ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME PROGRAMI

KÜÇÜK ÖRNEKLEMLERDE ÇOK KATEGORİLİ PUANLANAN MADDELERDEN OLUŞAN TESTLERDE KLASİK TEST EŞİTLEME YÖNTEMLERİNİN KARŞILAŞTIRILMASI

YÜKSEK LİSANS TEZİ

ALİ EMRE KARAGÜL

ANKARA
OCAK, 2020


ANKARA ÜNİVERSİTESİ
1946

ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ

EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME PROGRAMI

KÜÇÜK ÖRNEKLEMLERDE ÇOK KATEGORİLİ PUANLANAN MADDELERDEN OLUŞAN TESTLERDE KLASİK TEST EŞİTLEME YÖNTEMLERİNİN KARŞILAŞTIRILMASI

YÜKSEK LİSANS TEZİ

ALİ EMRE KARAGÜL

DANIŞMANI: DOÇ. DR. ERGÜL DEMİR

ANKARA
OCAK, 2020

ONAY

Ankara Üniversitesi Eğitim Bilimleri Enstitüsü Müdürlüğüne,

Ali Emre KARAGÜL adlı öğrencinin hazırladığı "**Küçük Örneklemlerde Çok Kategorili Puanlanan Maddelerden Oluşan Testlerde Klasik Test Eşitleme Yöntemlerinin Karşılaştırılması**" başlıklı bu çalışma Eğitim Bilimleri Anabilim Dalı / Eğitimde Ölçme ve Değerlendirme Tezli Yüksek Lisans Programı'nda jüri üyelerince oy birliği ile **Yüksek Lisans Tezi** olarak kabul edilmiştir.

	Jüri Üyeleri	İmza
Başkan	Prof. Dr. Ömay ÇOKLUK BÖKEOĞLU	.....
Üye	Prof. Dr. Nilüfer KAHRAMAN	.....
Üye	Doç. Dr. Ergül DEMİR (Danışman)	.....

ONAY

Bu tez Ankara Üniversitesi Lisansüstü Eğitim-Öğretim ve Sınav Yönetmeliği'nin ilgili maddeleri uyarınca jüri üyeleri tarafından 09/01/2020 tarihinde, Enstitü Yönetim Kurulunca .../.../20... tarihinde kabul edilmiştir.

Prof. Dr. Yasemin KEPENEKCİ
Eğitim Bilimleri Enstitü Müdürü

ETİK İLKELERE UYGUNLUK BİLDİRİMİ

Tez içindeki bütün bilgileri akademik yazım kurallarına uygun biçimde raporlaştırdığımı ve bunları etik ilkelere (atıfta bulunulan tüm yapıtlara kaynaklarda yer verilmesi, tezde kullanılan bilgi ve belgelere resmi yollarla ulaşılması ve bunların aslı bozulmadan kullanılması vb.) uygun olarak elde ettiğimi ve sunduğumu bildiririm.

Ali Emre KARAGÜL

ÖZET

Küçük Örneklemlerde Çok Kategorili Puanlanan Maddelerden Oluşan Testlerde Klasik Test Eşitleme Yöntemlerinin Karşılaştırılması

KARAGÜL, Ali Emre

Yüksek Lisans Tezi/ Eğitim Bilimleri Anabilim Dalı

Tez Danışmanı: Doç. Dr. Ergül DEMİR

Ocak, 2020, 107

Bu çalışma, çok kategorili puanlanan maddelerden oluşan testlerin küçük örneklemlere uygulanması durumunda kullanılabilecek test eşitleme yöntemlerini ortaya koymak amacıyla yürütülmüş bir olasılıksal simülasyon çalışmasıdır. Dört KTK temelli test eşitleme yönteminden (ortalamalar ile test eşitleme, Tucker, Braun-Holland ve düşük ağırlıklar doğrusal test eşitleme yöntemleri) elde edilen eşitlemenin standart hatası (ESH), yanlılık ve ortalama hata karekökü (OHKK) değerlerinin örneklem büyüklüğü (15, 20, 30, 60, 100, 150), madde sayısına bağlı puanlama biçimi (3 kategorili 50 madde, 6 kategorili 20 madde ve 11 kategorili 10 madde) ve formlar arası madde eşik parametresi farkı (.1, .4, .7) olmak üzere toplam 54 (6x3x3) koşulun nasıl değişim gösterdiği incelenmiştir.

Eşdeğer olmayan gruplar ile bağıntı testi deseni kullanılan simülasyonda veriler, genelleştirilmiş kısmi kredi modeli kullanılarak R yazılımı ltm paketi ile üretilmiştir. Elli dört koşul için toplam 108 test formu üretilmiştir. 100 tekrar uygulanmıştır. Test eşitleme işlemi, equate paketiyle yürütülmüştür. Bootstrap yöntemiyle hata puanları hesaplanmıştır. Bunların ortalamaları alınmış ve aralarındaki farkların test edilmesinde faktöriyel varyans analizi kullanılmıştır. Ayrıca analiz sonuçları grafikler ile yorumlanmıştır.

Çalışma, tüm test eşitleme yöntemlerinde, örneklem büyüdükçe ESH, yanlılık ve OHKK değerlerinin düştüğünü göstermektedir. Ayrıca, madde eşik parametresi farkları arttıkça, ESH, yanlılık ve OHKK değerleri artmaktadır. Ancak, madde sayısına bağlı puanlama biçimi koşulu değişimlerinin manidar bir etkisi gözlenmemiştir. Çalışma kapsamında ele alınan tüm koşullarda en düşük ESH, yanlılık ve OHKK değerlerini ortalamalar ile test eşitleme yöntemi, en yüksek değerleriyse düşük ağırlıklar doğrusal test eşitleme yöntemi üretmiştir. Tucker ve Braun-Holland doğrusal test eşitleme yöntemleri birbirlerine çok yakın değerler üretmiş. Aralarındaki fark istatiksel olarak manidar bulunmamıştır.

Anahtar sözcükler: Küçük örneklemler, test eşitleme, Klasik Test Kuramı, simülasyon, çok kategorili puanlanan testler.

ABSTRACT

Comparison Of Classical Test Equating Methods With Polytomously Scored Tests And Small Samples

KARAGÜL, Ali Emre

Master Degree / Department of Educational Sciences

Supervisor: Assoc. Prof. Dr. Ergül DEMİR

January, 2020, 107

This is a stochastic simulation study and is conducted to compare performances of Mean Equating and Tucker, Braun-Holland and Nominal Weights Linear Test Equating methods when tests are comprised of polytomously scored items and sample sizes are small. Standard error of equating (SEE), bias and root mean square error (RMSE) estimates were comparatively evaluated over 54 conditions using three factors: sample sizes (15, 20, 30, 60, 100, 150), scoring formats (50 items with 3 categories, 20 items with 6 categories and 10 items with 11 categories), and threshold differences between form A and B (.1, .4, .7).

The data of the simulation, which is carried on by using non-equivalent groups anchor test design, is generated via R programming language and ltm package program. 108 test forms, two forms for each of 54 conditions, are simulated. The number of replications is 100. Test equating process is carried via equate package program in R programming language. Errors of equating methods are calculated by bootstrapping. The means of these errors are calculated and the difference between them is tested by factorial variance analysis. The results are also summarized using plots.

The results reveal that SEE, bias and RMSE values decrease as sample size increases. SEE, bias and RMSE values also increase as the threshold differences increase. No significant differences are observed on SEE, bias and RMSE values due to variations

in the scoring formats. The highest SEE, bias and RMSE values are produced by Nominal Weights Linear Equating, whereas the lowest values are produced by Mean Equating in all conditions investigated in the study. Error estimates computed for the Tucker and Braun-Holland Linear Equating methods are similar and small, not amounting to any statistical significance.

Key words: small samples, test equating, Classical Test Theory, simulation studies, tests with polytomous items.

ÖNSÖZ

Öğretmenler, öğrencilerin öğrenme süreçlerini izlemek ve eğitim süreçleri ile ilgili önemli kararlar vermek için ölçme ve değerlendirme araçlarına sıklıkla başvururlar. Bir eğitim öğretim dönemi içerisinde birkaç kez testlerden faydalanırlar. Gerek aynı eğitim öğretim dönemi içerisinde, gerekse farklı dönemlerde benzer bilgi seviyelerine sahip öğrencilere benzer testler uygulanır. Test eşitleme, bu testlerden elde edilen puanları daha iyi anlamak ve değerlendirmek için kullanılabilen istatiksel bir araçtır. Ancak, sınıf içi uygulamalarda, çok kategorili puanlanan testlerin kullanılması ve öğrenci sayısının test eşitleme süreçlerini yürütmek için yeterli olmaması yaygın durumlardır. Bu çalışma ile, çok kategorili puanlanan testlerin az sayıda öğrencilere uygulandığı bu tür durumlar için öğretmenlere kullanabilecekleri en uygun test eşitleme yöntemini sunmak istenmektedir. Ayrıca, bu çalışmanın gelecekte test eşitleme ile ilgili yapılacak araştırmalara da katkı sağlayacağı düşünülmektedir.

Kendisini tanıdığım günden bu yana, büyük bir kısmı öğretmenlik ve araştırmacılık mesleklerini icra etmek ile geçen hayatımın zor günlerinde ve önemli karar anlarında; fikirleriyle, öğretileriyle ve kendisini gözlemlerken öğrendiğim davranış örüntüleriyle yoluma ışık tutan; bu araştırmanın konusu henüz bir fikir bile değilken, desteğini hiçbir şekilde esirgemeden varlığını hissettiren ve çalışmanın yürütülebilmesi için ihtiyaç duyduğum gücü ve bilgiyi bulmamda bana sabırla yol gösteren saygıdeğer danışman hocam Doç. Dr. Ergül Demir'e ne kadar teşekkür etsem azdır.

Akademik eğitimim boyunca benden yardımlarını ve birikimlerini esirgemeyen, her zaman yanımda olduklarını hissettiren saygıdeğer hocalarım Prof. Dr. Nizamettin Koç'a, Prof. Dr. Nükhet Demirtaşlı'ya, Prof. Dr. Ömay Çokluk'a, Doç. Dr. Deniz Gülleroğlu'na, Dr. Öğr. Gör. Ömer Kutlu'ya da teşekkürü bir borç bilirim.

Hem istatistik bilimi ile hem de çalışma kapsamındaki istatiksel süreçlerin tamamını yürüttüğüm R bilgisayar yazılım dili ile tanışmamda, bunların önemini kavramamda ve altında yatan temelleri anlamamda en az danışman hocam kadar katkısı olan saygıdeğer hocam Doç. Dr. C. Deha Doğan'a ayrıca teşekkür ederim.

Hayatı boyunca eğitimim için maddi manevi hiçbir fedakarlıktan kaçınmayan sevgili aileme; ayrıca, çalışmam boyunca yorulduğum, tıkandığım, bunaldığım her an manevi desteğini hissettiğim sevgili hayat arkadaşıma sonsuz teşekkürler. Onların önemini anlatmaya kelimeler kifayetsiz kalır.

Son olarak, şu an bulunduğum noktaya gelmemde emeği olan ancak burada isimlerini anamadığım tüm saygıdeğer hocalarıma, öğretmenlerime ve varlıklarıyla bana destek olan dostlarıma çok teşekkür ederim.

İÇİNDEKİLER

Sayfa

ONAY ... ii
ETİK İLKELERE UYGUNLUK BİLDİRİMİ ... iii
ÖZET ... iv
ABSTRACT ... vi
ÖNSÖZ ... viii
İÇİNDEKİLER ... x
TABLOLAR DİZİNİ ... xii
ŞEKİLLER DİZİNİ ... xiv
KISALTMALAR VE SİMGELER ... xv
BÖLÜM 1 ... 1
GİRİŞ ... 1
Problem ... 1
Amaç ... 6
Önem ... 7
Sınırlılıklar ... 8
Tanımlar ... 9
BÖLÜM 2 ... 10
KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR ... 10
Kuramsal Çerçeve ... 10
İlgili Araştırmalar ... 32
BÖLÜM 3 ... 42
YÖNTEM ... 42
Araştırmanın Modeli ... 42
Simülasyon Koşulları ... 43
Verilerin Üretilmesi ve Analizi ... 48
BÖLÜM 4 ... 51
BULGULAR VE YORUMLAR ... 51
Örneklem Büyüklüğü ve Madde Sayısına Bağlı Puanlama Biçimi Değişimlerinin ESH, Yanlılık ve OHKK Değerlerine Ortak Etkisi ... 53
Örneklem Büyüklüğü ve Madde Eşik Parametresi Farkı Değişimlerinin ESH, Yanlılık ve OHKK Değerlerine Ortak Etkisi ... 63
Madde Sayısına Bağlı Puanlama Biçimi ve Madde Eşik Parametresi Farkı Değişimlerinin ESH, Yanlılık ve OHKK Değerlerine Ortak Etkisi ... 74
BÖLÜM 5 ... 81
SONUÇ VE ÖNERİLER ... 81
Sonuç ... 81
Öneriler ... 84

KAYNAKLAR 87
EKLER 93
EK 1: Varsayımların Kontrolü ile İlgili Test Sonuçları 94
EK 2: Tukey HSD Testi Sonuçları 98
EK 3: Çalışma Sürecinde Yapılan Görüşmeler 104
EK 4: Etik Kurul Onayı 105
BENZERLİK BİLDİRİMİ 106
ÖZGEÇMİŞ 107

TABLOLAR DİZİNİ

Tablo **Sayfa**

Tablo 1. Yüzdelik dilimler ile test eşitlemede ayrıklık problemi 22
Tablo 2. MTK temelli modeller ve MKE oluşturma fonksiyonları..... 28
Tablo 3. MTK temelli eşitlemelerde parametre dönüşüm fonksiyonları..... 28
Tablo 4. Her puanlama biçimi koşulu için madde eşik parametreleri 47
Tablo 5. Çalışma kapsamında değişimlenen koşullar 49
Tablo 6. Çalışma kapsamında sabit tutulan koşullar 49
Tablo 7. Çalışma kapsamında ele alınan test eşitleme yöntemlerinden simülasyon koşullarında elde edilen ESH, yanlılık ve OHKK değerleri..... 52
Tablo 8. Örneklem büyüklüğüne göre ESH, yanlılık ve OHKK ortalama ve standart sapmaları..... 53
Tablo 9. Madde sayısına bağlı puanlama biçimi koşuluna göre ESH, yanlılık ve OHKK ortalama ve standart sapmaları..... 54
Tablo 10. Test eşitleme yöntemi, örneklem büyüklüğü ve madde sayısına bağlı puanlama biçimi koşullarının ESH değerleri üzerindeki etkileri 54
Tablo 11. Test eşitleme yöntemi, örneklem büyüklüğü ve madde sayısına bağlı puanlama biçimi koşullarının yanlılık değerleri üzerindeki etkileri..... 57
Tablo 12. Test eşitleme yöntemi, örneklem büyüklüğü ve madde sayısına bağlı puanlama biçimi koşullarının OHKK değerleri üzerindeki etkileri..... 59
Tablo 13. Örneklem büyüklüğüne göre ESH, yanlılık ve OHKK ortalama ve standart sapmaları 64
Tablo 14. Madde eşik parametre farkı koşuluna göre ESH, yanlılık ve OHKK ortalama ve standart sapmaları 64
Tablo 15. Test eşitleme yöntemi, örneklem büyüklüğü ve madde eşik parametre farkı koşullarının ESH değerleri üzerindeki etkileri..... 65
Tablo 16. Test eşitleme yöntemi, örneklem büyüklüğü ve madde eşik parametre farkı koşullarının yanlılık değerleri üzerindeki etkileri 67
Tablo 17. Test eşitleme yöntemi, örneklem büyüklüğü ve madde eşik parametre farkı koşullarının OHKK değerleri üzerindeki etkileri..... 70

Tablo 18. Madde sayısına bağlı puanlama biçimi koşuluna göre ESH, yanlılık ve OHKK ortalama ve standart sapmaları 75
Tablo 19. Madde eşik parametre farkı koşuluna göre ESH, yanlılık ve OHKK ortalama ve standart sapmaları 75
Tablo 20. Test eşitleme yöntemi, madde sayısına bağlı puanlama biçimi ve madde eşik parametresi farkı koşullarının ESH değerleri üzerindeki etkileri..... 76
Tablo 21. Test eşitleme yöntemi, madde sayısına bağlı puanlama biçimi ve madde eşik parametresi farkı koşullarının yanlılık değerleri üzerindeki etkileri 77
Tablo 22. Test eşitleme yöntemi, madde sayısına bağlı puanlama biçimi ve madde eşik parametresi farkı koşullarının OHKK değerleri üzerindeki etkileri..... 78
Tablo 23. Araştırma kapsamında incelenen koşullar için önerilen test eşitleme yöntemleri 84

ŞEKİLLER DİZİNİ

Şekil **Sayfa**

Şekil 1. Test ilişkilendirme yaklaşımları 3
Şekil 2. Tek grup deseni ile test eşitleme 13
Şekil 3. Eşdeğer gruplar deseni ile test eşitleme 14
Şekil 4. Denkleştirilmiş gruplar deseni ile test eşitleme 14
Şekil 5. EOGB deseni ile test eşitleme 15
Şekil 6. Güçlü ve Zayıf Gruplarda Doğrusal Test Eşitleme Problemi 19
Şekil 7. Yüzdelik Dilimler ile Test Eşitleme 20
Şekil 8. Örneklem büyüklüğü değişimlerine ve madde sayısına bağlı puanlama biçimi koşullarına göre test eşitleme yöntemlerinden elde edilen ortalama ESH, Yanlılık ve OHKK değerleri 62
Şekil 9. Örneklem büyüklüğü değişimlerine ve madde eşik parametre farkı koşullarına göre test eşitleme yöntemlerinden elde edilen ortalama ESH, yanlılık ve OHKK değerleri 73
Şekil 10. Madde Sayısına Bağlı Puanlama Biçimi ve Madde Eşik Parametre Farkı Koşullarına göre Test Eşitleme Yöntemlerinden Elde Edilen ortalama ESH, Yanlılık ve OHKK değerleri 80

KISALTMALAR VE SİMGELER

ESH: Eşitlemenin standart hatası
KTK: Klasik Test Kuramı
MEPF: Madde eşik parametre farkı
MSBP: Madde sayısına bağlı puanlama biçimi
MTK: Madde Tepki Kuramı
1PL: 1 Parametreli lojistik model
2PL: 2 Parametreli lojistik model
3PL: 3 Parametreli lojistik model
a: Ayırt edicilik parametresi,
b: Güçlük parametresi,
c: Şans parametresi,
D: Lojistik ve ogive model fonksiyonlarının birbirlerine çok yakın sonuç vermesi için kullanılan bir sabit değer,
GPCM: Generialized partial credit model (Genelleştirilmiş kısmi kredi modeli)
GRM: Graded response model (Derecelendirilmiş tepki modeli)
MKK: Madde karakteristik eğrisi
θ: Yetenek düzeyi.
OHKK: Ortalama hata karekökü
ÖBBS: Öğrenci Başarılarını Belirleme Sınavı
ÖSS: Öğrenci Seçme Sınavı
PISA: Program for Student Assessment
Test eşitleme desenleri:
TG: Tek grup deseni
EG: Eşdeğer gruplar deseni
DG: Denkleştirilmiş gruplar deseni
EOGB: Eşdeğer olmayan gruplar ile bağıntı testi deseni

BÖLÜM 1

GİRİŞ

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

Problem

Psikolojide ve eğitimde uygulanan ölçme ve değerlendirme süreçlerinde, bireyler ya da eğitim süreçleri hakkında karar vermek amacıyla, farklı ölçümlerden elde edilen sonuçlar arasında bir anlam ilişkisi kurma ihtiyacı duyulabilmektedir. Özellikle eğitim alanında, bir ölçme sürecinden elde edilen puanların başka bir ölçme süreci için manidar olup olmadığı ve manidar ise ne anlam taşıdığı merak konusu olmuştur. Çeşitli otoriteler, bireyler ve eğitim politikaları hakkında doğru kararlar almak amacıyla, testler arasında ilişki kurma süreçlerinden faydalanmak istemişlerdir. Örneğin 90'lı yılların sonunda, dönemin Amerika Birleşik Devletleri (ABD) Başkanı ve ABD'de eğitim alanında söz sahibi resmî kurumlar, ulusal olarak uygulanan ve farklı bilişsel alanları ölçen testlerin tek bir ölçek üzerinde gösterilip gösterilemeyeceği konusunda bir rapor oluşturulması talebinde bulunmuşlardır. Bu amaç doğrultusunda bir komite kurulmuş ve psikometri alanında çalışmalar yapmış olan 43 akademisyenin görüşlerine başvurularak bir rapor oluşturulmuştur (Feuer, Hollad, Green, Bertenthal ve Hemphill, 1999). Bu raporda, gerekli varsayımlar sağlandığı taktirde iki testin arasında ilişki kurulabileceği ve varsayımların karşılandığı ölçüde güvenilir yorumlar yapılabileceği vurgulanmıştır.

Psikometrik açıdan incelendiğinde, burada bahsedilen test ilişkilendirme kavramı; bir testten elde edilebilecek bir puanın, başka bir testten elde edilebilecek bir puana dönüştürülmesi olarak tanımlanmaktadır (Dorans, Moses ve Eignor, 2010). Kolen ve Brennan'a (2004) göre, iki veya daha fazla ölçmeden elde edilen puanları aynı ölçek üzerine yerleştirme sürecine *test ilişkilendirme* denir. Mislevy'e (1992) göre, bir ölçmeye ait sonuçların, aynı yapıyı ölçen bir başka ölçme ile karşılaştırılabilir hale getirilmesi için uygulanan yaklaşımların tümüne test ilişkilendirme denmektedir. Bu tanımlamalardan

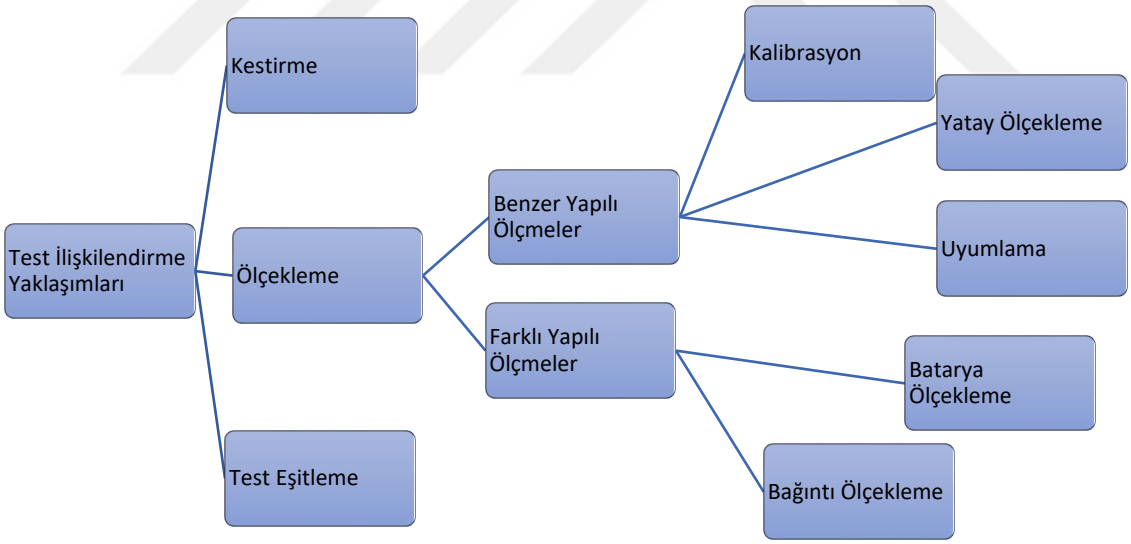
yola çıkarak test ilişkilendirme; iki veya daha fazla testten alınabilecek puanların ortak bir ölçek üzerinde gösterilebilmesi olarak tanımlanabilir.

Bu doğrultuda, Mislevy (1992) ve Linn (1993), test ilişkilendirmeyi yapılış amacına göre ve yapıldığı testlerin içeriğine ve uzunluğuna göre değişiklik gösteren yaklaşımlarla ele almaktadırlar. Bu bağlamda; test ilişkilendirme yaklaşımları, beş ana başlık altında derlenmiştir. Bunlar; (1) kestirme (projection or prediction), (2) istatiksel moderasyon, (3) sosyal moderasyon, (4) kalibrasyon ve (5) test eşitlemedir.

Kestirme yaklaşımı, bir grubun bir formdan elde ettiği puanlardan yola çıkarak, başka bir formdan elde etmesi muhtemel olan puanları hesaplama prensibine dayanmaktadır. Ancak, bu yaklaşımda bir asimetriklik söz konusudur. X testine ait puanlardan yola çıkarak Y testinin puanları tahmin edilebilirken, aynı formül kullanılarak Y testinden X'i tahmin etmek mümkün değildir (Dorans ve diğerleri, 2010). Bu nedenle kestirme yaklaşımı, test ilişkilendirme yaklaşımları içerisinde istatiksel olarak en güçsüz olan yaklaşımdır (Linn,1993). *Moderasyon yaklaşımı* ise, istatiksel moderasyon ve sosyal moderasyon olmak üzere iki farklı modelle tanımlanmaktadır. İstatiksel moderasyon, iki farklı gruba uygulanan iki farklı formun sonuçlarından, üçüncü bir form ya da ortak maddeler aracılığı ile karşılaştırılabilir puanlar elde etme süreci olarak tanımlanırken; sosyal moderasyon, uzman görüşlerine dayandırarak iki form arasında karşılaştırılabilir puanlar elde etme süreci olarak tanımlanmıştır. *Kalibrasyon yaklaşımı* ise, uzunlukları birbirinden farklı olan iki formun karşılaştırılması süreçlerinde kullanılan bir yöntemdir (Holland, 2007). *Test Eşitleme Yaklaşımı,* test ilişkilendirme yaklaşımları içerisinde en fazla teknik bilgi gerektiren ve istatiksel olarak en güçlü sonuçlar veren yaklaşım olarak tanımlanmaktadır (Mislevy, 1992; Linn, 1993; Kolen ve Brennan, 1995; Holland, 2007). Bunun en temel sebebi, test eşitlemenin ortak bir yapıyı ölçen testler arasında yapılıyor olmasıdır. Bir başka ifade ile test eşitleme süreci, sadece istatiksel bir süreç olarak ele alınmamalı, testlerin geliştirilme aşamasından ortak bir ölçek üzerinde puanlanmasına kadar bütün bir süreç olarak görülmelidir (Mislevy, 2018). Test eşitlemenin sadece aynı yapıyı ölçen testler için kullanılabiliyor olması, test eşitleme ile diğer test ilişkilendirme yaklaşımları arasındaki en temel farkı da ortaya koymaktadır. Test ilişkilendirme, farklı yapıları ölçen testler arasında da yapılabilirken; test eşitleme sadece ortak bir yapıyı ölçen testler arasında uygulanabilecek istatiksel bir yöntemdir (Kolen & Brennan, 2004; Holland, 2007; Dorans ve diğerleri, 2010). Yine de diğer test ilişkilendirme yaklaşımlarını ve test eşitlemeyi birbirinden tamamen bağımsız olgular olarak algılamak

doğru olmayacaktır, çünkü test eşitleme ve diğer test ilişkilendirme yaklaşımları hem yöntem hem de desen açısından ortak süreçler barındırmaktadır.

İstatiksel ilişkilendirme yaklaşımları, başka bir sınıflamada ise (1) kestirme, (2) ölçekleme ve (3) test eşitleme olarak üçe ayrılmıştır (Holland, 2007). *Kestirme* ve *test eşitleme*, genel hatlarıyla Mislevy (1992)'de yer alan sınıflamadaki gibi tanımlanmaktadır. *Ölçekleme* ise, iki farklı testin puanlarının bir ölçek üzerinde gösterilmesi olarak tanımlanmaktadır. Ölçekleme yaklaşımı, ölçülen yapının ortak olduğu ve olmadığı durumlar için farklı alternatifler sunan bir sınıflama barındırmaktadır. Buna göre yapının ortak olduğu durumlarda testler arasındaki güvenirlik düşük ise *kalibrasyon;* yüksek ise *yatay ölçekleme* ya da *uyumlama (concordance)* yaklaşımları sunulmaktadır. Ölçülen yapının ortak olmadığı durumlarda ise testin uygulandığı grup önem kazanmaktadır. Testler tek bir gruba uygulanmışsa, *batarya ölçekleme*; farklı gruplara uygulanmışsa, *bağıntı (anchor) ölçekleme* yaklaşımları sunulmaktadır (Holland, 2007). Holland (2007)'de yer alan test ilişkilendirme yaklaşımları sınıflamasının bir uyarlaması Şekil 1'de gösterilmektedir.



Şekil 1. Test ilişkilendirme yaklaşımları

Kaynak. Holland, 2007

Her ne kadar tüm bu yaklaşımlar, testler arasında mukayese yapma süreçlerinde kullanılsa da her birinin amacı ve kullanım alanları farklılık göstermektedir (Kolen ve Brennan, 1995). İlişkilendirme yapılmak istenen konuya ve ölçmelere göre en uygun

ilişkilendirme yaklaşımının seçilip uygulanması gerekmektedir. Ancak tüm bu yaklaşımların içerisinde en titiz çalışmayı test eşitleme gerektirmektedir.

Test eşitleme

Ölçme ve değerlendirme süreçlerinde; farklı zamanlarda elde edilen ölçme sonuçlarının birbirine eşdeğer şekilde karşılaştırılarak kullanılması, bunların tek bir testten elde edilen puanlar olduğu varsayımı üzerinden kararlar verilmesi veya eşdeğer formlar kullanma gibi durumlarla karşılaşılmaktadır. Farklı dönemlerde farklı formları uygulanan testler bu duruma örnektir.

Bir testin farklı formlarının oluşturulmasının bazı sebepleri bulunmaktadır. Zhu (1998), bu sebeplerin uygulamadan uygulamaya değişeceğini belirtmiştir. Sansivieri, Wiberg ve Matteucci (2017) ise bunları 3 maddede ele almaktadır. Bunlar; (1) güvenlik, (2) test maddelerinin kamu ile paylaşılması ve (3) test içeriklerindeki değişimdir. Burada bahsedilen ilk iki neden birbirleri ile ilişkilidir. Bir test uygulandıktan sonra, o teste ait maddeler gizliliğini yitirmiş olur. Ayrıca, bazı durumlarda test uygulayıcıları bu maddeleri kamu ile paylaşmaktadır. Bu nedenle bir sonraki uygulamada o teste ait başka bir formun kullanılması çok yaygın bir durumdur. Aynı amaçla kullanılan bu iki formdan elde edilen puanların aynı anlamı taşıdığından emin olmak, dolayısıyla farklı formları alan bireyler arasında adil bir karar verebilmek için uygulanan istatiksel süreçlere *test eşitleme* adı verilmektedir (Zhu, 1998).

Test eşitleme kavramı birçok araştırmacı tarafından benzer şekillerde tanımlanmaktadır. Örneğin; Dorans ve Holland (2000)'a göre, aynı yapıyı ölçen farklı testlerden elde edilen puanları karşılaştırabilmek amacı ile test puanlarının istatiksel ve/veya psikometrik tekniklerle ayarlanmasına test eşitleme denir. Başka bir tanımlamaya göre ise, bir teste ait iki formun maddelerinin ayırt edicilik ve güçlük düzeylerinin aynı olmamasından dolayı oluşan farklılıkları ortadan kaldırmak amacıyla yapılan istatiksel işlemlere test eşitleme denir (Dorans ve Puhan, 2017). Ancak, farklı formlardan elde edilen puanların eşdeğer kabul edilebilmesi ve test eşitleme süreçlerinin yürütülebilmesi için bu formların bazı ortak özellikleri taşıması gerekmektedir (Lord, 1980). Test eşitlemenin temel varsayımları olarak adlandırılan bu özellikler; (1) formların ortak bir yapıyı ölçmesi, (2) her bir puanın diğer formlarda bir karşılığının olması, (3) formların güvenirlik düzeyleri, madde güçlük ve ayırt edicilik endeksleri gibi psikometrik

özelliklerinin eşit olması ve (4) formlar arasındaki eşdeğerliğin uygulandıkları gruptan etkilenmemesi olarak tanımlanmaktadır. (Holland, 2007; Dorans ve diğerleri, 2010).

Test eşitleme yaklaşımı, formların güçlükleri açısından incelendiğinde ikiye ayrılmaktadır (Crocker ve Algina, 1986): *Yatay eşitleme* aynı yapıyı ölçen ve benzer zorluk düzeylerine sahip maddelerden oluşan testleri eşitlerken kullanılır. Bir başka ifade ile yatay eşitleme; kapsam, güvenirlik ve güçlük düzeyi benzer testler arasında test eşitleme yapıldığı zaman kullanılır. Bu durumlarda, bireylerin de ölçülmek istenen özellik açısından benzer seviyede olmaları beklenir (Felan, 2002). *Dikey eşitleme* ise aynı yapıyı farklı seviyelerde ölçen testleri eşitlerken kullanılır (Hambleton ve Swaminathan, 1985; Crocker ve Algina, 1986). Ancak, dikey eşitlemeyle ilgili yapılmış olan bu tanımın test eşitlemenin yürütülebilmesi için formların sahip olması gereken ortak özellikleri ile çeliştiği görülmektedir. Başka bir ifade ile, dikey eşitleme güçlük düzeyleri faklı olan testleri eşitlerken kullanılacağından, testlerin uygulandığı grupların da ölçülen özelliğe ait düzeyleri farklı olacaktır. Kolen (1988), böyle durumlarda, test eşitlemenin varsayımlarının ihlal edileceğini, bu nedenle dikey eşitlemenin bir test eşitleme sürecinden daha çok bir test ölçekleme süreci olduğunu ifade etmektedir.

Angoff (1953), test eşitleme sürecinde iki farklı hata kaynağından bahsetmektedir. Bunlardan birincisi formların uygulandığı gruplardan kaynaklanan hatalar, ikicisi ise formlardan kaynaklanan hatalardır. Burada bahsedilen her iki hata türünü de asgari düzeye indirebilmek için ya formların ortak maddeler barındırması ya da formların uygulandığı grupların aynı olması önerilmektedir. Bu durumlar, test eşitleme desenleri ve yöntemleri arasındaki farklılıkları doğurmaktadır.

Görüldüğü üzere, test eşitleme süreçlerinde, birçok değişken önem arz etmektedir. Bu değişkenlere uyum sağlayabilmek için önerilen veri toplama desenleri ve test eşitleme yöntemleri farklılaşmaktadır (Kolen ve Brennan, 2004). Örneğin; test formlarının uygulandıkları örneklemin büyüklüğü kullanılacak test eşitleme yönteminin karar verilmesinde önem arz etmektedir. Benzer bir şekilde; formların madde sayısı, testin uygulandığı bireylerin yetenek düzeyleri, formların maddelerinin güçlük düzeyleri, testin puanlama biçimi gibi değişkenlerin farklı koşulları için farklı veri toplama deseni ve test eşitleme yöntemi kullanmak gerekmektedir.

Test eşitleme alanında çalışmalar yapan araştırmacılar, bazı koşullarda kullanılacak desen ve yöntem açısından hemfikir iken, bazı koşullar için uygun desen ve yöntemler

tartışma konusu olmaya devam etmektedir. Formların uygulanacağı bireylerin sayısı bu tartışma konularından birisidir. Yapılan literatür taramasında örneklem büyüklüğünün küçük ve çok küçük olduğu durumlarda KTK temelli test eşitleme yöntemlerinin tercih edildiği görülmektedir. Kolen ve Brennan (2004) da MTK temelli test eşitleme yöntemlerinin büyük örneklemler ile uygun olduğunu, küçük örneklemlerle test eşitleme çalışmalarının KTK temelli test eşitleme yöntemleriyle yapılması gerektiğini belirtmişlerdir.

Benzer bir şekilde, testlerin *dikotom* (doğru-yanlış şeklinde iki kategorili puanlanan) ya da *politom* (çok kategorili puanlanan) maddelerden oluşması durumları da farklı test eşitleme yöntemlerinin kullanılmasını gerektirmektedir (Kolen ve Brennan, 1995). Örneğin, çok kategorili puanlanan maddelerden oluşan testlerin eşitlenmesinde MTK temelli test eşitleme yöntemleri önerilmektedir ve bu tür maddelerden oluşan testlerde KTK temelli yöntemlerle eşitleme ile ilgili çalışmalar sınırlıdır.

Tüm bunlar dikkate alındığında, küçük örneklem büyüklüklerinde KTK temelli test eşitleme yöntemleri önerilirken, çok kategorili puanlanan testlerin eşitlenmesinde MTK temelli test eşitleme yöntemlerinin önerildiği görülmektedir. Ancak, günümüz eğitim süreçlerinde ve sınıf içi ölçme süreçlerinde çok kategorili puanlanan testlerin kullanılmasına sıkça rastlanmaktadır. Bu gibi durumlarda, testlerin uygulandığı grupların genellikle çok küçük örneklemlerden oluştuğu da bilinmektedir. Burada bahsedilen eğitim süreçleri için, çok kategorili puanlanan maddelerden oluşan testlerde, örneklem boyutunun küçük olması durumunda hangi yöntemlerin kullanılabileceği konusunda, literatürde yeterli ve ayrıntılı düzeyde bilgiye rastlanmamıştır ve bu durum merak uyandırmaktadır. Bu nedenle, bu çalışma küçük örneklemlere uygulanan çok kategorili puanlanan maddelerden oluşan testlerin eşitlenmesinde kullanılabilecek test eşitleme yöntemlerini sınamak üzerine simülatif bir çalışmadır.

Amaç

Bu araştırma, çok kategorili maddelerden oluşan testlerin farklı formlarının küçük örneklemlere uygulanması durumunda kullanılabilecek en uygun klasik test eşitleme yöntemlerini ortaya koymak amacıyla yürütülmüştür. Bu amaç doğrultusunda kullanılabilecek en uygun, bir başka ifade ile en düşük test eşitleme hatasını üreten test eşitleme yönteminin belirlenmesi amaçlanmıştır. Bu nedenle, KTK temelli dört farklı test

eşitleme yönteminden (ortalamalar ile test eşitleme, Tucker doğrusal test eşitleme, düşük ağırlıklar doğrusal test eşitleme ve Braun-Holland doğrusal test eşitleme) elde edilen eşitlemenin standart hatası (ESH), yanlılık ve ortalama hata karekökü (OHKK) değerlerinin örneklem büyüklüğü koşullarında (N= 15, 20, 30, 60, 100, 150), madde sayısına bağlı puanlama biçimi koşullarında (3 kategorili puanlanan 50 madde, 6 kategorili puanlanan 20 madde ve 11 kategorili puanlanan 10 madde) ve maddelerin eşik parametresi farkı koşullarında (.1, .4, .7) nasıl değişim gösterdiği incelenmiştir. Şu sorulara yanıt aranmıştır:

1. Örneklem büyüklüğü ve madde sayısına bağlı puanlama biçimi değişimleri ayrı ayrı ve birlikte ele alındığında, çalışma kapsamındaki tüm test eşitleme yöntemlerinden elde edilen ESH, yanlılık ve OHKK değerleri nasıl etkilenmektedir?
2. Örneklem büyüklüğü ve maddelerin eşik parametresi değişimleri ayrı ayrı ve birlikte ele alındığında, çalışma kapsamındaki tüm test eşitleme yöntemlerinden elde edilen ESH, yanlılık ve OHKK değerleri nasıl etkilenmektedir?
3. Madde sayısına bağlı puanlama biçimi ve maddelerin eşik parametresi değişimleri birlikte ele alındığında, çalışma kapsamındaki tüm test eşitleme yöntemlerinden elde edilen ESH, yanlılık ve OHKK değerleri nasıl etkilenmektedir?

Önem

Bir okulda görev yapan bir öğretmen kendi branşının derslerini birden fazla şubede verebilmektedir. Bu öğretmen, öğrencilerin öğrenme düzeylerini ölçmek ya da onlar hakkında ders ile ilgili kararlar almak amacıyla eğitim öğretim yılı içerisinde farklı test uygulamaları yapabilmektedir. Bu gibi durumlarda test güvenliğini sağlamak ve öğrenciler arasında adil yargılara varmak için, farklı test formları hem farklı şubelerde hem de tek bir şube içerisinde uygulanabilmektedir. Ancak bu formların güçlük düzeyleri açısından eşdeğer olmaması durumunda, öğrencilerin birbirlerine karşı avantajlı ya da dezavantajlı olmaları kaçınılmazdır. Bu nedenle bu formların test eşitleme çalışmalarını yapmak; adil bir puanlama yapmak ve karşılaştırılabilir test puanları elde etmek için önem arz etmektedir.

Çok kategorili puanlanan testler, oluşturulması ve uygulanması açısından görece kolay görülen testlerdir. Bu nedenle, eğitimin tüm kademelerinde eğitimciler tarafından sıkça tercih edilir ve kullanılırlar. Bu testlerin test eşitleme çalışmalarında MTK temelli test eşitleme yöntemleri önerilmektedir (Gonzales ve Wiberg, 2017). Ancak gerçek eğitim öğretim durumları göz önünde bulundurulduğunda, sınıf içi test uygulamaları çoğu zaman MTK temelli bir test eşitleme çalışması yürütmek için yeterli örneklem ve madde sayısına sahip değildir. Ayrıca MTK temelli test eşitleme yöntemleri KTK temelli yöntemlere göre daha karmaşık istatiksel süreçler gerektirmektedir (Kolen ve Brennan, 2004). Lisans öncesi kademelerde eğitim veren bir öğretmenin bu süreçleri uygulayabilmesi için bu alanda uzmanlaşmış olması gerekmektedir. Tüm bunlar göz önünde bulundurulduğunda, hem küçük örneklemler ile kullanım kolaylığı sunması hem de uygulama açısından daha basit olması sebebiyle KTK temelli test eşitleme yöntemlerinin daha elverişli olduğu düşünülmektedir.

Bu çalışma, çok kategorili puanlanan testlerin küçük örneklemlere uygulandığı okul ve sınıf içi test uygulamalarında kullanılabilecek en uygun test eşitleme yöntemini sunmak amacıyla yapılmaktadır. Bu açıdan bakıldığında literatürde benzer bir çalışmaya rastlanmamıştır. Bu nedenle, bu çalışmanın ilgili literatüre önemli katkılar yapacağı düşünülmektedir.

Sınırlılıklar

1. Çalışma kapsamında sadece eşdeğer olmayan gruplar ile bağıntı testi deseni dikkate alınmıştır.

2. Çalışma; ortalamalar ile test eşitleme yöntemi, Tucker, düşük ağırlıklar ve Braun-Holland doğrusal test eşitleme yöntemleri ile sınırlıdır.

3. Çalışmada kullanılan veriler MTK temelinde genelleştirilmiş kısmi kredi modelinden (GPCM) faydalanılarak üretilmiştir.

4. Bu çalışma, simülasyonda değişimlenen faktörler olan örneklem büyüklüğü, madde sayısına bağlı puanlama biçimi ve test formları arasındaki eşik parametresi farkı ile sınırlıdır.

Tanımlar

Madde Sayısına Bağlı Puanlama Biçimi: Çalışmanın simülasyon koşullarından birisidir. Veriler, madde sayısına bağlı 3 puanlama biçimi ile üretilmiştir. Bunlar 3 kategorili (0-1-2), 6 kategorili (0 ile 5 arası) ve 11 kategorili (0 ile 10 arası) puanlanan testlerdir. Her bir koşulda madde sayıları, testten alınabilecek maksimum ve minimum puan sırasıyla 100 ve 0 olacak şekilde ayarlanmıştır. Buna göre 3 kategorili puanlanan test koşulu 50 madde, 6 kategorili puanlanan test koşulu 20 madde ve 11 kategorili puanlanan test koşulu 10 maddeden oluşmaktadır. Bu üç farklı durumdan oluşan simülasyon koşulu puanlama biçimi olarak adlandırılmıştır.

BÖLÜM 2

KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

Bu bölümde; test eşitlemenin temel kavramları, test eşitlemede veri toplama desenleri ve test eşitleme yöntemleri ile ilgili kavramsal açıklama ve tartışmalara değinilmiştir. Ayrıca test eşitleme alanında yapılmış çeşitli araştırmalara yer verilmiştir.

Kuramsal Çerçeve

Test eşitleme sürecinin işletilebilmesi için bazı süreçleri uygulamak gerekmektedir. Bu çalışmada bu süreçler "Test Eşitlemenin Temel Varsayımları", "Test Eşitlemede Veri Toplama Desenleri" ve "Test Eşitleme Yöntemleri" başlıkları altında ele alınmaktadır. Bunun yanı sıra, test eşitleme süreci ve test eşitlemede hata kavramı üzerinde de durulmuştur.

Test eşitlemenin temel varsayımları

Test eşitlemenin bazı temel varsayımları bulunmaktadır (Lord, 1950; Lord, 1980; Angoff, 1984; Petersen, Kolen ve Hoover, 1989; Dorans, 1990; Kolen ve Brennan, 2004; Holland, 2007; Kolen, Tong ve Brennan, 2011; Gonzales ve Wiberg, 2017). Bunlar; (1) ortak yapı varsayımı, (2) simetri varsayımı, (3) eşitlik varsayımı ve (4) grup değişmezliği varsayımı olarak tanımlanmaktadır. Bazı kaynaklarda test eşitleme varsayımları arasında beşinci varsayım olarak testlerin eşit güvenirlik düzeylerine sahip olması gerektiği de *(eşit güvenirlik varsayımı)* vurgulanmaktadır (von Davier, Holland ve Thayer; 2004). Ancak Dorans ve Holland (2000) tarafından belirtilene göre, bu varsayımın sağlanmadığı durumlarda da test eşitleme tatmin edici düzeylerde yapılabilmektedir. Ayrıca Lord (1980) ortak yapı varsayımı ya da eşit güvenirlik varsayımlarından herhangi birisinin sağlanmasının test eşitleme yapabilmek için yeterli olduğunu belirtmiştir. Bu nedenle eşit güvenirlik varsayımı bu çalışmada ele alınmamaktadır.

Ortak yapı varsayımı. Testlerin eşit kabul edilebilmesi için, bu testlerin ortak bir yapıyı, beceriyi ya da özelliği ölçüyor olması gerekmektedir (Angoff, 1984; Dorans, Moses ve Eignor, 2011). Örneğin ölçülmek istenen özelliği başarı testleri ile ölçeceksek, bu durumda kullanılan testlerin ortak bir belirtke tablosuna ve ortak bir yönergeye sahip olmaları beklenmektedir (Dorans ve Holland, 2000). Bu varsayımı daha somut bir örnek üzerinden açıklayacak olursak, bir fen bilimleri testini ele alabiliriz. Bu testin 10 maddesinin fizik bilgisi, 7 maddesinin kimya bilgisi ve 8 maddesinin de biyoloji bilgisi ölçen 25 maddeden oluştuğunu varsayalım. Bu testin eşiti kabul edilebilecek bir testin aynı sayıda maddeden oluşması, aynı sayıda kimya, biyoloji ve fizik maddelerini barındırması ve bu maddelerin aynı yapıyı ölçmeleri gerekmektedir.

Simetri varsayımı. Formlardan hangisinin X hangisinin ise Y olarak adlandırıldığından bağımsız olarak test eşitlemenin aynı sonucu vermesi gerekmektedir (Lord, 1980). Başka bir ifade ile X formundan elde edilen bir puanı Y formuna ait bir ölçek üzerinde gösterebilmek için kullanılacak fonksiyon, aynı şekilde Y formundan elde edilen bir puanı X formuna ait bir ölçek üzerinde gösterirken de kullanılabilmelidir (Kolen ve Brennan, 1995). Örneğin, X formundan elde edilen 70 puanın Y formunda 72 puana tekabül etmekte olduğunu varsayalım. Bu sonuca varmamızı sağlayan fonksiyonu kullanarak, Y formundan elde edilen 72 puanı aynı şekilde X formunda 70 puana eşitleyebilmemiz gerekmektedir. Aksi takdirde yapılan işlem bir test eşitleme olmaktan çıkacaktır (Bandalos, 2018).

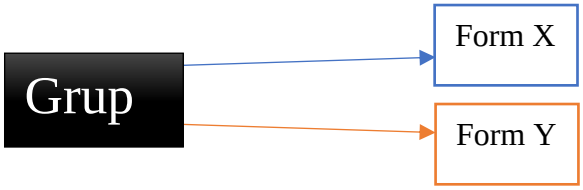
Eşitlik varsayımı. Bir bireyin test eşitleme işlemi yapılmış olan iki formdan hangisini aldığı, o kişinin gözlenen puanını etkilememelidir. Bunu sağlayabilmek için sadece formların aynı yapıyı ölçtüğünden emin olmak yeterli değildir (Dorans, 1990). Bunun yanı sıra, testlerin güçlük düzeylerinin ve diğer psikometrik özelliklerinin de benzerlik göstermesi gerekmektedir (Kolen ve Brennan, 2004). Ancak bu varsayımı pratikte uygulamak çok zordur (Kolen ve Brennan, 1995), çünkü bu varsayımın sağlanması eşitlenecek iki formun birebir aynı olmasını gerektirir ki bu durumda test eşitlemeye ihtiyaç kalmayacaktır. Daha önce de belirtildiği gibi, bu varsayımın da en mümkün şekilde sağlandığından emin olmak gerekmektedir.

Grup değişmezliği varsayımı. Test eşitlemek için kullanılan fonksiyonun, testlerin uygulandığı gruba bağlı olarak değişiklik göstermemesi beklenmektedir (Lord, 1980). Başka bir ifade ile test eşitleme sürecinin evrenden seçilen farklı gruplardan etkilenmeden aynı sonucu vermesi gerekmektedir (Holland, 2007). Örneğin, X formundan kadınların elde ettiği puanı Y formuna eşitlemek için kullanılan formül, aynı şekilde erkekler için de kullanılabilmelidir (Kolen ve Brennan, 2004). Aksi takdirde grup değişmezliği varsayımı sağlanmamış demektir. Holland (2007) grup değişmezliği varsayımı karşılanmadan da aynı istatiksel yöntemleri kullanarak test formlarını ilişkilendirebileceğimizi, ancak bu durumda yapılan işlemin bir test eşitlemeden ziyade test uyumlama olacağını vurgulamıştır.

Test eşitlemede veri toplama desenleri

Testler arasında eşitleme yapılabileceğine, yani testlerin benzer güvenirlik, güçlük düzeylerine sahip olduğuna ve aynı yapıyı ölçtüklerine karar verildikten sonra test eşitleme desenine karar vermek gerekmektedir (Bandalos, 2018). Literatürde tanımlanmış birçok veri toplama deseni bulunmaktadır. Desene karar vermek için testlerin kime uygulandığı dikkate alınmalıdır. Testler (1) aynı gruba ya da ölçülmek istenen özellik açısından benzer beceri düzeyine sahip farklı gruplara uygulanabilir. Eğer grupların beceri düzeyleri hakkında fikre sahip değilsek, (2) aynı yapıyı ölçen bir başka testi bütün gruplara uygulayarak grupların beceri düzeyleri hakkında bilgi edinir ve test eşitlemeyi bu temelde yapabiliriz (Livingston, 2014). İkinci alternatifte bahsedilen, test eşitleme sürecinde grupların ölçülmek istenen özellik açısından düzeylerini görmek amacıyla kullanılan bu test türüne *"*bağıntı testi *(anchor test, common equating test* or *common items)"* denmektedir (Lord, 1950; Angoff, 1984; Budescu, 1985; Kolen ve Brennan, 1995). Literatürde yaygın şekilde kullanılan veri toplama desenleri şu şekildedir:

Tek grup deseni (single group design). İki test formunu eşitlerken, uygulanabilecek en basit veri toplama deseni tek grup (TG) desenidir (Dorans ve diğerleri 2011). Bu desende, test eşitleme uygulanacak her iki test formu da belirli bir evrenden çekilen tek bir gruba uygulanır (Gonzales ve Wiberg, 2017). Bundan elde edilen veriler ile test eşitleme sürecine devam edilir. TG deseninde yer alan formlar ve gruplar Şekil 2'de gösterilmektedir.

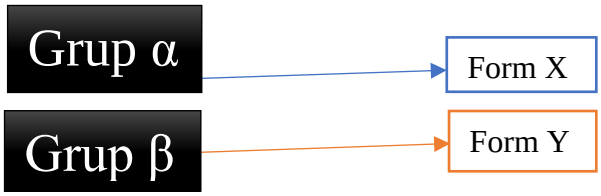


Şekil 2. Tek grup deseni ile test eşitleme

Gonzales ve Wiberg (2017), TG deseni ile ilgili iki dezavantaj tanımlamaktadır. Bunlardan ilki, her iki form da aynı gruba uygulandığı için diğer test eşitleme desenlerine nazaran daha fazla vakit alacaktır. İkinci dezavantaj ise, ilk testi alan grup ikinci testte ya yoruldukları için daha fazla hata yapacaklardır ya da maddelere yönelik aşinalık kazandıkları için gerçek puanlarından daha yüksek gözlenen puanlara sahip olacaklardır. Bu dezavantajların yanı sıra, Bandalos (2018) TG deseninin bir avantajından bahsetmektedir: iki test formunu da alan grup aynı olduğu için, elde edilen puanlara örneklemeden kaynaklanan bir hata karışmayacaktır.

Eşdeğer gruplar deseni (equivelant groups design). Eşdeğer gruplar deseninde (EG), aynı evrenden seçilen bağımsız iki grup ve iki farklı test formu bulunmaktadır. Bu desende, her iki gruba farklı bir form uygulanır (Gonzales ve Wiberg, 2017). Bunlardan elde edilen veriler ile test eşitleme sürecine devam edilir. Bir örnek üzerinde tartışılacak olursa; bir teste ait X ve Y formlarını eşitlemek amacı ile A evreninden bir 300 kişilik örneklem oluşturulduğunu varsayalım. Bu öğrenciler rastgele bir şekilde sıralara oturtulur ve oturma sırasına göre 1'den 300'e kadar numaralar verilir. Tek sayılı öğrencilere X formu, çift sayılı öğrencilere ise Y formu uygulanır. Kolen ve Brennan (1995), bu işlemi bir sarmallama işlemi (spiraling process) olarak adlandırmaktadır. Bu nedenle sarmallama işlemi, desene adını veren eşdeğer grupların oluşturulması için önerilen bir işlemdir (Kolen, 2007). Bunun yanı sıra, örneklem büyüdükçe grupların eşdeğerlik düzeylerinin de artacağı Dorans (1990) tarafından vurgulanmıştır.

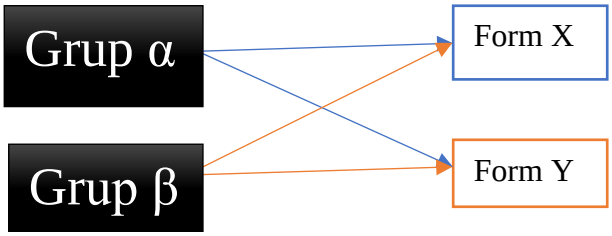
Sarmallama işlemi, Angoff (1984) tarafından sunulan testlerin bireylere seçkisiz olarak atanması işlemine bir alternatif olarak çıkmıştır. Bu nedenle EG deseni, Angoff (1984) ve Kolen ve Brennan (1995) tarafından "Rastgele Gruplar Deseni (Random Groups Design)" olarak adlandırılmaktadır. Ancak, sarmallama hem uygulanabilirlik açısından daha kolay olduğu için hem de örneklemeden kaynaklanan hataya daha az ortam sağlayacağından tercih edilmektedir (Von Davier ve diğerleri, 2004). EG deseninde yer alan formların ve grupların uygulanışı Şekil 3'te gösterilmektedir.



Şekil 3. Eşdeğer gruplar deseni ile test eşitleme

EG deseninin, uygulama açısından TG desenine göre bazı avantajları bulunmaktadır. Kolen ve Brennan (1995), EG deseninde her gruba yalnızca bir form uygulandığı için TG deseninde bahsedilen uzun uygulama süresi, yorgunluk ya da aşinalıktan dolayı oluşan problemlerin olmayacağını vurgulamaktadır. Ayrıca, EG deseninde gruplar eşdeğer oldukları sürece evreni temsil etmek zorunda değildir (Livingston, 2014). Ancak EG deseni, sadece temel bir varsayım olan *grupların eşdeğer beceri düzeyine sahip olması varsayımı* sağlandığı takdirde doğru sonuçlar vermektedir (Lyren ve Hambleton, 2011). Yine de bu varsayımı ihlal etme olasılığı örneklem büyüklüğü arttıkça azalmaktadır (Dorans ve dieğrleri, 2010).

Denkleştirilmiş gruplar deseni (counterbalanced groups design). Denkleştirilmiş gruplar (DG) deseni hem TG hem de EG desenlerinin aynı anda kullanılmasıyla oluşturulan bir desendir (Dorans ve dieğrleri, 2010). Teknik olarak, TG desenine benzer bir şekilde seçilmiş bir grubun rastgele ikiye bölünmesiyle yeni iki grup (α ve β grupları) oluşturulur. Grup α'ya önce ilk form hemen ardından da ikinci form uygulanır. Bu sırada grup β'ya ise önce ikinci form ve ardından ilk form uygulanır. Tüm bu işlemin sonunda, her iki gruba her iki form uygulanmış olacağından hem iki farklı TG deseninden hem de iki farklı EG deseninden elde edilecek veriyi elde etmiş oluruz (Gonzales ve Wiberg, 2017). DG deseni Şekil 4'te gösterilmektedir.

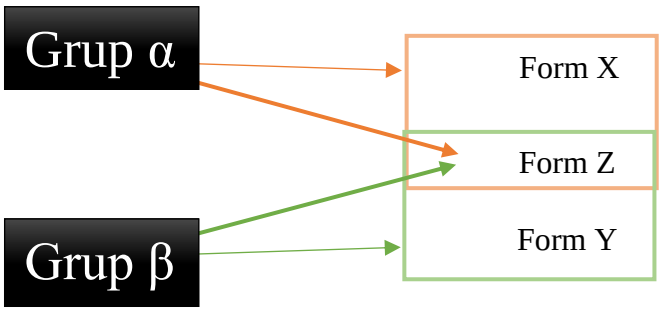


Şekil 4. Denkleştirilmiş gruplar deseni ile test eşitleme

DG desenin en büyük avantajı, EG deseninde de bahsedilen grupların eşdeğer beceri düzeyine sahip olması varsayımının karşılanması şartıyla, bireylerin uygulama esnasında edindikleri aşinalıktan ya da uzun test süresinin bireylerde oluşturduğu yorgunluk etkisinden kaynaklanan problemlerin görece çözüleceğidir. (Livingston, 2014). Ancak, EG deseninde veri toplama süreci özel bir çalışma gerektirdiği için her test eşitleme sürecinde kullanılamamaktadır (Dorans ve diğerleri, 2010).

DG deseninden birçok farklı veri elde edilebilmektedir. Von Davier ve diğerleri (2004), DG deseninden elde edilen verinin kullanılmasında dört farklı yöntem tanımlamıştır, ancak toplanan tüm verinin test eşitleme sürecinde kullanıldığı *iki bağımsız TG yöntemi (the two independent single group method)* tavsiye edilmektedir. Bu yöntemde, uygulamanın sonunda her iki form da grup değişkeni göz ardı edilerek tek bir gruba uygulanmış gibi kullanılır. Böylece DG deseninden elde edilen veri TG deseninden elde edilmiş gibi test eşitleme fonksiyonu oluşturulur.

Eşdeğer olmayan gruplar ile bağıntı testi (EOGB) deseni (Non-equivelant groups anchor test design-NEAT design). Eşdeğer olmayan gruplar ile yapılan test eşitleme desenlerinde, formlarda kullanılan bazı maddelerin birbirlerine benzerlik göstermesi gerekmektedir (von Davier ve diğerleri 2004). Bu desenler bir örnek üzerinden anlatılacak olursa; A evreninden seçilmiş olan α ve B evreninden seçilmiş olan β gruplarının olduğunu varsayalım. Bu durumda X ve Y formlarını eşitlemek için her iki gruba da yine aynı yapıyı ölçen bir üçüncü formun uygulanması gerekmektedir (Kolen ve Brennan, 1995). İki testi eşitlemek amacı ile kullanılan bu üçüncü forma bağıntı testi denmektedir (Livingston, 2014). Bu desen, Şekil 5'te gösterilmektedir.



Şekil 5. EOGB deseni ile test eşitleme

Bağıntı testine, X ve Y formlarının madde güçlük düzeylerini tam anlamıyla temsil ettiği durumlarda *minitest,* sadece orta düzey güçlük indekslerine (-2 ve +2 aralığı) sahip olduğu durumlarda ise *miditest* denmektedir (Sinharay ve Holland, 2006). Bu bağlamda, bir bağıntı testinin eşitlenmek istenen formlara kapsam ve madde güçlükleri açısından ne düzeyde benzemesi gerektiği üzerine tartışmalar olmuştur. Angoff (1984)'a göre, bağıntı testi ana formlara benzer kapsamda ve her iki gruba da aynı şartlarda uygulandığı sürece, nasıl olduklarının bir önemi yoktur. Sinharay ve Holland (2006), bağıntı testinin kapsamının eşitlenen formların kapsamına benzemesi gerektiğini ancak madde güçlük düzeyinin tam anlamıyla formları temsil etmesinin gerekli olmadığını belirtmişlerdir. Klein ve Jarjoura (1985), bağıntı testinin kapsamı eşitlenmek istenen formların kapsamına benzerlik gösterdikçe, test eşitleme hatasının azalacağını vurgulamışlardır. Bağıntı testinin uzunluğu ile alakalı bir çalışmada, Ricker ve von Davier (2007) bağıntı testi kısaldıkça, test eşitleme hatasının artacağını ortaya koymuşlardır. Araştırmacıların bağıntı testi kapsamının X ve Y formlarını temsil etmesi gerektiği üzerine hemfikir olduğu, ancak uzunluğu ve madde güçlükleri açısından ortak bir fikre henüz varamadıkları görülmektedir.

Bağıntı testinden elde edilen puanların, X ve Y formlarının puanlamalarında kullanılmasına ve formların içerisine yerleştirilişine göre iki farklı bağıntı testinden bahsedilmektedir. Eğer bağıntı testinden elde edilen puanlar toplam puanın hesaplanmasında kullanılıyorsa *iç bağıntı testi*, kullanılmıyorsa *dış bağıntı testi* olarak adlandırılmaktadır (Livingston, 2014; Gonzales ve Wiberg, 2017). İç bağıntı testi, X ve Y formlarının ortak maddelerinden oluşturulur ve bu maddeler her iki formda da benzer şekilde yer alan o formların parçalarıdır (von Davier ve diğerleri, 2004). Bu nedenle, iç bağıntı testi terimi yerine *ortak maddeler (common items)* tabiri de kullanılmaktadır.

Dış bağıntı testi ise çoğunlukla X ve Y formlarının sonunda uygulanır ve ne test süresine ne de puanlamasına dâhil edilir (von Davier ve diğerleri, 2004). İç bağıntı testi puanlamada kullanıldığı için tüm uygulamalar tamamlanana kadar gizlilik esastır ve maddeler açıklanmamalıdır. Ancak, dış bağıntı testinde böyle bir zorunluluk yoktur. Bu durum bir avantaj sunarken bir de sorun teşkil etmektedir; dış bağıntı testi ayrı bir test formu olduğu için, kapsam açısından eşitlenmek istenen formları tam anlamıyla temsil etmesi gerekmektedir (Livingston, 2014).

Test eşitleme yöntemleri

Test eşitleme süreci içerisinde, toplanan verinin işlenmesi ile bir eşitleme formülü oluşturulması gerekmektedir (Dorans, 1990). Bu amaç ile kullanılan yöntemlere test eşitleme yöntemleri denmektedir. Bu çalışmada test eşitleme yöntemleri; (1) *Klasik Test Kuramına dayalı test eşitleme yöntemleri* ve (2) *Madde Tepki Kuramına dayalı test eşitleme yöntemleri* olarak ele alınmaktadır.

Klasik test kuramına dayalı test eşitleme yöntemleri. Test eşitlemede veri toplama desenleri bölümünde bahsedilen TG, EG ve DG desenleri kullanılarak bir test eşitleme süreci uygulanacaksa, test eşitlemenin genel varsayımları dışında başka bir varsayımın kontrolüne gerek duyulmamaktadır (Gonzales ve Wiberg, 2017). Bu nedenle aşağıda belirtilen test eşitleme yöntemleri bu desenler söz konusu olduğunda doğrudan kullanılabilmektedir. Ancak EOGB deseni kullanılarak yapılan test eşitleme süreçlerinde üçüncü bir formdan ya da farklı bir madde grubundan elde edilen veriler aracılığıyla dolaylı bir şekilde test eşitleme süreci uygulanmaktadır. Bu nedenle EOGB deseninde kullanılan yöntemler farklı koşullar için çeşitlenmektedir. Bunlar "EOGB Deseninde Kullanılan KTK Temelli Test Eşitleme Yöntemleri" başlığı altında ele alınmaktadır.

Ortalamalar ile test eşitleme (mean equating). Bu yöntem X formu ile Y formu arasındaki farklılığın testi alan her bireyde aynı şekilde görüldüğü varsayımı üzerine kurulmuştur (Kolen ve Brennan, 2004). Başka bir ifade ile X formundan elde edilen bir puanın Y formunda kaç puana karşılık geleceğini anlamak için daha önceden belirlenmiş olan sabit bir sayının ham puana eklenmesi ya da ham puandan çıkarılması gerekmektedir (Livingston, 2014). Bu yöntemi Kolen ve Brennan (2004) aşağıdaki eşitlikle açıklamışlardır:

$$Y(X_i) = X_i + (\mu_{(Y)} - \mu_{(X)}) \quad (1)$$

Bu eşitlikte $Y(X_i)$, X formundan elde edilen herhangi bir puanın Y formu üzerinde alacağı değeri göstermektedir. X_i, X formundan elde edilen herhangi bir puanı göstermektedir. $\mu_{(X)}$ ve $\mu_{(Y)}$ ise sırasıyla X ve Y formlarından elde edilen puanların aritmetik ortalamasını göstermektedir. Buna göre, X formundan elde edilen bir puana, X ve Y formlarının ortalamaları arasındaki farkı eklersek, Y formunda alacağı değeri tahmin etmiş oluruz (Kolen ve Brennan, 2004). Ortalamalar ile test eşitleme işleminin

sonucunda, X formundan alınan her puanın Y formu üzerinde kaç puana denk geleceği görülebilmektedir. Bu işlemin sonucunda elde edilen veri setinin aritmetik ortalaması Y formundan elde edilen ortalama puan ile aynı olacaktır (Livingston, 2014).

Bu yöntem tüm eşitleme yöntemleri içerisinde en basit olanı olarak tanımlanmaktadır (Bandalos, 2018; Livingston, 2014). Bu avantajının yanı sıra bir dezavantajı da bulunmaktadır. X ve Y formlarının güçlük düzeyleri arasındaki olası bir farklılık yüzünden, X formunun uygulandığı bireylerin hepsinin ortalamadan aynı şekilde uzak puanlar elde edeceklerini varsaymak, eşitlenen puanların doğruluğunu düşürecektir (Livingston, 2014). Başka bir ifade ile X formundan elde edilen herhangi bir puana, sadece ortalamalardan elde edilen sabit bir değer eklenmesi test eşitleme hatasını arttıracaktır.

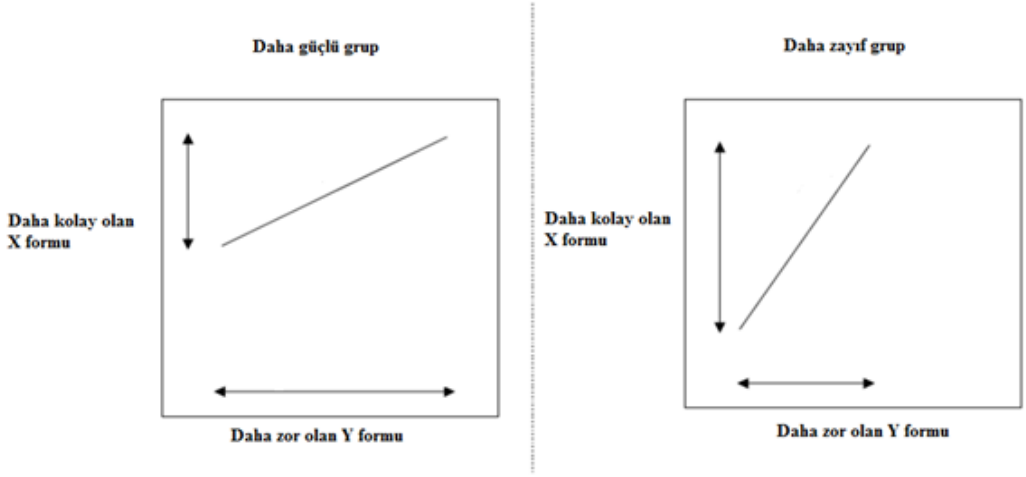
Doğrusal test eşitleme (linear equating). X formundan elde edilen bir puanın Y formundaki bir puana eşit olabilmesi için hem $\boldsymbol{Y(X_i)}$'nin (eşitlenmiş X puanı) hem de $\boldsymbol{Y_i}$'nin (Y formundan elde edilen ham puan) kendi ortalamalarına olan uzaklıklarının z puanı cinsinden aynı olması gerekmektedir (Livingston, 2014). Bu nedenle, doğrusal test eşitlemede, kullanılan eşitleme fonksiyonunun içerisine ortalamaların yanı sıra her iki formun standart sapma değerleri de girmektedir. X formundan elde edilen herhangi bir puanı ($\boldsymbol{X_i}$), Y formuna doğrusal test eşitleme yöntemi ile eşitlemenin temel mantığı Bandalos (2018)'a göre, $\boldsymbol{X_i}$'nin X formundan elde edilen puanların aritmetik ortalamasından ($\boldsymbol{\mu_{(X)}}$) standart sapma ($\boldsymbol{\sigma_x}$) cinsinden kaç birim uzaklaştığı hesabına dayandırılmaktadır. Angoff (1982) tarafından açıklanan doğrusal test eşitleme denklemi şu şekilde gösterilebilir.

$$Y(X_i) = \mu_{(X)} + \frac{\sigma_Y}{\sigma_X}\,(X_i - \mu_{(X)}) \qquad (2)$$

Bu denklemde bir önceki denklemden farklı olarak kullanılan $\frac{\sigma_Y}{\sigma_X}$ matematiksel işlemi; Y formu puanlarından elde edilen standart sapmanın X formu puanlarından elde edilen standart sapmaya oranını göstermektedir. Ancak bu denklem verilerin tek bir evrenden elde edildiği TG, EG ve DG desenleri ile kullanılabilmektedir (Kan, 2011). Doğrusal test eşitleme, ortalamalar ile eşitlemeye göre daha karışık bir işlem süreci gerektirmektedir, ancak yine ortalamalar ile eşitlemeye göre daha esnek bir eşitlik formülü sunmaktadır (Kolen ve Brennan, 1995). Yine de doğrusal test eşitleme

yönteminin de bazı dezavantajları bulunmaktadır (Livingston, 2014). Bunlardan birincisi; X formundan alınan bazı puanlar eşitleme işleminden sonra, Y formundan alınacak maksimum ya da minimum puanı geçebilmektedir. Böyle bir durumda, Kolen ve Brennan (1995) *budama (truncation)* adı verilen bir yöntem önermektedir. Bu yönteme göre, eşitleme fonksiyonundan elde edilen puan, alınabilecek maksimum değeri aşıyorsa maksimum değere sabitlenir, minimum değerin altında kalıyor ise, minimum değere sabitlenir (Kolen ve Brennan, 1995).

Bir diğer problem ise, doğrusal test eşitlemenin formların uygulandığı gruplara bağımlı olmasıdır. Özellikle formların güçlük düzeylerinin farklı olduğu durumlarda, grupta yer alan güçlü/başarılı ve zayıf/başarısız bireylerin arasındaki fark gözle görülür bir şekilde ortaya çıkmaktadır. Bu duruma bir örnek Şekil 6'da gösterilmektedir.



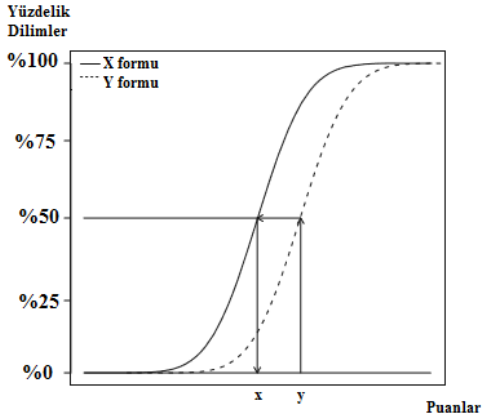
Şekil 6. Güçlü ve Zayıf Gruplarda Doğrusal Test Eşitleme Problemi
Kaynak. Livingston, 2014

İlk grafik, ikinci grafikte gösterilen bireylerden daha güçlü bireylerin doğrusal eşitleme çizgisini göstermektedir. Bu durumda Y formuna göre daha kolay olan X formundan elde edilen puanlar daha dar bir ranjda dağılırken, eşitleme çizgisi de daha yatık bir haldedir. İkinci grafikte ise, tam tersi bir durum görülmektedir.

Yüzdelik dilimler ile test eşitleme (equipercentile equating). Yüzdelik dilimler ile test eşitleme yöntemi ile X formundan elde edilen bir puanın Y formundan elde edilen bir puana eşitlenebilmesi için, aynı yüzdelikte olmaları gerekmektedir (Livingston, 2014). Bir başka ifadeyle, Gonzales ve Wiberg (2017) yüzdelik dilimler ile test eşitlemeyi; X formundan elde edilen x puanlarının Y formundan elde edilen y puanları ile aynı yüzdelik

dilimlere sahip olacak şekilde ayarlanması olarak tanımlamaktadır. Bu durum Şekil 7'de gösterilmektedir.

Şekil 7'de görüldüğü üzere, yüzdelik dilimler ile Y formunu X formuna eşitlerken, Y formundan elde edilen y puanının tekabül ettiği yüzdeliğe denk gelen x puanı bulunur ve eşitlenir (Dorans, 1990). Bir örnek üzerinden açıklayacak olursak; Y formundaki 47 puanın %50'ye tekabül ettiğini varsayalım. X formundaki %50'de yer alan puan ise 45 olsun. Bu durumda Y formundan elde edilen 47 puanın (Y_{47}) X formunda 45 puana (X_{45}) eşitlendiği söylenebilir. Aynı örnek üzerinde başka bir Y puanı 75 ise %83'ye tekabül ediyor olsun. X formunda %83'te yer alan puan ise 76 olsun. Bu durumda $Y_{75} = X_{76}$ eşitliğini elde etmiş oluruz (Bandalos, 2018).



Şekil 7. Yüzdelik Dilimler ile Test Eşitleme
Kaynak. Gonzales ve Wiberg, 2017

X ve Y formlarından elde edilen puanların dağılım eğrileri birbirlerine ne kadar benzerse, yüzdelik dilimler ile doğrusal test eşitleme yöntemleri o kadar benzer sonuçlar verecektir (Livingston, 2014). Bu nedenle, puanların dağılımlarının birbirine benzediği formları eşitlerken, kullanılan yöntemin yüzdelik dilimler ile test eşitleme ya da doğrusal test eşitleme olması büyük bir fark yaratmamaktadır. Yüzdelik dilimler ile test eşitleme yönteminin avantajı, bahsi geçen bu dağılımların birbirlerine benzemediği durumlarda test eşitlemeyi daha az test eşitleme hatası ile yapabilmemizi sağlamasıdır. Bir diğer avantaj ise, hiçbir puanın maksimum ya da minimum puanları geçmeyecek olmasıdır (Bandalos, 2018).

Yüzdelik dilimler ile test eşitleme yöntemlerinin doğası gereği bazı problemlerle karşılaşma olasılığı vardır (Kolen ve Brennan, 2004). Bu problemler (*düzensizlik* ve

ayrıklık) ve bu problemlere sunulmuş bazı çözüm önerileri (*düzleştirme* ve *ara değer hesaplama*) iki başlık altında anlatılacaktır.

1.Düzensizlik problemi (irregularity) ve Düzleştirme (Smoothing). Yüzdelik dilimlerle test eşitlemenin dezavantajlarından birisi puanların düzensiz dağılım göstermesidir (Livingston, 2014). Özellikle küçük gruplarla yapılan yüzdelik dilimler ile test eşitlemelerde, örneklemeden kaynaklanan hata düzensizlik problemini doğurmaktadır (Kolen, 1988). Livingston (2014) bu problemin kaynağını örneklemden elde edilen bazı puanlarda evrende olduğundan daha yüksek yüzdelik değerlerde bireyler yer alırken, bazılarında ise evrende olduğundan daha düşük yüzdelik değerlerde bireylerin yer almasına bağlamaktadır. Bu durum yüzdelik dağılım grafiğinde dalgalı bir görüntü oluşturmaktadır (Livingston, 2014). Lord (1982), böyle bir durumda örneklemin evreni daha fazla temsil edebilmesi için dağılım üzerinde ayarlamalar yapılmasını önermiştir. Bu ayarlamalara düzleştirme (smoothing) denir (Livingston, 1993). Örneklem büyüdükçe, evreni daha fazla temsil edeceği varsayıldığından; düzleştirme işlemi daha az etkili olmaktadır. Tam tersi durumlarda ise, yani örneklem küçüldükçe, düzleştirme işlemi daha başarılı sonuçlar vermektedir (Livingston, 1993).

Kolen (2007), dört farklı düzleştirme yönteminden bahsetmektedir. Bunlar; *pre-smoothing, post-smoothing, log-linear pre-smoothing* ve *cubic spline post-smoothing* olarak adlandırılmıştır, ancak bu yöntemler, bu çalışmanın konusu dışındadır. Bu nedenle detaylarına girilmeyecektir. Düzleştirme yöntemleri hakkında detaylı bilgi Livingston(1993), Kolen ve Brennan (2004) ve von Davier ve diğerleri (2004)'nden edinilebilir.

2. Ayrıklık problemi (Discreteness) ve Ara Değer Hesaplama (Interpolation). Yüzdelik dilimlerle test eşitlemenin bir diğer dezavantajı ise, X formundan elde edilen bir puanın yüzdelik değerine karşılık gelen Y formundan elde edilmiş bir puanın olmadığı durumlardır. Bu duruma ayrıklık problemi (discreteness) adı verilmektedir (Kolen ve Brennan, 2004). Bu durumla karşılaşma riski grup büyüdükçe azalsa da oldukça fazladır (Bandalos, 2018). Ayrıklık problemine bir çözüm önerisi Livingston (2014) tarafından Tablo 1'deki gibi bir örnek üzerinde açıklanmıştır.

Tablo 1.

Yüzdelik dilimler ile test eşitlemede ayrıklık problemi

X formu		Y formu	
Ham puan	Yüzdelik derece	Ham puan	Yüzdelik derece
52	78.07	52	68.96
51	74.95	51	65.09
50	71.64	50	61.12
49	68.18	49	57.07
48	64.60	48	52.99
47	60.92	47	48.93
46	57.18	46	44.93

Kaynak. Livingston, 2014

Tablo 1 incelendiğinde, X formundan alınan 47 puanın yüzdeliği %60.92 olarak görülmektedir. Ancak Y formunda %60.92'ye tekabül eden bir puanın olmadığı görülmektedir. Buna en yakın yüzdelik dilimin %61.12 ile %57.07 (49 ile 50 puan) aralığı olduğu görülmektedir. Bu durumda, X formundan alınan 47 puanın Y formundaki değeri şu hesapla tahmin edilebilir (Livingston, 2014):

$$49+\frac{60.92-57.07}{61.12-57.07}(50-49)=49.95$$

Bu hesaplamadan elde edilen sonuç, X formundan elde edilen 47 puanın Y formunda 49.95 puana denk geleceğidir. Bu hesaplamayı kullanarak, X formundan elde edilen herhangi bir puanın Y formunda tam olarak kaç puana tekabül edeceğini hesaplamak için, o puana ait ilgili değerlerin yukarıdaki formülde uygulanması gerekmektedir. Bu puan ayarlama sürecine ise *ara puan hesaplama (interpolation)* adı verilmektedir (Livingston, 2014; Bandalos, 2018).

Kernel test eşitleme. Kernel test eşitleme (KE) yöntemi, Holland ve Thayer (1989) tarafından geliştirilmiş, diğer geleneksel test eşitleme yöntemlerine göre daha karışık istatiksel süreçler barındıran bir yöntemdir. İçerisinde hem yüzdelik dilimler ile test eşitleme hem de doğrusal test eşitleme yöntemlerine ait bazı temel özellikleri barındırmaktadır (von Davier ve diğerleri, 2004). Mao, von Davier ve Rupp (2006), KE'nin KTK temelli yöntemleri de kapsayan genel bir çerçeve sunduğunu savunmuşlardır.

von Davier ve diğerleri (2004), KE yönteminin geleneksel test eşitleme yöntemlerine göre dezavantaj ve avantajlarını bir kitap bölümü olarak detaylı bir şekilde

aktarmışlardır. Buna göre, KE'nin tek dezavantajı geleneksel yöntemler kadar kolay hesaplanabilir olmamasıdır. Ancak, bilgisayarlar aracılığıyla istatiksel hesaplamaların artık çok hızlı yapılabiliyor olması bunu bir dezavantaj olmaktan çıkarmaktadır. KE'nin avantajlarından bazıları ise; (1) örneklemeden kaynaklanan hatanın en aza indirilmesi ve (2) her türlü örneklem büyüklüğünde çalışmaya elverişli olmasıdır. İlgili kitap bölümünde, KE'nin geleneksel yöntemlere göre sahip olduğu diğer avantajlar her yöntem için ayrı ayrı ele alınmıştır.

KE yöntemi, beş temel adımda ilerlemektedir (von Davier ve diğerleri, 2004). Bunlar: (1) düzleştirme (smoothing), (2) puan olasılık tahmini (estimation of the score probabilities), (3) sürekliliği sağlama (continuization), (4) eşitleme (equating) ve (5) eşitlemenin standart hatasını hesaplama (calculating the standard error of equating-calculating SEE). Bu bölümde bu beş adım kısaca açıklanacaktır.

1. *Düzleştirme.* KE'nin bu aşamasında, yüzdelik dilimler ile test eşitleme yönteminde bahsedildiği gibi, veri toplama sürecinde kullanılan test eşitleme desenine göre en uygun düzleştirme yöntemi seçilir ve veriye daha düzgün bir şekil kazandırılır (Holland ve Thayer 1989; Sansivieri ve Wiberg, 2017).

2. *Puan olasılık tahmini.* Bu aşamada hedef evrene ait puan dağılımları, düzleştirilmiş gözlenen puanlar aracılığıyla tahmin edilir.

3. *Sürekliliği sağlama.* Bu aşamada veri sürekliliği sağlamak amacıyla, yüzdelik dilimler ile test eşitleme yönteminde bahsedilen ara değer hesaplama sürecine benzer bir süreç işletilir (Mao ve diğerleri, 2006). Bu süreçte, veri seti arasındaki boşluklara sabit aralıklarla bir değer atamak tavsiye edilmektedir. Bu sabit değerin belirlenmesi için kullanılacak fonksiyonlar, Holland ve Thayer (1989) tarafından belirtilmiştir. KE'ye bu aşamada kullanılan Gaussian Kernel düzleştirme yönteminden dolayı Gaussian Kernel Test Eşitleme Yöntemi adı da verilmektedir (von Davier, 2007).

4. *Eşitleme.* Bu aşamaya kadar veriler eşitleme fonksiyonunu uygulamaya hazırlanmıştır. Eşitleme aşamasında ise, test eşitleme desenine göre en uygun test eşitleme fonksiyonu uygulanır. Örneğin; koşullu test eşitleme ve zincir test eşitleme yöntemlerine ait fonksiyonlar kullanılabilmektedir (Ricker ve von Davier, 2007).

5. *Eşitlemenin standart hatasını hesaplama.* Beşinci, yani son aşamada, test eşitleme hatası hesaplanır. Bununla ilgili detaylar, bu çalışmanın ilerleyen bölümlerinde ele alınacaktır.

Daire-yay test eşitleme (circle-arc equating). Bu yöntem ile, doğrusal olmayan dağılıma sahip verilerde ve küçük örneklemlerde diğer yöntemlere göre daha az eşitleme hatası hesaplanmaktadır (Livingston ve Kim, 2009). Altında yatan temel fikir test eşitlemede kullanılan veri dağılım grafiğinin üzerinde önceden belirlenmiş üç noktadan geçecek bir eğri çizerek eşitleme fonksiyonunu elde etmek üzerine kuruludur. Bu eğrinin başlangıç ve bitiş noktaları testten elde edilebilecek maksimum ve minimum puanlar olarak tanımlanmıştır. Orta nokta ise ortalama puanlar kullanılarak belirlenmektedir. Bu üç noktadan geçen yay şeklindeki bu çizgi eşitleme eğrisini oluşturmaktadır. Bu yöntemle, küçük örneklemlerde karşımıza çıkan en yaygın problem olan belirli puan noktalarında veri olmaması durumunun test eşitleme sürecine etkisi en düşük seviyeye indirmek amaçlanmaktadır. Livingston ve Kim (2009), bu yöntemle elde edilen hata puanlarına ait grafiğin yüzdelik dilimler ile test eşitleme yönteminden elde edilen hata puanları grafiğine çok benzediğini belirtmiştir. Bu nedenle, küçük örneklemler için en uyun yöntem olduğu düşünülmektedir (Livingston ve Kim, 2009).

Özdeş test eşitleme (ıdentity equating). Özdeş test eşitleme yöntemi eşitlenmek istenen formlardan elde edilebilecek maksimum ve minimum puanlarla doğrudan ilişkilidir. Eğer maksimum ve minimum puanlar her iki formda da aynı ise her iki formdan elde edilen puanlar eşdeğer olarak kabul edilir. Bu koşulda, Kolen ve Brennan (2004)'e göre eşitleme yapmamak ile özdeş eşitleme yapmak aynı sonucu verecektir. Ancak maksimum ve minimum puanlar eşit değilse, küçük örneklemlerde önerilen bir yöntemdir. Bu durumda özdeş test eşitleme ile Y formundan elde edilen $\boldsymbol{y}$ puanının X formunda eşdeğeri olan $\boldsymbol{y}_{(x)}$'i bulmak için- $\boldsymbol{y_1}$ ve $\boldsymbol{x_1}$sırasıyla Y ve X formlarından alınabilecek maksimum puanlar, $\boldsymbol{y_2}$ ve $\boldsymbol{x_2}$ ise minimum puanlar olmak üzere- şu eşitlik ile bulunabilmektedir:

$$y_{(x)} = \frac{y_2 - y_1}{x_2 - x_1} x + y_1 - \frac{y_2 - y_1}{x_2 - x_1} x_1 \qquad (3)$$

Sentetik test eşitleme. Kim, von Davier ve Haberman (2008), özdeş test eşitleme ve herhangi bir başka test eşitleme yönteminin ağırlıklandırılmış bir birleşimi olan sentetik test eşitleme yöntemini sunmuşlardır. Buna göre sentetik eşitleme ile eşitlenmiş bir $Sen_X(y)$ puanı şu eşitlikle bulunabilmektedir:

$$Sen_X(y) = we_X(y) + (1 - w)\ddot{O}z_X(y) \quad (4)$$

Burada, w 0 ile 1 arasında değişen ağırlıklandırma puanı, e özdeş eşitleme fonksiyonu dışında herhangi başka bir test eşitleme fonksiyonunu, $\ddot{O}z$ ise özdeş test eşitleme fonksiyonunu temsil etmektedir (Babcock, Albano ve Raymond, 2012).

EOGB deseninde kullanılan KTK temelli test eşitleme yöntemleri. Test eşitleme sürecinde verilerin elde edildiği gruplar aynı evrene ait gruplar değiller ise bağıntı testi kullanılarak bir ilişkilendirme yapılır. Eşdeğer olmayan gruplar ile bağıntı testi deseni ile test eşitleme uygulaması yapmak için birçok test eşitleme yöntemi sunulmuştur. Bunlar doğrusal eşitleme ve yüzdelik dilimler ile eşitleme temelinde geliştirilmiş yöntemlerdir. Bu çalışmada doğrusal eşitleme temelinde uygulanan "Levine doğrusal test eşitleme", "Tucker doğrusal test eşitleme", "düşük ağırlıklar test eşitleme", "zincirleme doğrusal test eşitleme" ve "Braun-Holland test eşitleme" yöntemleri ele alınmıştır. Ayrıca yüzdelik dilimler temelinde yürütülen "yüzdelik dilimler ile zincirleme test eşitleme" ve "yüzdelik dilimler ile frekans tahmini test eşitleme" yöntemleri de ele alınmıştır.

1. Levine doğrusal test eşitleme. EOGB deseni ile kullanılabilen doğrusal test eşitleme yöntemleri, eşitleme fonksiyonunda kullanılan parametreleri belirlerken benimsedikleri yaklaşımlar açısından birbirlerinden ayrılmaktadır. Levine (1955) tarafından sunulan ve hem gözlenen hem de gerçek puanlar üzerinden iki farklı yaklaşım sunan bu yöntemin KTK temelli bazı teorik varsayımları bulunmaktadır. Bunlar; (1) X ve Y formlarına ait gerçek puanlar ile bağıntı testine ait gerçek puanların arasındaki korelasyonun 1 olması, (2) hem bağıntı testi hem de X ve Y formlarına ait hata varyanslarının eşit olması ve (3) hem bağıntı testi hem de X ve Y formlarına ait doğrusal regresyon katsayılarının eşit olmasıdır.

2. Tucker doğrusal test eşitleme. Bir diğer doğrusal yöntem olan Tucker test eşitleme, iki temel varsayıma sahiptir. Bunlar; (1) X ve Y formlarının koşullu beklenen

değerleri her iki evren için de eşit olmalıdır ve (2) X ve Y formlarının koşullu varyansları her iki evren için de eşit olmalıdır (Gonzales ve Wiberg, 2017).

3. Düşük ağırlıklar doğrusal test eşitleme (nominal weights equating). Levine test eşitleme yöntemi gibi bu yöntem de formlar arasında kusursuz bir korelasyonun olması varsayımına dayanmaktadır (Gonzales ve Wiberg, 2017). Albano (2016)'ya göre küçük örneklemler için ideal bir test eşitleme yöntemidir.

4. Zincirleme test eşitleme (chained equating). Zincirleme test eşitleme yöntemi, EOGB deseni ile iki formu eşitlemek için kullanılan en eski yöntem olarak kabul edilmektedir (von Davier ve diğerleri, 2004). Bu yöntemde önce X formu ile bağıntı testi, sonra da Y formu ile bağıntı testi ilişkilendirilir (Livingston, 2014). Bu iki ilişkilendirme fonksiyonu kullanılarak, X ve Y formları arasında bir test eşitleme fonksiyonu elde edilir. Eğer bu fonksiyonlar doğrusal yöntemlerle elde edilirse "zincirleme doğrusal test eşitleme" yüzdelik dilimler ile elde edilirse "yüzdelik dilimler ile zincirleme test eşitleme" adını almaktadır.

5. Yüzdelik dilimler ile frekans tahmini test eşitleme (frequency estimation equating). Katmanlı test eşitleme (post-stratification equating - PSE) olarak da anılan (Doran ve diğerleri, 2011) bu yöntemde ise, bağıntı testinden elde edilen puanlar bir yordayıcı değişken gibi kullanılır. Burada, bağıntı testi aracılığı ile tüm evrenin X ve Y formlarına ait ortalamaları ve standart sapmaları tahmin edilir (von Davier ve diğerleri, 2004; Dorans ve diğerleri, 2011). Daha sonra bu tahminler, test eşitleme fonksiyonunun oluşturulmasında kullanılır (Livingston, 2014).

6. Braun-Holland doğrusal test eşitleme. Bu yöntem frekans tahmini test eşitleme yönteminin doğrusal bir yaklaşımı temel alması ile elde edilmektedir. Bu nedenle hem yüzdelik dilimler ile test eşitleme hem de doğrusal test eşitleme yöntemlerine ait özellikler barındırmaktadır (Gonzales ve Wiberg, 2017). Braun & Holland (1982), bu yöntemde dikkat edilmesi gereken iki varsayımdan bahsetmektedir. Bunlar, (1) regresyon fonksiyonunun doğrusallığı ve (2) varyansın homojenliğidir.

Madde tepki kuramına dayalı test eşitleme yöntemleri. MTK temelli test eşitleme dört aşamadan oluşmaktadır (Hambleton ve Swaminathan, 1985). Bunlar; (1) formların ve grupların durumlarına göre bir desenin belirlenmesi, (2) Uygun MTK temelli test eşitleme yöntemine karar verilmesi, (3) yetenek veya madde parametrelerine göre ortak bir ölçünün oluşturulması ve (4) test puanlarını raporlaştırırken bir ölçek üzerinde karar kılınması olarak tanımlanmıştır. Gözlenen ya da gerçek puanlar üzerinden değil de yetenek parametresi üzerinden bir raporlaştırma yapılacaksa, dördüncü aşamaya gerek kalmamaktadır.

Birinci aşama süreçleri, bu çalışmada "test eşitlemede veri toplama desenleri" başlığında ele alınmıştır. İkinci aşamada ise, test eşitleme deseni aracılığı ile toplanan verinin uygulanacak MTK temelli test eşitleme yöntemine uydurulması gerekmektedir (Hambleton ve Cook, 1977; Cook ve Eignor, 1989). *Model-veri uyumu* olarak adlandırılan bu varsayım sağlandıktan sonra o yönteme ait süreçler işleme konulmalıdır. Ancak, eğer bu varsayım sağlanamıyorsa ya veri setinin ya da modelin değiştirilmesi düşünülmelidir (Skaggs ve Lissitz, 1986).

Üçüncü aşamada, bir önceki aşamada bahsedilen varsayımın sağlandığı model kullanılarak iki formdan elde edilen yetenek düzeyleri yetenek düzeyleri (θ), ortalaması 0 ve standart sapması 1 olan standart puanlara dönüştürülerek ortak bir ölçek üzerinde gösterilir (Hambleton ve Swaminathan, 1985). Bu işlem için farklı modeller birçok kaynakta sunulmaktadır (Hambleton ve Swaminathan, 1985; Hambleton, Swaminathan ve Rogers, 1991; Kolen ve Brennan, 1995; Sansivieri ve diğerleri, 2017). Bu modeller, testlerin iki kategorili puanlanan ya da çok kategorili puanlanan maddelerden oluşmasına göre belirlenebilmektedir (Kolen ve Brennan, 1995).

MTK temelli test eşitleme yöntemleri arasında iki kategorili puanlanan testlerin eşitlenmesinde en yaygın kullanılan üç model bu çalışmanın kapsamı içerisinde kısaca açıklanmıştır (Skaggs ve Lissitz, 1986; Hambleton ve Jones, 1993). Bunlardan ilki *Rasch modeli* ya da *tek parametreli lojistik modeldir (1PL)*. Bir diğeri *iki parametreli lojistik model (2PL)* ve sonuncusu *üç parametreli lojistik modeldir (3PL)*. Ayrıca çok kategorili puanlanan maddelerden oluşan testlerin eşitlenmesinde kullanılan *genelleştirilmiş kısmi kredi modeli (generalized partial credit model- GPCM)* bu çalışmanın kapsamı içerisindedir.

İki kategorili maddelerde kullanılan üç modelde *madde karakteristik eğrisi- MKE (item characteristic curve- ICC)'*ni oluşturmak amacı ile farklı fonksiyonlar ve her fonksiyon için farklı parametreler devreye girmektedir. Bu parametreleri hesaplarken kullanılan fonksiyonlar Birnbaum (1968), Hambleton ve Swaminathan, (1985), Hambleton ve diğerleri (1991), Baker ve Kim (2004)'de bulunabilir. Bu modellere ait fonksiyonlar Tablo 2'de Hambleton ve Swaminathan, (1985)'dan uyarlanarak gösterilmiştir. Tabloda ele alınan modellerde kullanılan parametreler ve değişkenler ise şunlardır;

D = 1.7 : Lojistik ve ogive model fonksiyonlarının birbirlerine çok yakın sonuçlar vermesi için kullanılan bir sabit değer,
$a_m = e^{Da}$: m maddesinin ayırt edicilik parametresi,
$b_m = e^{Db}$: m maddesinin güçlük parametresi,
c_m : şans parametresi,
$\theta_b = e^{D\theta}$: b bireyinin yetenek düzeyi.

Tablo 2.

MTK temelli modeller ve MKE oluşturma fonksiyonları

Model	Fonksiyon
1PL	$P(\theta) = \frac{e^{D(\theta - b_m)}}{1+e^{D(\theta - b_m)}}$
2PL	$P(\theta) = \frac{e^{Da_m(\theta - b_m)}}{1+e^{Da_m(\theta - b_m)}}$
3PL	$P(\theta) = c_m + (1 - c_m)\frac{e^{Da_m(\theta - b_m)}}{1+e^{Da_m(\theta - b_m)}}$

Kaynak. Hambleton ve Swaminathan, 1985

Üçüncü aşamada Tablo 2'de belirtilen modellerden birisi ile bir madde karakteristik eğrisi oluşturulur. Bu işlemin ardından, bireylerin yetenek düzeyleri ve madde parametreleri belirli dönüşüm fonksiyonları kullanılarak tek bir ölçek üzerinde tahmin edilir. Bu dönüşüm fonksiyonları, Gonzales'ten (2014) uyarlanarak Tablo 3'de derlenmiştir.

Tablo 3.

MTK temelli eşitlemelerde parametre dönüşüm fonksiyonları

Parametre	Dönüşüm Fonksiyonu
a_{my}	$\frac{a_{mx}}{A}$
b_{my}	$A\frac{b_{mx}}{B}$
c_{my}	c_{mx}
θ_{by}	$A\,\theta_{bx} + B$

Kaynak. Gonzales, 2014

Bu fonksiyonlarda X ve Y değişkenleri, ölçekler arasındaki farkı belirtmek için eklenmiştir. A ve B değerleri ise doğrusal eşitleme fonksiyonunda eğim ve sabit değerleri olup farklı yöntemler ile hesaplanabilmektedir (Hambleton ve Swaminathan, 1985).

Çok kategorili puanlanan testlerde kullanılan modellerden birisi olan GPCM'nin altında yatan temel fikir ise, madde kaç kategorili olursa olsun, bireyin seçtiği kategori ve bir önceki kategori arasında iki kategorili bir ilişki kurmasıdır (Muraki, 1992). Bu nedenle art arda gelen her kategori çiftleri için ayrı ayrı güçlük parametreleri belirlenir. Bunlar *threshold parametre (eşik parametresi)* olarak adlandırılır. Bu nedenle her bir maddeye ait *k* sayısı kadar kategori varsa, *k-1* kadar eşik parametresi hesaplanır ve bunlar 2PL model gibi eşitleme süreci uygulanır.

Test eşitleme süreci

Test eşitleme sürecinde izlenmesi gereken birçok adım vardır. Bu adımlar uygulanırken, uygulayıcının hatırı sayılır miktarda karar vermesi ve yargıda bulunması gerekmektedir. Bu nedenle test eşitleme yapacak uygulayıcıların yeterli donanım ve tecrübeye sahip olması oldukça önemlidir (Kolen ve Brennan, 2004). Bunun yanı sıra, Kolen ve Brennan (2004), test eşitleme sürecini yedi adımda tanımlamışlardır. Eğer eşit iki test formu geliştirilmek isteniyorsa, ilk dört adımın form oluşturma sürecinde uygulanması gerektiği düşünülmektedir. Eğer oluşturulmuş ve uygulanmış iki formun test eşitleme çalışması yürütülecekse, ilk dört adımın atlanarak, son üç adımın izlenmesi gerektiği düşünülmektedir.

1. Amacın belirlenmesi. Daha önce de bahsedildiği gibi ölçme değerlendirme süreçlerinde; güvenlik, test maddelerinin kamu ile paylaşılması ve test içeriklerindeki değişimler gibi sebeplerden dolayı yeni test formları hazırlanmasına ihtiyaç duyulmaktadır. Bu bağlamda test eşitleme işleminin amacının da belirlenmesi gerekmektedir.

2. Taslak paralel formların oluşturulması. Günümüzde aynı yapıyı ölçmek amacıyla farklı maddeleri içeren test sürümleri oluşturulur. Bu farklı sürümlere "form" adı verilmektedir (Livingston, 2014). Test eşitleme sürecinin bu aşamasında aynı yapıyı ölçen formlar birebir aynı test geliştirme aşamaları izlenerek oluşturulur.

3. Veri toplama deseninin belirlenmesi. İstatiksel olarak paralel formların arasındaki farklılıkların ortaya konabilmesi için bu formların uygulanması gerekmektedir. Bu aşamada, test eşitlemede veri toplama desenlerinde bahsedilen desenler aracılığı ile yapılabilmektedir. Test eşitlemenin amacı doğrultusunda bir desen üzerinde karar kılınır.

4. Verilerin toplanması. Bir önceki aşamada belirlenen desen kullanılarak, ikinci aşamada oluşturulan paralel formlara ait veriler toplanır. Bu aşamanın tekrar edilmesi diğer aşamalara göre daha zordur. Bu nedenle sistematik bir çalışma gerekmektedir. Örneğin; bir teste ait X ve Y formlarını eşdeğer gruplar deseni aracılığı ile eşitlemek istediğimizi varsayalım. Bunun için 300 kişilik bir örneklem olduğunu düşünürsek; bu öğrenciler rastgele bir şekilde sıralara oturtulur ve oturma sırasına göre 1'den 300'e kadar numaralar verilir. Tek sayılı öğrencilere bir form, çift sayılı öğrencilere ise diğer form uygulanır. Kolen ve Brennan (1995), bu işlemi bir sarmallama süreci (spiraling process) olarak tanımlamakta ve eşdeğer gruplar deseni için seçkisizliği sağlamak adına önermektedir.

5. Varsayımların kontrolü. Bu aşamada, test eşitleme sürecinin hangi yöntemler ile yapılacağına karar verilir. Bu aşamada, daha önceki dört aşamanın uygulanıp uygulanmadığına bakılmadan, test eşitleme yöntemlerine ait varsayımlar kontrol edilir. Bu varsayımlar formlarla ya da formlardan elde edilen verilerle ilgili olabilir. Bu çalışmada, yöntemlere ait varsayımlara o yöntemin tartışıldığı başlıkta değinilmiştir.

6. İstatiksel yöntemin belirlenmesi. Bir önceki adım ile ilişkili olarak, varsayımların karşılandığı test eşitleme yöntemleri içerisinden bir ya da daha fazla istatiksel yöntem belirlenir. Örneğin, doğrusal yöntemler arasından Levine doğrusal gözlenen puan eşitleme yöntemi istatiksel yöntemlere bir örnektir. Belirlenen istatiksel yöntemler uygulanır ve test eşitleme hataları uygun yöntemlerle hesaplanır. Test eşitlemenin amacı doğrultusunda varılmak istenen karar verilir.

7. Test eşitleme sonuçlarının değerlendirilmesi. Burada belirtilen altı adımın sonucunda; formların oluşturulması, testlerin uygulanması, kullanılan istatiksel yöntemler gibi süreçlerin tamamının değerlendirilmesi gerekmektedir. Test eşitleme sürecinin değerlendirilmesi aşamasında faydalanılabilecek bazı istatiksel yöntemler ve

kriterler Kolen ve Brennan (2004) tarafından belirtilmiştir. Test eşitleme süreçlerinin uygulanma sırası amaç doğrultusunda değişiklik gösterebilmektedir. Ancak dikkat edilmesi gereken en önemli husus olarak varsayımların kontrolünden bahsedilmektedir (Bandalos, 2018). Denenen yöntemlerin varsayımlarının karşılandığından emin olmak, test eşitleme sürecine sistematik eşitleme hatasının karışmaması için önemlidir.

Test eşitlemede hata

Test eşitleme hatası iki başlık altında (1) *sistematik eşitleme hatası* ve (2) *rastgele eşitleme hatası* olarak tanımlanmaktadır (Kolen, 1988). Bir test eşitleme işlemine ait sistematik eşitleme hatası, yanlılık (*bias)* çalışmalarıyla tahmin edilebilir (Gonzales ve Wiberg, 2017). Rastgele eşitleme hatası ise, örneklem ile alakalıdır (Bandalos, 2018). Bu bağlamda incelendiğinde, bu çalışmada "örneklemeden kaynaklanan hatalar" olarak ele alınan hata türü, rastgele eşitleme hatasına sebep olmaktadır. Rastgele eşitleme hatası, *eşitlemenin standart hatası- ESH (standard error of equation - SEE)* hesaplanarak tahmin edilebilmektedir.

Bir test eşitleme sürecinde ortaya çıkan hatayı tahmin etmek için geliştirilen yöntemlerden birisi bootstrap yöntemidir (Kolen ve Brennan, 2004). Bu yöntemde, bir bilgisayar programı aracılığı ile bir test eşitleme verisi kullanılarak X formuna ait x_i puanının Y formuna ait ölçek üzerinde birçok tekrarlı tahmini puanı oluşturulur. Yapılan bu tekrarların sayısı R iken; her bir tekrardan elde edilen tahmini puanların ortalamasına $Y_{\cdot}(x_i)$, bu tahmini puan ortalamalarının tüm tekrarlardan elde edilen ortalamasına ise $Y_R(x_i)$ dersek, x_i puanının Y formu ölçeği üzerindeki değerinin eşitleme standart hatası (ESH_{boot}) şu eşitlikle tahmin edilebilmektedir (Kolen ve Brennan,2004):

$$ESH_{boot}\,[Y\,(x_i)] = \sqrt{\frac{\sum[Y_{\cdot}(x_i) - Y_R(x_i)]^2}{R-1}} \qquad (5)$$

Daha önce de belirtildiği üzere; gerçek veri setinde x_i puanında yer alan birey sayısı arttıkça, o puana ait ESH düşük olacaktır.

Sistematik eşitleme hatasını *(yanlılık- bias)* hesaplarken ise, şu eşitlik kullanılabilir (Gonzales ve Wiberg, 2017):

$$Bias[Y\,(x_i)] = \frac{\sum[Y_r(x_i) - Y(x_i)]}{R} \qquad (6)$$

Eşitlik içerisindeki x_i X formundan elde edilen herhangi bir puanı, $Y(x_i)$ bu puana ait eşitlenmiş gerçek puanı ve $Y_r(x_i)$ bootstrap yönteminden elde edilen tekrarlı tahmini puan örneklemlerinden herhangi bir r örneklemindeki x_i puanını temsil etmektedir.

Bu iki hata türü ve tahmin yöntemlerinden yola çıkarak, test eşitlemede oluşan tüm hata tahmin edilebilmektedir. *Ortalama hata karekökü-OHKK (Root mean squared errors- RMSE)* olarak adlandırılan bu bütünleşik hata türü ise şu denklemle hesaplanabilmektedir (Gonzales ve Wiberg, 2017):

$$OHKK(x_i) = \sqrt{Bias(x_i)^2 + ESH(x_i)^2} \quad (7)$$

Test eşitleme sürecinde örneklem büyüklüğü ESH açısından önem arz etmektedir. Test eşitleme sürecinin ideal şekilde ilerleyebilmesi için yeterli büyüklükte örneklemin bulunması gerekmektedir. Bu koşul, uygulanan testin türüne, uzunluğuna, amacına göre değişiklik gösterebilmektedir (Kolen ve Brennan, 2004).

İlgili Araştırmalar

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

Türkiye Dışında Yapılan Çalışmalar

Türkiye dışında küçük örneklemler kullanılarak yapılan test eşitleme çalışmalarının önemli bir kısmı iki kategorili puanlanan maddelerden oluşan testlerin eşitlenmesi üzerinedir (Kolen, 1985; Livingston, 1993; Hanson, Zeng ve Colton, 1994; Parshall, Houghton ve Kromrey, 1995; Skaggs, 2005; Mao ve diğerleri, 2006; Sinharay ve Holland, 2006; Kim ve diğerleri, 2008; Livingston ve Kim, 2009; Babcock ve diğerleri, 2012; Albano, 2015). Çok kategorili testlerin ilişkilendirilmesinde ise çoğunlukla MTK temelli yöntemlerin kullanıldığı görülmektedir (Baker, 1992; Li, Lissitz ve Yang, 1999; Tate, 1999; Lee, Kolen, Frisbie ve Ankenmann, 2001). Hem küçük örneklemlerin hem de çok kategorili puanlanan testlerin birlikte incelendiği çalışmalar sınırlıdır.

Kolen (1985), hem gerçek hem de simülatif veriler ile bir çalışma yürütmüştür. Çalışma kapsamında Tucker doğrusal test eşitleme yöntemi kullanılmıştır ve üç farklı test eşitleme hatası hesaplama yöntemi (Delta yöntemi, Bootstrap yöntemi ve Tucker Standart

hataları yöntemi) karşılaştırılmıştır. Çalışmanın birinci kısmında; bir başarı testinin simülasyonu yapılmış ve normallik varsayımını karşılayan ve karşılamayan iki koşul altında veriler üretilmiştir. 100 ve 250 kişilik iki örneklem durumu, 15 maddelik bir bağıntı testi ve 52'şer maddelik iki form üzerinden EOGB deseni ile simüle edilmiştir. Çalışmanın ikinci kısmında ise; 773 ve 795 bireyden oluşan iki farklı gruba EOGB deseni ile uygulanan 125 maddelik bir testin iki formunun eşitleme çalışması yürütülmüştür. Her iki formun 30'ar maddesi bağıntı testini oluşturmuş ve formlar bir yıl arayla uygulanmıştır. Çalışmanın sonucunda, Tucker standart hatalar yönteminin hem normallik varsayımının karşılandığı hem de karşılanmadığı durumlar için uygun olduğu belirtilmiştir.

Livingston (1993), log-lineer düzleştirme (log-linear smoothing) sonrası yüzdelik dilimler ile zincir test eşitleme kullanılan küçük örneklemlerde ne düzeyde doğru tahmin yapılabileceğini araştırmak amacı ile bir çalışma yürütmüştür. Çalışmasında desen olarak EOGB deseni ve ölçüt eşitleme olarak yüzdelik dilimler ile test eşitleme yöntemini kullanmıştır. Ayrıca 24'ü bağıntı testini oluşturan 58'er maddelik iki form kullanılmıştır. 93283 kişilik bir evrenden elde edilen veriler referans olarak kullanılmış, çalışma ise 25, 50, 100 ve 200 kişilik 4 farklı grup üzerinde yapılmıştır. Çalışmanın sonucunda ise, düzleştirmenin gerek olmadığı örneklem büyüklüğünün neredeyse yarısı kadar küçük örneklemlerde bile log-lineer düzleştirmenin daha doğru tahminler yapılmasını sağladığı gözlenmiştir.

Hanson ve diğerleri (1994), yüzdelik dilimler ile test eşitleme yönteminde kullanılan üç farklı ön düzgünleştirme yöntemini ve üç farklı test eşitleme yöntemini karşılaştırmak amacı ile bir çalışma yürütmüşlerdir. Kullanılan bu üç test eşitleme yönteminin birincisi son düzgünleştirme ile yüzdelik dilimler ile test eşitleme, ikicisi doğrusal test eşitleme ve üçüncüsü ise düzgünleştirilmemiş yüzdelik dilimler ile test eşitleme yöntemleridir. Çalışmada kullanılan veriler, beş farklı test uygulamasına ait evrenlerden 100, 250, 500, 1000 ve 3000'er kişilik örneklemler çekerek oluşturulmuştur. Beş test uygulamasının tamamı EOGB deseni ile yapılmış ve evrenlerden her örneklem büyüklüğü düzeyi için 500 farklı örnek seçilerek çalışma yürütülmüştür. Test eşitleme hatalarının kestiriminde Monte Carlo yöntemini kullanmışlardır. Çalışmanın sonucunda hem ön düzgünleştirilmiş hem de son düzgünleştirilmiş puanlar ile yapılan yüzdelik dilimler ile test eşitleme yöntemlerinin daha az test eşitleme hatası verdiğini ortaya koymuşlardır.

Parshall ve diğerleri (1995) ise, 15, 25, 50 ve 100'er kişilik örneklemlerden oluşan verileri kullanarak EOGB deseninde doğrusal test eşitleme yöntemlerini ESH ve yanlılık açısından karşılaştırmıştır. Veriler beş farklı öğretmen alan bilgisi testinden elde edilen verileri kullanarak 1000 tekrarlı şekilde simüle edilmiştir. Karşılaştırma yapabilmek için ölçüt olarak 500 örneklemli bir veri seti 1000 tekrarlı şekilde kullanılmıştır. Çalışmanın sonucunda örneklem büyüklüğü azaldıkça hem ESH'nin hem de yanlılığın arttığını ortaya koymuşlardır.

Skaggs (2005), eşdeğer gruplar deseni ile 25, 50, 75, 100, 150 ve 200'er kişilik örneklemler ile yaptığı simulatif çalışmada, iki farklı durum oluşturmuştur. Birincide güçlük açısından benzer maddelerden oluşan testler, ikincisinde ise güçlük açısından farklı maddelerden oluşan testler simüle edilmiştir. Sonucunda, ortalama eşitlemenin ortalamanın altındaki puanları eşitlemede en iyi sonucu verdiğini, özdeş test eşitlemenin ise 50 ve altı örneklem büyüklüklerinde kullanılabileceğini ortaya koymuştur. Ayrıca ortalamanın üzerindeki puanlarda, ön düzgünleştirilmiş yüzdelik dilimler ile test eşitleme yönteminin en iyi sonucu verdiği belirtilmiştir.

Mao ve diğerleri (2006), KE yönteminden elde edilen sonuçların geleneksel gözlenen puan eşitleme yöntemlerine (Tucker doğrusal gözlenen puan eşitleme yöntemi ve yüzdelik dilimler ile test eşitleme yöntemi) ne kadar benzer sonuçlar verdiğini araştırmak amacı ile bir çalışma yürütmüşlerdir. Çalışmada; birincisi EG deseni, ikincisi iç bağıntı testi ile eşdeğer olmayan gruplar deseni ve üçüncüsü dış bağıntı testi ile eşdeğer olmayan gruplar deseni olmak üzere üç farklı test eşitleme deseni kullanılmıştır. Çalışmada, Praxis (Professional Assessments for Beginning Teachers) testinden edinilen gerçek veriler kullanılmıştır. Çalışmanın verileri GENASYS ve KE Software bilgisayar programları ile analiz edilmiştir. Çalışmanın sonucunda, her yöntemden elde edilen hatalar doğrultusunda bir karşılaştırma yapılmıştır. Buna göre; KE'nin EG deseninde diğer geleneksel yöntemlere benzer sonuçlar verdiği gözlenmiştir. EOGB desenlerinde ise; düşük puan ranjı hariç, diğer puan ranjlarında KE yöntemi ile yüzdelik dilimler ile test eşitleme yöntemi benzer sonuçlar vermiştir. Yine EOGB desenlerinde Tucker doğrusal gözlenen puan eşitleme yöntemi ve KE yöntemi çok benzer sonuçlar vermiştir.

Sinharay ve Holland (2006), bağıntı testinin kapsamının ve madde güçlük parametresinin önemini ortaya koymak amacı ile simülatif bir çalışma yürütmüşlerdir. Yapılan çalışmada, 1000 kişi ve 40 maddeden oluşan bir test uygulaması simüle edilmiş

ve 5 farklı dış bağıntı testinin karşılaştırıldığı 100 farklı veri seti üretilmiştir. Bu 5 bağıntı testi; (1) kapsamı uyuşmayan bir minitest, (2) kapsamı uyuşmayan bir miditest, (3) kapsamı uyuşan bir minitest, (4) kapsamı uyuşan bir miditest ve (5) kapsamı uyuşan bir yarı-miditest olarak tanımlanmıştır. Yürütülen analizler sonucunda, kapsamı uyuşan miditest ve yarı-miditest bağıntı testlerinin ana form ile daha yüksek korelasyon gösterdiği bulunmuştur.

Kim ve diğerleri (2008), gerçek veriler ile yaptıkları çalışmada 10, 25, 50 ve 200'er kişilik örneklemlerle EOGB deseni ile üç farklı test eşitleme yöntemini (sentetik, özdeş ve zincirleme doğrusal test eşitleme) karşılaştırmışlardır. Örneklemleri 2000 ve 2001 yıllarında uygulanmış büyük ölçekli bir teste katılan 10634 ve 11321'er bireylik evrenlerden seçmişler ve her bir örneklem büyüklüğü için 1000 farklı örneklem seçmişlerdir. Yine aynı çalışma kapsamında 2004 yılında uygulanan başka bir teste ait iki formun eşitleme çalışmasını yürütmüşlerdir. Bu ikinci çalışmada ise 6019 ve 6386 kişilik iki evrenden ilk çalışmayla aynı büyüklüklerde örneklemler seçilmiştir. Çalışmaların sonucunda, 50 ve daha küçük örneklemlerde ve test güçlüklerinin benzeştiği durumlarda özdeş test eşitleme yönteminin daha az eşitleme hatası verdiğini ortaya koymuşlardır. Bu çalışma, literatüre sentetik test eşitleme yöntemini kazandırması açısından önem arz etmektedir.

Livingston ve Kim (2009) ise dört seçenekten çoktan seçmeli 107 maddeden oluşan bir teste ait iki formdan elde edilen verileri yeniden örnekleme yöntemi ile simüle etmişlerdir. EOGB deseni ve 25 ve 75'er kişilik örneklemler kullanarak yaptıkları çalışmada; Tucker doğrusal test eşitleme, Levine doğrusal test eşitleme ve zincirleme doğrusal test eşitleme yöntemlerini daire-yay test eşitleme yöntemi ile karşılaştırmışlardır. Çalışmada eşdeğer gruplar deseni kullanılmıştır. Çalışmanın sonucunda ise; daire-yay test eşitleme yönteminin özellikle 10 yüzdelik dilimin altında ve 90 yüzdelik dilimin üzerinde kalan puanları eşitlerken ortalamalar ile test eşitleme yöntemine göre çok daha az eşitleme hatası verdiğini, ayrıca bu iki yöntemin küçük örneklemlerde diğer yöntemlerden daha elverişli olduğunu ortaya koymuşlardır. Bu çalışma, literatüre daire-yay test eşitleme yöntemini kazandırması açısından önem arz etmektedir.

Babcock ve diğerleri (2012), EOGB deseni ile 20, 50 ve 80 kişilik örneklem durumlarını 3PL MTK modelini temele alarak simülatif bir çalışma yürütmüşlerdir.

Çalışma kapsamında 100 maddelik bir teste ait formlar simüle edilmiştir. Ayrıca; log-lineer ön düzgünleştirilmiş yüzdelik dilimler ile zincirleme test eşitleme, Tucker doğrusal test eşitleme, ortalamalar ile test eşitleme, düşük ağırlıklar test eşitleme, özdeş test eşitleme kullanılarak yapılan sentetik test eşitleme, .25 ve .75 ağırlıklandırmalar ile Tucker fonksiyonu ve daire-yay test eşitleme olmak üzere toplam yedi farklı test eşitleme yöntemi kullanmışlardır. Veri simülasyonunu üç farklı güçlük düzeyinde gerçekleştirmişlerdir. Çalışmanın sonucunda ise, düşük ağırlıklar test eşitleme yönteminin küçük örneklemler için önerilebileceğini ortaya koymuşlardır. Bu çalışma, literatüre düşük ağırlıklar test eşitleme yöntemini kazandırması açısından önem arz etmektedir.

Albano (2015), genel doğrusal test eşitleme yöntemini iki farklı çalışma üzerinden sunmaktadır. Araştırmanın birinci çalışması, DG deseni ile 2089 ortaokul öğrencisi üzerinden iki farklı akıcı okuma testi formlarını eşitlemek amacı ile yürütülmüştür. Bu testler, birer okuma metninden oluşmaktadır ve bir öğrencinin bir dakika içerisinde doğru bir şekilde okuduğu kelime sayısı o öğrenicinin testten aldığı puanı vermektedir. Çalışma 30, 50, 100 ve 300 öğrenciden oluşan alt gruplar üzerinden yürütülmüş ve yüzdelik dilimler ile test eşitleme yöntemi referans yöntem olarak kullanılmıştır. Birinci çalışmanın sonucunda, genel doğrusal test eşitleme yönteminin küçük örneklemlerde düşük eşitleme hatası verdiği gözlenmiştir. Diğer doğrusal ve yüzdelik dilimler ile test eşitleme yöntemlerinden elde edilen eşitleme hatalarının ancak 300 kişilik örneklem büyüklüğünde genel doğrusal test eşitleme yöntemine yakın olduğu rapor edilmiştir. Araştırmanın ikinci çalışması ise, bağıntı testi ile eşdeğer gruplar deseni 5000 kişi üzerinden kullanılmış ve 200 maddelik iki test eşitlenmeye çalışılmıştır. Bu çalışmada farklı örneklem büyüklüğü olan iki formun eşitlenmesi amaçlanmıştır, bu nedenle toplanan tüm verinin içerisinden, X formu için 20, 50, 100 ve 500 kişilik alt gruplar oluşturulmuş, Y formu ise 500 kişilik tek bir alt grup üzerinden çalışmaya dahil edilmiştir. İkinci çalışmanın sonucunda ise, genel doğrusal test eşitleme yönteminin örneklem büyüklüğü farklı olan uygulamalarda diğer doğrusal test eşitleme yöntemlerine benzer sonuçlar verdiği rapor edilmiştir. Çalışmalarda, R programlama dili ve *equate* paketi kullanılarak istatiksel analizler yapılmıştır. Çalışma, genel doğrusal test eşitleme yöntemini literatüre kazandırması açısından önem arz etmektedir.

Türkiye dışında çok kategorili testlerin ilişkilendirilmesi üzerine de birçok çalışma yapılmıştır. Baker (1992), tek grup deseni kullanarak simülatif bir çalışma yürütmüştür.

300 kişilik, 30 maddelik ve 4 kategorili simülatif verileri Graded Response Model (GRM) altında GENIRV programı ile üretmiştir. Üretilen bu veriler MULTILOG bilgisayar programı aracılığı ile analiz edilmiştir. Çalışmada yatay ve yatay olmayan test eşitleme süreçleri iki farklı yöntemle karşılaştırmıştır. Birinci yöntem her iki formdan elde edilen gerçek puanları eşleştirerek yapılırken, ikinci yöntem formlara ait test karakteristik eğrilerini eşleştirerek yapılmıştır. Çalışmanın sonucunda, her iki yöntemin de benzer sonuçlar verdiğini belirtmiştir.

Li ve diğerleri (1999) ise, karma maddelerden oluşan testlerin ilişkilendirme çalışmasını yürütmüştür. Çalışmasında iki kategorili puanlanan maddelerin yanı sıra, 5 maddelik ve 4 kategorili puanlanan maddelerden de oluşan bir test verisi simüle etmiş ve bunları yatay ve dikey eşitleme koşullarında karşılaştırmıştır. Çalışmasında test karakteristik eğrisi eşleştirme yöntemini kullanmıştır. İki kategorili puanlanan maddelerin simülasyonunda 3PL MTK modelinden faydalanılırken, çok kategorili puanlanan maddelerin simülasyonunda ise genelleştirilmiş kısmi kredi modelinden (GPCM) faydalanmışlardır. Çalışmanın sonucunda, çalışma kapsamında kullanılan yöntemin hem karma testlerde hem iki kategorili puanlanan testlerde hem de çok kategorili puanlanan testlerde doğru ilişkilendirme yapabildiğini belirtmişlerdir.

Lee ve diğerleri (2001), EG deseni kullanılarak elde edilen gerçek veriler üzerinden yaptıkları çalışmada; iki kategorili maddelerden oluşan testlerin MTK temelli test eşitleme yöntemleri ile eşitlenmesinde kullanılan 3PL model ile çok kategorili puanlanan testlerin eşitlenmesinde kullanılan dört farklı modeli (GRM, GPCM, *Nominal Model, Rating Scale Model*) karşılaştırmışlardır. Bu karşılaştırmayı (1) testletlerden elde edilen toplam puanlar üzerinden test eşitleme yaparak ve (2) madde bazında test eşitleme yaparak yürütmüşlerdir. Kullandıkları verileri, 1995 yılında 8. sınıf öğrencilerine uygulanan üç farklı testten (okuduğunu anlama, harita bilgisi ve matematik problem çözme becerisi) elde etmişlerdir. Okuduğunu anlama testi 49 maddeden oluşmaktadır ve 663 kişilik bir örnekleme uygulanmıştır. Harita bilgisi testi verileri, 33 madde ve 632 kişilik örneklemden oluşmaktadır. Matematik problem çözme becerisi testi verileri ise 537 kişilik örneklemden ve 36 maddeden oluşmaktadır. Bu çalışmada kriter referans olarak üç farklı KTK temelli test eşitleme yöntemini (ortalama, doğrusal ve yüzdelik dilimler) kullanmışlardır. Çalışmada kullanılan verilerin testletler üzerinden eşitleme yaparken tek boyutluluk varsayımını sağlamadığını belirtmişlerdir. Ancak madde bazında

yapılan test eşitlemelerin KTK temelli test eşitleme yöntemlerine benzer sonuçlar verdiğini belirtmişlerdir.

Tate (1999), birbirini izleyen yıllarda uygulanan testler arasında bir ilişkilendirme çalışması yapmak gerekliliğini vurgulamak amacıyla simülatif veriler ile iki ve çok kategorili maddelerden oluşan iki testin ilişkilendirme çalışmasını MTK temelli yöntemlerle yürütmüştür. Çalışmada kullanılan veriler, 2000 kişilik bir örneklem üzerinden, GRM modeli ile ve RESGEN programı aracılığıyla üretilmiş, kalibrasyon çalışması PARSCALE programı ile yapılmıştır. Çalışmanın sonucunda, farklı yıllarda uygulanan bu tür testler arasında güçlük ve ayırt edicilik adına farklılıkların olacağını belirtmiştir. Bu nedenle bir ilişkilendirme çalışmasının gerekliliğini vurgulamıştır.

Türkiye dışında küçük örneklemlerle yapılan test eşitleme çalışmaları incelendiğinde, genellikle EOGB deseninin kullanıldığı görülmektedir. Ayrıca, bunlar genellikle iki kategorili puanlanan maddelerden oluşan testlerin eşitlenmesi üzerine yürütülmüştür. Bu tür çalışmalarda KTK temelli yöntemlerin tercih edildiği dikkat çekmektedir. Bu çalışmalarda test eşitleme yöntemlerinin yanı sıra, düzgünleştirme yöntemleri, farklı bağıntı testi uzunlukları, içerikleri ve ortalama güçlük düzeyleri, farklı test uzunlukları ve içerikleri gibi koşullar da incelenmiştir. Bu çalışmalarda küçük örneklem olarak, 10 ile 500 örneklem büyüklüklerine kadar farklı örneklem boyutlarının incelendiği görülmektedir. EOGB deseni ile kullanılabilen KTK temelli test eşitleme yöntemleri arasından küçük örneklemler ile en iyi sonuçları Tucker doğrusal, düşük ağırlıklar doğrusal, daire-yay yöntemlerinin verdiği görülmektedir. Çok kategorili puanlanan maddelerden oluşan testlerin eşitleme çalışmalarında ise MTK temelli test eşitleme yöntemlerinin tercih edildiği görülmektedir. Bu çalışmalarda, PARSCALE, RESGEN, GENIRV, MULTILOG, GENAYSIS ve R programlama dili gibi yazılımlardan destek alındığı görülmektedir.

Türkiye'de Yapılan Çalışmalar

Türkiye'de de test eşitleme alanında birçok çalışma yapılmıştır (Kelecioğlu, 1994; Bozdağ ve Kan, 2010; Kan, 2010; Kilmen, 2010; Kan, 2011; Kahraman, 2012; Kelecioğlu ve Öztürk-Gübeş, 2013; Aşiret ve Sünbül, 2014; İnal ve Arıkan 2017).

Kelecioğlu (1994), Öğrenci Seçme Sınavı (ÖSS) sözel (Türkçe ve sosyal bilimler) ve sayısal (matematik ve fen bilimleri) testlerini yüzdelik dilimler, doğrusal, 1PL model

ve 2PL model yöntemleri ile eşitlemek amacıyla bir çalışma yürütmüştür. Çalışmada, EOGB deseni kullanılmıştır. Bu desende kullanılan bağıntı testi; aynı yapıyı ölçtüğü faktör analizi yöntemi ile kanıtlanmış olan farklı maddelerden oluşan bir iç bağıntı testidir. 1990, 1991 ve 1992 yıllarında uygulanan ÖSS testlerine katılan öğrenciler arasından seçkisiz olarak 2000'er kişilik örneklemler oluşturulmuştur. Her bir alt test belirtilen 4 yöntemle eşitlenmiş ve en uygun yöntemin; Türkçe alt testi için doğrusal, sosyal bilimler ve matematik alt testleri için 1PL ve fen bilimleri alt testi için ise yüzdelik dilimler ile test eşitleme yöntemi olduğu sonucuna varılmıştır.

Bozdağ ve Kan (2010), şans başarısının test eşitlemeye etkisinin olup olmadığını araştırmak üzere bir çalışma yürütmüşlerdir. Çalışma TG deseni ile doğrusal ve yüzdelik dilimler ile test eşitleme yöntemleri kullanılarak yürütülmüştür. Çalışmada, 2003 ve 2005 yıllarında uygulanan OKS Türkçe alt testleri, 1031 ilköğretim sekizinci sınıf öğrencisine uygulanarak veri elde edilmiştir. Araştırmanın sonucunda şanstan arındırılmış puanların eşitlendiği yüzdelik dilimler ile test eşitleme yönteminin en az test eşitleme hatasını verdiği görülmüştür.

Kan (2010), aynı yapıyı ölçen farklı formların eşitliğini sınayarak, farklı formların öğrenci performansına etkisini araştırmak amacı ile bir çalışma yürütmüştür. Çalışmasında TG deseni ile doğrusal test eşitleme yöntemini kullanılmıştır. Çalışma, farklı ilköğretim kurumlarında eğitim gören 420 adet altıncı sınıf öğrencisi üzerinde yürütülmüştür. Çalışmada aynı yapıyı ölçen sözel köklü ve sayısal köklü matematik testleri kullanılmıştır. Sonucunda ise, bu iki formun istatiksel olarak eşitlenemeyeceği ve bu nedenle birbirlerinin yerine kullanılamayacağı sonucuna varılmıştır.

Kilmen (2010), MTK temelli 4 test eşitleme yönteminin yetenek ve örneklem büyüklüğü değişkenleri açısından karşılaştırmak amacı ile bir çalışma yürütmüştür. Simülatif veriler üzerinden yapılan çalışma, EOGB deseni ile kurgulanmış ve veriler WinGen2 programı ile üretilmiştir. Test eşitleme çalışması ise IRTEQ programı aracılığıyla yürütülmüştür. Çalışmanın sonucunda, örneklem büyüdükçe daha az OHKK üretildiği, benzer yetenek düzeyindeki grupların ise farklı yetenek düzeyindeki gruplara nazaran daha az OHKK verdiği gözlenmiştir.

Kan (2011) tarafından yapılan araştırmada 2003 ve 2005 yılları arasında uygulanan OKS (Orta Öğretim Kurumları Sınavı) Türkçe alt testlerini eşitlemek amaçlanmıştır. Bu iki alt test birleştirilerek 50 maddelik bir test oluşturulmuş ve 1030 ilköğretim 8. sınıf

öğrencisine tek grup deseni ile uygulanmıştır. Elde edilen veriler, doğrusal test eşitleme yöntemi ile Statistica 7.0 paket programı kullanılarak analiz edilmiştir. Çalışmanın sonucunda, OKS 2003 ve OKS 2005 Türkçe alt testlerinin paralel olmadığı; bu nedenle bu iki form arasında test yansızlığından bahsedilemeyeceği belirtilmiştir.

Kahraman (2012), yüzdelik dilimler ile test eşitleme yönteminde kullanılan iki farklı düzleştirme yönteminin (loglineer pre-smoothing ve cubic spline post-smoothing düzleştirme yöntemleri) karşılaştırmasını yapmak amacı ile bir çalışma yürütmüştür. Çalışmada kullanılan veriler, 2010 ve 2011 yıllarında Hacettepe Üniversitesi'nde kullanılan iki farklı İngilizce yeterlik testlerini alan toplam 810 öğrenciden elde edilmiştir. Test formları; dinleme, temel bileşenler ve okuma becerisi olmak üzere 3 alt testten oluşmaktadır. Çalışmanın sonucunda, loglinear pre-smoothing yönteminin diğer düzleştirme yöntemine göre daha az hata verdiği gözlenmiştir. İlgili çalışma düzleştirme yöntemlerinin karşılaştırıldığı bir çalışma olması açısından önem arz etmektedir.

Kelecioğlu ve Öztürk-Gübeş (2013), 2009 yılında yapılan Öğrenci Başarılarının Belirlenmesi Sınavı (ÖBBS)'nın sosyal bilimler alt testinin iki farklı kitapçığından elde edilen puanların eşitleme çalışmasını yürütmüşlerdir. Çalışmada, 16670 ve 15743 öğrencilik iki gruptan elde edilen veriler ile EG deseni kullanılmıştır. Grupların eşdeğerliğinin sarmallama yöntemi ile sağlandığı belirtilmiştir. Formlardan elde edilen puanlar 4 farklı test eşitleme yöntemi kullanılarak eşitlenmiştir. Bunlar; doğrusal test eşitleme yöntemi ve yüzdelik dilimler ile test eşitleme yönteminin (1) düzleştirilmemiş puanlar ile, (2) pre-smoothing ile düzleştirilmiş puanlar ile ve (3) post-smoothing ile düzleştirilmiş puanlar ile uygulanması olarak tanımlanmıştır. Çalışmanın sonucunda, ham puanlar üzerinden yapılan eşitlemelerde verilerin doğrusal dağılımı ile alakalı olarak doğrusal test eşitlemenin daha az ESH verdiği; ancak düzleştirme sürecinden sonra en az ESH'nin post-smoothing yöntemi ile yapılan yüzdelik dilimler ile test eşitleme yöntemi tarafından üretildiği vurgulanmıştır. Çalışma merkezi bir testin eşitleme çalışması olması ve düzleştirme yöntemlerinin karşılaştırılması açısından önem arz etmektedir.

Aşiret ve Sünbül (2014), küçük örneklemlerde test eşitleme yöntemlerini karşılaştırmak amacıyla simülatif bir çalışma yürütmüşlerdir. Çalışmada kullanılan veriler, 3PL MTK modeli temelinde 10, 25, 50, 75, 100, 150 ve 200'er bireyden oluşan yedi farklı örneklem büyüklüğü koşulu altında 4 farklı güçlük düzeyinde ve seçkisiz grup deseninde oluşturulmuştur. Test eşitleme süreci, R 3.1 paket programı kullanılarak 5

farklı test eşitleme yöntemi (özdeş, ortalamalar, doğrusal, daire-yay ve ön düzgünleştirilmiş yüzdelik dilimler ile test eşitleme yöntemleri) ile yapılmıştır. Çalışmanın sonucunda, 50 ve üzeri örneklem büyüklüklerinde test eşitleme yapmanın manidar bir farklılık oluşturduğunu ortaya koymuşlardır. Ayrıca, daire-yay ve ortalamalar ile test eşitleme yöntemlerinin en az test eşitleme hatası verdiğini belirtmişlerdir.

İnal ve Arıkan (2017) Tucker ve Levine doğrusal gözlenen puan eşitleme yöntemlerini grup değişmezliği varsayımı açısından karşılaştırmak amacı ile bir çalışma yürütmüşlerdir. Çalışmada, 2012 yılında uygulanan PISA (Programme for International Student Assessment) testi Matematik alt testine ait iki kitapçığın eşitlemesi yapılmıştır. Kitapçıkların uygulandığı toplam 741 öğrenciden veri elde edilmiş ve EG deseni kullanılmıştır. Bu formlardan elde edilen veriler kullanılarak iki form hem tüm grup ile hem de cinsiyet değişkenine göre eşitlenmiştir. İstatistiksel işlemler Microsoft Excel programı ile yürütülmüştür. Çalışmanın sonucunda, tüm gruptan elde edilen veriler ile yapılan test eşitleme çalışmasının Tucker doğrusal gözlenen puan eşitleme yöntemi ile daha az eşitleme hatası verdiği gözlenmiştir. Ayrıca cinsiyet değişkenine göre oluşturulan gruplarda Levine doğrusal gözlenen puan eşitleme yönteminin Tucker doğrusal gözlenen puan eşitleme yöntemine göre daha fazla eşitleme hatası verdiği, bu hatanın da manidar bir farklılık barındırdığı gözlenmiştir. Bu durumda Tucker doğrusal gözlenen puan eşitleme yönteminin grup değişmezliği varsayımı açısından daha iyi sonuçlar verdiği belirtilmiştir.

Türkiye'de yapılan çalışmalar incelendiğinde, çeşitli örneklem büyüklüklerinde çalışmaların yürütüldüğü görülmektedir. Ancak, küçük örneklemlerde KTK temelli test eşitleme yöntemleri tercih edilmiştir. Yaygın bir şekilde, test eşitleme yöntemleri ve düzleştirme yöntemleri karşılaştırılmıştır. Bunun yanı sıra, farklı örneklem büyüklüğü koşulları, grup değişmezliği varsayımının ihlali, merkezi olarak uygulanan testlerin eşitlenmesi, farklı yetenek düzeyi koşulları gibi konular da ele alınmıştır. Farklı çalışmalarda incelenen küçük örneklem büyüklüklerinde; Tucker doğrusal, daire-yay, ortalamalar ile, yüzdelik dilimler ile test eşitleme yöntemlerinin daha iyi sonuçlar verdiği vurgulanmaktadır. Ayrıca, merkezi olarak uygulanan bazı testlerin eşitlenemediği de belirtilmektedir. Bu çalışmalarda, IRTEQ, Microsoft Excel ve R programlama dili gibi yazılımların kullanıldığı görülmektedir.

BÖLÜM 3

YÖNTEM

Bu bölümde, araştırmanın modeli, simülasyon koşulları, verilerin üretilmesi ve analizi ile ilgili açıklamalara yer verilmektedir.

Araştırmanın Modeli

Fraenkel, Wallen ve Hyun (2015), bir kuramın temelinde yatan süreçleri açıklamak amacı ile yapılan araştırmaları *temel araştırmalar (basic research)* olarak tanımlamaktadır. Buna göre temel araştırmalarda, araştırmacı bir kuramın içerisindeki bir ya da birkaç süreci düzenlemek, geliştirmek ve bu kurama katkı yapmak amacı gütmektedir. Karasar'a (2009) göre ise literatürde tanımlanmış yöntemlerin sınandığı çalışmalar temel araştırmalardır. Bu çalışmada ise, küçük örneklemlerde çok kategorili puanlanan test uygulamalarında kullanılabilecek en uygun test eşitleme yöntemini sunmak amaçlanmaktadır. Bu bağlamda, farklı test eşitleme yöntemleri karşılaştırılacaktır. Yukarıda belirtilen tanımlamalar doğrultusunda bu çalışmanın bir temel araştırma niteliğinde olduğu belirtilebilir. Bunun yanı sıra, bu çalışmada kullanılacak veriler bilgisayar ortamında simülatif olarak üretilecektir. Cohen, Manion ve Morrison (2007), bu türden bilgisayar tabanlı simülasyon çalışmalarını ikiye ayırmaktadır. Buna göre; simülasyon çalışması kapsamında ele alınan tüm değişkenlerin önceden belirlendiği çalışmalar *belirleyici simülasyon çalışmaları (deterministic simulations),* değişkenlerden en az birinin seçkisiz olarak atandığı simülasyon çalışmalarına ise *olasılıksal simülasyon çalışmaları (stochastic simulations)* olarak adlandırılmaktadır. Bu bağlamda incelendiğinde bu çalışma bir olasılıksal simülasyon çalışmasıdır.

Simülasyon Koşulları

Bu çalışma kapsamında yürütülen simülasyon çalışmasında bazı koşullar sabit tutulurken bazıları değişimlenmiştir. Bu bölümde, bu koşullar (1) değişimlenen koşullar ve (2) sabit tutulan koşullar olarak iki başlık altında ele alınmıştır.

Değişimlenen koşullar

Bu çalışmada küçük örneklemlerde çok kategorili puanlanan test uygulamalarında test eşitleme yöntemlerinden elde edilen ESH, yanlılık ve OHKK değerleri incelenmiştir. Çalışmada kullanılan simülasyon koşulları, daha önce yapılmış olan benzer çalışmalar dikkate alınarak oluşturulmuştur. Buna göre çalışma kapsamında değişimlenen 4 koşul şunlardır:

Örneklem büyüklüğü. Literatürde daha önce küçük örneklem koşullarında birçok test eşitleme çalışması yapılmıştır (Kolen, 1985; Livingston, 1993; Parshall ve diğerleri, 1995; Skaggs, 2005; Kim ve diğerleri, 2008; Livingston ve Kim, 2009; Babcock ve diğerleri, 2012; Albano, 2015). Bu çalışmalarda en az 10, en fazla ise 250 örneklem büyüklükleri kullanılmıştır. Bu çalışmada ise farklı sınıf içi ve okul içi öğrenci mevcutları simüle edileceğinden 15, 20, 30, 60, 100 ve 150 kişiden oluşan örneklem büyüklüklerinin test eşitleme sürecine etkisi incelenmiştir. 15, 20 ve 30 kişilik örneklem büyüklükleri sınıf içi; 60, 100 ve 150 kişilik örneklem büyüklükleri ise sınıflar arası ya da okul içi uygulamaları temsil etmek amacıyla tercih edilmiştir. Bu bağlamda tanımlanan örneklem büyüklüğü koşulu altı düzeyden oluşmaktadır.

Test formaları arasındaki eşik parametresi farkı. İlgili literatür incelendiğinde test formlarına ait maddelerin güçlük parametresi (b parametresi) ortalamaları arasında farkların olduğu görülmektedir. Bu durum hem gerçek hem de simülatif veriler ile yapılan çalışmalarda gözlenmektedir. Klein ve Jarjoura (1985) yaptıkları çalışmada güçlük parametresi ortalamaları farkının 0.76 ile 0.78 arasında değiştiğini belirtmişlerdir. Sinharay ve Holland (2006) ise yaptıkları simülatif çalışmada güçlük parametresi ortalamaları arasındaki farkı -0.2, 0, 0.1, 0.2, 0.5 olarak beş koşulda incelemişlerdir. Aşiret (2014) ise simülatif çalışmasında madde güçlük düzeyleri arasındaki ortalama farkı 0.1, 0.4 ve 0.7 olarak üç koşul altında incelemiştir.

Literatürde güçlük parametreleri ortalama farklarının -0.2 ile 0.8 arasında değişiklik gösterdiği görülmektedir. Bu çalışmada üretilen veriler genelleştirilmiş kısmi kredi modelinden faydalanılarak üretilmiştir. Bu modelde, 1PL, 2PL ve 3PL modellerden farklı olarak b parametresi yerine eşik parametresi kullanılır. Bu nedenle bu çalışmada simüle edilen iki formun eşik parametreleri arasındaki ortalama farklar 0.1, 0.4 ve 0.7 olarak üç koşul altında incelenmiştir.

Madde sayısına bağlı puanlama biçimi. İlgili literatür incelendiğinde, çok kategorili puanlanan testlerle ilgili çalışmaların az sayıda olduğu görülmektedir. Lee ve diğerleri (2001) yaptıkları çok kategorili testlerin eşitleme çalışmasında en fazla 10 kategorili puanlanan testler kullanmışlardır. Darrell Bock (1972) çalışmalarında kullandıkları çok kategorili puanlanan testlerin 3, 4 ve 5 kategorili olduğunu belirtmiştir. Her ne kadar çok kategorili puanlanan testlerle ilgili az sayıda çalışma varsa da bu testlerle eğitim durumlarının birçok alanında karşılaşılmaktadır. Özellikle ilk ve orta öğretim kademelerinde maksimum puanı 100, minimum puanı 0 olan ve çok kategorili puanlanan testler sıkça kullanılmaktadır. Bu nedenle bu çalışmada maksimum puanı 100 minimum puanı 0 olan üç koşul oluşturulmuştur. Bu üç koşul; (1) 0, 1 ve 2 şeklinde 3 kategorili puanlanan 50 maddelik test, (2) 0 ile 5 arasında 6 kategorili puanlanan 20 maddelik test ve (3) 0 ile 10 arasında 11 kategorili puanlanan 10 maddelik test koşulları şeklindedir.

Test eşitleme yöntemi. Küçük örneklemle de test eşitleme çalışmaları ile ilgili literatür incelendiğinde, çok sayıda çalışmaya rastlanmaktadır. Kolen (1985), yürüttüğü simülatif çalışmada, 100 ve 250 örneklem büyüklüklerinde Tucker doğrusal test eşitleme yöntemini simetrik ve simetrik olmayan test formları açısından karşılaştırmıştır. Kolen ve Brennan (2004), ortalamalar ile test eşitleme yöntemini küçük örneklemler için önermektedir. Benzer bir şekilde Skaggs (2005), çalışmasında ortalamalar ile test eşitlemenin küçük örneklemler için uygun olduğunu belirtmiştir. Kim ve diğerleri (2008), sentetik, özdeş ve Braun-Holland test eşitleme yöntemlerinin küçük örneklemlerde ürettikleri hatayı incelemek amacıyla gerçek verileri kullanarak bir çalışma yürütmüşlerdir. Livingston ve Kim (2009) ise çalışmalarında daire-yay test eşitleme yönteminin küçük örneklemler için uygun olduğunu belirtmişlerdir. Babcock ve diğerleri (2012) ise yaptıkları simülatif çalışmada düşük ağırlıklar test eşitleme yöntemini küçük

örneklemler için önermişlerdir. Lee ve diğerleri (2001) ise çok kategorili puanlanan testleri eşitlerken MTK temelli test eşitleme yöntemlerini kullanmışlardır ancak bunları karşılaştırmak amacıyla kriter eşitleme yöntemlerini ortalamalar, doğrusal ve yüzdelik dilimler ile test eşitleme yöntemleri olarak belirlemişlerdir. Kim ve diğerleri (2008) ise, özdeş test eşitlemenin küçük örneklemler için uygun bir yöntem olduğunu belirtmişlerdir. Ancak bu yöntem maksimum ve minimum puanları aynı olmayan testlerin eşitlenmesinde önerilmektedir (Kolen ve Brennan, 2004). Bunun yanı sıra, yüzdelik dilimlerdeki puanlar düzleştirme yöntemleri ile eşleştirilemeyebileceğinden, yüzdelik dilimler ile yapılan (zincirleme ve frekans tahmin) test eşitleme yöntemleri küçük örneklemler için önerilmemektedir (Kolen ve Brennan, 2004). Bu nedenle, bu üç test eşitleme yöntemi bu çalışmanın kapsamına dahil edilmemiştir.

İlgili literatürdeki çalışmalar göz önünde bulundurulduğunda, küçük örneklemler ile kullanılabilecek çeşitli test eşitleme yöntemlerinden dört tanesi (ortalamalar, Tucker doğrusal, düşük ağırlıklar doğrusal ve Braun-Holland doğrusal test eşitleme yöntemleri) bu çalışma kapsamına dahil edilmiştir. Çalışma kapsamında hesaplanan ESH, yanlılık ve OHKK değerleri bootstrap yöntemi ile elde edilmiştir. Bu yöntemde bir kriter yöntem belirlenerek, bunun üzerinden hesaplamalar yapılmaktadır. Bu nedenle daire-yay test eşitleme yöntemi test eşitleme hatası hesaplamalarında yöntem farketmeksizin kriter eşitleme yöntemi olarak kullanılmıştır.

Sabit tutulan koşullar

Çalışma kapsamında sabit tutularak ele alınan 6 koşul bulunmaktadır. Bunlar;

Model. İlgili literatür incelendiğinde, çok kategorili puanlanan testlerin eşitlenmesi çalışmalarında MTK temelli birçok modelin kullanıldığı görülmektedir. Baker (1992), GRM modeli doğrultusunda çalışmasını yürütmüştür. Li ve diğerleri (1999) GPCM modeli kullanarak çalışmalarını yürütmüştür. Lee ve diğerleri (2001) ise farklı modelleri karşılaştırdıkları çalışmada GPCM ve GRM modellerinin yanı sıra Nominal Model ve Rating Scale Modelleri de kullanmışlardır. Tüm bu çalışmalar incelendiğinde GRM ve GPCM modellerinin en yaygın kullanılan iki model olduğu görülmektedir. Bu çalışmada ise, kullanılan veriler GPCM modeli altında simüle edilmiştir.

Bağıntı testi madde sayısı. Test eşitleme çalışmalarında, EOGB deseni kullanıldığı durumlarda bağıntı testinin kaç maddeden oluşacağı sorusu araştırmacıların odağında olmuştur. Ricker ve von Davier (2007), bağıntı testinin uzunluğunun test eşitlemeye olan etkisini araştırdıkları çalışmada, 30 maddelik testlerin içerisinde yer alan 6, 5 ve 4 maddelik bağıntı testlerini incelemişlerdir. Bu durumda, toplam testin sırasıyla %20, %16 ve %13 oranlarında bağıntı testlerinin çalışma kapsamında ele alındığı görülmektedir. Sinharay ve Holland (2006) ise, bağıntı testinin özelliklerinin test eşitlemeye olan etkisini araştırdıkları çalışmada kullandıkları 45 maddelik testin 20 maddesinin, 78 maddelik testin 35 maddesinin ve 60 maddelik testin ise 25 maddesinin bağıntı testlerini oluşturduğunu belirtmişlerdir. Bu çalışmada ise bağıntı testinin toplam testin yaklaşık %40'ını oluşturduğu görülmektedir. Klein ve Jarjoura (1985), çalışmalarında bağıntı testinin toplam testi temsil etme gücünü inceledikleri çalışmada 250 madde için 60 maddelik bir bağıntı testi kullanmışlardır. Bunun ise toplam testin %24'ünü oluşturduğu görülmektedir. Angoff (1984) ise bağıntı testi uzunluğunun 20 madde ya da toplam testin %20'si kadar olması gerektiğini belirtmiştir. Bu nedenle bu çalışmada bağıntı testi uzunluğu toplam testin %20'sini oluşturacak şekilde ayarlanmıştır. Bu bağlamda birinci madde sayısına bağlı puanlama biçimi (3 kategorili puanlanan 50 madde) için 10 maddelik, ikinci madde sayısına bağlı puanlama biçimi(6 kategorili puanlanan 20 madde) için 4 maddelik ve üçüncü madde sayısına bağlı puanlama biçimi (11 kategorili puanlanan 10 madde) için 2 maddelik bağıntı testleri kullanılmıştır.

Madde eşik parametreleri. Çalışmanın puanlama biçimi koşuluna göre; 3, 6 ve 11 kategorili puanlanan testler üretilecektir. Bu testlerin her bir kategori için eşik parametreleri tek tek tanımlanmıştır. Baker (1992), çalışmasında güçlük parametrelerini ortalaması 0 standart sapması 1 olan normal dağılımdan elde etmiştir. Bu çalışmada ise bu parametrelerin kendinden önceki eşikten fazla olması amacıyla aritmetik olarak artan bir şekilde üretilmiştir. Bunlar; (1) 3 kategorili puanlanan test maddelerinin birinci eşik parametresi için ortalaması -0.2 ve standart sapması 1 olan normal dağılımdan, ikinci eşik parametresi için ortalaması 0.4 ve standart sapması 1 olan normal dağılımdan elde edilecektir. (2) 6 kategorili puanlanan test maddelerinin birinci eşik parametresi için ortalaması -0.4 ve standart sapması 1 olan normal dağılımdan, ikinci eşik parametresi için ortalaması -0.2 ve standart sapması 1 olan normal dağılımdan, üçüncü eşik parametresi için ortalaması 0 ve standart sapması 1 olan normal dağılımdan, dördüncü eşik

parametresi için ortalaması 0.2 ve standart sapması 1 olan normal dağılımdan ve beşinci eşik parametresi için ortalaması 0.4 ve standart sapması 1 olan normal dağılımdan elde edilecektir. (3) 11 kategorili puanlanan test maddelerinin birinci eşik parametresi için ortalaması -1 ve standart sapması 1 olan normal dağılımdan, ikinci eşik parametresi için ortalaması -0.8 ve standart sapması 1 olan normal dağılımdan, üçüncü eşik parametresi için ortalaması -0.6 ve standart sapması 1 olan normal dağılımdan, dördüncü eşik parametresi için ortalaması -0.4 ve standart sapması 1 olan normal dağılımdan, beşinci eşik parametresi için ortalaması -0.2 ve standart sapması 1 olan normal dağılımdan, altıncı eşik parametresi için ortalaması 0 ve standart sapması 1 olan normal dağılımdan, yedinci eşik parametresi için ortalaması 0.2 ve standart sapması 1 olan normal dağılımdan, sekizinci eşik parametresi için ortalaması 0.4 ve standart sapması 1 olan normal dağılımdan, dokuzuncu eşik parametresi için ortalaması 0.6 ve standart sapması 1 olan normal dağılımdan ve onuncu eşik parametresi için ortalaması 0.8 ve standart sapması 1 olan normal dağılımdan elde edilmiştir. Tüm eşik parametreleri Tablo 4'te gösterilmektedir.

Tablo 4.

Her puanlama biçimi koşulu için madde eşik parametreleri

	Madde Sayısına Bağlı Puanlama Biçimi		
Eşik	*3 kategori-50 madde*	*6 kategori-20 madde*	*11 kategori-10 madde*
1. eşik	N (-0.2, 1)	N (-0.4, 1)	N (-1,1)
2. eşik	N (0.4, 1)	N (-0.2, 1)	N (-0.8, 1)
3. eşik		N (0, 1)	N (-0.6, 1)
4. eşik		N (0.2, 1)	N (-0.4, 1)
5. eşik		N (0.4, 1)	N (-0.2, 1)
6. eşik			N (0, 1)
7. eşik			N (0.2, 1)
8. eşik			N (0.4, 1)
9. eşik			N (0.6, 1)
10. eşik			N (0.8, 1)

Madde ayırt edicilik düzeyleri. Literatürde yer alan çalışmalarda madde ayırt edicilik düzeyleri farklılık göstermektedir. Babcock ve diğerleri (2012) çalışmalarındaki verilerin ayırt edicilik parametrelerinin -0.15 ile 0.14 aralığında log normal dağılım gösterdiğini belirtmişlerdir. Klein ve Jarjaura (1985) ise ayırt edicilik parametrelerinin 0.23 ile 0.26 arası ortalama ve 0.8 standart sapmaya sahip bir normal dağılım gösterdiğini belirtmişlerdir. Wolf (2013) ise, çalışmasında kullandığı verilerin ayırt edicilik parametrelerini 0.7 ile 1.3 aralığında normal uniform dağılımdan elde etmiştir. Muraki

(1992), ayırt edicilik parametresinin GPCM için 0'dan büyük olması gerektiğini belirtmiştir. Ayrıca Muraki (1993), 6 maddeden oluşan bir test ile yaptığı çalışmada ayırt edicilik parametresinin 0.4 ile 1 aralığında değiştiğini ifade etmiştir. Baker (1992) GRM modeli altında yürüttüğü çalışmasında ayırt edicilik parametresini 1.34 ile 2.65 aralığında uniform dağılımdan elde etmiştir. Bu çalışmada ise tüm maddelerin ayırt edicilik düzeyleri 0.8 ile 1.8 aralığında normal uniform dağılımdan (U (0.8, 1.8)) elde edilmiştir.

Kriter test eşitleme yöntemi. Livingston ve Kim (2009) daire-yay test eşitleme yönteminin küçük örneklemler için en uygun yöntem olduğunu belirtmişlerdir. Bu nedenle, 100 tekrarlı bootstrap yöntemi ile ESH, yanlılık (bias) ve OHKK hesaplamaları yapılırken, daire-yay test eşitleme yöntemi kriter olarak kullanılmıştır.

Tekrar sayısı. Yapılan literatür incelemesinde, test eşitleme ile ilgili simülasyon çalışmalarının farklılaşan tekrar sayıları ile yapıldığı görülmektedir. Livingston (1993), 50 tekrarla çalışmasını yürütmüştür. Aşiret ve Sünbül (2014), 100 tekrarla çalışmalarını sürdürmüşlerdir. Kolen (1985) ise, çalışmasında 500 tekrar yapmıştır. Bazı araştırmacılar ise çalışmalarında 1000 tekrar kullanmışlardır (Kim ve diğerleri, 2008; Livingston ve Kim, 2009 ve Albana, 2015). Tekrar sayısının 1000 olduğu çalışmalara daha çok rastlanmıştır. Genel olarak araştırmalarda tercih edilen tekrar sayılarının 50 ile 1000 arasında değiştiği görülmektedir. Bu çalışmada da uygulanabilirliği göz önünde bulundurularak, 100 tekrarın yeterli olacağı değerlendirilmiştir.

Verilerin Üretilmesi ve Analizi

Bu çalışmada verilerin üretilmesinde MTK temelinde çok kategorili puanlanan testler için oluşturulan genelleştirilmiş kısmi kredi modeli (GPCM) kullanılmıştır. Çalışma kapsamında kullanılan veriler, R 3.6.1 programı (R core team, 2019) ltm paketi (Rizopoulos, 2006) kullanılarak üretilmiştir. Çalışma kapsamında kullanılacak verilerin üretilmesi sürecinde değişimlenen koşullar Tablo 5'te, sabit tutulan koşullar ise Tablo 6'da gösterilmektedir.

Tablo 5.

Çalışma kapsamında değişimlenen koşullar

Değişimlenen Koşul	Koşul Sayısı	Koşul Değerleri
Örneklem Büyüklüğü	6	15 20 30 60 100 150
Test Formları Arasındaki Eşik Parametreleri Farkı	3	0.1 0.4 0.7
Puanlama Biçimi	3	3 kategorili puanlanan 50 madde 6 kategorili puanlanan 20 madde 11 kategorili puanlanan 10 madde
Test Eşitleme Yöntemi	4	Ortalamalar ile Tucker doğrusal Düşük ağırlıklar doğrusal Braun-Holland doğrusal

Tablo 6.

Çalışma kapsamında sabit tutulan koşullar

Sabit Tutulan Koşul	Koşul Değerleri
Veri Üretiminde Kullanılan Model	Genelleştirilmiş Kısmi Kredi Modeli
Bağıntı Testi Madde Sayısı	Her madde sayısına bağlı puanlama biçimi için toplam madde sayısının %20'si
Madde Eşik Parametreleri	Tablo 4'te gösterilmiştir
Madde Ayırt Edicilik Düzeyleri	0.8 ile 1.8 aralığında normal uniform (U (0.8, 1.8))
Kriter Test Eşitleme Yöntemi	Daire-Yay
Tekrar Sayısı	100

Değişimlenen koşullar göz önünde bulundurulduğunda; örneklem büyüklüğü koşulu (6 koşul), formlar arası güçlük parametresi farkı koşulu (3 koşul) ve madde sayısına bağlı puanlama biçimi koşulu (3 koşul) birlikte 6x3x3=54 koşul oluşturmaktadır. Bu durumda üretilen her koşul için bir adet X formu ve bir adet Y formu olmak üzere toplam 108 test formu simüle edilmiştir. Çalışma verileri 100 tekrarla üretildiğinde 100x108=10800 adet cevap örüntüsü elde edilmiştir.

Elde edilen bu örüntüler arasından her koşul için altışar adet olmak üzere toplam 324 adet seçkisiz olarak seçilerek, ilgili koşula ait parametrik özelliklere sahip olup olmadığı kontrol edilmiştir. Yapılan bu kontrollerde, herhangi bir uyumsuzluğa rastlanmamıştır. Bu nedenle simülasyonun geçerli olduğuna karar verilmiştir. Bu nedenle,

simülasyondan elde edilen veriler 4 farklı test eşitleme yöntemiyle eşitlenmiştir. Test eşitleme analiz ve işlemlerinde R 3.6.1 yazılım dili (R core team, 2019) ve R ortamında kullanılabilen "Equate" paketi (Albano, 2016) kullanılmıştır.

Her bir yöntem için üretilen ESH, yanlılık ve OHKK değerleri her koşul için tek tek hesaplanmış ve ortalamaları alınmıştır. Bu ortalamalar arasındaki farkların istatiksel olarak manidar olup olmadığı sınanmıştır. Öncelikle verilerin normallik varsayımını sağlayıp sağlamadıklarını kontrol etmek için Shapiro-Wilk's Normallik Testi ve varyansların homojenliğini sınamak için, Fligner-Killeen varyansların homojenliği testi uygulanmıştır. Bu testlerin sonuçlarına göre verilerde, normallik ve varyansların homojenliği varsayımları açısından herhangi bir manidar ihlal gözlenmemiştir. Bu testlere ait analiz sonuçları Ek 1'de verilmiştir. Buna göre dikkate alınan simülasyon koşullarında, farklı test eşitleme yöntemlerine yönelik olarak hesaplanan ESH, yanlılık ve OHKK ortalama değerleri arasındaki farkların test edilmesinde faktöryel varyans analizleri uygulanmıştır. Analizlerde hem ana etkiler hem de etkileşim etkileri test edilmiştir. Manidar olmayan etkiler çıkarılarak en sade modeller tanımlanmıştır. Manidar düzeyde olduğu tespit edilen etkilerin kaynağını ve yönünün belirlemek için Tukey'in dürüst manidar fark testi (Tukey HSD) uygulanmıştır. Bu teste ait analiz sonuçları Ek 2'de gösterilmiştir. Bunun yanı sıra ESH, yanlılık ve OHKK değerlerinin simülasyon koşullarına göre değişimi, grafiklerle görselleştirilerek de sunulmuştur. Bu grafiklerin R ortamında oluşturulmasında ve düzenlenmesinde Demir'den (2019) faydalanılmıştır.

BÖLÜM 4

BULGULAR VE YORUMLAR

Bu bölümde; örneklem büyüklüğü, test formlarının maddeleri arasındaki eşik parametre farkları ve puanlama biçimi gibi faktörlerin bazı test eşitleme yöntemlerinden elde edilen eşitlemenin standart hatası (ESH - *SEE*), yanlılık (bias) ve OHKK (ortalama hata karekökü - *RMSE*) değerlerine yaptığı etkilere yönelik analiz çıktılarına yer verilmektedir. Bu bağlamda, çalışma kapsamında ele alınan test eşitleme yöntemlerinden farklı simülasyon koşullarında elde edilen ESH, yanlılık ve OHKK değerleri Tablo 7'de gösterilmiştir. Burada yer alan değerler, simüle edilen 100 tekrarın ortalamaları alınarak elde edilmiştir.

Tablo 7 incelendiğinde, çalışma kapsamında simülatif verilerden elde edilen ortalama ESH, yanlılık ve OHKK değerlerinin madde sayısına bağlı puanlama biçimi, örneklem büyüklüğü ve eşik parametresi farkı değişkenleri açısından sayısal bazı farklılıklar gösterdiği görülmektedir. Araştırma soruları kapsamında, bu farklılıkların istatiksel olarak manidar olup olmadığı incelenmiş ve başlıklar halinde yorumlanmıştır.

Tablo 7.

Çalışma kapsamında ele alınan test eşitleme yöntemlerinden simülasyon koşullarında elde edilen ESH, yanlılık ve OHKK değerleri

A	B	C	ESH				Yanlılık				OHKK			
			O	T	DA	BH	O	T	DA	BH	O	T	DA	BH
3	15	0.1	6.71	10.27	26.70	10.51	2.89	7.55	15.35	7.25	7.54	13.06	31.31	13.11
		0.4	6.66	11.46	31.40	11.02	4.31	8.57	16.91	8.74	8.24	14.64	36.33	14.43
		0.7	6.67	12.83	21.85	13.02	6.74	10.98	18.57	11.93	9.71	17.28	29.70	18.04
	20	0.1	5.94	8.43	19.78	8.60	3.05	6.35	13.06	6.76	6.92	10.84	24.35	11.26
		0.4	5.76	8.75	22.96	8.89	4.26	6.64	13.33	7.33	7.47	11.25	27.40	11.78
		0.7	5.71	10.28	24.75	10.65	6.81	9.87	19.01	10.78	9.06	14.45	32.46	15.43
	30	0.1	4.92	6.97	17.52	7.18	2.04	5.15	12.42	5.37	5.48	8.88	21.77	9.20
		0.4	4.78	7.16	19.26	7.19	3.97	5.58	13.42	6.07	6.41	9.26	24.31	9.59
		0.7	4.74	7.78	14.81	7.95	6.38	8.00	15.91	8.91	8.06	11.36	22.57	12.18
	60	0.1	3.53	4.58	8.49	4.72	2.21	3.64	7.43	3.99	4.30	6.00	11.58	6.33
		0.4	3.42	4.80	8.40	4.97	3.80	5.02	10.05	5.41	5.28	7.09	13.37	7.48
		0.7	3.39	5.40	8.04	5.71	6.69	7.64	14.40	8.18	7.55	9.45	16.74	10.09
	100	0.1	2.70	3.45	5.92	3.64	1.62	3.14	5.64	3.24	3.28	4.79	8.36	5.02
		0.4	2.67	3.60	5.70	3.78	4.00	4.17	8.35	4.39	4.90	5.59	10.24	5.88
		0.7	2.63	4.07	5.68	4.31	6.28	6.14	13.00	6.53	6.84	7.41	14.29	7.89
	150	0.1	2.21	2.89	4.72	2.94	1.59	2.84	4.68	2.92	2.88	4.15	6.80	4.26
		0.4	2.16	2.96	4.55	3.08	3.72	3.90	8.02	4.03	4.41	4.95	9.33	5.13
		0.7	2.16	3.27	4.50	3.48	6.49	6.19	12.73	6.52	6.86	7.03	13.55	7.44
6	15	0.1	8.12	11.43	33.66	11.40	3.89	7.79	17.26	7.91	9.30	14.11	38.67	14.20
		0.4	8.24	12.49	24.42	12.39	4.67	8.99	17.45	9.39	9.82	15.77	31.02	15.87
		0.7	7.95	17.12	27.80	17.50	7.41	14.01	20.67	15.67	11.19	22.54	36.98	23.93
	20	0.1	7.18	9.26	20.78	9.69	3.28	6.36	14.45	6.74	8.13	11.50	25.88	12.02
		0.4	7.14	10.67	17.20	10.90	4.63	8.07	14.60	8.95	8.80	13.60	23.38	14.37
		0.7	6.84	13.25	21.47	13.74	7.35	12.11	20.29	14.35	10.25	18.27	31.00	20.26
	30	0.1	6.02	7.40	14.06	7.71	2.76	5.42	11.00	5.75	6.80	9.36	18.34	9.90
		0.4	5.82	8.33	14.76	8.69	4.53	7.03	13.86	7.49	7.70	11.12	20.92	11.71
		0.7	5.60	9.53	15.17	9.80	7.18	9.71	18.54	10.62	9.23	13.84	25.15	14.69
	60	0.1	4.24	5.11	8.47	5.27	2.43	4.03	7.60	4.44	5.07	6.69	11.64	7.05
		0.4	4.13	5.47	8.21	5.69	5.05	5.57	11.71	5.96	6.70	7.93	14.67	8.37
		0.7	4.02	6.39	8.40	6.85	7.36	8.57	17.05	9.24	8.46	10.80	19.31	11.66
	100	0.1	3.28	3.82	6.15	3.97	2.19	3.40	5.73	3.52	4.11	5.26	8.59	5.46
		0.4	3.25	4.21	6.28	4.39	4.48	4.94	10.31	5.28	5.65	6.57	12.32	6.97
		0.7	3.13	4.98	6.03	5.20	7.34	8.15	17.05	8.45	8.01	9.59	18.24	9.98
	150	0.1	2.70	3.15	4.88	3.21	1.74	3.04	5.16	3.08	3.36	4.48	7.26	4.56
		0.4	2.60	3.43	5.02	3.61	4.92	5.04	9.65	5.24	5.67	6.18	10.99	6.44
		0.7	2.56	3.98	4.95	4.25	7.56	8.09	16.55	8.45	7.99	9.05	17.33	9.50
11	15	0.1	8.16	12.46	26.91	12.19	3.22	8.63	15.81	9.06	9.00	15.57	32.09	15.66
		0.4	8.13	13.56	30.21	13.88	4.83	9.44	17.71	10.28	9.69	16.89	36.12	17.69
		0.7	7.65	18.91	31.94	20.27	7.18	17.27	23.06	19.66	10.71	26.29	41.16	28.98
	20	0.1	7.34	9.86	23.93	10.01	3.41	6.51	14.81	6.92	8.32	12.08	28.67	12.48
		0.4	7.05	11.12	23.72	11.49	4.86	8.93	15.67	9.91	8.80	14.61	29.54	15.58
		0.7	6.84	14.24	22.57	15.13	7.32	11.93	20.73	13.52	10.19	18.92	32.06	20.58
	30	0.1	6.10	7.67	15.85	8.03	2.58	5.60	11.79	6.17	6.81	9.70	20.15	10.34
		0.4	5.85	8.64	25.40	8.94	4.56	7.45	15.19	8.12	7.66	11.68	31.38	12.41
		0.7	5.61	10.38	15.53	11.10	7.27	10.74	19.14	12.36	9.31	15.12	25.83	16.90
	60	0.1	4.31	5.27	9.13	5.47	2.47	4.37	7.93	4.81	5.15	7.01	12.42	7.41
		0.4	4.14	5.75	9.12	5.94	4.50	5.84	10.91	6.20	6.33	8.37	14.59	8.79
		0.7	3.98	7.21	8.30	7.54	7.27	9.87	18.71	10.82	8.38	12.35	20.75	13.35
	100	0.1	3.28	3.99	6.33	4.21	2.23	3.83	6.69	3.95	4.14	5.68	9.50	5.91
		0.4	3.26	4.40	6.58	4.59	4.32	5.47	10.82	5.80	5.57	7.15	12.94	7.56
		0.7	3.16	5.39	6.53	5.79	7.17	9.20	16.85	9.54	7.87	10.73	18.20	11.24
	150	0.1	2.73	3.21	5.10	3.31	1.83	3.36	5.43	3.46	3.45	4.79	7.67	4.92
		0.4	2.63	3.54	5.15	3.68	4.44	4.96	9.92	5.04	5.27	6.18	11.29	6.33
		0.7	2.56	4.35	5.17	4.60	7.01	8.56	16.86	8.85	7.48	9.64	17.70	10.02

A: Madde Sayısına bağlı Puanlama Biçimi (3: üç kategorili 50 madde, 6: altı kategorili 20 madde, 11: onbir kategorili 10 madde)
B: Örneklem büyüklüğü
C: Eşik parametreleri arasındaki fark

O: Ortalamalar ile test eşitleme
T: Tucker doğrusal test eşitleme
DA: Düşük ağırlıklar doğrusal test eşitleme
BH: Braun-Holland doğrusal test eşitleme

Örneklem Büyüklüğü ve Madde Sayısına Bağlı Puanlama Biçimi Değişimlerinin ESH, Yanlılık ve OHKK Değerlerine Ortak Etkisi

Örneklem büyüklüğü koşuluna göre her test eşitleme yöntemi tarafından üretilen ortalama ESH, yanlılık ve OHKK değerleri ve bunlara ait standart sapmalar Tablo 8'de verilmiştir. Madde sayısına bağlı puanlama biçimi koşuluna göre test eşitleme yöntemlerinden elde edilen ortalama ESH, yanlılık ve OHKK değerleri ve bunlara ait standart sapmalar ise Tablo 9'da verilmiştir.

Tablo 8.

Örneklem büyüklüğüne göre ESH, yanlılık ve OHKK ortalama ve standart sapmaları

						Örneklem Büyüklüğü	
Test Eşitleme	İstatisik		*15*		*20*		*30*
Yöntemi		X̄	S	X̄	S	X̄	S
Ortalamalar	*ESH*	7.59	.70	6.64	.65	5.49	.54
	Yanlılık	5.02	1.70	5.00	1.74	4.59	1.98
	OHKK	9.47	1.13	8.66	1.11	7.50	1.27
Tucker	*ESH*	13.39	2.82	10.65	1.97	8.21	1.14
	Yanlılık	10.36	3.26	8.53	2.34	7.19	2.00
	OHKK	17.35	4.32	13.95	2.97	11.15	2.16
Braun-Holland	*ESH*	13.58	3.25	11.01	2.17	8.51	1.29
	Yanlılık	11.10	4.08	9.47	2.92	7.87	2.39
	OHKK	17.99	5.21	14.86	3.52	11.88	2.55
Düşük ağırlıklar	*ESH*	28.32	3.82	21.91	2.38	16.93	3.56
	Yanlılık	18.09	2.42	16.22	2.98	14.59	2.87
	OHKK	34.82	3.94	28.30	3.30	23.38	3.86
						Örneklem Büyüklüğü	
Test Eşitleme	İstatistik		*60*		*100*		*150*
Yöntemi		X̄	S	X̄	S	X̄	S
Ortalamalar	*ESH*	3.91	.36	3.04	.29	2.48	.23
	Yanlılık	4.64	2.09	4.40	2.16	4.37	2.33
	OHKK	6.36	1.52	5.60	1.69	5.26	1.88
Tucker	*ESH*	5.55	.81	4.21	.63	3.42	.48
	Yanlılık	6.06	2.17	5.38	2.12	5.11	2.12
	OHKK	8.41	2.09	6.97	2.02	6.27	1.98
Braun-Holland	*ESH*	5.80	.90	4.43	.69	3.57	.55
	Yanlılık	6.56	2.34	5.63	2.20	5.29	2.23
	OHKK	8.95	2.33	7.32	2.11	6.51	2.11
Düşük ağırlıklar	*ESH*	8.51	.38	6.13	.33	4.89	.25
	Yanlılık	11.75	4.15	10.49	4.40	9.89	4.66
	OHKK	15.01	3.30	12.52	3.80	11.32	4.12

Tablo 9.

Madde sayısına bağlı puanlama biçimi koşuluna göre ESH, yanlılık ve OHKK ortalama ve standart sapmaları

Test Eşitleme Yöntemi	İstatistik	*Madde Sayısına Bağlı Puanlama Biçimi*					
		3 kategorili 50 madde		*6 kategorili 20 madde*		*11kategorili 10 madde*	
		X̄	S	X̄	S	X̄	S
Ortalamalar	*ESH*	4.26	1.69	5.16	2.06	5.15	2.03
	Yanlılık	4.27	1.87	4.93	2.01	4.80	1.96
	OHKK	6.40	1.93	7.57	2.15	7.45	2.10
Tucker	*ESH*	6.61	3.18	7.78	4.00	8.33	4.41
	Yanlılık	6.19	2.29	7.24	2.90	7.89	3.41
	OHKK	9.30	3.82	10.93	4.78	11.82	5.45
Braun-Holland	*ESH*	6.76	3.15	8.01	4.03	8.68	4.64
	Yanlılık	6.57	2.49	7.81	3.37	8.58	3.98
	OHKK	9.70	3.92	11.50	5.12	12.56	6.03
Düşük ağırlıklar	*ESH*	14.17	8.95	13.76	8.83	15.42	9.71
	Yanlılık	12.35	4.21	13.83	4.77	14.34	4.95
	OHKK	19.69	9.28	20.65	9.36	22.34	10.04

Tablo 8 ve 9'da görülen ESH, yanlılık ve OHKK değerlerine ait ortalamalardaki değişimlerin istatiksel manidarlığı sınanmıştır. Bunun için etkileşim etkilerini de dikkate alan faktöryel varyans analizleri uygulanmıştır. Bu yöntemle hem ana etkiler hem de etkileşim etkileri test edilmiştir. Manidar olmayan etkiler çıkarılarak en sade modeller tanımlanmıştır. ESH değerleri üzerine yapılan etkilerin incelendiği ANOVA testi sonuçları Tablo 10'da gösterilmiştir.

Tablo 10.

Test eşitleme yöntemi, örneklem büyüklüğü ve madde sayısına bağlı puanlama biçimi koşullarının ESH değerleri üzerindeki etkileri

	Varyansın kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p	η^2
Model 1*	Yöntem (Y)	2692	3	897.3	338.01	.00	.88
	Örneklem (Ö)	4213	5	842.6	317.41	.00	.92
	Madde sayısına bağlı puanlama (MSBP)	75	2	37.5	14.15	.00	.16
	Y*Ö	1409	15	93.9	35.38	.00	.79
	Y*MSBP	23	6	3.9	1.46	.19	.06
	Ö*MSBP	29	10	2.9	1.10	.37	.07
	Y*Ö*MSBP	31	30	1.0	0.39	.99	.08
	Hata	382	144	2.7			
Model 2**	Yöntem (Y)	2692	3	897.3	365.82	.00	.85
	Örneklem (Ö)	4213	5	842.6	343.53	.00	.90
	Madde sayısına bağlı puanlama (MSBP)	75	2	37.5	15.31	.00	.13
	Y*Ö	1409	15	93.9	38.29	.00	.75
	Hata	466	190	2.5			

*Tüm etkiler modeli **İndirgenmiş model

Cohen'e (1988) göre, eta-kare (η^2) değerleri .01 düşük, .06 orta ve .14 yüksek etki büyüklüğü olarak tanımlanmaktadır. Tablo 10'da yer alan tüm etkilerin dikkate alındığı ANOVA modeline (model 1) ait sonuçlar incelendiğinde, test eşitleme yöntemlerinden elde edilen ortalama ESH değerleri arasında yüksek bir etki ile istatiksel olarak manidar bir fark olduğu görülmektedir ($F_{(3, 144)} = 338.01$, $p < .05$, $\eta^2 = .88$). Örneklem büyüklüğü koşulları arasında da yüksek bir etki ile istatiksel olarak manidar bir fark olduğu ($F_{(5, 144)} = 317.41$, $p < .05$, $\eta^2 = .92$), madde sayısına bağlı puanlama biçimi koşulları arasında yüksek bir etki ile istatiksel olarak manidar bir farklılığın olduğu ($F_{(2, 144)} = 14.15$, $p < .05$, $\eta^2 = .16$) görülmektedir. Yani ana etkiler manidardır. Bunların yanı sıra, kullanılan test eşitleme yöntemi ve örneklem büyüklüğü koşullarının ortak etkisinin yüksek bir etki ile istatiksel olarak manidar olduğu da görülmektedir ($F_{(15, 144)} = 35.38$, $p < .05$, $\eta^2 = .79$). Ancak kullanılan test eşitleme yöntemi ve madde sayısına bağlı puanlama biçimi koşullarının ortalama ESH değerleri üzerinde manidar bir ortak etkiye sahip olmadığı görülmektedir ($F_{(6, 144)} = 1.46$, $p > .05$, $\eta^2 = .06$). Benzer bir şekilde; örneklem büyüklüğü ve madde sayısına bağlı puanlama biçimi koşullarının ortalama ESH değerleri üzerinde manidar bir ortak etkiye sahip olmadığı görülmektedir ($F_{(10, 144)} = 1.10$, $p > .05$, $\eta^2 = .07$). Son olarak; test eşitleme yöntemi, örneklem büyüklüğü ve madde sayısına bağlı puanlama biçimi koşullarının ortalama ESH değerleri üzerine yaptıkları ortak etkinin de istatiksel olarak manidar olmadığı görülmektedir ($F_{(30, 144)} = 0.39$, $p > .05$, $\eta^2 = .08$).

Manidar olmayan bu etkileri çıkartarak en sade modeli oluşturmak için analizler yenilenmiştir. Buna göre indirgenmiş model (model 2) incelendiğinde; test eşitleme yöntemlerinden elde edilen ortalama ESH değerleri arasında yüksek bir etkiyle istatiksel olarak manidar bir fark olduğu görülmektedir ($F_{(3, 190)} = 365.82$, $p < .05$, $\eta^2 = .85$). Örneklem büyüklüğü koşulları arasında da yüksek bir etkiyle istatiksel olarak manidar bir fark olduğu görülmektedir ($F_{(5, 190)} = 343.53$, $p < .05$, $\eta^2 = .90$). Madde sayısına bağlı puanlama biçimi koşulları incelendiğinde, aralarında orta düzey bir etki büyüklüğü ile istatiksel olarak manidar bir farklılığın olduğu görülmektedir ($F_{(2, 190)} = 15.31$, $p < .05$, $\eta^2 = .13$). Bunların yanı sıra, kullanılan test eşitleme yöntemi ve örneklem büyüklüğü koşullarının ortak etkisinin yüksek bir etkiyle istatiksel olarak manidar olduğu görülmektedir ($F_{(15, 190)} = 38.29$, $p < .05$, $\eta^2 = .75$).

Tablo 10'da görülen manidar farklılıkların hangi düzeyler arasında olduğunu tespit etmek için devam analizi olarak Tukey HSD testi uygulanmıştır. Bu testten elde

edilen manidarlık düzeyleri Ek 2'de gösterilmiştir. Ek 2. A incelendiğinde, Tucker ve Braun-Holland doğrusal test eşitleme yöntemleri arasında istatiksel olarak manidar bir fark bulunmazken, diğer tüm yöntemlerin ikili karşılaştırılması durumunda manidar bir farklılığın bulunduğu görülmektedir. Bunun yanı sıra, Ek 2.B incelendiğinde, 150 örneklem büyüklüğü koşulu hariç tüm test örneklem büyüklüğü koşullarında üretilen ESH değerleri açısından manidar fark bulunmaktadır. Ek 2.C incelendiğinde, tüm madde sayısına bağlı puanlama biçimi düzeylerinin birbirlerinde istatiksel olarak manidar düzeyde farklılaştığı görülmektedir. Örneklem büyüklüğü ve test eşitleme yöntemi koşullarının ortak etkilerinin araştırıldığı Ek 2.D-1 incelendiğinde, tüm örneklem büyüklüklerinde en düşük ESH değerlerini üreten yöntemin ortalamalar ile test eşitleme yöntemi olduğu dikkat çekmektedir. Bu yöntemden elde edilen ortalama ESH değerlerinin 15 ve 20 örneklem büyüklüklerinde diğer yöntemlerden manidar şekilde daha az olduğu dikkat çekmektedir. Tucker doğrusal test eşitleme yöntemi, 30 ve üzeri örneklem büyüklüklerinde ortalamalar ile test eşitleme yöntemine benzer sonuçlar elde edilmekte iken, 60 ve üzeri örneklem büyüklüğünde Braun-Holland test eşitleme yöntemiyle de ortalamalar ile test eşitleme yöntemine benzer ESH değerleri hesaplanmaktadır. Düşük ağırlıklar doğrusal test eşitleme yönteminden ise sadece 150 örneklem büyüklüğünde ortalamalar ile test eşitleme yöntemine benzer sonuçlar elde edilmektedir.

Yanlılık değerleri üzerine yapılan etkilerin incelendiği ANOVA testi sonuçları ise Tablo 11'de gösterilmektedir.

Tablo 11.

Test eşitleme yöntemi, örneklem büyüklüğü ve madde sayısına bağlı puanlama biçimi koşullarının yanlılık değerleri üzerindeki etkileri

	Varyansın kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p	η^2
Model 1*	Yöntem (Y)	2273.5	3	757.8	79.90	.00	.63
	Örneklem (Ö)	696.5	5	139.3	14.69	.00	.34
	Madde sayısına bağlı puanlama (MSBP)	92.4	2	46.2	4.87	.00	.06
	Y*Ö	220.6	15	14.7	1.55	.09	.13
	Y*MSBP	13.7	6	2.3	0.24	.96	.01
	Ö*MSBP	4.5	10	0.4	0.05	.99	.01
	Y*Ö*MSBP	7.9	30	0.3	0.03	.99	.01
	Hata	1365.9	144	9.5			
Model 2**	Yöntem (Y)	2273	3	757.8	103.4	.00	.62
	Örneklem (Ö)	697	5	139.3	19.01	.00	.33
	Madde sayısına bağlı puanlama (MSBP)	92	2	46.2	6.31	.00	.06
	Y*Ö	221	15	14.7	2.01	.02	.14
	Hata	1392	190	7.3			

*Tüm etkiler modeli **İndirgenmiş model

Tablo 11'de yer alan tüm etkilerin dikkate alındığı ANOVA modeline (model 1) ait sonuçlar incelendiğinde, test eşitleme yöntemlerinden elde edilen ortalama yanlılık değerleri arasında yüksek düzeyde istatiksel olarak manidar bir fark olduğu görülmektedir ($F_{(3,\ 144)}$ =79.90, p <.05, η^2=.63). Örneklem büyüklüğü koşulları arasında da yüksek düzeyde istatiksel olarak manidar bir fark olduğu ($F_{(5,\ 144)}$ =14.69, p <.05, η^2=.34), madde sayısına bağlı puanlama biçimi koşulları arasında orda düzeyde istatiksel olarak manidar bir farklılığın olduğu ($F_{(2,\ 144)}$ =4.87, p <.05, η^2=.06) görülmektedir. Ancak kullanılan test eşitleme yöntemi ve örneklem büyüklüğü koşullarının ortak etkisinin istatiksel olarak manidar olmadığı görülmektedir ($F_{(15,\ 144)}$ =1.55, p>.05, η^2=.13). Bunun yanı sıra, kullanılan test eşitleme yöntemi ve madde sayısına bağlı puanlama biçimi koşullarının ortalama yanlılık değerleri üzerinde manidar bir ortak etkiye sahip olmadığı da görülmektedir ($F_{(6,\ 144)}$ =0.24, p>.05, η^2=.01). Benzer bir şekilde; örneklem büyüklüğü ve madde sayısına bağlı puanlama biçimi koşullarının ortalama yanlılık değerleri üzerinde manidar bir ortak etkiye sahip olmadığı görülmektedir ($F_{(10,\ 144)}$ =0.05, p>.05, η^2=.01). Son olarak; test eşitleme yöntemi, örneklem büyüklüğü ve madde sayısına bağlı puanlama biçimi koşullarının ortalama yanlılık değerleri üzerine yaptıkları ortak etkinin de istatiksel olarak manidar olmadığı görülmektedir ($F_{(30,\ 144)}$ =0.03, p>.05, η^2=.01).

Manidar olmayan bu etkileri çıkartarak en sade modeli oluşturmak için üç yönlü ANOVA uygulanmıştır. İndirgenmiş modele (model 2) ait ANOVA çıktıları incelendiğinde test eşitleme yöntemlerinden elde edilen ortalama yanlılık değerleri arasında yüksek bir etkiyle istatiksel olarak manidar bir fark olduğu görülmektedir ($F_{(3, 190)}=103.4$, $p<.05$, $\eta^2=.62$). Örneklem büyüklüğü koşulları arasında da yüksek bir etkiyle istatiksel olarak manidar bir fark olduğu görülmektedir ($F_{(5, 190)}=19.01$, $p<.05$, $\eta^2=.33$). Madde sayısına bağlı puanlama biçimi koşulları incelendiğinde, aralarında küçük bir etki büyüklüğü ile istatiksel olarak manidar bir farklılığın olduğu görülmektedir ($F_{(2, 190)}=6.31$, $p<.05$, $\eta^2=.06$). Bunların yanı sıra, kullanılan test eşitleme yöntemi ve örneklem büyüklüğü koşullarının ortak etkisinin orta düzey bir etkiyle istatiksel olarak manidar olduğu görülmektedir ($F_{(15, 190)}=2.01$, $p<.05$, $\eta^2=.14$).

Tablo 11'de görülen manidar farklılıkların hangi düzeyler arasında olduğunu tespit etmek için devam analizi olarak Tukey HSD testi uygulanmıştır. Bu testten elde edilen manidarlık düzeyleri Ek 2'de gösterilmiştir. Ek 2. A incelendiğinde, yanlılık değerleri açısından Tucker ve Braun-Holland doğrusal test eşitleme yöntemleri arasında istatiksel olarak manidar bir fark bulunmazken, diğer tüm yöntemlerin ikili karşılaştırılması durumunda manidar bir farklılığın bulunduğu görülmektedir. Bunun yanı sıra, Ek 2.B incelendiğinde, 60 ve üzeri örneklem büyüklüğü koşulu hariç tüm test örneklem büyüklüğü koşullarında üretilen yanlılık değerleri açısından manidar fark bulunmaktadır. Ek 2.C incelendiğinde, 3 kategorili 50 madde koşulunun diğer madde sayısına bağlı puanlama biçimi koşullarından istatiksel olarak daha düşük ortalama yanlılık düzeyi verdiği görülmektedir. Örneklem büyüklüğü ve test eşitleme yöntemi koşullarının ortak etkilerinin araştırıldığı Ek 2.D-2 incelendiğinde, tüm örneklem büyüklüklerinde en düşük ortalama yanlılık değerlerini üreten yöntemin ortalamalar ile test eşitleme yöntemi olduğu dikkat çekmektedir. Bunun yanı sıra, örneklem büyüklüğü değişimleri bu yöntemden elde edilen yanlılık düzeylerine etki etmemektedir. 20 ve üzeri örneklem büyüklüğü koşullarında Tucker doğrusal, Braun-Holland doğrusal ve ortalamalar ile test eşitleme yöntemlerinin ürettikleri ortalama yanlılık düzeyleri arasında istatiksel olarak manidar bir farklılığın olmadığı görülmektedir. Bunun yanı sıra, düşük ağırlıklar doğrusal ve ortalamalar ile test eşitleme yöntemlerinden çalışma kapsamında ele alınan hiçbir örneklem büyüklüğü koşulunda benzer ortalama yanlılık düzeyleri elde edilmemektedir.

OHKK değerleri üzerine yapılan etkilerin incelendiği ANOVA testi sonuçları ise Tablo 12'de gösterilmektedir.

Tablo 12.

Test eşitleme yöntemi, örneklem büyüklüğü ve madde sayısına bağlı puanlama biçimi koşullarının OHKK değerleri üzerindeki etkileri

	Varyansın kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p	η^2
Model 1*	Yöntem (Y)	5617	3	1872.4	202.42	.00	.80
	Örneklem (Ö)	4508	5	901.7	97.48	.00	.77
	Madde sayısına bağlı puanlama(MSBP)	189	2	94.3	10.20	.00	.12
	Y*Ö	1445	15	96.3	10.41	.00	.52
	Y*MSBP	25	6	4.2	.45	.84	.02
	Ö*MSBP	28	10	2.8	.31	.98	.02
	Y*Ö*MSBP	36	30	1.2	.13	.99	.03
	Hata	1332	144	9.3			
Model 2**	Yöntem (Y)	5617	3	1872.4	250.22	.00	.80
	Örneklem (Ö)	4508	5	901.7	120.49	.00	.76
	Madde sayısına bağlı puanlama(MSBP)	189	2	94.3	12.60	.00	.12
	Y*Ö	1445	15	96.3	12.87	.00	.50
	Hata	1422	190	7.5			

*Tüm etkiler modeli **İndirgenmiş model

Tablo 12'de yer alan tüm etkilerin dikkate alındığı ANOVA modeline (model 1) ait sonuçlar incelendiğinde, test eşitleme yöntemlerinden elde edilen ortalama OHKK değerleri arasında yüksek bir etki ile istatiksel olarak manidar bir fark olduğu görülmektedir ($F_{(3, 144)}$ =202.42, p <.05, η^2=.80). Örneklem büyüklüğü koşulları arasında da yüksek bir etki ile istatiksel olarak manidar bir fark olduğu ($F_{(5, 144)}$ =97.48, p <.05, η^2=.77), madde sayısına bağlı puanlama biçimi koşulları arasında yüksek bir etki ile istatiksel olarak manidar bir farklılığın olduğu ($F_{(2, 144)}$ =10.20, p <.05, η^2=.12) görülmektedir. Bunların yanı sıra, kullanılan test eşitleme yöntemi ve örneklem büyüklüğü koşullarının ortak etkisinin yüksek bir etki ile istatiksel olarak manidar olduğu da görülmektedir ($F_{(15, 144)}$ =10.41, p <.05, η^2=.52). Ancak kullanılan test eşitleme yöntemi ve madde sayısına bağlı puanlama biçimi koşullarının ortalama OHKK değerleri üzerinde manidar bir ortak etkiye sahip olmadığı görülmektedir ($F_{(6, 144)}$ =.45, p>.05, η^2=.02). Benzer bir şekilde; örneklem büyüklüğü ve madde sayısına bağlı puanlama biçimi koşullarının ortalama OHKK değerleri üzerinde manidar bir ortak etkiye sahip olmadığı görülmektedir ($F_{(10, 144)}$ =.31, p>.05, η^2=.02). Son olarak; test eşitleme yöntemi, örneklem büyüklüğü ve madde sayısına bağlı puanlama biçimi koşullarının ortalama

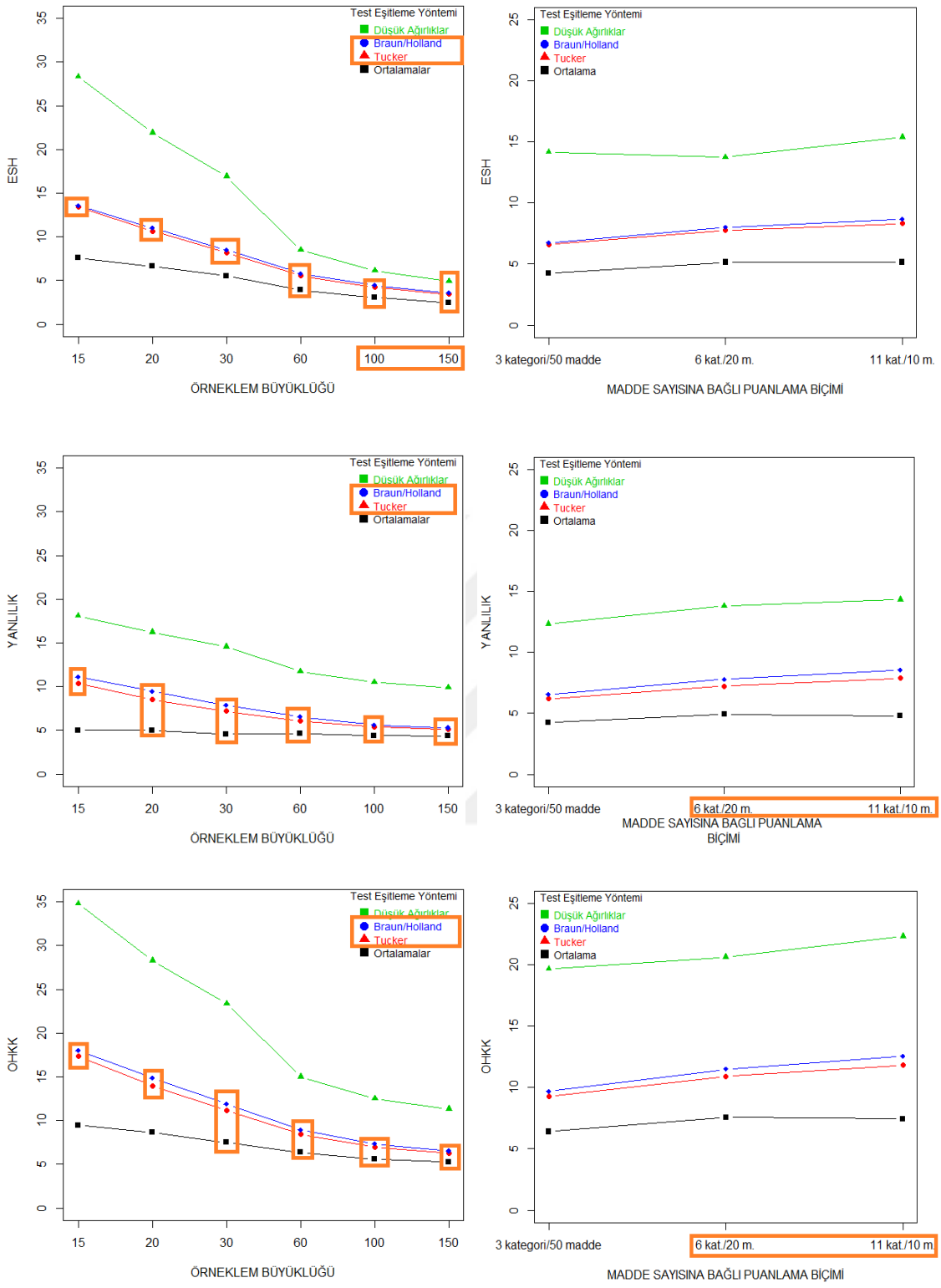
OHKK değerleri üzerine yaptıkları ortak etkinin de istatiksel olarak manidar olmadığı görülmektedir ($F_{(30, 144)} = .13$, $p > .05$, $\eta^2 = .03$).

Manidar olmayan bu etkileri çıkartarak en sade modeli oluşturmak için analizler yenilenmiştir. Bu indirgenmiş modele (model 2) ait ANOVA çıktıları incelendiğinde, test eşitleme yöntemlerinden elde edilen ortalama OHKK değerleri arasında yüksek bir etkiyle istatiksel olarak manidar bir fark olduğu görülmektedir ($F_{(3, 190)} = 250.22$, $p < .05$, $\eta^2 = .80$). Örneklem büyüklüğü koşulları arasında da yüksek bir etkiyle istatiksel olarak manidar bir fark olduğu görülmektedir ($F_{(5, 190)} = 120.49$, $p < .05$, $\eta^2 = .76$). Madde sayısına bağlı puanlama biçimi koşulları incelendiğinde, aralarında orta düzey bir etki büyüklüğü ile istatiksel olarak manidar bir farklılığın olduğu görülmektedir ($F_{(2, 190)} = 12.60$, $p < .05$, $\eta^2 = .12$). Bunların yanı sıra, kullanılan test eşitleme yöntemi ve örneklem büyüklüğü koşullarının ortak etkisinin yüksek bir etkiyle istatiksel olarak manidar olduğu görülmektedir ($F_{(15, 190)} = 12.87$, $p < .05$, $\eta^2 = .50$).

Tablo 12'de görülen manidar farklılıkların hangi düzeyler arasında olduğunu tespit etmek için devam analizi olarak Tukey HSD testi uygulanmıştır. Bu testten elde edilen manidarlık düzeyleri Ek 2'de gösterilmiştir. Ek 2. A incelendiğinde, OHKK değerleri açısından test eşitleme yöntemlerinin Tucker ve Braun-Holland doğrusal test eşitleme yöntemleri haricinde birbirlerinden manidar düzeyde farklılaştığı görülmektedir. Bunun yanı sıra, Ek 2.B incelendiğinde, test eşitleme yöntemlerinden elde edilen OHKK değerlerinin ortalaması, 60 ve üzeri örneklem büyüklüğünde istatiksel olarak benzerlik göstermektedir. Ek 2.C incelendiğinde, ortalama OHKK değerleri açısından, 6 kategorili 20 madde ve 11 kategorili 10 madde koşullarının aralarında istatiksel olarak manidar bir farkın bulunmadığı, bunun dışında kalan tüm ikili karşılaştırmalarda istatiksel olarak manidar bir farklılığın olduğu görülmektedir. Örneklem büyüklüğü ve test eşitleme yöntemi koşullarının ortak etkilerinin araştırıldığı Ek 2.D-3 incelendiğinde, ortalamalar ile test eşitleme yönteminin tüm örneklem büyüklüklerinde en düşük OHKK değerlerini üreten yöntem olduğu dikkat çekmektedir. Bu yöntemden elde edilen ortalama OHKK değerlerinin 15 ve 20 örneklem büyüklüklerinde diğer yöntemlerden manidar şekilde daha az olduğu göze çarpmaktadır. Tucker ve Braun-Holland doğrusal test eşitleme yöntemleri, 30 ve üzeri örneklem büyüklüklerinde ortalamalar ile test eşitleme yöntemine istatiksel olarak benzer sonuçlar vermektedir. Tucker doğrusal test eşitleme yönteminin ise, tüm örneklem büyüklüklerinde Braun-Holland doğrusal test eşitleme yöntemi ile

istatiksel olarak benzer ortalama OHKK değerleri verdiği dikkat çekmektedir. Düşük ağırlıklar doğrusal test eşitleme yöntemi ise hiçbir örneklem büyüklüğü koşulunda diğer test eşitleme yöntemlerine istatiksel olarak benzer ortalama OHKK değerleri vermemektedir.

Örneklem büyüklüğü koşuluna göre ve madde sayısına bağlı puanlama biçimi koşuluna göre çalışma kapsamında elde edilen ESH, yanlılık ve OHKK değerlerinin değişimlerinin daha kolay görülebilmesi için, bu değişimlere yönelik grafikler Şekil 8'de verilmiştir. Şekil 8'de, Tukey HSD testi sonuçlarına göre aralarında manidar farklılık bulunmayan koşullar varsa, aynı kutuların içerisinde gösterilmiştir.



Şekil 8. Örneklem büyüklüğü değişimlerine ve madde sayısına bağlı puanlama biçimi koşullarına göre test eşitleme yöntemlerinden elde edilen ortalama ESH, Yanlılık ve OHKK değerleri

Şekil 8 test eşitleme yöntemleri açısından incelendiğinde, ortalamalar ile test eşitleme yönteminin her koşulda en düşük ESH, yanlılık ve OHKK değerlerini üreten yöntem olduğu görülmektedir. Bunun yanı sıra, Tucker ve Braun-Holland doğrusal test eşitleme yöntemleri arasında istatiksel olarak bir farklılığın olmadığı, yani çakışıklığın bulunduğu dikkat çekmektedir. Düşük ağırlıklar doğrusal test eşitleme yöntemi ise her koşulda en yüksek eşitleme hatalarını üreten yöntem olarak dikkat çekmektedir.

Örneklem büyüklüğü değişimleri açısından incelendiğinde, ortalamalar ile, Tucker doğrusal ve Braun-Holland doğrusal test eşitleme yöntemlerinin 60 ve üzeri örneklem büyüklüklerinde benzer ESH, yanlılık ve OHKK değerleri verdiği görülmektedir. Ancak, düşük ağırlıklar doğrusal test eşitleme yöntemi 150 örneklem büyüklüğünde diğerleri ile benzer ESH üretirken, yanlılık ve OHKK açısından hiçbir koşulda diğer yöntemler ile benzerlik göstermemektedir.

Madde sayısına bağlı puanlama biçimi değişimleri açısından incelendiğinde ise, üretilen ESH, yanlılık ve OHKK değerlerinin paralel bir şekilde dağıldığı ancak sadece Tucker ve Braun-Holland test eşitleme yöntemleri arasında bir çakışıklığın olduğu görülmektedir. Ayrıca bu koşulun üretilen ESH, yanlılık ve OHKK değerlerine manidar bir etki yapmadığı da dikkat çekmektedir.

Bu araştırma sorusuna ait bulguların literatür ile uyumlu olduğu görülmektedir. Örneklem büyüklüğü arttıkça, çalışma kapsamında incelenen test eşitleme yöntemlerinden elde edilen ESH, yanlılık ve OHKK değerlerinin azaldığı görülmektedir. Bu durum literatürdeki birçok çalışmada da görülmektedir (Kolen, 1985; Livingston, 1993; Hanson ve diğerleri, 1994; Parshall ve diğerleri, 1995; Skaggs, 2005; Kim ve diğerleri, 2008, Livingston ve Kim, 2009; Babcock ve diğerleri, 2012; Albano, 2015).

Örneklem Büyüklüğü ve Madde Eşik Parametresi Farkı Değişimlerinin ESH, Yanlılık ve OHKK Değerlerine Ortak Etkisi

Örneklem büyüklüğü koşuluna göre her test eşitleme yöntemi tarafından üretilen ortalama ESH, yanlılık ve OHKK değerleri ve bunlara ait standart sapmalar Tablo 13'de verilmiştir. Madde eşik parametre farkı koşuluna göre test eşitleme yöntemlerinden elde edilen ortalama ESH, yanlılık ve OHKK değerleri ve bunlara ait standart sapmalar ise Tablo 14'de verilmiştir.

Tablo 13.

Örneklem büyüklüğüne göre ESH, yanlılık ve OHKK ortalama ve standart sapmaları

Test Eşitleme Yöntemi	İstatistik	*Örneklem Büyüklüğü* *15*		*20*		*30*	
		X̄	S	X̄	S	X̄	S
Ortalamalar	*ESH*	7.59	.70	6.64	.65	5.49	.54
	Yanlılık	5.02	1.70	5.00	1.74	4.59	1.98
	OHKK	9.47	1.13	8.66	1.11	7.50	1.27
Tucker	*ESH*	13.39	2.82	10.65	1.97	8.21	1.14
	Yanlılık	10.36	3.26	8.53	2.34	7.19	2.00
	OHKK	17.35	4.32	13.95	2.97	11.15	2.16
Braun-Holland	*ESH*	13.58	3.25	11.01	2.17	8.51	1.29
	Yanlılık	11.10	4.08	9.47	2.92	7.87	2.39
	OHKK	17.99	5.21	14.86	3.52	11.88	2.55
Düşük ağırlıklar	*ESH*	28.32	3.82	21.91	2.38	16.93	3.56
	Yanlılık	18.09	2.42	16.22	2.98	14.59	2.87
	OHKK	34.82	3.94	28.30	3.30	23.38	3.86
Test Eşitleme Yöntemi	İstatistik	Örneklem Büyüklüğü *60*		*100*		*150*	
		X̄	S	X̄	S	X̄	S
Ortalamalar	*ESH*	3.91	.36	3.04	.29	2.48	.23
	Yanlılık	4.64	2.09	4.40	2.16	4.37	2.33
	OHKK	6.36	1.52	5.60	1.69	5.26	1.88
Tucker	*ESH*	5.55	.81	4.21	.63	3.42	.48
	Yanlılık	6.06	2.17	5.38	2.12	5.11	2.12
	OHKK	8.41	2.09	6.97	2.02	6.27	1.98
Braun-Holland	*ESH*	5.80	.90	4.43	.69	3.57	.55
	Yanlılık	6.56	2.34	5.63	2.20	5.29	2.23
	OHKK	8.95	2.33	7.32	2.11	6.51	2.11
Düşük ağırlıklar	*ESH*	8.51	.38	6.13	.33	4.89	.25
	Yanlılık	11.75	4.15	10.49	4.40	9.89	4.66
	OHKK	15.01	3.30	12.52	3.80	11.32	4.12

Tablo 14.

Madde eşik parametre farkı koşuluna göre ESH, yanlılık ve OHKK ortalama ve standart sapmaları

Test Eşitleme Yöntemi	İstatistik	*Madde Eşik Parrametre Farkı* *0.1*		*0.4*		*0.7*	
		X̄	S	X̄	S	X̄	S
Ortalamalar	*ESH*	4.97	2.01	4.87	2.01	4.73	1.92
	Yanlılık	2.52	.66	4.44	.38	7.04	.38
	OHKK	5.78	2.11	6.91	1.68	8.73	1.30
Tucker	*ESH*	6.62	3.10	7.24	3.48	8.85	4.76
	Yanlılık	5.06	1.81	6.42	1.77	9.83	2.76
	OHKK	8.55	3.63	9.94	3.85	13.56	5.32
Braun-Holland	*ESH*	6.78	3.08	7.40	3.46	9.27	4.97
	Yanlılık	5.30	1.85	6.87	1.96	10.80	3.36
	OHKK	8.84	3.65	10.35	3.95	14.56	5.90
Düşük ağırlıklar	*ESH*	14.35	9.16	14.91	9.47	14.08	8.95
	Yanlılık	10.12	4.30	12.66	3.11	17.73	2.70
	OHKK	18.06	10.15	20.56	9.47	24.06	8.14

Tablo 13 ve 14'de görülen ESH, yanlılık ve OHKK değerlerine ait ortalamalardaki değişimlerin istatiksel manidarlığı sınanmıştır. Bunun için etkileşim etkilerini de dikkate alan faktöryel varyans analizleri uygulanmıştır. Bu yöntemle hem ana etkiler hem de etkileşim etkileri test edilmiştir. Manidar olmayan etkiler çıkarılarak en sade modeller tanımlanmıştır. ESH değerleri üzerine yapılan etkilerin incelendiği ANOVA testi sonuçları Tablo 15'de gösterilmiştir.

Tablo 15.

Test eşitleme yöntemi, örneklem büyüklüğü ve madde eşik parametre farkı koşullarının ESH değerleri üzerindeki etkileri

	Varyansın kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p	η^2
Model 1*	Yöntem (Y)	2692	3	897.3	386.86	.00	.89
	Örneklem (Ö)	4213	5	842.6	363.28	.00	.93
	Madde eşik parametre Farkı (MEPF)	40	2	20.2	8.18	.00	.11
	Y*Ö	1409	15	93.9	40.49	.00	.81
	Y* MEPF	75	6	12.5	5.37	.00	.18
	Ö* MEPF	32	10	3.2	1.36	.21	.09
	Y*Ö* MEPF	60	30	2.0	.87	.66	.15
	Hata	334	144	2.3			
Model 2**	Yöntem (Y)	2692	3	897.3	387.63	.00	.86
	Örneklem (Ö)	4213	5	842.6	364.00	.00	.91
	Madde eşik parametre Farkı (MEPF)	40	2	20.2	8.73	.00	.08
	Y*Ö	1409	15	93.9	40.57	.00	.77
	Y* MEPF	75	6	12.5	5.38	.00	.15
	Hata	426	184	2.3			

*Tüm etkiler modeli **İndirgenmiş model

Tablo 15'de yer alan tüm etkilerin dikkate alındığı ANOVA modeline (model 1) ait sonuçlar incelendiğinde, test eşitleme yöntemlerinden elde edilen ortalama ESH değerleri arasında yüksek bir etki ile istatiksel olarak manidar bir fark olduğu görülmektedir ($F_{(3, 144)}$ =386.86, p <.05, η^2=.89). Örneklem büyüklüğü koşulları arasında da yüksek bir etki ile istatiksel olarak manidar bir fark olduğu ($F_{(5, 144)}$ =363.28, p <.05, η^2=.93), madde eşik parametresi farkı koşulları arasında orta düzey bir etki ile istatiksel olarak manidar bir farklılığın olduğu ($F_{(2, 144)}$ =8.18, p <.05, η^2=.11) görülmektedir. Bunların yanı sıra, kullanılan test eşitleme yöntemi ve örneklem büyüklüğü koşullarının ortak etkisinin yüksek bir etki ile istatiksel olarak manidar olduğu ($F_{(15, 144)}$ =40.49, p <.05, η^2=.81) ve test eşitleme yöntemi ile madde eşik parametresi farkı koşullarının ortak etkisinin de yüksek bir etki ile istatiksel olarak manidar olduğu ($F_{(6, 144)}$ =5.37, p <.05, η^2=.18) görülmektedir. Ancak örneklem büyüklüğü ve madde eşik parametresi farkı

koşullarının ortalama ESH değerleri üzerinde manidar bir ortak etkiye sahip olmadığı görülmektedir ($F_{(10, 144)} = 1.36$, $p > .05$, $\eta^2 = .09$). Son olarak; test eşitleme yöntemi, örneklem büyüklüğü ve madde eşik parametresi farkı koşullarının ortalama ESH değerleri üzerine yaptıkları ortak etkinin de istatiksel olarak manidar olmadığı görülmektedir ($F_{(30, 144)} = 0.87$, $p > .05$, $\eta^2 = .15$).

Manidar olmayan bu etkileri çıkartarak en sade modeli oluşturmak için analizler yenilemiştir. Buna göre indirgenmiş model (model 2) incelendiğinde; test eşitleme yöntemlerinden elde edilen ortalama ESH değerleri arasında yüksek bir etkiyle istatiksel olarak manidar bir fark olduğu görülmektedir ($F_{(3, 184)} = 387.63$, $p < .05$, $\eta^2 = .86$). Örneklem büyüklüğü koşulları arasında da çok yüksek bir etkiyle istatiksel olarak manidar bir fark olduğu görülmektedir ($F_{(5, 184)} = 364.00$, $p < .05$, $\eta^2 = .91$). Madde eşik parametresi farkı koşulları incelendiğinde, aralarında orta düzey bir etki büyüklüğü ile istatiksel olarak manidar bir farklılığın olduğu görülmektedir ($F_{(2, 184)} = 8.73$, $p < .05$, $\eta^2 = .08$). Ayrıca, kullanılan test eşitleme yöntemi ve örneklem büyüklüğü koşullarının ortak etkisinin yüksek bir etkiyle istatiksel olarak manidar olduğu görülmektedir ($F_{(15, 184)} = 40.57$, $p < .05$, $\eta^2 = .77$). Bunların yanı sıra, kullanılan test eşitleme yöntemi ve madde eşik parametresi farkı koşullarının ortalama ESH değerleri üzerine yaptığı ortak etkinin de yüksek düzeyde ve istatiksel olarak manidar olduğu görülmektedir ($F_{(6, 184)} = 5.38$, $p < .05$, $\eta^2 = .15$).

Tablo 15'de görülen manidar farklılıkların hangi düzeyler arasında olduğunu tespit etmek için devam analizi olarak Tukey HSD testi uygulanmıştır. Bu testten elde edilen manidarlık düzeyleri Ek 2'de gösterilmiştir. Ek 2.A, Ek 2.B ve Ek 2.D-1'de görülen test eşitleme yöntemleri arasındaki, örneklem büyüklükleri arasındaki ve hem örneklem hem de test eşitleme koşullarının birlikte yaptığı etkiler bir önceki bölümde ("Örneklem Büyüklüğü ve Madde Sayısına Bağlı Puanlama Biçimi Değişimlerinin ESH, Yanlılık ve OHKK Değerlerine Ortak Etkisi") ele alınmıştır. Ek 2. E'de ise madde eşik parametre farkı koşullarının ikili karşılaştırmaları verilmiştir. Bu tablo incelendiğinde, ortalama ESH değerleri açısından, madde eşik parametresi farkının 0.1 olması ile 0.4 olması arasında istatiksel olarak manidar bir farklılığın olmadığı görülmektedir. Ancak, bu farkın 0.7 olması durumlarının diğer iki koşuldan istatiksel olarak manidar düzeyde farklılaştığı görülmektedir. Madde eşik parametre farkı ve test eşitleme yöntemi koşullarının ortak etkilerinin araştırıldığı Ek 2.F-1 incelendiğinde ise, ortalamalar ile test

eşitleme yönteminde madde eşik parametresi farkının istatiksel olarak manidar bir farklılık yaratmadığı görülmektedir. Ancak bu yöntemin tüm madde eşik parametresi farkı koşullarında diğer yöntemlerden istatiksel olarak manidar düzeyde daha düşük ESH değerleri verdiği görülmektedir. Tucker ve Braun-Holland doğrusal test eşitleme yöntemlerinin hiçbir madde eşik parametresi koşulunda birbirlerinden istatiksel olarak farklılaşmadığı görülmektedir. Tucker doğrusal test eşitleme yönteminde madde eşik parametresi farkının sadece 0.1 ve 0.7 olması koşullarının istatiksel olarak farklı ESH düzeyleri verdiği görülmektedir. Braun-Holland doğrusal test eşitleme yönteminde ise, madde eşik parametresi farkının sadece 0.1 ve 0.4 olması koşullarının istatiksel olarak benzer ESH düzeyleri verdiği görülmektedir. Düşük ağırlıklar doğrusal test eşitleme yönteminin ise bütün madde eşik parametresi farkı koşullarında verdiği ortalama ESH değerlerinin birbirlerinden istatiksel olarak faklı olmadığı görülmektedir. Ancak bu yöntemin tüm madde eşik parametresi farkı koşullarında diğer yöntemlerden istatiksel olarak manidar düzeyde daha yüksek ESH değerleri verdiği görülmektedir.

Yanlılık değerleri üzerine yapılan etkilerin incelendiği ANOVA testi sonuçları ise Tablo 16'da gösterilmektedir.

Tablo 16.

Test eşitleme yöntemi, örneklem büyüklüğü ve madde eşik parametre farkı koşullarının yanlılık değerleri üzerindeki etkileri

	Varyansın kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p	η^2
Model 1*	Yöntem (Y)	2273.5	3	757.8	597.26	.00	.93
	Örneklem (Ö)	696.5	5	139.3	109.79	.00	.79
	Madde eşik parametre Farkı (MEPF)	1173.3	2	586.7	462.36	.00	.87
	Y*Ö	220.6	15	14.7	11.59	.00	.55
	Y* MEPF	59.0	6	9.8	7.75	.00	.24
	Ö* MEPF	13.2	10	1.3	1.04	.41	.07
	Y*Ö* MEPF	56.1	30	1.9	1.48	.07	.23
	Hata	182.7	144	1.3			
Model 2**	Yöntem (Y)	2273.5	3	757.8	553.28	.00	.90
	Örneklem (Ö)	696.5	5	139.3	101.70	.00	.73
	Madde eşik parametre Farkı (MEPF)	1173.3	2	586.7	428.31	.00	.82
	Y*Ö	220.6	15	14.7	10.74	.00	.46
	Y* MEPF	59.0	6	9.8	7.18	.00	.19
	Hata	252.0	184	1.4			

*Tüm etkiler modeli **İndirgenmiş model

Tablo 16'da yer alan tüm etkilerin dikkate alındığı ANOVA modeline (model 1) ait sonuçlar incelendiğinde, test eşitleme yöntemlerinden elde edilen ortalama yanlılık değerleri arasında yüksek bir etki ile istatiksel olarak manidar bir fark olduğu görülmektedir ($F_{(3, 144)}$ =597.26, p <.05, η^2=.93). Örneklem büyüklüğü koşulları arasında da yüksek bir etki ile istatiksel olarak manidar bir fark olduğu ($F_{(5, 144)}$ =109.79, p <.05, η^2=.79), madde eşik parametresi farkı koşulları arasında yüksek bir etki ile istatiksel olarak manidar bir farklılığın olduğu ($F_{(2, 144)}$ =462.36, p <.05, η^2=.87) görülmektedir. Bunların yanı sıra, kullanılan test eşitleme yöntemi ve örneklem büyüklüğü koşullarının ortak etkisinin yüksek bir etki ile istatiksel olarak manidar olduğu ($F_{(15, 144)}$ =11.59, p <.05, η^2=.55) ve test eşitleme yöntemi ile madde eşik parametresi farkı koşullarının ortak etkisinin de yüksek bir etki ile istatiksel olarak manidar olduğu ($F_{(6, 144)}$ =7.75, p <.05, η^2=.24) görülmektedir. Ancak örneklem büyüklüğü ve madde eşik parametresi farkı koşullarının ortalama yanlılık değerleri üzerinde manidar bir ortak etkiye sahip olmadığı görülmektedir ($F_{(10, 144)}$ =1.04, p>.05, η^2=.07). Son olarak; test eşitleme yöntemi, örneklem büyüklüğü ve madde eşik parametresi farkı koşullarının ortalama yanlılık değerleri üzerine yaptıkları ortak etkinin de istatiksel olarak manidar olmadığı görülmektedir ($F_{(30, 144)}$ =1.48, p>.05, η^2=.23).

Manidar olmayan bu etkileri çıkartarak en sade modeli oluşturmak için analizler yenilenmiştir. İndirgenmiş modele (model 2) ait ANOVA çıktıları incelendiğinde test eşitleme yöntemlerinden elde edilen ortalama yanlılık değerleri arasında yüksek bir etkiyle istatiksel olarak manidar bir fark olduğu görülmektedir ($F_{(3, 184)}$ =553.28, p <.05, η^2=.90). Örneklem büyüklüğü koşulları arasında da çok yüksek bir etkiyle istatiksel olarak manidar bir fark olduğu görülmektedir ($F_{(5, 184)}$ =101.70, p <.05, η^2=.73). Madde eşik parametresi farkı koşulları incelendiğinde, aralarında yüksek düzey bir etki büyüklüğü ile istatiksel olarak manidar bir farklılığın olduğu görülmektedir ($F_{(2, 184)}$ =428.31, p <.05, η^2=.82). Ayrıca, kullanılan test eşitleme yöntemi ve örneklem büyüklüğü koşullarının ortak etkisinin yüksek bir etkiyle istatiksel olarak manidar olduğu görülmektedir ($F_{(15, 184)}$ =10.74, p <.05, η^2=.46). Bunların yanı sıra, kullanılan test eşitleme yöntemi ve madde eşik parametresi farkı koşullarının ortalama yanlılık değerleri üzerine yaptığı ortak etkinin de yüksek düzeyde ve istatiksel olarak manidar olduğu görülmektedir ($F_{(6, 184)}$ =7.18, p <.05, η^2=.19).

Tablo 16'da görülen manidar farklılıkların hangi düzeyler arasında olduğunu tespit etmek için devam analizi olarak Tukey HSD testi uygulanmıştır. Bu testten elde edilen manidarlık düzeyleri Ek 2'de gösterilmiştir. Ek 2.A, Ek 2.B ve Ek 2.D-2'de görülen test eşitleme yöntemleri arasındaki, örneklem büyüklükleri arasındaki ve hem örneklem hem de test eşitleme koşullarının birlikte yaptığı etkiler bir önceki bölümde ("Örneklem Büyüklüğü ve Madde Sayısına Bağlı Puanlama Biçimi Değişimlerinin ESH, Yanlılık ve OHKK Değerlerine Ortak Etkisi") ele alınmıştır. Ek 2. E'de ise madde eşik parametre farkı koşullarının ikili karşılaştırmaları verilmiştir. Bu tablo incelendiğinde, ortalama yanlılık değerleri açısından, madde eşik parametresi farkı koşullarının birbirlerinden istatiksel olarak manidar düzeyde farklılaştığı görülmektedir. Madde eşik parametre farkı ve test eşitleme yöntemi koşullarının yanlılık değerleri üzerine yaptığı ortak etkilerinin gösterildiği Ek 2.F-2 incelendiğinde ise, ortalamalar ile test eşitleme yönteminde madde eşik parametresi farkının her koşulu için istatiksel olarak manidar bir farklılığın olduğu görülmektedir. Yine bu yöntemin tüm madde eşik parametresi farkı koşullarında diğer yöntemlerden istatiksel olarak manidar düzeyde daha düşük yanlılık değerleri verdiği görülmektedir. Tucker doğrusal test eşitleme yönteminden her madde eşik parametresi farkı koşulunda elde edilen yanlılık değerlerinin birbirlerinden istatiksel olarak farklı olduğu görülmektedir. Tucker ve Braun-Holland doğrusal test eşitleme yöntemlerinin hiçbir madde eşik parametresi koşulunda birbirlerinden istatiksel olarak farklılaşmadığı da görülmektedir. Tucker doğrusal test eşitleme yöntemine benzer bir şekilde, Braun-Holland doğrusal test eşitleme yönteminden de her madde eşik parametresi farkı koşulunda elde edilen yanlılık değerlerinin birbirlerinden istatiksel olarak farklı olduğu görülmektedir. Yine benzer bir şekilde, düşük ağırlıklar doğrusal test eşitleme yönteminden de her madde eşik parametresi farkı koşulunda elde edilen yanlılık değerlerinin birbirlerinden istatiksel olarak farklı olduğu görülmektedir. Bu yöntemin tüm madde eşik parametresi farkı koşullarında diğer yöntemlerden istatiksel olarak manidar düzeyde daha yüksek yanlılık değerleri verdiği de görülmektedir.

OHKK değerleri üzerine yapılan etkilerin incelendiği ANOVA testi sonuçları ise Tablo 17'de gösterilmektedir.

Tablo 17.

Test eşitleme yöntemi, örneklem büyüklüğü ve madde eşik parametre farkı koşullarının OHKK değerleri üzerindeki etkileri

	Varyansın kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p	η^2
Model 1*	Yöntem (Y)	5617	3	1872.4	545.70	.00	.92
	Örneklem (Ö)	4508	5	901.7	262.78	.00	.90
	Madde eşik parametre Farkı (MEPF)	905	2	452.3	131.83	.00	.65
	Y*Ö	1445	15	96.3	28.07	.00	.75
	Y* MEPF	60	6	9.9	2.90	.01	.11
	Ö* MEPF	24	10	2.4	.71	.71	.05
	Y*Ö* MEPF	127	30	4.2	1.24	.20	.21
	Hata	494	144	3.4			
Model 2**	Yöntem (Y)	5617	3	1872.4	533.39	.00	.90
	Örneklem (Ö)	4508	5	901.7	256.84	.00	.87
	Madde eşik parametre Farkı (MEPF)	905	2	452.3	128.86	.00	.58
	Y*Ö	1445	15	96.3	27.44	.00	.69
	Y* MEPF	60	6	9.9	2.83	.01	.09
	Hata	646	184	3.5			

*Tüm etkiler modeli **İndirgenmiş model

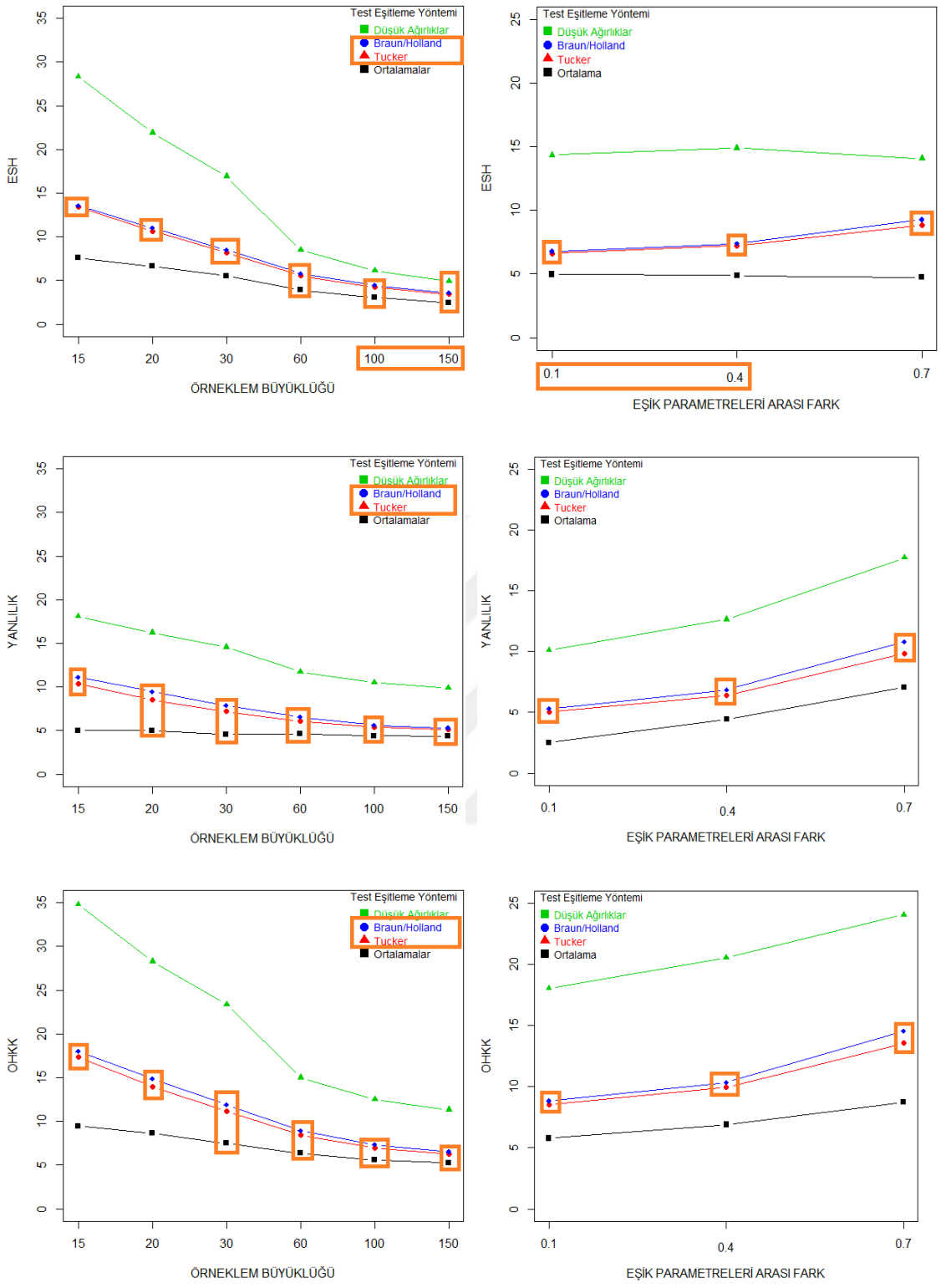
Tablo 17'de yer alan tüm etkilerin dikkate alındığı ANOVA modeline (model 1) ait sonuçlar incelendiğinde, test eşitleme yöntemlerinden elde edilen ortalama OHKK değerleri arasında yüksek bir etki ile istatiksel olarak manidar bir fark olduğu görülmektedir ($F_{(3, 144)}$ =545.70, p <.05, η^2=.92). Örneklem büyüklüğü koşulları arasında da yüksek bir etki ile istatiksel olarak manidar bir fark olduğu ($F_{(5, 144)}$ =262.78, p <.05, η^2=.90), madde eşik parametresi farkı koşulları arasında yüksek bir etki ile istatiksel olarak manidar bir farklılığın olduğu ($F_{(2, 144)}$ =131.83, p <.05, η^2=.65) görülmektedir. Bunların yanı sıra, kullanılan test eşitleme yöntemi ve örneklem büyüklüğü koşullarının ortak etkisinin yüksek bir etki ile istatiksel olarak manidar olduğu ($F_{(15, 144)}$ =28.07, p <.05, η^2=.75) ve test eşitleme yöntemi ile madde eşik parametresi farkı koşullarının ortak etkisinin orta düzey bir etki ile istatiksel olarak manidar olduğu ($F_{(6, 144)}$ =2.90, p <.05, η^2=.11) görülmektedir. Ancak örneklem büyüklüğü ve madde eşik parametresi farkı koşullarının ortalama OHKK değerleri üzerinde manidar bir ortak etkiye sahip olmadığı görülmektedir ($F_{(10, 144)}$ =.71, p>.05, η^2=.05). Son olarak; test eşitleme yöntemi, örneklem büyüklüğü ve madde eşik parametresi farkı koşullarının ortalama OHKK değerleri üzerine yaptıkları ortak etkinin de istatiksel olarak manidar olmadığı görülmektedir ($F_{(30, 144)}$ =1.24, p>.05, η^2=.21).

Manidar olmayan bu etkileri çıkartarak en sade modeli oluşturmak için analizler yenilenmiştir. Bu indirgenmiş modele (model 2) ait ANOVA çıktıları incelendiğinde, test eşitleme yöntemlerinden elde edilen ortalama OHKK değerleri arasında çok yüksek bir etkiyle istatiksel olarak manidar bir fark olduğu görülmektedir ($F_{(3, 184)} = 533.39$, $p < .05$, $\eta^2 = .90$). Örneklem büyüklüğü koşulları arasında da çok yüksek bir etkiyle istatiksel olarak manidar bir fark olduğu görülmektedir ($F_{(5, 184)} = 256.84$, $p < .05$, $\eta^2 = .87$). Madde eşik parametresi farkı koşulları incelendiğinde, aralarında yüksek düzey bir etki büyüklüğü ile istatiksel olarak manidar bir farklılığın olduğu görülmektedir ($F_{(2, 184)} = 128.86$, $p < .05$, $\eta^2 = .58$). Ayrıca, kullanılan test eşitleme yöntemi ve örneklem büyüklüğü koşullarının ortak etkisinin yüksek bir etkiyle istatiksel olarak manidar olduğu görülmektedir ($F_{(15, 184)} = 27.44$, $p < .05$, $\eta^2 = .69$). Bunların yanı sıra, kullanılan test eşitleme yöntemi ve madde eşik parametresi farkı koşullarının ortalama OHKK değerleri üzerine yaptığı ortak etkinin de yüksek düzeyde ve istatiksel olarak manidar olduğu görülmektedir ($F_{(6, 184)} = 2.83$, $p < .05$, $\eta^2 = .09$).

Tablo 17'de görülen manidar farklılıkların hangi düzeyler arasında olduğunu tespit etmek için devam analizi olarak Tukey HSD testi uygulanmıştır. Bu testten elde edilen manidarlık düzeyleri Ek 2'de gösterilmiştir. Ek 2.A, Ek 2.B ve Ek 2.D-3'de görülen test eşitleme yöntemleri arasındaki, örneklem büyüklükleri arasındaki ve hem örneklem hem de test eşitleme koşullarının birlikte yaptığı etkiler bir önceki bölümde ("Örneklem Büyüklüğü ve Madde Sayısına Bağlı Puanlama Biçimi Değişimlerinin ESH, Yanlılık ve OHKK Değerlerine Ortak Etkisi") ele alınmıştır. Ek 2. E'de ise madde eşik parametre farkı koşullarının ikili karşılaştırmaları verilmiştir. Bu tablo incelendiğinde, ortalama OHKK değerleri açısından, değerleri açısından, madde eşik parametresi farkı koşullarının birbirlerinden istatiksel olarak manidar düzeyde farklılaştığı görülmektedir. Ek 2.F-3 incelendiğinde ise, ortalamalar ile test eşitleme yönteminde madde eşik parametresi farkının sadece 0.1 ile 0.7 olduğu koşullar arsında istatiksel olarak manidar bir farklılık oluşturduğu görülmektedir. Ancak bu yöntemin tüm madde eşik parametresi farkı koşullarında diğer yöntemlerden istatiksel olarak manidar düzeyde daha düşük OHKK değerleri verdiği görülmektedir. Tucker ve Braun-Holland doğrusal test eşitleme yöntemlerinin hiçbir madde eşik parametresi koşulunda birbirlerinden istatiksel olarak farklılaşmadığı görülmektedir. Tucker doğrusal test eşitleme yönteminde madde eşik parametresi farkının sadece 0.1 ve 0.4 olması koşullarının arasında istatiksel olarak farklılaşma olduğu görülmektedir. Buna benzer bir durum, Braun-Holland doğrusal test

eşitleme yönteminde de görülmektedir. Düşük ağırlıklar doğrusal test eşitleme yönteminden ise bütün madde eşik parametresi farkı koşullarında elde edilen ortalama OHKK değerlerinin birbirlerinden istatiksel olarak faklı olmadığı görülmektedir. Ancak bu yöntemin tüm madde eşik parametresi farkı koşullarında diğer yöntemlerden istatiksel olarak manidar düzeyde daha yüksek OHKK değerleri verdiği görülmektedir.

Örneklem büyüklüğü koşuluna göre ve madde eşik parametre farkı koşuluna göre çalışma kapsamında elde edilen ESH, yanlılık ve OHKK değerlerinin değişimlerinin daha kolay görülebilmesi için, bu değişimlere yönelik grafikler Şekil 9'da verilmiştir. Şekilde, Tukey HSD testi sonuçlarına göre aralarında manidar farklılık bulunmayan koşullar varsa, aynı kutuların içerisinde gösterilmiştir. Ayrıca Şekil 9'un sol tarafında yer alan örneklem büyüklüğü ile ilgili grafikler bir önceki ("Örneklem Büyüklüğü ve Madde Sayısına Bağlı Puanlama Biçimi Değişimlerinin ESH, Yanlılık ve OHKK Değerlerine Ortak Etkisi") bölümde yorumlanmıştır.



Şekil 9. Örneklem büyüklüğü değişimlerine ve madde eşik parametre farkı koşullarına göre test eşitleme yöntemlerinden elde edilen ortalama ESH, yanlılık ve OHKK değerleri

Madde eşik parametre farkı değişimleri açısından incelendiğinde ise, üretilen ESH, yanlılık ve OHKK değerlerinin paralel bir şekilde dağıldığı ancak sadece Tucker ve Braun-Holland test eşitleme yöntemleri arasında bir çakışıklığın olduğu görülmektedir. Ayrıca madde eşik parametre farkı 0.1 olması ile 0.4 olması arasında üretilen ESH açısından bir farklılık bulunmazken, diğer tüm koşullar arasında tüm hata türleri için manidar farklılık bulunmaktadır.

Bu bulgular, literatürle uyumludur. ESH açısından incelendiğinde; en düşük ortalama ESH düzeyine sahip olan ortalamalar ile test eşitleme yönteminden madde eşik parametresi farkı, yani maddeler arasındaki güçlük farkı, arttıkça daha düşük ESH elde edildiği görülmektedir. Livingston ve Kim (2009) de benzer bir şekilde, maddeler arasındaki güçlük farkının arttığı durumlarda ortalamalar ile test eşitleme yöntemini önermektedir. Bu nedenle bu bulgu literatürle benzerlik göstermektedir. Yanlılık açısından incelendiğinde, madde eşik parametresi farkı koşulu arttıkça tüm yöntemlerden elde edilen yanlılık değerlerinin de manidar düzeyde arttığı görülmektedir. Benzer bir şekilde, Babcock ve diğerleri (2012) de hem ortalamalar ile test eşitleme hem Tucker doğrusal test eşitleme yönteminin hem de düşük ağırlıklar doğrusal test eşitleme yönteminin maddeler arasındaki güçlük parametresi farkı arttıkça daha yüksek yanlılık düzeyleri verdiğini belirtmişlerdir. Bu nedenle bu bulgu literatürle benzerlik göstermektedir. OHKK değerleri açısından incelendiğinde, madde eşik parametresi farkı koşulu arttıkça tüm yöntemlerden elde edilen OHKK değerlerinin de arttığı görülmektedir. Benzer bir şekilde, Babcock ve diğerleri (2012) de hem ortalamalar ile test eşitleme hem Tucker doğrusal test eşitleme yönteminin hem de düşük ağırlıklar doğrusal test eşitleme yönteminin maddeler arasındaki güçlük parametresi farkı arttıkça daha yüksek OHKK düzeyleri verdiğini belirtmişlerdir. Bu nedenle bu bulgu literatürle benzerlik göstermektedir.

Madde Sayısına Bağlı Puanlama Biçimi ve Madde Eşik Parametresi Farkı Değişimlerinin ESH, Yanlılık ve OHKK Değerlerine Ortak Etkisi

Madde sayısına bağlı puanlama biçimi koşuluna göre her test eşitleme yöntemi tarafından üretilen ortalama ESH, yanlılık ve OHKK değerleri ve bunlara ait standart sapmalar Tablo 18'de verilmiştir. Madde eşik parametre farkı koşuluna göre test eşitleme yöntemlerinden elde edilen ortalama ESH, yanlılık ve OHKK değerleri ve bunlara ait standart sapmalar ise Tablo 19'da verilmiştir.

Tablo 18.

Madde sayısına bağlı puanlama biçimi koşuluna göre ESH, yanlılık ve OHKK ortalama ve standart sapmaları

Test Eşitleme Yöntemi	İstatistik	*Madde Sayısına Bağlı Puanlama Biçimi*					
		3 kategorili 50 madde		*6 kategorili 20 madde*		*11kategorili 10 madde*	
		X̄	S	X̄	S	X̄	S
Ortalamalar	*ESH*	4.26	1.69	5.16	2.06	5.15	2.03
	Yanlılık	4.27	1.87	4.93	2.01	4.80	1.96
	OHKK	6.40	1.93	7.57	2.15	7.45	2.10
Tucker	*ESH*	6.61	3.18	7.78	4.00	8.33	4.41
	Yanlılık	6.19	2.29	7.24	2.90	7.89	3.41
	OHKK	9.30	3.82	10.93	4.78	11.82	5.45
Braun-Holland	*ESH*	6.76	3.15	8.01	4.03	8.68	4.64
	Yanlılık	6.57	2.49	7.81	3.37	8.58	3.98
	OHKK	9.70	3.92	11.50	5.12	12.56	6.03
Düşük ağırlıklar	*ESH*	14.17	8.95	13.76	8.83	15.42	9.71
	Yanlılık	12.35	4.21	13.83	4.77	14.34	4.95
	OHKK	19.69	9.28	20.65	9.36	22.34	10.04

Tablo 19.

Madde eşik parametre farkı koşuluna göre ESH, yanlılık ve OHKK ortalama ve standart sapmaları

Test Eşitleme Yöntemi	İstatistik	*Madde Eşik Parrametre Farkı*					
		0.1		*0.4*		*0.7*	
		X̄	S	X̄	S	X̄	S
Ortalamalar	*ESH*	4.97	2.01	4.87	2.01	4.73	1.92
	Yanlılık	2.52	.66	4.44	.38	7.04	.38
	OHKK	5.78	2.11	6.91	1.68	8.73	1.30
Tucker	*ESH*	6.62	3.10	7.24	3.48	8.85	4.76
	Yanlılık	5.06	1.81	6.42	1.77	9.83	2.76
	OHKK	8.55	3.63	9.94	3.85	13.56	5.32
Braun-Holland	*ESH*	6.78	3.08	7.40	3.46	9.27	4.97
	Yanlılık	5.30	1.85	6.87	1.96	10.80	3.36
	OHKK	8.84	3.65	10.35	3.95	14.56	5.90
Düşük ağırlıklar	*ESH*	14.35	9.16	14.91	9.47	14.08	8.95
	Yanlılık	10.12	4.30	12.66	3.11	17.73	2.70
	OHKK	18.06	10.15	20.56	9.47	24.06	8.14

Tablo 18 ve 19'da görülen ESH, yanlılık ve OHKK değerlerine ait ortalamalardaki değişimlerin istatiksel manidarlığı sınanmıştır. Bunun için etkileşim etkilerini de dikkate alan faktöryel varyans analizleri uygulanmıştır. Bu yöntemle hem ana etkiler hem de etkileşim etkileri test edilmiştir. Manidar olmayan etkiler çıkarılarak en sade modeller tanımlanmıştır. ESH değerleri üzerine yapılan etkilerin incelendiği ANOVA testi sonuçları Tablo 20'de gösterilmiştir.

Tablo 20.

Test eşitleme yöntemi, madde sayısına bağlı puanlama biçimi ve madde eşik parametresi farkı koşullarının ESH değerleri üzerindeki etkileri

Varyansın kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p	η^2
Yöntem (Y)	2692	3	897.3	27.38	.00	.31
Madde sayısına bağlı puanlama (MSBP)	75	2	37.5	1.15	.32	.01
Madde eşik parametre Farkı (MEPF)	40	2	20.2	.62	.54	.01
Y*MSBP	23	6	3.9	.12	.99	.00
Y*MEPF	75	6	12.5	.38	.89	.01
MSBP*MEPF	17	4	4.4	.13	.97	.00
Y*MSBP*MEPF	33	12	2.7	.08	1.00	.00
Hata	5899	180	32.8			

Tablo 20'de yer alan tüm etkilerin dikkate alındığı ANOVA modeline ait sonuçlar incelendiğinde, test eşitleme yöntemlerinden elde edilen ortalama ESH değerleri arasında yüksek bir etki ile istatiksel olarak manidar bir fark olduğu görülmektedir ($F_{(3, 180)}$=27.38, $p < .05$, η^2=.31). Ancak madde sayısına bağlı puanlama biçimi koşulları arasında istatiksel olarak manidar bir farklılığın olmadığı görülmektedir ($F_{(2, 180)}$=1.15, $p > .05$, η^2=.01). Benzer bir şekilde, madde eşik parametre farkı koşulları arasında da istatiksel olarak manidar bir farklılık bulunmamaktadır ($F_{(2, 180)}$=.62, $p > .05$, η^2=.01). Test eşitleme yöntemi ve madde sayısına bağlı puanlama biçimi koşullarının birlikte yaptığı ortak etkinin de istatiksel olarak anlamsız olduğu görülmektedir ($F_{(6, 180)}$=.12, $p > .05$, η^2=.00). Benzer bir şekilde, test eşitleme yöntemi ve madde eşik parametre farkı koşullarının ortak etkisinin de istatiksel olarak manidar olmadığı görülmektedir ($F_{(6, 180)}$=.38, $p > .05$, η^2=.01). Madde sayısına bağlı puanlama biçimi ve madde eşik parametresi farkı koşullarının ortak etkisinin de manidar olmadığı görülmektedir ($F_{(4, 180)}$=.13, $p > .05$, η^2=.00). Son olarak, her üç koşul birden yaptığı ortak etkinin de istatiksel olarak manidar olmadığı dikkat çekmektedir ($F_{(12, 180)}$=.08, $p > .05$, η^2=.00). Tablo 20'de görülen tek manidar etkinin test eşitleme yöntemleri arasında olduğu dikkat çekmektedir. Bu etkinin yönünü belirlemek için gerekli devam analizleri "Örneklem Büyüklüğü ve Madde Sayısına Bağlı Puanlama Biçimi Değişimlerinin ESH, Yanlılık ve OHKK Değerlerine Ortak Etkisi" başlığı altında incelenmiştir.

Yanlılık değerleri üzerine yapılan etkilerin incelendiği ANOVA testi sonuçları ise Tablo 21'de gösterilmektedir.

Tablo 21.

Test eşitleme yöntemi, madde sayısına bağlı puanlama biçimi ve madde eşik parametresi farkı koşullarının yanlılık değerleri üzerindeki etkileri

Varyansın kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p	η^2
Yöntem	2273.5	3	757.8	133.28	.00	.69
Madde sayısına bağlı puanlama biçimi (MSBP)	92.4	2	46.2	8.13	.00	.08
Madde Eşik Parametre Farkı (MEPF)	1173.3	2	586.7	103.18	.00	.53
Y*MSBP	13.7	6	2.3	0.40	.88	.01
Y*MEPF	59.0	6	9.8	1.73	.12	.06
MSBP*MEPF	31.0	4	7.8	1.36	.25	.03
Y*MSBP*MEPF	8.5	12	0.7	0.12	.99	.01
Hata	1023.5	180	5.7			

Tablo 21'de yer alan tüm etkilerin dikkate alındığı ANOVA modeline ait sonuçlar incelendiğinde, test eşitleme yöntemlerinden elde edilen ortalama yanlılık değerleri arasında yüksek bir etki ile istatiksel olarak manidar bir fark olduğu görülmektedir ($F_{(3, 180)}$ =133.28, p <.05, η^2=.69). Benzer bir şekilde, madde sayısına bağlı puanlama biçimi koşulları arasında da orta düzey bir etki ile istatiksel olarak manidar bir farklılığın olduğu görülmektedir ($F_{(2, 180)}$ =8.13, p>.05, η^2=.08). Madde eşik parametre farkı koşulları arasında ise yüksek düzey bir etki ile istatiksel olarak manidar bir farklılık bulunmaktadır ($F_{(2, 180)}$ =103.18, p>.05, η^2=.53). Test eşitleme yöntemi ve madde sayısına bağlı puanlama biçimi koşullarının birlikte yaptığı ortak etkinin ise istatiksel olarak anlamsız olduğu görülmektedir ($F_{(6, 180)}$ =.40, p>.05, η^2=.01). Benzer bir şekilde, test eşitleme yöntemi ve madde eşik parametre farkı koşullarının ortak etkisinin de istatiksel olarak manidar olmadığı görülmektedir ($F_{(6, 180)}$ =1.73, p>.05, η^2=.06). Madde sayısına bağlı puanlama biçimi ve madde eşik parametresi farkı koşullarının ortak etkisinin de manidar olmadığı görülmektedir ($F_{(4, 180)}$ =1.36, p>.05, η^2=.03). Son olarak, her üç koşulun birden yaptığı ortak etkinin de istatiksel olarak manidar olmadığı dikkat çekmektedir ($F_{(12, 180)}$ =.12, p>.05, η^2=.01). Tablo 21'de görülen tek manidar etkinin test eşitleme yöntemleri arasında olduğu dikkat çekmektedir. Bu etkinin yönünü belirlemek için gerekli devam analizleri "Örneklem Büyüklüğü ve Madde Sayısına Bağlı Puanlama Biçimi Değişimlerinin ESH, Yanlılık ve OHKK Değerlerine Ortak Etkisi" başlığı altında incelenmiştir.

OHKK değerleri üzerine yapılan etkilerin incelendiği ANOVA testi sonuçları ise Tablo 22'de gösterilmektedir.

Tablo 22.

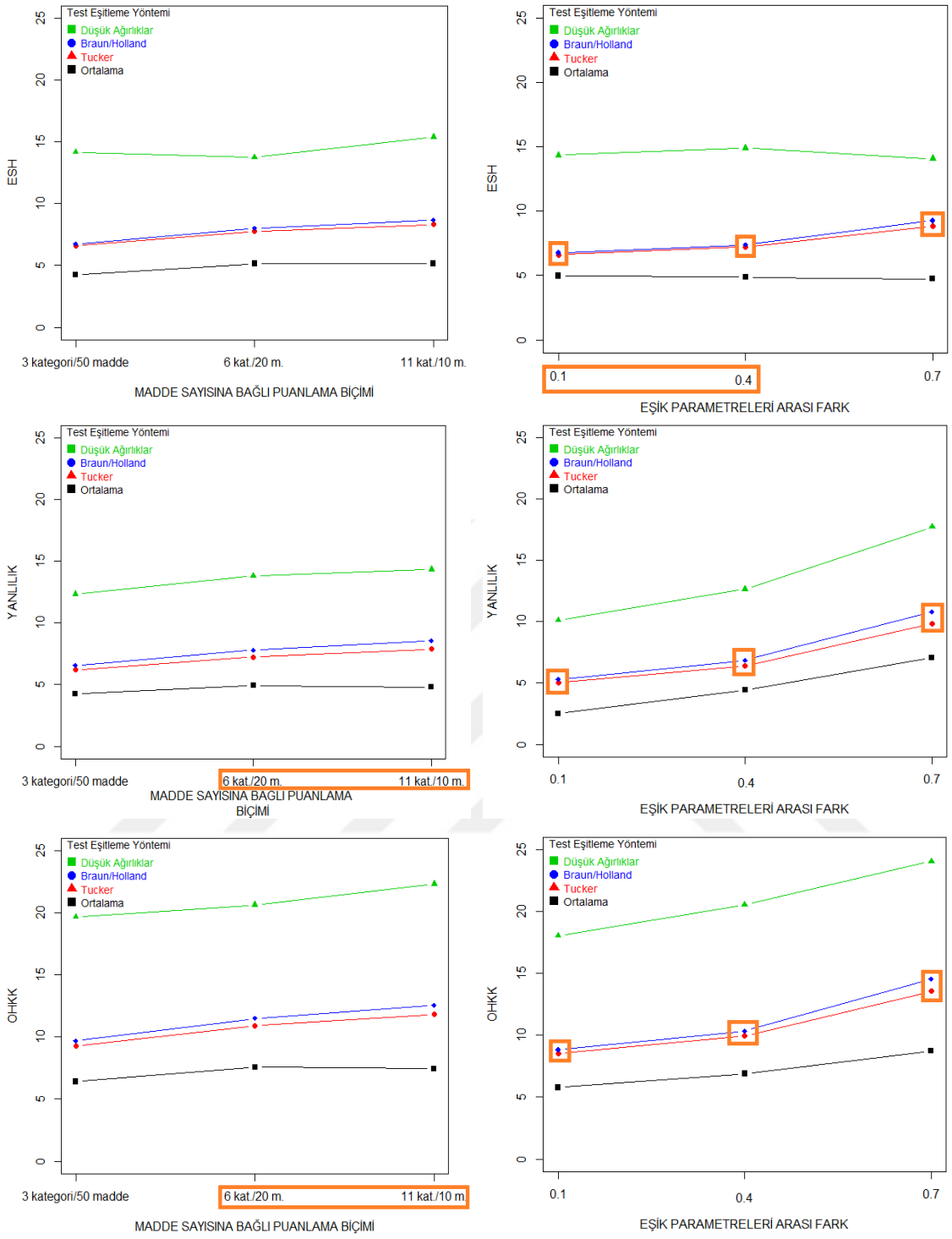
Test eşitleme yöntemi, madde sayısına bağlı puanlama biçimi ve madde eşik parametresi farkı koşullarının OHKK değerleri üzerindeki etkileri

Varyansın kaynağı	Kareler Toplamı	sd	Kareler Ortalaması	F	p	η^2
Yöntem	5617	3	1872.4	53.49	.00	.47
Madde sayısına bağlı puanlama biçimi (MSBP)	189	2	94.3	2.69	.07	.03
Madde Eşik Parametre Farkı (MEPF)	905	2	452.3	12.93	.00	.13
Y*MSBP	25	6	4.2	.12	.99	.00
Y*MEPF	60	6	9.9	.28	.94	.01
MSBP*MEPF	48	4	12.0	.34	.85	.01
Y*MSBP*MEPF	37	12	3.1	.09	1.00	.01
Hata	6300	180	35.0			

Tablo 22'de yer alan tüm etkilerin dikkate alındığı ANOVA modeline ait sonuçlar incelendiğinde, test eşitleme yöntemlerinden elde edilen ortalama OHKK değerleri arasında yüksek bir etki ile istatiksel olarak manidar bir fark olduğu görülmektedir ($F_{(3, 180)}$ =53.49, p <.05, η^2=.47). Ancak madde sayısına bağlı puanlama biçimi koşulları arasında istatiksel olarak manidar bir farklılığın olmadığı görülmektedir ($F_{(2, 180)}$ =2.69, p>.05, η^2=.03). Bunun yanı sıra, madde eşik parametre farkı koşulları arasında orta düzey bir etki ile istatiksel olarak manidar bir farklılık bulunmaktadır ($F_{(2, 180)}$ =12.93, p>.05, η^2=.13). Test eşitleme yöntemi ve madde sayısına bağlı puanlama biçimi koşullarının birlikte yaptığı ortak etkinin de istatiksel olarak anlamsız olduğu görülmektedir ($F_{(6, 180)}$ =.12, p>.05, η^2=.00). Benzer bir şekilde, test eşitleme yöntemi ve madde eşik parametre farkı koşullarının ortak etkisinin de istatiksel olarak manidar olmadığı görülmektedir ($F_{(6, 180)}$ =.28, p>.05, η^2=.01). Madde sayısına bağlı puanlama biçimi ve madde eşik parametresi farkı koşullarının ortak etkisinin de manidar olmadığı görülmektedir ($F_{(4, 180)}$ =.34, p>.05, η^2=.01). Son olarak, her üç koşul birden yaptığı ortak etkinin de istatiksel olarak manidar olmadığı dikkat çekmektedir ($F_{(12, 180)}$ =.09, p>.05, η^2=.01). Tablo 22'de görülen tek manidar etkinin test eşitleme yöntemleri arasında olduğu dikkat çekmektedir. Bu etkinin yönünü belirlemek için gerekli devam analizleri "Örneklem Büyüklüğü ve

Madde Sayısına Bağlı Puanlama Biçimi Değişimlerinin ESH, Yanlılık ve OHKK Değerlerine Ortak Etkisi" başlığı altında incelenmiştir.

Madde sayısına bağlı puanlama biçimi koşuluna göre ve madde eşik parametre farkı koşuluna göre çalışma kapsamında elde edilen ESH, yanlılık ve OHKK değerlerine yönelik grafikler Şekil 10'da verilmiştir. Şekil 10'da, Tukey HSD testi sonuçlarına göre aralarında manidar farklılık bulunmayan koşullar varsa, aynı kutuların içerisinde gösterilmiştir. Şekil 10 incelendiğinde, her iki koşulda da ESH, yanlılık ve OHKK değerlerinin paralel bir şekilde değiştiği görülmektedir. Bunun yanı sıra, Tucker ve Braun-Holland doğrusal test eşitleme yöntemlerinin çakışık oldukları da görülmektedir.



Şekil 10. Madde Sayısına Bağlı Puanlama Biçimi ve Madde Eşik Parametre Farkı Koşullarına göre Test Eşitleme Yöntemlerinden Elde Edilen ortalama ESH, Yanlılık ve OHKK değerleri

BÖLÜM 5

SONUÇ VE ÖNERİLER

Sonuç

Bu bölümde; çalışma kapsamında elde edilen bulgular ilgili literatür ile karşılaştırılmış ve yorumlanmıştır.

Örneklem büyüklüğü değişimlerinin ESH, yanlılık ve OHKK değerlerine etkisi

Çalışma bulguları ESH açısından incelendiğinde; 100'den daha küçük örneklem büyüklükleri için en kullanışlı yöntemin ortalamalar ile test eşitleme olduğu ve 100 ve üzeri örneklem büyüklüklerinde çalışma kapsamında incelenen test eşitleme yöntemlerinin manidar bir farklılık oluşturmadığı görülmektedir.

Yanlılık açısından incelendiğinde; 20 ve üzeri örneklemlerde, ortalamalar, Tucker doğrusal ve Braun-Holland doğrusal test eşitleme yöntemlerinin manidar bir farklılık oluşturmadığı; ancak düşük ağırlıklar doğrusal test eşitleme yönteminden her örneklem büyüklüğü koşulunda diğerlerinden daha yüksek yanlılık değerleri elde edildiği görülmektedir.

OHKK açısından incelendiğinde ise, 30 ve üzeri örneklemlerde yine ortalamalar, Tucker doğrusal ve Braun-Holland doğrusal test eşitleme yöntemlerinin manidar bir farklılık oluşturmadığı; ancak düşük ağırlıklar doğrusal test eşitleme yönteminden her örneklem büyüklüğü koşulunda diğerlerinden daha yüksek OHKK elde edildiği görülmektedir.

Bu bulgular doğrultusunda; çalışma kapsamında ele alınan test eşitleme yöntemleri içinde en düşük ESH, yanlılık ve OHKK değerleri ortalamalar ile test eşitleme yöntemiyle elde edilmiştir. Bu bulgu, bu yöntemin literatürde küçük örneklemler için en çok önerilen yöntem olması ile uyuşmaktadır (Skaggs, 2005; Livingston ve Kim, 2009). Ayrıca, düşük ağırlıklar doğrusal test eşitleme yönteminin en yüksek ESH, yanlılık ve

OHKK değerlerini verdiği de bulunmuştur. Babcock ve diğerleri (2012) de benzer bir şekilde, bu yöntemin 50 ve daha küçük örneklem koşullarında Tucker doğrusal ve ortalamalar ile test eşitleme yönteminden daha yüksek OHKK ve yanlılık değerleri verdiğini belirtmişlerdir. Ancak bu çalışmada incelenen diğer yöntemlerle düşük ağırlıklar test eşitleme yönteminden elde edilen ESH, yanlılık ve OHKK değerleri arasındaki farkın çok yüksek olmasının çalışma kapsamında çok kategorili puanlanan testlerin simüle edilmesinden kaynaklandığı düşünülmektedir.

Madde sayısına bağlı puanlama biçimi değişimlerinin ESH, Yanlılık ve OHKK değerlerine etkisi

Çalışma kapsamında; her birinden alınabilecek maksimum puan 100 olacak şekilde ayarlanmış 3 farklı madde sayısına bağlı puanlama biçimi koşulu ele alınmıştır. 3 kategorili puanlanan 50 maddelik testlerin eşitlenmesinde tüm yöntemlerin en düşük ESH, yanlılık ve OHKK değerlerini verdiği görülmektedir. Madde sayısına bağlı puanlama biçimi koşullarının diğer iki düzeyinde üretilen ESH, yanlılık ve OHKK değerlerinin düzensiz dağıldığı dikkat çekmektedir. Ancak bu koşuldan elde edilen tüm ESH, yanlılık ve OHKK değerleri değişimlerinin istatiksel olarak manidar düzeyde farklılaşmadığı görülmektedir. Başka bir ifade ile; eşitlenmek istenen testlerin 50 maddeden oluşan ve 3 kategorili puanlanan (0-2) testler olması, 20 maddeden oluşan ve 6 kategorili puanlanan (0-5) testler olması ya da 10 maddeden oluşan ve 11 kategorili puanlanan (0-10) testler olması koşulları arasında istatiksel olarak manidar bir farklılık yoktur.

Madde sayısına bağlı puanlama biçimi koşulu ile ilgili elde edilen bulgular, bu koşulun çok kategorili puanlanan testlerin test eşitleme çalışmalarında bir önem arz etmediğini göstermektedir. Bu durumun madde sayısı değişimlerinden çok testten elde edilebilecek maksimum puanla alakalı olduğu düşünülmektedir. Başka bir ifade ile; çok kategorili puanlanan bir testten elde edilebilecek maksimum puan eşit olduğu sürece, o testte kaç madde olduğu test eşitleme açısından bir önem arz etmemekte, tüm koşullardan benzer ESH, yanlılık ve OHKK değerleri elde edilmektedir.

Madde eşik parametresi farkı değişimlerinin ESH, Yanlılık ve OHKK değerlerine etkisi

Madde eşik parametreleri arasındaki farklar açısından incelendiğinde, test eşitleme yöntemlerinden elde edilen ESH, yanlılık ve OHKK değerlerinin farklı eğilimler gösterdiği görülmektedir. ESH açısından incelendiğinde; en düşük ortalama ESH düzeyine sahip olan ortalamalar ile test eşitleme yönteminden madde eşik parametresi farkı, yani maddeler arasındaki güçlük farkı, arttıkça daha düşük ESH elde edildiği görülmektedir. Yanlılık açısından incelendiğinde, madde eşik parametresi farkı koşulu arttıkça tüm yöntemlerden elde edilen yanlılık değerlerinin de manidar düzeyde arttığı görülmektedir. OHKK değerleri açısından incelendiğinde, madde eşik parametresi farkı koşulu arttıkça tüm yöntemlerden elde edilen OHKK değerlerinin de arttığı görülmektedir.

Çalışma kapsamında elde edilen tüm bulgular doğrultusunda; çok kategorili puanlanan testlerin eşitlenmesinde, test eşitleme yöntemi ve örneklem büyüklüğü koşullarının madde sayısına bağlı puanlama biçimi koşulundan ve madde eşik parametresi farkı koşulundan daha fazla etkiye sahip olduğu ve daha fazla önem arz ettiği düşünülmektedir. Çalışma kapsamında incelenen tüm koşullar birlikte ele alındığında, önerilen test eşitleme yöntemleri Tablo 23'de gösterilmektedir. Tablo incelendiğinde, her koşul için ortalamalar ile test eşitleme yönteminin kullanılabileceği görülmektedir. Diğer yöntemler ise belirli koşullar altında ortalamalar ile test eşitleme yöntemine benzer sonuçlar vermektedir. Ancak çalışma kapsamında incelenen hiçbir koşul altında ortalamalar ile test eşitleme yönteminden daha düşük ESH, yanlılık ve OHKK değerleri elde edilen bir test eşitleme yöntemi bulunmamaktadır. Otuz ve üzeri örneklem büyüklüklerinde tüm madde sayısına bağlı puanlama biçimi koşullarında ve madde eşik parametresi farkı koşullarında Tucker doğrusal test eşitleme yöntemi önerilmektedir. Atmış ve üzeri örneklem büyüklüklerinde ise, Braun-Holland doğrusal test eşitleme yöntemi kullanılabilmektedir. Hiçbir koşulda, düşük ağırlıklar doğrusal test eşitleme yöntemi önerilmemektedir.

Tablo 23.

Araştırma kapsamında incelenen koşullar için önerilen test eşitleme yöntemleri

		MEPF		
MSBP	N	*0.1*	*0.4*	*0.7*
3 kategorili 50 madde	*15*	O	O	O
	20	O	O	O
	30	O, T	O, T	O, T
	60	O, T, BH	O, T, BH	O, T, BH
	100	O, T, BH	O, T, BH	O, T, BH
	150	O, T, BH	O, T, BH	O, T, BH
6 kategorili 20 madde	*15*	O	O	O
	20	O	O	O
	30	O, T	O, T	O, T
	60	O, T, BH	O, T, BH	O, T, BH
	100	O, T, BH	O, T, BH	O, T, BH
	150	O, T, BH	O, T, BH	O, T, BH
11 kategorili 10 madde	*15*	O	O	O
	20	O	O	O
	30	O, T	O, T	O, T
	60	O, T, BH	O, T, BH	O, T, BH
	100	O, T, BH	O, T, BH	O, T, BH
	150	O, T, BH	O, T, BH	O, T, BH

O: Ortalamalar ile test eşitleme
T: Tucker doğrusal test eşitleme
BH: Braun-Holland doğrusal test eşitleme
MSBP: Madde sayısına bağlı puanlama biçimi
MEPF: Madde eşik parametresi farkı

Öneriler

Uygulayıcılara öneriler

Bu çalışmanın günlük eğitim durumlarında öğrenciler hakkında daha adil yargılarda bulunmak amacı ile test eşitleme uygulamak isteyen eğitimcilere sunacağı öneriler farklı koşullar için maddeler halinde sıralanmıştır. Bu öneriler, tüm testin %20'si oranında ortak madde ya da bağıntı testi kullanılması durumlarında geçerlidir. Ayrıca, testten elde edilebilecek maksimum puanın 100 olduğu üç, altı veya 11 kategorili puanlanan testlerde geçerlidir.

1. Bir formun 15'ten daha az öğrenciye uygulanacağı durumlarda test eşitleme yapılmaması önerilmektedir. Ancak, her formun 15 ve üzeri sayıda öğrenciye uygulanması durumlarında, eşitleme çalışmalarının ortalamalar ile test eşitleme yöntemi ile yapılması önerilmektedir.

2. Her bir formu alan öğrenci sayısı 30 ile 60 aralığında ise, ortalamalar ile test eşitleme yönteminden elde edilen sonuçları karşılaştırmak ve desteklemek amacıyla Tucker doğrusal test eşitleme yönteminin kullanılması da önerilmektedir.
3. Test uygulamasının 60 ile 150 arası sayıda öğrenciye uygulanacağı durumlarda ortalamalar ile test eşitleme yönteminden elde edilen sonuçları karşılaştırmak ve desteklemek amacıyla hem Tucker hem de Braun-Holland doğrusal test eşitleme yöntemlerinin kullanılması önerilmektedir.
4. Testten elde edilebilecek maksimum puan 100 olduğu sürece hangi kategori koşulunun kullanıldığı ya da formlar arasında ne düzeyde güçlük farkı olduğu fark etmeksizin ortalamalar ile test eşitleme yöntemi önerilmektedir.

Araştırmacılara öneriler

1. Bu çalışma bilgisayar ortamında üretilmiş olan simülatif veriler ile yürütülmüştür. Aynı çalışmanın gerçek bir test uygulamasından elde edilen veriler ile yapılması da mümkündür.
2. Çalışmada, ortalamalar ile test eşitleme yöntemi, Tucker, Braun-Holland ve düşük ağırlıklar doğrusal test eşitleme yöntemleri incelenmiştir. Bunların dışında kalan klasik test eşitleme yöntemleri ya da MTK temelli test eşitleme yöntemleri de farklı koşullar altında tercih edilebilir.
3. Çalışma kapsamında 15 ile 150 arasında değişen örneklem büyüklüğü koşulları incelenmiştir. Daha büyük örneklem koşulları da incelenerek araştırma genişletilebilir.
4. Bu çalışmada, simüle edilen veriler GPCM modeli ile üretilmiştir. Farklı bir model ile de üretilebilir.
5. Bu çalışmada, bağıntı testi madde sayısı toplam testin %20'sini oluşturacak şekilde sabit tutulmuştur. Bu koşulun da değişimlenerek incelenmesi mümkündür.
6. Bu çalışmada, bir maddeye ait puanlama kategorileri arasındaki eşik parametreleri, madde ayırt edicilik ve şans parametreleri sabit tutulmuştur. Bu koşullar değişimlenerek bir çalışma yürütülebilir.
7. Bu çalışmada, daire-yay test eşitleme yöntemi kriter test eşitleme yöntemi olarak tercih edilmiştir. Farklı koşullar altında, başka bir test eşitleme yöntemi tercih edilebilir.

8. Çalışma sonuçları genel olarak incelendiğinde ortalamalar ile test eşitleme yönteminin en düşük hata puanları üreten yöntem olduğu görülmektedir. Bu nedenle daire-yay test eşitleme yöntemi ve ortalamalar ile test eşitleme yönteminin karşılaştırıldığı bir çalışma yürütülebilir.
9. Bu çalışmada incelenen faktörler dışında kalan geçme/kalma puanı ve bağıntı testinin toplam testi temsil etme düzeyi gibi farklı koşullar da araştırılabilir.

KAYNAKLAR

Angoff, W. H. (1953). *Equating of the ACE psychological examinations for high school students* (Research Bulletin No. RB-53-03). Princeton: Educational Testing Service. http://dx.doi.org/10.1002/j.2333-8504.1953.tb00887.x

Angoff, W. H. (1982). Summary and derivation of equating methods used at ETS. P. W. Holland ve D. B. Rubin (Eds.), *Test equating.* (pp. 55–70). New York: Academic Press.

Angoff, W. H. (1984). *Scales, norms, and equivalent scores*. Princeton, NJ: Educational Testing Service.

Albano, A. D. (2015). A general linear method for equating with small samples. *Journal of Educational Measurement*, 52 (1), 55-69.

Albano, A. D. (2016). Equate: An R package for observed-score linking and equating. *Journal of Statistical Software*, 74 (8), 1-36. doi:10.18637/jss.v074.i08

Aşiret, S. (2014). Küçük Örneklemlerde Test Eşitleme Yöntemlerinin Çeşitli Faktörlere Göre İncelenmesi (Yayınlanmamış Yüksek Lisans Tezi). Mersin Üniversitesi, Mersin, Türkiye.

Aşiret, S. ve Sünbül, S. Ö. (2014). Investigating test equating methods in small samples through various factors. *Educational Sciences: Theory & Practice,* 16, 647-668.

Babcock, B., Albano, A. D., & Raymond, M. (2012). Nominal weights mean equating: A method for very small samples. *Educational and Psychological Measurement*, 72, 608– 628.

Baker, F. G. (1992). Equating tests under the grade response model. *Applied Psychological Measurement,* 16, 87-96.

Baker F.B., & Kim S.H. (2004). *Item response theory: Parameter estimation techniques*. Marcel Dekker: New York.

Bandalos, D. L. (2018). *Measurement theory and applications for the social sciences*. Newyork: Guilford Press.

Birnbaum A (1968). Some latent trait models and their use in ınferring any examinee's ability. In FM Lord, MR Novick (eds.), *Statistical theories of mental test scores*, (pp. 395–479). Adison-Wesley.

Bozdağ, S. ve Kan, A. (2010). Şans Başarısının Test Eşitlemeye Etkisi. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi, 39,* 91-108.

Braun, H. I., & Holland, P. W. (1982). Observed score test equating: A mathematical analysis of some ets equating practices. P. W. Holland ve D. B. Rubin (Eds.), *Test equating.* (pp. 9-50). New York: Academic Press.

Budescu, D. (1985). Efficiency of linear equating as a function of the lenght of the anchor test. *Journal of Educational Statistics*, 12 (1), 33-43.

Cohen J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). NY: Lawrence Erlbaum Associates.

Cohen, L., Manion, L., & Morrison, K. (2007*) Research Methods in Education* (6th ed.). London: Routledge Falmer.

Cook, L. L., & Eignor, D. R. (1989). Using item response theory in test score equating. *International Journal of Educational Reserarch*, 13 (2), 161-173.

Crocker, L. & Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. New York: Holt, Rinehart and Winston.

Demir, E. (2019). *R Dili ile İstatistik Uygulamaları* (1. Baskı). Ankara: Pegem Akademi.

Darrell Bock, R. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37(1), 29–51.

Dorans, N.J. (1990). Equating methods and sampling designs. *Applied Measurement in Education*, 3(1), 3–17.

Dorans, N.J., & Holland, P.W. (2000). Population invariance and equitability of tests: Basic theory and the linear case. *Journal of Educational Measurement*, 37, 281–306.

Dorans, N.J., Moses, T.P., & Eignor, D.R. (2010*) Principles and practices of test score equating*. Princeton, NJ: Educational Testing Service.

Dorans, N.J., Moses, T.P., & Eignor, D.R. (2011). Equating test scores: Towards best practices. A. von Davier (ed.), *Statistical models for test equating, scaling and linking*. (pp. 43-58) New York: Springer.

Dorans, N.J., & Puhan, G. (2017). Contributions to score linking theory and practice. R. Bennett ve M. von Davier (Eds.), *Advancing human assessment:Methodological, psychological, and policy contributions* (pp. 79-133). New York, NY: Springer.

Felan, G. D. (2002). T*est equating: mean, linear, equipercentile, and ıtem response theory*. Paper presented in annual meeting of Southwest Educational Research Association. Austin, TX. (ERIC)

Feuer, M. J., Holland, P. W., Green, B. F., Bertenthal, M. W., & Hemphill, F. C. (1999). *Uncommon measures: Equivalence and linkage among educational tests* (Report of the Committee on Equivalency and Linkage of Educational Tests, National Research Council). Washington, DC: National Academy Press.

Fraenkel, J. R., Wallen, N. E., & Hyun, H. H. (2015). *How to design and evaluate research in education* (9th ed.). New York, NY: McGraw-Hill Education.

Gonzales, J. (2014). SNSequate: Standard and nonstandard statistical models and methods for test equating. *Journal of Statistical Software,* 59 (7), 1-30.

Gonzales, J., & Wiberg, M. (2017). *Applying test equating methods using R*. New York, NY: Springer

Hambleton, R. K, & Swaminathan, H. (1985*). Item response theory: Principles and applications*. Boston, MA: KluwerNijhoff.

Hambleton, R. K, Swaminathan, H., & Rogers, H.J. (1991). *Fundamentals of item response theory*. Newbury Park, CA:Sage.

Hambleton, R. K., & Jones, R. W. (1993). An NCME instructional module on comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice,* 12, 38-47.

Hambleton, R. K., & Cook, L. L. (1977). Latent trait models and their use in the analysis of educational test data. *Journal of Educational Measurement,* 14, 75-99.

Hanson, B. A., Zeng, L., & Colton, D.(1994). *A comparison of pre-smoothing and post-smoothing methods in equipercentile equating.* (ACT Research Report 94 – 4). Iowa City, IA : American College Testing, Inc.

Holland, P. W., & Thayer, D. T. (1989). The kernel method of equating score distributions. *ETS Research Report No. RR - 89 - 07*. Princeton, New Jersey: Educational Testing Service.

Holland, P. W. (2007). A framework and history for score linking. N. J. Dorans, M. Pommerich, ve P. W. Holland (Eds.), *Linking and aligning scores and scales* (pp. 5-30). New York, NY: Springer-Verlag.

İnal, H. ve Arıkan, Ç. A. (2017). Test eşitlemede grup değişmezliğinin cinsiyete göre incelenmesi. *Eğitimde ve Psikolojide Ölçme ve Değerlendirme Dergisi, 8 (1),* 128-145.

Kahraman, H. (2012). Düzgünleştirilmiş Puanların Eşitleme Hatasına Etkisi. (Yayınlanmamış Yüksek Lisans tezi). *Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü, Ankara.* https://tez.yok.gov.tr/UlusalTezMerkezi/ internet adresinden alınmıştır.

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 İstatiksel Eşitliğinin Sınanması. *Eğitim ve Bilim.* 36(160), s. 38-51

Karasar, N. (2009). *Bilimsel Araştırma Yöntemi* (19. Baskı). Ankara: Nobel Yayınları.

Kelecioğlu, H. (1994). Öğrenci Seçme Sınavı Puanlarının Eşitlenmesi Üzerine Bir Çalışma. (Yayımlanmış doktora tezi). *Hacettepe Üniversitesi, Sosyal Bilimler Enstitüsü,* Ankara, Türkiye.

Kelecioğlu, H. ve Öztürk-Gübeş, N. (2013). Random grup deseni ile yapılan doğrusal eşitleme ve eşit yüzdelikli eşitleme yöntemlerinin karşılaştırılması. *International Online Journal of Educational Sciences, 5 (1),* 227-241.

Kilmen, S. (2010). Madde Tepki Kuramına Dayalı Test Eşitleme Yöntemlerinden Kestirilen Eşitleme Hatalarının Örneklem Büyüklüğü ve Yetenek Dağılımına Göre Karşılaştırılması. (Yayınlanmış Doktora Tezi). *Ankara Üniversitesi, Eğitim Bilimleri Enstitüsü, Ankara.* https://tez.yok.gov.tr/UlusalTezMerkezi/ internet adresinden alınmıştır.

Kim, S., von Davier, A. A., & Haberman, S. (2008). Small-sample equating using synthetic linking function. *Journal of Educational Measurement*, 45, 325–342.

Klein, L. W., & Jarjoura, D. (1985). The importance of content representation for common-item equating with nonrandom groups. *Journal of Educational Measurement.* 22 (3), 197-206.

Kolen, M. J. (1985). Standard errors of Tucker equating. *Applied Psychological Measure- merit*, 9, 209-223.

Kolen, M. J. (1988). Traditional equating methodology. *Educational Measurement: Issues and Practice,* 7 (4), 29-36.

Kolen, M. J., & Brennan, R. L. (1995). *Test equating: Methods and practices*. New York, NY: Springer-Verlag

Kolen, M. J., & Brennan, R. L. (2004). *Test equating, scaling and linking: Methods and practices*. New York, NY: Springer-Verlag

Kolen M.J. (2007). Data Collection Designs and Linking. N. J. Dorans, M. Pommerich, & P. W. Holland (Ed.), *Linking and aligning scores and scales* (pp. 5-30). New York, NY: Springer-Verlag.

Kolen, M. J., Tong, Y., Brennan, R. L. (2011). Scoring and Scaling Educational Tests. In A. von Davier (ed.), *Statistical Models for Test Equating, Scaling and Linking*. (pp. 43-58) New York: Springer.

Linn, R. L. (1993) *Linking results of distinct assessments, Applied Measurement in Education,* 6 (1), 83-102, doi: 10.1207/s15324818ame0601_5

Lee, G., Kolen, M. J., Frisbie, D. A., & Ankenmann, R. D. (2001). Comparison of dichotomous and polytomous ıtem response models in equating scores from tests composed of testlets. *Applied Psychological Measurement*, 25(4), 357–372

Levine, R. S. (1955). *Equating the score scales of alternate forms administered to samples of different ability* (ETS Research Bulletin No. RB-55-23). Princeton, NJ: Educational Testing Service.

Li, Y. H., Lissitz, R. W., & Yang, Y. N. (1999, April). *Estimating ırt equating coefficients for tests with polytomously and dichotomously scored ıtems*. Paper Presented at the Annual Meeting of the National Council on Measurement in Education (NCME). Montreal, Quebec, Canada.

Livingston, S. A. (1993). Small-sample equating with log-linear smoothing. *Journal of Educational Measurement,* 30, 23–29.

Livingston, S. A., & Kim, S. (2009). The circle-arc method for equating in small samples. *Journal of Educational Measurement,* 46, 330–343.

Livingston, S. A. (2014). *Equating test scores (without IRT)* (2nd ed.). Princeton, NJ: Educational Testing Service.

Lord, F. M. (1950). *Notes on comparable scales for test scores* (ETS Research Bulletin No. RB-50-48). Princeton, NJ: Educational Testing Service.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems. Hillsdale*, NJ: Lawrence Erlbaum Associates.

Lord, F. M. (1982). The standard error of equipercentile equating. *Journal of Education Statistics*, 7, 165–174.

Mao, X., von Davier, A. A., & Rupp, S. (2006). Comparisons of the kernel equating method with the traditional equating methods on PRAXIS data *(ETS Research Report No. RR-06-30).* Princeton, NJ: Educational Testing Service.

Mislevy, R. J. (1992). *Linking educational assessments: Concepts, issues, methods, and prospects (Policy Information Rep.)*. Princeton, NJ: Educational Testing Service.

Mislevy, R.J. (2018). *Sociocognitive foundations of educational measurement*. New York, NY: Routhledge.

Muraki, E. (1992). A generalized partial credit model: application of an em algorithm. *Applied Psychological Measurement,* 16(2), 159–176.

Muraki, E. (1993). Information functions of the generalized partial credit model. *Applied Psychological Measurement,* 17(4), 351–363.

Parshall, C.G., Houghton, P. D. B., & Kromrey, J. D. (1995). Equating error and statistical bias in small sample linear equating. *Journal of Educational Measurement*, 32, 37–54.

Petersen, N. S., Kolen, M. J., &Hoover, H. D. (1989). Scaling, norming, and equating. R. L. Linn (Ed.), *Educational measurement* (pp. 221-262). New York: Macmillan

R Core Team (2019). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/.

Ricker, K. L., & von Davier, A. A. (2007). *Impact of anchor test length on equating results in a nonequivalent groups design.* Princeton, NJ: Educational Testing Service.

Rizopoulos, D. (2006). Ltm: An R package for latent varibale modeling and ıtem response theory analyses. *Journal of Statistical Software.* 17(5), 1-24. URL http://www.jstatsoft.org/v17/i05/

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.

Sansivieri, V., & Wiberg, M. (2017). Item response theory equating with the non-equivalent groups with covariates design. L. A. van der Ark, M. Wiberg, S. A. Culpepper, J. A. Douglas, W. Wang (eds.), *Quantitative Psychology*. (pp. 275-285). New York, NY: Springer.

Sinharay, S., & Holland, P. (2006). *Choice of anchor test in equating.* Princeton, NJ: Educational Testing Service.

Skaggs, G., & Lissitz, R. W. (1986). IRT test equating: Relevant issues and a review of recent research. *Review of Educational Research.* 56 (4), 495-529.

Skaggs, G. (2005). Accuracy of random groups equating with very small samples. *Journal of Educational Measurement,* 42, 309–330

Tate, R. L. (1999). A cautionary note on irt-based linking of tests with polytomous ıtems. *Journal of Educational Measurement,* 36(4), 336–346. doi:10.1111/j.1745-3984.1999.tb00560.x

Von Davier, A., Holland, P.W., & Thayer, D.T. (2004). *The kernel method of test equating.* New York, NY: Springer-Verlag

Von Davier, A.(2007). Potential solutions to practical equating ıssues. N. J. Dorans, M. Pommerich, & P. W. Holland (Ed.), *Linking and aligning scores and scales* (pp. 89-105). New York, NY: Springer-Verlag.

Wolf, R. (2013). Assessing the impact of characteristics of the test, common-items, and examinees on the preservation of equity properties in mixed-format test equating (Unpublished doctoral dissertation). University of Pittsburgh, Pittsburgh, the USA.

Zhu, W. (1998). Test equating: what, why, how?, *Research Quarterly for Exercise and Sport*, 69 (1), 11-23.

EKLER

EK 1: Varsayımların Kontrolü ile İlgili Test Sonuçları

EK 1.A: Test eşitleme yöntemlerinden elde edilen ortalama ESH değerlerinin örneklem büyüklüğü ve madde sayısına bağlı puanlama biçimi koşullarına göre normallik ve varyansların homojenliği testi sonuçları

		Shapiro-Wilk's			Fligner-Killeen		
Yöntem	Koşul	W-test	sd	p	Med. X^2	sd	p
Ortalamalar ile test eşitleme	15	0.76	9	.04	57.91	5	3.29
	20	0.83	9	.08			
	30	0.85	9	.09			
	60	0.85	9	.08			
	100	0.76	9	.06			
	150	0.82	9	.03			
Tucker doğrusal test eşitleme	15	0.85	9	.09	10.55	5	.06
	20	0.91	9	.32			
	30	0.91	9	.34			
	60	0.92	9	.40			
	100	0.93	9	.51			
	150	0.90	9	.31			
Braun / Holland doğrusal test eşitleme	15	0.84	9	.05	10.92	5	.05
	20	0.91	9	.29			
	30	0.91	9	.28			
	60	0.92	9	.37			
	100	0.92	9	.38			
	150	0.92	9	.38			
Düşük ağırlıklar doğrusal test eşitleme	15	0.97	9	.89	22.83	5	.05
	20	0.94	9	.58			
	30	0.75	9	.06			
	60	0.85	9	.06			
	100	0.94	9	.61			
	150	0.91	9	.29			
Ortalamalar ile test eşitleme	3 kategorili 50 madde	0.89	18	.04	2.51	2	.29
	6 kategorili 20 madde	0.90	18	.06			
	11 kategorili 10 madde	0.90	18	.06			
Tucker doğrusal test eşitleme	3 kategorili 50 madde	0.92	18	.12	1.33	2	.51
	6 kategorili 20 madde	0.92	18	.13			
	11 kategorili 10 madde	0.92	18	.12			
Braun / Holland test eşitleme	3 kategorili 50 madde	0.92	18	.15	1.45	2	.48
	6 kategorili 20 madde	0.93	18	.16			
	11 kategorili 10 madde	0.92	18	.11			
Düşük ağırlıklar test eşitleme	3 kategorili 50 madde	0.88	18	.03	1.02	2	.60
	6 kategorili 20 madde	0.88	18	.03			
	11 kategorili 10 madde	0.86	18	.01			

Ek 1.B: Test eşitleme yöntemlerinden elde edilen ortalama yanlılık değerlerinin örneklem büyüklüğü ve madde sayısına bağlı puanlama biçimi koşullarına göre normallik ve varyansların homojenliği testi sonuçları

		Shapiro-Wilk's			Fligner-Killeen		
Yöntem	Koşul	W-test	sd	p	Med. X^2	sd	p
Ortalamalar ile test eşitleme	15	0.90	9	.26	3.93	5	0.56
	20	0.86	9	.11			
	30	0.91	9	.32			
	60	0.88	9	.17			
	100	0.91	9	.32			
	150	0.90	9	.26			
Tucker doğrusal test eşitleme	15	0.81	9	.03	0.42	5	0.99
	20	0.85	9	.06			
	30	0.89	9	.19			
	60	0.92	9	.37			
	100	0.90	9	.27			
	150	0.90	9	.23			
Braun / Holland doğrusal test eşitleme	15	0.84	9	.06	1.70	5	0.89
	20	0.86	9	.11			
	30	0.90	9	.28			
	60	0.91	9	.31			
	100	0.92	9	.35			
	150	0.89	9	.22			
Düşük ağırlıklar doğrusal test eşitleme	15	0.90	9	.23	5.39	5	0.37
	20	0.85	9	.08			
	30	0.94	9	.53			
	60	0.90	9	.27			
	100	0.90	9	.25			
	150	0.90	9	.23			
Ortalamalar ile test eşitleme	3 kategorili 50 madde	0.89	18	.05	0.71	2	.70
	6 kategorili 20 madde	0.89	18	.05			
	11 kategorili 10 madde	0.88	18	.03			
Tucker doğrusal test eşitleme	3 kategorili 50 madde	0.97	18	.77	1.73	2	.42
	6 kategorili 20 madde	0.95	18	.46			
	11 kategorili 10 madde	0.92	18	.13			
Braun / Holland test eşitleme	3 kategorili 50 madde	0.96	18	.68	1.91	2	.39
	6 kategorili 20 madde	0.93	18	.20			
	11 kategorili 10 madde	0.92	18	.10			
Düşük ağırlıklar test eşitleme	3 kategorili 50 madde	0.96	18	.55	0.81	2	.66
	6 kategorili 20 madde	0.94	18	.33			
	11 kategorili 10 madde	0.97	18	.79			

Ek 1.C: Test eşitleme yöntemlerinden elde edilen ortalama OHKK değerlerinin örneklem büyüklüğü ve madde sayısına bağlı puanlama biçimi koşullarına göre ANOVA varsayımları

		Shapiro-Wilk's			Fligner-Killeen		
Yöntem	Koşul	W-test	sd	p	Med. X^2	sd	p
Ortalamalar ile test eşitleme	15	0.97	9	.91	7.69	5	.17
	20	.95	9	.71			
	30	.95	9	.72			
	60	.92	9	.39			
	100	.93	9	.53			
	150	.93	9	.48			
Tucker doğrusal test eşitleme	15	.84	9	.06	2.31	5	.80
	20	.88	9	.14			
	30	.89	9	.21			
	60	.92	9	.39			
	100	.89	9	.22			
	150	.89	9	.19			
Braun / Holland doğrusal test eşitleme	15	.83	9	.05	3.61	5	.60
	20	.86	9	.09			
	30	.90	9	.26			
	60	.91	9	.32			
	100	.90	9	.27			
	150	.89	9	.19			
Düşük ağırlıklar doğrusal test eşitleme	15	.93	9	.50	1.27	5	.94
	20	.95	9	.64			
	30	.94	9	.61			
	60	.90	9	.26			
	100	.89	9	.20			
	150	.89	9	.22			
Ortalamalar ile test eşitleme	3 kategorili 50 madde	.97	18	.80	.17	2	.92
	6 kategorili 20 madde	.97	18	.85			
	11 kategorili 10 madde	.97	18	.74			
Tucker doğrusal test eşitleme	3 kategorili 50 madde	.95	18	.42	1.01	2	.61
	6 kategorili 20 madde	.95	18	.36			
	11 kategorili 10 madde	.93	18	.36			
Braun / Holland test eşitleme	3 kategorili 50 madde	.96	18	.60	1.55	2	.46
	6 kategorili 20 madde	.94	18	.26			
	11 kategorili 10 madde	.92	18	.13			
Düşük ağırlıklar test eşitleme	3 kategorili 50 madde	.94	18	.27	.50	2	.78
	6 kategorili 20 madde	.95	18	.48			
	11 kategorili 10 madde	0.97	18	.79			

Ek 1.D: Test eşitleme yöntemlerinden elde edilen ortalama ESH değerlerinin madde eşik parametresi farkı koşuluna göre ANOVA varsayımları

		Shapiro-Wilk's			Fligner-Killeen		
Yöntem	Koşul	W-test	sd	p	Med. X^2	sd	p
Ortalamalar	0.1	.92	18	.12	.23	2	.89
ile test	0.4	.92	18	.15			
eşitleme	0.7	.92	18	.13			
Tucker	0.1	.92	18	.12	2.02	2	.36
doğrusal test	0.4	.91	18	.09			
eşitleme	0.7	.91	18	.08			
Braun /	0.1	.92	18	.12	2.42	2	.29
Holland test	0.4	.92	18	.11			
eşitleme	0.7	.91	18	.10			
Düşük	0.1	.90	18	.04	.23	2	.89
ağırlıklar test	0.4	.87	18	.02			
eşitleme	0.7	.88	18	.03			

Ek 1.E: Test eşitleme yöntemlerinden elde edilen ortalama yanlılık değerlerinin madde eşik parametresi farkı koşuluna göre ANOVA varsayımları

		Shapiro-Wilk's			Fligner-Killeen		
Yöntem	Koşul	W-test	sd	p	Med. X^2	sd	p
Ortalamalar	0.1	.96	18	.67	7.48	2	.03
ile test	0.4	.96	18	.68			
eşitleme	0.7	.89	18	.05			
Tucker	0.1	.92	18	.14	.94	2	.62
doğrusal test	0.4	.92	18	.12			
eşitleme	0.7	.92	18	.11			
Braun /	0.1	.94	18	.26	2.47	2	.29
Holland test	0.4	.93	18	.22			
eşitleme	0.7	.91	18	.10			
Düşük	0.1	.90	18	.06	9.04	2	.02
ağırlıklar test	0.4	.94	18	.31			
eşitleme	0.7	.97	18	.79			

Ek 1.F: Test eşitleme yöntemlerinden elde edilen ortalama OHKK değerlerinin madde eşik parametresi farkı koşuluna göre ANOVA varsayımları

		Shapiro-Wilk's			Fligner-Killeen		
Yöntem	Koşul	W-test	sd	p	Med. X^2	sd	p
Ortalamalar	0.1	.93	18	.18	5.69	2	.06
ile test	0.4	.94	18	.32			
eşitleme	0.7	.96	18	.52			
Tucker	0.1	.92	18	.13	.61	2	.74
doğrusal test	0.4	.92	18	.11			
eşitleme	0.7	.92	18	.11			
Braun /	0.1	.93	18	.16	1.50	2	.47
Holland test	0.4	.92	18	.15			
eşitleme	0.7	.91	18	.10			
Düşük	0.1	.90	18	.06	1.38	2	.50
ağırlıklar test	0.4	.89	18	.03			
eşitleme	0.7	.93	18	.17			

EK 2: Tukey HSD Testi Sonuçları

Ek 2.A: Test eşitleme yöntemlerinin ESH, yanlılık ve OHKK değerleri üzerine etkisini gösteren Tukey HSD testine ait p değerleri.

Hata	Yöntemler	Ortalamalar	Tucker	Braun/Holland	Düşük Ağırlıklar
ESH	*Ortalamalar*				
	Tucker	.00			
	Braun/Holland	.00	.85		
	Düşük Ağırlıklar	.00	.00	.00	
Yanlılık	*Ortalamalar*				
	Tucker	.00			
	Braun/Holland	.00	.72		
	Düşük Ağırlıklar	.00	.00	.00	
OHKK	*Ortalamalar*				
	Tucker	.00			
	Braun/Holland	.00	.70		
	Düşük Ağırlıklar	.00	.00	.00	

Ek 2.B: Örneklem büyüklüğü koşullarının ESH, yanlılık ve OHKK değerleri üzerine etkisini gösteren Tukey HSD testine ait p değerleri.

	Örn. Büy.	15	20	30	60	100	150
ESH	*15*						
	20	.00					
	30	.00	.00				
	60	.00	.00	.00			
	100	.00	.00	.00	.00		
	150	.00	.00	.00	.00	.18	
Yanlılık	*15*						
	20	.29					
	30	.00	.37				
	60	.00	.00	.32			
	100	.00	.00	.02	.83		
	150	.00	.00	.00	.53	.99	
OHKK	*15*						
	20	.00					
	30	.00	.00				
	60	.00	.00	.00			
	100	.00	.00	.00	.15		
	150	.00	.00	.00	.01	.84	

Ek 2.C: Madde sayısına bağlı puanlama biçimi koşullarının ESH, yanlılık ve OHKK değerleri üzerine etkisini gösteren Tukey HSD testine ait p değerleri.

	Puanlama Biçimi	3 kategori-50 madde	6 kategori-20 madde	11 kategori-10 madde
ESH	*3 kategori-50 madde*			
	6 kategori-20 madde	.02		
	11 kategori-10 madde	.00	.02	
Yanlılık	*3 kategori-50 madde*			
	6 kategori-20 madde	.04		
	11 kategori-10 madde	.00	.58	
OHKK	*3 kategori-50 madde*			
	6 kategori-20 madde	.01		
	11 kategori-10 madde	.00	.13	

Ek 2.D-1: Örneklem büyüklüğü ve test eşitleme yöntemlerinin ESH değerleri üzerine yaptığı ortak etkilere ait Tukey HSD testi p değerleri.

Yöntem		Ortalamalar						Tucker						Braun/ Holland						Düşük Ağırlıklar					
	Örneklem	15	20	30	60	100	150	15	20	30	60	100	150	15	20	30	60	100	150	15	20	30	60	100	150
Ortalamalar	15																								
	20	.99																							
	30	.42	.99																						
	60	.00	.05	.89																					
	100	.00	.00	.14	.99																				
	150	.00	.00	.01	.96	.99																			
Tucker	15	.00	.00	.00	.00	.00	.00																		
	20	.01	.00	.00	.00	.00	.00	.05																	
	30	.99	.90	.05	.00	.00	.00	.00	.15																
	60	.48	.99	.99	.85	.12	.01	.00	.00	.07															
	100	.00	.15	.99	.99	.99	.78	.00	.00	.00	.98														
	150	.00	.00	.44	.99	.99	.99	.00	.00	.00	.38	.99													
Braun / Holland	15	.00	.00	.00	.00	.00	.00	.99	.02	.00	.00	.00	.00												
	20	.00	.00	.00	.00	.00	.00	.18	.99	.04	.00	.00	.00	.10											
	30	.99	.65	.01	.00	.00	.00	.00	.37	.99	.02	.00	.00	.00	.12										
	60	.73	.99	.99	.63	.04	.00	.00	.00	.17	.99	.89	.19	.00	.00	.05									
	100	.01	.31	.99	.99	.97	.56	.00	.00	.00	.99	.99	.99	.00	.00	.00	.97								
	150	.00	.01	.60	.99	.99	.99	.00	.00	.00	.53	.99	.99	.00	.00	.00	.30	.99							
Düşük Ağrılıklar	15	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00						
	20	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00					
	30	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00				
	60	.99	.66	.01	.00	.00	.00	.00	.37	.99	.02	.00	.00	.00	.12	.99	.05	.00	.00	.00	.00	.00			
	100	.95	.99	.99	.30	.01	.00	.00	.00	.44	.99	.60	.05	.00	.00	.19	.99	.81	.10	.00	.00	.00	.19		
	150	.06	.77	.99	.99	.67	.16	.00	.00	.00	.99	.99	.94	.00	.00	.00	.99	.99	.98	.00	.00	.00	.00	.99	

Ek 2.D-2: Örneklem büyüklüğü ve test eşitleme yöntemlerinin yanlılık değerleri üzerine yaptığı ortak etkilere ait Tukey HSD testi p değerleri.

Yöntem		Ortalamalar						Tucker						Braun/ Holland						Düşük Ağırlıklar					
	Örneklem	15	20	30	60	100	150	15	20	30	60	100	150	15	20	30	60	100	150	15	20	30	60	100	150
Ortalamalar	15																								
	20	.99																							
	30	.99	.99																						
	60	.99	.99	.99																					
	100	.99	.99	.99	.99																				
	150	.99	.99	.99	.99	.99																			
Tucker	15	.01	.01	.00	.00	.00	.00																		
	20	.48	.47	.25	.28	.18	.17	.99																	
	30	.99	.99	.93	.94	.87	.86	.68	.99																
	60	.99	.99	.99	.99	.99	.99	.13	.96	.99															
	100	.99	.99	.99	.99	.99	.99	.03	.70	.99	.99														
	150	.99	.99	.99	.99	.99	.99	.01	.53	.99	.99	.99													
Braun / Holland	15	.00	.00	.00	.00	.00	.00	.99	.94	.27	.02	.00	.00												
	20	.09	.09	.03	.04	.02	.02	.99	.99	.98	.54	.19	.11	.99											
	30	.84	.84	.62	.65	.51	.48	.96	.99	.99	.99	.95	.88	.65	.99										
	60	.99	.99	.99	.99	.99	.99	.32	.99	.99	.99	.99	.99	.08	.82	.99									
	100	.99	.99	.99	.99	.99	.99	.05	.83	.99	.99	.99	.99	.01	.30	.99	.99								
	150	.99	.99	.99	.99	.99	.99	.02	.64	.99	.99	.99	.99	.00	.16	.93	.99	.99							
Düşük Ağrılıklar	15	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00						
	20	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.99					
	30	.00	.00	.00	.00	.00	.00	.15	.00	.00	.00	.00	.00	.00	.02	.00	.00	.00	.00	.49	.99				
	60	.00	.00	.00	.00	.00	.00	.99	.65	.07	.00	.00	.00	.99	.98	.28	.01	.00	.00	.00	.09	.86			
	100	.01	.01	.00	.00	.00	.00	.99	.99	.60	.10	.00	.01	.99	.99	.93	.26	.03	.01	.00	.00	.19	.99		
	150	.03	.03	.01	.01	.01	.01	.99	.99	.90	.31	.01	.04	.99	.99	.99	.59	.14	.07	.00	.00	.05	.99	.99	

Ek 2.D-3: Örneklem büyüklüğü ve test eşitleme yöntemlerinin OHKK değerleri üzerine yaptığı ortak etkilere ait Tukey HSD testi p değerleri.

Yöntem		Ortalamalar						Tucker						Braun/ Holland						Düşük Ağırlıklar					
	Örneklem	15	20	30	60	100	150	15	20	30	60	100	150	15	20	30	60	100	150	15	20	30	60	100	150
Ortalamalar	15																								
	20	.99																							
	30	.99	.99																						
	60	.74	.98	.99																					
	100	.30	.76	.99	.99																				
	150	.17	.57	.99	.99	.99																			
Tucker	15	.00	.00	.00	.00	.00	.00																		
	20	.10	.01	.00	.00	.00	.00	.57																	
	30	.99	.96	.42	.05	.01	.00	.00	.88																
	60	.99	.99	.99	.99	.87	.72	.00	.01	.90															
	100	.96	.99	.99	.99	.99	.99	.00	.00	.18	.99														
	150	.69	.97	.99	.99	.99	.99	.00	.00	.04	.99	.99													
Braun / Holland	15	.00	.00	.00	.00	.00	.00	.99	.23	.00	.00	.00	.00												
	20	.01	.00	.00	.00	.00	.00	.96	.99	.38	.00	.00	.00	.73											
	30	.97	.68	.12	.01	.00	.00	.01	.99	.99	.53	.03	.00	.01	.80										
	60	.99	.99	.99	.94	.60	.40	.00	.03	.99	.99	.99	.92	.00	.00	.83									
	100	.77	.99	.99	.99	.99	.99	.00	.00	.33	.99	.99	.99	.00	.00	.08	.99								
	150	.89	.99	.99	.99	.99	.99	.00	.00	.07	.99	.99	.99	.00	.00	.01	.97	.99							
Düşük Ağırlıklar	15	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00						
	20	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00					
	30	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01	.00	.00	.00	.00	.00	.00	.03				
	60	.01	.00	.00	.00	.00	.00	.98	.99	.31	.00	.00	.00	.80	.99	.73	.00	.00	.00	.00	.00	.00			
	100	.77	.31	.03	.00	.00	.00	.04	.99	.99	.20	.01	.00	.01	.98	.99	.47	.02	.00	.00	.00	.00	.96		
	150	.99	.92	.33	.03	.00	.00	.00	.93	.99	.83	.13	.02	.00	.49	.99	.98	.25	.04	.00	.00	.00	.40	.99	

Ek 2.E: Madde eşik parametre farkı koşullarının ESH, yanlılık ve OHKK değerleri üzerine etkisini gösteren Tukey HSD testine ait p değerleri.

	Madde Eşik Parametre Farkları	0.1	0.4	0.7
ESH	*0.1*			
	0.4	.22		
	0.7	.00	.04	
Yanlılık	*0.1*			
	0.4	.00		
	0.7	.00	.00	
OHKK	*0.1*			
	0.4	.00		
	0.7	.00	.00	

Ek 2.F-1: Madde eşik parametre farkı ve test eşitleme yöntemlerinin ESH değerleri üzerine yaptığı ortak etkilere ait Tukey HSD testi p değerleri

Yöntemler		Ortalamalar			Tucker			Braun / Holland			Düşük Ağırlıklar		
	Madde Eşik Parametresi Farkı	0.1	0.4	0.7	0.1	0.4	0.7	0.1	0.4	0.7	0.1	0.4	0.7
Ortalamalar	0.1												
	0.4	.99											
	0.7	.99	.99										
Tucker	0.1	.04	.03	.01									
	0.4	.00	.00	.00	.98								
	0.7	.00	.00	.00	.00	.07							
Braun / Holland	0.1	.02	.01	.00	.99	.99	.00						
	0.4	.00	.00	.00	.93	.99	.16	.98					
	0.7	.00	.00	.00	.00	.01	.99	.00	.02				
Düşük Ağrılıklar	0.1	.00	.00	.00	.00	.00	.00	.00	.00	.00			
	0.4	.00	.00	.00	.00	.00	.00	.00	.00	.00	.99		
	0.7	.00	.00	.00	.00	.00	.00	.00	.00	.00	.99	.90	

Ek 2.F-2: Madde eşik parametre farkı ve test eşitleme yöntemlerinin yanlılık değerleri üzerine yaptığı ortak etkilere ait Tukey HSD testi p değerleri

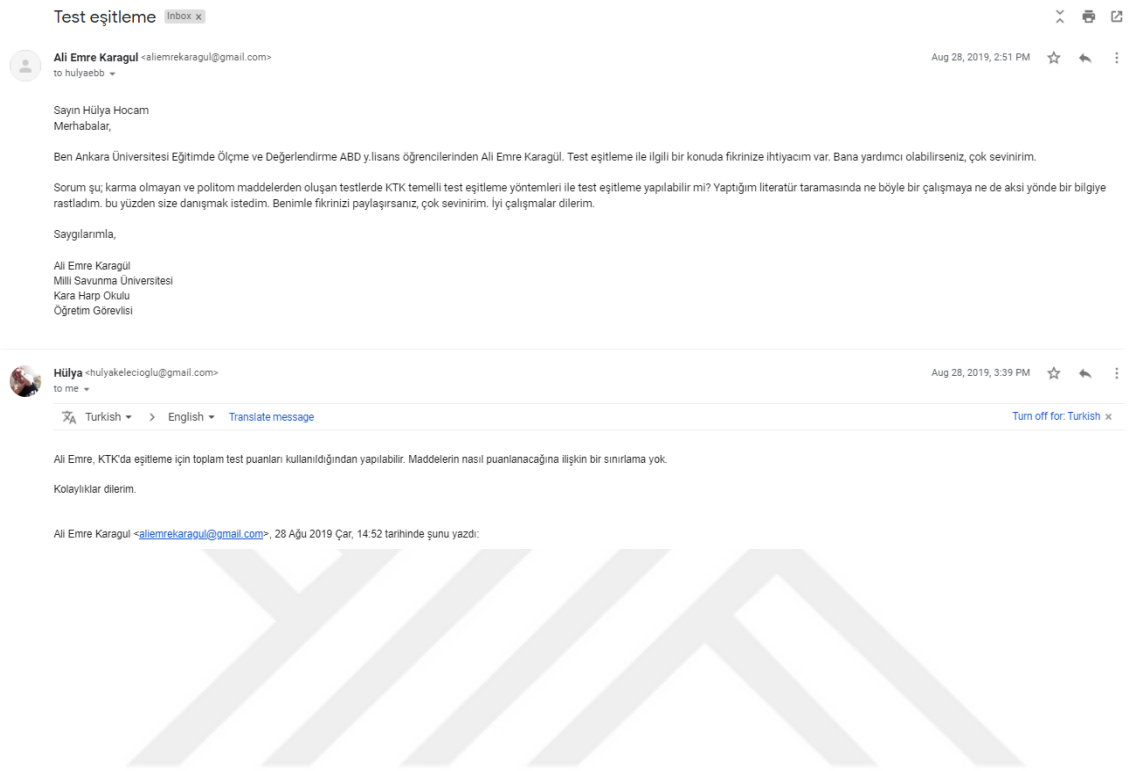
Yöntemler		Ortalamalar			Tucker			Braun / Holland			Düşük Ağırlıklar		
	Madde Eşik Parametresi Farkı	0.1	0.4	0.7	0.1	0.4	0.7	0.1	0.4	0.7	0.1	0.4	0.7
Ortalamalar	0.1												
	0.4	.00											
	0.7	.00	.00										
Tucker	0.1	.00	.91	.00									
	0.4	.00	.00	.91	.03								
	0.7	.00	.00	.00	.00	.00							
Braun / Holland	0.1	.00	.55	.00	.99	.16	.00						
	0.4	.00	.00	.99	.00	.99	.00	.00					
	0.7	.00	.00	.00	.00	.00	.37	.00	.00				
Düşük Ağırlıklar	0.1	.00	.00	.00	.00	.00	.99	.00	.00	.85			
	0.4	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00		
	0.7	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	

Ek 2.F-3: Madde eşik parametre farkı ve test eşitleme yöntemlerinin OHKK değerleri üzerine yaptığı ortak etkilere ait Tukey HSD testi p değerleri

Yöntemler		Ortalamalar			Tucker			Braun / Holland			Düşük Ağırlıklar		
	Madde Eşik Parametresi Farkı	0.1	0.4	0.7	0.1	0.4	0.7	0.1	0.4	0.7	0.1	0.4	0.7
Ortalamalar	0.1												
	0.4	.81											
	0.7	.00	.15										
Tucker	0.1	.00	.27	.99									
	0.4	.00	.00	.74	.54								
	0.7	.00	.00	.00	.00	.00							
Braun / Holland	0.1	.00	.09	.99	.99	.84	.00						
	0.4	.00	.00	.29	.15	.99	.00	.39					
	0.7	.00	.00	.00	.00	.00	.91	.00	.00				
Düşük Ağırlıklar	0.1	.00	.00	.00	.00	.00	.00	.00	.00	.00			
	0.4	.00	.00	.00	.00	.00	.00	.00	.00	.00	.01		
	0.7	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	.00	

EK 3: Çalışma Sürecinde Yapılan Görüşmeler

Ek 3.A: Prof. Dr. Hülya Kelecioğlu ile yapılan e-posta görüşmesi.



Test eşitleme Inbox ×

Ali Emre Karagul <aliemrekaragul@gmail.com> Aug 28, 2019, 2:51 PM
to hulyaebb

Sayın Hülya Hocam
Merhabalar,

Ben Ankara Üniversitesi Eğitimde Ölçme ve Değerlendirme ABD y.lisans öğrencilerinden Ali Emre Karagül. Test eşitleme ile ilgili bir konuda fikrinize ihtiyacım var. Bana yardımcı olabilirseniz, çok sevinirim.

Sorum şu; karma olmayan ve politom maddelerden oluşan testlerde KTK temelli test eşitleme yöntemleri ile test eşitleme yapılabilir mi? Yaptığım literatür taramasında ne böyle bir çalışmaya ne de aksi yönde bir bilgiye rastladım. bu yüzden size danışmak istedim. Benimle fikrinizi paylaşırsanız, çok sevinirim. İyi çalışmalar dilerim.

Saygılarımla,

Ali Emre Karagül
Milli Savunma Üniversitesi
Kara Harp Okulu
Öğretim Görevlisi

Hülya <hulyakelecioglu@gmail.com> Aug 28, 2019, 3:39 PM
to me

Turkish › English Translate message Turn off for: Turkish ×

Ali Emre, KTK'da eşitleme için toplam test puanları kullanıldığından yapılabilir. Maddelerin nasıl puanlanacağına ilişkin bir sınırlama yok.

Kolaylıklar dilerim.

Ali Emre Karagul <aliemrekaragul@gmail.com>, 28 Ağu 2019 Çar, 14:52 tarihinde şunu yazdı:

EK 4: Etik Kurul Onayı

ANKARA ÜNİVERSİTESİ
SOSYAL BİLİMLER ALT ETİK KURULU
KARAR ÖRNEĞİ

Karar Tarihi : **07/10/2019**
Toplantı Sayısı : **12**
Karar Sayısı : **354**

354- Üniversitemiz Eğitim Bilimleri Enstitüsü Ölçme ve Değerlendirme Anabilim Dalı yüksek lisans öğrencisi **Ali Emre Karagül**'ün "Küçük Örneklemlerde Çok Kategorili Puanlanan Maddelerden Oluşan Testlerde Klasik Test Eşitleme Yöntemlerinin Karşılaştırılması" başlıklı tezi ile ilgili 06/09/2019 tarihli "İnsan Üzerinde Yapılan Klinik Dışı Araştırmalar Başvuru Formu" Etik Kurulumuzca incelendi.

Üniversitemiz Eğitim Bilimleri Enstitüsü Ölçme ve Değerlendirme Anabilim Dalı yüksek lisans öğrencisi **Ali Emre Karagül**'ün "Küçük Örneklemlerde Çok Kategorili Puanlanan Maddelerden Oluşan Testlerde Klasik Test Eşitleme Yöntemlerinin Karşılaştırılması" başlıklı tezinin, insan üzerinde yapılan bir çalışma olmaması nedeniyle etik kurul onayına ihtiyaç duymadığına oy birliği ile karar verildi.

ASLININ AYNIDIR
07/10/2019

Prof. Dr. Muharrem ÖZEN
Ankara Üniversitesi
Etik Kurulu Başkanı

BENZERLİK BİLDİRİMİ

"Küçük Örneklemlerde Çok Kategorili Puanlanan Maddelerden Oluşan Testlerde Klasik Test Eşitleme Yöntemlerinin Karşılaştırılması" başlıklı tezimin ana bölümü (ön bölüm, kaynaklar ve ekler hariç) Turnitin İntihali Engelleme Programı aracılığıyla incelenmiş ve ilgili rapor danışmanım tarafından da kontrol edilmiştir. Kontrol sırasında (1) "Kapak, Önsöz, Özet, İçindekiler, Kısaltmalar/ Simgeler Sayfası, Tablolar Dizini, Şekiller Dizini, Dipnotlar, Kaynaklar, Ekler, Özgeçmiş" (2) "Tırnak içinde ve blok olarak verilen (doğrudan) alıntılar" (3) "Yedi sözcüğe kadar olan benzerlik/örtüşme içeren metin kısımları" (4) "Öğrencinin lisansüstü tezi ile ilgili yapmış olduğu yayınlar ile yararlandığı mevzuat metinleri" dışarıda tutulmuştur. Benzerlik kontrolüne ilişkin rapordan elde edilen sonuçlar aşağıda sunulmuştur.

Rapor Tarihi	: 30.12.2019
Gönderim Numarası	: 1238745234
Sayfa Sayısı	: 63
Sözcük Sayısı	: 23521
Karakter Sayısı	: 151313
Benzerlik Oranı	: %5
Savunma Tarihi	: 09.01.2020

Yukarıda belirtilen sonuçları gösteren Turnitin İntihali Engelleme Programı'na ilişkin orijinal raporu, sonuçlarda herhangi bir değişiklik yapmaksızın bu bildirimin ekinde Enstitüye teslim ettiğimi, tezimin **%10**'dan fazla benzerlik oranı içerdiğinin ve tek bir kaynakla eşleşme oranının **%2**'yi geçtiğinin belirlenmesi durumunda, bundan doğabilecek tüm yasal sorumluluğu kabul ettiğimi bildirir, saygılarımı sunarım.

Öğrencinin Adı Soyadı: Ali Emre KARAGÜL

Tarih: .....

İmza: .....

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı ve Soyadı **:** Ali Emre KARAGÜL

E-posta **:** aliemrekaragul@gmail.com

Öğrenim Durumu

Lisans **:** Gazi Üniversitesi- İngilizce Öğretmenliği

Yüksek Lisans **:** Ankara Üniversitesi- Eğitimde Ölçme ve Değerlendirme

İş Deneyimi

Görevi	Görev yeri	Yıl
İngilizce Öğretmeni	Şırnak Anadolu Lisesi- ŞIRNAK	2014-2015
İngilizce Öğretmeni	İncek Amerikan Kültür Koleji- ANKARA	2015-2016
Okutman	Ufuk Üniversitesi- ANKARA	2016-2017
Öğretim Görevlisi	Milli Savunma Üniversitesi- ANKARA	2017-2019
İngilizce Öğretmeni	AKN Yabancı Dil Kursu- ANKARA	2019-...