

AKDENİZ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

EĞİTİM BİLİMLERİ ANA BİLİM DALI EĞİTİMDE ÖLÇME VE DEĞERLENDİRME
TEZLİ YÜKSEK LİSANS PROGRAMI

ÇOK KATEGORİLİ PUANLANAN MADDELERDEN OLUŞAN TESTLERDE KLASİK
TEST KURAMI VE MADDE TEPKİ KURAMI'NA DAYALI TEST EŞİTLEME
YÖNTEMLERİNİN KARŞILAŞTIRILMASI

YÜKSEK LİSANS TEZİ

MERVE ÇÖRTÜK

ANTALYA, 2022

AKDENİZ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ

EĞİTİM BİLİMLERİ ANA BİLİM DALI EĞİTİMDE ÖLÇME VE DEĞERLENDİRME
TEZLİ YÜKSEK LİSANS PROGRAMI

ÇOK KATEGORİLİ PUANLANAN MADDELERDEN OLUŞAN TESTLERDE KLASİK
TEST KURAMI VE MADDE TEPKİ KURAMI'NA DAYALI TEST EŞİTLEME
YÖNTEMLERİNİN KARŞILAŞTIRILMASI

YÜKSEK LİSANS TEZİ

MERVE ÇÖRTÜK

DANIŞMAN: DOÇ. DR. ALPER SİNAN

ANTALYA, 2022

ONAY

T.C.

AKDENİZ ÜNİVERSİTESİ

EĞİTİM BİLİMLERİ ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Merve Çörtük ‘ün bu çalışması 21.06.2022 tarihinde jürimiz tarafından Eğitim Bilimleri Ana Bilim Dalı Eğitimde Ölçme ve Değerlendirme Tezli Yüksek Lisans Programında Yüksek Lisans Tezi olarak oy birliği/oy çokluğu ile kabul edilmiştir.

İMZA

Başkan:

.....

Üye:

.....

Üye (Danışman):

.....

YÜKSEK LİSANS TEZİNİN ADI: Çok Kategorili Puanlanan Maddelerden Oluşan Testlerde Klasik Test Kuramı ve Madde Tepki Kuramı’na Dayalı Test Eşitleme Yöntemlerinin Karşılaştırılması

ONAY: Bu tez, Enstitü Yönetim Kurulunca belirlenen yukarıdaki jüri üyeleri tarafından uygun görülmüş ve Enstitü Yönetim Kurulunun tarihli ve sayılı kararıyla kabul edilmiştir.

DOĞRULUK BEYANI

Yüksek lisans / Doktora tezi / Dönem Projesi olarak sunduğum bu çalışmayı, bilimsel ahlak ve geleneklere aykırı düşecek bir yol ve yardıma başvurmaksızın yazdığımı, yararlandığım eserlerin kaynakçalardan gösterilenlerden oluştuğunu ve bu eserleri her kullanışmada alıntı yaparak yararlandığımı belirtir; bunu onurumla doğrularım. Enstitü tarafından belli bir zamana bağlı olmaksızın, tezimle ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçlara katlanacağımı bildiririm.

28 / 06 / 2022

Merve Çörtük



İÇİNDEKİLER

TABLolar DİZİNİ.....	iii
ŞEKİLLER DİZİNİ.....	iv
KISALTMALAR VE SİMGELER	v
TEŞEKKÜR	vi
ÖZET	vii
ABSTRACT	viii

BÖLÜM I

GİRİŞ

1.1. Araştırmanın Problemi.....	1
1.2. Araştırmanın Amacı.....	3
1.3. Araştırmanın Önemi	4
1.4. Varsayımlar.....	5
1.5. Sınırlılıklar	5

BÖLÜM II

KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

2.1. Klasik Test Kuramı	6
2.2. Madde Tepki Kuramı.....	7
2.2.1 İkili Puanlanan Maddeler için MTK Modelleri	8
2.2.2 Çoklu Puanlanan Maddeler için MTK Modelleri	9
2.3. Test Eşitleme.....	11
2.3.1. Test Eşitleme Türleri.....	11
2.3.2. Test Eşitleme Adımları	11
2.3.3. Test Eşitleme Koşulları.....	12
2.3.4. Test Eşitleme Desenleri	13
2.3.5. Test Eşitleme Yöntemleri	16

2.3.5.1. Geleneksel Eşitleme Yöntemleri.....	17
2.3.5.2. Madde Tepki Kuramı'na Dayalı Test Eşitleme Yöntemleri	25
2.3.6. Test Eşitlemede Hata Kavramı	29
2.4. İlgili Araştırmalar	29
2.4.1. Türkiye’de Yapılan Çalışmalar.....	29
2.4.2. Türkiye Dışında Yapılan Çalışmalar	31

BÖLÜM III

YÖNTEM

3.1. Araştırmanın Modeli.....	33
3.2. Simülasyon Koşulları.....	33
3.2.1. Değişimlenen Koşullar.....	33
3.2.2. Sabit Tutulan Koşullar	34
3.3. Verilerin Üretilmesi	35
3.4. Verilerin Analizi	36

BÖLÜM IV

BULGULAR

4.1. 10 Maddeye İlişkin Bulgular	37
4.1. 20 Maddeye İlişkin Bulgular	38
4.1. 30 Maddeye İlişkin Bulgular	40
4.1. 40 Maddeye İlişkin Bulgular	41
4.1. 50 Maddeye İlişkin Bulgular	43

BÖLÜM V

SONUÇ, TARTIŞMA VE ÖNERİLER

5.1. Sonuç ve Tartışma	45
5.2. Öneriler	48
KAYNAKÇA.....	50
ÖZGEÇMİŞ.....	56
İNTİHAL RAPORU.....	57

TABLÖLAR DİZİNİ

Tablo 3.1. Sabit Tutulan Koşullar	35
Tablo 3.2. Değişimlenen Koşullar	35
Tablo 4.1. 10 Madde için Hata Miktarları	35
Tablo 4.2. 20 Madde için Hata Miktarları	38
Tablo 4.3. 30 Madde için Hata Miktarları	40
Tablo 4.4. 40 Madde için Hata Miktarları	42
Tablo 4.5. 50 Madde için Hata Miktarları	43



ŞEKİLLER DİZİNİ

Şekil 1.1. Test Eşitleme Yöntemleri.....	2
Şekil 2.1. Madde Karakteristik Eğrisi	7
Şekil 2.2. Test Eşitleme Türleri	14
Şekil 2.3. Rastgele Eşdeğer Gruplar	14
Şekil 2.4. Dengelenmiş Tek Grup	15
Şekil 2.5. Ortak Madde Eşdeğer Olmayan Gruplar.....	15
Şekil 2.6. Tek Grup Deseni	16
Şekil 2.7. Daire Yayı Kullanılarak Dönüştürülmüş Noktalar.....	25
Şekil 4.1. 10 Madde İçin Hata Miktarları.....	38
Şekil 4.2. 20 Madde İçin Hata Miktarları.....	39
Şekil 4.3. 30 Madde İçin Hata Miktarları.....	41
Şekil 4.4. 40 Madde İçin Hata Miktarları.....	42
Şekil 4.5. 50 Madde İçin Hata Miktarları.....	44

KISALTMALAR VE SİMGELER

GRM: Graded Response Model

GPCM: Genelleştirilmiş Kısmi Kredi Modeli

KTK: Klasik Test Kuramı

MKE: Madde Karakteristik Eğrisi

MTK: Madde Tepki Kuramı

NRM: Nominal Response Model

PCM: Partial Credit Model

RSM: Rating Scale Model

1PL: 1 Parametreli Lojistik Model

2PL: 2 Parametreli Lojistik Model

3PL: 3 Parametreli Lojistik Model

TEŞEKKÜR

Yüksek lisans eğitimim boyunca benden yardımlarını ve birikimlerini esirgemeyen, desteğini her zaman hissettiğim değerli danışman hocam Doç. Dr. Alper Sinan’a bana ve tezime sunduğu katkılar için teşekkür ederim.

Madde Tepki Kuramı konusunda ilgimi hoşgörüyü karşılayan, tüm sorularıma sabırla cevap veren ve bana her zaman destek olan değerli hocam Doç. Dr. Hakan Koğar’a teşekkür ederim. R ve tez çalışmam konusunda görüş ve yardımlarını aldığım değerli hocam Doç. Dr. Bilal Barış Alkan’a destekleri için teşekkür ederim. Tavsiye ve görüşleri ile tezime katkı sunan değerli hocalarım Prof. Dr. Bayram Bıçak ve Dr. Öğr. Üyesi Aydın Karakoca’ya teşekkür ederim.

Eğitimim için maddi manevi hiçbir fedakarlıktan kaçınmayan sevgili annem Züleyha Çörtük ve babam Aydın Çörtük’e, bana ikinci bir aile olan ablam Fadime Kılıç ve eşi Yusuf Kılıç’a, bilgisayar başında geçen günlerimde benden desteklerini esirgemeyen kardeşlerim Emine Çörtük ve Berra Çörtük’e teşekkür ederim. Son olarak, şu an bulunduğum noktaya gelmemde emeği olan ancak burada isimlerini anmadığım tüm saygıdeğer hocalarıma ve bana her zaman destek olan değerli dostlarıma çok teşekkür ederim.

ÖZET

Çok Kategorili Puanlanan Maddelerden Oluşan Testlerde Klasik Test Kuramı ve Madde Tepki Kuramı'na Dayalı Test Eşitleme Yöntemlerinin Karşılaştırılması

Çörtük, Merve

Yüksek Lisans Tezi/ Eğitim Bilimleri Anabilim Dalı

Tez Danışmanı: Doç. Dr. Alper Sinan

Haziran, 2022, 55

Test eşitleme, aynı örtük özelliği ölçen ve benzer güçlük dercesine sahip iki formun birbirinin yerine kullanılabilmesi için yürütülen istatistiksel süreçtir. Bu çalışma, çok kategorili puanlanan maddelerden oluşan testlerde Klasik Test Kuramı (KTK) ve Madde Tepki Kuramı'na (MTK) dayalı test eşitleme yöntemlerinin karşılaştırılmasını amaçlayan bir olasılıksal simülasyon çalışmasıdır. Üç KTK yönteminden (doğrusal, ortalama ve eşit yüzdelikli eşitleme) ve üç MTK yönteminden (ortalama-ortalama, ortalama-standart sapma ve test karakteristik eğrisi) elde edilen hata miktarlarının, örneklem büyüklüğü (50, 100, 300, 1000 ve 3000 kişi) ve test uzunluğu (10, 20, 30, 40 ve 50 madde) değişkenlerine dayalı olarak karşılaştırılması amaçlanmıştır. Çalışmada, 0-1-2 şeklinde puanlanan genelleştirilmiş kısmi kredi modeline uyumlu simülatif veriler kullanılmıştır. Çalışmanın verilerini üretmede WinGen3 programı kullanılmıştır. Bu veri setlerini eşitleme için R programından faydalanılmıştır.

Araştırma sonunda, genelleştirilmiş kısmi kredi modeline uygunluk gösteren testler için; farklı örneklem büyüklüğü ve test uzunluğuna sahip testlerin eşitlenmesinde en az hata ile eşitleme yapan yöntemin eşit yüzdelikli eşitleme olduğu saptanmıştır. Bu bulgu literatürle de desteklenmektedir. Araştırmada en yüksek eşitleme hatalarını veren yöntemler ise ortalama-standart sapma ve test karakteristik eğrisi yöntemleridir. Örneklem büyüklüğüne dayalı olarak test eşitleme hataları incelendiğinde, örneklem büyüklüğü arttıkça test eşitleme hatasının azaldığı saptanmıştır. Test uzunluğu değişkenine göre test eşitleme yöntemleri karşılaştırıldığında ise, madde sayısı arttıkça hesaplanan hata miktarının da arttığı bulunmuştur.

Anahtar kelimeler: test eşitleme, Madde Tepki Kuramı, Klasik Test Kuramı, eşitleme hatası, çok kategorili puanlanan testler.

ABSTRACT

Comparison of Test Equating Methods Based on Classical Test Theory and Item Response Theory in Polytomously Scored Tests

Çörtük, Merve

Master Degree / Department of Educational Sciences

Supervisor: Assoc. Prof. Dr. Alper Sinan

June, 2022, 55

Test equating is a statistical process that is carried out to use two forms interchangeably which measure same latent trait and have similar difficulty levels. This work is a stochastic simulation study aiming to compare test equating methods based on Classical Test Theory (CTT) and Item Response Theory (IRT) in tests consisting of polytomously scored items. It was aimed to compare the error, the sample sizes (50, 100, 300, 1000 and 3000 people) and the test length (10, 20, 30, 40 and 50 items) that were obtained from three CTT methods (linear, mean and equipercentile equating) and three IRT methods (mean-mean, mean-sigma and test characteristic curve). In this study, simulative data compatible with Generalized Partial Credit Model (GPCM) scored as 0-1-2 were used. In producing the data of the study, WinGen3 programme was used, to Equate this data sets, it was utilized from the R programme.

At the end of the research, it was determined that the method which equates with the least error in equating test with different sample sizes and test lengths is equipercentile equating for test conformity to Generalized Partial Credit Model. This finding is also supported by the literature. The methods that give the highest equating errors in the research are the mean-sigma and test characteristic curve methods. When the test errors were analyzed based on sample size, it was found that as the sample sized increased, the error decreased. When the test equating methods were compared according to the test lenght variable, it was found that as the amount of item increased, the calculated error also increased.

Key words: test equating, Item Response Theory, Classical Test Theory, equating error, polytomously scored tests.

BÖLÜM I

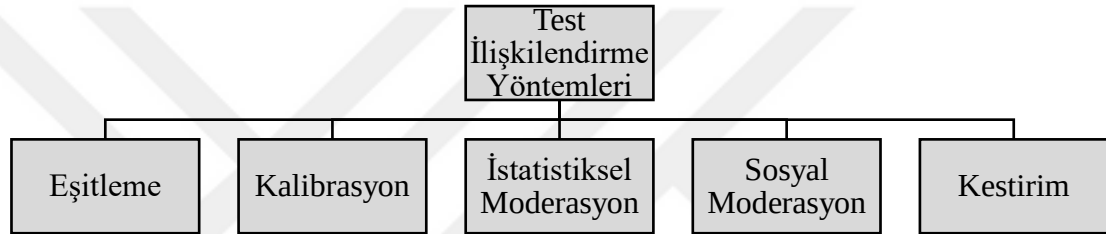
GİRİŞ

Bu bölümde, araştırmaya ait problem, amaç, önem, varsayımlar ve sınırlılıklar başlıkları yer almaktadır.

1.1. Araştırmanın Problemi

Eğitimde ve psikolojide bireyler ve toplum hakkında karar verebilmek amacıyla ölçme ve değerlendirme süreci yürütülür. Bu sürecin ölçme kısmı ölçülmesi istenen özellik genellikle gizil bir özellik olduğu için çeşitli testler aracılığıyla yapılır. Özellikle eğitim bilimlerinde testler önemli bir yere sahiptir. Okullarda hemen hemen her tür mevcut test kullanılmaktadır. Zekâ, özel yetenek, çoklu yetenek ve kişilik testleri bunlardan sadece birkaçıdır. Öğretmenler ve eğitim yöneticileri sıklıkla birkaç farklı test türüyle elde edilen sonuçlara göre hareket etmektedir. Bu testlerin yanısıra belirli bir eğitim veya öğretim programının etkilerini ölçmek için de başarı testleri kullanılmaktadır (Anastasi, 1976). Sadece okullar değil ülke genelinde de her yıl onlarca sınav uygulanmaktadır. Lise ve üniversite gibi eğitim kademelerine geçiş için yapılan sınavlar, bursluluk sınavları, meslek hayatına geçişte kurumların yaptığı sınavlar bunların sadece bir kısmını oluşturmaktadır. Yapılan bu sınavların sonuçlarıyla önemli kararlar alınmaktadır. Dolayısıyla da sonuçların geçerli ve güvenilir olması istenmektedir. Sonuçların güvenilirliğini etkileyen en önemli faktörlerden birisi bu sınavlarda yer alan maddelerin yalnızca bir kez kullanılıyor oluşu ve yılda bir ya da daha fazla yapılması halinde farklı test formlarının yer almasıdır. Üniversitelerin hazırlık sınavları gibi bazı sınavları farklı bireylerin farklı tarihlerde alması ya da üniversite sınavı, ALES (Akademik Personel ve Lisansüstü Eğitim Giriş Sınavı), YDS (Yabancı Dil Sınavı) gibi sınavlara bireylerin yıl içerisinde birden fazla girme hakkının olması aynı test formunun kullanılmasını imkânsız hale getirmektedir. Bu duruma bağlı olarak da farklı test formları kullanılmaktadır. Aynı testin farklı formlarını alan bireylerden birinin diğerine karşı avantaj ya da dezavantaj sahibi olmaması için bu formların benzer olması gerekmektedir. Formların benzerliğinin kontrol edilmesi test ilişkilendirme işlemi ile mümkün olmaktadır (Kolen & Brennan, 2014).

von Davier'e (2011) göre ilişkilendirme kavramı üç farklı anlamda kullanılmaktadır: Farklı test formları arasında toplam puan ya da madde parametrelerine dayalı olarak bağlantı belirleme, eşitleme işleminin daha zayıf bir türü ve MTK (Madde Tepki Kuramı) kalibrasyonu olarak da adlandırılabilen iki testin puanlarını aynı ölçeğe yerleştirme işlemidir. Mislevy (1992) ve Linn (1993) ilişkilendirmeyi bir testin sonucun diğer testlerle karşılaştırabilmek için çeşitli yaklaşımların uygulanmasını içeren genel bir terim olarak kullanmışlardır. Bu yaklaşımları beş başlık altında toplamışlardır: eşitleme, kalibrasyon, istatistiksel moderasyon, kestirim ve sosyal moderasyon.



Şekil 1.1. Test Eşitleme Yöntemleri

Eşitleme, bir testin diğer bir test yerine kullanılabilmesini sağlayan ilişkilendirme yöntemidir. Diğer ilişkilendirme yöntemlerine göre daha zor olmasına rağmen daha iyi anlaşılması ve en güçlü ilişkilendirme yöntemi olması avantaj sağlamaktadır. Varsayımların sağlandığı durumda, eşitlenmiş iki test birbirinin yerine kullanılabilir. Bir başka deyişle, Form X için yapılan yorum Form Y için de geçerlidir. Örneğin, bir üniversiteye yerleşmek için her yıl sınav yapılmaktadır ve bu sınavların birbirine eşdeğer olması istenir. Kalibrasyon, bir testin daha kısa ve dolayısıyla daha az güvenilir olduğu düşünülebilen bir formunun daha uzun bir formu ile karşılaştırılması için kullanılan bir ilişkilendirme yöntemidir. İstatistiksel moderasyon, farklı kaynaklar tarafından oluşturulan ya da farklı özellikleri ölçen testlerin karşılaştırılmasını sağlayan bir ilişkilendirme yöntemidir. Karşılaştırma yapabilmek için çoğunlukla ankor test kullanılır. Örneğin, bir üniversite öğrencilerinin KPSS eğitim bilimleri puanları ile bu sınavda yer alan derslerde yapılan sınavların puanlarını karşılaştırmak istediğinde bu yöntemi kullanabilir. Kestirim, ilişkilendirme yöntemleri arasında en zayıf olan yöntemdir. Bir formdan elde edilen puanlar ile başka bir formun puanlarının tahmin edilmesidir. Gruba ve zamana bağlı olması olumsuz bir özelliğidir. Örneğin sadece erkeklerin bulunduğu

bir gruptan elde edilen puanlar sadece kızların olduđu bir grup için geçerli olmayacaktır. Bu olumsuz özellik nedeniyle de testlerin sık sık tekrar edilmesi gerekmektedir. Kestirim yöntemi, iki test arasında güçlü bir ilişki olduđu zaman doğru sonuçlar verebilen bir yöntemdir. Sosyal moderasyon, formlardan farklı kişiler tarafından ve standart bir puanlama düzeniyle elde edilen puanların karşılaştırılmasıdır (Kolen & Brennan, 2014).

Test puanlarından elde edilen bilgi önemli kararlar almak için kullanılabilir. Bu kararlar hangi okula gideceğine dair bireysel; üniversite, askeriye gibi kurumların kişi yaptığı kuruma ait; ya da ülkenin eğitim seviyesinin belirlendiği toplumsal kararlar olabilir. Daha doğru bilgiye ulaşmak için eşitleme yapılırsa daha iyi kararlar alınabilir. Uygulanan test için farklı seçenekler de olabilir. Kişilere farklı günler verilerek test birkaç güne yayılabilir, eğitim testleri yıllarca yapılabilir. Aynı test için kişiler farklı bilgilere sahip olabilir. Bu güvenlik sorununu çözmek için aynı içerik ve istatistiksel özelliklere sahip farklı test formları geliştirilir. Farklı test formlarının kullanılması, formların farklı güçlülere sahip olabileceği endişesini beraberinde getirir. Eşitleme, farklı test formlarının birbirlerinin yerine kullanılabilmesi için yürütülen istatistiksel bir süreçtir. Bir kişi aynı testi iki kez alabilir. Bu durumda ikincisinde aldığı yüksek puan başarının bir göstergesi mi yoksa aynı teste maruz kaldığı için mi karar vermek adına farklı test formları kullanılır. Farklı test formları için de bu problem söz konusudur. Test formlarının kullanılmasıyla ilgili bir diğer problem ise örneğin iki farklı öğrenci iki farklı test formu aldığıında, kolay olan formu alan öğrencinin daha yüksek puana sahip olmasıdır. Bu problemin çözümü de test eşitleme sürecinin yürütülmesidir (Kolen & Brennan, 2014).

1.2. Araştırmanın Amacı

Bu araştırmanın genel amacı, farklı test uzunluğu ve örneklem büyüklüğü koşulları altında, çeşitli test eşitleme yöntemlerini (ortalama, eşit yüzdelikli, doğrusal, ortalama-ortalama, ortalama-standart sapma, Haebara ve Stocking-Lord) eşitleme hatalarına göre karşılaştırmak ve bu yöntemlerden en uygun olanını önermektir.

Bu genel amaç doğrultusunda farklı test uzunluğuna (10, 20, 30, 40 ve 50 madde) sahip testlerde ve çeşitli örneklem büyüklüğüne (50, 100, 300, 1000 ve 3000) sahip gruplarda, farklı test formlarının eşitlenmesinde en az hata veren yönteminin hangisi olduğu sorusuna cevap aranmıştır.

1.3. Araştırmanın Önemi

Yapılan sınavların çoğunda güvenlik gibi sebeplerden dolayı aynı sorular tekrar sorulmamaktadır. Bunun yerine aynı testin farklı formları kullanılmaktadır. Aynı amaç için sınava giren bireyler farklı formlardan aldıkları puanları kullanabilmektedir. Örneğin, akademide bir kadro açıldığında bilim sınavına girmeye hak kazanan bireyler ALES ve dil puanlarıyla belirlenmektedir. ALES, YDS gibi sınavlar yılda birden fazla yapılan dolayısıyla farklı formları kullanılan sınavlardır. Bu da sonucun bu kadar önemli olduğu bir durumda adaletin sağlanıp sağlanamadığı şüphesini beraberinde getirmektedir. Üniversitelerin hazırlık sınavları ilk ders döneminin başında ve verilen eğitim sonunda yapılmaktadır. Çoğu üniversite belirli bir geçme notu belirlemektedir. Örneğin geçme notu 70 olan bir hazırlık programı için aynı bilgi birikimine sahip iki kişiden birisi dönem başındaki daha kolay olan sınavdan 70 üzeri bir not alarak geçebilirken diğer kişi daha zor olan dönem sonu ya da sonraki yıla ait sınavdan 70'in altında bir not alarak kalabilir. Bu gibi durumların birçok benzer örneği mevcuttur. Aynı testin farklı formlarının birbirlerinin yerine kullanılabilmesi için test eşitleme işlemi ile mümkün olmaktadır.

Test eşitleme işleminin yapılmasının genel olarak üç nedeni vardır. İlk ve en önemli nedeni test puanı elde edebilmek için aynı testin kullanılmasının güvenlik sorunlarına yol açmasıdır. Test eşitlemeyle birlikte aynı amaç için benzer formlar geliştirilebilir. Böylelikle sınava farklı zamanlarda giren bireylere adil davranılmış olunur. İkinci neden test puanlarının değişebilir olmasıdır. Testi alan bireylerin karşılaştırılabilmesi için testlerin eşitlenmesi gereklidir. Üçüncü neden ise testlerin sürekliliğinin sağlanmasıdır. Herhangi bir bilgi ya da yetenek için tasarlanan testler farklı düzeyler için farklı zorluklarda hazırlanabilir. Aynı özelliği ölçen ancak farklı düzeylere hitap eden bu testlerin istatistiksel olarak eşitliği dikey eşitleme ile kontrol edilir (Zhu, 1998).

Test eşitleme çalışmalarında çeşitli yöntemler kullanılmaktadır. Bu çalışmada test uzunluğu ve örneklem büyüklüğü değişken olarak belirlenmiştir. Araştırmanın amacı, farklı test uzunluğu ve örneklem büyüklüğüne sahip gruplarda en az hata veren eşitleme yöntemini belirleyebilmektir. Bu çalışma, gelecek araştırmalar için kuramsal katkı sağlamak ve test eşitlemenin önemini vurgulamak amacıyla yürütülmüştür.

1.4. Varsayımlar

Simülasyon çalışmaları, değişkenlerin kontrol edilebilmesi açısından değerli çalışmalardır. Ancak verilerin gerçeğe yakın üretilmediği durumlarda genelleme yapabilmek pek mümkün değildir. Bu nedenle bu çalışmada, üretilen verilerin gerçek durumları yansıttığı ve dolayısıyla genellenebileceği varsayılmaktadır.

1.5. Sınırlılıklar

Bu çalışma KTK'ya (Klasik Test Kuramı) dayalı olarak ortalama, eşit yüzdelikli ve doğrusal; MTK'ye dayalı olarak ortalama-ortalama, ortalama-standart sapma, Haebara ve Stocking-Lord yöntemleri ile sınırlı tutulmuştur. Veriler MTK temelinde çoklu puanlanan maddeler için kullanılan genelleştirilmiş kısmi puan modeli ile üretilmiştir. Çoklu puanlanan maddeler için kullanılan diğer modeller ve 1-0 olarak puanlanan maddeler için kullanılan modeller çalışmanın dışında tutulmuştur. Simülasyonda değişen faktörler olarak test uzunluğu ve örneklem büyüklüğü ele alınmıştır. Test uzunluğu 10, 20, 30, 40 ve 50 madde ile; örneklem büyüklüğü 50, 100, 300, 1000 ve 3000 kişi ile sınırlandırılmıştır.

BÖLÜM II

KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

Bu bölümde; test eşitlemenin varsayımlarına ve test eşitleme yöntemlerine değinilmiştir. Ayrıca test eşitlemeyle ilgili yapılan bazı çalışmalara yer verilmiştir.

2.1. Klasik Test Kuramı

Klasik Test Kuramı (KTK) eğitim ve psikolojide ölçülen yapıların açıklanması konusunda yardımcı olan bir ölçme teorisidir. Ölçmede, ölçülmek istenen davranış her zaman doğrudan gözlemlenmeyebilir. Bu durumda bir ölçme aracı kullanarak bu davranışı bireyin ne kadar kazandığına karar vermek istenir. Ancak ölçmeye her zaman hata karışacağı için bireyin sahip olduğu gerçek puana ulaşmak oldukça güçtür. Bu yüzden gerçek puana dair kestirimler yapılır. Bu kestirimler çeşitli ölçme kuramları kullanılarak elde edilir, bunlardan biri de KTK'dir. KTK, gözlenen puanlar ve hataların hesaplanması aracılığıyla gerçek puanın kestirilmesi sağlar (Baykul, 2015). KTK'de kullanılan denklem X gerçek puan, T gözlenen puan ve ϵ hata olmak üzere eşitlik 2.1'de verilmiştir:

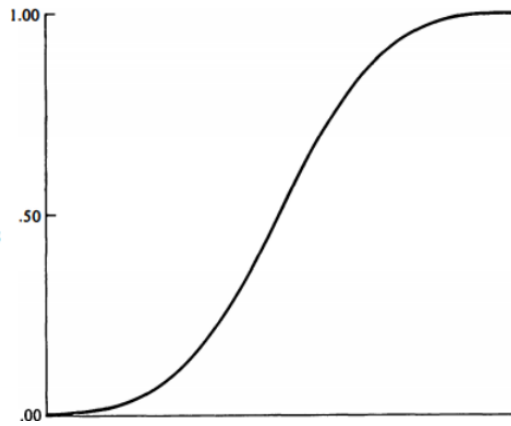
$$X = T + \epsilon \quad (2.1)$$

KTK'nin birtakım sınırlılıkları vardır. Gözlenen ve gerçek puanın teste bağımlı olması bu sınırlılıkların başında gelmektedir. Ölçmenin standart hatasının tüm test üzerinden hesaplanarak sınava giren tüm bir bireyler için sabit olduğunun varsayılması, güvenilirliği arttırmak için madde sayısını artırma yoluna başvurulması gerekliliği, test eşitleme çalışmaları için testlerin paralel formlarda olmasına ihtiyaç duyulması, madde parametrelerinin yanlı olması, ölçek puanlarının ancak normal dağılım sağlandığı zaman eşit aralık düzeyinde olabilmesi, farklı başlangıç seviyesine sahip grupları karşılaştırmanın anlamlı olmayışı ve ikili puanlanan maddelere uygulanan faktör analizinin faktörler yerine yapay bir yapıyı ortaya çıkarması KTK'nin sözü edilen sınırlılıkları içerisinde yer alabilmektedir (Koğar, 2020).

2.2. Madde Tepki Kuramı

Madde Tepki Kuramı bir ölçme aracı tarafından ölçülen yetenek ya da özellik ile bir maddeye verilen cevap arasında ilişkiyi göstermek için kullanılan bir ölçme kuramıdır. Madde Tepki Kuramı modelleri bireyin maddeye doğru yanıt verme olasılığını, ölçülen yeteneğe bağlayan matematiksel bir fonksiyondur. Yani diğer bir deyişle MTK modelleri maddeye doğru yanıt verme olasılığı ile yetenek arasındaki ilişkinin doğrusal olmayan regresyonudur (Sünbül, 2020).

Madde puanının yetenek üzerindeki regresyonu madde karakteristik eğrisi (MKE) olarak adlandırılır. MKE Madde Tepki Kuramı'nın temel yapı taşı niteliğindedir. Eğri, maddeye doğru yanıt verme ve yetenek arasındaki ilişkinin gösterilmesine olanak sağlar. Maddeye doğru yanıt verme olasılığının yorumlanmasında çeşitli görüşler vardır. Bunlardan en yaygın olanı ve literatürde genellikle yer alan örtük özellik ölçeğinin her bir noktasında yer alan bireylerin sınava giren tüm bireylerin bir alt popülasyonu ve homojen olduğudur. Bu herhangi bir yetenek düzeyinde doğru yanıtlama olasılığını aynı zamanda bir doğru yanıtlama olasılığının da belirli bir yetenek düzeyini ifade etmesi olarak yorumlanabilir. Örneğin, $\theta=1.5$ yetenek düzeyindeki bir bireyin doğru yanıtlama olasılığı .5 ise bu doğru yanıtlama olasılığı .5 olan herhangi bir bireyin aynı zamanda $\theta=1.5$ yetenek düzeyinde olması olarak da ifade edilebilir (Crocker & Algina, 2006b). Yatay eksen θ (örtük özellik) ve dikey eksen bireylerin doğru cevap verme olasılığı olarak tanımlanan MKE şekil 2.1'de verilmiştir.



Kaynak. (Crocker & Algina, 2006b).

Şekil 2.1. Madde Karakteristik Eğrisi

MTK temelde iki varsayıma sahiptir: tek boyutluluk ve yerel bağımsızlık. Tek boyutluluk testin tek bir örtük özelliği ölçmesi olarak tanımlanabilir. Bu varsayımın katı bir

şekilde karşılanması mümkün değildir çünkü testi alan bireyin bilişsel ya da kişisel özellikleri test performansını etkileyecektir. Bu nedenle tek bir örtük özellik yerine baskın bir özellik aranır (Hambleton, Swaminathan, & Rogers, 1991). Tek boyutluluk varsayımının karşılanıp karşılanmadığı faktör analizi yapılarak kontrol edilir. Yapılan faktör analizi doğrultusunda özdeğerler incelenerek baskın bir faktör arayışı içine girilir. Baskın bir faktörün varlığı tek boyutluluğun karşılandığına işarettir. Stout'un temel tek boyutluluk testi de varsayımın karşılanıp karşılanmadığını kontrol etmek için bir başka alternatiftir (DeMars, 2010).

Yerel bağımsızlık, test performansını etkileyen yetenek düzeyi sabit tutulduğunda, sınava giren bireylerin maddelere verecekleri tepkilerin istatistiksel olarak bağımsız olmasıdır. Tek boyutluluk ve yerel bağımsızlık eş kavramlar değildir. Ancak yerel bağımsızlığın sağlanabilmesi için tek boyutluluk varsayımının karşılanmış olması gerekir. Çünkü maddenin temelini oluşturan birden fazla faktör varsa bu durumda yeteneklerden birini sabitlemek yeterli olmayacak ve istatistiksel olarak bir bağımlılık oluşturacaktır. Bu yüzden çoğu zaman MTK uygulamalarında tek boyutluluk varsayımının karşılanması yerel bağımsızlık varsayımının da karşılanması olarak da kabul edilir. Yerel bağımsızlık ayrı olarak kontrol edilmek istenirse Yen'in Q3 testi kullanılabilir (Sünbül, 2020).

MTK'de a, b ve c olmak üzere üç parametre bulunmaktadır. a parametresi madde ayırt edicilik parametresidir. Maddenin kalitesini yani elde edilen θ yetenek ölçüsünün gerçek θ hakkında ne kadar bilgi vermekte olduğunu gösterir. Bu açıdan bakıldığında bir geçerlik ölçüsüdür. Literatürde MKE'nin eğimi olarak da bilinmektedir. MTK'de b parametresi yani madde güçlüğü maddenin hangi noktada işlevsel olduğuyla ilgilidir. Düşük yetenekli katılımcılarda kolay maddeler işe yararken yüksek yetenekli katılımcılarda daha güç maddeler işe yaramaktadır. Bu açıdan bakıldığında madde güçlük parametresi bir konum parametresidir. MTK'de madde güçlük parametresi b parametresi olarak ifade edilmektedir. Çoktan seçmeli maddelerle ilgili en büyük sorunlardan birisi sınava giren bireyin o maddeyi doğru yanıtlamak için gerekli özelliğe sahip olmasa bile soruyu tahmin yoluyla doğru yanıtlama olasılığıdır. Bu durum çeldiricilerin dikkatli bir şekilde hazırlanmasıyla azalsa da madde seçeneklere sahip olduğu sürece her zaman karşılaşılabileceğimiz bir durumdur. MTK'de c parametresi olarak da ifade edilen şans parametresi MKE için düşük asimptot olarak ifade edilmektedir (Baker, 2001).

2.2.1 İkili Puanlanan Maddeler için MTK Modelleri

MTK'de ikili puanlanan (1-0) maddeler için üç lojistik model bulunmaktadır: bir parametrelili lojistik model, iki parametrelili lojistik model ve üç parametrelili lojistik model. Bir

parametrelili model Rasch modeli olarak da bilinmektedir. Bir parametrelili modelde şans parametresinin 0 olduđu ve ayırt edicilik parametresi olan a değeri için 1'e eşit olduđu varsayılır. Bir parametrelili modelde MKE'yi oluşturan temel kavram b parametresidir (Crocker & Algina, 2006b). Modelin denklemi eşitlik 2.2'de verilmiştir:

$$P(\theta) = \frac{1}{1 + e^{-1(\theta-b)}} \quad (2.2)$$

İki parametrelili modelde de Rasch modelinde olduđu gibi şans parametresi olan c değeri 0 olarak varsayılır. Ancak Rasch modelinden farklı olarak bu sefer her madde için a parametresi de hesaplanır (Crocker & Algina, 2006b). Modelin denklemi eşitlik 2.3'te verilmiştir:

$$P(\theta) = \frac{1}{1 + e^{-a(\theta-b)}} \quad (2.3)$$

Üç parametrelili modelde a, b ve c parametrelerinin tümü madde için hesaplanır (Crocker & Algina, 2006b). Modelin denklemi eşitlik 2.4'te verilmiştir:

$$P(\theta) = c + (1 - c) \frac{1}{1 + e^{-1(\theta-b)}} \quad (2.4)$$

2.2.2 Çoklu Puanlanan Maddeler için MTK Modelleri

MTK'de çoklu puanlanan (0-1-2, 1-2-3-4-5 vb.) maddeler için temelde beş model bulunmaktadır: dereceleme ölçeđi modeli, sınıflamalı tepki modeli, kısmi puan modeli, genelleştirilmiş kısmi puan modeli, aşamalı tepki modeli.

Dereceleme ölçeđi modeli (Rating Scale Model-RSM) Likert tipi ölçekler olarak bilinen sıralama düzeyindeki ölçeklerin analizinde kullanılmaktadır. Modelde bütün maddelerin a parametresi eşit kabul edilir ve kategoriler eşit aralıklardan oluşur (Embretson & Reise, 2000). Modelde b ve c parametresi hesaplanmaktadır. Modelin denklemi eşitlik 2.5'te verilmiştir:

$$P(X_{jk} = k | \theta_j) = \frac{e^{\sum_{h=0}^k (\theta_j - b_i - d_h)}}{\sum_{c=0}^k m_i e^{\sum_{h=0}^k (\theta_j - b_i - d_h)}} \quad (2.5)$$

Sınıflamalı tepki modeli (Nominal Response Model-NRM) yanıtların özellik sürekliliđi boyunca zorunlu olarak sıralanmadıđı durumlarda madde yanıtlarını karakterize etmek için kullanılmaktadır. Bir başka deyişle seçeneklerin özelliklerine göre dođru yanıt ve çeldirici

yanıtlar olarak ayrıldığı analiz süreci yürütülmektedir (Embretson & Reise, 2000). Modelde a ve b parametresi hesaplanmaktadır. Modelin denklemi eşitlik 2.6’da verilmiştir:

$$P(X_{jk} = k|\theta_j) = \frac{e^{a_{ik}\theta_j - c_{ik}}}{\sum_{h=1} m_{i-1} e^{a_{ih}\theta_j - c_{ih}}} \quad (2.6)$$

Kısmi puan modeli (Partial Credit Model-PCM) başarı testlerinin 1-0 olarak puanlaması durumundaki sınırlılıkları gidermek adına geliştirilmiştir. Modele göre maddeler 0 ‘yanlış’, 1 ‘kısmen doğru’, 2 ‘doğru’ vb. şekillerde puanlanabilmektedir (de Ayala, 2009). Modelde yalnızca b parametresi her madde için hesaplanmaktadır. a parametresi bütün maddeler için aynıdır. Modelin denklemi eşitlik 2.7’de verilmiştir:

$$Pr(Y_{ij} = y, y > 0|\theta_j, \beta_i) = \frac{\exp \sum_{s=1} y (\theta_j - \beta_{is})}{1 + \sum_{k=1} m_i \exp \sum_{s=1} y (\theta_j - \beta_{is})} \quad (2.7)$$

Genelleştirilmiş kısmi puan modeli (Generalized Partial Credit Model-GPCM), kısmi puan modelinin a parametresinin bütün maddeler için aynı olması sınırlılığını ortadan kaldırmak için geliştirilmiş bir modeldir. Modelde a parametresi de her madde için hesaplanmaktadır (de Ayala, 2009). Modelin denklemi eşitlik 2.8’de verilmiştir:

$$P(X_{ikj} = k|\theta_j) = \frac{e^{\sum_{h=0}^k [a_i(\theta_j - b_{ih})]}}{\sum_{c=0}^{m_i} e^{\sum_{h=0}^c [a_i(\theta_j - b_{ih})]}} \quad (2.8)$$

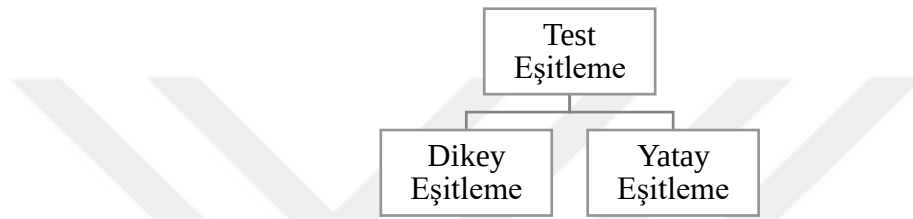
Aşamalı tepki modeli (Graded Response Model-GRM) yığılmalı kategori yanıt fonksiyonlarının şeklinin aynı olduğu ve birbirleriyle asla kesişmediği durumlarda kullanılan bir modeldir (Paek & Cole, 2020). Modelde a ve b parametresi hesaplanmaktadır. Modelin denklemi eşitlik 2.9’da verilmiştir:

$$P(X_{ij} \geq k|\theta_j) = \frac{e^{a_i(\theta_j - b_{ik})}}{1 + e^{a_i(\theta_j - b_{ik})}} \quad (2.9)$$

2.3. Test Eşitleme

Geniş ölçekli testlerde, test güvenliği ve genellenebilirliği korumak için testin birden fazla formu oluşturulur. Oluşturulan test formlarının zorluk ve değişkenlik açısından benzer olmalıdır. Formların bu benzerliği test eşitleme ile kontrol edilir (Jaeger, 1981). Aynı testin benzer zorluklara sahip iki formunun puanlarının birbirine dönüştürülmesini içeren istatistiksel süreç test eşitleme olarak tanımlanır (Kolen & Brennan, 2014).

2.3.1. Test Eşitleme Türleri



Şekil 2.2. Test Eşitleme Türleri

Şekil 2.2’de görüldüğü üzere iki çeşit test eşitleme vardır: dikey ve yatay. Crocker ve Algina (2006a) dikey eşitlemeyi bir testin eğitimin farklı aşamalarında uygulanan formlarının eşitlenmesi olarak tanımlarken yatay eşitlemeyi ise bir testin benzer özelliklere sahip formları arasındaki eşitleme olarak tanımlamıştır. Yatay eşitlemede kullanılan test formları, içerik ve zorluk olarak benzer olmasının yanı sıra benzer bir grup için hazırlanmaktadır. Dikey eşitlemede ise test formları benzer içeriğe sahip olsa da farklı yetenek dağılımına sahip gruplar için kasıtlı olarak farklı zorluk seviyelerinde tasarlanmaktadır (Loyd & Hoover, 1980).

2.3.2. Test Eşitleme Adımları

Test eşitleme yedi adımdan oluşan bir süreçtir (Kolen & Brennan, 2014):

1. Eşitle amacına karar verilmesi: Test eşitleme süreci genel olarak adaleti sağlama isteğiyle yürütülmektedir. Test eşitleme sürecine başlamadan önce kişi ya da kurum ilk olarak bu işlemi hangi amaçla yapacağını belirlemelidir. Bu amaç yöntem karşılaştırması yaparak benzer durumlarda hangi yöntemin kullanılmasının daha sağlıklı olacağına karar vermek veya bir kurumun hangi bireyi öğrenci ya da çalışan olarak alacağı olabilir.

2. Alternatif formlar oluşturulması: Testin birden çok kez yapılması ya da farklı tarihlerde yapılması alternatif formlar oluşturmayı gerektirir. Bu formların oluşturulması test eşitleme işleminin önemli bir adımını oluşturmaktadır. Formlar benzer içerik ve güçlüğe sahip olacak şekilde oluşturulur.
3. Veri toplamak için desen seçilmesi: Formların oluşturulmasının ardından veri toplamak için desen seçilmelidir. Bu desen seçilirken ankor testin varlığı ve örneklem büyüklüğü göz önünde bulundurulur.
4. Veri toplama deseninin uygulanması: Uygun desenin seçilmesinin ardından formlar bireylere uygulanır.
5. Varsayımların kontrol edilmesi: Test eşitlemenin istatistiksel bir süreç olması beraberinde bazı varsayımların kontrol edilmesini de gerektirmektedir. Oluşturulan test formlarını kullanarak eşitleme yapabilmek için birtakım varsayımlar kontrol edilir.
6. Bir ya da daha fazla istatistiksel yöntem seçilmesi: Varsayımların karşılanmasıyla beraber test eşitlemenin amacına uygun olarak bir ya da daha fazla yöntem belirlenir.
7. Eşitleme sonuçlarının değerlendirilmesi: Test eşitleme sürecinin son adımı sonuçların değerlendirilmesidir. Seçilen yöntemlerle testlerin farklı formları eşitlenerek yeni puanlar elde edilir. Elde edilen puanlar ve hata miktarlarının hesaplanması sonrasında eşitleme sonuçları değerlendirilir.

2.3.3. Test Eşitleme Koşulları

Test eşitlemenin amacı iki teste ait puanların birbirlerinin yerine kullanılması olarak da tanımlanabilir. Dolayısıyla bu durum iki testin eşitlenebilmesi için bazı gereklilikleri de beraberinde getirmektedir. İki testin eşitlenebilmesi için beş koşul vardır (Dorans, Moses, & Eignor, 2011). Bu koşullara aşağıda yer verilmiştir:

1. Eşit yapı gereksinimi: İki test de aynı örtük özelliği ölçmelidir. Ayrıca, iki test benzer istatistiksel özelliklere sahip olmalıdır.
2. Eşit güvenilirlik şartı: İki testin güvenilirliği aynı olmalıdır. Aynı zamanda iki testin de güvenilirlikleri yüksek olmalıdır.
3. Simetri gereksinimi: Form X'in puanlarının Form Y puanlarına dönüştürülmüş hali ile Form Y'nin puanlarının Form X puanlarına dönüştürülmüş hali eşit olmalıdır. Bir başka deyişle, Y'nin X'e eşitlenmesi ve bir puanın diğerine dönüştürülmesi arasındaki farkın vurgulanması amacıyla eşitleme, hangi

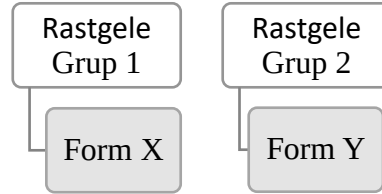
formun X ve hangi formun Y olarak isimlendirilmesinde bağımsız olmayı gerektirir (van der Linden & Wiberg, 2010). Örneğin, Form X'den alınan 50 puan, eşitleme sonucu Form Y'den alınan 53 puana denk geliyorsa; Form Y'den alınan 53 puan eşitleme sonucu Form X için puana denk gelmelidir.

4. Eşitlik gereksinimi: Sınava giren bireylerin hangi formu aldığı eşitleme sürecini etkilememelidir. Lord'un önerdiği eşitlik gereksinimine göre sınava giren bir birey iki test formundan da aynı puanı almalıdır. Bir diğer deyişle, Form X'deki a puanı Form Y için de a puanı olmalıdır. Bu gereksinimin gerçekleşme ihtimali çok düşük olduğu ve gerçekleştiği durumda eşitleme yapmak gerek gereksiz olduğu için Lord'un eşitlik gereksinimi pratikte kriter olarak alınmamaktadır.
5. Grup değişmezliği gereksinimi: Grup değişmezliği gereksinimi test eşitleme işleminin puanlar arasında etkili bir denklik üretme amacından ortaya çıkmaktadır. İki test formundan alınan puanlar eşit ise, iki popülasyon arasında bire bir benzeşme vardır. Bu eşitlemenin benzersiz olduğu anlamına gelmektedir. Bir başka deyişle eşitleme ele alınan gruptan bağımsız olarak aynı olmalıdır (Petersen & ACT, 2008).

2.3.4. Test Eşitleme Desenleri

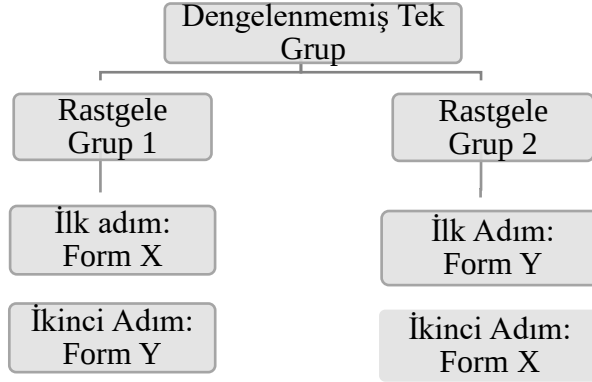
Rastgele eşdeğer grup deseni. Rastgele eşdeğer grup deseni, popülasyondan iki veya daha fazla rastgele örneklem almayı ve ardından her birine tek bir test uygulamayı içerir. Bu şekilde, gruplar arasındaki farklı puanlar testlerin zorluk derecelerindeki farklılıklara bağlanabilir ve bu nedenle eşitlenmeleri gerekir. Bu modelin uygulanması rastgele eşdeğer gruplar gerçekten oluşturulmuşsa geçerlidir; bu çok büyük örneklemle çalışırken yerine getirilen bir koşuldur (Guilera & Gomez, 2008). Rastgele eşdeğer grup deseni uygulanırken spiral yöntemi kullanılabilir. Spiral yöntemi, oturma düzenine göre test formlarının verildiği bir yöntemdir. İlk kişiye Form X verilirken ikinci kişiye Form Y, üçüncü kişi Form X ve dördüncü kişi Form Y olmak üzere katılımcılara testler verilir. Sistematiik oturma düzeninin olduğu durumlarda bu desenin uygulanması eşitlemenin bozulmasına neden olabilir. Örneğin, bir sınıfta kız-erkek-kız-erkek şeklinde bir oturma düzeni varsa katılımcılar otomatik olarak kızlar ve erkekler olarak bölünecektir. Bu durum da test sonuçlarında olası bir yanlılığa neden olabilir. Büyük örneklemle çok fazla test sınıfının kullanılmasını gerektirebilir. Bu durumda da

sistematik oturma düzeninden kaynaklı sıra etkisine ve katılımcıların benzer sayıda olmasına dikkat edilmesi gerekmektedir.



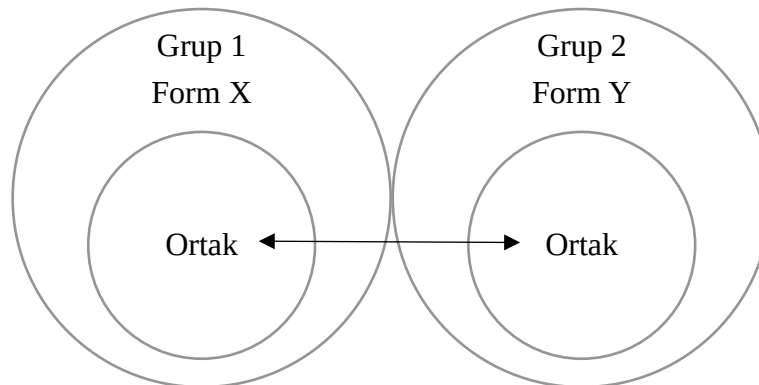
Şekil 2.3. Rastgele Eşdeğer Gruplar

Dengelenmiş tek grup deseni. Tek grup tasarımında sıra etkileri olasılığına izin vermek için, çalışma grubu bazen rastgele ikiye bölünür ve her gruba iki test farklı sıralarda yapılır: İlk gruba önce X, sonra Y veya önce Y ve sonra X. Bu desen, dengelenmiş tek grup deseni olarak adlandırılır. Dengelenmiş tek grup deseni, içerisinde hem tek grup hem de eşdeğer gruplar desenini içerir. Dengelenmiş tek grup deseni küçük örneklemelerde daha doğru bir eşitleme sağlar. Ancak bu avantajın yanısıra dezavantaj olarak verileri toplamak için özel bir çalışma gerektirir (Dorans, Moses, & Eignor, 2010). Tek grup deseni olumsuz etkileri en aza indirmek için geliştirilen bir yöntem olmasına rağmen bazı olumsuzluklar içermektedir. Örneğin, iki alt grubun da testleri sırayla alması zaman konusunda bir olumsuzluktur. Yeterince zamanın olmadığı durumlarda kısa aralıklarla iki testin alınması katılımcılarda yorgunluk hissi yaratabilir. Bu durum da test puanları ve dolayısıyla eşitleme sonuçlarında olumsuz etki yaratabilir.



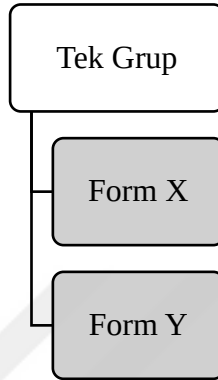
Şekil 2.4. Dengelenmiş Tek Grup

Ortak madde eşdeğer olmayan gruplar deseni. Ortak madde eşdeğer olmayan gruplar deseninde X ve Y olmak üzere iki test formu ve P ve Q olmak üzere iki popülasyon bulunmaktadır. P popülasyonu X formunu ve Q popülasyonu ise Y formunu almaktadır. Eşdeğer olmayan gruplar kullanıldığında test eşitleme işlemini yapabilmek için ankor madde adı verilen maddelere ihtiyaç duyulmaktadır. V ankor maddelerinin X ve Y formlarına dahil olması durumunda iç ankor, ayrı bir test formu olarak verilmesi durumuna ise dış ankor denilmektedir (von Davier, Wilson, & Educational Testing Service, 2008). Diğer desenlerde yer alan güvenlik sorunu gibi kaygıların giderilmesi amacıyla geliştirilmiş bir desendir.



Şekil 2.5. Ortak Madde Eşdeğer Olmayan Gruplar

Tek grup deseni. Katılımcılar sırayla X ve Y formunu alırlar. Sırayla iki formun da katılımcılara uygulanması yorgunluk etkisi oluşturulabilir. Ayrıca formların benzer olması yine bu desenin uygulanmasındaki olumsuzluklardan birisidir. Katılımcıların ikinci sırada aldıkları formda daha yüksek puan almaları durumunda ilk test mi daha kolay aşinalık mı buna sebep oldu belirlemek güçtür.



Şekil 2.6. Tek Grup Deseni

2.3.5. Test Eşitleme Yöntemleri

Test eşitleme yöntemleri genel olarak geleneksel eşitleme yöntemleri ve Madde Tepki Kuramı'na dayalı eşitleme yöntemleri olmak üzere iki başlık altında yer almaktadır. Geleneksel eşitleme yöntemleri Klasik Test Kuramı'na dayalı eşitleme yöntemleri olarak da bilinmektedir. Geleneksel test eşitleme yöntemleri aşağıdaki gibidir:

1. Eşit yüzdelikli eşitleme,
2. Doğrusal eşitleme,
3. Ortalama eşitleme,
4. Kernel Yöntemi,
5. Tucker Yöntemi,
6. Levine Gözlenen Puan Yöntemi,
7. Levine Gerçek Puan Yöntemi,
8. Chained Doğrusal Yöntem,
9. Sıklık Kestirim Yöntemi,
10. Değiştirilmiş Sıklık Kestirimi,
11. Chained Eşit Yüzdelikli Eşitleme,
12. Circle-Arc Yöntemi.

Madde Tepki Kuramı'na dayalı test eşitleme yöntemleri aşağıdaki gibidir:

1. Ortalama-ortalama,
2. Ortalama-standart sapma,
3. Haebara,
4. Stocking-Lord,
5. Gözlenen Puan yöntemi,
6. Gerçek Puan Yöntemi.

2.3.5.1. Geleneksel Eşitleme Yöntemleri

2.3.5.1.1. Eşit Yüzdelikli Eşitleme

Form X'de yer alan bir puan ve Form Y'de yer alan bir puanın yüzdelik sıralaması aynıysa testi alan grup içerisinde eşit kabul edilir. Buna göre, puanları ayarlamak için Form X'de yer alan puanları Form Y'de yer alan puanlara eşitlemek için Form X puanları aynı yüzdelik dilime sahip olduğu Form Y puanlarına dönüştürülür. Bu işlem eşit yüzdelikli eşitleme işlemi ifade eder. Testin zor bir formu, yetenek düzeyi yüksek test katılımcılarının puanlarını dağıtma eğiliminde olacaktır; düşük yetenekli sınava girenlerin puanları altta bir araya toplanacaktır. Bu durumda, yüksek yüzdelik dilimler birbirinden daha uzak ve alt yüzdelik dilimler birbirine daha yakın olacaktır. Testin kolay bir formu, yetenek düzeyi düşük olan teste girenlerin puanlarını dağıtma eğiliminde olacaktır; yüksek yetenekli sınav katılımcılarının puanları en üstte bir araya gelecektir. Bu durumda, alt yüzdelik dilimler birbirinden daha uzak ve yüksek yüzdelik dilimler birbirine daha yakın olacaktır. Form X'deki her puan, hedef popülasyonda aynı yüzdelik sıralamasına sahip ya da mümkün olduğunca yakın Form Y'deki ham puana eşit olacak şekilde ayarlanır. Eşit yüzdelikli eşitleme, Form X'deki düzeltilmiş puanların, Form Y'deki puanlarla hemen hemen aynı dağılıma sahip olmasını sağlayacaktır (Livingston, 2014). Eşit yüzdelik eşitlemede, zorlukta biçimden biçime farklılıkları tanımlamak için bir eğri kullanılır, bu da eşit yüzdelik eşitlemeyi doğrusal eşitmeye göre daha genel hale getirir. Örneğin, eşit yüzdelik eşitlemeyi kullanmak, Form X, yüksek ve düşük puanlarda Form Y'den daha zor olabilir, ancak orta puanlarda daha az zor olabilir. Eşitleme işlevi, Form Y ölçeğine dönüştürülen Form X'deki puanların dağılımı, popülasyondaki Form Y'deki puanların dağılımına eşitse, eşit yüzdelikli bir eşitleme işlevidir. Eşit yüzdelik eşitleme, Form Y'deki puanlarla aynı yüzdelik derecelere sahip olan Form X üzerindeki puanların belirlenmesiyle geliştirilen rastgele grup deseni uygulanarak kullanılan bir gözlenen puan yöntemidir (Kolen &

Brennan, 2014). Eşit yüzdelikli eşitleme yöntemi için kullanılan dağılım eşitlikleri eşitlik 2.10 ve eşitlik 2.11’de verilmiştir (Kolen & Jarjoura, 1987):

$$F_s(x) = w_a F_a(x) + w_b F_b(x) \quad (2.10)$$

$$G_s(y) = w_a G_a(y) + w_b G_b(y) \quad (2.11)$$

Eşitlikte a, b ve s sırasıyla a ve b gruplarını ve sentetik evreni temsil etmektedir. w_a ve w_b toplamı 1 olan ağırlık oranıdır.

Rastgele gruplar eşitleme tasarımında eşit yüzdelikli eşitleme sonuçlarını iyileştirmek için iki tür düzgünleştirme yöntemi kullanılmaktadır. Ön düzgünleştirme yöntemleri, eşit yüzdelik eşitleme yapılmadan önce eşitlenecek iki test formu için puan dağılımlarının düzgünleştirilmesini içerir. Son düzgünleştirme yöntemleri, düzgünleştirilmemiş puan dağılımlarından üretilen eşit yüzdelikli eşitleme fonksiyonunun düzgünleştirilmesini içerir. Hem ön düzgünleştirme hem de son düzgünleştirme düzgünleştirilmemiş eşitlemeye göre daha iyi sonuçlar vermektedir (Hanson, Zeng, & Colton, 1994).

2.3.5.1.2. Doğrusal Eşitleme

Doğrusal eşitleme yönteminde, ortalamalar ve standart sapma dışında iki formun eşit olduğu varsayılır (Muraki, Hombo, & Lee, 2000). İki form arasındaki farkları sabit olarak düşünmek yerine, iki test formu arasındaki zorluk farklılıklarının puan ölçeği boyunca değişmesine izin verir. Örneğin, Form X’in, düşük başarılı sınavlar için Form Y’den daha zor olmasına, ancak yüksek başarılı sınavlar için daha az zor olmasına izin verir. Doğrusal eşitlemede, ortalamalarına eşit uzaklıkta olan puanlar standart sapma birimlerinde eşit olarak ayarlanır. Bu nedenle, doğrusal eşitleme, iki formun araçlarının yanı sıra ölçek birimlerinin de farklı olmasına izin veriyor olarak görülebilir. Bu duruma bağlı olarak rastgele gruplarda sıklıkla kullanılan gözlenen puan yöntemlerinden birisidir. Kullanılan denklem eşitlik 2.12’de verilmiştir:

$$\frac{X - \mu(X)}{\sigma(X)} = \frac{Y - \mu(Y)}{\sigma(Y)}$$

(2.12)

2.3.5.1.3. Ortalama Eşitleme

Ortalama eşitleme, doğrusal eşitleme ve eşit yüzdelikli eşitlemeye göre daha az katıdır. Form X ve Form Y'den alınan puan ortalamalarının eşit olmasını gerektirir (Sansivieri, Wiberg, & Matteucci, 2017). Ortalama eşitleme, büyük örneklemelerde kullanılması ve daha az varsayıma sahip olması açısından daha avantajlı olan rastgele gruplar deseninin uygulandığı gözlenen puan yöntemlerinden birisidir. Kullanılan denklem eşitlik 2.13'te verilmiştir:

$$m_Y(x) = x - \mu(X) + \mu(Y)$$

(2.13)

$m_Y(x)$, Form X'te yer alan x puanının Form Y ölçeğinde yer alan bir puana dönüştürülmüş halidir.

2.3.5.1.4. Kernel Yöntemi

Kernel yöntemi, eşit yüzdelikli eşitlemeye benzer bir eşitleme yöntemidir. Kernel yöntemi beş adımdan oluşur (von Davier, Holland, & Thayer, 2004):

1. Ön-düzgünleştirme: Bu adımda, veri toplama deseni ile elde edilen ham verilere uygun istatistiksel modeller uydurularak ilgili tek değişkenli ve/veya iki değişkenli puan olasılıklarının tahminleri elde edilir. Bu, veriler için çeşitli modellerin denendiği ve verilere uygun bir uyum sağlamak için seçildiği tamamen istatistiksel bir aşamadır.
2. Puan Olasılıklarının Kestirimi: Hedef popülasyonun puan olasılıkları, ilk adımda tahmin edilen puan olasılıkları kullanılarak kestirilir. Bu adımda yer alan tasarım fonksiyonu, ilk adımdaki puan dağılımlarından puan kestiriminin yapıldığı doğrusal ya da doğrusal olmayan bir dönüşümdür.
3. Süreklileştirme: Bu adımda elde edilen kesikli puan dağılımları bir ölçek boyunca sürekli puan dağılımlarına dönüştürülür.
4. Eşitleme: İlk üç adım tamamlandıktan sonra eşitleme işlemi yürütülür.

5. Eşitlemenin Standart Hatasının Hesaplanması: Son adım olarak hem eşitlemenin standart hatası hem de eşitlemenin standart hatası farkı hesaplanmaktadır.

2.3.5.1.5. Tucker Yöntemi

Tucker yöntemi, ortak madde seti olan V'nin üzerinde X'in doğrusal bir regresyonu olduğunu ve her iki katılımcı grubu için de koşullu varyansın aynı olduğu varsayımına sahiptir. Benzer bir varsayım Y'nin V üzerinde doğrusal bir regresyonu olduğuna dair de olabilir. Bu varsayımlarla yapay grubun ortalama ve varyansları, verilerden eşitlik 2.14, 2.15, 2.16 ve 2.17 yardımıyla elde edilebilir (Peabody, 2020):

$$\mu_s(X) = \mu_1(X) - w_2\gamma_1[\mu_1(V) - \mu_2(V)] \quad (2.14)$$

$$\mu_s(Y) = \mu_2(Y) - w_1\gamma_2[\mu_1(V) - \mu_2(V)] \quad (2.15)$$

$$\sigma_s^2 = \sigma_1^2(X) - w_2\gamma_1^2[\sigma_1^2(V) - \sigma_2^2(V)] + w_1w_2\gamma_1^2[\mu_1(V) - \mu_2(V)]^2 \quad (2.16)$$

$$\sigma_s^2 = \sigma_2^2(Y) - w_2\gamma_2^2[\sigma_1^2(V) - \sigma_2^2(V)] + w_1w_2\gamma_2^2[\mu_1(V) - \mu_2(V)]^2 \quad (2.17)$$

γ terimleri, ortak madde seti olan V üzerindeki toplam test puanları (X ve Y) için regresyon eğimleridir.

2.3.5.1.6. Levine Gözlenen Puan Yöntemi

Levine gözlenen puan yöntemi sentetik evreni dikkate almayan, daha çok eşdeğer olmayan gruplar deseni uygulanarak kullanılan bir doğrusal eşitleme yöntemidir. Bu yöntem, bir denklem ile Form X'den alınan gözlenen puanları Form Y'den alınan gözlenen puanlar ölçeği ile ilişkilendirir. Klasik Test Kuramı'na göre gözlenen puanlar ile ilişkilendirilen, T_X , T_Y ve T_V gerçek puanlarının kullanıldığı denklem eşitlik 2.18, 2.19 ve 2.20'de verilmiştir:

$$X = T_X + E_X \quad (2.18)$$

$$Y = T_Y + E_Y \quad (2.19)$$

$$V = T_V + E_V \quad (2.20)$$

Eşitliklerde yer alan E_X , E_Y ve E_V gerçek puanla ilişkisiz olan hata değerleridir.

2.3.5.1.7. Levine Gerçek Puan Yöntemi

Levine gözlenen puan yöntemi ile aynı varsayımları kullanarak gerçek puanlar için bir yöntem geliştirmiştir. Gözlenen puan yöntemi, gözlenen puanlardan yola çıkarak bir denklem ile eşitleme yaparken; gerçek puan yönteminde gerçek puanlar eşitlikte yer almaktadır. Levine gerçek puan de daha çok eşdeğer olmayan gruplar deseni uygulanarak kullanılan bir doğrusal eşitleme yöntemidir (Kolen & Brennan, 2014). Kullanılan denklem eşitlik 2.21’de verilmiştir:

$$l_{Y_S} = \frac{\sigma_X(T_Y)}{\sigma_X(T_X)} [t_X - \mu_S(T_X)] + \mu_S(T_Y) \quad (2.21)$$

Klasik Test Kuramı’na göre, gözlenen puanların ortalaması gerçek puan ortalamasına eşittir. Bu yüzden,

$$l_{Y_S} = \frac{\sigma_X(T_Y)}{\sigma_X(T_X)} [t_X - \mu_S(X)] + \mu_S(Y) \quad (2.22)$$

2.3.5.1.8. Chained Doğrusal Yöntem

Doğrusal eşitleme gerçekleştirilirken;

1. X, V ölçeğine yerleştirilerek $l_V(x)$ adını alır,
2. V, Y ölçeğine yerleştirilerek $l_Y(v)$ adını alır,
3. Y eşitliği $l_Y(x) = l_Y[l_V(x)]$ olmak üzere üç adımda tamamlanır.

Chained doğrusal yöntem adını üç aşamada gerçekleşmesinden almaktadır. Üçüncü adım birleştirme fonksiyonudur. V'nin hem X hem de Y ile bağlantılı olduğu düşünülerek son adımda X ve Y arasında ilişki kurulur. Levine ve Tucker yöntemleri kadar yaygın olarak kullanılmayan bir yöntemdir. İlk adımda uzun formdaki puanlar kısa formdaki puanlarla, ikinci adımda ise kısa formdaki puanlar uzun formdaki puanlarla ilişkilendirilir. Biçim farklılığı yöntemin olumsuzluğu olarak görülmektedir. Ayrıca bir diğer durum ilk adımda grup 1'in ve ikinci adımda grup 2'nin yer almasıdır. Bu da eşitlemenin hangi grup için yapıldığı problemini beraberinde getirmektedir. Bu problem de iki grup ağırlığının önceden değişmez olduğu varsayımıyla çözülmektedir (Kolen & Brennan, 2014).

2.3.5.1.9. Sıklık Kestirim Yöntemi

Sıklık kestirim yöntemi, ankor madde eşdeğer olmayan gruplar deseni ile toplanan verilerden sentetik evren oluşturmak için Form X ve Form Y'deki puanların yığılmalı dağılımlarını kestirmek için kullanılan bir yöntemdir. Yüzdelik sıralamalar bu yığılmalı dağılımlardan elde edilir ve daha sonra eşit yüzdelikli eşitleme yöntemi ile eşitleme işlemi gerçekleştirilir (Kolen & Brennan, 2014).

2.3.5.1.10. Değiştirilmiş Sıklık Kestirimi Yöntemi

Sıklık kestirim yöntemi sonuçları belirli koşullarda yanlı sonuçlar verebilmektedir. Bu problemi çözmek için sıklık kestirim yönteminin varsayımları, ortak maddeler için gerçek puanlar üzerinde koşullandırmaya dayalı karşılık gelen varsayımlarla değiştirilmiştir (Kolen & Brennan, 2014).

2.3.5.1.11. Chained Eşit Yüzdelikli Eşitleme Yöntemi

Eşdeğer olmayan gruplarda kullanılan eşit yüzdelikli eşitleme yöntemlerinden birisidir. Üç adımda gerçekleştirilir (Kolen & Brennan, 2014):

1. Form X'den alınan puanları ilk gruptan alınan ortak maddelere dayalı puanlara dönüştürmek için aralarındaki yüzdelik ilişki bulunur. Bu ilişki $ev_1(x)$ olarak ifade edilir.

2. İkinci gruptan elde edilen ortak maddelere dayalı puanları Form Y'den alınan puanlara dönüştürmek için yüzdelik ilişki bulunur. Bu ilişki de $eY_2(v)$ olarak ifade edilir.
3. Form X puanlarını Form Y puanlarına dönüştürmek için, önce Form X puanını ortak madde puanına dönüştürmek için $ev_1(x)$ kullanılır. Ardından ortak madde puanlarını Form Y puanlarına dönüştürmek için $eY_2(v)$ kullanılır.

2.3.5.1.12. Circle-Arc Yöntemi

Divgi tarafından ortaya atılan bir yöntemde, tahmin edilen eşitleme eğrisi önceden belirlenen iki uç nokta ve deneysel olarak belirlenen orta nokta ile sınırlandırılmaktadır. Eşitlenecek test puanlarından alınabilecek minimum ve maksimum nokta uç noktaları, testin ortalama puanı ise orta noktayı oluşturmaktadır. Tahmini oluşturulan eşitleme dönüşümü, standart sapma ile belirlenen orta noktadaki eğrinin eğimi ile bu üç noktadan geçen kübik bir polinom eğrisidir. Çoktan seçmeli bir test için minimum nokta şans puanı ile belirlenmektedir. Böylece, soruları okumadan yanıtlayan bir katılımcının alabileceği puan minimum nokta seçilmiş olur. Eğri üzerindeki orta nokta, puan dağılımının ortasındaki bir noktada eşitlenen veriler ile belirlenir. Bu üç nokta düz bir çizgi üzerinde bulunuyorsa, bu doğru bir tahmini eşitleme eğrisidir. Üç nokta düz bir çizgi üzerinde yer almıyorsa, bir daire yayı belirlerler. Bu yay, uç noktalar arasında yer alan aralıktaki eşitleme eğrisinin bir tahmini olarak kullanılır (Livingston & Kim, 2009). Bu yöntemden yola çıkılarak eşitleme fonksiyonu doğrusal ve eğrisel bileşen olarak ayrılmıştır. Eğrisel bileşen için model olarak bir daire yayı kullanılmıştır. Daire yayı, eşitleme eğrisinin doğrusallıktan sapması için eğrinin kendisinin şeklinden daha iyi bir modeldir. Eşitleme fonksiyonunu en düşük anlamlı puanın altına genişletmek için, eğrinin alt uç noktası, iki formdaki olası minimum puanları temsil eden noktaya doğrusal olarak bağlanabilir. Minimum ve maksimum nokta ile oluşturulan düz çizgi, eşitleme fonksiyonunun doğrusal bileşeni olan $L(x)$ ile ifade edilir. Kullanılan denklem eşitlik 2.23'te verilmiştir:

$$L(x) = y_1 \frac{y_3 - y_1}{x_3 - x_1} (x - x_1) \quad (2.23)$$

$L(x)$ doğrusu çıkarılarak üç nokta dönüştürülür:

$$y^* = y - L(x)$$

(2.24)

Dönüştürülmüş üç nokta daire yayı belirlemek için kullanılır. Bu yay eşitlenmiş puanları belirler. r dairenin yarıçapı, x_c ve y_c merkez noktalar olduğu zaman dairenin denklemi eşitlik 2.25'te verilmiştir:

$$(X - x_c)^2 + (Y - y_c)^2 = r^2$$

(2.25)

y_2^* pozitif ise yay denklemi

$$Y^* = y_c + \sqrt{r^2 - (X - x_c)^2}$$

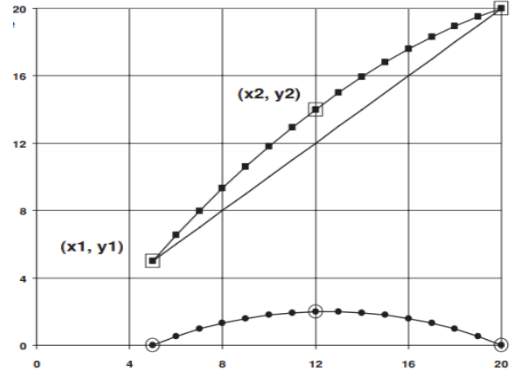
(2.26)

y_2^* negatif ise yay denklemi

$$Y^* = y_c - \sqrt{r^2 - (X - x_c)^2}$$

(2.27)

olur.



Kaynak. (Livingston & Kim, 2009).

Şekil 2.7. Daire Yayı Kullanılarak Dönüştürülmüş Noktalar

2.3.5.2. Madde Tepki Kuramı'na Dayalı Test Eşitleme Yöntemleri

Bir MTK modeli verilere uyduğunda, değişmezlik özelliğinden dolayı farklı testlere giren bireylerin yetenek parametrelerinin doğrudan karşılaştırılması mümkündür. Teknik olarak, MTK'de test puanlarını eşitlemek gereksizdir çünkü madde yanıtları, test formlarındaki puanların tek bir adımda ilişkilendirileceği şekilde ölçeklenebilir. Ancak iki test formundan elde edilen madde ve yetenek parametrelerinin aynı ölçekte olması esastır. Bu nedenle, MTK'de esas olan, bir eşitleme çalışması yürütmek yerine test yanıtlarını uygun şekilde ölçeklemektir. Ancak bir başka açıdan bakıldığında ise MTK'ye dayalı eşitleme yöntemleri geleneksel yöntemlerinin eksikliklerinin tamamlanarak geliştirildiği yöntemler olduğu için bazı durumlarda daha sağlıklı sonuçlar verebilmektedir (Price, 2017).

2.3.5.2.1. Ortalama-Ortalama Yöntemi

Ortak maddeli eşdeğer olmayan gruplar deseninde ölçekleri dönüştürmenin en temel yolu, MTK eşitliklerinde yer alan parametreler için ortak maddelerin madde parametre kestirimlerinin ortalama ve standart sapmalarını yerine koymaktır. Dönüşümden sonra madde parametreleri genellikle kalibre edilmiş olarak varsayılır. A sabiti için MTK eşitliklerinde yer alan kişi ve madde parametreleri yerine ortak maddelerden elde edilen a parametresinin kestirimlerinin ortalamaları kullanılır. B sabiti yerine ise ortak maddelerden elde edilen b parametresinin kestirimlerinin ortalamaları kullanılır. Böylece temelde $Y = A + BX$ şeklinde

bir denklem elde edilerek eşitleme işlemi gerçekleştirilir. A ve B sabiti için kullanılan denklem eşitlik 2.28’de verilmiştir (Gonzalez, 2014):

$$A = \frac{\mu(a_x)}{\mu(a_y)} \quad B = \mu(\beta_y) - A\mu(\beta_x) \quad (2.28)$$

2.3.5.2.2. Ortalama-Standart Sapma Yöntemi

Ortalama-standart sapma yöntemi ortalama-ortalama yöntemine benzer bir yöntemdir. A sabiti için ortak maddelerden elde edilen parametre kestirimlerinin standart sapmaları kullanılır. B sabiti için MTK eşitliklerinde yer alan kişi ve madde parametreleri yerine ortak maddelerden elde edilen b parametresinin kestirimlerinin ortalamaları ve A sabiti kullanılır. Böylece temelde $Y = A + BX$ şeklinde bir denklem elde edilerek eşitleme işlemi gerçekleştirilir. A ve B sabiti için kullanılan denklem eşitlik 2.29’da verilmiştir:

$$A = \frac{\sigma(b_x)}{\sigma(b_y)} \quad B = \mu(\beta_y) - A\mu(\beta_x) \quad (2.29)$$

2.3.5.2.3. Haebara Yöntemi

Ortalama-ortalama ve ortalama-standart sapma yöntemleriyle ilgili olası bir problem, a, b ve c parametre kestirimlerinden elde edilen çeşitli yanıt örüntüleri, çoğu öğrencinin puan aldığı yetenek aralığında hemen hemen aynı madde karakteristik eğrileri oluştuğunda ortaya çıkar. Örneğin, iki kestirimde çok farklı b parametresi değerlerine sahip bir madde çok benzer madde karakteristik eğrilerine sahip olabilir. Bu durumda, iki kestirimdeki maddeler için madde karakteristik eğrileri çok benzer olsa bile, ortalama-standart sapma yöntemi, b parametresi kestirimleri arasındaki farktan aşırı derecede etkilenebilir. Bu problem, ölçek dönüştürme yöntemlerinin tüm madde parametre kestirimlerini aynı anda dikkate almamasından kaynaklanmaktadır. Bu problemin giderilmesi için tüm madde parametrelerini aynı anda dikkate alan bir yöntem geliştirmiştir (Kolen & Brennan, 2014). Haebara tarafından madde karakteristik eğrileri arasındaki farkı ifade etmek için kullanılan yöntem, belirli bir yeteneğe sahip kişilerde her bir madde için madde karakteristik eğrileri arasındaki farkın karesinin

toplamı alınarak eşitleme işleminin gerçekleşmesini sağlamaktadır. Yöntemde kullanılan denklem eşitlik 2.30’da verilmiştir (Gonzalez, 2014):

$$H_{crit} = \sum_i \sum_{j:V} \left[p_{ij}(\theta_{yi}, \hat{a}_{yj}, \hat{\beta}_{yj}, \hat{c}_{yj}) - p_{ij} \left(\theta_{yi}, \frac{\hat{a}_{xj}}{A}, A\hat{\beta}_{xj} + B, \hat{c}_{xj} \right) \right]^2 \quad (2.30)$$

2.3.5.2.4. Stocking-Lord Yöntemi

Ortalama-ortalama ve ortalama-standart sapma yöntemlerinde ortaya çıkabilecek problemleri gidermek adına Stocking ve Lord da Haebara’nın yöntemine benzer bir yöntem geliştirmişlerdir. Hem kendi yöntemlerini hem de Haebara yöntemini karakteristik eğri yöntemleri olarak adlandırmışlardır. Madde karakteristik eğrileri arasındaki farkın temel alındığı bu yöntemde kullanılan denklem eşitlik 2.31’de verilmiştir (Gonzalez, 2014):

$$SL_{crit} = \sum_i \left[\sum_{j:V} p_{ij} p_{ij}(\theta_{yi}, \hat{a}_{yj}, \hat{\beta}_{yj}, \hat{c}_{yj}) - \sum_{j:V} p_{ij} \left(\theta_{yi}, \frac{\hat{a}_{xj}}{A}, A\hat{\beta}_{xj} + B, \hat{c}_{xj} \right) \right]^2 \quad (2.31)$$

2.3.5.2.5. Gözlenen Puan Yöntemi

Gözlenen puan yöntemi, her formda gözlenen doğru yanıtların puanlarının tahmini bir dağılımını elde etmek için MTK modeli kullanılan bir eşitleme yöntemidir. Dağılım elde edildikten sonra eşit yüzdelikli eşitleme yöntemi kullanılarak puanlar eşitlenir. Form X, bileşik binom dağılımı belirli bir yeteneğe sahip olması olası olan bireyler için gözlemlenen doğru puanların dağılımını oluşturmak için kullanılır. Daha sonra, bu gözlenen puan dağılımları, Form X için bir doğru gözlenen puan dağılımı üretmek üzere bir sınava girenler üzerinde toplanır. Form Y için, doğru gözlenen puan dağılımını oluşturmak için benzer yöntemler izlenir. Ortaya çıkan doğru puan dağılımları daha sonra geleneksel eşit yüzdelikli yöntemler kullanılarak eşitlenir. MTK tarafından gözlenen puan eşitleme, sınava girenlerdeki yetenek dağılımının açık bir şekilde belirtilmesini gerektirir (Kolen & Brennan, 2014).

2.3.5.2.6. Gerçek Puan Yöntemi

Gerçek puan yöntemi, X ve Y formunun aynı gizil yeteneği ölçtüğünü, tek boyutlu olduğunu ve iki form için de madde parametrelerinin aynı ölçek üzerinde olduğunu varsayar (T. Han, Kolen, & Pohlman, 1997). n madde içeren Form X için doğru puan sayısı eşitlik 2.32’de verilmiştir:

$$\varepsilon = \sum_{i=1}^n P(a_i, b_i, c_i, \theta)$$
(2.32)

m madde içeren Form Y için doğru puan sayısı eşitlik 2.33’teki gibidir:

$$\eta = \sum_{j=1}^m P(a_j, b_j, c_j, \theta)$$
(2.33)

Yetenek düzeyi θ_0 olduğunda, Form X için gerçek puan ε_0 ile gösterilir ve eşitlik 2.34’te verilmiştir:

$$\left(\sum_{i=1}^n P(a_i, b_i, c_i, \theta) \right) - \varepsilon_0 = 0$$
(2.34)

Yetenek düzeyi θ_0 olduğunda, Form Y için gerçek puan η_0 ile gösterilir ve eşitlik 2.35’teki gibi olur:

$$\left(\sum_{j=1}^m P(a_j, b_j, c_j, \theta) \right) - \eta_0 = 0$$
(2.35)

2.3.6. Test Eşitlemede Hata Kavramı

Bootstrap yöntemi, çeşitli istatistiklerin standart hatalarını tahmin etmek için kullanılan bir yöntemdir (Efron, 1982). Bootstrap yöntemi ile standart hatanın hesaplanması beş adımda gerçekleşmektedir (Kolen & Brennan, 2014).

1. N büyüklüğünde bir örneklem ile başlanır.
2. Bu örnek veriden N boyutunda çeşitli örnek veriler oluşturulur.
3. Her bir örneklem için ortalama hesaplanır.
4. Adım 2 ve 3 R kez tekrarlanır.
5. Örnek verilerin standart sapması hesaplanır. Bu standart sapma, istatistiğin tahmini bootstrap standart hatasıdır.

Bootstrap yöntemi, standart hatası matematiksel olarak anlaşılmaz olduğunda, standart hatayı tahmin etmek için uygulanabilen bir yeniden örnekleme yaklaşımıdır. Bu yöntem tipik olarak çok sayıda bootstrap örneği üzerinde tekrarlanan analizleri içerir. Bu örnekler, parametrik olmayan veya parametrik bir şekilde, orijinal veri kümesinden rastgele alınabilir veya oluşturulabilir (Zhang, 2020). Bu çalışmada sadece parametrik bootstrap yöntemi dikkate alınmıştır.

2.4. İlgili Araştırmalar

Bu bölümde incelenen çalışmalar, Türkiye’de yapılan çalışmalar ve Türkiye dışında yapılan çalışmalar olarak iki başlık altında ele alınmaktadır.

2.4.1. Türkiye’de Yapılan Çalışmalar

Şahhüseyinoğlu (2006) 2000, 2001 ve 2002 yıllarına ait İngilizce yeterlik puanlarını ardışık iki yıla ait puanlar şeklinde alarak bir çalışma yürütmüş. Yaptığı çalışmada doğrusal eşitleme, eşit yüzdelikli eşitleme ve Rasch yöntemlerini kullanmıştır. Çalışmasının ilk aşamasında betimsel istatistikleri hesaplamış, ikinci aşamasında Madde Tepki Kuramı’nın varsayımlarını kontrol etmiş, üçüncü aşamasında test eşitlemenin varsayımlarını kontrol ederek eşitleme işlemini yapmıştır. 2001 yılının puanlarını 2000 yılının puanlarına eşitlemiş ve Rasch yönteminin daha düşük hatayla eşitleme yaptığı sonucuna ulaşmıştır. 2002 yılının puanlarını 2001 yılına eşitlemiş ve eşit yüzdelikli eşitleme ile Rasch yönteminin doğrusal eşitleme yöntemine göre daha düşük hataya sahip olduğu sonucuna ulaşmıştır.

Kan (2010) 6. sınıfa giden 420 matematik öğrencisine matematik becerisini ölçen, 5 seçenekli 25 sorudan oluşan, sözel köklü ve sayısal köklü testler uygulayarak bir test eşitleme çalışması yürütmüştür. Çalışmasında tek grup düzeneği ve Klasik Test Kuramı'na dayalı yöntemlerden birisi olan doğrusal eşitleme yöntemini kullanmıştır. Test grup düzeneği için eşitleme hatasını (SHE) kestirmiş ve iki testin istatistiksel olarak eşit olmadığı, bir başka deyişle birbirleri yerine kullanılamayacağı sonucuna ulaşmıştır.

Kan (2011) çalışmasında 2003 ve 2005 yıllarında yapılan OKS Türkçe testini 1030 sekizinci sınıf öğrencisine uygulamıştır. İlk aşamada test eşitleme koşullarının sağlanıp sağlanmadığını kontrol etmiş ve sonra doğrusal eşitleme yöntemi kullanarak 2003 Türkçe alt test puanlarının 2005 Türkçe alt test puanlarına eşitlemiştir. İki testin istatistiksel olarak eşit olmadığı, birbirleri yerine kullanılmasının öğrencilerin performanslarında değişiklik oluşturabileceği sonucuna ulaşmıştır.

Öztürk ve Anıl (2012) 2008 yılının ilkbahar ve sonbahar dönemlerinde yapılan ALES'in sayısal testi puanlarının kullanarak bir test eşitleme çalışması yürütmüşlerdir. 2040 kişi ile yaptıkları çalışmada ilk aşamada testlerin betimsel istatistiklerini hesaplamış ve ikinci aşamada test eşitlemenin varsayımlarının kontrol etmişlerdir. Üçüncü aşamada doğrusal ve eşit yüzdelikli eşitleme yöntemlerini kullanarak test eşitleme işlemini yapmışlar ve son aşamada hata miktarlarını hesaplamışlardır. İki yöntemle yapılan eşitleme hatalarını karşılaştırmışlar ve eşit yüzdelikli eşitleme yönteminin daha düşük hatalı eşitleme yaptığı dolayısıyla bu yöntemin kullanılmasının daha doğru sonuçlar vereceği yargısına ulaşmışlardır.

Gök ve Kelecioğlu (2014) test uzunluğu, örneklem büyüklüğü, yetenek dağılımı ve model faktörlerinin yer aldığı simülasyon verinin kullanıldığı bir çalışma yürütmüştür. Çalışmalarında denk olmayan gruplarda ortak test deseni uygulayarak Madde Tepki Kuramı'na dayalı olan ortalama-ortalama, ortalama-standart sapma ve Stocking Lord yöntemlerini kullanarak karşılaştırma yapmışlardır. Test eşitleme işleminin ardından hata kestirimleri yapılmış ve a parametresi için en düşük hatanın ortalama-ortalama yönteminde, en yüksek hatanın ortalama-standart sapma yönteminde; b parametresi için en düşük hatanın ortalama-ortalama yönteminde, en yüksek hatanın ise Stocking Lord yönteminde olduğu sonucuna varmışlardır.

Gübeş Öztürk ve Kelecioğlu (2015) TIMSS 2011 Türkiye örnekleminde sınava giren 8. sınıf öğrencilerinin 5. ve 6. kitapçıkta yer alan matematik testlerine verdiği cevaplar ile test eşitleme çalışması yürütmüşlerdir. Çalışmalarında eşdeğer olmayan gruplarda ortak test deseni kullanmışlardır. MTK gerçek puan ve MTK gözlenen puan eşitleme yöntemleri ile yapılan

eşitlemeler eşitlik özelliğine göre karşılaştırılmıştır. Yapılan karşılaştırma sonucu MTK gerçek puan eşitleme yönteminin birinci sıra eşitlik özelliğini korumada daha iyi performans gösterdiği, MTK gözlenen puan eşitleme yönteminin ise ikinci sıra eşitlik özelliğini korumada daha iyi olduğu belirlenmiştir.

Pektaş ve Kılınç (2016) yaptıkları çalışmada PISA 2012 Türkiye örnekleminde 7 numaralı kitapçığı cevaplayan 370 kişi ve 6 numaralı kitapçığı cevaplayan 372 kişiden oluşan çalışma grubundan elde edilen verilerle ortak maddeli denk grup deseni kullanılarak gözlenen puan eşitleme yöntemlerini karşılaştırmayı amaçlamışlardır. Ortalama eşitleme, doğrusal eşitleme ve eşit yüzdelikli eşitleme yöntemleri kullanılarak test eşitleme işlemi yapılmış ve elde edilen puanların dağılımlarına yer verilmiştir. Yapılan test eşitleme işlemi sonucunda puan dağılımlarının en doğru şekilde eşitlenmesini gözlenen puan eşitleme yönteminin eşit yüzdelikli eşitleme yöntemi olduğu sonucuna ulaşılmıştır.

2.4.2. Türkiye Dışında Yapılan Çalışmalar

Kolen (1981) Iowa Tests of Educational Development verilerini kullanarak bir test eşitleme çalışması yürütmüştür. İki aşamalı lise sınavının formları için en iyi eşitleme yöntemini belirlemeyi amaçlamıştır. Aşama bir için eşit yüzdelikli eşitleme ve bir parametrelili gerçek puan ve modifiye edilmiş bir parametrelili gerçek puan yöntemini karşılaştırdığı zaman eşit yüzdelikli eşitlemenin daha iyi olduğu sonucuna varmıştır. Aşama iki için modifiye edilmiş bir parametrelili yöntem ile bir parametrelili yöntemin benzer sonuçlar verdiğini belirtmiştir. Üç parametrelili model için gözlenen puan yöntemi ile eşitleme yapılmıştır. Çalışmada kullanılan yöntemler karşılaştırıldığında bir parametrelili gerçek puan yönteminin yetersiz olduğu, üç parametrelili gözlenen puan yöntemi ve eşit yüzdelikli eşitleme yönteminin sorunlar olsa da daha uygun olduğu belirtilmiştir.

Jaeger (1981) çeşitli kolej ve üniversitelerden topladığı verilerle sekiz test formu oluşturarak bir çalışma yürütmüştür. Çalışmasında random gruplar desenini kullanmıştır. İlk aşamada kümülatif puan dağılımı ve Kolmogorov-Smirnov testine yer vermiştir. Çalışmasında doğrusal eşitleme ve eşit yüzdelikli eşitleme yöntemlerini kullanmış ve benzer sonuçlara varmıştır. Eşitleme işlemini üç adım olarak yapmış, madde güçlük ve ayırt edicilik indekslerinin dağılımlarını da incelemiştir.

Kolen ve Whitney (1982) GED testine ait on bir formun beş tanesi anchor test olmak üzere bir çalışmaya yürütmüşlerdir. Çalışma grubunu 1980 yılında testi alan iki yüz birey oluşturmuştur. Çalışmalarında bir parametrelili model, üç parametrelili model, doğrusal ve eşit

yüzdelikli eşitleme yöntemlerini karşılaştırmışlardır. Geleneksel yöntemler olan doğrusal ve eşit yüzdelikli eşitleme, Madde Tepki Kuramı'na dayalı yöntemler olan bir parametrelili ve üç parametrelili modeller ile test eşitleme işlemi yapılmıştır. Üç parametrelili model ve eşit yüzdelikli eşitleme yöntemleri ile daha büyük çalışma gruplarına ihtiyaç duyulduğu belirtilmiştir.

Angoff ve Cook (1988) sözel ve sayısal kısımları olan SAT ve PAA verileriyle bir test eşitleme çalışması yürütmüşlerdir. Orijinal dili İspanyolca olan bazı maddeler seçilerek uzmanlar tarafından İngilizceye, orijinal dili İngilizce olan maddeler ise İspanyolcaya çevrilerek bireylere verilmiştir. Çalışma ortak madde deseni ile yapılmıştır. İlk aşamada betimsel istatistikler ve madde parametrelerine yer verilmiştir. Çalışmada doğrusal eşitleme yöntemlerinden Tucker ve Levine, eğrisel eşitleme yöntemlerinden ise eşit yüzdelikli ve MTK eşitleme kullanılmıştır. Farklı dillerde verilen testlerin birbirine eşit olmadığı sonucuna ulaşmışlardır.

Little ve Rubin (1994) yaptıkları çalışmada askeriye giriş sınavı olan ASVAB'ın farklı formlarıyla test eşitleme çalışması yürütmüşlerdir. Çalışmalarında ilk aşamada betimsel istatistiklere ve madde yanlılığına ilişkin sonuçlara yer vermişlerdir. Olası yanlılık miktarının önemli olmadığı sonucuna ulaşmışlardır. İkinci aşamada doğrusal eşitleme ve eşit yüzdelikli eşitleme yöntemlerini kullanarak test eşitleme işlemi yapmışlardır. Eşit yüzdelikli eşitleme yönteminde hata miktarının daha fazla olması nedeniyle ön düzeltme yapılarak işlem tekrarlanmıştır. Düzeltme yapılan eşit yüzdelikli eşitleme ve doğrusal eşitleme yönteminin düzeltme yapılmayan eşit yüzdelikli eşitleme yöntemine göre daha kullanışlı olduğu sonucuna varmışlardır.

Von Davier ve Wilson (2007) eşdeğer olmayan gruplarda ortak madde deseni kullanarak bir çalışma yürütmüşlerdir. 2001 ve 2003 Calculus AB sınavının verileri kullanılmıştır. Sınav formlarında çoklu puanlanan ve açık puanlanan maddeler yer almıştır. Çalışmalarının amacı Madde Tepki Kuramı yöntemlerinin gerekliliklerini ve uygulamasını tartışmaktır. İlk aşamada varsayımların kontrolü için betimsel istatistikler, faktör analizi, hız testi olup olmadığının kontrolü ve DIF belirleme uygulamalarına yer verilmiştir. İkinci aşamada model-veri uyumu ve madde parametrelerine yer verilmiştir. Son aşamada test eşitleme işlemi yapılmış ve karışık formatlı formların çoklu puanlananlara göre daha iyi eşitlendiği sonucuna varılmıştır. MTK gözlenen puan ve MTK gerçek puan yöntemlerinin benzer olduğu belirtilmiştir.

BÖLÜM III

YÖNTEM

Bu bölümde, araştırmanın modeli, simülasyon koşulları, verilerin üretilmesi ve analizi başlıklarına yer verilmektedir.

3.1. Araştırmanın Modeli

Temel araştırma, bir kuram ya da hipotezin altında yatan süreçleri açıklamakla ilgilenen çalışma olarak tanımlanabilir (Fraenkel, Wallen, & Hyun, 2012). Bu çalışmada, farklı test uzunluğu ve örneklem büyüklüğüne sahip durumlar temel alınarak test eşitleme yöntemleri karşılaştırılmıştır. Yapılan çalışma sonucunda en az hatayı verecek olan yöntem belirlenmek ve kuramsal çalışmalara katkı sağlanmak istenmiştir. Dolayısıyla bu çalışma temel araştırma niteliği taşımaktadır. Eğitime önemli katkılar sağlayan simülasyon çalışmaları iki başlık altında toplanabilir: belirleyici ve olasılıksal simülasyon çalışmaları. Belirleyici simülasyon çalışmalarında bütün koşullar sabit tutulurken olasılıksal simülasyon çalışmalarında en az bir koşul değişkendir (Cohen, Manion, & Morrison, 2007). Buna göre bu çalışma ayrıca olasılıksal simülasyon çalışmasıdır.

3.2. Simülasyon Koşulları

Bu araştırma kapsamında yer alan simülasyon çalışmasında bazı koşullar sabit tutulurken bazıları değişimlenmiştir. Bu bölümde, bu koşullar değişimlenen koşullar ve sabit tutulan koşullar olarak iki başlık altında ele alınmıştır.

3.2.1. Değişimlenen Koşullar

Bu çalışmada değişimlenen iki koşul vardır: test uzunluğu ve örneklem büyüklüğü.

Örneklem büyüklüğü. Literatürde çeşitli örneklem koşullarında birçok test eşitleme çalışması yapılmıştır. Bu çalışmada ise Klasik Test Kuramı ve Madde Tepki Kuramı'na dayalı eşitleme yöntemleri karşılaştırılmak istendiği için 50, 100, 300, 1000 ve 3000 örneklem büyüklükleri seçilmiştir. Örneklem büyüklüklerinin geniş bir aralıkta olması amaçlanmıştır.

Test uzunluğu. İlgili literatür incelendiğinde test uzunluğunun çeşitli çalışmalarda yer aldığı özellikle ankor madde kullanılarak yapılan eşitlemelerde değişimlenen koşul olduğu görülmüştür. Bu çalışmada ise Türkiye’de yapılan YDS, ALES gibi sınavlar göz önünde bulundurularak testler 10, 20, 30, 40 ve 50 maddeden oluşacak şekilde üretilmiştir.

3.2.2. Sabit Tutulan Koşullar

Çalışmada kişilerin yetenek dağılımı, a ve b parametrelerinin dağılımı ve tekrar sayısı sabit tutulmuştur. Ayrıca tüm veri setleri genelleştirilmiş kısmi puan modeline göre üretilmiştir.

Yetenek dağılımı. Üretilen verilerin genel durumu yansıtması adına geniş bir aralık seçilmiştir. Yetenek parametreleri b parametresi ile uyumlu olarak -3 ve +3 arasında değişmektedir.

Madde ayırt edicilik parametresi. MTK’ye göre ayırt edicilik parametresi olan a parametresi 0.2’den düşük olduğu zaman madde testten atılmalıdır. Bu durum göz önünde bulundurularak a parametresi için en düşük değer 0.2 olarak belirlenmiştir. Yetenek ve b parametresi gibi a parametresinin de geniş bir aralıkta yer alması amaçlanarak üst sınır da pratik üst sınır olarak görülen 2.8 olarak alınmıştır.

Madde güçlük parametresi. MTK’ye göre madde güçlük parametresi olarak da bilinen b parametresi teorik olarak $-\infty$ ve $+\infty$ arasında değişmektedir. Pratikte ise -3 ve +3 olarak alınmaktadır. Bu durum ve verilerin geneli temsil etmesi amacı göz önünde bulundurularak b parametresi -3 ve +3 aralığında olacak şekilde üretilmiştir.

Tekrar sayısı. Yürütülen simülasyon çalışmalarında elde edilen bulguların genellenebilirliğini arttırmak için replikasyonlar yapılmaktadır. İlgili literatürde Madde Tepki Kuramı’na dayalı çalışmalarda 25 replikasyonun yeterli olduğu belirtilmektedir (Harwell, Stone, Hsu, & Kirisci, 1996). Bu çalışmada da replikasyon sayısı her alt koşul için 25 olarak belirlenmiştir.

Model. Literatür incelendiğinde ikili puanlanan ve çok kategorili puanlanan modelleri içeren birçok test eşitleme çalışması yürütüldüğü görülmektedir. Bu çalışmada da çok kategorili puanlanan modellerden olan genelleştirilmiş kısmi kredi modeline yer ver verilmiştir. Kısmi kredi modeli, başarı testleri için doğru ve yanlış cevabının yanısıra kısmen doğru puanlamasına izin vermektedir ancak a parametresi hesaplanmamaktadır. GPCM ise kısmen doğru puanlamasının olmasıyla beraber a parametresinin hesaplanıyor olması avantajına da sahiptir (de Ayala, 2009). Bu durumlar göz önünde bulundurularak bu çalışma için GPCM seçilmiştir.

Madde sayısına bağlı puanlama biçimi. Yürütülen bu çalışmada maddeye verilen yanıtın doğru ve yanlış olarak puanlandırılmasının yanısıra kısmen doğru olarak da puanlandırılabilmesi için üretilen yanıt örüntülerinde puanlar üç kategoriden oluşmaktadır. 0 yanlış cevabı, 1 kısmen doğru cevabı ve 2 doğru cevabı göstermektedir.

3.3. Verilerin Üretilmesi

Bu çalışmada verilerin üretilmesinde MTK temelinde çok kategorili puanlanan testler için oluşturulan genelleştirilmiş kısmi kredi modeli kullanılmıştır. Çalışma kapsamında kullanılan veriler, ücretsiz olarak ulaşılması nedeniyle WinGen3 programı kullanılarak üretilmiştir (K. T. Han, 2007). Çalışma kapsamında kullanılacak verilerin üretilmesi sürecinde değişimlenen koşullar Tablo 3.1’de gösterilmektedir.

Tablo 3.1. Sabit Tutulan Koşullar

	En küçük değer	En büyük değer
a parametresi	0.2	2.8
b parametresi	-3	3
c parametresi	-	-
Yetenek	-3	3
Tekrar Sayısı	25	
Model	Genelleştirilmiş Kısmi Kredi Modeli	
Kategori Sayısı	3	

Sabit tutulan koşullar Tablo 3.2’de gösterilmektedir.

Tablo 3.2. Değişimlenen Koşullar

	Değerler				
Örneklem	50	100	300	1000	3000
Büyüklüğü					
Test	10	20	30	40	50
Uzunluğu					

Değişimlenen koşullar göz önünde bulundurulduğunda; örneklem büyüklüğü koşulu (5 koşul) ve test uzunluğu koşulu (3 koşul) olmak üzere 25 koşul olmaktadır. Bu durumda üretilen her koşul için bir adet X formu ve bir adet Y formu olmak üzere toplam 50 test formu oluşturulmuştur. Çalışma verileri 25 tekrarla üretildiğinde 1250 adet cevap örüntüsü elde edilmiştir.

3.4. Verilerin Analizi

Bu araştırmada kullanılan veriler WinGen3 programı aracılığıyla üretilmiştir (K. T. Han & Hambleton, 2007). Verilerin üretilmesinde MTK temelinde çok kategorili puanlanan testler için oluşturulan genelleştirilmiş kısmi kredi modeli (GPCM) kullanılmıştır. Çalışma kapsamında üretilen verilerin KTK'ye dayalı eşitleme yöntemleri ile eşitlenmesinde R 4.1.2 programı (R core team, 2021) Equate paketi (Albano, 2016) kullanılmıştır. MTK'ye dayalı test eşitleme yöntemleri için ise IRTEQ programı kullanılmıştır. Eşitleme süreci boyunca verilerin düzenlenmesi ve bootstrap hata miktarlarının hesaplanması Microsoft Excel aracılığıyla yapılmıştır.

BÖLÜM IV

BULGULAR

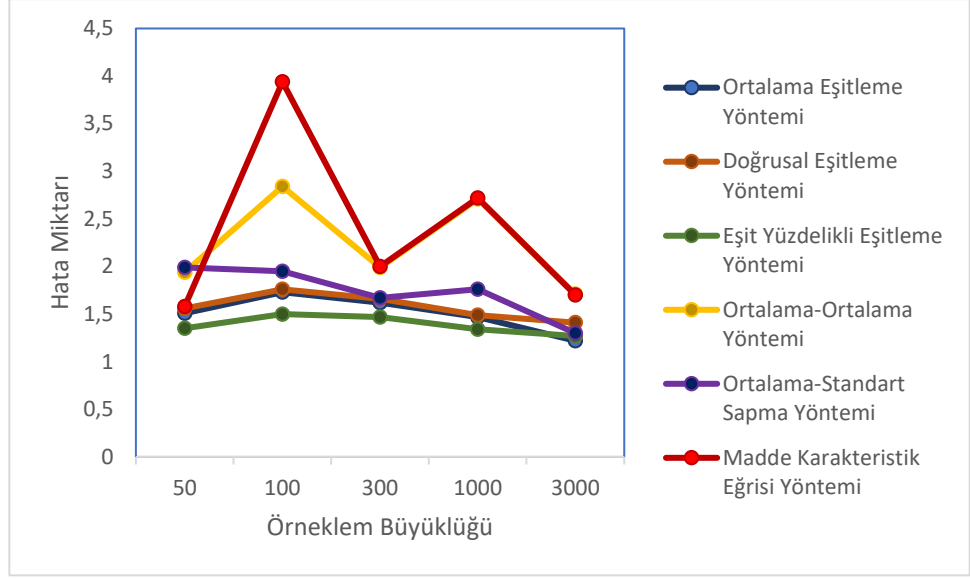
Farklı test uzunluğuna (10, 20, 30, 40 ve 50) maddelerden oluşan testler için hata miktarları tablo ve şekil olarak verilmiştir.

4.1. 10 Maddeye İlişkin Bulgular

Tablo 4.1. 10 Madde için Hata Miktarları

Örneklem büyüklüğü	ESH					
	O	D	E.Y.	O-O	O-S	TCC
50	1.51	1.56	1.35	1.94	1.99	1.58
100	1.73	1.76	1.50	2.84	1.95	3.94
300	1.62	1.66	1.47	1.99	1.67	2.00
1000	1.47	1.49	1.34	2.71	1.76	2.72
3000	1.42	1.41	1.27	1.71	1.30	1.70

Tablo 4.1 incelendiğinde, benzer madde parametreleri ve yetenek dağılımına sahip gruplarda, geliştirilmiş kısmi kredi modeline uyumlu ve 10 maddeden oluşan farklı test formlarının eşitlenmesinde kullanılabilecek en uygun test eşitleme yönteminin tüm örneklem büyüklükleri için ‘eşit yüzdelikli eşitleme’ yöntemi olduğu görülmektedir. En fazla hata içeren yöntem ise ‘test karakteristik eğrisi’ yöntemidir.



Şekil 4.1. 10 Madde için Hata Miktarları

10 madde için örneklem büyüklüğü arttıkça genel olarak eşit yüzdelikli eşitleme, ortalama, doğrusal ve ortalama-standart sapma yönteminden elde edilen hatada azalma görülürken madde karakteristik eğrisi yöntemleri ve ortalama-ortalama yönteminde daha yüksek hata ve hata miktarının yönünde değişkenlik gözlemlenmiştir.

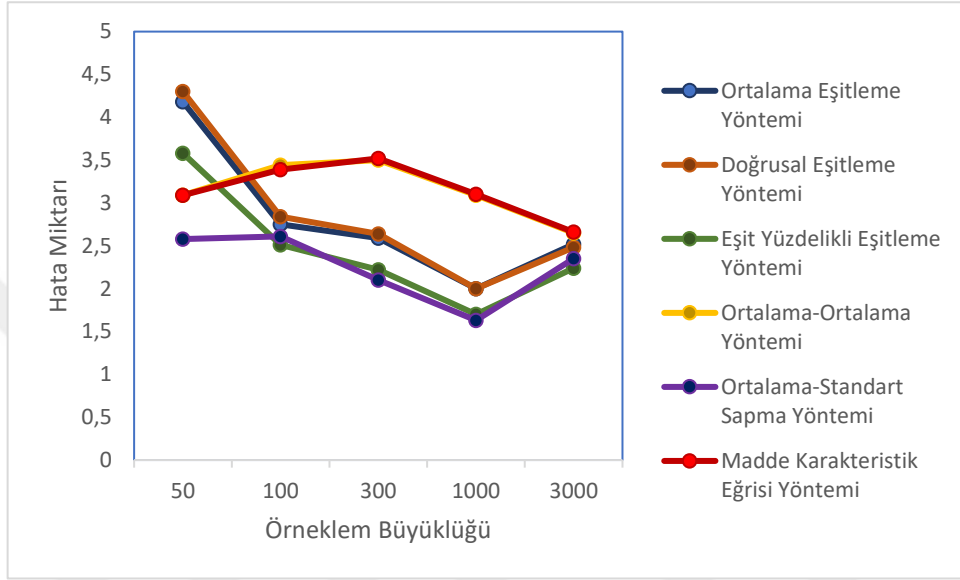
4.1. 20 Maddeye İlişkin Bulgular

Tablo 4.2. 20 Madde için Hata Miktarları

Örneklem büyüklüğü	ESH					
	O	D	E.Y.	O-O	O-S	TCC
50	4.18	4.30	3.58	3.09	2.58	3.09
100	2.75	2.84	2.51	3.44	2.61	3.39
300	2.59	2.64	2.22	3.50	2.10	3.52
1000	2.00	2.00	1.70	3.09	1.63	3.10
3000	2.52	2.48	2.24	2.65	2.35	2.66

Tablo 4.2 incelendiğinde, benzer madde parametreleri ve yetenek dağılımına sahip gruplarda, geliştirilmiş kısmi kredi modeline uyumlu ve 20 maddeden oluşan farklı test

formlarının eşitlenmesinde kullanılabilecek en uygun test eşitleme yönteminin 50 kişilik gruplarda ‘ortalama-standart sapma’, 100 kişilik gruplarda ‘eşit yüzdelikli eşitleme’, 300 kişilik gruplarda ‘ortalama-standart sapma’, 1000 kişilik gruplarda ‘ortalama-standart sapma’ ve 3000 kişilik gruplarda ‘eşit yüzdelikli eşitleme’ yöntemi olduğu görülmektedir. En fazla hata içeren yöntem ise genel olarak ‘test karakteristik eğrisi’ yöntemidir.



Şekil 4.2. 20 Madde için Hata Miktarları

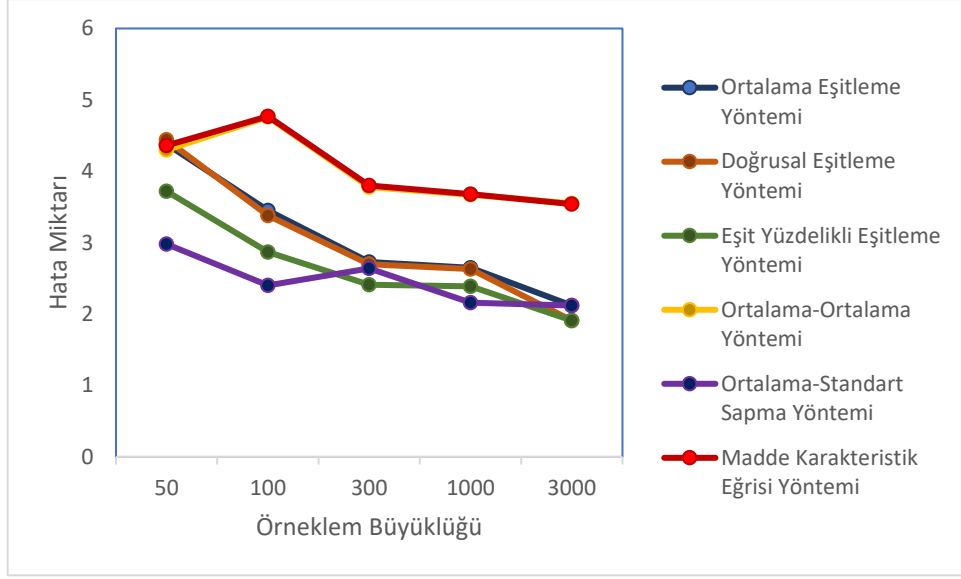
20 madde için örneklem büyüklüğü arttıkça genel olarak eşit yüzdelikli eşitleme, ortalama, doğrusal ve ortalama-standart sapma yönteminden elde edilen hatada azalma görülürken madde karakteristik eğrisi yöntemleri ve ortalama-ortalama yönteminde daha yüksek hata ve hata miktarının yönünde değişkenlik gözlemlenmiştir.

4.1. 30 Maddeye İlişkin Bulgular

Tablo 4.3. 30 Madde için Hata Miktarları

Örneklem büyüklüğü	ESH					
	O	D	E.Y.	O-O	O-S	TCC
50	4.38	4.44	3.72	4.39	2.98	4.36
100	3.45	3.38	2.87	4.76	2.40	4.77
300	2.73	2.70	2.41	3.78	2.64	3.80
1000	2.65	2.63	2.39	3.67	2.16	3.68
3000	2.12	2.13	1.91	3.55	2.12	3.54

Tablo 4.3 incelendiğinde, benzer madde parametreleri ve yetenek dağılımına sahip gruplarda, genelleştirilmiş kısmi kredi modeline uyumlu ve 30 maddeden oluşan farklı test formlarının eşitlenmesinde kullanılabilecek en uygun test eşitleme yönteminin 50 kişilik gruplarda ‘ortalama-standart sapma’, 100 kişilik gruplarda ‘ortalama-standart sapma’, 300 kişilik gruplarda ‘eşit yüzdellikli eşitleme’, 1000 kişilik gruplarda ‘ortalama-standart sapma’ ve 3000 kişilik gruplarda ‘eşit yüzdellikli eşitleme’ yöntemi olduğu görülmektedir. En fazla hata içeren yöntem ise genel olarak ‘test karakteristik eğrisi’ yöntemidir.



Şekil 4.3. 30 Madde için Hata Miktarları

30 madde için örneklem büyüklüğü arttıkça genel olarak eşit yüzdelikli eşitleme, ortalama, doğrusal ve ortalama-standart sapma yönteminden elde edilen hatada azalma görülürken madde karakteristik eğrisi yöntemleri ve ortalama-ortalama yönteminde daha yüksek hata ve hata miktarının yönünde değişkenlik gözlemlenmiştir.

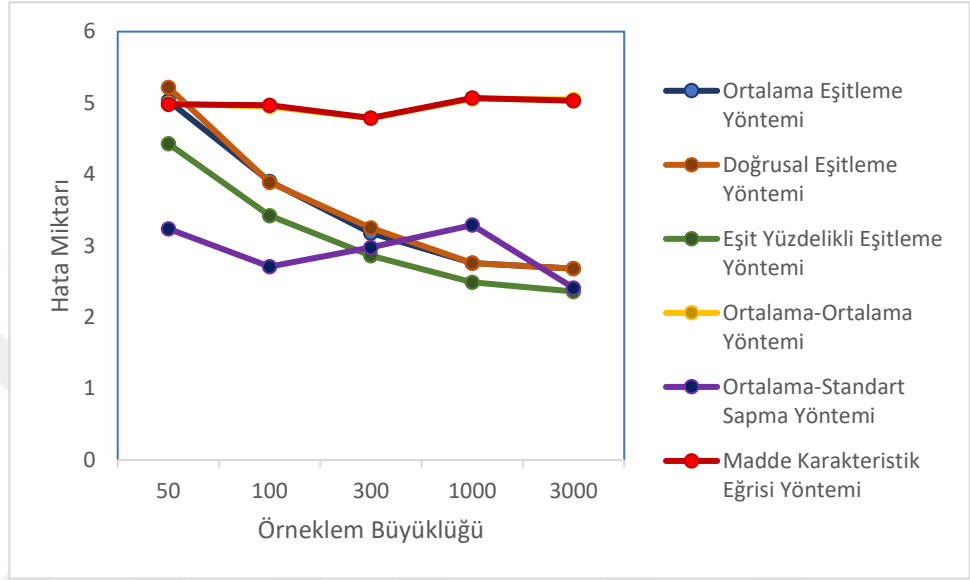
4.1. 40 Maddeye İlişkin Bulgular

Tablo 4.4. 40 Madde için Hata Miktarları

Örneklem büyüklüğü	ESH					
	O	D	E.Y.	O-O	O-S	TCC
50	5.03	5.22	4.43	4.99	3.24	4.98
100	3.90	3.89	3.42	4.95	2.71	4.97
300	3.18	3.25	2.86	4.79	2.98	4.79
1000	2.76	2.76	2.49	5.06	3.29	5.07
3000	2.68	2.68	2.36	5.05	2.41	5.03

Tablo 4.4 incelendiğinde, benzer madde parametreleri ve yetenek dağılımına sahip gruplarda, genelleştirilmiş kısmi kredi modeline uyumlu ve 40 maddeden oluşan farklı test

formlarının eşitlenmesinde kullanılabilecek en uygun test eşitleme yönteminin 50 kişilik gruplarda ‘ortalama-standart sapma’, 100 kişilik gruplarda ‘ortalama-standart sapma’, 300 kişilik gruplarda ‘eşit yüzdelikli eşitleme’, 1000 kişilik gruplarda ‘eşit yüzdelikli eşitleme’ ve 3000 kişilik gruplarda ‘eşit yüzdelikli eşitleme’ yöntemi olduğu görülmektedir. En fazla hata içeren yöntem ise genel olarak ‘test karakteristik eğrisi’ yöntemidir.



Şekil 4.4. 40 Madde için Hata Miktarları

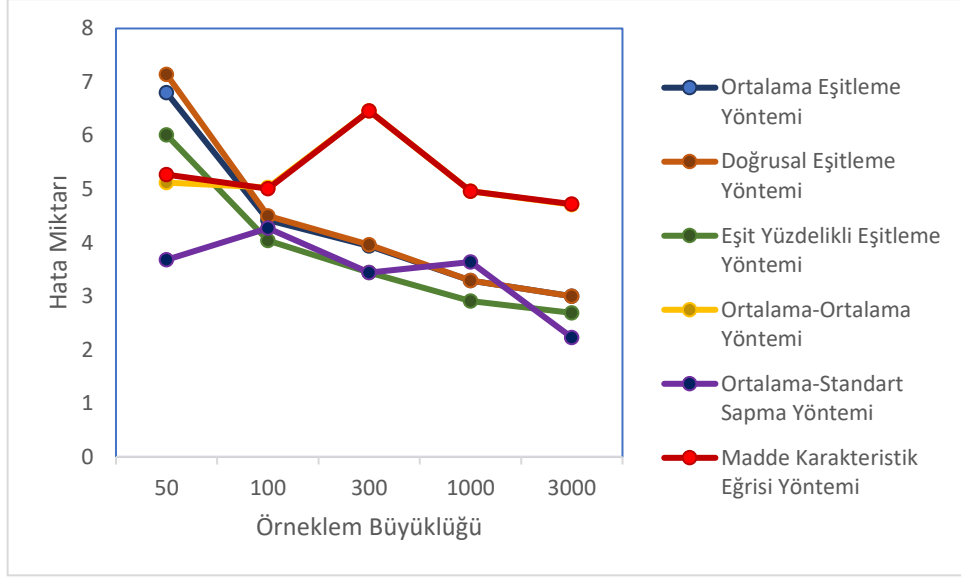
40 madde için örneklem büyüklüğü arttıkça genel olarak eşit yüzdelikli eşitleme, ortalama, doğrusal ve ortalama-standart sapma yönteminden elde edilen hatada azalma görülürken madde karakteristik eğrisi yöntemleri ve ortalama-ortalama yönteminde daha yüksek hata ve hata miktarının yönünde değişkenlik gözlemlenmiştir.

4.1. 50 Maddeye İlişkin Bulgular

Tablo 4.5. 50 Madde için Hata Miktarları

Örneklem büyüklüğü	ESH					
	O	D	E.Y.	O-O	O-S	TCC
50	6.80	7.14	6.01	5.12	3.68	5.27
100	4.42	4.50	4.04	5.03	4.27	5.01
300	3.94	3.96	3.44	6.46	3.44	6.46
1000	3.30	3.29	2.91	4.96	3.64	4.96
3000	2.98	3.00	2.69	4.71	2.23	4.72

Tablo 4.5 incelendiğinde, benzer madde parametreleri ve yetenek dağılımına sahip gruplarda, geliştirilmiş kısmi kredi modeline uyumlu ve 50 maddeden oluşan farklı test formlarının eşitlenmesinde kullanılacak en uygun test eşitleme yönteminin 50 kişilik gruplarda ‘ortalama-standart sapma’, 100 kişilik gruplarda ‘eşit yüzdelikli eşitleme’, 300 kişilik gruplarda ‘ortalama-standart sapma’ ve ‘eşit yüzdelikli eşitleme’, 1000 kişilik gruplarda ‘eşit yüzdelikli eşitleme’ ve 3000 kişilik gruplarda ‘ortalama-standart sapma’ yöntemi olduğu görülmektedir. En fazla hata içeren yöntem ise genel olarak ‘test karakteristik eğrisi’ yöntemidir.



Şekil 4.5. 50 Madde için Hata Miktarları

50 madde için örneklem büyüklüğü arttıkça genel olarak eşit yüzdelikli eşitleme, ortalama, doğrusal ve ortalama-standart sapma yönteminden elde edilen hatada azalma görülürken madde karakteristik eğrisi yöntemleri ve ortalama-ortalama yönteminde daha yüksek hata ve hata miktarının yönünde değişkenlik gözlemlenmiştir.

BÖLÜM V

SONUÇ, TARTIŞMA VE ÖNERİLER

Bu bölümde, araştırmadan elde edilen sonuçlara ve bu sonuçlara dayalı olarak sunulabilecek önerilere yer verilmiştir.

5.1. Sonuç ve Tartışma

Bu araştırmada farklı örneklem büyüklüğü ve test uzunluğu koşullarında, Klasik Test Kuramı'na dayalı ortalama, doğrusal, eşit yüzdelikli eşitleme yöntemleri ve Madde Tepki Kuramı'na dayalı ortalama-ortalama, ortalama-standart sapma, Haebara ve Stocking-Lord yöntemlerinin karşılaştırılması amaçlanmıştır. Haebara ve Stocking-Lord yöntemlerinin A ve B sabitlerinin eşit olması nedeniyle bu yöntemler test karakteristik eğrisi yöntemleri başlığı altında beraber verilmiş ve yorumlanmıştır. Bu araştırmanın sonuçları belirtilen koşullar altında aşağıda verilmiştir:

10 maddeden oluşan testlerin eşitlenmesinden elde edilen hatalar incelendiğinde, 50 örneklem büyüklüğü için en düşük hata eşit yüzdelikli eşitleme yönteminden ve en yüksek hata ise ortalama-standart sapma yönteminden elde edilmiştir. 100 örneklem büyüklüğü için tüm yöntemlerde hata miktarının arttığı, genel olarak ise en düşük hatanın eşit yüzdelikli eşitleme yönteminden ve en yüksek hatanın ise test karakteristik eğrisi yöntemlerinden elde edildiği görülmüştür. 300, 1000 ve 3000 örneklem büyüklükleri için en düşük hata eşit yüzdelikli eşitleme yönteminden ve en yüksek hata ise test karakteristik eğrisi yöntemlerinden elde edilmiştir. Klasik Test Kuramı'na dayalı ortalama, doğrusal ve eşit yüzdelikli eşitleme yöntemlerinde örneklem büyüklüğü sırasıyla 300, 1000 ve 3000 olduğunda hata miktarlarının azaldığı gözlemlenirken, Madde Tepki Kuramı'na dayalı ortalama-ortalama, ortalama-standart sapma ve test karakteristik eğrisi yöntemlerinden elde edilen hata miktarlarının önce artma sonra azalma eğiliminde oldukları gözlemlenmiştir.

20 maddeden oluşan testlerin eşitlenmesinden elde edilen hatalar incelendiğinde, 50 örneklem büyüklüğü için en düşük hata ortalama-standart sapma yönteminden ve en yüksek hata ise doğrusal eşitleme yönteminden elde edilmiştir. 100 örneklem büyüklüğü için en düşük

hatanın eşit yüzdelikli eşitleme yönteminden ve en yüksek hatanın ise ortalama-ortalama yönteminden elde edildiği görülmüştür. 300 örneklem büyüklüğü için en düşük hata ortalama-standart sapma yönteminde en yüksek hata ise test karakteristik eğrisi yöntemlerinde gözlemlenmiştir. 1000 örneklem büyüklüğünde yöntemlerden elde edilen hata miktarları örneklem büyüklüğü 300 olduğunda elde edilen hata miktarları ile benzerlik göstermektedir. 3000 örneklem büyüklüğü için en düşük hata eşit yüzdelikli eşitleme yönteminden ve en yüksek hata miktarı ise test karakteristik eğrisi yöntemlerinde gözlemlenmiştir. Genel olarak en düşük hatanın 20 madde için de eşit yüzdelikli eşitleme yönteminden elde edildiği söylenebilir.

30 maddeden oluşan testlerin eşitlenmesinden elde edilen hatalar incelendiğinde, 50 örneklem büyüklüğü için en düşük hata ortalama-standart sapma yönteminden ve en yüksek hata ise doğrusal eşitleme yönteminden elde edilmiştir. 100 örneklem büyüklüğü için en düşük hatanın ortalama-standart sapma yönteminden ve en yüksek hatanın ise test karakteristik eğrisi yöntemlerinden elde edildiği görülmüştür. 300 örneklem büyüklüğü için en düşük hata eşit yüzdelikli eşitleme yönteminde en yüksek hata ise test karakteristik eğrisi yöntemlerinde gözlemlenmiştir. 1000 örneklem büyüklüğü için en düşük hatanın ortalama-standart sapma yönteminden ve en yüksek hatanın ise test karakteristik eğrisi yöntemlerinden elde edildiği görülmüştür. 3000 örneklem büyüklüğü için en düşük hata ortalama ve ortalama-standart sapma yöntemlerinde ve en yüksek hata miktarı ise ortalama-ortalama yöntemlerinde gözlemlenmiştir. 3000 örneklem büyüklüğü için ortalama ve ortalama-ortalama yöntemi doğrusal eşitleme yöntemi ile çok yakın bir sonuç verirken ortalama-ortalama ve test karakteristik eğrisi yöntemlerinden de yakın sonuçlar elde edilmiştir. Genel olarak en düşük hatanın 30 madde için de ortalama-standart sapma eşitleme yönteminden elde edildiği söylenebilir.

40 maddeden oluşan testlerin eşitlenmesinden elde edilen hatalar incelendiğinde, 50 örneklem büyüklüğü için en düşük hata ortalama-standart sapma yönteminden ve en yüksek hata ise doğrusal eşitleme yönteminden elde edilmiştir. 100 örneklem büyüklüğü için en düşük hatanın ortalama-standart sapma yönteminden ve en yüksek hatanın ise ortalama-ortalama ve test karakteristik eğrisi yöntemlerinden elde edildiği görülmüştür. 300 örneklem büyüklüğü için en düşük hata eşit yüzdelikli eşitleme yönteminde en yüksek hata ise ortalama-ortalama ve test karakteristik eğrisi yöntemlerinde gözlemlenmiştir. 1000 örneklem büyüklüğü için en düşük hatanın ortalama-standart sapma yönteminden ve en yüksek hatanın ise test karakteristik eğrisi yöntemlerinden elde edildiği görülmüştür. 3000 örneklem büyüklüğü için en düşük hatanın ortalama-standart sapma yönteminde ve en yüksek hatanın ise test karakteristik eğrisi

yöntemlerinde gözlemlenmiştir. Genel olarak en düşük hatanın 40 madde için de ortalama-standart sapma eşitleme yönteminden elde edildiği söylenebilir.

50 maddeden oluşan testlerin eşitlenmesinden elde edilen hatalar incelendiğinde, 50 örneklem büyüklüğü için en düşük hata ortalama-standart sapma yönteminden ve en yüksek hata ise doğrusal eşitleme yönteminden elde edilmiştir. 100 örneklem büyüklüğü için en düşük hatanın eşit yüzdelikli eşitleme yönteminden ve en yüksek hatanın ise ortalama-ortalama ve test karakteristik eğrisi yöntemlerinden elde edildiği görülmüştür. 300 örneklem büyüklüğü için en düşük hata eşit yüzdelikli eşitleme ve ortalama-standart sapma yöntemlerinde en yüksek hata ise ortalama-ortalama ve test karakteristik eğrisi yöntemlerinde gözlemlenmiştir. 1000 örneklem büyüklüğü için en düşük hatanın eşit yüzdelikli eşitleme yönteminden ve en yüksek hatanın ise ortalama-ortalama ve test karakteristik eğrisi yöntemlerinden elde edildiği görülmüştür. 3000 örneklem büyüklüğü için en düşük hata ortalama ve ortalama-standart sapma yöntemlerinde ve en yüksek hata miktarı ise ortalama-ortalama ve test karakteristik eğrisi yöntemlerinde gözlemlenmiştir. Genel olarak en düşük hatanın 50 madde için de ortalama-standart sapma eşitleme yönteminden elde edildiği söylenebilir.

Literatürde test eşitleme süreci için yöntemlerin karşılaştırıldığı benzer çalışmalar yer almaktadır. Kolen (1981) KTK'ye dayalı eşitleme yöntemi olan eşit yüzdelikli eşitleme ile MTK'ye dayalı 1PL ve 3PL yöntemlerini karşılaştırmış ve eşit yüzdelikli ve 3PL eşitleme yönteminin 1PL'den daha iyi olduğunu belirtmiştir. Kolen ve Whitney (1982) benzer şekilde eşit yüzdelikli eşitlemenin daha iyi olduğu sonucuna ulaşmışlardır. Şahhüseyinoğlu (2006) 3 yıla ait İngilizce puanlarını doğrusal, eşit yüzdelikli ve Rasch eşitleme yöntemine göre eşitlemiş ve eşit yüzdelikli eşitleme yönteminin Rasch kadar güvenilir bir sonuç verdiğine ulaşmıştır. Kilmen ve Demirtaşlı (2012) farklı örneklem büyüklüğüne sahip gruplarda MTK'ye dayalı test eşitleme yöntemlerini karşılaştırmış ve test karakteristik eğrisi yöntemlerinin daha az hata eşitleme süreci yürüttüğü ayrıca örneklem büyüklüğünün artmasının da hesaplanan hata miktarının azalmasını sağladığı sonucuna ulaşmışlardır. Gök ve Kelecioğlu (2014) yaptıkları çalışmada test uzunluğu ve örneklem büyüklüğü koşullarına bağlı olarak test eşitleme süreci yürütmüşlerdir. Ortalama-ortalama yönteminin daha az hata ile eşitleme yaptığı en fazla hata veren yöntemin ortalama-standart sapma yöntemi olduğunu gözlemlemişlerdir. Madde sayısının ve örneklem büyüklüğünün arttığı durumlarda daha kararlı eşitlemeler yapıldığı sonucuna ulaşmışlardır. Uysal ve Kilmen (2016) 3PL ve GPCM'ye göre üretilen veriler ile test eşitleme çalışması yürütmüşlerdir. Test karakteristik eğrisi yönteminin daha az hata ile eşitleme yaptığı sonucuna ulaşmışlardır. Nisa ve Retnawati (2018) yaptıkları test eşitleme

alışmasına MTK'ye dayalı yöntemleri karşılaştırmışlar ve en az hata ile eşitleme yapan yöntemin ortalama-ortalama yöntemi olduğunu gözlemlemişlerdir.

İlgili literatür incelendiğinde KTK'ye ve MTK'ye dayalı eşitleme yöntemlerinin karşılaştırıldığı ve genellikle daha az hata ile eşitleme yapan yöntemin KTK yöntemlerinden eşit yüzdelikli eşitleme ve MTK yöntemlerinden de ortalama-ortalama ve test karakteristik eğrisi yöntemleri olduğu görülmüştür. Bu çalışmada da eşit yüzdelikli eşitlemenin daha az hata ile eşitleme yaptığı sonucuna ulaşılmıştır. Ancak ortalama-ortalama ve test karakteristik eğrisi ile yapılan eşitlemelerde diğer yöntemlere göre hesaplanan hata miktarı daha fazladır. Örneklem büyüklüğünün değişimlenen koşul olarak alındığı çalışmalarda örneklem büyüklüğü arttıkça hesaplanan hata miktarının da azaldığı sonucuna ulaşılmıştır. Bu çalışmada da benzer bir sonuç elde edilmiştir. Test uzunluğunun değişimlenen koşul olarak ele alındığı çalışmalarda madde sayısı arttıkça hata miktarının azaldığı gözlemlenirken bu çalışmada tam tersi bir sonuca ulaşılmıştır.

5.2. Öneriler

Bu araştırmadan elde edilen bulgulara dayanarak, pratikte test eşitleme çalışmaları için KTK'ye dayalı yöntemlerden eşit yüzdelikli eşitleme yöntemi ve MTK'ye dayalı yöntemlerden ortalama-standart sapma yönteminin kullanılması önerilebilir. Bundan sonra test eşitleme alanında yapılabilecek araştırmalar için ise aşağıdaki öneriler sunulabilir:

- Bu çalışma bilgisayar ortamında üretilmiş olan simülatif veriler ile yapılmıştır. Aynı çalışma gerçek bir test uygulamasından elde edilen veriler ile yürütülebilir.
- Bu çalışmada KTK'ye dayalı test eşitleme yöntemlerinden doğrusal, ortalama ve eşit yüzdelikli eşitleme; MTK'ye dayalı test eşitleme yöntemlerinden ise ortalama-ortalama, ortalama-standart sapma ve test karakteristik eğrisi yöntemleri ile yürütülmüştür. Benzer bir çalışmada daha farklı test eşitleme yöntemlerinden faydalanılabilir.
- Bu çalışmada veriler MTK'ye dayalı genelleştirilmiş kısmi puan modeli seçilerek yürütülmüştür. Aynı çalışma farklı bir model kullanılarak yapılabilir.
- Bu çalışmada veriler MTK'ye dayalı GPCM kullanılarak 3 kategori şeklinde yürütülmüştür. Çalışmada kullanılan yöntemler çeşitli kategori sayıları temel alınarak karşılaştırılabilir.
- Bu çalışmada madde ve yetenek parametreleri için geniş bir aralık tercih edilmiştir. Çeşitli aralıklar alınarak test eşitleme çalışması yürütülebilir.

- Bu çalışmada hata kestirim yöntemlerinden bootstrap yöntemi kullanılmıştır. Çeşitli eşitleme süreçleri için hata yöntemleri karşılaştırılabilir.
- Bu çalışmada eşit yüzdelikli eşitleme yöntemi için bir düzgünleştirme işlemi yapılmamıştır. Ön-düzgünleştirme ve son-düzgünleştirme yapılarak hata miktarları karşılaştırılabilir.



KAYNAKÇA

- Albano, A. D. (2016). {equate}: An {R} package for observed-score linking and equating. *Journal of Statistical Software*, 74(8), 1-36. doi:10.18637/jss.v074.i08
- Anastasi, A. (1976). *Psychological testing* (4. Baskı). London: Collier Macmillan Publishers.
- Angoff, W. H., & Cook, L. L. (1988). Equating the scores of the prueba de aptitude académica and the scholastic aptitude test. *College Board Report*, 88(2).
- Baker, F. B. (2001). *The basics of item response theory* (Cilt 2). USA: ERIC Clearinghouse on Assessment and Evaluation.
- Baykul, Y. (2015). Klasik test teorisiyle ilgili temel kavramlar. İçinde *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması* (s. 92-107). Ankara: Pegem Akademi.
- Cohen, L., Manion, L., & Morrison, K. (2007). Internet-based research and computer usage. İçinde *Research Methods in Education* (6. Baskı, s. 226-252). New York: Routledge.
- Crocker, L., & Algina, J. (2006a). Equating scores from different tests. İçinde *Introduction to Classical and Modern Test Theory* (s. 456-481). Ohio: Cengage Learning.
- Crocker, L., & Algina, J. (2006b). Introduction to item response theory. İçinde *Introduction to Classical and Modern Test Theory* (s. 339-375). Mason, Ohio Cengage Learning.
- de Ayala, R. J. (2009). *The theory and practice of item response theory*. New York: The Guilford Press.
- DeMars, C. (2010). *Item response theory*. New York: Oxford.
- Dorans, N. J., Moses, T. P., & Eignor, D. R. (2010). *Principles and practices of test score equating*. Retrieved from New Jersey:

- Dorans, N. J., Moses, T. P., & Eignor, D. R. (2011). Equating test scores: toward best practices. İçinde A. A. von Davier (Ed.), *Statistical models for test equating, scaling, and linking* (s. 21-42). New York: Springer.
- Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Retrieved from Philadelphia:
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahway, New Jersey: Lawrence Erlbaum Associates, Inc.
- Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2012). The nature of research. İçinde *How to design and evaluate research in education* (8. Baskı). New York, NY: McGraw-Hill Education.
- Gonzalez, J. (2014). SNSequate: standard and nonstandard statistical models and methods for test equating. *Journal of Statistical Software*, 59(7). Retrieved from <http://www.jstatsoft.org/>
- Gök, B., & Kelecioğlu, H. (2014). Denk olmayan gruplarda ortak madde deseni kullanılarak madde tepki kuramına dayalı eşitleme yöntemlerinin karşılaştırılması. *Mersin Üniversitesi Eğitim Fakültesi Dergisi*, 10(1), 120-136.
- Guilera, G., & Gomez, J. (2008). Item Response Theory test equating in health sciences education. *Advances in Health Sciences Education* 13, 3-10. doi:10.1007/s10459-006-9020-8
- Gübeş Öztürk, N., & Kelecioğlu, H. (2015). Karma testlerin eşitlenmesinde MTK eşitleme yöntemlerinin eşitlik özelliği korunumu ölçütüne göre karşılaştırılması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 6(1), 117-125.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. London: Sage Publications.
- Han, K. T. (2007). WinGen: Windows software that generates IRT parameters and item responses. *Applied Psychological Measurement*, 31(5), 457-459.

- Han, K. T., & Hambleton, R. K. (2007). *User's manual: WinGen* (642). Retrieved from Amherst, MA:University of Massachusetts:
- Han, T., Kolen, M. J., & Pohlman, J. (1997). A comparison among IRT true and observedscore equatings and traditional equipercentile equating. *Applied Measurement in Education*, 10(2), 105-121. doi:10.1207/s15324818ame1002_1
- Hanson, B. A., Zeng, L., & Colton, D. (1994). *A comparison of presmoothing and postsmoothing methods in equipercentile equating*. Retrieved from Iowa City:
- Harwell, M., Stone, C. A., Hsu, T., & Kirisci, L. (1996). Monte Carlo studies in Item Response Theory. *Applied Psychological Measurement*, 20(2), 101-125.
- Jaeger, R. M. (1981). Some exploratory indices for selection of a test equating method. *Journal of Educational Measurement*, 18(1), 23-38.
- Kan, A. (2010). Test eşitleme : aynı davranışları ölçen, farklı madde formlarına sahip testlerin istatistiksel eşitliğinin sınanması. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi*, 1(1), 16-21.
- Kan, A. (2011). Test eşitleme: OKS testlerinin istatistiksel eşitliğinin sınanması. *Eğitim ve Bilim*, 36(160), 38-51.
- Kilmen, S., & Demirtaşlı, N. (2012). Comparison of test equating methods based on item response theory according to the sample size and ability distribution *Procedia - Social and Behavioral Sciences*, 46, 130-134. doi: 10.1016/j.sbspro.2012.05.081
- Koğar, H. (2020). *R ile Geçerlik ve Güvenirlik Analizleri*. Ankara: Pegem Akademi.
- Kolen, M. J. (1981). Comparison of traditional and item response theory methods for equating tests. *Journal of Educational Measurement Spring*, 18(1), 1-11.
- Kolen, M. J., & Brennan, R. L. (2014). *Test equating, scaling and linking: methods and practices* (3. Baskı). New York: Springer.

- Kolen, M. J., & Jarjoura, D. (1987). Analytic smoothing for equipercentile equating under the common item nonequivalent populations design. *Psychometrika*, 52(1), 43-59.
- Kolen, M. J., & Whitney, D. R. (1982). Comparison of four procedures for equating the tests of general educational development. *Journal of Educational Measurement*, 19(4), 279-293.
- Linn, L. L. (1993). Linking results of distinct assessments. *Applied Measurement in Education*, 6(1), 83-102. doi:10.1207/s15324818ame0601_5
- Little, R. J. A., & Rubin, D. B. (1994). Test equating from biased samples, with application to the Armed Services vocational aptitude battery. *Journal of Educational and Behavioral Statistics*, 19(4), 309-335.
- Livingston, S. A. (2014). Equipercentile equating. İçinde *Equating Test Scores (Without IRT)* (2. Baskı, s. 17-23): Educational Testing Service.
- Livingston, S. A., & Kim, S. (2009). The circle-arc method for equating in small samples. *Journal of Educational Measurement*, 46(3), 330-343.
- Loyd, B. H., & Hoover, H. D. (1980). Vertical equating using the Rasch model. *Journal of Educational Measurement*, 17(3), 179-193. Retrieved from <https://www.jstor.org/stable/1434833>
- Mislevy, R. J. (1992). *Linking educational assessments: concepts, issues, methods and prospects*. Retrieved from Princeton,NJ:
- Muraki, E., Hombo, C. M., & Lee, Y. (2000). Equating and linking of performance assessments. *Psychological Measurement*, 24(4), 325-337.
- Nisa, C., & Retnawati, H. (2018). Comparing the methods of vertical equating for the math learning achievement tests for junior high school students. *REiD (Research and Evaluation in Education)*, 4(2), 164-174.

- Öztürk, N., & Anıl, D. (2012). Akademik personel ve lisansüstü eğitimi giriş sınavı puanlarının eşitlenmesi üzerine bir çalışma. *Eğitim ve Bilim*, 37(165), 180-193.
- Paek, I., & Cole, K. (2020). *Using R for item response theory model applications*. London: Routledge.
- Peabody, M. R. (2020). Some methods and evaluation for linking and equating with small samples. *Applied Measurement İçinde Education*, 33(1), 3-9. doi:10.1080/08957347.2019.1674304
- Pektaş, S., & Kılınç, M. (2016). PISA 2012 matematik testlerinden iki kitapçığın gözlenen puan eşitleme yöntemleri ile eşitlenmesi *Mehmet Akif Ersoy Üniversitesi Eğitim Fakültesi Dergisi* (40), 432-444. doi:10.21764/efd.49376
- Petersen, N. S., & ACT. (2008). A discussion of population invariance of equating. *Applied Psychological Measurement*, 32(1), 98-101. doi: 10.1177/0146621607311581
- Price, L. R. (2017). Norms and test equating. İçinde T. D. Little (Ed.), *Psychometric Methods: Theory into Practise* (s. 407-449). New York: The Guilford Press.
- R core team. (2021). R: a language and environment for statistical computing. Retrieved from <https://www.R-project.org/>
- Sansivieri, V., Wiberg, M., & Matteucci, M. (2017). A review of test equating methods with a special focus on IRT-based approaches. *Statistica*, 77(4), 329-352.
- Sünbül, Ö. (2020). Madde tepki kuramı (MTK). İçinde *Psikolojide Ölçme ve Ölçek Geliştirme II: Ölçme Araçlarının Psikometrik Nitelikleri ve Ölçme Kuramları* (pp. 122-141). Ankara: Pegem Akademi.
- Şahhüseyinoğlu, D. (2006). İngilizce yeterlik sınavı puanlarının üç farklı eşitleme yöntemine göre karşılaştırılması. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*(31), 115-125.

- Uysal, İ., & Kilmen, S. (2016). Comparison of item response theory test equating methods for mixed format tests. *International Online Journal of Educational Sciences*, 8(2), 1-11.
- van der Linden, W. J., & Wiberg, M. (2010). Local observed-score equating with anchor-test designs. *Applied Psychological Measurement*, 34(8), 620-640. doi:10.1177/0146621609349803
- von Davier, A. A. (2011). A statistical perspective on equating test scores. İçinde A. A. von Davier (Ed.), *Statistical Models for Test Equating, Scaling, and Linking* (s. 1-20). New York: Springer.
- von Davier, A. A., Holland, P. W., & Thayer, D. T. (2004). The Kernel method of test equating: theory. İçinde *The Kernel Method of Test Equating* (s. 17-94). New York: Springer.
- von Davier, A. A., & Wilson, C. (2007). IRT true-score test equating: a guide through assumptions and applications. *Educational and Psychological Measurement*, 67(6), 940-957.
- von Davier, A. A., Wilson, C., & Educational Testing Service. (2008). Investigating the population sensitivity assumption of Item Response Theory true-score equating across two subgroups of examinees and two test formats. *Applied Psychological Measurement*, 32(1), 11-26. doi:10.1177/0146621607311560
- Zhang, Z. (2020). Estimating standard errors of IRT true score equating coefficients using imputed item parameters. *The Journal of Experimental Education*. doi:10.1080/00220973.2020.1751579
- Zhu, W. (1998). Test equating: What, why, how? *Research Quarterly for Exercise and Sport*, 69(1), 11-23. doi:10.1080/02701367.1998.10607662

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı Soyadı: [REDACTED]

Doğum Yeri ve Tarihi: [REDACTED]

Eğitim Durumu

Lisans Öğrenimi: Akdeniz Üniversitesi/Sınıf Öğretmenliği

Yüksek Lisans Öğrenimi: Akdeniz Üniversitesi/Eğitimde Ölçme ve Değerlendirme Tezli

Yüksek Lisans

Bildiği Yabancı Diller: İngilizce

İletişim

E-Posta Adresi: [REDACTED]

Tarih: 27.06.2022

İNTİHAL RAPORU

ÇOK KATEGORİLİ PUANLANAN MADDELERDEN OLUŞAN TESTLERDE KLASİK TEST KURAMI VE MADDE TEPKİ KURAMI'NA DAYALI TEST EŞİTLEME YÖNTEMLERİNİN KARŞILAŞTIRILMASI

ORJİNALLİK RAPORU

% 15	% 14	% 4	% 4
BENZERLİK ENDEKSİ	İNTERNET KAYNAKLARI	YAYINLAR	ÖĞRENCİ ÖDEVLERİ

BİRİNCİL KAYNAKLAR

1	acikbilim.yok.gov.tr İnternet Kaynağı	% 3
2	www.openaccess.hacettepe.edu.tr:8080 İnternet Kaynağı	% 3
3	acikarsiv.ankara.edu.tr İnternet Kaynağı	% 1
4	dergipark.ulakbim.gov.tr İnternet Kaynağı	% 1
5	dspace.gazi.edu.tr İnternet Kaynağı	% 1
6	docplayer.biz.tr İnternet Kaynağı	% 1
7	Submitted to Akdeniz University Öğrenci Ödevi	% 1
8	hdl.handle.net İnternet Kaynağı	<% 1

openaccess.hacettepe.edu.tr:8080

BİLDİRİM

Hazırladığım tezin/raporun tamamen kendi çalışmam olduğunu ve her alıntıya kaynak gösterdiğimi taahhüt eder, tezimin/raporumun kâğıt ve elektronik kopyalarının Akdeniz Üniversitesi Eğitim Bilimleri Enstitüsü arşivlerinde aşağıda belirttiğim koşullarda saklanmasına izin verdiğimi onaylarım:

☐ Tezimin/Raporumun tamamı her yerden erişime açılabilir.

☐ Tezim/Raporum sadece Akdeniz Üniversitesi yerleşkelerinden erişime açılabilir. Tezimin/Raporumun yıl süreyle erişime açılmasını istemiyorum. Bu sürenin sonunda uzatma için başvuruda bulunmadığım takdirde, tezimin/raporumun tamamı her yerden erişime açılabilir.

28/06/2022

Merve Çörtük