu koşul altında elde edilirken, en büyük \bar{D}_1 çok boyutluluğun derecesinin en yüksek olduğu ($\rho_{\theta_1\theta_2} = 0.50$) koşul altında elde edilmiştir. Örneğin; O-O ölçek dönüştürme yönteminin kullanıldığı, grupların eşdeğer ve ortak madde setinin karma olduğu koşul altında en düşük \bar{D}_1 değeri (0.068) tek boyutlu koşul altında elde edilirken en yüksek \bar{D}_1 değeri (0.163) çok boyutluluğun oluştuğu iki yetenek arasındaki korelasyonun 0.50 olduğu durumda elde edilmiştir.

Boyutluluğun GR_PE yöntemi ile elde edilen eşitleme sonuçlarının İSE özelliğinin korunumuna etkisini değerlendirmek üzere Tablo 4.2. verilen \bar{D}_2 değerleri incelenmiştir. Tablo 4.2.' de verilen \bar{D}_2 değerlerine göre, karakteristik eğri ölçek dönüştürme yöntemleri uygulanarak elde edilen GR_PE sonuçlarının genel olarak

(EOG-ÇSOM koşulu hariç) tek boyutlu koşullar altında en düşük, çok boyutlu koşullar altında ise en yüksek \bar{D}_2 değerine sahip olduğunu söyleyebiliriz. Moment ölçek dönüştürme yöntemleri uygulanarak elde edilen GR_PE sonuçlarının ise grupların eşdeğer olduğu durumda, kullanılan ortak madde seti türüne bağlı olmaksızın en düşük \bar{D}_2 değerine çok boyutlu koşullar altında sahip olduğunu; grupların eşdeğer olmadığı durumda ise genel olarak (O-S/CSOM koşulu hariç) tek boyutlu koşullar altında en düşük \bar{D}_2 değerine sahip olduğunu söyleyebiliriz.

Tablo 4.3.'te verilen ANOVA testi sonuçları da boyutluluğun istatistiksel olarak, BSE ve İSE özelliğinin korunması üzerinde anlamlı ve büyük etkisinin olduğunu göstermiştir ($\eta_{BSE}^2 = 0.675$, $\eta_{ISE} = 0.428$, $p < 0.0001$). Boyutluluğa dair oluşan farklılığın hangi çok boyutluluk dereceleri arasında olduğunu tespit etmek için ikili karşılaştırma testi yapılmış, elde edilen sonuçlar ve marjinal ortalamalar sırayla Tablo 4.4. ile Tablo 4.5.'te sunulmuştur.

Tablo 4.4. Boyutluluğa İlişkin İkili Karşılaştırma Testi Sonuçları

<i>Boyutluluk</i>	<i>D₁ Ortalama Fark</i>	<i>D₂ Ortalama Fark</i>
<i>1.00-0.80</i>	-0.040*	0.000
<i>1.00-0.61</i>	-0.123*	-0.013*
<i>1.00-0.50</i>	-0.186*	-0.031*
<i>0.80-0.61</i>	-0.083*	-0.012*
<i>0.80-0.50</i>	-0.146*	-0.031*
<i>0.61-0.50</i>	-0.063*	-0.018*

*p < 0.0001

Tablo 4.5. Marjinal Ortalamalar

<i>Boyutluluk</i>	<i>D₁ Ortalama</i>	<i>D₂ Ortalama</i>
<i>1.00</i>	0.087	0.041
<i>0.80</i>	0.127	0.042
<i>0.61</i>	0.210	0.054
<i>0.50</i>	0.273	0.072

BSE özelliği değerlendirme ölçütü alındığında Tablo 4.4.'te yer alan ikili karşılaştırma testi sonuçlarının tümünün istatistiksel olarak anlamlı olduğu görülmektedir. Ortalamalar arasındaki en büyük farklılık (-0.186) tek boyutlu koşul ile çok boyutluluğun derecesinin en yüksek olduğu ($\rho_{\theta_1\theta_1} = 0.50$) durumda elde edilmiştir. İSE özelliği değerlendirme ölçütü olarak alındığında ise sadece tek boyutlu koşul ile çok boyutluluğun derecesinin en az olduğu ($\rho_{\theta_1\theta_1} = 0.80$) koşulların ikili karşılaştırmaları arasında istatistiksel olarak anlamlı farklılık

bulunmamaktadır, diğer tüm karşılaştırma sonuçları istatistiksel olarak anlamlıdır. Tablo 4.5.'te verilen marjinal ortalamalara dayalı olarak genel olarak boyutluluk arttıkça \bar{D}_1 ve \bar{D}_2 değerlerinin de arttığını, boyutluluğun oluşmasından BSE ve İSE özelliklerinin korunumunun olumsuz yönde etkilendiğini söyleyebiliriz.

4.1.2. Ortak madde seti formatı eşitlik özelliğinin korunumunu nasıl etkilemektedir?

Tablo 4.1.' de yer alan \bar{D}_1 değerleri incelendiğinde tek boyutlu koşul altında gruplar eşdeğer olsun ya da olmasın moment ölçek dönüştürme yöntemleri uygulandığında elde edilen GR_PE sonuçlarının KOM seti kullanıldığında daha düşük \bar{D}_1 değerine sahip olduğunu söyleyebiliriz. Aynı koşullar altında (tek boyutlu-EG/EOG) SL ölçek dönüştürme yöntemi her iki ortak test için aynı \bar{D}_1 değerine sahip olurken, Haebara yöntemi ÇSOM seti kullanıldığında daha düşük \bar{D}_1 değerine sahip olmuştur. Çok boyutluluğun oluştuğu tüm koşullar (EG-Haebara koşulu hariç) altında ise gruplar eşdeğer olsun ya da olmasın elde edilen tüm eşitleme sonuçlarının KOM seti kullanıldığında ÇSOM seti kullanılan koşullara nazaran daha düşük \bar{D}_1 değerine sahip olduğu görülmektedir.

Tablo 4.2.' de verilen \bar{D}_2 değerlerine göre tek boyutlu-eşdeğer gruplar koşulu altında, moment ölçek dönüştürme yöntemleri kullanılarak elde edilen GR_PE sonuçları KOM seti kullanıldığında daha düşük \bar{D}_2 değerlerine sahip iken karakteristik eğri ölçek yöntemleri uygulandığında her iki ortak madde seti kullanılarak elde edilen eşitleme sonuçları aynı \bar{D}_2 değerine (0.037) sahip olmuştur. Grupların eşdeğer olup çok boyutluluğun oluştuğu koşullarda ise O-S yöntemi hariç diğer tüm ölçek dönüştürme yöntemleri kullanılarak elde edilen eşitleme sonuçları ÇSOM seti kullanıldığında daha düşük \bar{D}_2 değerine sahip olmuştur. Grupların eşdeğer olmadığı tek boyutlu ve çok boyutlu koşullar altında yine benzer şekilde O-S ölçek dönüştürme yöntemi hariç diğer tüm ölçek dönüştürme yöntemleri ile elde edilen GR_PE sonuçları ÇSOM seti kullanılan koşullarda daha düşük \bar{D}_2 değerine sahip olma eğiliminde olmuştur.

Ortak madde seti formatının BSE ve İSE özelliği korunumu üzerinde istatistiksel olarak anlamlı bir etkiye sahip olup olmadığını belirlemek üzere Tablo 4.3.'te verilen ANOVA sonuçları incelenmiştir. Tablo 4.3.'te verilen bilgilere dayalı olarak format etkisinin BSE özelliği korunumu ölçütüne göre GR_PE sonuçları üzerinde

anlamli ve buyuk etkisinin oldugunu, İSE ozelligi korunumu olcütüne göre ise anlamli bir etkisinin olmadigini söyleyebiliriz ($\eta_{BSE}^2 = 0.153$, $p < 0.0001$; $\eta_{İSE}^2 = 0.000$, $p > 0.0001$). Yapilan ikili karřılařtırma testi sonucu, KOM ve ÇSOM ortak madde seti kullanımı sonucu elde edilen D_1 deęerleri arasındaki farkın ÇSOM seti kullanımı lehine istatistiksel olarak manidar olduęu görölmüřtür ($\Delta\mu_{BSE} = 0.043$, $p < 0.0001$). Bu bulguya dayalı olarak, BSE özellięinin KOM setinin kullanıldıęı kořullarda daha iyi korunduęunu söyleyebiliriz.

4.1.3. Grupların yetenek daęılımları eřitlik özellięi korunumunu nasıl etkilemektedir?

Grupların yetenek daęılımının BSE özellięinin korunumu üzerine etkisi Tablo 4.1.'de verilen \bar{D}_1 deęerlerine dayalı olarak deęerlendirilmiřtir. Tablo 4.1.'de verilen bilgilere dayalı olarak boyutluluęa, ortak madde setine ve kullanılan ölçek dönüřtürme yöntemine baęlı olmaksızın grupların eřdeęer olduęu kořullarda edilen GR_PE sonuçlarının grupların eřdeęer olmadıęı kořullara nazaran daha düşük \bar{D}_1 deęerine sahip olduęunu söyleyebiliriz.

Grupların yetenek daęılımının GR_PE sonuçlarının İSE özellięinin korunumu üzerine etkisini deęerlendirmek üzere Tablo 4.2.'de yer alan \bar{D}_2 deęerleri incelenmiřtir. Tek boyutlu-KOM kořulu altında, moment ölçek dönüřtürme yöntemleri uygulandıęında hem eřdeęer hem eřdeęer olmayan gruplarda aynı \bar{D}_2 deęeri elde edilirken, karakteristik eęri yöntemleri kullanıldıęında ise eřdeęer gruplarla yapılan GR_PE sonucu daha düşük \bar{D}_2 elde edilmiřtir. Tek boyutlu-ÇSOM kořulu altında ise O-O ölçek dönüřtürme yöntemi hariç dięer tüm ölçek dönüřtürme yöntemleri elde edilen GR_PE sonuçlarının grupların eřdeęer olduęu kořulda daha düşük \bar{D}_2 deęerine sahip olmuřtur. Çok boyutluluęun oluřtuęu kořullar altında (ÇSOM setinin kullanılıp O-S yöntemi uygulandıęı $\rho_{\theta_1\theta_2} = 0.80$ olduęu kořul hariç) hem KOM seti hem ÇSOM kullanıldıęı kořullarda grupların eřdeęer olduęu durumda daha düşük \bar{D}_2 deęerleri elde edilmiřtir.

Tablo 4.3.'te verilen ANOVA sonuçları da grupların yetenek daęılımı faktörünün GR_PE sonuçlarının BSE ve İSE özelliklerinin korunması üzerinde istatistiksel olarak anlamli ve buyuk etkisinin oldugunu göstermiřtir ($\eta_{BSE}^2 = 0.629$, $\eta_{İSE}^2 = 0.433$, $p < 0.0001$). Yapilan ikili karřılařtırmalar testi her iki deęerlendirme

ölçütüne göre bu farklılığın eşdeğer olmayan gruplar lehine istatistiksel olarak anlamlı olduğunu ($\Delta\mu_{BSE} = 0.131$, $\Delta\mu_{ISE} = 0.025$, $p < 0.0001$), eşdeğer olmayan gruplarla elde edilen \bar{D}_1 ve \bar{D}_2 değerlerinin daha büyük olduğunu göstermiştir. Tüm bulgular ışığında grupların eşdeğer olduğu koşul altında elde edilen GR_PE sonuçlarının BSE ve İSE özelliklerini daha iyi koruduğunu söyleyebiliriz.

4.1.4. Ölçek dönüştürme yöntemleri eşitlik özelliğinin korunumunu nasıl etkilemektedir?

Ölçek dönüştürme yöntemlerinin BSE eşitlik özelliği korunumu üzerine etkisi Tablo 4.1.' de verilen \bar{D}_1 değerleri incelenerek değerlendirilmiştir. Tek boyutlu-KOM-EG koşulu altında en büyük \bar{D}_1 değerine (0.072) O-S ölçek dönüştürme yöntemi sahip olurken en düşük \bar{D}_1 değerine (0.067) SL yöntemi ile elde edilen GR_PE sonuçları sahip olmuştur. Aynı koşullar altında (tek boyutlu-KOM-EG) genel olarak BSE özelliğini korumada yakın \bar{D}_1 değerlerine sahip olarak SL, HA ve O-O yöntemlerinin benzer davranış sergilediğini söyleyebiliriz. Yine grupların eşdeğer olduğu koşul altında, kullanılan ortak madde setine bağlı olmaksızın çok boyutluluğun derecesi arttıkça karakteristik eğri ölçek dönüştürme yöntemlerinin moment ölçek dönüştürme yöntemlerinden daha düşük \bar{D}_1 değerlerine sahip olduğu görülmektedir. Benzer bulgular grupların eşdeğer olmadığı koşullarda da elde edilmiştir. Grupların eşdeğer olmadığı tek boyutlu ve çok boyutlu koşullar altında, kullanılan ortak madde setine bağlı olmaksızın karakteristik eğri ölçek dönüştürme yöntemleri sonucu elde edilen GR_PE sonuçlarının moment yöntemleri ile karşılaştırıldığında daha düşük \bar{D}_1 değerlerine sahip olmuştur.

Ölçek dönüştürme yöntemlerinin GR_PE sonucu elde edilen eşitleme sonuçlarının İSE özelliğini korumasını nasıl etkilediğini değerlendirmek amacıyla Tablo 4.2.'de verilen \bar{D}_2 değerleri incelenmiştir. Tek boyutlu-EG koşulları altında kullanılan ortak madde setine bağlı olmaksızın karakteristik eğri yöntemlerinin daha düşük \bar{D}_2 değerlerine sahip olduğunu söyleyebiliriz. Fakat EG koşulu altında çok boyutluluğun derecesi arttıkça tam tersi yönde bulgular elde edilmiştir. Grupların eşdeğer olduğu ve çok boyutluluğun olduğu koşullar altında KOM seti kullanıldığında O-S ölçek dönüştürme yöntemi, ÇSOM seti kullanıldığı koşullarda ise O-O yöntemi daha düşük \bar{D}_2 değerlerine sahip olmuştur. Grupların eşdeğer olmadığı ve KOM seti kullanıldığı çok boyutlu koşullar altında, O-S yöntemi en

düşük \bar{D}_2 değerlerine sahip olmuştur. Aynı koşullar altında (EOG) ÇSOM seti kullanıldığında ise her ne kadar tek boyutlu koşul altında O-O yöntemi en düşük \bar{D}_2 değerine sahip olsa da özellikle çok boyutluluğun derecesinin ciddi olduğu koşullarda karakteristik eğri yöntemleri daha düşük \bar{D}_2 değerlerine sahip olma eğiliminde olmuştur.

Tablo 4.3.'te verilen ANOVA testi sonucuna dayalı olarak, ölçek dönüştürme yöntemi faktörünün BSE özelliği ölçüt olarak alındığında GR_PE yöntemi ile elde edilen eşitleme sonuçları üzerinde anlamlı ve büyük etkiye sahip olduğunu; İSE özelliği ölçüt olarak alındığında ise anlamlı ve orta düzey etkiye sahip olduğunu söyleyebiliriz ($\eta_{BSE}^2 = 0.345$, $\eta_{ISE}^2 = 0.096$, $p < 0.0001$). Ölçek dönüştürme yöntemlerinin BSE ve İSE özelliğini koruma ölçütüne göre aralarında oluşan farklılığının hangi yöntemler arasında oluştuğunu belirlemek amacıyla ikili karşılaştırma testi yapılmış ve elde edilen bulgular Tablo 4.6.'da, her bir ölçek dönüştürme yöntemine ilişkin hesaplanan marjinal ortalamalar Tablo 4.7.'de verilmiştir.

Tablo 4.6. Ölçek Dönüştürme Yöntemlerine İlişkin İkili Karşılaştırma Testi Sonuçları

Ölçek Dönüştürme Yöntemi	D_1 Ortalama Fark	D_2 Ortalama Fark
OO-OS	0.019*	0.006*
OO-SL	0.087*	0.010*
OO-HA	0.740*	0.012*
OS-SL	0.067*	0.004*
OS-HA	0.055*	0.007*
SL-HA	-0.012*	0.002*

* $p < 0.0001$; OO: Ortalama-Ortalama, OS: Ortalama-Sigma, SL: Stocking-Lord, HA: Haebara

Tablo 4.7. Marjinal Ortalamalar

Ölçek Dönüştürme Yöntemleri	D_1 Ortalama	D_2 Ortalama
OO	0.219	0.059
OS	0.200	0.054
SL	0.133	0.049
HA	0.145	0.047

OO: Ortalama-Ortalama, OS: Ortalama-Sigma, SL: Stocking-Lord, HA: Haebara

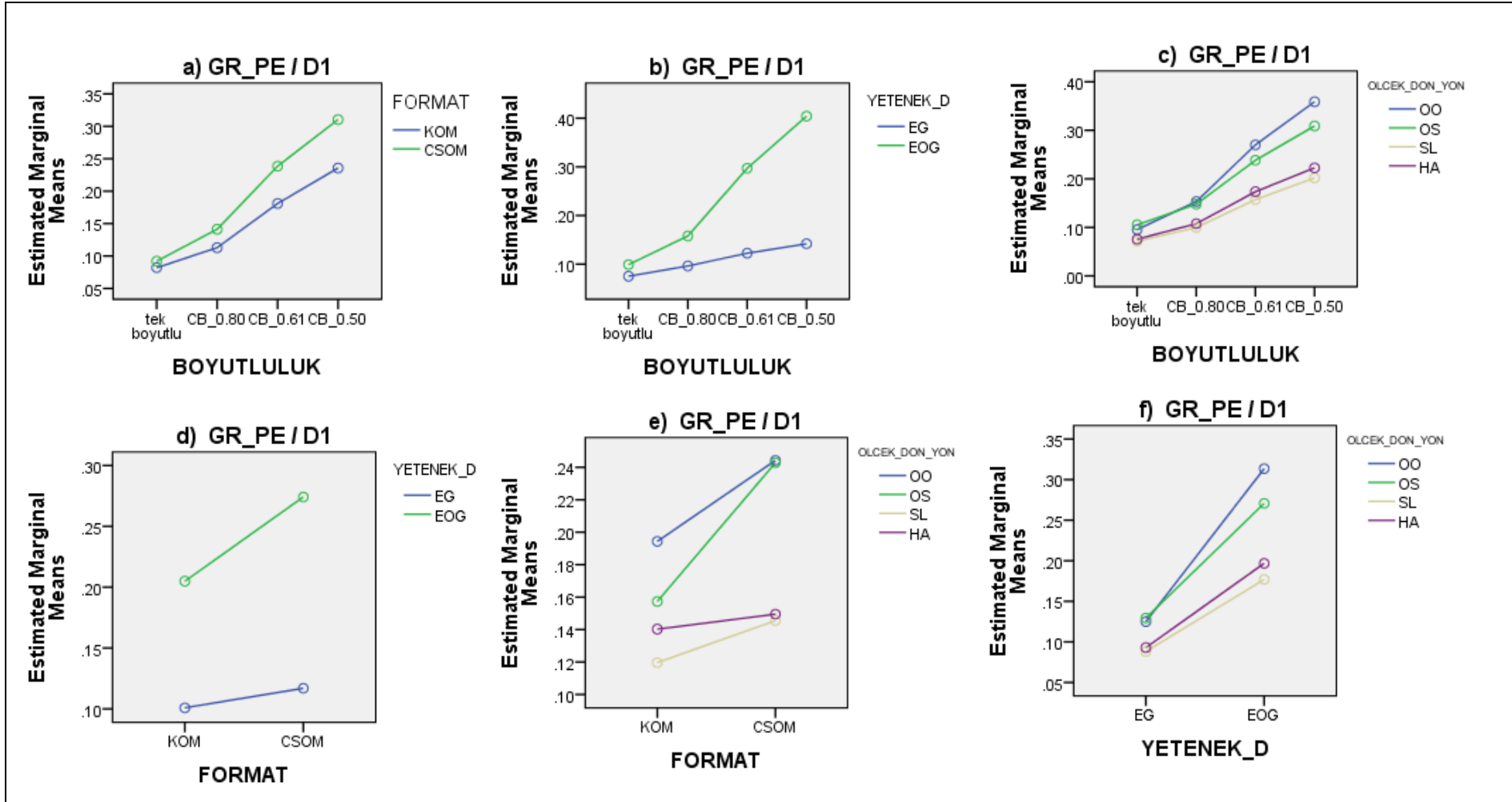
Tablo 4.6. incelendiğinde, bütün ikili karşılaştırmalar arasındaki farkın istatistiksel olarak manidar olduğu görülmektedir. Her iki değerlendirme ölçütüne göre en büyük fark O-O ölçek dönüştürme yöntemi ile HA ölçek dönüştürme yöntemi arasında elde edilirken en küçük farklılık SL ve HA yöntemleri arasında elde

edilmiştir. Tablo 4.7.'deki marjinal ortalamalar incelendiğinde, genel olarak BSE özelliğini koruma ölçütüne göre en iyi performans gösteren yöntemin SL, en kötü performans gösteren yöntemin ise O-O ölçek dönüştürme yöntemi olduğunu söyleyebiliriz. İSE eşitliği ölçüt olarak alındığında ise en kötü performans gösteren yöntem yine O-O yöntemi iken en iyi performans gösteren yöntem HA yöntemi olmuştur. Tablo 4.7.'de verilen marjinal ortalamalara dayalı olarak karakteristik eğri ölçek dönüştürme yöntemleri ile elde edilen GR_PE sonuçlarının daha düşük \bar{D}_1 ve \bar{D}_2 değerlerine sahip olduğunu dolayısıyla BSE ve İSE özelliklerine göre korumada moment ölçek dönüştürme yöntemleri ile karşılaştırıldığında daha iyi performans gösterdiğini söyleyebiliriz.

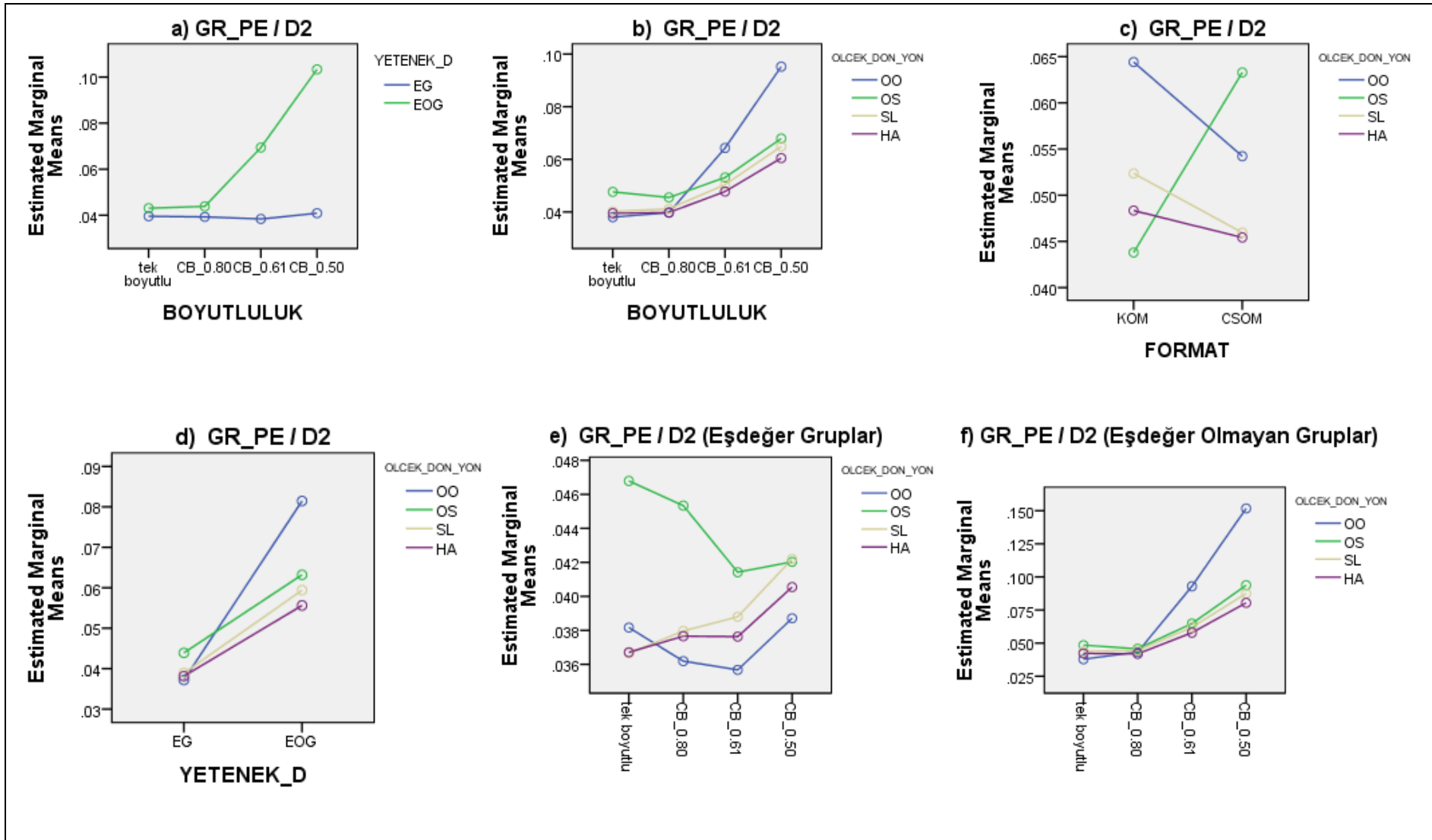
4.1.5. Faktörlerinin etkileşimi eşitlik özelliğinin korunumunu nasıl etkilemektedir?

Araştırma kapsamında incelenen boyutluluk (B), ortak madde seti formatı (F), grupların yetenek dağılımı (YD) ve ölçek dönüştürme yöntemleri (ÖDY) faktörlerinin etkileşimlerinin GR_PE sonuçlarının BSE ve İSE özelliğini koruması üzerinde etkisine dair elde edilen bulgular da Tablo 4.3.'te sunulmuştur. Her iki değerlendirme ölçütüne göre tüm etkileşimler istatistiksel olarak anlamlı olmasına rağmen etki büyüklükleri incelendiğinde BSE ölçütüne göre B*YD ile B*ÖDY etkileşimlerinin büyük; B*F, F*YD, F*ÖDY ve YD*ÖDY etkileşimlerinin orta etki büyüklüğüne sahip olduğunu söyleyebiliriz. İSE ölçütüne göre ise B*YD, B*ÖDY, F*ÖDY ve B*YD*ÖDY etkileşimleri büyük; YD*ÖDY etkileşimi orta büyüklükte etkiye sahiptir. Her iki değerlendirme ölçütüne göre orta ve büyük etkiye sahip etkileşimlere ilişkin çizilen grafikler Şekil 4.1. ve Şekil 4.2.'de verilmiş olup bulgular aşağıda sunulmuştur.

*Boyutluluk*format (B*F):* Tablo 4.3.'te verilen bilgilere dayalı olarak B*F etkileşiminin BSE özelliğinin korunumu ölçütüne göre GR_PE sonuçları üzerinde anlamlı ve orta düzey etkiye, İSE özelliği ölçütüne göre ise anlamlı fakat küçük etkiye sahip olduğunu söyleyebiliriz ($\eta_{BSE}^2 = 0.059, \eta_{ISE}^2 = 0.016, p < 0.0001$). Şekil 4.1.'de verilen grafiğe (a) göre tek boyutlu koşul altında her iki ortak madde seti kullanılarak elde edilen GR_PE sonuçlarının BSE özelliğini korumada benzer performans gösterdiğini söyleyebiliriz.



Şekil 4.1. GR_PE yöntemi ile D_1 ölçütüne göre büyük ve orta etkiye sahip ikili etkileşimlerin grafikleri



Şekil 4.2. GR_PE yöntemi ile D_2 ölçütüne göre büyük ve orta etkiye sahip ikili etkileşimlerin grafikleri

Fakat çok boyutluluğun derecesi arttıkça KOM seti kullanılarak elde edilen eşitleme sonuçları ÇSOM kullanılan koşullara göre daha düşük D_1 değerlerine sahip olmuş dolayısıyla BSE özelliğini korumada daha iyi performans göstermiştir.

*Boyutluluk*yetenek dağılımı (B*YD):* Tablo 4.3.'te yer alan bilgilere göre, GR_PE sonuçlarının BSE ve İSE özelliklerinin korunumu üzerinde B*YD etkileşiminin en büyük etki büyüklüğüne sahip etkileşim olduğunu söyleyebiliriz ($\eta_{BSE}^2 = 0.468$, $\eta_{ISE}^2 = 0.407$, $p < 0.0001$). Şekil 4.1.'de verilen grafik (b) incelendiğinde tek boyutlu koşul altında eşdeğer ve eşdeğer olmayan gruplarla elde edilen GR_PE sonuçlarının benzer D_1 değerlerine sahip olduğunu söyleyebiliriz. Çok boyutluluğun derecesi arttıkça, grupların eşdeğer olduğu durumda elde edilen GR_PE eşitleme sonuçlarının eşdeğer olmayan gruplarla karşılaştırıldığında daha düşük D_1 değerlerine sahip olarak BSE özelliğini korumada daha iyi performans göstermiştir. İSE özelliği ölçütüne göre çizilen Şekil 4.2.'de verilen grafik (a) incelendiğinde ise tek boyutlu ve çok boyutluluğun derecesinin en az olduğu ($\rho_{\theta_1\theta_2} = 0.80$) koşul altında hem eşdeğer hem de eşdeğer olmayan gruplarla elde edilen eşitleme sonuçları benzer D_2 değerlerine sahip olmuştur. Oluşan çok boyutluluğun derecesi arttıkça eşdeğer gruplarla elde edilen eşitleme sonuçları daha düşük D_2 değerlerine sahip olarak İSE özelliğini korumada daha iyi performans göstermiştir.

*Boyutluluk*ölçek dönüştürme yöntemi (B*ÖDY):* Tablo 4.3.'te verilen ANOVA sonuçlarına göre, B*ÖDY etkileşimi, GR_PE sonuçlarının BSE ve İSE özelliklerini korumada büyük etkiye sahip etkileşimlerden biridir ($\eta_{BSE}^2 = 0.141$, $\eta_{ISE}^2 = 0.154$, $p < 0.0001$). Şekil 4.1.'de verilen grafiğe (c) dayalı olarak tek boyutluluk koşulu altında dört ölçek dönüştürme yönteminin çok yakın D_1 değerine sahip olduğunu dolayısıyla BSE özelliğini korumada benzer performans gösterdiğini söyleyebiliriz. Fakat çok boyutluluğun derecesi arttıkça, karakteristik eğri ölçek dönüştürme yöntemleri kullanılarak elde edilen GR_PE sonuçları daha düşük D_1 değerlerine sahip olarak BSE özelliğini korumada moment yöntemleri ile karşılaştırılınca daha iyi performans göstermiştir. Şekil 4.2.'de verilen İSE özelliği ölçütüne göre çizilen grafik (b) incelendiğinde, tek boyutlu koşul altında O-S dışındaki tüm yöntemlerin birbirine yakın ve daha küçük D_2 değerlerine sahip olarak İSE özelliğini korumada benzer performans gösterdiğini söyleyebiliriz. Fakat

çok boyutluluğun derecesi arttıkça O-O yöntemi en büyük D_2 değerine sahip olarak İSE özelliğini korumada en kötü performansı gösteren yöntem olmuştur. Çok boyutluluğun derecesi arttıkça genel olarak karakteristik eğri yöntemlerinin daha düşük D_2 değerlerine sahip olarak İSE özelliğini korumada moment yöntemleri ile karşılaştırıldığında daha iyi performans gösterdiğini söyleyebiliriz.

*Format*yetenek dağılımı (F*YD):* Tablo 4.3.'te verilen bilgilere dayalı olarak F*YD etkileşimi BSE özelliği değerlendirme ölçütüne göre GR_PE sonuçları üzerinde anlamlı ve orta etki büyüklüğüne sahip iken İSE özelliğine göre anlamlı fakat küçük etki büyüklüğüne sahiptir ($\eta_{BSE}^2 = 0.065$, $\eta_{ISE}^2 = 0.007$, $p < 0.0001$). Şekil 4.1.'de verilen grafik (d) incelendiğinde, hem eşdeğer hem eşdeğer olmayan gruptan elde edilen GR_PE sonuçlarının KOM seti kullanıldığı koşullarda daha düşük D_1 değerlerine sahip olduğunu dolayısıyla BSE özelliğini korumada daha iyi performans gösterdiğini söyleyebiliriz.

*Format*ölçek dönüştürme yöntemi (F*ÖDY):* Tablo 4.3'te yer alan ANOVA sonuçlarına dayalı olarak, BSE özelliği ölçütüne göre F*ÖDY etkileşiminin GR_PE sonuçları üzerinde anlamlı ve orta düzey, İSE özelliği ölçütüne göre ise büyük etkisinin olduğunu söyleyebiliriz ($\eta_{BSE}^2 = 0.075$, $\eta_{ISE}^2 = 0.136$, $p < 0.0001$). Şekil 4.1.'de yer alan grafik (e) incelendiğinde, tüm ölçek dönüştürme yöntemleri ile elde edilen GR_PE sonuçlarının KOM seti kullanıldığı koşullarda daha düşük D_1 değerlerine sahip olarak BSE özelliğini daha iyi koruduğunu söyleyebiliriz. Şekil 4.2.'de verilen grafik (c) incelendiğinde ise O-S dışındaki tüm ölçek dönüştürme yöntemleri kullanılarak elde edilen GR_PE sonuçlarının ÇSOM seti kullanıldığında daha düşük D_2 değerlerine sahip olarak İSE özelliğini daha iyi koruduğunu söyleyebiliriz. O-S ölçek dönüştürme yöntemi ise tam ters yönde davranış göstermiş ve KOM seti kullanıldığı koşullarda daha düşük D_2 değerlerine sahip olarak İSE özelliğini daha iyi korumuştur.

*Yetenek dağılımı*ölçek dönüştürme yöntemi (YD*ÖDY):* Tablo 4.3'te verilen bilgilere dayalı olarak YD*ÖDY etkileşiminin her iki değerlendirme ölçütüne göre GR_PE sonuçları üzerinde orta düzey etkisinin olduğunu söyleyebiliriz ($\eta_{BSE}^2 = 0.129$, $\eta_{ISE}^2 = 0.124$, $p < 0.0001$). Şekil 4.1.'de verilen grafik (f) ve Şekil 4.2.'de verilen grafik (d) incelendiğinde, tüm ölçek dönüştürme yöntemlerinin grupların eşdeğer olduğu koşullar altında daha düşük D_1 ve D_2 değerlerine sahip olarak

BSE ve İSE özelliklerinin eşdeğer gruplarla yapılan GR_PE' de daha iyi korunduğunu söyleyebiliriz.

*Boyutluluk*yetenek dağılımı*ölçek dönüştürme yöntemi (B*YD*ÖDY):* Tablo 4.3'te verilen bilgilere göre, B*YD*ÖDY etkileşiminin BSE özelliği ölçütüne göre GR_PE sonuçları üzerinde anlamlı fakat küçük etkiye sahip iken İSE özelliği ölçütüne göre ise anlamlı ve büyük etkiye sahip olduğunu söyleyebiliriz ($\eta_{BSE}^2 = 0.049, \eta_{ISE}^2 = 0.158, p < 0.0001$). Üçlü etkileşimi aynı anda gösteren grafiği çizmek mümkün olmadığından yetenek dağılımına göre B*ÖDY etkileşiminin iki ayrı grafiği çizilerek incelenmiştir. Şekil 4.2.'de verilen grafik (e) incelendiğinde karakteristik eğri yöntemi ile elde edilen GR_PE sonuçlarının en düşük D_2 değerine tek boyutlu koşul altında sahip olduğu çok boyutluluğun derecesi arttıkça genel olarak D_2 değerlerinin de arttığı görülmektedir. Fakat moment ölçek dönüştürme yöntemleri tam tersi yönde davranış göstermiş ve en küçük D_2 değerine çok boyutlu koşul altında ($\rho_{\theta_1\theta_2} = 0.61$) sahip olarak İSE özelliğini en iyi tek boyutlu koşul altında değil de çok boyutlu koşul altında korumuştur. Şekil 4.2.'de verilen grafik (f) incelendiğinde ise grupların eşdeğer olmadığı koşul altında genel olarak tüm ölçek dönüştürme yöntemlerinin en düşük D_2 değerine tek boyutlu koşul ile çok boyutluluğun derecesinin en düşük olduğu koşul ($\rho_{\theta_1\theta_2} = 0.80$) altında, en büyük değere çok boyutluluğun en ciddi olduğu koşul altında ($\rho_{\theta_1\theta_2} = 0.50$) sahip olduğu görülmektedir. Bu bulguya dayalı olarak grupların eşdeğer olmadığı koşullarda çok boyutluluğun derecesi arttıkça GR_PE sonuçlarının İSE özelliğini korumadaki performansının da azaldığını söyleyebiliriz.

4.2. Alt Problem 2'ye İlişkin Bulgular

Madde Tepki Kuramı gözlenen puan eşitleme yöntemi ile yapılan eşitlemelerde, eşitlik özelliğinin korunumunu;

- a) Testlerin boyutluluğu,
- b) Ortak madde seti formatı,
- c) Grupların yetenek dağılımları,
- d) Ölçek dönüştürme yöntemleri ve
- e) Bu faktörlerin etkileşimleri nasıl etkilemektedir?

Bu alt problemin çözümünde değerlendirme ölçütü olarak BSE ve İSE eşitlik özelliği ölçütleri kullanılmıştır. MTK gözlenen-puan eşitleme (GZ_PE) yöntemi

sonucu elde edilen eşitleme sonuçlarının BSE ve İSE özelliklerine dayalı olarak değerlendirmek üzere hesaplanan D_1 indeksi ve D_2 indeksi ortalamaları (\bar{D}_1 ve \bar{D}_2) sırayla Tablo 4.8. ve Tablo 4.9.'da sunulmuştur.

Tablo 4.8. MTK Gözlenen Puan Eşitleme Yöntemine Göre Elde Edilen D_1 İndeksi Ortalamaları

Yetenek Dağılımı	Ortak Madde Seti	Boyutluluk	Ölçek Dönüştürme Yöntemleri			
			O-O	O-S	SL	HA
Eşdeğer Gruplar (EG)	KOM	1.00	0.073	0.073	0.068	0.068
		0.80	0.098	0.090	0.075	0.080
		0.61	0.132	0.115	0.083	0.096
		0.50	0.160	0.135	0.089	0.109
	ÇSOM	1.00	0.078	0.130	0.069	0.069
		0.80	0.113	0.154	0.079	0.079
		0.61	0.157	0.170	0.093	0.095
		0.50	0.194	0.181	0.102	0.104
Eşdeğer Olmayan Gruplar (EOG)	KOM	1.00	0.112	0.100	0.072	0.082
		0.80	0.172	0.130	0.088	0.110
		0.61	0.362	0.224	0.127	0.196
		0.50	0.541	0.308	0.160	0.265
	ÇSOM	1.00	0.124	0.139	0.072	0.078
		0.80	0.228	0.215	0.122	0.129
		0.61	0.496	0.451	0.229	0.240
		0.50	0.723	0.663	0.305	0.321

*KOM: Karma ortak madde, ÇSOM: Çoktan seçmeli ortak madde, O-O: Ortalama-Ortalama, O-S: Ortalama-Sigma, SL: Stocking-Lord, HA: Haebara

Tablo 4.9. MTK Gözlenen Puan Eşitleme Yöntemine Göre Elde edilen D_2 İndeksi Ortalamaları

Yetenek Dağılımı	Ortak Madde Seti	Boyutluluk	Ölçek Dönüştürme Yöntemleri			
			O-O	O-S	SL	HA
Eşdeğer Gruplar (EG)	KOM	1.00	0.027	0.028	0.027	0.027
		0.80	0.029	0.026	0.027	0.028
		0.61	0.031	0.025	0.027	0.027
		0.50	0.035	0.027	0.029	0.029
	ÇSOM	1.00	0.028	0.039	0.027	0.027
		0.80	0.027	0.037	0.026	0.026
		0.61	0.027	0.033	0.024	0.024
		0.50	0.030	0.032	0.025	0.026
Eşdeğer Olmayan Gruplar (EOG)	KOM	1.00	0.031	0.027	0.030	0.031
		0.80	0.039	0.027	0.032	0.032
		0.61	0.081	0.034	0.043	0.042
		0.50	0.142	0.047	0.057	0.053
	ÇSOM	1.00	0.030	0.036	0.030	0.030
		0.80	0.034	0.037	0.028	0.028
		0.61	0.065	0.056	0.035	0.036
		0.50	0.100	0.083	0.048	0.047

*KOM: Karma ortak madde, ÇSOM: Çoktan seçmeli ortak madde, O-O: Ortalama-Ortalama, O-S: Ortalama-Sigma, SL: Stocking-Lord, HA: Haebara

Göreceli olarak düşük \bar{D}_1 ve \bar{D}_2 değerleri BSE ve İSE eşitlik özelliklerinin daha iyi korunduğunun bir göstergesi olarak ele alınmıştır. Boyutluluk, ortak madde seti formatı, yetenek dağılımı ve ölçek dönüştürme yöntemi faktörlerinin ve bu faktörlerin etkileşimlerinin BSE ve İSE özelliklerinin korunması üzerinde istatistiksel olarak anlamlı etkisinin olup olmadığı 4-yönlü ANOVA yapılarak test edilmiş ve elde edilen bulgular Tablo 4.10.' da sunulmuştur.

Tablo 4.10. MTK Gözlenen Puan Eşitleme Sonuçlarının BSE ve İSE Korunumu Ölçütlerine Göre Dört Yönlü ANOVA Testi Sonuçları

<i>Etkiler</i>	<i>sd</i>	<i>D₁</i>		<i>D₂</i>	
		<i>F</i>	η^2	<i>F</i>	η^2
<i>Boyutluluk (B)</i>	3	3436.087*	0.619	1147.953*	0.352
<i>Format (F)</i>	1	1592.789*	0.201	3.825	0.001
<i>Yetenek</i>	1	8470.809*	0.572	3687.503*	0.368
<i>Dağılımı (YD)</i>					
<i>Ölçek Dönüştürme</i>	3	1649.341*	0.438	618.367*	0.226
<i>Yöntemi (ÖDY)</i>					
<i>B*F</i>	3	199.423*	0.086	17.762*	0.008
<i>B*YD</i>	3	1722.635*	0.449	1115.370*	0.346
<i>B*ÖDY</i>	9	231.549*	0.248	199.204*	0.221
<i>F*YD</i>	1	658.045*	0.094	11.990	0.002
<i>F*ÖDY</i>	3	237.457*	0.101	301.002*	0.125
<i>YD*ÖDY</i>	3	584.803*	0.217	462.859*	0.180
<i>B*F*YD</i>	3	176.472*	0.077	1.056	0.000
<i>B*F*ÖDY</i>	9	17.278*	0.024	31.471*	0.043
<i>B*YD*ÖDY</i>	9	108.081*	0.133	146.697*	0.172
<i>F*YD*ÖDY</i>	3	43.970*	0.020	73.541*	0.034
<i>B*F*YD*ÖDY</i>	9	21.554*	0.030	31.074*	0.042
<i>Hata</i>	6336				
<i>Toplam</i>	6399				

*p<0.0001

4.2.1. Testlerin boyutluluğu eşitlik özelliği korunumunu nasıl etkilemektedir?

Tablo 4.8.'de verilen \bar{D}_1 değerleri incelendiğinde grupların yetenek dağılımına, kullanılan ortak madde setine ve uygulanan ölçek dönüştürme yöntemine bağlı olmaksızın boyutluluk arttıkça \bar{D}_1 değerlerinin de arttığı görülmektedir. Örneğin grupların eşdeğer olup, karma ortak madde setinin kullanıldığı koşullar altında O-O ölçek dönüştürme yöntemi uygulanarak elde edilen GZ_PE sonuçlarının en düşük \bar{D}_1 değerine (0.073) tek boyutluluk varsayımının sağlandığı ($\rho_{\theta_1\theta_2} = 1.00$) koşul altında sahip olurken en büyük \bar{D}_1 değerine (0.160) çok boyutluluğun derecesinin ciddi sayılabileceği iki yetenek arasındaki korelasyonun 0.50 olduğu koşul altında sahip olmuştur. Bu bulguya dayalı olarak boyutluluk ile GZ_PE yöntemi ile elde

edilen eşitleme sonuçlarının BSE özelliğini koruma performansı arasında ters orantı olduğunu söyleyebiliriz.

Boyutluluk faktörünün GZ_PE sonuçlarının İSE özelliğini korunumuna etkisini değerlendirmek üzere Tablo 4.9.'de yer alan \bar{D}_2 değerleri incelenmiştir. Grupların eşdeğer olup KOM madde setinin kullanıldığı koşullar altında, O-O ölçek dönüştürme yöntemi uygulandığında çok boyutluluğun derecesi arttıkça \bar{D}_2 değerleri de artmıştır. Aynı koşullar altında (EG-KOM), diğer ölçek dönüştürme yöntemleri uygulandığında ise GZ_PE sonuçları tek boyutlu ve çok boyutlu koşullarda genel olarak aynı ya da artıp azalan \bar{D}_2 değerlerine sahip olarak çok boyutluluğun oluşmasından çok fazla etkilenmemiştir. Hatta O-S ölçek dönüştürme yöntemi uygulandığında en küçük \bar{D}_2 değeri (0.025) çok boyutluluğun olduğu ($\rho_{\theta_1\theta_2} = 0.61$) koşul altında elde edilirken, en büyük değer (0.028) tek boyutlu koşul altında elde edilmiştir. Grupların eşdeğer olduğu ve ÇSOM kullanıldığı koşullarda yine benzer şekilde en küçük D_2 değeri çok boyutluluğun olduğu koşullar altında elde edilirken en büyük D_2 değeri (O-O hariç) tek boyutlu koşul altında elde edilmiştir.

Grupların eşdeğer olmadığı koşullar altında KOM seti kullanıldığında ölçek dönüştürme yöntemine bağlı olmaksızın çok boyutluluğun derecesi arttıkça elde edilen \bar{D}_2 değerleri de artış göstermiştir. Fakat aynı koşullar altında (EOG), ÇSOM seti kullanılıp moment ölçek dönüştürme yöntemleri uygulandığında en küçük \bar{D}_2 değeri tek boyutlu koşul altında elde edilirken, karakteristik eğri ölçek dönüştürme yöntemleri uygulandığında en küçük \bar{D}_2 değeri çok boyutluluğun derecesinin en düşük olduğu ($\rho_{\theta_1\theta_2} = 0.80$) koşul altında elde edilmiştir. Grupların eşdeğer olmadığı koşullar altında genel olarak çok boyutluluğun derecesi arttıkça elde edilen \bar{D}_2 değerlerinin de arttığını söyleyebiliriz.

Tablo 4.10'da verilen ANOVA sonuçları da boyutluluk faktörünün GZ_PE sonuçlarının BSE ve İSE özelliğinin korunması üzerinde istatistiksel olarak anlamlı ve büyük etkisinin olduğunu göstermiştir ($\eta_{BSE}^2 = 0.619$, $\eta_{ISE}^2 = 0.352$, $p < 0.0001$). Boyutluluğa dair GZ_PE sonuçlarında oluşan farklılığın hangi çok boyutluluk dereceleri arasında olduğunu tespit etmek amacıyla ikili karşılaştırmalar

testi yapılmıştır. Elde edilen bulgular ve marjinal ortalamalar sırayla Tablo 4.11. ve Tablo 4.12.'de sunulmuştur.

Tablo 4.11. Boyutluluk Faktörünün İkili Karşılaştırma Testi Sonuçları

<i>Boyutluluk</i>	<i>D₁</i> <i>Ortalama Fark</i>	<i>D₂</i> <i>Ortalama Fark</i>
<i>1.00-0.80</i>	-0.035*	0.000
<i>1.00-0.61</i>	-0.116*	-0.009*
<i>1.00-0.50</i>	-0.185*	-0.021*
<i>0.80-0.61</i>	-0.082*	-0.008*
<i>0.80-0.50</i>	-0.150*	-0.020*
<i>0.61-0.50</i>	-0.068*	-0.012*

*p < 0.0001

Tablo 4.12. Marjinal Ortalamalar

<i>Boyutluluk</i>	<i>D₁</i> <i>Ortalama</i>	<i>D₂</i> <i>Ortalama</i>
<i>1.00</i>	0.088	0.030
<i>0.80</i>	0.123	0.030
<i>0.61</i>	0.204	0.038
<i>0.50</i>	0.273	0.051

Tablo 4.11.'de verilen bilgilere dayalı olarak BSE özelliği ölçüt olarak alındığında yapılan ikili karşılaştırmaların hepsinin istatistiksel olarak anlamlı olduğu görülmektedir. İSE özelliği ölçüt olarak alındığında ise tek boyutlu koşul ile iki yetenek arasındaki korelasyonun 0.80 olduğu bir diğer deyişle çok boyutluluğun en düşük derecede olduğu koşul dışında yapılan tüm ikili karşılaştırmaların istatistiksel olarak anlamlı olduğu görülmektedir. Tablo 4.12.'de verilen marjinal ortalamalar incelendiğinde her iki değerlendirme ölçütüne göre genel olarak çok boyutluluğun derecesi arttıkça \bar{D}_1 ve \bar{D}_2 değerlerinin arttığını, dolayısıyla GZ_PE sonuçlarının BSE ve İSE özelliğini koruma performansının çok boyutluluğun oluşmasından olumsuz yönde etkilendiğini söyleyebiliriz.

4.2.2. Ortak madde seti formatı eşitlik özelliğinin korunumunu nasıl etkilemektedir?

Ortak madde seti formatının, GZ_PE sonuçlarının BSE özelliğini korunumuna etkisini değerlendirmek için Tablo 4.8.'de verilen \bar{D}_1 değerleri incelenmiştir. Tablo 4.8.'deki değerlere dayalı olarak boyutluluğa, grupların dağılımına ve kullanılan ölçek dönüştürme yöntemine bağlı olmaksızın KOM seti kullanılarak elde edilen eşitleme sonuçlarının ÇSOM seti kullanılarak elde edilenlerle karşılaştırılınca (HA yöntemi hariç) daha düşük \bar{D}_1 değerlerine sahip olduğunu söyleyebiliriz. Örneğin,

grupların eşdeğer olup tek boyutluluk varsayımının sağlandığı koşullar altında kullanılan ölçek dönüştürme yöntemine bağlı olmaksızın KOM seti kullanılarak elde edilen GZ_PE sonuçlarının, aynı koşullar altında (tek boyutlu-EG) ÇSOM seti kullanılarak elde edilenlerden daha düşük \bar{D}_1 değerlerine sahip olmuştur.

Ortak madde seti formatının GZ_PE sonuçlarının İSE özelliğinin korunumuna etkisini değerlendirmek üzere Tablo 4.9.'da verilen \bar{D}_2 değerleri incelenmiştir. Grupların eşdeğer olduğu tek boyutlu koşullar altında moment ölçek dönüştürme yöntemleri uygulandığında KOM seti kullanılarak elde edilen eşitleme sonuçları ÇSOM setinin kullanıldığı aynı koşullar ile karşılaştırıldığında daha düşük \bar{D}_2 değerlerine sahip olmuştur. Tek boyutluluk koşulu altında, grupların dağılımına bağlı olmaksızın karakteristik eğri ölçek dönüştürme yöntemleri uygulandığında ise her iki ortak madde seti kullanılarak elde edilen eşitleme sonuçları genel olarak aynı \bar{D}_2 değerine sahip olarak İSE özelliğinin korunmasında benzer performans göstermiştir. Çok boyutluluğun olduğu koşullar altında ise gruplar eşdeğer olsun ya da olmasın O-S ölçek dönüştürme yöntemi hariç diğer ölçek dönüştürme yöntemleri ile elde edilen GZ_PE sonuçları genel olarak ÇSOM seti kullanıldığında KOM seti kullanılan aynı koşullar ile karşılaştırıldığında daha düşük \bar{D}_2 değerlerine sahip olmuştur.

Tablo 4.10.' da verilen ANOVA sonuçları BSE özelliğinin korunumu ölçütüne göre format ana etkisinin GZ_PE sonuçları üzerinde anlamlı ve büyük etkisinin olduğunu fakat İSE özelliğinin korunumu ölçütüne göre anlamlı etkisinin olmadığını göstermektedir ($\eta_{BSE}^2 = 0.201, p < 0.0001$; $\eta_{ISE}^2 = 0.001; p > 0.0001$). Yapılan iki karşılaştırma testi \bar{D}_1 değerleri arasında oluşan farklılığın ÇSOM seti lehine istatistiksel olarak anlamlı olduğunu göstermiştir ($\Delta\mu_{BSE} = 0.057, p < 0.0001$). Tüm bu bulgular ışığında KOM seti ile elde edilen GZ_PE sonuçlarının BSE özelliğini korumada daha iyi performans gösterdiğini söyleyebiliriz.

4.2.3. Grupların yetenek dağılımı eşitlik özelliğinin korunumunu nasıl etkilemektedir?

Grupların yetenek dağılımının GZ_PE sonuçlarının BSE özelliği korunumunu nasıl etkilediğini değerlendirmek üzere Tablo 4.8.'de verilen \bar{D}_1 değerleri incelenmiştir. Boyutluluğa, kullanılan ortak madde setine ve uygulanan ölçek dönüştürme yöntemine bağlı olmaksızın grupların eşdeğer olduğu koşullar altında elde edilen

\bar{D}_1 değerleri aynı koşullar altında eşdeğer olmadığı koşullardan elde edilen \bar{D}_1 değerlerinden daha küçüktür. Örneğin tek boyutluluğun sağlandığı, KOM setinin kullanılıp O-O ölçek dönüştürme yöntemi uygulandığında eşdeğer olmayan gruplar deseni altında elde edilen \bar{D}_1 değeri 0.073 iken aynı koşullar altında eşdeğer olmayan gruplar deseni ile elde edilen \bar{D}_1 değeri 0.112'dir.

Grupların yetenek dağılımının GZ_PE sonuçlarının İSE özelliği korunumuna etkisini değerlendirmek üzere Tablo 4.9.'da verilen \bar{D}_2 değerleri incelenmiştir. Tablo 4.9.'da verilen bilgilere göre, O-S ölçek dönüştürme yöntemi kullanıldığı koşullar hariç diğer tüm ölçek dönüştürme yöntemleri ile elde edilen eşitleme sonuçları grupların eşdeğer olduğu koşullarda eşdeğer olmadığı aynı koşullar ile karşılaştırıldığında daha düşük \bar{D}_2 değerlerine sahip olmuştur. O-S ölçek dönüştürme yöntemi kullanılarak elde edilen eşitleme sonuçları ise tek boyutluluğun sağlandığı koşullar dışında kalan diğer tüm koşullarda, grupların eşdeğer olduğu durumda eşdeğer olmadığı durumla karşılaştırıldığında genel olarak daha düşük \bar{D}_2 değerlerine sahip olmuştur.

Tablo 4.10.' da verilen ANOVA sonuçları da yetenek dağılımı faktörünün GZ_PE sonuçlarının BSE ve İSE özelliklerinin korunumu üzerinde istatistiksel olarak anlamlı ve büyük etkisinin olduğunu göstermiştir ($\eta_{BSE}^2 = 0.572$, $\eta_{ISE}^2 = 0.368$, $p < 0.0001$). Yapılan ikili karşılaştırmalar testi, her iki değerlendirme ölçütüne göre bu farklılığın eşdeğer olmayan gruplar lehine olduğunu göstermiştir ($\Delta\mu_{BSE} = 0.130$, $\Delta\mu_{ISE} = 0.018$, $p < 0.0001$). Tüm bulgular ışığında, eşdeğer gruplarla elde edilen GZ_PE sonuçlarının daha düşük \bar{D}_1 ve \bar{D}_2 değerlerine sahip olduğunu dolayısıyla BSE ve İSE özelliğini daha iyi koruduğunu söyleyebiliriz.

4.2.4. Ölçek dönüştürme yöntemleri eşitlik özelliğinin korunumunu nasıl etkilemektedir?

Ölçek dönüştürme yöntemlerinin, GZ_PE sonuçları BSE özelliği korunumuna etkisini değerlendirmek üzere Tablo 4.8.'de verilen \bar{D}_1 değerleri incelenmiştir. Tablo 4.8' deki bilgilere dayalı olarak boyutluluğa, grupların yetenek dağılımına ve kullanılan ortak madde setine bağlı olmaksızın karakteristik eğri ölçek dönüştürme yöntemleri uygulanarak elde edilen ölçek dönüştürme yöntemlerinin moment

yöntemleri uygulanarak elde edilenlerle karşılaştırdığında daha düşük \bar{D}_1 değerlerine sahip olduğunu söyleyebiliriz.

Ölçek dönüştürme yöntemlerinin, GZ_PE sonuçlarının İSE özelliğinin korunması üzerindeki etkisi Tablo 4.9.' da sunulan \bar{D}_2 değerleri incelenerek değerlendirilmiştir. Tablo 4.9.'daki bilgilere göre tek boyutlu-KOM-EG koşulu altında tüm ölçek dönüştürme yöntemleri yaklaşık olarak eşit \bar{D}_2 değerlerine sahip olmuştur fakat çok boyutluluğun olduğu koşullarda en küçük \bar{D}_2 değerine sahip olan yöntem O-S yöntemi olmuştur. Benzer bulgu grupların eşdeğer olmadığı koşulda da elde edilmiştir. Grupların eşdeğer olmayıp KOM setinin kullanıldığı tek boyutlu ve çok boyutlu koşullar altında O-S yöntemi en düşük \bar{D}_2 değerine sahip olan yöntem olmuştur. Fakat ÇSOM setinin kullanıldığı koşullarda, gruplar eşdeğer olsun ya da olmasın özellikle çok boyutlu koşullar altında karakteristik eğri ölçek dönüştürme yöntemleri uygulanarak elde edilen GZ_PE sonuçları moment ölçek dönüştürme yöntemlerinden elde edilen sonuçlarla karşılaştırılınca daha düşük \bar{D}_2 değerlerine sahip olmuştur.

Tablo 4.10.'da yer alan ANOVA sonuçları da ölçek dönüştürme yöntemi faktörünün BSE ve İSE özellikleri ölçütlerine göre GZ_PE sonuçları üzerinde istatistiksel olarak anlamlı ve büyük etkiye sahip olduğunu göstermiştir ($\eta_{BSE}^2 = 0.438$, $\eta_{ISE}^2 = 0.226$, $p < 0.0001$). Ölçek dönüştürme yöntemlerine göre oluşan farklılığın hangi yöntemler arasında olduğunu tespit etmek amacıyla ikili karşılaştırma testi yapılmıştır. Elde edilen sonuçlar Tablo 4.13.'te, marjinal ortalamalar Tablo 4.14.'te sunulmuştur.

Tablo 4.13. Ölçek Dönüştürme Yöntemleri Faktörünün İkili Karşılaştırma Testi Sonuçları

Ölçek Dönüştürme Yöntemi	D_1 Ortalama Fark	D_2 Ortalama Fark
OO-OS	0.030*	0.010*
OO-SL	0.121*	0.015*
OO-HA	0.102*	0.015*
OS-SL	0.090*	0.005*
OS-HA	0.072*	0.005*
SL-HA	-0.018*	0.000

* $p < 0.0001$; OO: Ortalama-Ortalama, OS: Ortalama-Sigma, SL: Stocking-Lord, HA: Haebara

Tablo 4.14. Marjinal Ortalamalar

Ölçek Dönüştürme Yöntemleri	D_1 Ortalama	D_2 Ortalama
OO	0.235	0.047
OS	0.205	0.037
SL	0.114	0.032
HA	0.133	0.032

OO: Ortalama-Ortalama, OS: Ortalama-Sigma, SL: Stocking-Lord, HA: Haebara

Tablo 4.13.'te incelendiğinde BSE özelliği ölçüt olarak alındığında ölçek dönüştürme yöntemleri arasında yapılan ikili karşılaştırma testlerinin tümünün anlamlı olduğu görülmektedir. BSE özelliği ölçütüne göre ortalamalar arasındaki en büyük fark (0.121) O-O ile SL yöntemleri arasında elde edilirken en küçük farklılık ise SL ile HA (-0.018) ölçek dönüştürme yöntemleri arasında elde edilmiştir. Tablo 4.14..'te verilen marjinal ortalamalara dayalı olarak, en düşük ortalamaya sahip olan SL yönteminin GZ_PE sonuçlarının BSE özelliğini korumasında en iyi performans gösteren ölçek dönüştürme yöntemi olduğunu, O-O yönteminin ise en büyük ortalamaya sahip olarak en kötü performans gösteren ölçek dönüştürme yöntemi olduğunu söyleyebiliriz.

Tablo 4.13.'te İSE özelliği ölçütüne göre SL ve HA yöntemleri hariç diğer tüm ölçek dönüştürme yöntemleri arasında yapılan ikili karşılaştırmaların istatistiksel olarak anlamlı olduğu görülmektedir. Tablo 4.14.'te verilen İSE özelliğine dair ölçek dönüştürme yöntemlerinin sahip olduğu marjinal ortalamalar incelendiğinde, yine en yüksek ortalamaya sahip olan yöntem O-O yöntemidir dolayısıyla İSE özelliğini korumada en kötü performans gösteren yöntem olmuştur. HA ve SL yöntemleri aynı marjinal ortalamaya sahip olarak İSE özelliğini korumada benzer performans göstermişlerdir.

Genel olarak, karakteristik eğri yöntemleri kullanılarak elde edilen GZ_PE sonuçlarının moment yöntemleri ile karşılaştırıldığında daha düşük \bar{D}_1 ve \bar{D}_2 ortalamalarına sahip olduğunu dolayısıyla BSE ve İSE özelliğini korumada daha iyi performans gösterdiğini söyleyebiliriz.

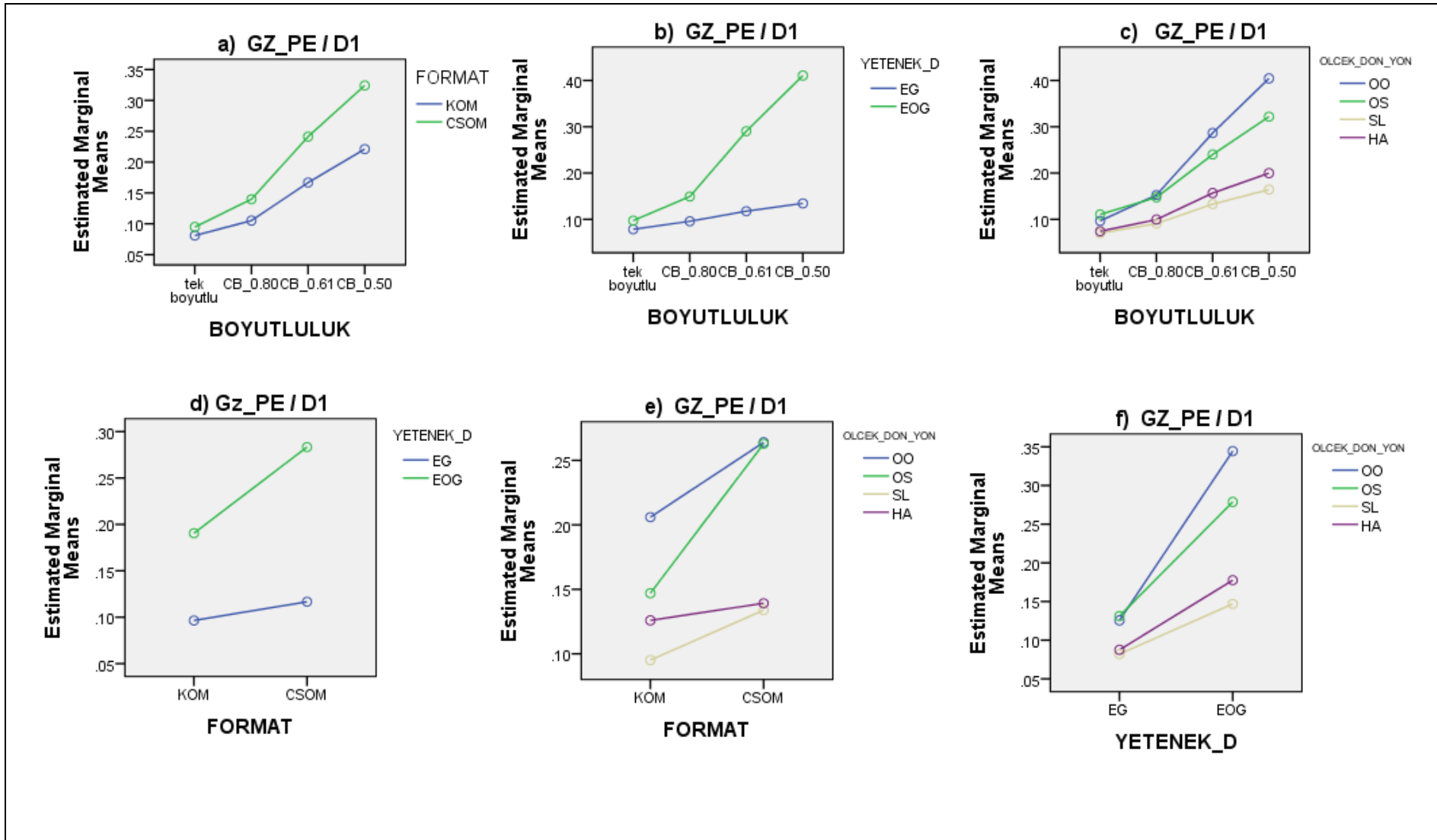
4.2.5. Faktörlerin etkileşimi eşitlik özelliği korunumunu nasıl etkilemektedir?

Araştırma kapsamında incelenen boyutluluk (B), ortak madde seti formatı (F), grupların yetenek dağılımı (YD) ve ölçek dönüştürme yöntemlerinin (ÖDY)

etkileşimlerinin GZ_PE sonuçlarının BSE ve İSE özelliğini koruması üzerinde etkisine dair elde edilen bulgular Tablo 4.10.'da sunulmuştur. Tablo 4.10.'da BSE özelliği ölçütüne göre tüm etkileşimlerin istatistiksel olarak anlamlı olduğu; İSE özelliği ölçütü olarak alındığında ise F*YD ile B*F*YD dışında kalan tüm etkileşimlerin istatistiksel olarak anlamlı olduğu görülmektedir. Her iki değerlendirme ölçütüne göre orta ve büyük etkiye sahip etkileşimlere ilişkin çizilen grafikler sırayla Şekil 4.3., Şekil 4.4. ve Şekil 4.5.'te sunulurken, etkileşimlere ilişkin bulgular ve yorumlar aşağıda sunulmuştur.

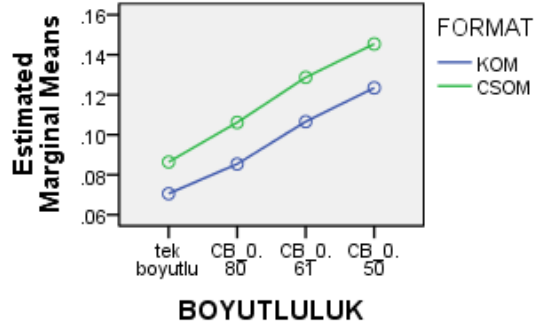
*Boyutluluk*format etkileşimi (B*F):* Tablo 4.10.'da verilen bilgilere dayalı olarak B*F etkileşiminin GZ_PE sonuçları üzerinde BSE özelliği ölçütüne göre anlamlı ve orta düzey; İSE özelliği ölçütüne göre ise anlamlı olmasına rağmen küçük etkisinin olduğunu söyleyebiliriz ($\eta_{BSE}^2 = 0.086, \eta_{ISE}^2 = 0.008, p < 0.0001$). Şekil 4.3.'te verilen grafik (a) incelendiğinde, GZ_PE sonuçlarının tek boyutlu koşul altında ÇSOM ve KOM seti kullanıldığında BSE özelliğini korumada benzer performans gösterdiğini söyleyebiliriz. Çok boyutluluğun derecesi arttıkça ise KOM seti kullanılarak elde edilen GR_PE sonuçları, ÇSOM setinin kullanıldığı koşullarla karşılaştırıldığında daha düşük D_1 değerlerine sahip olarak BSE özelliğini daha iyi korumuştur.

*Boyutluluk*yetenek dağılımı etkileşimi (B*YD):* Tablo 4.10.'da verilen bilgilere dayalı olarak B*YD etkileşiminin GZ_PE sonuçlarının BSE ve İSE özelliğini koruması üzerinde büyük etkiye sahip olduğunu söyleyebiliriz ($\eta_{BSE}^2 = 0.449; \eta_{ISE}^2 = 0.346$). Şekil 4.3.'te verilen grafik (b) incelendiğinde tek boyutlu koşul altında eşdeğer ve eşdeğer olmayan gruplarla elde edilen GZ_PE sonuçlarının BSE özelliğini korumada benzer performans gösterdiğini söyleyebiliriz. Şekil 4.5.'te verilen grafik (a) incelendiğinde, eşdeğer gruplarla elde edilen eşitleme sonuçları tek boyutlu ve çok boyutlu koşullarda genel olarak İSE özelliğini korumada benzer performans göstermiştir. Grupların eşdeğer olmadığı durumda ise özellikle çok boyutluluğun derecesi arttıkça D_2 değerleri de artmıştır. Genel olarak, çok boyutluluğun derecesi arttıkça grupların eşdeğer olduğu koşulda elde edilen eşitleme sonuçlarının daha düşük D_1 ve D_2 değerlerine sahip olduğunu, BSE ve İSE özelliğini korumada grupların eşdeğer olmadığı koşullarla karşılaştırılınca daha iyi performans gösterdiğini söyleyebiliriz.

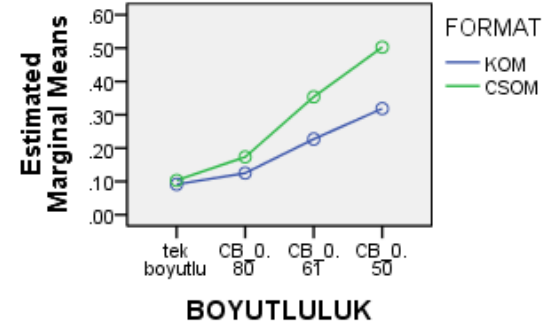


Şekil 4.3. GZ_PE yöntemi ile D_1 ölçütüne göre büyük ve orta etkiye sahip ikili etkileşimlerin grafikleri

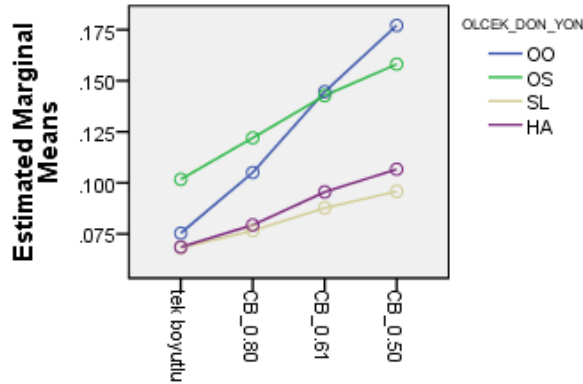
g) GZ_PE / D1 (Eşdeğer Gruplar)



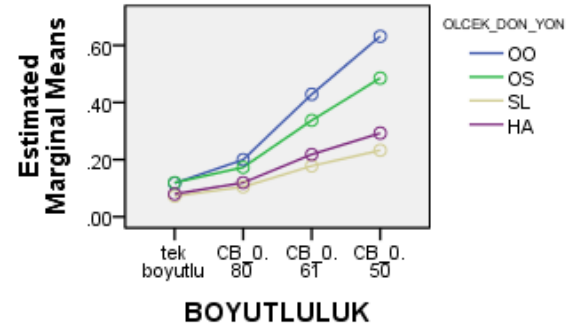
h) GZ_PE / D1 (Eşdeğer Olmayan Gruplar)



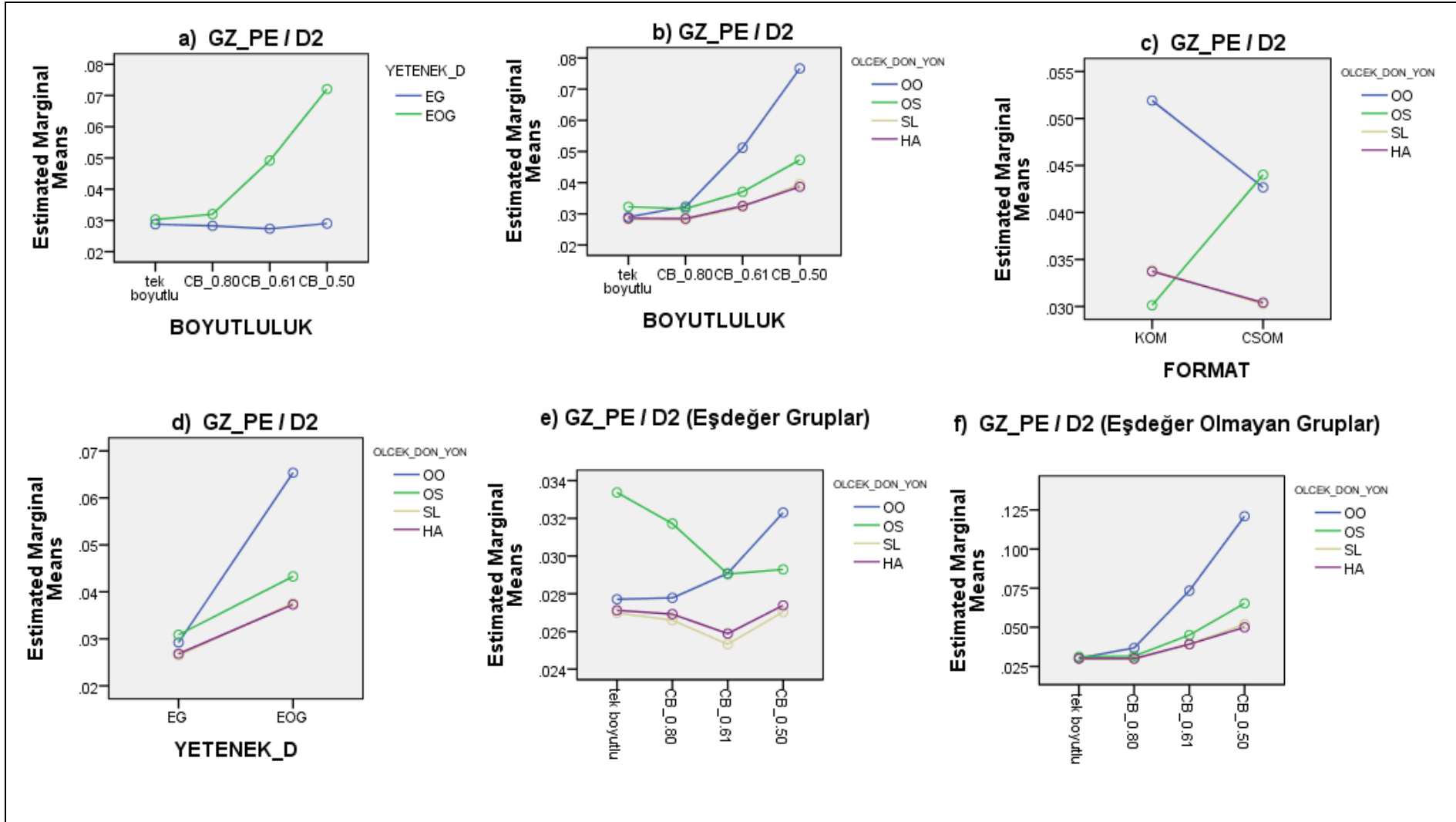
i) GZ_PE / D1 (Eşdeğer Gruplar)



j) GZ_PE / D1 (Eşdeğer Olmayan Gruplar)



Şekil 4.4. GZ_PE yöntemi ile D₁ ölçütüne göre büyük ve orta etkiye sahip ikili etkileşimlerin grafiklerinin devamı



Şekil 4.5. GZ_PE yöntemi ile D_2 ölçütüne göre büyük ve orta etkiye sahip ikili etkileşimlerin grafikleri

*Boyutluluk*ölçek dönüştürme yöntemi (B*ÖDY):* Tablo 4.10.'da verilen bilgilere göre B*ÖDY etkileşiminin GZ_PE sonuçlarının BSE ve İSE özelliklerini koruması üzerinde anlamlı ve büyük etkiye sahiptir ($n_{BSE} = 0.248; n_{ISE} = 0.221, p < 0.0001$). Şekil 4.3.'te verilen grafik (c) incelendiğinde tek boyutluluğun sağlandığı koşullar altında dört ölçek dönüştürme yönteminin BSE özelliğini koruma ölçütüne göre benzer performans gösterdiğini, fakat boyutluluk arttıkça karakteristik eğri yöntemlerinin daha düşük D_1 değerlerine sahip olarak moment yöntemlerinden daha iyi performans gösterdiğini söyleyebiliriz. Şekil 4.5.'te verilen grafik (b) incelendiğinde yine tek boyutluluk koşulu sağlandığında, tüm ölçek dönüştürme yöntemlerinin İSE özelliğini korumada benzer performans gösterdiğini söyleyebiliriz. Fakat çok boyutluluğun derecesi arttıkça, O-S yöntemi her ne kadar O-O yönteminden daha iyi performans gösterse de karakteristik eğri ölçek dönüştürme yöntemlerinin İSE özelliğini korumada moment ölçek dönüştürme yöntemleri ile karşılaştırılınca çok daha iyi performans gösterdiğini söyleyebiliriz.

*Format*yetenek dağılımı (F*YD):* Tablo 4.10.'da verilen bilgilere dayalı olarak F*YD etkileşiminin BSE özelliği ölçütüne göre GZ_PE sonuçları üzerinde anlamlı ve orta düzey etkisi var iken İSE özelliği ölçütüne göre anlamlı etkisinin olmadığını söyleyebiliriz ($\eta_{BSE}^2 = 0.094, p < 0.0001; \eta_{ISE}^2 = 0.002, p > 0.0001$). Şekil 4.3.'te verilen grafik (d) incelendiğinde hem eşdeğer hem de eşdeğer olmayan gruplarla elde edilen GZ_PE sonuçlarının KOM seti kullanıldığı koşullarda ÇSOM setinin kullanıldığı koşullara göre daha düşük D_1 değerlerine sahip olarak BSE özelliğini daha iyi koruduğunu söyleyebiliriz. Özellikle grupların eşdeğer olmadığı koşullarda KOM setinin kullanımı ile ÇSOM seti kullanımının D_1 değerleri arasında oluşturduğu fark ÇSOM seti lehine hızla artış göstermiştir.

*Format*ölçek dönüştürme yöntemi (F*ÖDY):* Tablo 4.10.'daki bilgilere dayalı olarak F*ÖDY etkileşiminin GZ_PE sonuçlarının BSE ve İSE özelliklerini koruması üzerinde anlamlı ve orta düzey etkiye sahip olduğunu söyleyebiliriz ($\eta_{BSE}^2 = 0.101, \eta_{ISE}^2 = 0.125, p < 0.0001$). Şekil 4.3.'te verilen grafik (e) incelendiğinde, tüm ölçek dönüştürme yöntemlerinin KOM seti kullanıldığı koşullarda ÇSOM seti kullanılan koşullara göre daha düşük D_1 değerlerine sahip olma eğiliminde olduğunu dolayısıyla BSE özelliğini daha iyi koruduğunu söyleyebiliriz. Şekil 4.5.'te verilen grafik (c) incelendiğinde sadece O-S ölçek dönüştürme yönteminin KOM

seti kullanıldığı koşullarda daha düşük D_2 değerlerine sahip olarak İSE özelliğini ÇSOM setinin kullanıldığı koşullarla karşılaştırıldığında daha iyi koruduğunu söyleyebiliriz. Diğer üç ölçek dönüştürme yöntemi ÇSOM setinin kullanıldığı koşullarda KOM setinin kullanıldığı koşullara göre daha düşük D_2 değerlerine sahip olarak İSE özelliğini daha iyi korumuştur.

*Yetenek dağılımı*ölçek dönüştürme yöntemi (YD*ÖDY):* Tablo 4.10.'da verilen bilgilere dayalı olarak YD*ÖDY etkileşiminin GZ_PE sonuçlarının BSE ve İSE özelliği korunumu üzerinde istatistiksel olarak anlamlı ve büyük etkiye sahip olduğunu söyleyebiliriz ($\eta_{BSE}^2 = 0.217, \eta_{ISE}^2 = 0.180, p < 0.0001$). Şekil 4.3.'te verilen grafik (f) ve Şekil 4.5.'te verilen grafik (d) incelendiğinde tüm ölçek dönüştürme yöntemlerinin grupların eşdeğer olduğu koşullar altında daha düşük D_1 ve D_2 değerlerine sahip olduğunu, BSE ve İSE özelliklerinin eşdeğer olan gruplarla yapılan GZ_PE sonuçlarında daha iyi korunduğunu söyleyebiliriz.

*Boyutluluk*format*yetenek dağılımı (B*F*YD):* Tablo 4.10.'da verilen bilgilere dayalı olarak, B*F*YD etkileşiminin BSE özelliği ölçütüne göre GZ_PE sonuçları üzerinde orta düzey etkiye sahip iken İSE özelliğine göre anlamlı etkisi yoktur ($\eta_{BSE}^2 = 0.077, p < 0.0001; \eta_{ISE}^2 = 0.000, p > 0.0001$). Üçlü etkileşimi aynı grafik üzerinde göstermek mümkün olmadığından yetenek dağılımına göre çizilen grafik (g) ve grafik (h) Şekil 4.4.'te sunulmuştur. Grupların eşdeğer olduğu durumda, GZ_PE sonuçlarının BSE özelliğini en iyi tek boyutlu ve KOM seti kullanıldığı durumlarda korumuştur. Grupların eşdeğer olmadığı tek boyutluluğun sağlandığı koşulda ise KOM ve ÇSOM seti kullanıldığında GZ_PE sonuçları BSE özelliğini korumada benzer performans göstermiştir. Fakat çok boyutluluğun derecesi arttıkça eşdeğer ve eşdeğer olmayan gruplarla elde edilen GZ_PE sonuçları BSE özelliğini KOM seti kullanıldığında daha iyi korumuştur.

*Boyutluluk*yetenek dağılımı*ölçek dönüştürme yöntemi etkileşimi (B*YD*ÖDY):* Tablo 4.10.'da verilen bilgilere dayalı olarak B*YD*ÖDY etkileşiminin BSE özelliği ölçütüne göre GZ_PE sonuçları üzerinde istatistiksel olarak anlamlı ve orta düzey, İSE özelliği ölçütüne göre ise büyük etkisinin olduğunu söyleyebiliriz ($\eta_{BSE}^2 = 0.133, \eta_{ISE}^2 = 0.172, p < 0.0001$). Üçlü etkileşimi aynı grafikte göstermek mümkün olmadığından yetenek dağılımına göre iki ayrı grafik çizilmiş Şekil 4.4. ve Şekil 4.5.'te sunulmuştur. Şekil 4.4.'te verilen grafik (i) ve (j)'ye dayalı olarak hem

eşdeğer hem eşdeğer olmayan gruplar koşulu altında tüm ölçek dönüştürme yöntemleri sonucu elde edilen eşitleme sonuçlarının BSE özelliğini en iyi tek boyutlu koşul altında, en kötü ise çok boyutluluğun derecesinin en yüksek olduğu koşul altında koruduğunu söyleyebiliriz. Şekil 4.5.'te verilen grafik (e) incelendiğinde grupların eşdeğer olduğu durumda, O-O dışındaki tüm ölçek dönüştürme yöntemleri en küçük D_2 değerine çok boyutluluğun olduğu koşulda ($\rho = 0.61$) sahip olmuştur. Grupların eşdeğer olmadığı koşul için çizilen grafik (f) incelendiğinde ise tüm ölçek dönüştürme yöntemlerinin en düşük D_2 değerlerine tek boyutlu ve çok boyutluluğun derecesinin en düşük olduğu koşul altında sahip olduğu ve boyutluluğun derecesi arttıkça D_2 değerlerinin de arttığı görülmektedir (Bkz. Şekil 4.5.).

4.3.Tartışma ve Yorum

Elde edilen bulgular sonucunda, boyutluluk faktörünün MTK GR_PE ve GZ_PE sonuçlarının BSE ve İSE özelliklerinin korunumu üzerinde anlamlı ve büyük etkisinin olduğu görülmüştür. BSE özelliği korunumu ölçütüne göre tüm koşullar altında en küçük \bar{D}_1 değeri tek boyutlu ($\rho_{\theta_1\theta_2} = 1.00$) koşul altında elde edilirken en büyük değer çok boyutluluğun derecesinin en yüksek ($\rho_{\theta_1\theta_2} = 0.50$) olduğu koşul altında elde edilmiştir. Marjinal ortalamalara dayalı olarak BSE özelliğinin en iyi tek boyutlu koşul altında korunurken en kötü çok boyutluluğun derecesinin en yüksek olduğu koşul altında korunduğu söylenebilir. İSE özelliği ölçütüne göre tek boyutlu ve çok boyutluluğun derecesinin en düşük olduğu koşul altında ($\rho_{\theta_1\theta_2} = 0.80$) elde edilen \bar{D}_2 değerleri arasında anlamlı bir fark olmadığı fakat çok boyutluluğun derecesi arttıkça \bar{D}_2 değerlerinin arttığı ve oluşan farkın anlamlı olduğu görülmüştür. Kestirilen marjinal ortalamalara dayalı olarak, her iki eşitleme yöntemine göre elde edilen eşitleme sonuçlarının İSE özelliğini en iyi tek boyutlu ve çok boyutluluğun derecesinin en düşük olduğu koşul altında korunurken en kötü çok boyutluluğun derecesinin en yüksek olduğu ($\rho_{\theta_1\theta_2} = 0.50$) koşul altında koruduğu söylenebilir.

Literatürde yapılan diğer araştırmalar da boyutluluğun BSE ve İSE özelliklerinin korunumu üzerinde etkili olduğunu göstermiştir. Benzer şekilde Wolf (2013) çalışmasında MTK eşitleme yöntemlerinin de içinde bulunduğu eşitleme yöntemlerinin tek boyutlu ve çok boyutlu koşullardaki performansını çeşitli

faktörlere göre eşitlik özelliğinin korunumunu ölçüt olarak değerlendirmiştir. Wolf (2013) çalışmasında, çok boyutlu koşullarla karşılaştırıldığında tek boyutlu koşul altında GR_PE sonuçlarının daha düşük \bar{D}_1 değerlerine sahip olduğunu, GZ_PE sonuçlarının ise daha düşük \bar{D}_2 değerlerine sahip olduğunu bulmuştur. Wolf (2013)'ün elde ettiği bu bulgular araştırma sonuçlarıyla tutarlılık göstermektedir. Fakat Wolf (2013) GR_PE sonuçlarını İSE özelliğine göre karşılaştırdığında tek boyutlu koşul altında daha büyük \bar{D}_2 değerleri; benzer şekilde GZ_PE sonuçlarını BSE özelliği ölçütüne göre karşılaştırdığında yine tek boyutlu koşul altında daha büyük D_1 değerleri elde etmiştir. Wolf'un bu bulgusu araştırmamızın bulguları ile tutarlılık göstermemektedir. Oluşan bu farklılığın sebebi Wolf (2013)'ün çalışmasında çok boyutluluğu bifaktör modeline göre oluşturup, genel bir faktörün her iki madde formatının ölçtüğü yeteneği ölçüyor şeklinde kurgulaması olabilir.

Literatürde karma testleri etkileyen faktörlerin farklı eşitleme yöntemi ve ölçütlerine göre değerlendirildiği araştırmalar da madde formatına dayalı oluşan çok boyutluluğun eşitleme sonuçlarını etkilediğini göstermiştir (Cao, 2008; Lee vd., 2012; Wang, 2013).

Bu araştırma kapsamında elde edilen bir diğer bulgu ortak madde seti formatı faktörünün BSE özelliğinin korunumu ölçütüne göre GR_PE ve GZ_PE sonuçları üzerinde istatistiksel olarak anlamlı ve büyük etkiye sahip iken İSE özelliğinin korunumu ölçütüne göre anlamlı bir etkisinin olmadığıdır. Boyutluluk*format etkileşiminin BSE özelliği ölçütüne göre GR_PE ve GZ_PE sonuçları üzerinde anlamlı ve orta düzey etkisinin olduğu, İSE özelliği ölçütüne göre ise anlamlı olmasına rağmen küçük etkisinin olduğu görülmüştür. Bulgular özellikle çok boyutlu koşullar altında KOM seti kullanılarak elde edilen eşitleme sonuçlarının BSE özelliğini daha iyi koruduğunu göstermiştir.

Wolf (2013), çalışmasında KOM seti kullandığı koşullarda ÇSOM setinin kullanıldığı koşullarla karşılaştırılınca daha düşük \bar{D}_1 ve \bar{D}_2 değerleri elde etmiş fakat ANOVA testi sonucu bu farklılıkların anlamlı olmadığını bulmuştur. Wang (2013) geleneksel eşitleme yöntemlerini çeşitli ölçütlere göre değerlendirmiş ve ortak madde setinin kompozisyonunun özellikle çok boyutluluğun derecesinin en yüksek olduğu koşullarda RMSE ve bias değerlerini yükselttiğini fakat tesadüfî hatanın sabit kaldığını bulmuştur. Yine Wang (2013) araştırmasında tek

boyutluluğun sağlandığı ve grupların yetenek dağılımının benzer olduğu koşullarda KOM ve ÇSOM setinin benzer performans gösterdiğini bulmuştur.

Cao (2008) RMSE, yanlılık ve sınıflama tutarlılığı ölçütlerine göre eşitleme sonuçlarını değerlendirdiği çalışmasında özellikle çok boyutluluğun madde formatına dayalı olarak oluşturduğu koşullarda ortak madde setinin format olarak testin bütününe temsil etmesinin önemli bir faktör haline geldiğini bulmuştur. Hagge (2010) araştırmasında, eğer testi alan bireyler belirli bir madde formatını diğer madde formatına göre daha zor buluyorlar ise karma testlerin sadece çoktan seçmeli ortak maddelerle eşitlenmesinin büyük yanlılık (bias) değerlerine neden olabileceğini göstermiştir.

Hem bu araştırmadan hem diğer araştırmalardan format etkisine dair elde edilen bulguları dikkatli yorumlamak gerekir. Eğer test tek boyutlu ise farklı madde formatlarının aynı yapıyı ölçtüğü varsayılabilir. Bu durumda ortak madde setinin testin bütününe temsil etmesi ya da etmemesi eşitleme sonuçlarını önemli ölçüde etkileyemeyebilir. Fakat eğer test çok boyutlu ise ve öğrenciler çoktan seçmeli ve açık uçlu maddelerde farklı performans gösteriyorsa daha doğru eşitleme sonuçları elde etmek için ortak madde setinin testin bütününe temsil etmesi önemli hale gelebilir. Ortak madde setinin, grup farklılıklarını tam olarak yansıtabilmesi için Kolen ve Brennan (2004) ortak madde setinin testin bütününe küçük bir minyatürü olması gerektiğini belirtmişlerdir.

Grupların yetenek dağılımı faktörünün GR_PE ve GZ_PE sonuçlarının BSE ve İSE özelliklerini korunumu üzerinde istatistiksel olarak anlamlı ve büyük etkisinin olduğu sonucuna ulaşılmıştır. Genel olarak, grupların eşdeğer olduğu koşullar altında elde edilen eşitleme sonuçları grupların eşdeğer olmadığı koşullarda elde edilenlerle karşılaştırılınca daha düşük \bar{D}_1 ve \bar{D}_2 değerlerine sahip olma eğiliminde olduğu görülmüştür. Genel olarak tüm koşullar altında grupların eşdeğer olduğu durumda elde edilen GR_PE ve GZ_PE sonuçlarının eşitlik özelliğini daha iyi koruduğunu söyleyebiliriz. Literatürde yapılan diğer araştırmalar da eşitlik özelliğinin grupların eşdeğer olduğu koşullarda daha iyi korunduğunu göstermiştir (Andrews, 2011; Wolf, 2013). Ayrıca test eşitleme sonuçlarının farklı ölçütlere göre değerlendirildiği birçok araştırma farklı formları alan grupların aynı ya da benzer evrenden geldiği durumlarda, elde edilen eşitleme sonuçlarının daha az eşitleme

hatası içerdiğini göstermiştir (Cao, 2008; Hagge, 2010; Kim ve Lee, 2006; Kirkpatrick, 2005; Sinharay ve Holland, 2007). Grupların yetenek dağılımlarının birbirinden farklı olduğu koşullarda gruplardan oluşan farklılık formlardan oluşan farklılığa karıştırıcı etki oluşturabilir. Bu da formlar arasındaki güçlük farklılıklarını düzenlemek için yürütülen eşitleme çalışmasını daha karmaşık hale getirebilmektedir.

Bu araştırma kapsamında ölçek dönüştürme yöntemi faktörünün BSE özelliği ölçüt olarak alındığında GR_PE ve GZ_PE yöntemi ile elde edilen eşitleme sonuçları üzerinde anlamlı ve büyük etkiye sahip olduğu görülmüştür. İSE özelliği ölçüt olarak alındığında GR_PE sonuçları üzerinde orta düzey etkiye sahip iken GZ_PE sonuçları üzerinde büyük etkiye sahip olduğu görülmüştür. Boyutluluk*ölçek dönüştürme etkileşimine ilişkin bulgulara göre tek boyutlu koşullar altında tüm ölçek dönüştürme yöntemleri sonucu elde edilen eşitlenmiş puanlar BSE ve İSE özelliğini korumada benzer performans göstermiş fakat özellikle çok boyutluluğun derecesinin arttığı koşullarda karakteristik eğri ölçek dönüştürme yöntemleri moment yöntemlerinden çok daha iyi performans göstermiştir.

Literatürde ölçek dönüştürme yöntemlerinin eşitlik özelliği korunumu ölçütüne göre karşılaştırıldığı herhangi bir çalışmaya rastlanmamıştır. Ölçek dönüştürme yöntemlerinin çeşitli ölçütlere göre değerlendirildiği araştırmalarda genel olarak karakteristik eğri ölçek dönüştürme yöntemleri sonucu elde edilen eşitleme sonuçlarının daha az hata içerdiği bulunmuştur (Baker ve Al-Karni, 1991; Kim ve Lee, 2004; Kim ve Lee, 2006; Hanson ve Beguin, 2002). Kim (2004) araştırmasında, format etkisine rağmen karakteristik eğri ölçek dönüştürme yöntemlerinin moment yöntemlerinden daha iyi performans gösterdiği ve iki karakteristik eğri ölçek dönüştürme yöntemi arasındaki farkın önemsiz olacağını ölçüde küçük olduğu sonucuna ulaşmıştır.

Bu durumun oluşmasının bir sebebi karakteristik eğri ölçek dönüştürme yöntemlerinin bağlanma katsayılarını belirlerken a-, b- ve c- parametrelerini eş zamanlı olarak dikkate alması, moment yöntemlerinin ise madde parametrelerini ayrı ayrı dikkate alması olabilir (Tian, 2011). Bir diğer sebebini Ogasawara (2002) madde parametreleri kararsız kestirilse bile madde karakteristik eğrilerinin kararlı kestirilebilmesiyle açıklamıştır. Kolen ve Brennan (2004) test eşitleme

alıřmalarında ayrı kalibrasyon yapılıp karakteristik eđri lek dnřtrme yntemlerinin kullanılmasının daha gvenli olduđunu belirtmiřlerdir. Arařtırmamız bu bulgusu ile literatrde yer alan karakteristik eđri lek dnřtrme yntemlerinin daha az eřitleme hatası ierdiđi gerekesi ile tercih edilmesini eřitlik zelliđi deđerlendirme ltne gre de desteklemiřtir.

5. SONUÇ ve ÖNERİLER

Bu bölümde araştırmanın bulgu ve yorumlarına dayalı olarak ulaşılan sonuçların özetine ve bu sonuçlardan yola çıkarak geliştirilen önerilere yer verilmiştir.

5.1. Sonuçlar

Bu araştırma kapsamında boyutluluğun, ortak madde seti formatının, grupların yetenek dağılımının ve ölçek dönüştürme yöntemlerinin Madde Tepki Kuramı gerçek puan eşitleme (GR_PE) ve gözlenen puan eşitleme (GZ_PE) yöntemlerine göre karma testlerin eşitlenmesinde eşitlik özelliğinin korunumu üzerine etkisi incelenmiştir. Eşitlik özelliğinin korunumu birinci-sıra eşitlik (BSE) ve ikinci-sıra eşitlik (İSE) olmak üzere iki ölçüte göre değerlendirilmiştir. GR_PE ve GZ_PE yöntemleri sonucu elde edilen sonuçlar aşağıda sunulmuştur.

- 1) GR_PE ve GZ_PE yöntemlerine göre elde edilen eşitleme sonuçlarının BSE özelliğini en iyi tek boyutlu en kötü ise çok boyutluluğun derecesinin en yüksek olduğu ($\rho_{\theta_1\theta_2} = 0.50$) koşulda koruduğu sonucuna ulaşılmıştır. Her iki eşitleme yöntemine göre elde edilen eşitleme sonuçlarının İSE özelliğini ise en iyi tek boyutlu ve çok boyutluluğun derecesinin en düşük olduğu ($\rho_{\theta_1\theta_2} = 0.80$) koşul altında, en kötü ise yine çok boyutluluğun derecesinin en yüksek olduğu koşul altında koruduğu sonucuna ulaşılmıştır.
- 2) GR_PE ve GZ_PE yöntemleri kullanılarak elde edilen eşitleme sonuçlarının BSE özelliğini KOM setinin kullanıldığı koşullarda daha iyi koruduğu sonucuna ulaşılmıştır. Tek boyutlu koşul altında karma ve çoktan seçmeli ortak madde seti kullanılarak elde edilen eşitleme sonuçları BSE özelliğini korumada benzer performans gösterirken çok boyutluluğun derecesi arttıkça KOM seti kullanılarak elde edilen eşitleme sonuçları BSE özelliğini korumada daha iyi performans göstermiştir. Grupların eşdeğer olduğu ve olmadığı koşullarda yine KOM seti kullanılarak elde edilen eşitleme sonuçlarının BSE özelliğini daha iyi koruduğu sonucuna ulaşılmıştır. İSE özelliği ölçüt olarak alındığında ise her iki ortak madde seti kullanılarak elde edilen eşitleme sonuçları benzer \bar{D}_2 değerlerine sahip olmuş ve aralarında istatistiksel olarak anlamlı fark oluşmamıştır. Dolayısıyla karma ya da çoktan seçmeli ortak madde seti kullanımının GR_PE ve GZ_PE

sonuçlarının İSE özelliği korunumu üzerinde fark oluşturmadığı sonucuna ulaşılmıştır.

- 3) Tek boyutlu koşul altında, eşdeğer ve eşdeğer olmayan gruplarla elde edilen GR_PE ve GZ_PE sonuçları BSE özelliğini korumada benzer performans göstermiştir. Fakat çok boyutluluğun derecesi arttıkça eşdeğer gruplarla elde edilen eşitleme sonuçları, eşdeğer olmayan gruplarla karşılaştırıldığında BSE özelliğini daha iyi korumuştur. İSE özelliği ölçütüne göre ise tek boyutlu ve çok boyutluluğun derecesinin en düşük olduğu koşul altında eşdeğer ve eşdeğer olmayan gruplarla elde edilen GR_PE ve GZ_PE sonuçları benzer performans göstermiştir. Fakat çok boyutluluğun derecesi arttıkça benzer şekilde eşdeğer gruplarla elde edilen eşitleme sonuçları İSE özelliğini daha iyi korumuştur. Genel olarak grupların eşdeğer olduğu koşul altında her iki eşitleme yöntemine dayalı olarak elde edilen sonuçların BSE ve İSE özelliklerini daha iyi koruduğu sonucuna ulaşılmıştır.
- 4) Tek boyutluluk koşulu altında, dört ölçek dönüştürme yöntemi kullanılarak elde edilen GR_PE ve GZ_PE sonuçlarının BSE ve İSE özelliklerini korumada benzer performans gösterdiği ancak çok boyutluluğun derecesi arttıkça karakteristik eğri ölçek dönüştürme yöntemlerinin moment ölçek dönüştürme yöntemleri ile karşılaştırıldığında her iki eşitlik özelliğini daha iyi koruduğu sonucuna varılmıştır.
- 5) Tüm ölçek dönüştürme yöntemleri ile elde edilen GR_PE ve GZ_PE sonuçlarının KOM seti kullanıldığı koşullarda BSE özelliğini daha iyi koruduğu sonucuna ulaşılmıştır. O-S dışındaki tüm ölçek dönüştürme yöntemleri kullanılarak elde edilen GR_PE ve GZ_PE sonuçları ÇSOM seti kullanıldığında, O-S ölçek dönüştürme yöntemi ile elde edilen eşitleme sonuçları ise KOM seti kullanıldığında İSE özelliğini daha iyi korumuştur.
- 6) Tüm ölçek dönüştürme yöntemleri kullanılarak elde edilen GR_PE ve GZ_PE sonuçlarının grupların eşdeğer olduğu koşullar altında BSE ve İSE özelliklerini daha iyi koruduğu sonucuna ulaşılmıştır.

5.2. Öneriler

Bu arařtırmadan elde edilen bulguların karma testlerin eřitlenmesine ynelik yapılacak uygulamalara ve alıřmalara ışık tutacağı dřnlmektedir.

5.2.1. Arařtırmanın Bulgularına Dayalı Olarak Öneriler

- 1) Bu arařtırma kapsamında BSE ve İSE zelliklerinin en iyi Madde Tepki Kuramı'nın tek boyutluluk varsayımı saęlandığı kořullarda korunduęu zellikle ok boyutluluęun derecesi arttıka eřitlik zellięinin korunumunun olumsuz ynde etkilendięi grlmřtr. Eęer tek boyutlu MTK test eřitleme yntemleri kullanılarak test eřitleme alıřması yrtlecek ise testlerin tek boyutluluk varsayımını saęlamasına dikkat edilmesi gerekmektedir.
- 2) Arařtırmanın sonucunda, tek boyutlu kořul altında KOM ya da SOM seti kullanıldıęında eřitlik zellięini korumada GR_PE ve GZ_PE yntemlerinin benzer performans gsterdięi fakat zellikle ok boyutluluęun derecesi arttıka KOM seti kullanarak elde edilen eřitleme sonularının eřitlik zellięini (zellikle BSE zellięini) daha iyi koruduęu grlmřtr. Eęer eřitlenecek karma testler tek boyutluluk varsayımını saęlıyor ise ortak madde seti karma formatta ya da sadece oktan semeli formatta kullanılabilir. Fakat eęer karma testler, tek boyutluluk varsayımını ihlal ediyor ise ortak madde setinin testin btn yansıtacak Őekilde karma formatta hazırlanmasına dikkat edilmelidir.
- 3) Arařtırmanın sonucunda, grupların eřdeęer olduęu durumda elde edilen GR_PE ve GZ_PE sonularının eřitlik zellięini her kořulda eřdeęer olmayan gruplarla elde edilen eřitleme sonularına nazaran daha iyi koruduęu grlmřtr. Fakat NEAT deseni zellikle grupların eřdeęer olmadıęı kořullarda testleri eřitleyebilmek iin geliřtirilen bir desendir. Gerek hayatta grupların eřdeęer olduęu kořullarla karřılařmak ok zordur. Dolayısıyla grupların eřdeęer olmadıęı kořullarda eřitleme alıřmasını en iyi Őekilde yrtebilmek iin dięer hususlara dikkat edilmelidir. rneęin, bu arařtırmada grupların eřdeęer olmadıęı kořullar altında KOM seti kullanılarak elde edilen eřitleme sonuları BSE ve İSE zellięini daha iyi koruduęu grlmřtr. Dolayısıyla grupların eřdeęer olmadıęı kořullar

altında özellikle ortak madde setinin testin bütününe yansıtacak şekilde hazırlanmasına dikkat edilebilir.

- 4) Bu araştırmanın sonucunda seçilen ölçek dönüştürme yönteminin eşitlik özelliği korunumu ölçütüne göre eşitleme sonuçları üzerinde etkili olduğu görülmüştür. Her ne kadar tek boyutluluk koşulu sağlandığında moment ölçek dönüştürme yöntemleri eşitlik özelliğini korumada karakteristik eğri ölçek dönüştürme yöntemlerine yakın performans göstermiş olsa da özellikle çok boyutluluğun olduğu koşullarda karakteristik eğri ölçek dönüştürme yöntemleri önemli ölçüde ve büyük farkla daha iyi performans göstermiştir. Dolayısıyla karma formattaki testler NEAT deseni kullanılarak eşitlenirken karakteristik eğri ölçek dönüştürme yöntemlerinden (SL ve HA) biri seçilebilir.
- 5) Test eşitleme literatüründe, eşitlemenin doğruluğunu değerlendirmek için bir ölçüt bulunmamaktadır. Örneğin, “ D_1 ya da D_2 indeksinin büyüklüğü ne olursa eşitleme sonuçlarının eşitlik özelliği korunur ya da korunmaz?” bu sorunun cevabı olacak herhangi bir ölçüt bulunmamaktadır. Dolayısıyla “Standartlar” (AERA, NCME, APA) çerçevesinde bu ölçütün belirlenip rapor edilmesinde fayda vardır.

5.2.2. İleride Yapılacak Araştırmalara Dönük Öneriler

- 1) Bu çalışmada boyutluluk, ortak madde seti formatı, grupların yetenek dağılımı ve ölçek dönüştürme yöntemi olarak dört faktörün karma formattaki testlerin MTK eşitleme yöntemleri sonucu elde edilen eşitleme sonuçlarını nasıl etkilediği incelenmiştir. İleride yapılacak olan araştırmalarda, bu araştırma kapsamında bir faktör olarak alınmayan örneklem büyüklüğü, test uzunluğu, ortak madde seti uzunluğu gibi faktörlerin karma formattaki testlerin eşitlenmesine etkisi araştırılabilir.
- 2) Bu araştırma kapsamında değerlendirme ölçütü olarak birinci-sıra eşitlik ve ikinci-sıra eşitlik özellikleri kullanılmıştır. İleride yapılacak araştırmalarda, farklı değerlendirme ölçütlerine göre (RMSE, yanlılık, sınıflama tutarlılığı...) karma formattaki testleri etkileyen faktörler incelenebilir. Ayrıca bu araştırma sadece MTK eşitleme yöntemleri kullanılarak yürütülmüştür.

ileride yapılacak olan arařtırmalarda geleneksel ve MTK eřitleme yontemleri karřılařtırılabilir.

- 3) Bu arařtırma simülasyon verileri kullanılarak yürütölmüřtür. Her ne kadar simülasyon çalıřmasının gerçeęi yansıtması için çaba gösterilse de gerçek durumu birebir yansıtması beklenemez. Dolayısıyla bu arařtırma kapsamında incelenen ve incelenmeyen faktörlerin karma formattaki testlerin eřitlenmesini etkisi gerçek veri kullanılarak arařtırılabilir.
- 4) Bu arařtırmada tek boyutlu ve çok boyutlu kořullarda tek boyutlu MTK test eřitleme yontemleri kullanılarak çeřitli faktörlerin eřitlik özellięi korunumunu nasıl etkiledięi incelenmiřtir. Çok boyutlu eřitleme yontemleri kullanıldıęında eřitlik özellięinin çeřitli faktörlere göre nasıl etkilendięi incelenebilir.
- 5) Bu arařtırmada çok boyutluluk sadece madde formatına dayalı oluřturulmuřtur. Oysaki gerçek test durumlarında çok boyutluluęun oluřmasının birçok sebebi olabilir. Örneęin, ölçölen kapsama ya da ölçölen özellięe dayalı olarak oluřan çok boyutluluęun karma formattaki testleri nasıl etkiledięi arařtırılabilir.

KAYNAKÇA

- Ackerman, T. (1989). Unidimensional IRT calibration of compensatory and noncompensatory items. *Applied Psychological Measurement*, 13, 113-127.
- Andrews, B.J. (2011). Assessing first-and second-order equity for the common item nonequivalent groups design using multidimensional IRT. Unpublished doctoral dissertation, University of Iowa.
- Angoff, W.H. (1971). Scales, norms, and equivalent scores. In R.L. Thorndike (Ed.), *Educational measurement* (2nd ed., pp.508-600).Washington DC: American Council on Education.
- Angoff, W.H. (1984). *Scales, norms and equivalent scores*. New Jersey: Educational Testing Service.
- Atalay-Kabasakal, K. (2014). Değişen madde fonksiyonunun test eşitlemeye etkisi. Yayınlanmamış doktora tezi. Hacettepe Üniversitesi, Eğitim Bilimler Enstitüsü, Ankara.
- Baghi, H., Bent, P., DeLain, M., & Hennings, S. (1995). *A comprison of the results from two equatings for performance-based student assessments*. Paper presented at the Annual Meeting of the National Council on Measurement in Education, San Francisco, CA.
- Baker, F.B.,& Al-Karni, A. (1991). A comparison of two procederes for computing IRT equating coefficients. *Journal of Educational Measurement*, 28(2), 147-162.
- Bastari, B. (2000). Linking multiple-choice and constructed-response items a common proficiency scale. Unpublished doctoral dissertation, University of Massachusetts.
- Beguin, A.A., Hanson, B.A., & Glas, C.A. (2000). *Effect of multidimensionality on separate and concurrent estimation in IRT equating*. Paper presented at the annual meeting of the National council on Measurement in Education, New Orleans, LA.
- Beguin, A.A., Hanson, B.A. (2001). Effect of noncompensatory multidimensionality on separate and concurrent estimation in IRT observed score equating. Paper presented at the annual meeting of the National council on Measurement in Education, Seattle, WA.
- Bolt, D.M. (1999). Evaluating the effects of multidimensionality on IRT true-score equating. *Applied Measurement in education*, 12(4), 383-407.
- Bozdağ, S. & Kan, A. (2010). Şans başarısının eşitlemeye etkisi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 39, 91-108.
- Braun, H.I., & Holland, P.W. (1982). Observed-score test equating: A mathematical analysis of some ETS equating procederes. In P.W. Holland and D.B. Rubin (Eds.), *Test equating* (pp. 9-49). New York: Academic.
- Cao, L. (2008). Mixed-format test equating: Effects of test dimensionality and common item sets. Unpublished doctoral dissertation, University of Maryland, College Park.

- Childs, R.A., & Chen, W-H. (1999). Obtaining comparable item parameter estimates in MULTILOG and PARSCALE for two polytomous IRT models. *Applied Psychological Measurement*, 23(4), 371-379.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Divgi, D.R. (1981). Two direct procedures for scaling and equating tests with item response theory. Paper presented at the Annual Meeting of American Educational Research Association, Los Angeles.
- Doğan, N. (2009). Çoktan Seçmeli Testler. H. Atılgan (Ed.), *Eğitimde ölçme ve değerlendirme* (s.223-268). Ankara: Anı Yayıncılık.
- de Ayala, R.J. (2009). *The theory and practice of item response theory*: Guilford Press.
- Duong, M.Q. (2011). Evaluating equating results in the non-equivalent groups with anchor test design using equipercentile and equity criteria. Unpublished doctoral dissertation, Michigan State University.
- Embretson, S.E., & Reise, S.P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.
- Ercikan, K., Schwarz, R., Julian, M.W., Burket, G.R., Weber, M.W., & Link, V. (1998). Calibration and scoring of tests with multiple-choice and constructed response test item type. *Journal of Educational Measurement*, 35(2), 137-154.
- Gök, B. (2012). Denk olmayan gruplarda ortak madde deseni kullanılarak madde tepki kuramına dayalı eşitleme yöntemlerinin karşılaştırılması. Yayınlanmamış doktora tezi. Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Gulliksen, H. (1950). *Theory of mental tests*. New York: Wiley.
- Haebara, T. (1980). Equating logistic ability scales by a weighted least squares method. *Japanese Psychological Research*, 22, 144-149.
- Hagge, S.L. (2010). The impact of equating method and format representation of common items on the adequacy of mixed-format test equating using nonequivalent groups. Unpublished doctoral dissertation, University of Iowa.
- Haladyna, T.M. (1997). *Writing test items to evaluate higher order thinking*. Boston: Allyn & Bacon.
- Hanson, B. A., & Beguin, A.A. (2002). Obtaining a common scale for item response theory item parameters using separate versus concurrent estimation in the common-item equating design. *Applied Psychological Measurement*, 26, 3-24.
- Harris, D.J. & Crouse, J.D. (1993). A study of criteria used in equating. *Applied Measurement in Education*, 6, 195-240.
- He, Y., & Kolen, M.J. (2011). Equity and same distributions properties for test equating. CASMA reports.

- He, Y. (2011). *Evaluating equating properties for mixed-format tests*. Unpublished doctoral dissertation, University of Iowa.
- Huggins, A.C. (2012). The effect of differential item functioning on population invariance of item response theory true score equating. Unpublished doctoral dissertation, University of Miami.
- Hu, H., Rogers, W.T., & Vukmirovic, Z. (2008). Investigation of IRT-based equating methods in the presence of outlier common items. *Applied Psychological Measurement*, 32, 311-333.
- Kang, T. (2006). *Model selection methods for unidimensional and multidimensional IRT models*. Unpublished doctoral dissertation, University of Wisconsin-Madison.
- Karasar, N. (2009). *Bilimsel araştırma yöntemi*. Ankara:Nobel Basımevi.
- Kelecioğlu, H. (1994). Öğrenci seçme sınavı puanlarının eşitlenmesi üzerine bir çalışma. Yayınlanmamış doktora tezi. Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Kilmen, S. (2010). Madde tepki kuramına dayalı eşitleme yöntemlerinden kestirilen eşitleme hatalarının örneklem büyüklüğü ve yetenek dağılımına göre karşılaştırılması. Yayınlanmamış doktora tezi, Ankara Üniversitesi, Eğitim Bilimleri Üniversitesi, Ankara.
- Kim, S. (2004). Unidimensional IRT scale linking procedures for mixed-format tests and their robustness to multidimensionality. Unpublished doctoral dissertation, University of Iowa, Iowa City.
- Kim, D., Brennan, R., & Kolen, M.J. (2005). A comparison of IRT equating and beta 4 equating. *Journal of Educational Measurement*, 42(1), 77-99.
- Kim, S., & Kolen, M.J. (2006). Robustness to format effects of IRT linking methods for mixed-format tests, *Applied Measurement in Education*, 19 (4), 357-381.
- Kim, S., & Lee, W. (2004). IRT scale linking methods for mixed-format tests (ACT Research Report 2004-5). Iowa City, IA: ACT, Inc.
- Kim, S., & Lee, W.C. (2006). An extension of four IRT linking methods for mixed format tests. *Journal of Educational Measurement*, 43 (1), 53-76.
- Kim, S., & Walker, M. (2012). Determining the anchor composition for a mixed format test: Evaluation of subpopulation invariance of linking functions. *Applied Measurement in Education*, 25, 178-195.
- Kim, S., Walker, M.E. ve McHale, F. (2010). Comparison among designs for equating mixed-format tests in large-scale assessments. *Journal of Educational Measurement*, 47,1,36-53.
- Kirkpatrick, R.K. (2005). *The effects of item format in common item equating*. Unpublished doctoral dissertation, University of Iowa.

- Koh, K.H., & Zumbo, B.D. (2008). Multi-group confirmatory factor analysis for testing measurement invariance in mixed format item format data. *Journal of Modern Applied Statistical Methods*, 7(2), 471-477.
- Kolen, M.J. (2004). *POLYCSEM* [computer program]. Iowa City, IA: The Center for Advanced Studies in Measurement and Assessment (CASMA), The University of Iowa.
- Kolen, M.J. (2004). *POLYEQUATE* [computer program]. Iowa City, IA: The Center for Advanced Studies in Measurement and Assessment (CASMA), The University of Iowa.
- Kolen, M.J., & Brennan, R.L. (2004). *Test equating: Methods and practices* (2nd ed.). New York, NY: Springer-Verlag.
- Kolen, M.J., & Harris, D.J. (1990). Comparison of item pre-equating using IRT and equipercentile methods. *Journal of Educational Measurement*, 27(1), 27-39.
- Kolen, M.J., & Lee, W.C. (2011). Psychometric properties of raw and scale scores on mixed format tests. *Educational Measurement: Issues and Practice*, 30(2), 15-24.
- Lee, E., Lee, W., & Brennan, R.L. (2012). *Exploring equity properties in equating using AP examinations* (2012-4 ed): College Board.
- Levine, R. (1955). *Equating the score scales of alternate forms administered to samples of different ability* (Research Bulletin 55-23). Princeton, NJ: Educational Testing Service.
- Li, Y.H., & Lissitz, R.W. (2000). An evaluation of multidimensional linking. *Applied Psychological Measurement*, 24, 115-138.
- Li, Y.H., Lissitz, R.W., & Yang, Y.N. (1999). *Estimating IRT equating coefficients for tests with polytomously and dichotomously scored items*. Paper presented at the Annual Meeting of the national Council in measurement in Education, Montreal Canada.
- Livingston, S.A. (2009). Constructed-Response test questions: Why we use them; How we score them (R & D Connections, No. 11). Princeton, NJ: Educational Testing Service. Retrieved from http://www.ets.org/Media/Research/pdf/RD_Connections11.pdf .
- Lord, F.M. (1980). *Applications of item response theory to practical testing problems*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.
- Loyd, B.H., & Hoover, H. D. (1980). Vertical equating using Rasch model. *Journal of Educational Measurement*, 17, 179-193.
- Marco, G.L. (1977). Item characteristic curve solutions to three intractable testing problems. *Journal of Educational Measurement*, 14, 139-160.
- Masters, G.N. (1982). *A Rasch model for partial credit scoring*. *Psychometrika*, 47, 149-174.
- Morris, C.N. (1982). On the foundations of test equating. In P. W. Holland & D. B. Rubin (Eds.) *Test equating* (pp. 169-191). New York: Academic.

- Mullis, I.V.S, & Martin, M.O. (2011). *TIMSS 2011 Item Writing Guidelines*. Boston College.
- Mullis, I. V.S., Martin, M. O., Ruddock, G. J., O'Sullivan, C. Y., & Preuschoff, C. (2009). *TIMSS 2011 Assessment Frameworks*. Chesnut Hill, MA: Boston College.
- Muraki, E. (1992). A generalized partial credit model: Application of an EM algorithm. *Applied Psychological Measurement*, 16 (2), 159-176.
- Muraki, E., & Bock, R. (2003). *PARSCALE 4.1*. Chicago, IL:Scientific Software International.
- Muraki, E., Hombo, C. M., & Lee, Y.-W. (2000). Equating and linking of performance assessment. *Applied Psychological Measurement*, 24, 325-337.
- Ogasawara, H. (2002). Stable response functions with unstable item parameter estimates. *Applied Psychological Measurement*, 26(3), 239-254.
- Öztürk, N. (2010). *Akademik personel ve lisansüstü eğitimi giriş sınavı puanlarının eşitlenmesi üzerine bir çalışma*. Yayınlanmamış yüksek lisans tezi, Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.
- Öztürk, N. & Anıl, D. (2012). Akademik personel ve lisansüstü eğitimi giriş sınavı puanlarının eşitlenmesi üzerine bir çalışma. *Eğitim ve Bilim*, 37(165), 180-193.
- R Development Core Team, 2005. *R: A language and environment for statistical computing, reference index version 2.2.1*. R foundation for Statistical Computing. Vienna: Austria.
- Reckase, M.D. (2009). *Multidimensional item response theory*. New York: Springer.
- Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika Monograph*, No. 17.
- Sinharay, S., & Holland, P. (2006). *The correlation between the scores of a test and an anchor test*. ETS Research Report.
- Stocking, M.L., & Lord, F. M. (1983). Developing a common metric in item response theory. *Applied Psychological Measurement*, 7, 201-210.
- Tankersley, K. (2007). *Tests that teach: Using standardized tests to improve instruction*. An ASCD production.
- Tian, F. (2011). *A comparison of equating/linking using the Stocking-Lord method and concurrent calibration with mixed-format tests in the non-equivalent groups common-item design under IRT*. Unpublished doctoral dissertation, Boston College.
- Tong, Y. & Kolen, M.J. (2005). Assessing equating results on different equating criteria. *Applied Psychological Measurement*, 29, 418-432.
- Traub, R.E. (1993). On the equivalence of traits assessed by multiple-choice and constructed response tests. In R. E. Bennet & W. C. Ward (Eds.), *Construction versus choice in cognitive measurement (pp. 29-44)*. Hillsdale, NJ: Erlbaum.
- von Davier, A.A., Holland, P.W., & Thayer, D.T. (2004). *The kernel method of equating*. New York: Springer.

- Xu, Y. (2009). *Measuring change in jurisdiction achievement over time: Equating issues in current international assessment programs*. Unpublished doctoral dissertation, University of Toronto.
- Wang, W. (2013). *Mixed-format test score equating: Effect of item-type multidimensionality, length and composition of common-item set, and group ability difference*. University of Iowa, Iowa City.
- Wolf, R. (2013). *Assessing the impact of characteristics of the test, common items, and examinees on the preservation of equity properties in mixed-format test equating*. Unpublished doctoral dissertation, University of Pittsburgh.
- Yao, L. (2003). *SimuMIRT*. [Computer Software]. Monterey, CA: Defense Manpower Data Center.
- Yao, L., & Schwarz, R.D. (2006). A multidimensional partial credit model with associated item and test statistics: An application to mixed-format tests. *Applied Psychological Measurement*, 30, 469-492.
- Yen, W. M. (1983). Tau-equivalence and equipercentile equating. *Psychometrika*, 48, 353-369.

EKLER DİZİNİ

EK 1. VERİ TÜRETMEDE KULLANILAN MADDELERİN EVREN PARAMETRE DEĞERLERİ

	Madde Adı	Konu Alanı	Madde Türü	a	b	c	d1	d2	TEST	
									X	Y
1	M052024	N	ÇS	1.422	0.569	0.217			x	x
2	M042016	N	ÇS	0.887	0.523	0.220			x	x
3	M032673	A	ÇS	1.554	0.405	0.216			x	x
4	M042100	A	ÇS	1.403	0.22	0.277			x	x
5	M042150	G	ÇS	0.877	0.886	0.189			x	x
6	M032679	G	ÇS	1.222	0.291	0.227			x	x
7	M042252	DC	ÇS	1.178	0.761	0.162			x	x
8	M042222	DC	ÇS	1.015	0.334	0.124			x	x
9	M042302B	N	AU	0.886	0.457		-0.665	0.665	x	x
10	M032760A	A	AU	0.999	0.590		-1.122	1.122	x	x
11*	M052079	N	ÇS	1.106	0.57	0.326			x	x
12*	M052066	A	ÇS	1.38	0.437	0.232			x	x
13	M042032	N	ÇS	0.858	-0.513	0.241				x
14	M032160	N	ÇS	2.050	1.171	0.133				x
15	M052147	N	ÇS	1.573	0.823	0.258				x
16	M022066	N	ÇS	1.560	0.227	0.116				x
17	M032273	A	ÇS	1.088	-0.167	0.27				x
18	M022050	A	ÇS	1.049	1.022	0.148				x
19	M042234	A	ÇS	1.529	0.316	0.171				x
20	M052130	A	ÇS	1.247	0.89	0.143				x
21	M032205	G	ÇS	0.609	0.23	0.254				x
22	M042137	G	ÇS	1.229	0.634	0.227				x
23	M032100	G	ÇS	1.039	0.322	0.127				x
24	M022101	DC	ÇS	0.895	-0.379	0.155				x
25	M032507	DC	ÇS	1.771	1.048	0.185				x
26	M052161	DC	ÇS	1.071	-0.096	0.199				x
27	M022223B	N	AU	0.934	1.128		-1.462	1.462		x
28	M052002	A	AU	1.002	1.28		-0.582	0.582		x
29	M042300Z	G	AU	0.771	0.267		-0.426	0.426		x
30	M032753B	DC	AU	1.224	0.863		-0.011	0.011		x
31	M042022	N	ÇS	0.995	0.761	0.276				x
32	M032662	N	ÇS	1.975	1.299	0.138				x
33	M052078	N	ÇS	1.118	0.952	0.190				x
34	M052142	N	ÇS	1.07	0.86	0.134				x
35	M042238	A	ÇS	1.317	1.101	0.192				x
36	M042245	A	ÇS	1.736	1.116	0.130				x
37	M052092	A	ÇS	1.161	1.536	0.129				x
38	M052090	A	ÇS	1.265	0.911	0.231				x
39	M052046	G	ÇS	1.155	1.571	0.173				x
40	M032324	G	ÇS	1.185	0.732	0.163				x
41	M032397	G	ÇS	1.195	0.711	0.228				x
42	M052429	DC	ÇS	0.879	0.504	0.152				x
43	M052418B	DC	ÇS	1.725	0.652	0.251				x
44	M052410	DC	ÇS	0.761	0.585	0.291				x
45	M032755	N	AU	1.084	1.143		-0.175	0.175		x
46	M042229Z	A	AU	1.089	0.836		-0.500	0.500		x
47	M032692	G	AU	0.635	0.952		-1.400	1.400		x
48	M032753A	DC	AU	1.093	0.716		-0.222	0.222		x

*11 ve 12. Madde her iki testte de ortak olmasına rağmen sadece ortak test-2 (sadece çoktan seçmeli durumda) 9 ve 10 numaralı açık uçlu maddelerin yerine ortak madde setine alınmıştır.

EK 2. FORM X'İN KALİBRASYONU İÇİN PARSCALE KODU

Unidimensional mixed format test

```
>COMMENT;
>FILE DFNAME='sim1.dat', SAVE;
>SAVE PARM='sim1.PAR';
>INPUT NIDW=4, NTOTAL=30, NTEST=1, LENGTH=30,NFMT=1;
(4A1,30A1)
>TEST TNAME=MIXTEST, ITEM=(1(1)30), NBLOCK=30;
>BLOCK BNAME=DSITEMS, NITEMS=1, NCAT=2,
ORI=(0,1), MOD=(1,2), GPARAM=0.25, GUESS=(2,EST),REP=8;
>BLOCK BNAME=PSITEMS, NITEMS=1, NCAT=3,
ORI=(0,1,2), MOD=(1,2,3), REP=2;
>BLOCK BNAME=DITEMS, NITEMS=1, NCAT=2,
ORI=(0,1), MOD=(1,2), GPARAM=0.25, GUESS=(2,EST),REP=16;
>BLOCK BNAME=PSITEMS, NITEMS=1, NCAT=3,
ORI=(0,1,2), MOD=(1,2,3), REP=4;
>CALIB PAR, LOG, SCALE=1, NQPTS=40, ESTORDER,
CYCLES=(200,10,10,1,1),NEWTON=5,SPRI,GPRI,ITEMFIT=10;
>QUADP POINTS=(-4.0000,-3.7950,-3.5900,-3.3850,-3.1790,
-2.9740,-2.7690,-2.5640,-2.3590,-2.1540,
-1.9490,-1.7440,-1.5380,-1.3330,-1.1280,
-0.9231,-0.7179,-0.5128,-0.3077,-0.1026,
0.1026,0.3077,0.5128,0.7179,0.9231,
1.1280,1.3330,1.5380,1.7440,1.9490,
2.1540,2.3590,2.5640,2.7690,2.9740,
3.1790,3.3850,3.5900,3.7950,4.0000);
```


EK 3. ETİK KURUL İZİN MUAFİYETİ FORMU

	Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü	Form: 40
Tez Çalışması Etik Kurul İzin Muafiyeti Formu		
11 / 08 / 2014		
Hacettepe Üniversitesi Eğitim Bilimleri Enstitüsü Eğitim Bilimleri Anabilim Dalı Başkanlığı'na		
Tez Başlığı / Konusu:	Testlerin Boyutluluğunun, Ortak Madde Formatının, Yetenek Dağılımının ve Ölçek Dönüştürme Yöntemlerinin Karma Testlerin Eşitlenmesine Etkisi / Test Eşitleme	
Yukarıda başlığı/konusu gösterilen tez çalışmam:		
<ol style="list-style-type: none">1. İnsan ve hayvan üzerinde deney niteliği taşımamaktadır,2. Biyolojik materyal (kan, idrar vb. biyolojik sıvılar ve numuneler) kullanılmasını gerektirmemektedir.3. Beden bütünlüğüne müdahale içermemektedir.4. Gözlemsel ve betimsel araştırma (anket, ölçek/skala çalışmaları, dosya taramaları, veri kaynakları taraması, sistem-model geliştirme çalışmaları) niteliğinde değildir.		
Hacettepe Üniversitesi Etik Kurullar ve Komisyonlarının Yönergelerini inceledim ve bunlara göre tez çalışmamın yürütülebilmesi için herhangi bir Etik Kuruldan izin alınmasına gerek olmadığını; aksi durumda doğabilecek her türlü hukuki sorumluluğu kabul ettiğimi ve yukarıda vermiş olduğum bilgilerin doğru olduğunu beyan ederim.		
Gereğini saygılarımla arz ederim.		
 Neşe ÖZTÜRK GÜBEŞ (Öğrencinin Adı Soyadı, İmzası)		
Öğrenci Bilgileri		
Adı Soyadı	Neşe ÖZTÜRK GÜBEŞ	
Öğrenci No	N10144538	
Anabilim Dalı	Eğitim Bilimleri	
Programı	Eğitimde Ölçme ve Değerlendirme	
Statüsü	<input type="checkbox"/> Yüksek Lisans <input checked="" type="checkbox"/> Doktora <input type="checkbox"/> Bütünleşik Dr.	
Danışman Görüşü ve Onayı		
Etik kurul izni gerekmemektedir		
Prof. Dr. Hülya KELECİOĞLU (İmza) (Danışmanın Ünvanı, Adı ve Soyadı) 		
Eğitim Bilimleri Enstitüsü: e-Ağ: http://ebe.hacettepe.edu.tr/ Tel: 0(312) 297-85 72 Belgegeçer: 0(312) 297-85 66 e-Posta: ebe@hacettepe.edu.tr		

EK 4. ORJİNALLİK RAPORU

The screenshot displays the iThenticate Professional Plagiarism Prevention web interface. At the top, there are navigation tabs for 'Folders', 'Settings', and 'Account Info'. The iThenticate logo is prominently displayed. Below the logo, there is a search bar and a 'Trash' button. The main content area is divided into two sections: 'My Folders' on the left and 'My Documents' on the right. The 'My Documents' section shows a list of documents with columns for 'Title', 'Report', 'Author', 'Processed', and 'Actions'. A single document is listed with a 10% match rate, authored by Neşe Öztürk Gübeş, and processed on August 12, 2014, at 5:16:48 PM EEST. The document title is 'Testlerin Boyutluluğunun, Ortak Madde Formatının, Yetenek Dağılımının ve Ölçek Dönüştürme Yöntemlerinin Karma Testlerin Eşitlenmesine Etkisi'. The document consists of 1 part with 27,538 words. The interface also includes a 'page 1 of 1' indicator in the bottom right corner.

iThenticate®
Professional Plagiarism Prevention

My Folders
My Documents
Trash

My Documents

Title	Report	Author	Processed	Actions
Testlerin Boyutluluğunun, Ortak Madde Formatının, Yetenek Dağılımının ve Ölçek Dönüştürme Yöntemlerinin Karma Testlerin Eşitlenmesine Etkisi 1 part - 27,538 words	10%	Neşe Öztürk Gübeş	August 12, 2014 5:16:48 PM EEST	

ÖZGEÇMİŞ

Adı Soyadı	Neşe Öztürk Gübeş
Doğum Yeri	Antalya
Doğum Yılı	1985
Medeni Hali	Evli

Eğitim ve Akademik Durumu

Lise	Korkuteli Anadolu Lisesi, Antalya	1997-2004
Lisans	Hacettepe Üniversitesi, Eğitim Fakültesi, Fen Bilgisi Öğretmenliği	2004-2008
Yüksek Lisans	Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı	2008-2010
Doktora	Hacettepe Üniversitesi, Eğitim Bilimleri Enstitüsü, Eğitim Bilimleri Anabilim Dalı, Eğitimde Ölçme ve Değerlendirme Bilim Dalı	2010-2014
Yabancı Dil	İngilizce	
İş Deneyimi	Mehmet Akif Ersoy Üniversitesi, Eğitim Fakültesi, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı, Araştırma Görevlisi	2009-(devam ediyor)