

**ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

**EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME PROGRAMI**

**KOPYA BELİRLEMEDE BENZERLİK İNDEKSLERİNİN
BİREY-UYUM İSTATİSTİKLERİ ARACILIĞIYLA
AŞAMALI KULLANIMININ I. TİP HATALARININ VE
GÜCÜNÜN İNCELENMESİ**

DOKTORA TEZİ

ARZU UÇAR

**ANKARA
OCAK, 2021**



**ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

**EĞİTİM BİLİMLERİ ANABİLİM DALI
EĞİTİMDE ÖLÇME VE DEĞERLENDİRME PROGRAMI**

**KOPYA BELİRLEMEDE BENZERLİK İNDEKSLERİNİN
BİREY-UYUM İSTATİSTİKLERİ ARACILIĞIYLA
AŞAMALI KULLANIMININ I. TİP HATALARININ VE
GÜCÜNÜN İNCELENMESİ**

DOKTORA TEZİ

ARZU UÇAR

DANIŞMAN: DOÇ. DR. CELAL DEHA DOĞAN

**ANKARA
OCAK, 2021**

Ankara Üniversitesi Eğitim Bilimleri Enstitüsü Müdürlüğüne,

Arzu UÇAR adlı öğrencinin hazırladığı “Kopya Belirlemede Benzerlik İndekslerinin Birey-Uyum İstatistikleri Aracılığıyla Aşamalı Kullanımının I. Tip Hatalarının Ve Gücünün İncelenmesi” başlıklı bu çalışma Eğitim Bilimleri Anabilim Dalı / Eğitimde Ölçme ve Değerlendirme Doktora Programı’nda jüri üyelerince oy birliği ile **Doktora Tezi** olarak kabul edilmiştir.

	<u>Jüri Üyeleri</u>	<u>İmza</u>
Başkan	Prof. Dr. Nuri DOĞAN
Üye (Danışman)	Doç. Dr. Celal Deha DOĞAN
Üye	Prof.Dr. Ömay ÇOKLUK BÖKEOĞLU
Üye	Doç. Dr. Önder SÜNBÜL
Üye	Doç.Dr. Seher YALÇIN

ONAY

Bu tez Ankara Üniversitesi Lisansüstü Eğitim Öğretim Yönetmeliği’nin ilgili maddeleri uyarınca jüri üyeleri tarafından 28/01/2021 tarihinde, Enstitü Yönetim Kurulunca .../.../20... tarihinde kabul edilmiştir.

Prof. Dr. Yasemin KARAMAN KEPENEKÇİ

Eğitim Bilimleri Enstitüsü 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.

Arzu UÇAR

ÖZET

KOPYA BELİRLEMEDE BENZERLİK İNDEKSLERİNİN BİREY-UYUM İSTATİSTİKLERİ ARACILIĞIYLA AŞAMALI KULLANIMININ I. TİP HATALARININ VE GÜCÜNÜN İNCELENMESİ

UÇAR, Arzu

Doktora Tezi, Ölçme ve Değerlendirme Anabilim Dalı

Tez Danışmanı: Doç. Dr. Celal Deha DOĞAN

Ocak, 2021, xiii+94

Bu çalışma, bilgisayar temelli ve kağıt-kalem testi uygulamalarında, sınav maddelerine ilişkin madde ön bilgisine sahip olmayan ve cevap kopyalamadan kaynaklı test hilesi yapan bireyleri tespit etmek için aşamasız ve iki aşamalı olarak kullanılan indekslerin çeşitli koşullar altında I. tip hata oranları ve güçlerinin nasıl etkilendiğini ortaya koymayı amaçlamaktadır. Bu amaç doğrultusunda, I. tip hata oranı ortalamaları elde edilmesi için Rasch modele dayalı, test hilesi yapan bireyin yeteneği (çok düşük yetenek, düşük yetenek), testin ortalama gücü (kolay, zor) koşulları altında 100 yinleme ile 400 veri kümesi üretilmiştir. Kopya ve benzerlik indekslerinin, birey uyum istatistiklerinin, Kullback-Leibler uzaklık indeksinin güç ortalamaları için yetenek düzeyi, ortalama test gücü ve kopya çekilen madde oranı (0.1, 0.4 ve 0.6) koşulları altında 100 yinleme yapılarak 1200 veri kümesi oluşturulmuştur. Aşamasız analizde kopya ve benzerlik indekslerinin, birey uyum istatistiklerinin ve Kullback-Leibler uzaklık indeksinin değişimlenen şartlar altında I. tip hata ve güç ortalamaları hesaplanmıştır. İki aşamalı analizde önce birey uyum/Kullback-Leibler indeksleri ile test hilesi yapan bireyler belirlenmiştir. Belirlenen bireylerin sınav salonundaki yerlerine göre cevaplarını kopyalayabileceği potansiyel kaynak bireyler ile olan analizi için ise kopya ve benzerlik indeksleri kullanılarak, kopya ve benzerlik indekslerinin I. tip hata oranı ve güç ortalamaları elde edilmiştir. İki aşamalı analizde aşamasız analize göre daha düşük I. tip hata oranı ortalamaları, yine aşamasız analize göre daha yüksek güç ortalamaları elde edilmiştir. Kullback-Leibler indeksinin aşamasız kullanıldığında en düşük güç ve en yüksek hata ortalamasına sahip olduğu görülmüştür.

Anahtar Sözcükler: Kopya ve benzerlik indeksleri, birey uyum istatistikleri, kullback-leibler indeksi, test hilesi, cevap kopyalama



ABSTRACT

INVESTIGATION OF TYPE-I-ERROR AND POWER OF SIMILARITY INDICES BY USING TWO-STAGE ANALYSIS VIA PERSON-FIT STATISTICS

UÇAR, Arzu

Dissertation, Department of Measurement and Evaluation

Supervisor: Assoc. Prof. Dr. Celal Deha DOĞAN

Janury, 2021, xii+94

This study aimed to how the rate of Type I error and the power of statistics are affected under various conditions. No-step analysis and two-steps analysis were used to detect copier who does not have any item pre-knowledge and copy from the source. 400 data sets were generated with 100 replication under the conditions of the copiers' ability (very low ability, low ability) and the average of test difficulty (easy, difficult). Then the rates of type I error were calculated for these conditions. 1200 data sets were simulated with 100 replication under the conditions of ability level, test difficulty, and copying ratio (0.1, 0.4, and 0.6) to calculated the power of answer copying and similarity indices, person-fit statistics and Kullback-Leibler divergence index. In the no-step analysis, type I error and power were calculated for each index and statistic. For the two-steps analysis, the person fit statistic/Kullback-Leibler divergence index was used to detect copiers in the first step. In the second step, answer copying and similarity indexes were used to detect copiers by comparing only copiers' potential sources in the classroom. After that, the rate of type I error and power of answer copying and similarity indices were calculated under varying conditions. In two-steps analysis, the rates of type I error of answer copying and similarity indices were lower. Besides, the power of answer copying and similarity indices were higher. It has been observed that the Kullback-Leibler divergence index has the lowest power and the highest error when used no-step analysis.

Key Words: Answer copying and similarity indices, person-fit statistics, kullback-leibler index, cheating, answer copying



ÖNSÖZ

Hem bireysel hem de toplumsal sonuçlarıyla dikkatleri üzerine çeken geniş ölçekli testlerin güvenliği son derece önemlidir. Ülkemizde test güvenliği konusunda son dönemde çeşitli çalışmalar yapılmış olsa da, test hilesinde bulunma davranışının kontrol edilemediği gibi hileyi yapan/ları en iyi şekilde sınıflandıran tek bir yöntem yoktur. Günlük hayatta sınav esnasında karşılaşılabilecek durum ve test hilesini belirlemede kullanılan yöntemleri etkileyen şartlar göz önünde bulundurularak bireysel test hilesini belirlemeye yönelik iki aşamalı analiz yaklaşımıyla, test hilesi belirlemeye ek kanıtlar ortaya koymayı amaçlayan çalışma gerçekleştirilmiştir. Çalışmadan elde edilen sonuçlar ışığında uygulamaya ve araştırmaya yönelik önerilerde bulunulmuştur.

İnsana ve mesleğine saygılı, özgür bir çalışma süreci ve ortamı sunan, hoş görümlü ve sabırlı değerli Doç. Dr. Celal Deha Doğan hocama, doktora eğitimim boyunca görüş ve önerileriyle bana yol gösterdiği, çalışmamın planlanıp bir ürün olarak ortaya çıkmasına kadar olan her aşamasında titizlikle takip ettiği ve destek verdiği için çok teşekkür ederim.

Tezime görüş ve önerileriyle katkıda bulunan Prof. Dr. Ömay ÇOKLUK BÖKEOĞLU'na, çalışma konumun belirlenmesinde ve çalışmam süresince görüş ve önerileriyle katkıda bulunan Doç. Dr. Önder SÜNBU'L'e, ayrıca tez jürimde yer alarak görüş ve önerileriyle katkı sağlayan Prof. Dr. Nuri DOĞAN'a ve Doç. Dr. Üyesi Seher YALÇIN'a teşekkürlerimi sunarım.

Yüksek lisans ve doktora eğitimim boyunca desteğini esirgemeyen değerli arkadaşım Araş. Gör. Ebru BALTA'ya teşekkür ederim.

Bütün hayatım boyunca maddi ve manevi destekleriyle yanımda olan, bu günlere gelmemi sağlayan aileme, çalışma sürecinde her zaman yanımda olan ve destekleyen eşime sonsuz teşekkürlerimi sunarım.

İÇİNDEKİLER

	Sayfa
ETİK İLKELERE UYGUNLUK BİLDİRİMİ.....	iii
ÖZET	iv
ABSTRACT	vi
ÖNSÖZ.....	viii
İÇİNDEKİLER.....	ix
TABLolar DİZİNİ.....	xi
ŞEKİLLER DİZİNİ.....	xii
KISALTMALAR.....	xiii
BÖLÜM 1.....	1
GİRİŞ.....	1
Problem.....	1
Amaç.....	7
Önem.....	8
Sınırlılıklar.....	9
Tanımlar.....	10
BÖLÜM 2.....	11
KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR.....	11
Kopya ve Benzerlik İndeksleri.....	11
Birey Uyum İstatistiği.....	15
Kullback Leibler Uzaklığı İndeksi.....	18
İlgili Araştırmalar.....	19
BÖLÜM 3.....	28
YÖNTEM.....	28
Araştırmanın Modeli.....	28
Simülasyon Deseni.....	28
Araştırmanın Değişkenleri.....	28
Manipüle Edilmeyen Değişkenler.....	29
Manipüle Edilen Değişkenler.....	30
Verilerin Üretilmesi.....	31
Verilerin Analizi.....	33
İndekslerin I. Tip Hata Oranının Belirlenmesi.....	33
İndekslerin Güç Oranının Belirlenmesi.....	35
BÖLÜM 4.....	37
BULGULAR VE YORUMLAR.....	37
İndekslerin Aşamasız Kullanımı Sonucunda Elde Edilen I. Tip Hata Oranlarına Yönelik Bulgular.....	37
İndekslerin Aşamasız Kullanımı Sonucunda Elde Edilen Güç Oranlarına Yönelik Bulgular.....	42
İndekslerin İki Aşamalı Kullanımı Sonucunda Elde Edilen I. Tip Hata Oranlarına Yönelik Bulgular.....	47
İndekslerin İki Aşamalı Kullanımı Sonucunda Elde Edilen Güç Oranlarına Yönelik Bulgular.....	50
BÖLÜM 5.....	56
SONUÇLAR VE ÖNERİLER.....	56
Sonuçlar.....	56

Öneriler	58
Uygulayıcılara Öneriler.....	58
Araştırmacılara Öneriler.....	59
KAYNAKLAR.....	61
EKLER	70
EK 1. Etik Kurul Onayı	71
EK 2. Aşamasız ve İki aşamalı Analizlere Ait I. Tip Hata ve Güç Oranları	72
EK 3. Rasch Modelde 150 Örneklem Büyüklüğü Ve 30 Test Uzunluğu Kullanıldığında Elde Edilen Faktör Analizi Ve Paralel Analiz Sonuçları.....	83
EK 4. Kullback Leibler Uzaklık İndeksi İçin Elde Edilen Kesme Puanları	84
EK 5. Farklı Kullback Leibler Uzaklık İndeksi Kesme Puanlarının I. Tip Hata ve Güç Oranları	85
EK 6. Güçlük Parametresi Ortalaması	86
EK 7. ω ve GBT indekslerinin Aşamasız ve İki Aşamalı Analizleri I. Tip Hata Oranları Temel Etki ve Ortak Etki Grafikleri	87
EK 8. ω ve GBT indekslerinin Aşamasız ve İki Aşamalı Analizleri Güç Oranları Temel Etki ve Ortak Etki Grafikleri	89
BENZERLİK BİLDİRİMİ	92
ÖZGEÇMİŞ.....	93

TABLolar DİZİNİ

Tablo	Sayfa
Tablo 1 Birey Uyum İstatistikleri.....	4
Tablo 2 Simülasyon Deseni.....	31
Tablo 3 İndeksin I. Tip Hatası ve Kopya Belirleme Gücü.....	33
Tablo 4 Yetenek Düzeyi Altındaki Yöntemlerin I. Tip Hata Oranları Ortalaması.....	72
Tablo 5 Test Güçlüğü Düzeyleri Altındaki Yöntemlerin I. Tip Hata Oranları Ortalaması.....	72
Tablo 6 Değişimlenen Faktörler Altındaki I. Tip Hata Oranları.....	73
Tablo 7 Yetenek Düzeyi Altındaki Güç Oranı Ortalaması.....	73
Tablo 8 Ortalama Test Güçlüğü Düzeyleri Altındaki Güç Oranı Ortalaması.....	74
Tablo 9 Kopya Oranı Düzeyleri Altındaki Güç Oranı Ortalamasıemel Etkisi.....	74
Tablo 10 Değişimlenen Faktörler Altındaki Güç Oranları.....	75
Tablo 11 ω ve GBT İndekslerinin Yetenek Düzeyi Altındaki I. Tip Hata Oranı Ortalaması.....	76
Tablo 12 ω ve GBT İndekslerinin Test Güçlüğü Düzeyleri Altında I. Tip Hata Oranları Ortalaması.....	77
Tablo 13 ω ve GBT İndekslerinin Değişimlenen Faktörler Altındaki I. Tip Hata Oranları.....	78
Tablo 14 ω ve GBT İndekslerinin Yetenek Düzeyi Altındaki Güç Oranı Ortalaması... 79	79
Tablo 15 ω ve GBT İndekslerinin Ortalama Test Güçlüğü Düzeyleri Altındaki Güç Oranı Ortalaması.....	79
Tablo 16 ω ve GBT İndekslerinin Kopya Oranı Düzeyleri Altındaki Güç Oranı Ortalaması.....	80
Tablo 17 ω ve GBT İndekslerinin Değişimlenen Faktörler Altındaki Güç Oranı.....	81
Tablo 18 Paalizden Elde Edilen Öz Değerler.....	83
Tablo 19 Zor ve Kolay Test İçin Doğrulayıcı Faktör Analizi Sonuçları.....	83
Tablo 20 KL Değerlerinin Kesme Puanları.....	84
Tablo 21 Güç ve I. Tip Hata Oranları.....	85
Tablo 22 Güçlük Parametresi Ortalaması.....	86

ŞEKİLLER DİZİNİ

Şekil	Sayfa
Şekil I. Yetenek Düzeyinin I. Tip Hata Oranı Ortalamasına Olan Temel Etkisinin Grafiği.....	38
Şekil II. Testin Ortalama Güçlüğünün I. Tip Hata Oranı Ortalamasına Olan Temel Etkisinin Grafiği.	39
Şekil III. Değişimlenen Faktörlerin I. Tip Hata Oranları Olan Ortak Etkisinin Grafiği	41
Şekil IV. Yetenek Düzeyinin Güç Oranı Ortalamasına Olan Temel Etkisinin Grafiği..	42
Şekil V. Testin Ortalama Güçlüğünün Güç Oranı Ortalamasına Olan Temel Etkisinin Grafiği.....	43
Şekil VI. Kopya Oranının Güç Oranı Ortalamasına Olan Temel Etkisinin Grafiği.....	44
Şekil VII. Değişimlenen Faktörlerin Güç Oranına Olan Ortak Etkisinin Grafiği.....	45
Şekil VIII. Yetenek Düzeyinin I. Tip Hata Oranları Ortalamasına Olan Temel Etkisinin Grafiği.....	47
Şekil IX. Ortalama Test Güçlüğünün I. Tip Hata Oranları Ortalamasına Olan Temel Etkisinin Grafiği.	48
Şekil X. Değişimlenen Faktörlerin I. Tip Hata Oranları Olan Ortak Etkisinin Grafiği .	49
Şekil XI. Yetenek Düzeyinin Güç Oranı Ortalamasına Olan Temel Etkisinin Grafiği..	51
Şekil XII. Ortalama Test Güçlüğünün Güç Oranı Ortalamasına Olan Temel Etkisinin Grafiği.....	52
Şekil XIII. Kopya Oranının Güç Oranına Ortalaması Olan Temel Etkisinin Grafiği... ..	53
Şekil XIV. Değişimlenen Faktörlerin Güç Oranına Olan Ortak Etkisinin Grafiği..	54
Şekil XV. Değişimlenen Faktörlerin I. Tip Hata Oranları Olan Ortak Etkisinin Grafiği	88
Şekil XVI. Yetenek Düzeyinin Güç Oranı Ortalamasına Olan Temel Etkisinin Grafiği	89
Şekil XVII. Testin Ortalama Güçlüğünün Güç Oranı Ortalamasına Olan Temel Etkisinin Grafiği	90
Şekil XVIII. Değişimlenen Faktörlerin Güç Oranları Olan Ortak Etkisinin Grafiği	91

KISALTMALAR

BHLM	Bayesian Hiyerarşik Lineer Model
BOBUT	Bilgisayar Ortamında Bireye Uyarlanmış Test
BTF	Birey Tepki Fonksiyonu
GBT	Genelleştirilmiş Binom Testi
KL	Kullback Leibler
MTK	Madde Tepki Kuramı
2PLM	İki Parametrelı Logistik Model
3PLM	Üç parametrelı Lojistik Model

BÖLÜM 1

GİRİŞ

Bu bölümde araştırmanın problem durumu açıklanmıştır. Ayrıca araştırmanın amaç, önem, sınırlılıklar ve tanımlarına yer verilmiştir.

Problem

Gelişen bilim ve teknolojiden yararlanan kurum ve kuruluşlar bireylere kağıt-kalem testleri dışında bilgisayar ortamında da testler uygulamaktadır. Sınav salonlarında gerçekleştirilen test uygulamaları dışında gözetmensiz çevrimiçi testlerin uygulanması test güvenliğinin önemini daha arttırmaktadır. Çünkü sertifika ve lisans sınavları gibi her türlü testlerde bireylerin test hilesinde bulunma motivasyonları her zaman vardır. İyi puan alma çabası, bir üst seviyeye geçme isteği ya da rekabete yönelik testlerin uygulanması gibi birçok neden bireylerin testte hile yapmasına neden olmaktadır.

Eğitimde bireylerin ilgi, performansları ve başarıları gibi doğrudan gözlenemeyen özelliklerin gözlenmesi ve ölçülebilmesi amacıyla geliştirilen ölçme araçlarına test adı verilir (Baykul, 2010). Testlerden elde edilen bilgilerin kullanılarak karara varılması ya da çalışmaların yapılabilmesi için testlerin geçerli ve güvenilir olması gerekmektedir.

Ölçme çalışmalarında gerçek puanlara yakın gözlenen puanlar elde etmek amaçlanmaktadır. Bu durum ise güvenilir ölçmelerdeki kaynağı ve yönü bilinmeyen seçkisiz hataların azlığı ölçüsünde gerçekleşmektedir (Baykul, 2010). Bu sebeple güvenilirlik, ölçme sonuçlarının tesadüfi hatalardan arınık olma derecesi olarak tanımlanmaktadır (Turgut ve Baykul, 2010).

Bir diğer özellik testlerin geçerli olmasıdır. Geçerlik, “ölçme aracının ölçmek istediği özelliği başka değişkenlerle karıştırmadan ölçebilme derecesidir” (APA, 1974; Akt: Atılgan, Kan ve Aydın, 2014). Başka bir ifadeyle, testin ölçmek istediği özelliği hatasız olarak ölçebilme derecesidir. Testin amacı dışında ölçme sonuçlarına; puanlamada yanlış davranılması, yazı güzelliğine puan verilmesi, bireylerin hile yapması (kopya çekmesi) gibi başka değişkenler de karışabilmektedir. Testin geçerliğini düşüren bu durumları düzeltmek ya da engellemek daha kolay olabilmektedir. Bireyler hakkında karar

verilirken kullanılan ölçme sonuçlarının geçerli bir testten elde edilmesi ya da ölçümlerin geçerli olması önemlidir. Çünkü bireyler hakkında karar ya da geri dönüt verilirken bireyle ilgili elde edilen ölçme sonuçlarından yararlanılmaktadır.

Geniş ölçekli testler, sertifika testleri gibi akademik testlerden elde edilen ölçme sonuçları kullanarak bireyle ilgili kararlar alınmaktadır. Alınan kararlarsa bireyin hayatını etkilemektedir. Böylesi önemli durumlarda bireyin kopya çekme (hile yapma) motivasyonu artmaktadır. Kopya ile ilgili temel sorun, bireylerin yeterlilik, beceri ve yetkinlikleri ile ilgili çıkarımların geçerliliğinin tehlikeye girmesi olarak görülmektedir. Sonuçta geçersiz yeterlilik tahminleri sonraki karar vermede temel olarak kullanılmaktadır. Bu soruna bir cevap olarak, iki kişinin cevap vektörü arasında yalnızca şans eseri gerçekleşmesi mümkün olmayan olağandışı uyuma derecesini tespit etmek için çeşitli yöntemler geliştirilmiştir.

Bir testte yer alan maddelere verilen cevaplar ve testi alan bireyin birleştirilmiş test puanı, bireyin yetenek düzeyi, bilgi ve beceri türü açısından uygunluğu da dahil olmak üzere testin ölçmeyi amaçladığı çeşitli faktörlerden etkilenmektedir. Belirli bir yetenek seviyesine sahip bireyin teste verdiği tepkiler sonucunda tepki örüntüsünün uygun olup olmaması, temel alınan madde tepki modeli veya test puanı, kullanılan ölçme modeline ya da örneklemdaki/popülasyondaki benzer seviyede yeteneğe sahip test katılımcılarının tepki modellerine ve puanlarına dayanarak belirli bir kesinlik derecesi ile belirlenmektedir. Bireyin gözlenen tepki örüntüsü ve test puanı, beklenen tepki örüntüsü ve puana uygun değil ise, tepki örüntüsü ve puanı beklenmeyen (anormal/uygun olmayan) olarak ifade edilir. Beklenmeyen tepkiler ve test puanlarına neden olacak birçok faktör bulunmaktadır. Bunlar; bireylerin maddeleri yanıtlamadaki alışılmadık davranışlarını, gözetmen ya da test uygulayıcısının, test hakkında, test öncesi ve sonrasında bilgi paylaşımında bulunması veya hile gibi uygunsuz davranışları veya hile gibi diğerlerini içermektedir (Karabatsos, 2003; Meijer, 1994, 1996; Thiessen, 2008). Beklenmeyentepekiler, bireyler için beklenmedik yüksek veya beklenmedik düşük test puanlarına neden olmaktadır. Test sonuçlarında sapmanın varlığı test puanlarına hata karıştırmakta ve önerilen yorum ve kullanımlarını geçersiz kılmaktadır (Cizek ve Wollack, 2017).

Meijer (1994, 1996) çalışmasında beklenmeyen davranış olarak adlandırdığı tepkileri, beklenmeyen yüksek ya da düşük puanlar, dikkatsiz cevaplama (bireyin doğru cevaplayabileceği maddeleri yanlış cevaplama), şanslı tahminde bulunma (bireyin doğru cevabını bilmediği maddeleri, doğru cevapları tahmin ederek cevaplama),

yaratıcı cevaplama (yüksek yetenekli bireylerin maddelerin yaratıcı ve karmaşık yorumları nedeniyle bazı kolay maddeleri yanlış cevaplaması) ve rastgele cevaplama (bireyin bazı çoktan seçmeli maddelerin alternatiflerini rastgele seçerek cevaplaması), test hilesinde bulunma şeklinde beş faktöre ayırarak tanımlamıştır. Bu çalışmada beklenmeyen davranışlardan biri olan test hilesinde bulunma (kopya çekme, kopya) faktörü ele alınmıştır. Test hilesi, bireylerin, testte yer alan maddeleri önceden bilmesi veya diğer sınava giren bireylerin cevaplarının kopyalaması veya öğretmenleri tarafından verilen cevapların kopyalaması yoluyla doğru şekilde cevap veremedikleri maddeler üzerinde yasal olmayan bir şekilde doğru cevaplar verdikleri davranışlardır.

Hile, aynı sınıftan (veya okuldan) veya farklı sınıflardan (veya okullardan) çok sayıda bireyi içeren hem bireysel düzeyde hem de grup düzeyinde gerçekleşebilmektedir. Bireysel düzeyde gerçekleşen test hileleri, bireyin doğru cevabını bilmediği maddeleri, doğru cevapları tahmin ederek cevaplaması (şanslı tahminde bulunma), testte yer alan maddeleri önceden bilmesi veya diğer sınava giren bireylerin cevaplarının kopyalaması şeklinde meydana gelebilmektedir. Grup düzeyinde gerçekleşen test hilesi bazı durumlarda öğretmenlerin veya sınava girenlerin sınav puanlarını hileli bir şekilde yükseltmelerine yardımcı olan diğer ilgili kişileri içerebilmektedir. Bu test hilesi türü eğitimci hilesi olarak tanımlanmaktadır (Thiessen, 2008). Grup düzeyinde yapılan test hilesi, bir gözetmeni ya da bir sınıf öğretmenini, test sırasında sistematik cevap paylaşımını ve uygun olmayan işaretleme gibi durum ve davranışları içermektedir (Belov, 2013; Wollack ve Maynes, 2011). Belov (2014b) grup düzeyinde yapılan test hilesini, bireylerin aynı merkezden, aynı gruptan ya da aynı sosyal ağa ait olunması gibi durumları belirterek coğrafi konumla sınırlı olmadığını ifade etmiştir. Beklenmeyen tepkileri ve beklenmeyen test puanlarını incelemek için, testlerde yapılan farklı hile türleriyle ilişkili anomaliyi tespit etmeye odaklanan çok sayıda istatistiksel teknik geliştirilmiştir. Her ne kadar test hilesini tespit etmek için kullanılan yöntemler üzerine araştırmalar öncelikle bireysel düzeye odaklanmış olsa da, son yıllarda grup düzeyinde testte yapılan hileyi tespit etmek için kullanılan yöntemler üzerinde odaklanan çalışmalar dikkat çekmektedir (Belov, 2013; Wollack ve Maynes, 2011).

Bireysel ve grup düzeyinde yapılan test hilesini tespit etmek için birden fazla istatistiksel yöntemin kullanılmasını içeren çok sayıda çalışma vardır (Karabatsos, 2003; Krimpen-Stoop ve Meijer, 2001; Meijer ve Sijtsma, 2001; Meijer ve Tendeiro, 2014; Sideridis ve Zopluoğlu, 2018; Wise ve Kong, 2005; Wise ve DeMars, 2006; Wise, Ma, Kingsbury ve Hauser, 2010; Wollack, 2006; Zopluoğlu, 2017). Bireysel düzeyde yapılan

test hilesini tespit etmek için kullanılan yöntemler cevap kopyalama ve benzerlik analizleri, birey uyum istatistikleri, testin iki farklı alt maddeleri kümesindeki performanslarının karşılaştırılması ve model yaklaşımları şeklinde sınıflandırılmıştır (He, Meadows ve Black, 2018). Cevap kopyalama ve benzerlik analizleri; Angoff'un B ve H indeksleri, K indeksi, g_2 indeksi, ω indeksi, S_1 ve S_2 indeksleri, VMI_s (Variable Match) indeksleri olan ξ ve ξ^* indeksleri, Wesolowsky' nin Z benzerlik indeksi, Genelleştirilmiş Binom Testi (GBT) indeksi ve M_4 benzerlik indeksidir. Birey uyum istatistikleri maddelerin ikili-çoklu puanlamalarına ve parametrik-parametrik olmayan model türüne göre farklılaşmaktadır. Aşağıdaki tabloda birey uyum istatistiklerine yer verilmiştir.

Tablo 1

Birey Uyum İstatistikleri

Model Türü	Puanlama Türü	
	İkili	Çoklu
İstatistik		
Parametrik Olmayan Model	$r.pbis$ C G, G_n A, D, E $U3, ZU3$ C^* NCI H^T	G^P G_N^P $U3^P$
	Parametrik Model	l_z l_z^P l_z^*

Kaynak. He ve vd., 2018, 28.

Testin iki farklı alt maddeleri kümesindeki performanslarının karşılaştırılması yöntemleri ise Kullback-Leibler uzaklığı indeksi (KL), MPI (Matched Percentile) indeksi, IRI (Irregularity) indeksi, farklı puanlar için Z-testi istatistiği, model yaklaşımli yöntemde ise DG (Deterministic, Gated IRT Model) modeli yer almaktadır.

Grup bazında test hilesini belirlemek için kullanılan yöntemler tepki örüntüsü ve test puanına bağlı analizler çatısı altında toplanmıştır. Bu analizler cevap değiştirme (yanlış cevaplanmış maddeyi doğru cevaplama) analizleri, tepki örüntüsü ve test puanına dayalı analizler, tepki puanı benzerliği ve değişkenlere dayalı analizler şeklinde

gruplandırılmıştır (He, Meadows ve Black, 2018). Tepki örüntüsü ve test puanına dayalı analizler; benzerlik indekslerine dayalı olanlar, birey uyum istatistiklerine dayalı olanlar, madde tepki ve grup içi test puanına dayalı olanlar ve testte yer alan maddelerin alt küme puanları arasındaki ilişkilere dayalı olan analizler şeklinde sınıflandırılmıştır. Tepki puanı benzerliği ve değişkenlere dayalı analizler Jacob ve Levitt yaklaşımı, iki oranlı Z-puanı yaklaşımı ve çok düzeyli lojistik regresyon yaklaşımı; değişkenlerle ilişkiye dayalı analizler ise kümülatif logit regresyon modelinin kullanımı, regresyon ve kümeleme temelli yaklaşımlar, çok düzeyli modelleme, basit doğrusal regresyon analizi ve standartlaştırılmış puanlar ve temel eksen kullanan artık analiz yaklaşımı şeklinde gruplandırılmıştır. Bununla birlikte son yıllarda kullanımı yaygınlaşan bilgisayar destekli testlerde, bireylerin test maddelerine verdikleri tepkilerle daha ayrıntılı bilgiler elde edilmektedir. Test maddesinin ekranda görülme anı, bireyin test maddesini çözme süresi ve maddeye cevap verme anı gibi zaman bilgilerini de elde etmeyi sağlamaktadır. Bu bilgiler ışığında bireylerle ilgili uygun olmayan tepki örüntüleri tespit edildiğinde bu tespiti desteklemek ya da tanımlamak için bireyin test maddeleri karşısındaki kayıtlı zaman bilgileri de kullanılmaktadır. Hauser, Kingsbury ve Houser (2011) tarafından yapılan bir çalışmada, madde düzeyinde cevap verme süresi ile ilgili olarak, test hilesinin madde düzeyinde ortaya çıktığı şeklinde kavramsallaştırmanın uygun olduğu sonucuna varmıştır. Wang ve Xu (2015) tarafından yapılan bir çalışmada ise, bireylerin hızlı tahminde bulunma (beklenmeyen) davranışlarını tanımlamak için, maddelere verdikleri tepkiler ve tepki süreleri bilgilerini kullanarak karışık hiyerarşik bir model kullanılmıştır.

Hız-doğruluk ilişkileri, hız testleri, test alma stratejileri (örneğin ilerleme hızı ve tahmin etme davranışı) ve alt grup farklılıkları gibi konuları araştırmak için birkaç tepki süresi modeli önerilmektedir (Schnipke ve Scrams, 1999). van der Linden (2006) çalışmasında geliştirdiği log-normal tepki süresi modeli; van der Linden (2007) süresi modelini hiyerarşik bir çerçeveye genişletmiş ve van der Linden ve Guo (2008) ise hiyerarşik çerçeveye türetilmiş bu tepki süresi modelini bireyin testte yer alan belirli bir maddenin cevaplama süresini tahmin etmek için testin diğer maddelerindeki tepki sürelerinden yararlanılmasını sağlayan bir modele yükseltmişlerdir. Araştırmalar, madde ve birey parametre tahminlerini iyileştirmede, bireyselleştirilmiş bilgisayar destekli testlerde madde seçim kurallarının etkinliğini arttırmada tepki süresi verilerinin kullanımına odaklanan tepki süresi modellerini içerecek şekilde genişletilmiştir (van der Linden, 2009). Bireyin sınav motivasyonu veya sınava girerken bireylerin düşük çaba düzeylerine odaklanan tepki süresi modelleri üzerinde çalışmalar gerçekleştirilmiştir

(Wise & Kong, 2005; Wise ve De Mars, 2006; Wise, Ma, Kingsbury ve Hauser, 2010). Bu tür davranışlar, hızlı tepki verme veya test maddelerine tahminle tepki verme şeklinde ortaya çıkmaktadır. van der Linden ve van Krimpen-Stoop (2003) çalışmalarında, bir madde cevap vektörünün Madde Tepki Kuramı uyum analizini tamamlamadaki beklenmeyen tepkiyi kontrol etmek için loglineer tepki süresi modeli uygulamışlardır. Tepki süresini lognormal dağılıma sahip bir değişken olarak modellemiş ve bireysel tepki süresi vektörlerinin uygunluğunu değerlendirmek için bir dizi klasik ve Bayesian artık kontrol kullanmıştır. Meijer ve Sotaridona (2006) çalışmalarında sınava girmeden önce testte yer alan maddeleri önceden bilen bireyleri tespit etmek için bir cevap süresi geliştirmişlerdir. Etkili Tepki Süresi (ERT) adı verilen ölçü, bireyin maddeyi doğru cevaplama için gereken süre olarak tanımlanmıştır. Böylelikle madde ön bilgisine sahip birey ile madde ön bilgisine sahip olmayan birey bir birinden ayırt edilmiştir.

Bu bölümde ifade edildiği gibi, kopya çeken (kopyacı, olası kopyacı) bireylerin belirlenmesi için birçok istatistiksel yöntem geliştirilmiş ve geliştirilmeye devam etmektedir. Kullanılan istatistiksel yöntemlerde her ne kadar amaç olası kopyacı bireyin belirlenmesi olsa da bu yöntemler bazen kopya çekmeyen bireyleri de olası kopyacı birey olarak belirlemektedir. İstatistiklerin bu yanlış tespiti I. tip hata olarak adlandırılmaktadır. Bununla birlikte kopya çeken bireyleri de olası kopyacı birey olarak belirlemesi istatistiklerin gücünü göstermektedir. Alanyazında test hilesini belirlemeye yönelik istatistiklerin I. tip ve güçlerine dayalı çalışmalar yaygın olmakla birlikte en iyi performansı veren tek bir istatistiğin olmadığı görülmektedir (Armstrong ve Shi, 2009; Belov ve Armstrong, 2010; Belov, 2013, 2014a, 2016; Belov, Pashley, Lewis ve Armstrong, 2007; Karabatsos, 2003; Krimpen-Stoop ve Meijer, 2001; Shu, 2010; Sotaridona ve Meijer, 2002, 2003; Sünbül ve Yormaz, 2018a; Wollack, 1997; Wollack ve Cohen, 1998; Wollack, 2003, 2006; Yormaz, 2014; Yormaz ve Sünbül, 2017; Yormaz, 2019; Zopluoğlu ve Davenport, 2012; Zopluoğlu, 2016). Dolayısıyla olası kopyacı bireyleri belirlerken tek bir istatistiğin sonuçlarına bağlı kalınmamakta ve daha fazla kanıt ortaya koyulmasına önem verilmektedir. Alanyazındaki senaryoların çeşitliliği göz önünde bulundurulduğunda gerçek hayatta sık karşılaşılabilecek durumlar göz önünde bulundurulurken, olası kopyacıları belirlemede kullanılan yöntemlerin farklı şekillerde kullanıldığı senaryolar karşısındaki performanslarının incelenmesi önemli görülmektedir. Test uzunluğu, örneklem büyüklüğü, kopya çeken birey sayısı, kopya çekilen madde sayısı gibi değişkenlerin ve değişken koşullarının çeşitlendirildiği durumlar, test hilesini belirlemede kullanılan yöntemlerin farklılaştığı ve bu yöntemlerin aşamalı kullanıldığı

durumlara yer verilmesine önem verilmektedir. Ayrıca şüpheli bireylerin sınav salonundaki yerleri göz önünde bulundurularak olası kaynaklarla olan karşılaştırması da test hilesini belirleme analizindeki inceleme süresini azalttığı ve ortaya konulan kanıtın artmasını sağladığı için bu durumların da simüle edilmesi önemli görülmektedir. Belirtilen durumların içinde yer aldığı senaryolara dayalı simülasyonların yapılması, böylece simülasyonlardaki veriler kullanılarak istatistiklerin performansları ile ilgili bilgiler elde edilmesi ve elde edilen bilgiler ışığında analizlerin yapılması gerekli görülmektedir.

Amaç

Bu çalışmanın amacı madde ön bilgisine sahip olmadan cevap kopyalayarak test hilesi yapan bireyleri belirlemede, indekslerin aşamasız ve iki aşamalı olarak kullanımının I. tip hata ve güç düzeyleri üzerindeki etkisini çeşitli koşullar altında incelemektir. Bu genel amaç doğrultusunda aşağıdaki araştırma sorularına cevap aranmıştır.

1. Testin ortalama güçlüğü ve kopyacının yetenek düzeyinin; kopya ve benzerlik indeksleri, birey uyum istatistikleri ve *KL* indeksinin aşamasız olarak kullanıldığı durumda, indekslerin I. tip hata oranına olan
 - a. temel etkisi nasıldır?
 - b. ortak etkisi nasıldır?
2. Testin ortalama güçlüğü, kopyacının yetenek düzeyi ve kopya oranının; kopya ve benzerlik indeksleri, birey uyum istatistikleri ve *KL* indeksinin aşamasız olarak kullanıldığı durumda, indekslerin gücüne olan
 - a. temel etkisi nasıldır?
 - b. ortak etkisi nasıldır?
3. Testin ortalama güçlüğü ve kopyacının yetenek düzeyinin; kopya ve benzerlik indeksleri, birey uyum istatistikleri ve *KL* indeksinin iki aşamalı olarak kullanıldığı durumda, kopya ve benzerlik indekslerinin I. tip hata oranına olan
 - a. temel etkisi nasıldır?
 - b. ortak etkisi nasıldır?
4. Testin ortalama güçlüğü, kopyacının yetenek düzeyi ve kopya oranının); kopya ve benzerlik indeksleri, birey uyum istatistikleri ve *KL* indeksinin iki aşamalı olarak kullanıldığı durumda, kopya ve benzerlik indekslerinin gücüne olan

- a. temel etkisi nasıldır?
- b. ortak etkisi nasıldır?

Önem

Testlerin bu psikometrik özelliklerini etkileyen birçok faktör mevcuttur. Testte hile yapma da bu faktörler arasında yer almaktadır. Birey testte hile yaptığı zaman, testin ölçtüğü özellik bakımından kendisinde var olan durumu yansıtmayıp ilgili özelliğe dair yanıltıcı bilgiler edinmemize neden olmaktadır. Elde edilen bilgiler ışığında da verilen kararlar hatalı olmaktadır. Bu durum testin geçerliğine gölge düşürmektedir.

Alanyazında bireysel düzeyde test hilesini belirlemeye yönelik hem gerçek veriler hem de simülasyon çalışmaları yürütülerek kopya ve benzerlik indekslerinin ya da birey uyum istatistiklerinin test hilesini belirlemedeki I. tip hataları ve güçlerine yönelik gerçekleştirilmiş çok fazla çalışma olduğu görülmektedir (Armstrong ve Shi, 2009; Belov, 2011; Belov ve Armstrong, 2010; Krimpen-Stoop ve Wollack, 2003; Meijer ve Sijtsma, 2001; Shu, 2010; Sotaridona ve Meijer, 2002, 2003; Sünbül ve Yormaz, 2018; Wollack, 1997; Wollack ve Maynes, 2017; Wollack ve Cohen, 1998; Zopluoğlu, 2016; Zopluoğlu ve Davenport, 2012). Bununla birlikte test hilesini belirlemeye yönelik kullanılan indeks ve istatistiklerden en iyi olanının hangisi olduğuna yönelik genel bir kabul yoktur. Bunun nedeni ise test hilesini kontrol altına alabilecek kesin kurallarının olmamasıdır. Sık rastlanabilecek tepki örüntüleri ya da durumlar göz önünde bulundurularak senaryolar üretilmektedir. Üretilen bu senaryolara dayalı olarak indekslerin olası kopyacı bireyleri belirlemede gösterdikleri performansları ile ilgili bilgilere dayalı olarak analizlerin yapılması önemli görülmektedir.

Geçekleştirilen simülasyon çalışmalarında test hilesini belirlemeye yönelik indekslere ait senaryolarda olası kopyacı-kaynak cevapları eşleşmesi dışında olası kopyacı bireylerin kendi cevap örüntüsünü baz alarak analiz eden birey uyum istatistiklerine, bireylerin sınav salonundaki oturma planları dikkate alınarak yer verildiği görülmüştür (Magis, Raïche ve Beland, 2012; Sünbül ve Yormaz, 2018a, 2018b; Yormaz ve Sünbül, 2017; van der Ark, Emons ve Sijtsma, 2008). İfade edilen durumlar dikkate alınarak bu araştırmada, çeşitli kopya ve benzerlik indekslerinin (ω indeksi ve *GBT* indeksi) yanı sıra birey uyum istatistikleri (l_z ve l_z^*) ve testin iki farklı alt maddeleri kümesindeki performanslarının karşılaştırılması yöntemlerinden biri olan *KL* indeksi kullanılmıştır. Ayrıca çalışmada kopya çeken bireylerin belirlenmesinde eldeki kanıtın

arttırılması için birey uyum indeksleri ve *KL* indeksi ile birlikte kopya ve benzerlik indeksleri de kullanılarak iki aşamadan oluşan bir tespit de yapılmıştır. Başka bir ifadeyle, çalışmada üretilen çeşitli senaryolar altında, alanyazında cevap kopyalamadan kaynaklı test hilesini belirlemeye yönelik kullanılan istatistiklerin gücünü ve I. tip hatasını etkileyen faktörlerle birlikte kopya ve benzerlik indekslerinin (ω indeksi ve *GBT* indeksi), birey uyum istatistiklerinin (l_z ve l_z^*) ve *KL* indeksinin aşamasız kullanıldığı durumlarda istatistiklerin kopya çeken bireyin belirlenmesindeki I.Tip hata ve güçlerinin; iki aşamalı kullanıldığı durumlarda ise kopya ve benzerlik indekslerinin (ω indeksi ve *GBT* indeksi) I.Tip hata ve güçlerinin nasıl değiştiği incelenmiştir. Çalışmanın hem geniş ölçekli sınavlar hem de tüm sınavlardaki bireysel test hilesini tespit etmede kullanılan istatistiklerin hangilerinin, hangi şartlar altında kullanılmasının iyi olacağına yönelik bilgileri sunduğundan dolayı ölçme ve değerlendirme çalışmalarına katkıda bulunulacağı düşünülmektedir.

Sınırlılıklar

Çalışmada testin iki farklı alt maddeleri kümesindeki performanslarının karşılaştırılması yöntemlerinden biri olan Kullback-Leibler uzaklığı (*KL*) indeksi kullanılarak bireyler cevap kopyalamadan kaynaklı test hilesi göstermiştir/göstermemiştir şeklinde sınıflandırılmıştır. *KL* indeksinin sınıflaması için ise değişimlenen faktörler altında R programında yer alan “*OptimalCutpoints*” (Raton-Lopez, Rodriquez-Alvarez, Suarez- Cadarso ve Sampedro-Gude, 2014) paketinde var olan çeşitli kesme puanları elde edilerek bu kesme puanları kullanıldığında *KL* indeksinin gücü ve I. tip hatasına yönelik araştırmacı tarafından inceleme yapılmıştır (Bu konuda daha ayrıntılı bilgi için bkz. EK4, EK5). Elde edilen sonuçlar ışığında *KL* indeksi için “*OptimalCutpoints*” paketinde yer alan Youden İndeksi ile elde edilen kesme puanlarının kullanılmasına karar verilmiştir.

Çalışmada *KL* indeksi değerlerinin hesaplanması için öncelikle yetenek parametreleri kullanılmıştır. Yetenek parametrelerinin kestirilmesi için R programında yer alan “*irtoys*” paketinden yararlanılmıştır. *KL* indeksi değerlerinin hesaplanması için yetenek parametreleri hesaplanması adımı “*irtoys*” paket programıyla sınırlandırılmıştır.

Gerçek hayatta sınav salonu büyüklükleri 80, 100, 120, 150, 170, 200 kişilik v.b. şeklinde farklılaşmaktadır. Bununla beraber test uzunlukları kopya çalışmalarında özellikle 40 ve 80 olarak değişimlenmesinin yanı sıra 10, 20, 30 v.b. maddeden oluşmaktadır. Ayrıca testlerin ortalama güçlükleri kolay, orta ve zor olacak şekilde

değişmektedir. Yukarıda belirtilen koşullarla birlikte Şahin (2012) çalışmasında elde ettiği sonuçlar göz önünde bulundurularak, bu çalışmanın yürütüldüğü verilerin üretilmesi için gerçek hayatta karşılaşılan durumların yanı sıra kullanılan indekslerin temel aldıkları ölçme modellerine olan uygunlukları dikkate alınarak, örneklem büyüklüğü ve testin uzunluğunun model veri uyumuna etkisini ortaya çıkarmak amacıyla araştırmacı tarafından inceleme yapılmıştır (Bu konuda daha ayrıntılı bilgi için bkz,EK3.). Elde edilen sonuçlara dayalı olarak çalışma Rasch model kullanılarak 150 örneklem büyüklüğü ve 30 test uzunluğu ile sınırlandırılmıştır.

Çalışmada olası kopyacının oturduğu yere göre önünde, sağında ve solunda yer alan bireylerden yalnızca biri kaynak olarak belirlenmiştir. Kaynak olarak seçilen bireyin cevapları çalışmada yer alan kopya çekme oranları düzeyinde kopyalanarak olası kopyacının cevapları değiştirilmiştir. Çalışma sadece olası kopyacının yakınlarındaki (önünde, sağında ve solunda yer alan) kaynaktan cevap kopyaladığı senaryosu ile sınırlandırılmıştır.

Çalışma, kullanılan bilgisayar programı ve paket programlarıyla üretilen simülasyon verileri; araştırmaya dahil edilen kopya ve benzerlik indeksleri, birey uyum indeksleri ve KL indeksi ile analiz yöntemleri sınırlandırılmıştır. Bununla birlikte seçilen koşullar (kopyacının yetenek düzeyi, test maddelerinin güçlüğü ve kopya çekme oranları) ve bu koşulların düzeyleri ile sınırlandırılmıştır.

Tanımlar

Aşamasız: Kopya ve benzerlik indeksleri (ω indeksi ve *GBT* indeksi) ve birey uyum istatistikleri (l_z ve l_z^*) ve testin iki farklı alt maddeleri kümesindeki performanslarının karşılaştırılması yöntemi olan Kullback Leibler uzaklığı (*KL*) indeksinin her biri ayrı ayrı kullanılarak kopya çeken bireyleri belirlediği durum aşamasız olarak adlandırılmıştır.

İki aşamalı: Önce birey uyum istatistikleri ve *KL* indeksi sonra kopya ve benzerlik indekslerinin kullanıldığı durum ise iki aşamalı olarak adlandırılmıştır. Başka bir ifade ile önce *KL* sonra ω indeksi, önce *KL* sonra *GBT* indeksi, önce l_z sonra ω indeksi, önce l_z sonra, *GBT* indeksi, önce l_z^* sonra ω indeksi, önce l_z^* sonra, *GBT* indeksi olacak şekilde kullanılan aşamalı olarak adlandırılmıştır.

BÖLÜM 2

KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

Bu bölümde çalışmada bireysel düzeyde test hilesini tespit etmek için kopya ve benzerlik indeksleri (ω indeksi ve GBT indeksi), birey uyum istatistikleri (l_z ve l_z^*) ve testin iki farklı alt maddeleri kümesindeki performanslarının karşılaştırılması yöntemlerine kısaca yer verilmiştir. Ayrıca çalışmada kullanılan kopya ve benzerlik indeksleri (ω indeksi ve GBT indeksi), birey uyum istatistikleri (l_z ve l_z^*) ve Kullback-Leibler uzaklığı indeksi kullanılacağı için bu indekslere yönelik daha fazla bilgi yer almıştır.

Kopya ve Benzerlik İndeksleri

Testteki potansiyel test hilesini belirlemeye yönelik çalışmalar Bird'e (1927) dayanmaktadır. Ampirik sıfır dağılımını ilk olarak Bird (1927, 1929) kullanmıştır. Yanlış sayısı aynı olan farklı yerlerde oturan bireyleri seçkisiz olarak ikili karşılaştırıp tespit ettikten sonra aynı yanlış sayısını kullanarak sıfır dağılımı oluşturmuştur. Bird oluşturduğu bu dağılımın ortalamasını referans olarak kullanmıştır. Bu ortalamaya göre karşılaştırılan ikililerin aynı yanlış sayıları ortalamadan daha yüksek bir değere sahipse, bu durumu bireylerin umulmadık bir uyuşma içinde olduğunun göstergesi olarak yorumlamıştır. Angoff (1972) çalışmasında sekiz farklı indeks geliştirmiştir. Ancak yapılan araştırmalar ve kabul edilen olası kopyacılar baz alındığında Angoff'un (1972) çalışmasındaki B ve H indekslerinin kopya belirlemede daha güçlü olduğu görülmüştür. Angoff'un B indeksi; W_i ve W_j , i'inci ve j'inci bireyler için yanlış sayısı/puanı olmak üzere W_{ij} 'yi bağımlı değişken olarak, $W_i * W_j$, ise bağımsız değişken olarak kullanılmıştır. Angoff (1972) çalışmasında bir regresyon yaklaşımı kullanmak yerine, $W_i * W_j$ 'yi kopya çekmeyen ikililer için hesaplamayı, $W_i * W_j$ aralığını eşit aralıklara bölerek, her aralık için W_{ij} 'nin ortalama ve standart sapmasını hesaplamıştır. Kopya çeken bir ikili olduğunda, bu ikilinin $W_i * W_j$ 'si hangi aralık grubuna ait olduklarını belirlemek için hesaplanarak, iki birey arasında gözlenen W_{ij} , elde edilen belirli aralığın

W_{ij} 'sinin ortalama ve standart sapması t istatistiği ile karşılaştırılır. Angoff'un H indeksi, Angoff'un B 'sine benzemektedir. Ancak ilgili bağımlı ve bağımsız değişkenler farklılaşmaktadır. Çünkü Angoff'un H 'si art arda birçok maddeden kopya çekildiği belirli bir tür test hilesini hedeflemektedir. İlgili bağımlı değişken, iki bireyin arasında aynı şekilde işaretlenmiş en uzun yanlış cevapları şeklinde ifade edilirken; bağımsız değişken, olası kaynağın yanlış ve ihmal edilen cevaplarının sayısı olarak belirtilmektedir.

Bu indeksler sıfır dağılımının oluşturulmasının zor olması nedeniyle araştırmacılar tarafından kullanılmamaktadır. Çünkü sıfır dağılımı oluşturabilmek için bireyler hakkında gerekli bilgiye (sınav yerleri vb.) sahip oldukları büyük bir veriye ihtiyaçlarının olmasının yanı sıra oluşturulan sıfır dağılımı sadece o testte kullanılmaktadır. Başka bir ifadeyle genelleştirilmiş ya da bir başka durum yerine kullanılamamaktadır.

Sotaridona ve Meijer (2002), herhangi bir test modeline dayanmayan, Frederick Kling geliştirilen K indeksi ve K indeksinin farklı türevlerinin istatistiksel özelliklerini araştırdığı ve bu indekslerle birlikte madde tepki kuramı modellemesine dayanan ω indeksinin de şüpheli bireylerin tespit oranlarını karşılaştırmıştır. Çalışmada olası kopyacı ile kaynağın eşleşen doğru cevap sayısı arttığında K indeksinin duyarlılığının azaldığını ve indeksin olası kopyacı belirleme gücünün düşük olduğunu bulmuşlardır.

Sotaridona ve Meijer (2003) çalışmasında eşleşen doğru ve yanlış cevapları dikkate alan S_2 indeksini, K indeksiyle matematiksel olarak benzer forma sahip olası kopyacı ile kaynağın eşleşen yanlış cevap sayılarını Poisson dağılımıyla modelleyen S_1 indeksi ve K_2 indekslerini karşılaştırmıştır. S_1 indeksinin olası kopyacıları tespit etmedeki oranı K_2 indeksine göre daha yüksek olduğu sonucuna ulaşılmıştır. Bununla birlikte S_2 indeksinin S_1 indeksine göre performansının daha iyi olduğu görülmüştür. Ancak Sınıflamalı Tepki Modeli parametrelerinin güvenilir bir şekilde kestirildiği durumda ω indeksinin küçük örneklemelerde, olası kopyacıların yer aldığı tüm yetenek seviyelerindeki cevap kopyalayan bireyleri tespit etmede en iyi sonuçların elde edildiği ifade edilmiştir.

Klasik Test Kuramı'na dayanan kopya belirleme indeksleri, iki birey arasındaki beklenen tepki benzerliğini karşılaştırma üzerine geliştirilmiştir. Bununla birlikte tüm bireylerin tepkilerini birbirleriyle ikili olarak karşılaştıran indeksler dışında, bireyleri olası kopyacı ve kaynak şeklinde ele alarak karşılaştıran g_2 indeksi geliştirilmiştir. g_2 indeksi, maddenin çeldiricilerini seçme olasılıklarını hesaba kattığından dolayı Klasik Test Kuramı'na dayalı diğer indekslere göre daha avantajlıdır. Ancak her ne kadar test puanına bağlı olarak olasılıkların hesaplanmasına uygun olsa da g_2 indeksi, veri

kümesinden bağımsız değildir. Bununla birlikte g_2 indeksi her örneklem büyüklüğü için hesaplanmaktadır (Wollack, 1997). Ancak, Klasik Test Kuramı'nda madde istatistikleri bireylerin yetenek düzeylerine bağımlı olduğu için iki birey arasındaki beklenen benzerlik karşılaştırması dışında testi alan bireylerin performanslarından etkilenmektedir. ω indeksi, g_2 indeksi gibi her bir tepkinin olasılığını hesaplayan ve Sınıflamalı Tepki Modeli'ne dayalı bir test istatistiğidir. İkili puanlanan madde tepki modellerinde madde verilen doğru cevabın (yanlış cevabın) olasılığı bireyin yetenek düzeyi ve maddenin parametresini baz alan bir fonksiyon yardımıyla hesaplanmaktadır. Cevap kopyalamasından kaynaklı test hilesi çalışmalarında bireyin doğru ya da yanlış cevabı dışında aynı seçeneği seçmiş olmasıyla da ilgilenilmektedir. Dolayısıyla Sınıflamalı Tepki Modeli'nin baz alınarak her bir maddenin k alternatifinin seçilme olasılığı bireyin yetenek düzeyi kullanılarak kestirilebilmektedir. Bununla beraber ikili puanlamanın yapıldığı (1-0) durumları için de lojistik modeller çerçevesinde uyarlanmış türevleri de bulunmaktadır (Zopluoğlu, 2016).

g_2 ve ω indekslerinin güç ve I. tip hata oranları sonuçlarını içeren simülatif çalışmaların bir kısmında g_2 indeksinin I. tip hatası nominal değer üzerinde bulunmuştur (Hanson, Harris, ve Brennan, 1987; Wollack, 1997; Wollack ve Cohen, 1998), ω indeksinin I. tip hata oranı üzerinde genellikle iyi bir kontrole sahip olduğu ve olası kopyacı bireylerin tespit edilmesinde iyi bir güce sahip olduğuna ulaşılmıştır (Sotaridona ve Meijer, 2002, 2003; Wollack, 1997; Wollack ve Cohen, 1998).

g_2 ve ω indekslerinin ikisi de tüm maddelerden elde edilen bilgileri kullanmaktadırlar. İki indekste de olası kopyacı bireylerin tespiti için birey çiftlerinin (olası kopyacı (C)-kaynak(S)) eşleşen cevaplarından yararlanılmaktadır. Kopya çektiğinden şüphelenilen bireyin kaynaktan kopyaladığı i maddesini cevaplama olasılığı $P_C(U_{is})$ (u_{is} kaynağın i 'inci maddeye verdikleri cevabı) olasılığının hesaplanmasında ki fark iki indeksi birbirinden ayırmaktadır. g_2 indeksi bu olasılığı hesaplama prensibi Klasik Test Kuramına dayalı olarak, çeldiricinin güçlüğü ve şüpheli bireyin doğru sayısının sınavdaki tüm bireylerin doğru cevaplama sayısı ortalamasına oranına dayalı gerçekleştirilirken, ω indeksinde ise Madde Tepki Kuramına dayalı gerçekleştirilmektedir (Wollack, 2003).

ω indeksiyle olası kopyacı ve kaynağın gözlenen aynı cevapları ile şansa bağlı benzerliklerin karşılaştırılarak cevap kopyalamadan kaynaklı test hilesinin olasılığı belirlenmektedir. Bir başka ifade ile ω indeksi, iki bireyin (C- S) cevap eşleşmeleri (hem

doğru hem de yanlış) arasındaki standartlaştırılmış farkın bir ölçüsünü sunmaktadır. Bunun için olası kopyacı (C) ve kaynak (S) çifti için aynı cevaplanan maddelerin sayısı;

$$I_i = \begin{cases} 1, & u_{iC} = u_{iS} \\ 0, & u_{iC} \neq u_{iS} \end{cases} \quad (1)$$

$$h_{CS} = \sum_{i=1}^n I_i \quad (2)$$

eşitlik (2)' de ifade edildiği şekilde bulunmaktadır. Eşitlik (1)'de yer alan i ; i 'inci maddeyi, u_{iC} ve u_{iS} sırasıyla olası kopyacının ve kaynağın i 'inci maddeye verdikleri cevabı ifade etmektedir.

$P_C(u_{iS})$; olası kopyacının yetenek düzeyi, kaynağın cevapları ve madde parametreleri koşulu altında olası kopyacının kaynaktan aldığı cevapların olasılığını göstermekle beraber kaynağın yeteneği ve doğru bir şekilde kestirilen olası kopyacının yetenek düzeyi ile madde parametreleri koşulları altında h_{CS} bağımsız Benoulli değişkenlerinin toplamı olup (Wollack, 1997), olası kopyacının kaynaktan aldığı cevapların olasılığı ve varyans aşağıda yer alan eşitlik (3) yardımıyla hesaplanmaktadır (Wollack, 2006).

$$\sigma_{h_{CS}} = \sqrt{\sum_{i=1}^n P_C(u_{iS})[1 - P_C(u_{iS})]} \quad (3)$$

Böylece ω aşağıdaki eşitlik (4) ile hesaplanmaktadır;

$$\omega = \frac{h_{CS} - \sum_{i=1}^n P_C(u_{iS})}{\sigma_{h_{CS}}} \quad (4)$$

Genelleştirilmiş Binom Testi (*GBT*) (Sotaridona ve van der Linden, 2006), sınava giren iki bireyin cevapları arasındaki aynı doğru ve yanlış cevapların sayısı için bileşik binom dağılımını kullanır. P_{M_k} ; i 'inci ve j 'inci bireyin k 'inci maddedeki eşleşme olasılığı olmak üzere bu olasılık eşitlik (5)'teki gibi ifade edilmektedir.

$$P_{M_k} = \sum_{o=1}^0 P_{iko} * P_{jko} \quad (5)$$

Eşitlik (5)'te yer alan P_{iko} ve P_{jko} terimleri sırasıyla i 'inci ve j 'inci bireylerin k 'ncı maddedeki O 'uncu alternatifi seçme olasılığıdır. İki tepki vektörünün K maddeleri üzerindeki n eşleşmenin gözlenme olasılığı ise eşitlik (6)'da gösterilmiştir.

$$f_K(n) = \sum \left(\prod_{k=1}^K P_{M_k}^{u_k} (1 - P_{M_k})^{1-u_k} \right) \quad (6)$$

u_k eşitlikte yer alan u_k terimi iki bireyin k maddesine aynı cevabı verdiklerinde 1 değerini alırken aksi durumda 0 değerini almaktadır. $w_{ij} + R_{ij}$ gözlenme olasılığı ya da K maddelerinde daha fazla eşleşmenin gözlenme olasılığı eşitlik (7)'de görülmektedir.

$$\sum_{n=w_{ij}+R_{ij}}^K f_K(n) \quad (7)$$

ω indeksi, iki tepki vektörü arasındaki aynı doğru ve yanlış cevapların sayısı için bileşik binom dağılımına normal bir yaklaşımdır (Cizek ve Wollack, 2017). ω indeksi bir kopya istatistiğidir (asimetrik), GBT ise bir benzerlik istatistiğidir (simetrik). Başka bir ifade ile ω iki yönde hesaplanarak benzerlik için kullanılabilir (X olası kopyacı-Y kaynak, Y olası kopyacı-X kaynak).

Birey Uyum İstatistiği

Test hilesini belirlemede kullanılan bir başka yöntem ise birey uyum istatistikleridir. Aşağıda bu çalışma için kullanılacak olan birey uyum istatistiklerine yer verilmiştir.

Birey-uyum istatistikleri veya uygunluk ölçümü, bir tepki modelinin belirli bir test modeline uyumunu değerlendirmek için kullanılan istatistiksel yöntemleri ifade etmektedir. Madde Tepki Kuramı'na dayanarak beklenen tepki örüntüsü ile uyuşmayan madde puanı örüntülerini tanımlamak için bir dizi birey-uyum istatistiği önerilmiştir (Drasgow, Levine, ve Williams, 1985; Molenaar ve Hoijtink, 1990). Meijer ve Sijtsma (2001) çalışmalarında, kırktan fazla birey-uyum istatistiği arasından en fazla ilgi gören istatistiğin l_z istatistiği olduğunu belirtmişlerdir. Karabatsos (2003) çalışmasında l_z istatistiğinin beklenmeyen tepki örüntüsünü belirlemede en güçlü birey-uyum istatistiği olduğunu ifade etmiştir. Drasgow, Levine ve McLaughlin (1987) çalışmalarında üç

parametrelili lojistik (*3PLM*) modelini kullanarak l_z , *ZU*, *ZW*, *C*, *JK*, *O/E*, *ECI2z* ve *ECI4z* birey-uyum istatistiklerinin performansını karşılaştırmışlardır. Çalışmanın sonucunda beklenmeyen hem yüksek hem düşük puan alan bireyleri belirlemede en uygulanabilir birey-uyum istatistiğinin l_z olduğunu göstermişlerdir. Li ve Olejnik (1997) ise Rasch modelini kullanarak l_z , *ZU*, *ZW*, *ECI2z* ve *ECI4z* istatistiklerinin performanslarını karşılaştırmışlardır. Sonuç olarak l_z istatistiğinin uyumsuz madde-puan örüntüsünü belirlemede en güçlü istatistik olduğunu belirtmişlerdir. Nering ve Meijer (1998) l_z istatistiği ile birey tepki fonksiyonu (*BTF (PRF)*) yönteminin performansını karşılaştırmış ve l_z istatistiğinin çoğu durumda *BTF* yönteminden daha iyi performans gösterdiğini bulmuştur. Gerçekleştirilen çalışmalara rağmen l_z istatistiğinin gücünü sunan en etkili senaryonun nasıl olacağı hala net değildir. Reise ve Due (1991) çalışmalarında l_z istatistiğinin, çeşitli güçlük seviyesi ve küçük tahmin parametrelerine sahip uzun bir test için en etkili olduğunu belirtmişlerdir. Reise (1995), l_z istatistiğinin modele uymayan tepkileri belirlemedeki gücünü araştırmak için iki parametrelili lojistik (*2PLM*) modeli kullanmış ve l_z 'nin farklı puanlama stratejilerinden nasıl etkilendiğini incelemiştir. Yetenek seviyeleri ile madde güçlük parametreleri arasındaki fark büyük olduğunda en iyi belirleme oranlarına ulaşıldığı bulunmuştur. Ancak Amstrong, Stoumbos, Kung ve Shi (2007) çalışmalarında bu durumun her zaman böyle olmadığını göstermiştir.

Bu çalışmada alan yazındaki çalışmalar dikkate alınarak beklenmeyen tepki örüntülerini belirlemek için birey uyum indekslerinden l_z ve l_z^* istatistiği seçilmiştir.

Birey uyum analizi, daha önce bahsedildiği gibi bir örneklemede beklenmeyen cevap davranışı sergileyen test katılımcılarını tespit etmek için kullanılabilir. Bir ölçme aracına verilen beklenmeyen cevapları tetikleyebilecek birçok psikolojik süreç vardır. Hile, tepki kurulumu, maddelerin son derece yaratıcı yorumu şeklinde beklenmeyencevaplama davranışlarının bazı örnekleridir (Meijer, 1996). Beklenmeyen farklı cevap davranışlarının var olmasıyla birlikte bu psikolojik süreçlerin birbirinden ayırt edilmesi kolay olmamaktadır. Başka bir ifadeyle, farklı psikolojik süreçlerin benzer madde-puanı örüntülerine yol açması mümkün olabilmektedir. Bu nedenle birey-uyum sonuçların yorumlanması genellikle kolay olmamaktadır. Genel bir kural olarak, birey-uyum analizleri belli bir beklenmeyen cevap davranışına dair kesin bir kanıt sağlamamaktadır. Dolayısıyla, beklenmeyen cevap davranışları gözlemlendiğinde, testi alan bireylerle görüşmeler yapılması (Meijer ve Tendeiro. 2014) veya birey-uyum

indekslerini başka deęişkenlerle ilişkilendirme (Conijn, 2013) gibi çözümlere gidilmesi gerektięi belirtilmektedir.

x_n madde puan örüntüsü ve P_i 1-2 ve 3 parametrelili lojistik modeli göstermek üzere, iki kategorili tepkilerde olabilirlik logaritması eşitlik (8)'de ifade edilmiştir.

$$\log L(\theta_n) = \sum_{i=1}^I x_{ni} \ln P_i(\theta_n) + (1 - x_{ni}) \ln(1 - P_i(\theta_n)) \quad (8)$$

Eşitlik (9)'da ifade edilen l_0 birey uyum indeksidir (Levine ve Rubin, 1979). l_0 indeksinin bazı sınırlamaları vardır, çünkü standart değildir veya bilinen bir madde tepki kuramı modeline uygun bir dağılımı yoktur. Ancak, Drasgow ve diğerleri (1985) l_z olarak gösterilen standartlaştırılmış normal bir l_0 versiyonu önermiştir. l_z deęerleri l_0 'a benzer şekilde yorumlanır, yani (negatif) l_z deęerleri ne kadar küçükse uyumsuzluk göstergesi o kadar güçlü olmaktadır.

$$l_z = \frac{l_0 - E(l_0)}{[Var(l_0)]^{1/2}} = \sum_{i=1}^n \frac{l_0 - \bar{l}}{[Var(l_0)]^{1/2}} \quad (9)$$

olmak üzere, $\bar{l} = E(l_0)/n$ 'dir. Eşitlik (9)'da yer alan l_0 istatistięinin beklenen ve varyansı eşitlik (10) ve (11)'de ifade edilmiştir.

$$E(l_0) = \sum_{i=1}^n [P_i(\theta) \ln(P_i(\theta)) + (1 - P_i(\theta)) \ln(1 - P_i(\theta))] \quad (10)$$

$$Var(l_0) = \sum_{i=1}^n P_i(\theta)(1 - P_i(\theta)) \left\{ \ln \left[\frac{P_i(\theta)}{1 - P_i(\theta)} \right] \right\}^2 \quad (11)$$

l_z indeksi, asimptotik normal dağılımlıdır. Ayrıca l_z istatistięi sadece gerçek deęerler kullanıldığında geçerli olmaktadır (Molenaar ve Hoijtink, 1990), bu pratikte gerçekçi olmayan bir gereklilik özellięi taşımaktadır. Denklem (8)'de kestirilen parametreler ile l_z varyansının azaldığı gözlenmiştir (Molenaar ve Hoijtink, 1990; Nering, 1995; Reise, 1995). Dolayısıyla, Snijders (2001) çalışmasında kestirilen parametrelerinin örnekleme deęişkenliğini dikkate alan l_z^* olarak bilinen düzeltilmiş bir

istatistik önermiştir. Magis, Raïche ve Béland (2012) çalışmalarında l_z^* istatistiğini eşitlik (12)'deki gibi ifade etmiştir.

$$l_z^* = \frac{l_0(\hat{\theta}) - E(l_0(\hat{\theta})) + c_t(\hat{\theta})r_0(\hat{\theta})}{[Var(l_0)]^{1/2}} \quad (12)$$

Aşağıdaki eşitlik (13), eşitlik (14) ve eşitlik (15)'te l_z^* istatistiğinin hesaplanmasında kullanılan sırasıyla varyans ve değiştirilmiş ağırlıklar terimlerinin eşitlikleri verilmiştir.

$$Var(l_0(\hat{\theta})) = \sum_{i=1}^n P_i(\theta)(1 - P_i(\theta))\tilde{w}_i(\theta)^2 \quad (13)$$

$$\tilde{w}_i(\hat{\theta}) = w_i(\hat{\theta}) - c_t(\hat{\theta})r_0(\hat{\theta}) \quad (14)$$

$$c_t(\theta) = \frac{\sum_{i=1}^n a_i P_i(\theta)(1 - P_i(\theta))w_i(\theta)}{\sum_{i=1}^n a_i^2 P_i(\theta)(1 - P_i(\theta))} \quad (15)$$

Kullback Leibler Uzaklığı İndeksi

Bu çalışmada test hilesi yapan bireyi belirlemek için testin iki farklı alt maddeleri kümesindeki performanslarının karşılaştırılması yöntemlerinden biri olan Kullback Leibler uzaklığı indeksi (Kullback Leibler Divergence (*KL*)) kullanılmıştır. Aşağıda *KL* indeksine yönelik bilgilere yer verilmiştir.

KL uzaklığı indeksi teori ve pratikte yaygın olarak uygulanmaktadır. *KL* negatif değer almaz ve yalnızca iki dağılım aynı ise sifıra eşittir. Uzaklığın tanımı, ayrık ve sürekli dağılımlar için geçerli olma özelliğine sahiptir. *KL* ayrıca bir bilgi ölçüsü olarak da düşünülebilir. Bilinen birçok bilgi ölçüsü *KL*'den türetilmektedir. Fisher Bilgileri (*FI*) (Chang ve Ying, 1996) *KL*'nin ikinci türevi değerlendirilerek bulunmaktadır. *KL* genellikle simetrik değildir, ancak Karşılıklı Bilgi (*MI*) (Cover ve Thomas, 1991) durumunda simetri geçerlidir. Chang ve Ying (1996) çalışmalarında, bilgisayar ortamında bireye uyarlanmış testte (BOBUT) madde seçme kriteri olarak *KL* uzaklığını kullanmışlardır. Maksimum olabilirlik yöntemi (MLE) kullanıldığında, madde sayısı arttıkça *KL* ve Fisher Bilgisi (*FI*) madde seçme yöntemlerinin yakınsadığını ifade etmişlerdir. BOBUT' un başlangıç aşamasında MLE yönteminin genellikle yeteneklerin

iyi bir tahmincisi olmadığını bununla birlikte KL 'nin bilgiyi tahmin etmek için testin tamamında bütünsel bir yaklaşım sağladığı sonucunda varmıştır. Eggen (1999) çalışmasında, BOBUT ile sıralı olasılık oranı testi kullanılarak sınıflandırma yaparken madde seçimi için KL 'yi kullanmıştır. Xu ve Douglas (2006) BOBUT'ta parametrik olmayan MTK modellerinde madde seçimi için Fisher Bilgisi yerine KL 'yi kullanmıştır. Çünkü Fisher İndeksi için gerekli olan türevin parametrik olmayan MTK ile problemlili olduğunu, ancak KL 'nin bir türev almayı gerektirmediğini belirtmişlerdir. Bireylerin uygun olmayan tepki örüntülerini tespit etmek için KL 'nin kullanımı ilk olarak Belov, Pashley, Lewis ve Armstrong (2007) tarafından önerilmiştir (Belov, 2014a). Belov (2013) toplu kopya üzerine yaptığı çalışmasında ise KL yaklaşımını genişletmiştir. Ayrıca cevap kopyalayan bireyleri tespit etmek için Belov ve Armstrong (2010) KL 'yi kullanmışlardır.

$g(x)$ ve $h(x)$ iki olasılık yoğunluk fonksiyonu olmak üzere; bu iki fonksiyon arasındaki farklılığı değerlendirmek için çeşitli yaklaşımlar kullanılabilir (Pardo, 2006). İki dağılım arasındaki farklılık eşitlik (16)'da yer alan KL ifadesi (Cover ve Thomas, 1991; Kullback ve Leibler, 1951) ile hesaplanır:

$$D(g||h) = \int_{-\infty}^{+\infty} g(x) \ln \frac{g(x)}{h(x)} dx \quad (16)$$

İlgili Araştırmalar

Bu bölümde alanyazında gerçekleştirilmiş test hilesi yapan bireyleri tespit etme önce yurt dışı ardından Türkiye'de yapılan çalışmalara kronolojik olarak yer verilmiştir.

Karbatsos (2003) çeşitli koşullar altında birey uyum istatistiklerinin performansını belirlemeye yönelik bir çalışma gerçekleştirmiştir. Çalışmada beklenmeyen davranış türü olarak adlandırılan (kopya çekme, yaratıcı cevap verme, tahmin ederek cevaplama ve rastgele cevap verme) beş farklı beklenmeyen davranışı sergileyenlerin veri kümelerindeki dört farklı oranı (%5, %10, %25 ve %50) ve test uzunluğu 17, 33 ve 65 madde olmak üzere bu değişen koşullardan 60 farklı veri kümesi hem normal davranış hem de beklenmeyen davranışı içirecek şekilde 500 yineleme ile simüle edilmiştir. Bu veri kümelerinin her birinde 36 farklı birey uyum parametresinin performansları incelenmiştir. HT , $D(\theta)$, C , MCI ve $U3$ birey uyum istatistiklerinin çeşitli koşullar altında beklenmeyen davranışı belirlemeye yönelik en iyi performansı gösterdiği sonucuna ulaşmıştır.

Sotaridona ve Meijer (2003) çalışmalarında örneklem büyüklüğü, test uzunluğu ve kopya çekme oranına göre geliştirdikleri benzerlik indeksleri $S1$ ve $S2$ ile birlikte K , \overline{K}_2 , ve ω indekslerini de kullanarak karşılaştırma yapmışlardır. Çalışmaları sonucunda $S1$ indeksinin \overline{K}_2 indeksine göre, $S2$ ve ω indeksinin diğerlerine göre kopya belirlemede daha güçlü olduğu ortaya çıkmıştır. Madde tepki kuramına dayalı Sınıflamalı Tepki Modeli (NRM) kullanılarak madde parametreleri güvenilir bir şekilde kestirildiği durumlarda ise kopyacının her yetenek düzeyi için ω indeksinin diğer indekslere göre kopya belirlemede daha duyarlı sonuçlar verdiğini gözlenmiştir. Ayrıca ω indeksinin küçük örneklerde kopya belirlemede kullanılabilir bir indeks olduğu sonucuna ulaşmışlardır. $S1$ ve $S2$ indekslerinin küçük örneklerde kopya belirleme gücünün düşük olduğu vurgulanmıştır.

Wollack (2006) çalışmasında çeşitli α düzeylerinde (0.01, 0.005, 0.001 ve 0.0005) \overline{K}_2 , $S1$, $S2$, ω , H ve B indekslerini karşılaştırmıştır. \overline{K}_2 , $S1$, $S2$, ω ve B indekslerinin tüm α düzeylerinde I. tip hata oranının düşük olduğunu, H indeksinin ise yüksek değerler aldığını belirtmiştir. Ayrıca ω indeksinin karşılaştırılan tüm indekslerden cevap kopyalamadan kaynaklı test hilesi yapan bireyi belirleme gücünün daha yüksek olduğunu, $S2$ indeksinin ise ω indeksi dışında diğerlerine göre daha güçlü olduğunu ortaya koymuştur. Tüm indeksler bir arada dikkate alındığında ise $\omega - H^*$ ikilisinin diğer ikililere göre daha güçlü olduğunu göstermiştir. Bununla birlikte, madde tepki kuramına dayanan ω indeksinin kullanılmadığı durumlarda tercih edilebilecek en güçlü indeksin $S2$ indeksi olduğunu belirtmiştir.

Küçük ölçekli sınavlarda cevap kopyalamadan kaynaklı test hilesinde bulunan bireylerin belirlenmesinde iki türlü indeks kullanılmaktadır. Bunlardan biri olası kopyacı ve kaynağı bilinen belirleme yaparken diğerinde ise olası kopyacının ve kaynağın kim olduğunu ortaya çıkararak gerçekleştiren istatistiklerdir. Bu iki istatistiğin aynı test sonucunda kullanılmasına yönelik olarak van der Ark, Emons ve Sijtsma (2008) çalışmalarında alternatif testleri kullanmışlardır. Çalışmada hem gerçek hem de simülasyon çalışması gerçekleştirilmiştir. Gerçek veriler için kurgulanan senaryoda bireylerin oturma yerlerine göre aldıkları test formları farklılaşmaktadır. Ayrıca farklı testlerde ortak maddeler de yer almıştır. Böylece kimin cevap kopyalama ile kopya çeken ve kimin kaynak olduğu yönünde bilgi elde edilmiştir. Gerçek çalışma baz alınarak çeşitli faktörlerin değişimlendiği simülasyon çalışması gerçekleştirilmiştir.

Belov ve Armstrong'un (2010) birlikte gerçekleştirdiği çalışmada, bireylerin cevaplayacağı testi iki bölümden oluşturmuşlardır. İlk bölüm testi alacak tüm bireyler için

aynı iken ikinci bölüm her aday için farklılaşmaktadır. İkinci bölüm bireylerin toplam test puanını etkilememektedir. Kopya çeken ya da uygun olmayan davranışı gerçekleştiren bireyleri belirlemek için böyle bir desen oluşturmuşlardır. Bu desene göre kopya çeken bireyler ortak maddelerin olduğu bölümden ancak kopyadan kaynaklı test hilesinde bulunmuştur. Gerçekleştirilen çalışma bir simülasyon çalışmasıdır. Çalışmaya göre uygun olmayan davranışı gösterme oranı %60, %70, %80, %90 ve %100 olarak değişimlenmiştir. Testin iki bölümü ise kolay ya da zor, sözel ve sayısal şeklinde gerçek hayatta uygulanabilir olarak değiştirilebilir. Böylelikle testin iki bölümündeki bireyin yetenek dağılımı arasında bir farklılaşmanın olup olmadığı *KL* uzaklığı ile belirlenmiştir. Çalışmada kopya çeken bireyler öncelikle *K* benzerlik istatistiği ile belirlenmekle beraber ayrıca kopya çeken bireyler *KL* uzaklığı ile belirlendikten sonra bu belirlenen bireylerin buldukları sınıflarda tekrar *K* benzerlik istatistiği kullanılarak kopya çeken bireyler belirlenmiştir. *KL* uzaklığı ile *K* indeksinin birlikte kullanıldığı durumda olası kopyacı olarak belirlenen bireyler gerçek sayı ile örtüşürken, sadece *K* indeksi kullanılarak belirlenen olası kopyacı birey sayılarıyla gerçekten kopya çeken birey sayıları arasında farklılaşma olduğu görülmüştür. Başka bir ifadeyle *K*-indeksinin I. tip hataya sahip olduğu elde edilmiştir.

Clark (2010) çalışmasında, birey uyum istatistiklerine katkı getiren bir birey uyum istatistiğini geliştirmiştir. Birey uyum istatistikleri uygun olmayan davranışı belirlemek için kullanılan birey uyum istatistiklerinin çoğu tek boyutlu ölçme modellerini baz alarak değerlendirmede bulunmaktadır. Oysa Clark kopyayı çok boyutlu bir yapı olarak ele alınmasını öne sürerek tek ve çift faktörlü açımlayıcı faktör analizi modellerinde birey uyum istatistiklerindeki değişikliklerin karşılaştırılmasını inceleyen yeni bir teknik geliştirmiştir. Bu uyum istatistiğine göre modele cevap kopyalamadan kaynaklı test hilesi ikinci bir faktör olarak eklenmiştir. Bu ekleme sonucunda birey uyum istatistiğinde anlamlı bir farklılaşma sonucu kopya çekilmiş olduğunun kanıtı olarak ele alınmıştır. Çalışmasında çok kategorili maddeler kullanmıştır. İki farklı simülasyon uygulamıştır. İlk simülasyon da I. tip hatanın incelenmesi için gerçekleştirilirken ikinci simülasyonda olası kopyacı sayısı (10, 50, 100 ve 250) ve açığa çıkmış madde sayıları (3, 7 ve 13) değişimlenerek gerçekleştirilmiştir. Her iki simülasyon deseninde test uzunlukları (10, 18 ve 25) olarak değişimlenmiştir. Çalışmasında küçük ve orta büyüklükteki örneklerde kopyanın tespiti için kullandığı modelin yararlı olduğu sonucuna varmıştır.

Benzerlik indekslerinin çoğunluğu yanlış cevapların eşleşmesini baz aldığı için olası kopyacı ve kaynağın tepki vektörlerini karşılaştırırken yanlış cevap örüntüsü uyuşması

baz alınmaktadır. Ancak kaynağın yetenek düzeyi arttıkça bu indekslerin II. tip hatalarında artma olduğu gözlenmiştir. Bunun nedeni ise yanlış cevap sayısının azalmasıyla birlikte indekslerin gücünü kaybetmesidir. Bu durumdan yola çıkarak Belov (2011) çalışmasında yanlış cevapların yanı sıra doğru cevapları da baz alan *VM* indeksini geliştirmiştir. Çalışmasında en iyi sonuçlar veren *K* indeksini kullanarak *VM* ve *K* indekslerini I. tip ve II. tip hatasını incelemiştir. Elde ettiği sonuca göre, *K* indeksine göre *VM* indeksinin I. tip hatasında büyük bir düşüş, II. tip hatasında ise küçük bir düşüş olduğunu elde etmiştir.

St-Onge, Valois, Abdous ve Germain (2011) yüksek beklenmeyen davranış sergileyen cevap örüntülerini tespit etmek için kullanılan birey uyum istatistiklerinin doğruluğunu incelemeyi amaçladıkları çalışmalarında birey uyum istatistiklerinin doğruluğunu etkileyen faktörleri kullanarak tamamen çapraz deneysel desen (2 (madde güçlük ranjı) x 2 (madde ayırt edicilik ranjı) x 21 (beklenmeyen tepki oranı) x 2 (beklenmeyen tepki türü) x 4 (test uzunluğu)) kullanmışlardır. Maddelerin güçlük parametreleri [-1.5, 1.5] ve [-2.5, 2.5] aralıklarında olan uniform dağılımlardan elde edilmiştir. Madde ayırt edicilik parametreleri de [0.5, 1.0] ve [1.0, 1.5] aralıklarında olan uniform dağılımlardan türetilmiştir. Aralıkların kombinasyonlarından yararlanılarak madde karakteristik eğrileri elde edilmiştir. Bireylerin yetenekleri ise standart normal dağılımdan çekilmiştir. Beklenmeyen tepki örüntüsü oranı ise %10 ile %60 aralığında %2.5 artışla 21 farklı değerle değişimlenmiştir. Cevap kopyalamadan kaynaklı test hilesi ve şansız tahmin ile tepki verme olmak üzere iki farklı uygun olmayan tepki türü kullanılmıştır. Cevap kopyalamadan kaynaklı test hilesi için yetenek düzeyi sıfırdan küçük bireyler seçilmiştir ve seçilen bireylerin tepkileri uygun olmayan tepki oranlarına yüksek yetenekli bireylerin cevaplarıyla aynı olacak şekilde kopyalanmıştır. Şansız tahmin davranışı için ise yüksek yetenekli bireylerin teste verdikleri doğru tepkiler belirlenmiştir. Daha sonra belirlenen bu maddeler rastgele olacak şekilde beklenmeyen davranış oranlarına göre seçilerek, bireylerden doğru cevabı alma olasılığı 0.25 olarak manipüle edilmiştir. Çalışmada bireylere uygulanan testlerin uzunluğu ise 20, 40, 60 ve 80 madde uzunluğunda farklılaşmıştır. Çalışmada l_z ve *ECI2Z*, grup temelli birey uyum istatistikleri için ise *HT* ve *U3* birey uyum istatistikleri kullanılmıştır. %40 uygun olmayan tepki oranı %40 kadar olan artışta birey uyum istatistiklerinin beklenmeyen tepki gösteren bireyleri belirlemedeki doğruluğun arttığı sonucuna ulaşılmıştır. Bununla birlikte beklenmeyen tepki oranı arttıkça birey uyum istatistiklerinin bireyleri belirlemedeki doğruluğundaki artış bir yere kadar olduğu sonrasında ise bu artışın

azaldığı gözlenmiştir ancak en iyi performansı *ECI2Z* istatistiğinin gösterdiği görülmüştür. Özel olarak, kopya için, l_z ve grup temelli birey uyum istatistikleri için sırasıyla yaklaşık %30 ve %40'lık bir sapma oranında zirveye çıktığı, ancak şanssız tahmin için yaklaşık %35 ve %55 oranında zirveye çıktığı gözlemlenmiştir.

Zopluoğlu ve Davenport (2012) çalışmalarında *GBT* ve ω indekslerinin çeşitli koşullar altında gücü ve I. tip hatalarını incelemeyi amaçlamışlardır. 1440 koşulun her biri için cevap kopyalayarak kopya çeken 500 ikili karşılaştırılarak *GBT* ve ω indekslerinin gücü incelenmiştir. Benzer şekilde 1440 koşulun her biri için 10000 ikili karşılaştırılarak iki indeksin I. tip hatası incelenmiştir. İki indeksin I. tip hatasını şişirmediği ve pratikte kullanıma uygun olduğu sonucuna ulaşılmıştır. Bununla birlikte güç testi incelemesinde ise *GBT* indeksinin küçük bir farkla daha iyi performans gösterdiği gözlenmiştir. Ayrıca, tepkilerde kopya miktarının ana etkisi ve kaynağın yetenek düzeyi ile kopyanın miktarı arasındaki etkileşim etkisi çok güçlü olarak bulunurken, diğer tüm ana ve etkileşim etkileri ihmal edilebilir düzeyde olduğuna ulaşılmıştır.

Tendeiro ve Meijer (2012) gerçekleştirdikleri çalışmada madde güçlük ve birey yetenekleri standart normal dağılımdan elde edilmiştir. Madde ayırt edicilik ve tahmin parametreleri ise sırasıyla (0.5, 1.5) ve (0, 0.25) aralıklarında, uniform dağılımdan üretilmiştir. Testte yer alan madde sayısı 100 ve birey sayısı 10000 olarak değişimlenmiştir. Çalışmada 3 parametrelili lojistik model kullanılmıştır. Rastgele tahmin ve kopya olmak üzere iki türlü beklenmeyen tepki türü kullanılmıştır. Beklenmeyen davranış türleri altındaki her istatistiğin performansını daha iyi karakterize etmek için her beklenmeyen davranışı ayrı ayrı ele almışlardır. Böylelikle her durum için bağımsız veri setleri oluşturulmuştur. Rastgele tahminleri belirlemek için C^- , C_{VM}^{LR} , C^L , C^{LR} , C^{RR} , U , W ve l_z ; kopyayı belirlemek için ise C^+ , C_{VM}^{LR} , C^U , C^{LR} , C^{Ch} , U , W ve l_z istatistikleri kullanılmıştır. Olası kopyacı birey olarak seçilen bireylerin yetenekleri 0.5'in altında olacak şekilde belirlenmiştir. Olası kopyacıların oranı 0.01, 0.05, 0.10 olarak değişimlenirken, şüpheli bireylerin maddelerindeki manipülasyon oranı ise, 0.05, 0.10 ve 0.25 olacak şekilde farklılaştırılmıştır. Değişimlenen maddeler testte ardışık olacak şekilde seçilmiştir. Madde ve yetenek parametreleri manipülasyon öncesinde ve sonrasında tüm koşullar için hesaplanmıştır. Beklenmeyen tepkinin parametre kestirimini etkileyip etkilemediğini kontrol etmek için korelasyonlara bakılmıştır. Rastgele tahmin davranışı durumuna *CUSUM* tabanlı istatistiklerden biri olan C_{VM}^{LR} haricindeki diğer

İstatistikler üzerinde olası kopyacı birey oranının bir etkisinin olmadığı görülmüştür. Ancak U , W ve l_z istatistiklerinin tespit oranları üzerinde ise bir etkiye sahip olduğu gözlenmiştir. Beklenmeyen tepki örüntüsünün miktarı arttıkça U , W ve l_z 'nin olası kopyacı belirleme gücünün düştüğü görülmüştür. Bununla birlikte, bu çalışmada ele alınan birey-uyum istatistikleri için, beklenmeyen tepki örüntüsü oranının tespit oranlarındaki değişimi açıklamadığı belirtilmiştir. Rastgele tahmin ve kopya çekme türlerinin tespit oranları genel olarak *CUSUM* temelli birey-uyum istatistikleri tepki örüntüsündeki değişim oranının artmasıyla l_z istatistiğinden daha iyi performans gösterdiği gözlenmiştir. *CUSUM*'ların beklenmeyen puan dizilerini tespit etmede daha hassas olduğu görülmüştür. Sıraların uzunluğunun artırılması, *CUSUM*'un beklenmeyen davranışı tespit etme yaklaşımlarına fayda sağladığı bulunmuştur. Beklenmeyen madde puanları dizilerine odaklandıkları için *CUSUM*'lar U , W ve l_z istatistiklerinden daha iyi performans göstermiştir. Sıralar ne kadar uzun olursa, geleneksel birey-uyum istatistikleriyle karşılaştırıldığında *CUSUM*'lar o kadar iyi performans göstermiştir. Beklenmeyen puan dizileri küçük olduğunda, U , W ve l_z istatistikleri *CUSUM*'lardan daha iyi performans sergilemiştir.

Belov (2016) çalışmasında sınava katılan bireylerin cevaplarındaki beklenmeyen değiştirmeleri belirlemeyi amaçlamıştır. Sınava katılan bireylerin cevaplarındaki örüntüyü değiştirmeksizin ve test maddelerine verilen tepkileri değiştirerek (performansın nihai cevapları) bu iki tepki örüntüsü arasındaki performans farkını ölçerek cevaplardaki beklenmeyen değişimleri belirlemiştir. Çalışmadaki yaklaşım Belov'un (2015) madde önbilgisini belirlemeye yönelik olarak gerçekleştirdiği çalışmayla aynıdır. Cevaplarında beklenmeyen değişimlere giden bireyleri belirlemek için *KL*, *EDI*, *PS* ve *NPL* olmak üzere dört istatistik kullanmıştır. Beklenmeyen cevap değişimlerini belirlemeye yönelik olarak kullanılan istatistikler testlerin kağıt-kalem ve bilgisayar destekli olmasına, beklenmeyen davranışın olma miktarına, C ve U belirsizliği olmak üzere farklı koşullar altında incelenmiştir. Çalışmada *PS* ve *KL* istatistiklerinin daha iyi sonuç verdiği görülmüştür.

Brussow, Skorupski ve Thompson (2018), bireysel düzeyde test hilesinden öte grup halinde yapılan test hilesini tespit etmeye yönelik çalışmalarını gerçekleştirmişlerdir. Madde Tepki kuramı bazında gerçekleştirilen simülasyon çalışmalarını, madde tepki kuramının bir hiyerarşik modeli olan Bayesian Hiyerarşik Lineer Model (BHLM) kullanarak grubun beklenmeyen davranış düzeyini belirlemeyi amaçlamışlardır. 300 sınıfın içine yuvalanmış öğrencilerin iki yıllık standartlaştırılmış test puanları

kullanılacak şekilde simüle edilmiştir. Çalışmada grup düzeyinde bilgiler yalnızca değerlendirmenin ikinci yılı için dahil edilmiştir; bu durum öğrencilerin her iki yıl boyunca aynı gruplarda kalmalarına gerek kalmamasını sağlamıştır. İlk yılın bireysel puanları sadece yeteneği belirlemek için kullanılmıştır. Böylece araştırmacılar bireylerle ilgili geçmiş yıllar hakkında bilgi toplamak zorunda kalmamışlardır. Simülasyon düzeni planlanırken gerçek hayata uygulaması kolay olacak şekilde senaryo planlanmıştır. Sonuç olarak grup düzeyinde meydana gelen büyümeleri göstermenin yanı sıra bireylerin yeteneklerindeki değişimleri de göstermiştir.

Man, Haring, Ouyang ve Thomas (2018) çalışmalarında bireylerin maddelere verdikleri tepki sürelerini kullanan yeni bir parametrik olmayan birey-uyum istatistiği kullanarak istatistiğin kopya belirleme performansına odaklanmıştır. Çalışmada geniş ölçekli bir değerlendirme yanı sıra Monte Carlo simülasyonu da kullanılmıştır. Kullanılan istatistiğin araştırmada yer alan kopya türlerini belirlemede başarılı olduğu görülmüştür. Çalışmada tepki süresi modelinde yer alan tepki süreleri kopya türlerine göre müdahale edilerek değiştirilmiştir. Madde ön bilgisine sahip olduğu durumlarda gerçekleşen test hilesi için maddelere verilen tepki süreleri gerçek hayattaki durumlar göz önünde bulundurularak 10 ile 15 saniye aralığında uniform dağılımdan rastgele seçilerek değiştirilmiştir. Kullanılan testlerde yer alan madde sayıları ise 15 ve 30 olarak değişmektedir. Cevap kopyalama davranışı için çalışmada şüpheli bireylerin kopya çektikleri maddelere verdikleri tepkilerin sürelerine ise “30, 120” saniye aralığında uniform dağılımdan çekilen süreler eklenerek müdahale edilmiştir. Kopya çekilen maddelerin sürelerindeki manipülasyonlar için eğer maddeler testin sonunda yer alıyorsa manipülasyon süresi testin zaman sınırlılığına bağlı olarak azaltılmıştır. Parametrik olmayan *KL* indeksi neredeyse her simülasyon koşulunda parametrik indeksten daha iyi performans göstermiştir. Özellikle *KL* indeksinin madde ön bilgisini belirlemede çok daha hassas olduğu görülmüştür.

Aşağıda son dönemde ülkemizde gerçekleştirilen test hilesi belirlemeye yönelik çalışmalara yer verilmiştir.

Sünbül ve Yormaz (2018a) ω ve *GBT* indeksinin performanslarını test uzunluğu, testte yer alan maddelerin güçlük ve ayırt edicilik parametrelerini farklı düzeylerde değişimleyerek incelemiştir. ω ve *GBT* indekslerinin I. tip hata oranları genel olarak kabul edilebilir nominal alfa düzeyinin altında bulunmuştur. Testte yer alan maddelerin güçlük düzeyi arttığında indekslerinin güç performanslarının arttığını ancak testte yer alan indekslerin ayırt edicilikleri arttığında ise ω ve *GBT* indekslerinin güç performansları için

benzer sonucun elde edilmediği belirtilmiştir. Testte yer alan madde sayısının artmasıyla da ω ve *GBT* indeksinin test hilesi yapan bireyi belirleme güçlerinin arttığı gözlenmiştir. Sünbül ve Yormaz (2018b) ω indeksinin yetenek düzeyi, testte yer alan maddelerin güçlük ve ayırt edicilik parametrelerine göre I. tip hatasını ve gücünün nasıl etkilendiğine yönelik bir simülasyon çalışması gerçekleştirmiştir. Olası kopyacı ve kaynağın yetenek düzeyleri arasındaki fark arttığında ω indeksinin gücünün arttığı gözlenmiştir. Zor ve yüksek ayırt edici maddeler içeren testte ω test hilesi yapan bireyi belirlemede performansının 0.56 ile 1 arasında değiştiğini elde etmişlerdir. Bununla birlikte ω indeksinin I. tip hatasının ise genel olarak düşük değerler aldığı belirtilmiştir.

Son dönemde bazı kurumlar, yurtdışı adaylarını internet üzerinden gözetimsiz bir şekilde bilgisayar ortamında bireye uyarlanmış test (BOBUT) uygulamaktadır. Gözetimsiz test puanları dikkate alınan adaylara, kampüs içerisinde yeniden bir çevrim içi test uygulanmaktadır. Çevrim içi uygulanan test gözetimsiz yapılan bilgisayar ortamında bireye uyarlanmış teste paralel ve BOBUT'u doğrulama niteliği taşımaktadır. Balta, Uçar ve Şahin'in (2019) çalışmalarında kampüs içerisinde bireylere uygulanan çevrim içi testi kullanarak, BOBUT'ta test hilesi yapanları tespit etmek için aşamasız ve iki aşamalı olarak analiz gerçekleştirmiştir. Çalışmada kopyacıların yetenek düzeyi (düşük, orta) ve test hilesi yapılan madde oranı (%60 , %70) faktörlerine dayalı olarak simülatif verilere aşamasız olarak l_z istatistiği kullanılarak test hilesi yapan bireyler belirlenmiş. Aynı verilerle test hilesi yapan bireyleri belirlemek için *KL* indeksi değerleri dikkate alınmıştır. Belirlenen bireylere ikinci olarak birey uyum analizi yapılarak l_z istatistiği değerleri hesaplanmıştır. l_z istatistiği değerlerine göre hile yapan bireyler belirlenmiştir. Çalışmanın sonunda l_z istatistiğinin aşamalı kullanıldığında test hilesi yapan bireylerin belirlenme performansının arttığı sonucuna ulaşılmıştır.

Yormaz (2019), çoktan seçmeli maddelerden oluşan testlerdeki, olası işbirliği ile test hilesi yapan bireyleri belirlemek için *M4* benzerlik istatistiğini kullanarak denetimsiz makine öğrenme türü olan hiyerarşik kümeleme ve yarı-denetimli makine öğrenme türü olan kendi kendine öğrenme ve *SETRED* (Self Training with Editing) yöntemleri ile bireyleri ve grupları belirleme performanslarını incelemiştir. Çalışmasında, iş birliği yapan grup sayısını (2), grupta yer alan bireylerin sayısını (5, 10, 20) ve kopya oranını (0.10, 0.15, 0.20, 0.40) olarak değişimlemiştir. *Ward* kümeleme yönteminin kopya oranı 0.40 olduğunda iş birliği yapan bireyleri belirlemede güçlü, kopya oranı 0.20 olduğunda

ise iş birliği yapan bireyleri belirleme performansının yetersiz kaldığını belirtmiştir. Gruplarda iş birliği yapan birey sayısının artmasıyla kümeleme yönteminin belirleme performansının da arttığını, kendi kendine öğrenme ve *SETRED* yöntemlerinin sınıflandırma performanslarının değişmediğini elde etmiştir. Kullandığı yöntemlerde genel olarak kopya oranının 0.40 ve üzerinde olduğu şartlarda, yarı-denetimsiz yöntemlerin iş birliği yapan bireyleri ve grupları ayırt etme performansının iyi, iş birliği yapan birey sayısının arttığında ise bu performansın da arttığını belirtmiştir.

Çiçek (2020) çalışmasında, örneklem büyüklüğü (500, 5000, 50000, 99990), beş farklı yetenek düzeyi, test uzunluğu (50, 160) şartları altında *Silme Belirleme İndeksi*'nin test tahrifatını belirleme performansını I. tip hata ve gücünü temel alarak çeşitli α seviyelerinde (0.00001, 0.0001, 0.0005, 0.001, 0.005, 0.01, 0.05) incelemiştir. Süreklilik düzeltmesinin tüm α seviyelerinde I. tip hata ile birlikte güç ortalamalarını da düşürdüğünü belirtmiştir. Örneklem büyüklüğünün artmasıyla *Silme Belirleme İndeksi*'nin I. tip hata ortalamasının azaldığını ifade etmiştir. Ayrıca, sabit silme oranının artmasıyla test tahrifatının belirleme gücünün arttığı, yetenek düzeyinin de arttığında indeksin belirleme gücünün arttığı sonucuna ulaşmıştır.

Yurt içinde ve yurt dışında yapılan çalışmalar incelendiğinde, farklı senaryolar altında test hilesi yapan bireyleri belirlemede kullanılan farklı yöntemlerin performanslarına yönelik olarak çeşitli araştırmaların gerçekleştirildiği görülmüştür. Sonuçlar baz alındığında test hilesi yapan bireyi belirlemeye yönelik kanıtların artırılması amacıyla kullanılan farklı yaklaşımlı yöntemler ve analizlerin önerildiği görülmektedir (Balta v.d., 2019; Belov ve Armstrong, 2010; Man v.d., 2018; Soel ve Rubright, 2013). Çalışmalarda *K* indeksinin performansı *KL* indeksi ile aşamalı kullanım durumuna göre incelenmiştir. Yurtdışında ve yurt içinde iki aşamalı yaklaşım kullanılarak gerçekleştirilen çalışmalarda *K* indeksi ve l_z birey uyum istatistiklerinin performansları incelenmiştir (Balta v.d., 2019; Belov ve Armstrong, 2010; Soel ve Rubright, 2013). Bu çalışmada ise *K* indeksi gibi küçük örneklemelerde de iyi performans gösteren kopya benzerlik indekslerinden ω ve alanyazında belirtildiği gibi bazı durumlarda ω dan daha iyi performans gösteren *GBT* indeksinin *KL* indeksi ve birey uyum istatistikleriyle birlikte iki aşamalı kullanımı sonucunda sergiledikleri performansları sunulmuştur.

BÖLÜM 3

YÖNTEM

Bu bölümde araştırmanın modeli, araştırmanın değişkenleri, verilerin üretilmesi ve araştırma kapsamında verilerin çözümlenmesine yer verilmiştir.

Araştırmanın Modeli

Bu çalışmada simülasyon verileri kullanılarak çeşitli koşullar altında ω indeksi, *GBT* indeksi, l_z ve l_z^* ve *KL* indeksinin I. tip hata oranları ile cevap kopyalamadan kaynaklı test hilesi yapan bireyleri belirleme güçleri hakkında bilgi edinilmeye çalışılmıştır. Böylece farklı test güçlükleri, farklı yetenek düzeyleri ve farklı kopya çekilen madde oranları kullanıldığında indekslerin I. tip hata ve güç oranları ile ilgili yapılan çalışmalara katkıda bulunması amaçlanmaktadır. Bu çerçevede, yapılmış olan bu çalışma, betimsel araştırma niteliğinde bir simülasyon çalışması olarak değerlendirilebilir.

Simülasyon Deseni

Çalışma simülatif verilerle gerçekleştirilmiştir. Bu bölümde simülatif çalışmanın değişkenleri ve bu değişkenlere uygun olarak gerçekleştirilmiş veri üretim sürecine yer verilmiştir.

Araştırmanın Değişkenleri

Bu çalışmada incelenen bağımlı değişkenler; test hilesini belirlemede kullanılan yöntemlerin I. tip hatası ve bu yöntemlerin gücüdür. Araştırmanın bağımsız değişkenleri test hilesi belirleme yöntemleri, testin ortalama güçlüğü, kopya çekilen madde oranı, kopya çeken bireylerin yetenek düzeyleridir. İfade edilen değişkenler ve düzeyleri ilgili alanyazında yer alan benzer çalışmalar, bu çalışmalardan elde edilen bulgular dikkate alınarak ve gerçek durumları yansıtacak şekilde belirlenmiştir.

Manipüle Edilmeyen Değişkenler

Örneklem büyüklüğü. İlgili alanyazında, örneklem büyüklüğünün test hilesi belirleme yöntemlerinin I. tip hatası ve gücünü etkileyen bir değişken olduğu görülmüştür. Sotoaridona ve Meijer (2002) çalışmalarında örneklem büyüklüğünü 100 (küçük), 500 (orta), 2000 (büyük) şeklinde değişimlemiştir. Çalışmada küçük örneklemelerde K , ancak büyük örneklemelerde ω 'nın daha iyi performans gösterdiği elde edilmiştir. K indeksinin hatası diğer indekslere göre daha azdır dolayısıyla küçük örneklemelerde iyi sonuçlar verdiği belirtilmiştir. Wollack (2006) ve Armstrong ve Shi (2009) çalışmalarında 10.000 bireyle çalışmasını yürütmüştür. Bununla birlikte Wollack (2006) çalışmasında 10.000 bireyi; örneklem büyüklüğü 400 (25 adet), 200 (50 adet), 100 (100 adet), 40 (250 adet), 20 (500 adet), 10 (1000 adet) şeklinde değişimlemiştir. Madde parametrelerinin kestiriminde örnek büyüklüğünün önemli olmasından dolayı ölçme modellerini değişimlediği çalışmasında Zopluoğlu (2016) örneklem büyüklüğünü 100, 500 ve 2.500 birey olacak şekilde değişimlemiştir.

Bu çalışmada örneklem büyüklüğü faktörü değişimlenmemiştir. Sınav salonlarının mevcut kapasiteleriyle birlikte çalışmada baz alınan ölçme modeli (en küçük örneklem büyüklüğü) göz önünde bulundurularak örneklem büyüklüğü 150 olarak sabitlenmiştir.

Test uzunluğu. Bireylerin ölçülen özellikteki potansiyelleri hakkındaki bilgilere test maddelerine verdikleri tepkiler kullanılarak ulaşılmaktadır. Dolayısıyla yeterli sayıda maddeye verdikleri tepkilerle ilgili sonuca ulaşılabilir. Birey uyum istatistikleri, bireylerin maddelere verdikleri tepkileri kullanarak bireyin potansiyeline ya da potansiyeline uygun olmayan örüntü sergilemiş olduğuna ya da olmadığına yönelik sonuçlar üretmektedir. Dolayısıyla bireyin testte hile yaptığına ya da yapmadığına yönelik bilginin edinilmesinde test uzunluğu etkili olmaktadır. Literatürde testteki hileyi belirlemeye yönelik çalışmalarda test uzunluğu 40 ve 80 (Sotaridona ve Meijer, 2002, 2003; Wollack, 2006), Karabatsos (2003) çalışmasında ise testin uzunluğunu 17 (kısa), 33 (orta), 65 (uzun) madde olmak üzere toplamda üç farklı düzeyde değişimlemiştir.

Bu çalışmada sabitlenen örneklem büyüklüğü değişkeni ile baz alınan ölçme modeli göz önünde bulundurularak test uzunluğu 30 madde olarak sabitlenmiştir.

Kopya çeken birey oranı. Steinkamp (2017) çalışmasında olası kopyacı oranını %5 ve %15 olarak değişimlemiştir. Bu çalışmada 150 örneklem büyüklüğü kullanılmıştır. Analiz sırasında parametre kestirimi hatasının artmaması amacıyla verileri manipüle edilen bireylerin oranı %5 olarak sabitlenmiştir.

Manipüle Edilen Değişkenler

Kopya oranı. Bireylerin testte hile yaptıkları madde sayısı ya da oranı test hilesi belirleme istatistiklerinin gücünü ve I. Tip hatasını etkileyen bir diğer önemli faktördür. Zopluoğlu (2016) çalışmasında kopya çekilen madde miktarını testin uzunluğunun %20'si (düşük), %40'ı (orta) ve %60'ı (yüksek) olmak üzere üç seviyede değişimlemiştir. Test hilesinin yapıldığı madde oranı miktarı, sınıflandırma performansını etkileyen en önemli bir değişken olduğunu belirtmiştir. Wollack (2006) ve Sotaridona ve Meijer (2002) ise test hilesinin yapıldığı madde oranı %10, %20, %30 ve %40 olarak farklılaştırmıştır. Karabatsos (2003) çalışmasında kopya çeken bireylerin tepki örüntüsünü uygun olmayan tepki şeklinde adlandırarak uygun olmayan tepki miktarını %5, %10, %25 ve %50 şeklinde değişimlemiştir.

Test hilesini belirlemede kullanılan indeks ve istatistiklerin I. tip hatasını ve gücünü incelenmesi alanyazında değişimlenen kopya oranları da dikkate alınarak bu çalışmada, test hilesi yapılan madde oranı %10 (düşük), %40 (orta) ve %60 (yüksek) şeklinde değişimlenmiştir.

Kopya çeken bireyin yetenek düzeyi. Sotaridona ve Meijer (2002) çalışmasında kaynağın düzeyini %90 'ıncı ve %60'ıncı yüzdelerdeki bireylerin yeteneği şeklinde değişimlerken, Yormaz (2017) ise kaynağın yetenek düzeyinin yüzdeliğine %40-49, %50-59, %60-69, %70-79, %80-90 göre 5 farklı aralıkta belirtmiştir. Kaynağın yetenek yüzdeliğinin artmasıyla kullandığı istatistiklerden S1 ve S2 'nin daha güçlü olduğunu gözlemlemiştir. Sünbül ve Yormaz (2018a) çalışmasında kaynak ve olası kopyacının yetenek düzeyini (-3,-1.5), (-1.51, 0), (0.01, 1.5), (1.51, 3) şeklinde değişimlenmişlerdir.

İndekslerin I. tip hata ve güçlerinin kopyadan kaynaklı test hilesi yapan bireylerin yetenek düzeyinden etkilenebilmektedir. Bununla birlikte bireylerin yüksek yetenekli bireylerin testte kendi verdikleri cevaplara güvendikleri için test hilesi yapma motivasyonları daha olduğundan düşük yetenekli bireylerin test hilesinde bulunma motivasyonları yüksektir (Voncken, 2014). Dolayısıyla bu çalışmada kopya çeken bireyin yetenek düzeyi ise çok düşük (-3,-1.51) ve düşük (-1.50, 0) şeklinde değişimlenmiştir.

Testin ortalama güçlüğü. Bireylerin kopyadan kaynaklı test hilesinde buldukları maddenin güçlüğüne kopya ve benzerlik indekslerinin olası kopyacı belirleme performansını etkilediği görülmüştür (Sünbül ve Yormaz, 2018a). Özellikle güçlük düzeyi düşük olan maddelerde olası kopyacı belirleme gücünün azaldığı gözlenmiştir. Alan yazında Sünbül ve Yormaz (2018) çalışmalarında test maddelerinin güçlüğüne (-

2.5, 0.00) ve (0.01, 2.5) Zopluoğlu (2016) testin ortalama güçlüğünü 0.8 ve 0.5 olarak deęişimlemiřlerdir.

Bu alıřmada birey uyum istatistiklerinin ve *KL* indeksinin’de kolay testteki performanslarına yönelik bulgular elde edilmesi iin testte yer alan maddelerin güçlük parametresi düzeyi (-2.5, 0.00) (kolay test) ve (0.01, 2.5) (zor test) olarak deęişimlenmiřtir.

Tablo 2’de alanyazıdaki alıřmalar göz önünde bulundurularak alıřmanın deęişkenleri ve deęişimlenen düzeylerine yer verilmiřtir.

Tablo 2

Simülasyon Deseni

	Deęişkenler	Düzeyley	Düzeyley Sayısı
	Kopya Oranı	% 10-%40-%60	3
Manipüle Edilenler	Testin Ortalama Güçlüğü	Kolay Test-Zor Test	2
	Yetenek Düzeyi	ok Düşük- Düşük	2
	Örneklem Büyüklüğü	150	1
Manipüle Edilmeyenler	Test Uzunluęu	30	1
	Kopya eken birey oranı	%5	1
Toplam Koşul Sayısı			3x2x2x1x1x1=12

Bu bilgiler ışığında ařaęıda alıřmanın veri üretim sürecine dair ayrıntılı bilgilere yer almaktadır.

Verilerin Üretilmesi

alıřmada deęişimlenen faktör ve düzeyler iin gerekleřtirilen veri üretimi ve verilerin manipülasyonu ile ilgili bilgilere ařaęıda yer verilmiřtir.

Güç analizleri iin, yineleme sayısı (100) (Sünbül ve Yormaz, 2018a,2018b), kopya oranı (0.1, 0.4, 0.6), kopya eken bireyin yetenek düzeyi (ok düşük (-3, -1.51) ve düşük (-1.5, 0)) ve testin ortalama güçlük düzeyi ((kolay(-2.5, 0) ve zor (0.01,2.5)) olmak üzere 1200 veri dosyası, I. tip hata analizleri iin, yineleme sayısı (100), kopya eken bireyin yetenek düzeyi (ok düşük (-3, -1.51) ve düşük(-1.5, 0)) ve testin ortalama güçlük düzeyi ((kolay(-2.5, 0) ve zor (0.01,2.5)) olmak üzere 400 veri dosyası simüle edilmiřtir. Her bir veri dosyası iin normal daęılımdan 1500 bireye ait yetenek düzeyleri üretilmiřtir.

Bireylerin yetenekleri göz önünde bulundurularak Rasch Modeline uygun 1-0 matrisi üretilmiştir.

Olası kopyacının kaynak olarak belirlediği bireyin yetenek düzeyinin, kendisinden daha yüksek olma olasılığını arttırmak için sınıf oluşturulurken, yüksek ve çok yüksek yetenekli bireylerin sınıftaki daha fazla olacak şekilde ağırlıklandırma yapılmıştır. Değişimlenen faktörler ve düzeylerinin kopya ve benzerlik indeksleri ve birey uyum istatistikleri ve KL indeksi I. tip hata ve güçlerine olan etkisini gözlemlemek için, çok düşük yetenekli bireylerin örneklemin %5, düşük yetenekli bireylerin örneklemin %15, orta yetenekli bireylerin örneklemin %60 ve yüksek yetenekli bireylerin ise örneklemin %20'si oranında olacak şekilde 1500 kişilik evrenden seçkisiz seçilerek 150'şer kişilik sınıflar oluşturulmuştur. Sınıflar oluşturulduktan sonra, çok düşük ve düşük yeteneğe sahip olası kopyacı bireyler ayrı ayrı durumlar olarak ele alınmış ve olası kopyacı bireyler belirlenmiştir. Sınıflardaki olası kopyacı bireyler belirlendikten sonra, bireylerin oturma şekli de seçkisiz olarak belirlenmiştir.

İndekslerin güç analizleri için, bireylerin yerleştirme işlemi gerçekleştirildikten sonra olası kopyacı bireylerin cevaplarını kopyaladıkları kaynak bireylerin belirlenmesi aşamasına geçilmiştir. Bu aşama için, sınıfta olası kopyacı olarak seçilen bireylerin önünde, sağında ve solunda yer alan bireyler kopya çeken bireyler için potansiyel kaynak olarak varsayılmıştır. Kopya çeken birey bu kaynak bireylerden yalnızca birisini kaynak olarak seçtiği sayılısı altında, kaynak bireylerden biri seçkisiz olarak belirlenmiştir. Olası kopyacı olarak seçilen bireyin testte verdiği cevaplarının belirli oranı (0.1, 0.4, 0.6) kaynağın cevaplarıyla aynı olacak şekilde belirlenmiştir. Böylece olası kopyacı olarak belirlenen bireyin cevapları, kaynağın kopyalanan cevaplarıyla değiştirilerek manipüle edilmiştir.

İndekslerin I.tip hata analizleri için bireylerin yerleştirme işlemi gerçekleştirildikten sonra olası kopyacıların kaynak olarak seçtikleri bireylerin belirlenmesi aşamasına geçilmiştir. Bu aşama için, sınıfta olası kopyacı olarak seçilen bireylerin önünde, sağında ve solunda yer alan bireyler kopya çeken bireyler için potansiyel kaynak olarak varsayılmıştır. Kopya çeken birey bu kaynak bireylerden yalnızca birisini kaynak olarak seçtiği sayılısı altında, kaynak bireylerden biri seçkisiz olarak belirlenmiştir. Bireylerin cevap örüntülerinde herhangi bir müdahale yapılmaksızın veriler üretilmiştir.

Verilerin Analizi

Alanyazında yer alan çalışmalar incelendiğinde kopyadan kaynaklı test hilesini tespit etmede kullanılan indekslerin performanslarına yönelik yararlanılan yöntemlerin birinin de I. tip hata ve güç uygulamaları olduğu görülmektedir. İndekslerin gücü kopyadan kaynaklı test hilesinin var olduğu durumda olası kopyacının tespit edilmesidir. Bununla birlikte kopyadan kaynaklı test hilesinin var olmadığı durumda ise yanlışlıkla olası kopyacının belirlenmesi ise indeksin I. tip hatasıdır. Tablo 3'te gerçekte olan durum ile indekslerin kullanılması sonucunda elde edilen durumda verilen karar sonucu elde edilen indeksin gücüne ve I. tip hatasına yönelik ilişkiye yer verilmiştir.

Tablo 3. İndeksin I. Tip Hatası ve Kopya Belirleme Gücü

		Hipotez Testi Sonucunda Verilen Karar	
		Ho Rededildi “Cevap kopyalamadan kaynaklı test hilesi var”	Ho Kabul edildi “Cevap kopyalamadan kaynaklı test hilesi yok”
Gerçek Durum	H_0 doğru “Cevap kopyalamadan kaynaklı test hilesi yok”	I. Tip Hata	Doğru Karar
	H_0 yanlış “Cevap kopyalamadan kaynaklı test hilesi var”	Doğru Karar (Cevap Kopyalamadan Kaynaklı Test Hilesini Belirleme Gücü)	II. Tip Hata

İndekslerin I. Tip Hata Oranının Belirlenmesi

Kopya ve benzerlik indekslerinin (ω ve GBT) kopya belirlemede I. Tip hata oranlarının belirlenmesi için kopya çekilmediği bilinen verilerle tekrarlanan her yinelemede, seçilen olası kopyacı ve kaynak çifti için R programında yer alan “CopyDetect” paketi (Zopluoğlu, 2018) kullanılarak kopya ve benzerlik indekslerinin olasılık değerleri hesaplanmıştır. Hesaplanan olasılık değerleri için α düzeyinden küçük ve eşit olanlara (“cevap kopyalamadan kaynaklı test hilesi var” kararları verilen çiftler için) 1, büyük olanlara ise (“cevap kopyalamadan kaynaklı test hilesi yok” kararları verilen çiftler için) 0 değeri verilerek bir 1-0 matrisi oluşturulmuştur. Matriste yer alan

“1” değerlerinin toplamı tüm seçilen çift sayısına oranı hesaplanmıştır. Böylece indekslerin seçilen çiftler için verdiği kararın yanlış olma olasılığı (I. tip hata oranı) elde edilmiştir.

Birey uyum (l_z ve l_z^*) istatistiklerinin I. tip hata oranlarının belirlenmesi için kopya çekilmediği bilinen verilerle tekrarlanan her yinelemede tüm bireyler için birey uyum puanları R programında yer alan “*PerFit*” paketi (Tendeiro, 2018) kullanılarak hesaplanmıştır. “*PerFit*” paketinde yer alan kesme puanı algoritması kullanılarak $\alpha = 0.05$ düzeyi için kesme puanı elde edilmiştir. Elde edilen kesme puanından büyük olanlara (“cevap kopyalamadan kaynaklı test hilesi var” kararı verilen olası kopyacı için) 1, küçük olanlara ise (“cevap kopyalamadan kaynaklı test hilesi yok” kararı) 0 değeri verilerek bir 1-0 matrisi oluşturulmuştur. Matriste yer alan “1” değerlerinin toplamı çalışmanın değişkenlerinden biri olan olası kopyacı sayısına (kopya oranı (0.05) x örneklem büyüklüğü (150)) oranı hesaplanmıştır. Böylece indekslerin verdiği kararın yanlış olma olasılığı (I. tip hata oranı) ortaya konmuştur.

KL indeksinin I. tip hata oranını ortaya koymak için manipülasyonun yapılmadığı verilerle tekrarlanan her yinelemede tüm bireylerin cevap örüntülerinin yer aldığı (1-0) cevap matrisi, testin parametreleri kullanılarak R programında yer alan “*irtoys*” paketi (Partchev, 2016) ile yetenek kestirimi yapılmıştır. Yetenek kestirimleri yapıldıktan sonra R programında yer alan “*LaplaceDemon*” paketi (Singmann, 2020) kullanılarak bireylerin *KL* indeksi değerleri hesaplanmıştır. Bu işlemden sonra *KL* indeksi için R programında yer alan “*OptimalCutpoints*” paketinde (Raton, 2014) yer alan *Youden İndeksi* ile $\alpha = 0.05$ düzeyi için kesme puanı elde edilmiştir. Elde edilen kesme puanından büyük olanlara (“cevap kopyalamadan kaynaklı test hilesi var” kararı verilen olası kopyacı için) 1, küçük olanlara ise (“cevap kopyalamadan kaynaklı test hilesi yok” kararı) 0 değeri verilerek bir 1-0 matrisi oluşturularak, matriste yer alan “1” değerlerinin toplamı çalışmanın değişkenlerinden biri olan olası kopyacı sayısına (kopya oranı (0.05) x örneklem büyüklüğü (150)) oranı hesaplanmıştır. Böylece *KL* indeksinin verdiği kararın yanlış olma olasılığı (I. tip hata oranı) elde edilmiştir.

Aşamalı adımda ise öncelikle birey uyum (*KL* indeksi) istatistikleri kullanılarak olası kopyacılar belirlenmiştir. Olası kopyacıların sınıfta oturduğu yere göre geriye kalan potansiyel kaynak (bireyin önünde, sağında ya da solunda yer alan bireyler) olarak görülen bireylerden biri seçilerek, olası kopyacı birey ile kopya analizine dahil edilmiştir. Kopya ve benzerlik indekslerinin (ω ve *GBT*) cevap kopyalamadan kaynaklı test hilesini belirlemede I. tip hata oranlarının belirlenmesi için kopya çekilmediği bilinen verilerle

tekrarlanan her yinelemede, seçilen olası kopyacı ve kaynak çifti için R programında yer alan “*CopyDetect*” paketi (Zopluoğlu, 2018) kullanılarak kopya ve benzerlik indekslerinin olasılık değerleri hesaplanmıştır. Hesaplanan olasılık değerleri için α düzeyinden küçük ve eşit olanlara (“cevap kopyalamadan kaynaklı test hilesi var” kararları verilen çiftler için) 1, büyük olanlara ise (“cevap kopyalamadan kaynaklı test hilesi yok” kararları verilen çiftler için) 0 değeri verilerek bir 1-0 matrisi oluşturulmuştur. Matriste yer alan “1” değerlerinin toplamı iki aşamalı analizde birey uyum ya da *KL* indeksini kullandıktan sonra elde edilen olası kopyacı bireyler (başlangıçta seçkisiz olarak belirlenmiş olası kopyacılar arasından tespit edilenler)-kaynak çifti sayısına oranı hesaplanmıştır. Böylece indekslerin seçilen çiftler için verdiği kararın yanlış olma olasılığı (I. tip hata oranı) elde edilmiştir.

İndekslerin Güç Oranının Belirlenmesi

Kopya ve benzerlik indekslerinin (ω ve *GBT*) kopya belirlemede güç oranlarının belirlenmesi için kopya çekilen madde oranının manipüle edildiği/kopya çekildiği bilinen verilerle tekrarlanan her yinelemede, seçilen olası kopyacı ve kaynak çifti için R programında yer alan “*CopyDetect*” paketi (Zopluoğlu, 2018) kullanılarak kopya ve benzerlik indekslerinin olasılık değerleri hesaplanmıştır. Hesaplanan olasılık değerleri için $\alpha = 0.05$ düzeyinden küçük ve eşit olanlara (“cevap kopyalamadan kaynaklı test hilesi var” kararları verilen çiftler için) 1, büyük olanlara ise (“cevap kopyalamadan kaynaklı test hilesi yok” kararları verilen çiftler için) 0 değeri verilerek bir 1-0 matrisi oluşturulmuştur. Matriste yer alan “1” değerlerinin toplamı tüm seçilen çift sayısına oranı hesaplanmıştır. Böylece indekslerin seçilen çiftler için verdiği kararın doğru olma olasılığı (güç oranı) elde edilmiştir.

Birey uyum (l_z ve l_z^*) istatistiklerinin güç oranlarının belirlenmesi için kopya çekilen madde oranının manipüle edildiği/kopya çekildiği bilinen verilerle tekrarlanan her yinelemede tüm bireyler için birey uyum puanları R programında yer alan “*PerFit*” paketi (Tendeiro, 2018) kullanılarak hesaplanmıştır. “*PerFit*” paketinde yer alan kesme puanı algoritması kullanılarak $\alpha = 0.05$ düzeyi için kesme puanı elde edilmiştir. Elde edilen kesme puanından büyük olanlara (“cevap kopyalamadan kaynaklı test hilesi var” kararı verilen olası kopyacı için) 1, küçük olanlara ise (“cevap kopyalamadan kaynaklı test hilesi yok” kararı) 0 değeri verilerek bir 1-0 matrisi oluşturulmuştur. Matriste yer alan

“1” değerlerinin toplamı seçilen olası kopyacı sayısına oranı hesaplanmıştır. Böylece indekslerin verdiği kararın doğru olma olasılığı (güç oranı) ortaya konmuştur.

KL indeksinin güç oranını ortaya koymak için hem manipülasyonun yapılmadığı verilerle hem de kopya çekilen madde oranının manipüle edildiği/kopya çekildiği bilinen verilerle tekrarlanan her yinelemede tüm bireylerin cevap örüntülerinin yer aldığı (1-0) cevap matrisi, testin parametreleri kullanılarak R programında yer alan “*irtoys*” paketi (Partchev, 2016) ile yetenek kestirimi yapılmıştır. Yetenek kestirimleri yapıldıktan sonra R programında yer alan “*LaplaceDemon*” paketi (Singmann, 2020) kullanılarak bireylerin *KL* indeksi değerleri hesaplanmıştır. Bu işlemden sonra *KL* indeksi için R programında yer alan “*OptimalCutpoints*” paketinde (Raton, 2014) yer alan *Youden İndeksi* ile $\alpha = 0.05$ düzeyi için kesme puanı elde edilmiştir. Elde edilen kesme puanından büyük olanlara (“cevap kopyalamadan kaynaklı test hilesi var” kararı verilen olası kopyacı için) 1, küçük olanlara ise (“cevap kopyalamadan kaynaklı test hilesi yok” kararı) 0 değeri verilerek bir 1-0 matrisi oluşturularak, matriste yer alan “1” değerlerinin toplamı seçilen olası kopyacı sayısına oranı hesaplanmıştır. Böylece *KL* indeksinin verdiği kararın doğru olma olasılığı (güç oranı) elde edilmiştir.

Aşamalı adımda ise öncelikle birey uyum (*KL* indeksi) istatistikleri kullanılarak olası kopyacılar belirlenmiştir. Olası kopyacıların sınıfta oturduğu yere göre geriye kalan potansiyel kaynak (bireyin önünde, sağında ya da solunda yer alan bireyler) olarak görülen bireylerden biri seçilerek, olası kopyacı birey ile kopya analizine dahil edilmiştir. Kopya ve benzerlik indekslerinin (ω ve *GBT*) kopya belirlemede güç oranlarının belirlenmesi için kopya çekilen madde oranının manipüle edildiği/kopya çekildiği bilinen verilerle tekrarlanan her yinelemede, seçilen olası kopyacı ve kaynak çifti için R programında yer alan “*CopyDetect*” paketi (Zopluoğlu, 2018) kullanılarak kopya ve benzerlik indekslerinin olasılık değerleri hesaplanmıştır. Hesaplanan olasılık değerleri için $\alpha = 0.05$ düzeyinden küçük ve eşit olanlara (“cevap kopyalamadan kaynaklı test hilesi var” kararları verilen çiftler için) 1, büyük olanlara ise (“cevap kopyalamadan kaynaklı test hilesi yok” kararları verilen çiftler için) 0 değeri verilerek bir 1-0 matrisi oluşturulmuştur. Matriste yer alan “1” değerlerinin toplamı, iki aşamalı analizde birey uyum (*KL* indeksi) istatistiği kullandıktan sonra elde edilen olası kopyacı bireyler (başlangıçta manipüle edilerek seçilmiş olası kopyacılar arasından belirlenenler)-kaynak çifti sayısına oranı hesaplanmıştır. Böylece indekslerin seçilen çiftler için verdiği kararın doğru olma olasılığı (güç oranı) elde edilmiştir.

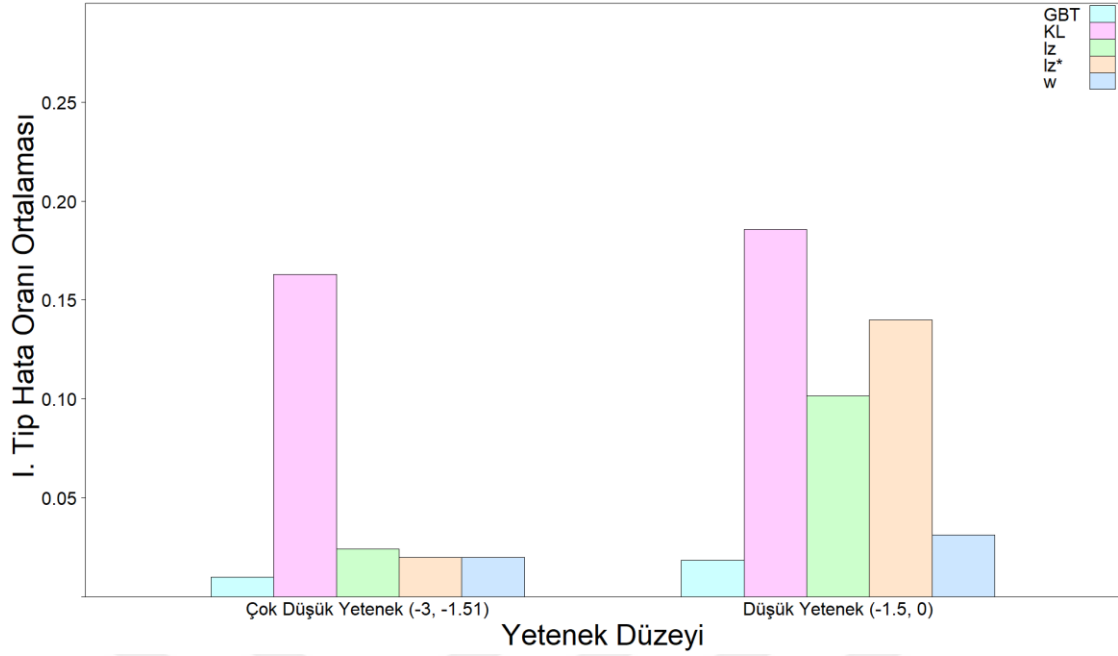
BÖLÜM 4

BULGULAR VE YORUMLAR

Bu bölümde, verilerin analizi sonucunda elde edilen bulgulara ve bulgulara ilişkin yorumlara yer verilmiştir. İlgili araştırma soruları okuyucuya kolaylık olması için bulgulardan önce yeniden yazılarak sonrasında bulgular sunulmuştur (Bu konuda daha ayrıntılı bilgi için bkz. EK2)

İndekslerin Aşamasız Kullanımı Sonucunda Elde Edilen I. Tip Hata Oranlarına Yönelik Bulgular

Testin ortalama güçlüğü ve kopyacının yetenek düzeyinin; kopya ve benzerlik indeksleri, birey uyum istatistikleri ve KL indeksinin aşamasız olarak kullanıldığı durumda, indekslerin I. tip hata oranına olan temel etkisi. Çalışmanın birinci alt probleminde müdahale edilmemiş verilere yönelik test hilesini belirleme amacıyla analizler gerçekleştirilmiştir. Analizde kullanılan testin ortalama güçlüğü ve kopyacının yetenek düzeyi değişkenlerinin indekslerin I. tip hatasına olan temel etkilerine yönelik bulgular elde edilmiştir. Şekil I'de yetenek düzeyinin, Şekil II'de ise testin ortalama güçlüğü değişkenlerinin indekslerin I. Tip hata oranlarına olan temel etkilerine yönelik elde edilen bulgular sunulmuştur.

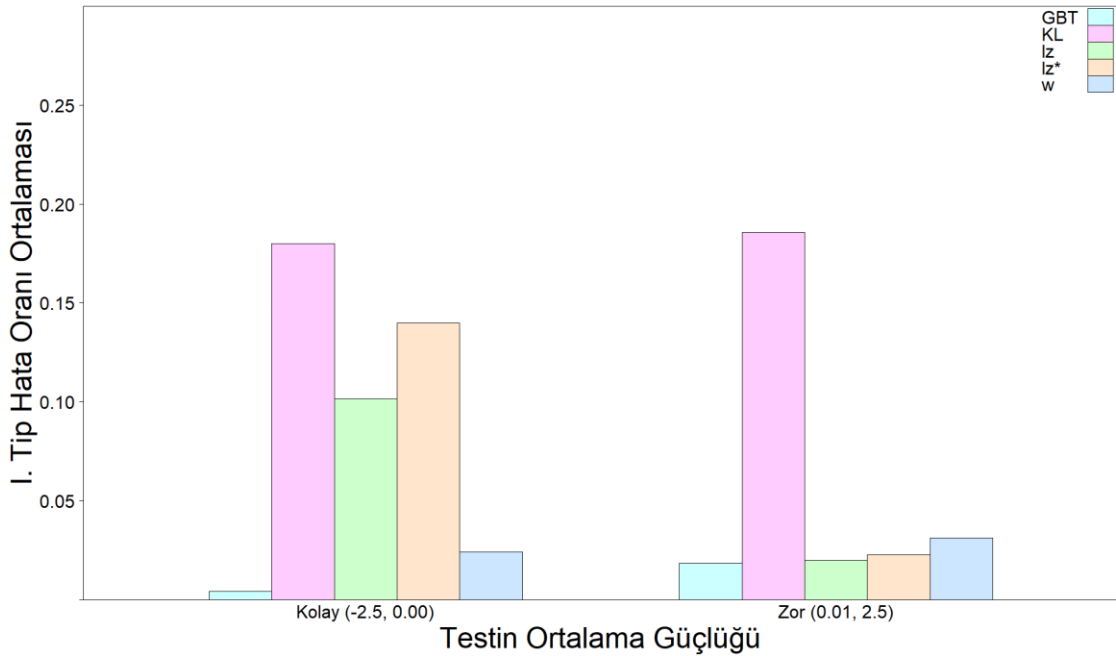


Şekil 1. Yetenek Düzeyinin I. Tip Hata Oranı Ortalamasına Olan Temel Etkisinin Grafiği. Farkların daha görülebilir olması için (0,0.3) arasında ölçeklendirme yapılmıştır.

Şekil 1'e göre yetenek düzeyi arttıkça KL , ω , GBT , l_z ve l_z^* istatistiklerinin I. tip hata oranları ortalamasının arttığı görülmektedir. l_z istatistiği çok düşük yetenek düzeyinde l_z^* istatistiği ile yakın I. tip hata oranı ortalamalarına sahipken, düşük yetenek düzeyinde ise l_z^* istatistiği l_z istatistiğine göre fazla I. tip hata oranı ortalaması sergilemiştir. Tenderio ve Meijer (2012) çalışmasında düşük yetenek düzeyindeki bireylerde l_z istatistiğinin I. tip hatasını yüksek olduğunu ifade etmiştir. Bu çalışmada l_z istatistiğinin I. tip hatasını yüksek elde edilmiştir. l_z^* istatistiğinin, testteki maddelerin güçlük düzeyi artacak şekilde ve 1 (doğru) cevabının solunda önemli miktarda 0 (yanlış) cevabın olduğu tepki örüntülerinden oldukça etkilendiği ifade edilmiştir (Meijer ve Tenderio, 2014). Tenderio ve Meijer (2014) çalışmalarında parametrik olmayan madde tepki modellerine ait birey uyum istatistikleriyle birlikte l_z^* istatistiğinin de kullanarak I. tip hatalarını karşılaştırdıkları bir çalışma gerçekleştirmişlerdir. Çalışma sonucunda istatistiğinin I. tip hata değerini yüksek bulmuşlardır. Bu çalışmada kullanılan örneklem büyüklüğünün etkisinden dolayı benzer bulgular elde edilmiştir. Armstrong, Stoumbos, Kung ve Shi (2007) çalışmasında l_z istatistiğinin düşük yetenek düzeyinde ve kolay testte iyi performans göstermediği sonucuna ulaşılmıştır. Ayrıca Armstrong, Stoumbos, Kung ve Shi (2007) l_z istatistiğinin

I. tip hatasının yüksek çıkmasına neden olarak tepki örüntüsünde yanlış cevabın çok olduğu durumlarda, istatistiğin bireyi hile yapmıştır olarak gösterdiği ifade edilmiştir. van Krimpen-Stoop ve Meijer (1999), l_z^* istatistiğinin 2-parametrelili logistik modele uygun olmakla birlikte yetenek kestirimi için Warm'ın ağırlıklı olasılık kestirimi yöntemine uygun yapılandırıldığını belirtmişlerdir. Bunun yanı sıra, l_z asimptotik standart normal dağılım için uygun iken l_z^* istatistiği ise standart normal dağılım için uygundur (Magis, Raïche ve Beland, 2012). Bu çalışmada kullanılan dağılım tam olarak standart normal dağılım göstermemektedir. Çalışmada Rasch model kullanılmıştır ve yetenek kestirimleri ise maksimum olasılık yöntemiyle yapılmıştır. Bu sebeplerden dolayı, l_z^* istatistiği çok daha fazla beklenmeyen tepki örüntüsü belirlemiştir. Çok düşük yetenek düzeyinde KL indeksi haricinde genel olarak ω , GBT , l_z ve l_z^* istatistiklerinin I. tip hata oranı ortalamaları değeri bir birine yakın değerler almıştır.

Kopya ve benzerlik indeksleri (ω ve GBT) her iki koşul altında düşük I. Tip hata oranı ortalamaları üretmiştir. Bununla birlikte her iki yetenek düzeyinde de KL indeksinin I. tip hata oranları ortalaması en yüksek, GBT indeksinin ise düşük elde edilmiştir.

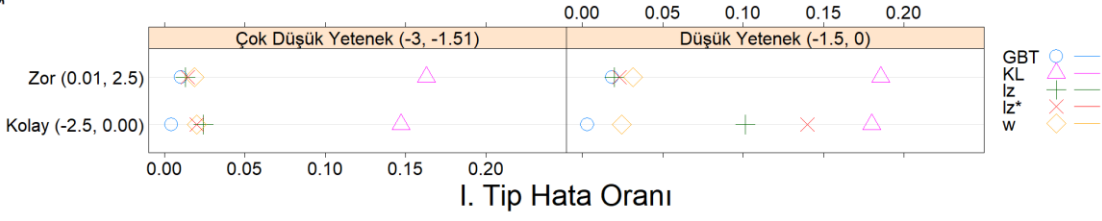


Şekil II. Testin Ortalama Güçlüğü'nün I. Tip Hata Oranı Ortalamasına Olan Temel Etkisinin Grafiği. Farkların daha görülebilir olması için (0,0.3) arasında ölçeklendirme yapılmıştır.

Şekil II'ye göre testin ortalama güçlüğü arttıkça *KL* indeksi haricinde ω , *GBT*, l_z ve l_z^* istatistiklerinin I. tip hata oranları ortalamasının azaldığı görülmektedir. Birey uyum istatistiklerinde durum; l_z istatistiği kolay testte l_z^* istatistiğinden düşük I. tip hata oranı ortalamalarına sahipken, zor testte ise l_z^* ve l_z istatistiklerinin I. tip hata oranı ortalaması birbirine yakın değerler sergilediği gözlenmiştir. Zor testte *KL* indeksi haricinde genel olarak I. Tip hata oranı ortalamaları değeri bir birine yakın değerler almıştır. Kopya ve benzerlik indeksleri (ω ve *GBT*) her iki koşul altında düşük I. tip hata oranı ortalamaları üretmiştir. Bununla birlikte her iki yetenek düzeyinde de *KL* indeksinin I. tip hata oranları ortalaması en yüksek, *GBT* indeksinin ise düşük elde edilmiştir.

KL indeksi değerlerinin hesaplanabilmesi için bireylere ait yetenekler, ayrı bir paket kullanılarak hesaplanmıştır. Hesaplanan yetenek parametreleri kullanılarak önsel ve sonsal dağılım arasındaki fark ile *KL* indeksi değeri hesaplanmıştır. *KL* indeksi fark yaratan tüm durumlardan etkilenmektedir. Yetenek parametrelerinin kestiriminden kaynaklanan farklılıklarla birlikte, özgüllük ve duyarlık kombinasyonunun maksimizasyonu/minimizasyonu dengeleyerek sınıflamayı amaçlayan *Youden İndeksi* yaklaşımıyla elde edilen kesme puanından dolayı yüksek birinci tip hata değerleri ortaya çıkmış olabilir. Bu durum göz önünde bulundurulduğunda manipüle edilen değişkenlerin temel etkilerinde en yüksek I. tip hata oranı ortalaması *KL* indeksi ait olduğu gözlenmiştir.

Testin ortalama güçlüğü ve kopyacının yetenek düzeyinin; kopya ve benzerlik indeksleri, birey uyum istatistikleri ve KL indeksinin aşamasız olarak kullanıldığı durumda, indekslerin I. tip hata oranına olan ortak etkisi. Analizde kullanılan testin ortalama güçlüğü kopyacının yetenek düzeyi değişkenlerinin indekslerin I. tip hatasına olan ortak etkilerine yönelik bulgular elde edilmiştir. Şekil III'de değişimlenen faktörlerin indekslerin I. tip hata oranlarına olan ortak etkilerine yönelik elde edilen bulgular sunulmuştur.



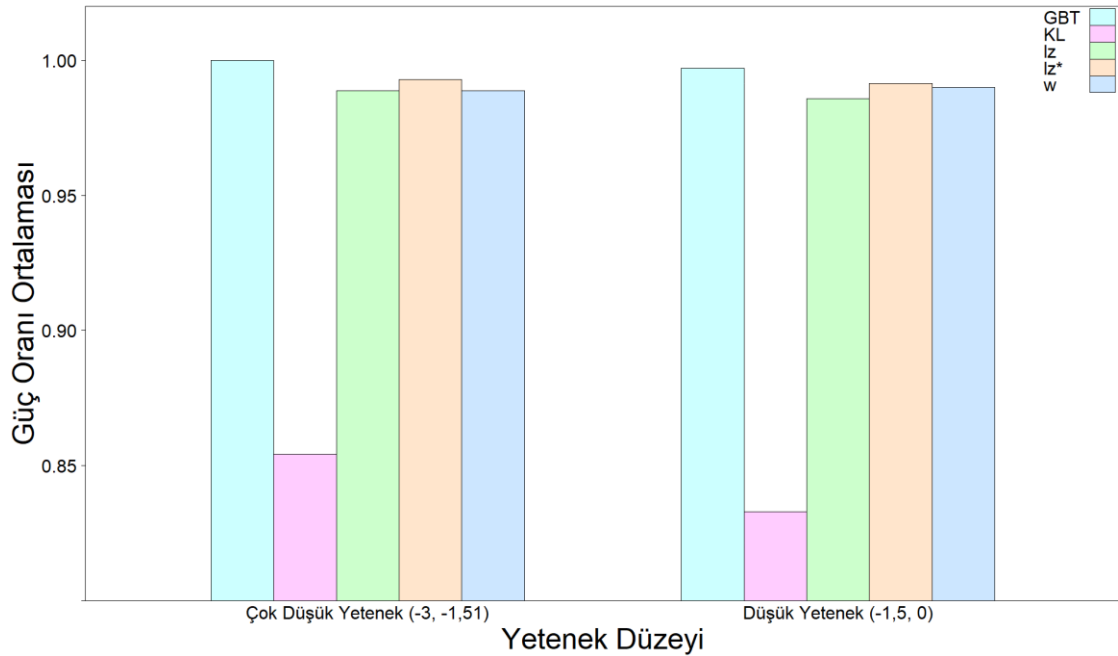
Şekil III. Değişimlenen Faktörlerin I. Tip Hata Oranları Olan Ortak Etkisinin Grafiği. Farkların daha görülebilir olması için (-0.01, 0.25) arasında ölçeklendirme yapılmıştır.

Şekil III'e göre testin kolay ve yetenek düzeyinin düşük olduğu koşul dışında *KL* indeksi haricinde ω ve *GBT* indekslerinin, l_z ve l_z^* istatistiklerinin I. tip hata oranlarının düşük ve birbirine yakın değerler aldığı gözlenmiştir. ω ve *GBT*'nin güç çalışmalarında benzer performans gösterdiğini belirten Zopluoğlu (2012) bu nedenle çalışmasında ω ve *GBT*'nin I. tip hatalarını karşılaştırmıştır. Zopluoğlu'nun (2012) çalışmasındaki sonuçlara benzer olarak bu çalışmanın bulgularında elde edildiği üzere yetenek düzeylerinde *GBT*'nin ω 'ya göre daha az I. tip hataya sahip olduğu görülmüştür. Bununla birlikte tüm şartlar altında *GBT* indeksinin I. tip hata ortalaması ω indeksinin I. tip hata oranına göre daha düşük elde edilmiştir. Sünbül ve Yormaz (2018a, 2018b) çalışmasıyla benzer bulgu elde edilmiştir.

KL indeksi ise tüm koşullar altında ω ve *GBT*, l_z ve l_z^* istatistiklerine göre yüksek I. tip hata üretmiştir. Kolay testte gerçekleştirilen analiz sonuçlarında l_z ve l_z^* istatistiğinin I. tip hatalarının arttığı gözlenmiştir. Testin kolay olduğu ve yetenek düzeyinin düşük olduğu koşulda indekslerin I. tip hata oranları değerleri en düşükten en yükseğe *GBT*, ω , l_z , l_z^* ve *KL* indeksleri şeklinde sıralandığı görülmüştür.

İndekslerin Aşamasız Kullanımı Sonucunda Elde Edilen Güç Oranlarına Yönelik Bulgular

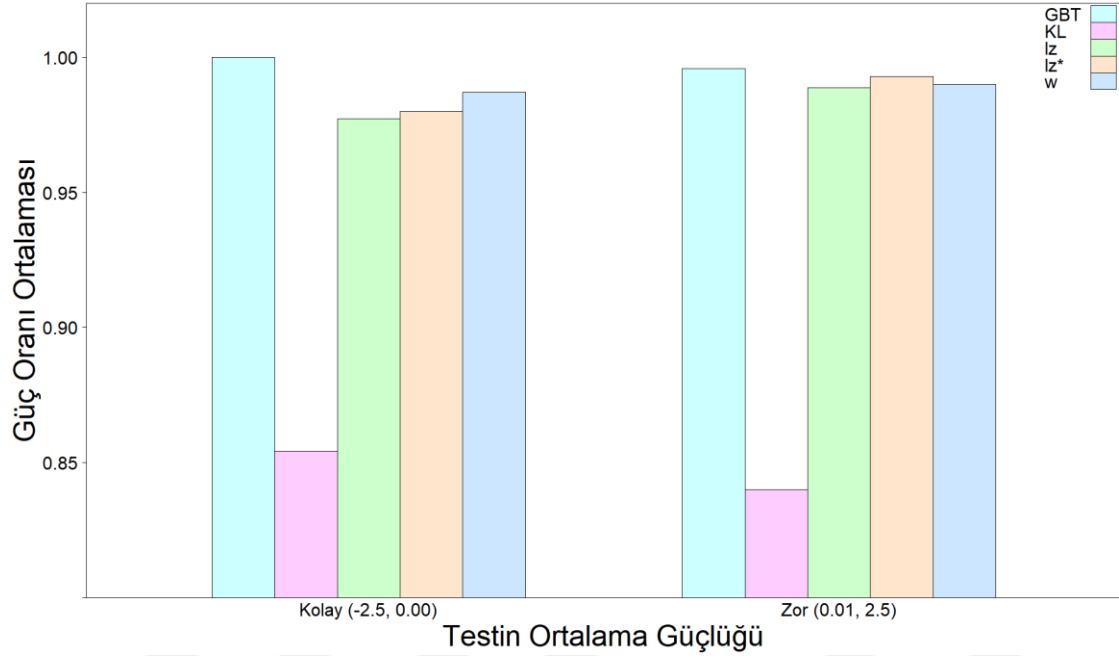
Testin ortalama güçlüğü ve kopyacının yetenek düzeyinin; kopya ve benzerlik indeksleri, birey uyum istatistikleri ve KL indeksinin aşamasız olarak kullanıldığı durumda, indekslerin güç oranına olan temel etkisi. Çalışmanın ikinci alt probleminde müdahale edilmiş verilere yönelik test hilesini belirleme amacıyla analizler gerçekleştirilmiştir. Analizde kullanılan testin ortalama güçlüğü, kopyacının yetenek düzeyi ve kopya oranı değişkenlerinin indekslerin gücüne olan temel etkilerine yönelik bulgular elde edilmiştir. Şekil IV’de kopyacının yetenek düzeyinin, Şekil V’de testin ortalama güçlüğü, Şekil VI’da kopya oranı değişkenlerinin indekslerin gücüne olan temel etkilerine yönelik elde edilen bulgular sunulmuştur.



Şekil IV. Yetenek Düzeyinin Güç Oranı Ortalamasına Olan Temel Etkisinin Grafiği. Farkların daha görülebilir olması için (0.8,1.02) arasında ölçeklendirme yapılmıştır.

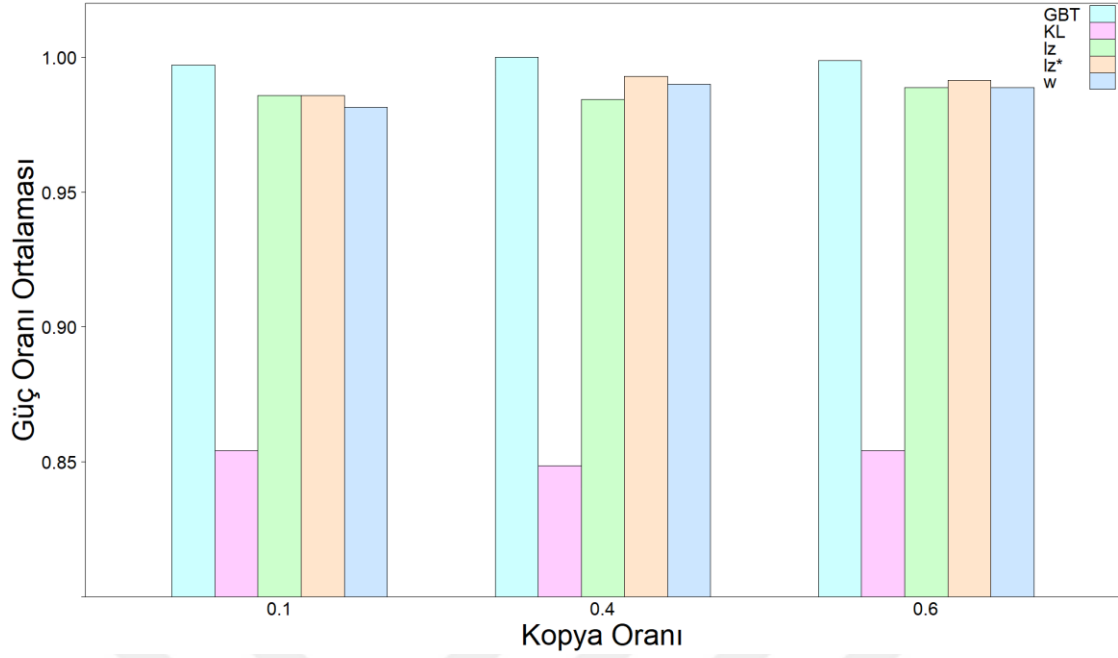
Şekil IV’e göre çok düşük ve düşük yetenek düzeylerinde kullanılan GBT , ω , l_z ve l_z^* istatistiklerinin teste hile yapan bireyleri belirleme güçleri ortalamalarının yüksek olduğu görülmüştür. Tüm yetenek düzeylerinde GBT indeksi en yüksek belirleme gücüne sahip olmasıyla birlikte, GBT , ω , l_z ve l_z^* istatistiklerinin hile yapan bireyleri belirleme güçleri birbirlerine yakın performans göstermiştir. Birey uyum istatistikleri kendi

aralarında karşılaştırıldığında ise l_z^* istatistiğinin l_z istatistiğine göre belirleme gücü fazladır. Ayrıca KL indeksi, GBT , ω , l_z ve l_z^* istatistiklerine göre hile yapan bireyleri belirlemede düşük performans sergilemiştir.



Şekil V. Testin Ortalama Güçlüğü'nün Güç Oranı Ortalamasına Olan Temel Etkisinin Grafiği. Farkların daha görülebilir olması için (0.8,1.02) arasında ölçeklendirme yapılmıştır.

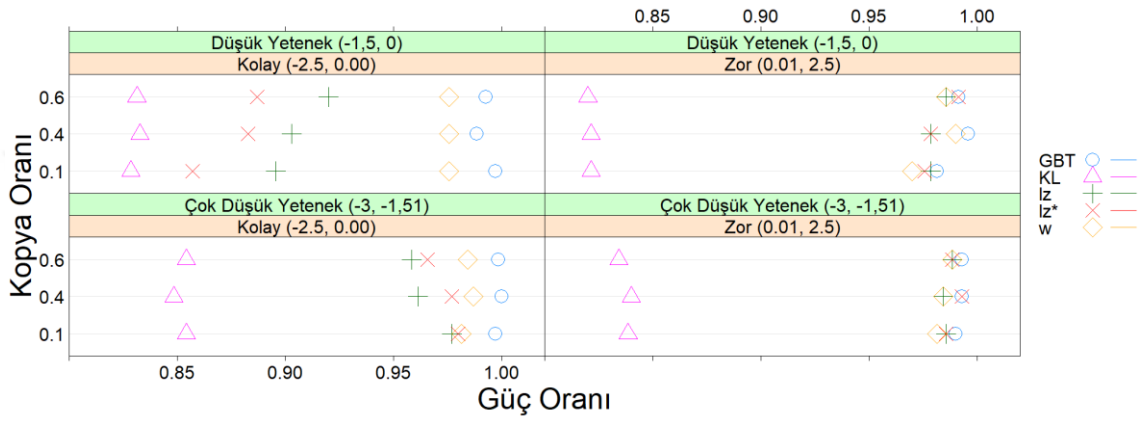
Şekil V'e göre testin ortalama güçlüğü düzeylerinde, testte hile yapan bireyleri belirlemede kullanılan GBT , ω , l_z , l_z^* ve KL indeksinin güçleri genel olarak yüksek değerler almıştır. Ayrıca KL indeksi dışında GBT , ω , l_z ve l_z^* istatistiğinin hile yapan bireyi belirleme güçleri birbirine yakın değerler aldığı gözlenirken, kolay testte, l_z , l_z^* ve ω indekslerinin sırasıyla test belirleme güçleri küçükten büyüğe sıralı değerler almıştır. Hile yapan bireyi belirleme gücü en düşük olan indeksin KL indeksi olduğu görülmüştür. KL indeksinin belirleme gücü testin ortalama güçlüğü arttığında azalmıştır. Bununla birlikte ω , l_z ve l_z^* istatistiğinin belirleme gücünde ise testin ortalama gücü arttığında bir artış olmuştur. Her iki düzeyde de hile yapan bireyleri belirlemede en yüksek performansı GBT göstermiştir.



Şekil VI. Kopya Oranının Güç Oranı Ortalamasına Olan Temel Etkisinin Grafiği. Farkların daha görülebilir olması için (0.8,1.02) arasında ölçeklendirme yapılmıştır.

Şekil VI'ya göre kopya oranının tüm değişimlenen düzeylerinde *GBT* indeksinin en yüksek belirleme gücüne sahip olduğu görülmüştür. *GBT* indeksinin tüm kopya oranı düzeylerinde yüksek güçlüğü sahip olmakla birlikte, kopya oranının arttığında ω indeksinin hile yapan bireyi belirleme gücünün de arttığı görülmüştür. Bu sonuç Zopluoğlu (2012, 2016), Sotaridona ve Meijer (2002, 2003) çalışmalarında elde ettiği sonuçlarla benzerlik göstermiştir. Bununla birlikte *KL* indeksi dışında *GBT*, ω , l_z ve l_z^* istatistiğinin test hilesi yapan bireyleri belirleme güçleri birbirine yakın değerler almıştır. Ayrıca kopya oranı en düşük seviyede iken birey uyum istatistiklerinin belirleme güçleri aynı olduğu, kopya oranı orta düzeye olduğunda ise l_z^* istatistiğinin belirleme gücünün l_z istatistiğinin belirleme gücünden fazla olduğu, kopya oranı yüksek iken l_z istatistiğinin belirleme gücünün arttığı ancak l_z^* istatistiğinin belirleme gücünden düşük kaldığı görülmüştür. Kopya oranının tüm düzeylerinde *KL* indeksinin test hilesi yapan bireyi belirleme gücü *GBT*, ω , l_z ve l_z^* istatistiğinin belirleme gücüne göre düşük performans göstermiştir.

Testin ortalama g¼c¼l¼g¼, kopyacının yetenek d¼zeyi ve kopya oranının; kopya ve benzerlik indeksleri, birey uyum istatistikleri ve KL indeksinin aşamasız olarak kullanıldıđı durumda, indekslerin g¼c¼ne olan ortak etkisi. Analizde kullanılan testin ortalama g¼c¼l¼g¼, kopyacının yetenek d¼zeyi ve kopya oranı deđiřkenlerinin indekslerin g¼c¼ne olan ortak etkilerine y¼nelik bulgular elde edilmiřtir. řekil VII’de deđiřimlenen fakt¼rlerin indekslerin g¼c¼ne olan ortak etkilerine y¼nelik elde edilen bulgular sunulmuřtur.



řekil VII. Deđiřimlenen Fakt¼rlerin G¼c¼ Oranına Olan Ortak Etkisinin Grafiđi. Farkların daha g¼r¼lebilir olması için (0,8,1,02) arasında ¼lçeklendirme yapılmıřtır.

řekil VII incelendiđinde, *KL* indeksinin hile yapan bireyi tespit etmedeki g¼c¼ çok düşük yetenek d¼zeyinde Belov’un (2014a) elde ettiđi sonuçlara benzer olarak daha yüksek olduđu g¼r¼lm¼řt¼r. Bununla birlikte *KL* indeksi t¼m kořullar altında alıřmada kullanılan diđer indekslerin gerisinde bir g¼ce sahip olduđu g¼zlenmiřtir. T¼m kořullar altında *GBT* indeksinin en y¼ksek g¼c¼ oranına sahip olduđu g¼r¼lm¼řt¼r. Sotaridona ve Meijer (2003) k¼c¼k (100) ¼rneklemelerde *K* indeksi kadar ω indeksinin de iyi performans g¼sterdiđini belirtmiřtir. řekil VII’ ye g¼re *GBT* indeksinden sonra en y¼ksek g¼c¼ ω indeksinde elde edilmiřtir. Elde edilen bu bulgulara g¼re t¼m kořullar altında en y¼ksek g¼c¼ deđerlerini kopya ve benzerlik istatistikleri ¼retmiřtir. Test hilesi belirleme alıřmalarında kopya eken bireyin yetenek d¼zeyi ile kaynađın yetenek d¼zeyi arasındaki fark arttıka indekslerin testte hile yapanları belirleme g¼c¼lerinin arttıđı g¼r¼lm¼řt¼r (Yormaz, 2017; Zopluođlu, 2012). Gerekleřtirilen bu alıřmada da genel olarak ok d¼ř¼k yeteneđe sahip ve hile yapan bireylerin indeksler tarafından tespit oranı *KL* indeksi dıřında *GBT*, ω , l_z ve l_z^* istatistiđinde y¼ksek g¼r¼lm¼řt¼r.

ω ve *GBT* indekslerinin performanslarına yönelik sonuçların olduğu çalışmada *GBT* indeksinin bazı koşullarda test hilesi yapan bireyi belirleme gücünün performansı ω indeksinin belirleme gücü performansından daha iyi olduğu belirtilmiştir (Zopluoğlu, 2012). Bununla birlikte geri kalan şartlarda ω indeksi ile benzer güce sahip olduğu ifade edilmiştir. Bu çalışmada da zor testte ve yetenek düzeyi ile kopya oranının tüm şartlarında *GBT* ve ω indeksinin genelde benzer sonuçlar göstermesiyle birlikte kolay testte ve yetenek düzeyi ile tüm kopya oranlarında *GBT* indeksinin belirleme gücünün daha iyi performans gösterdiği elde edilmiştir.

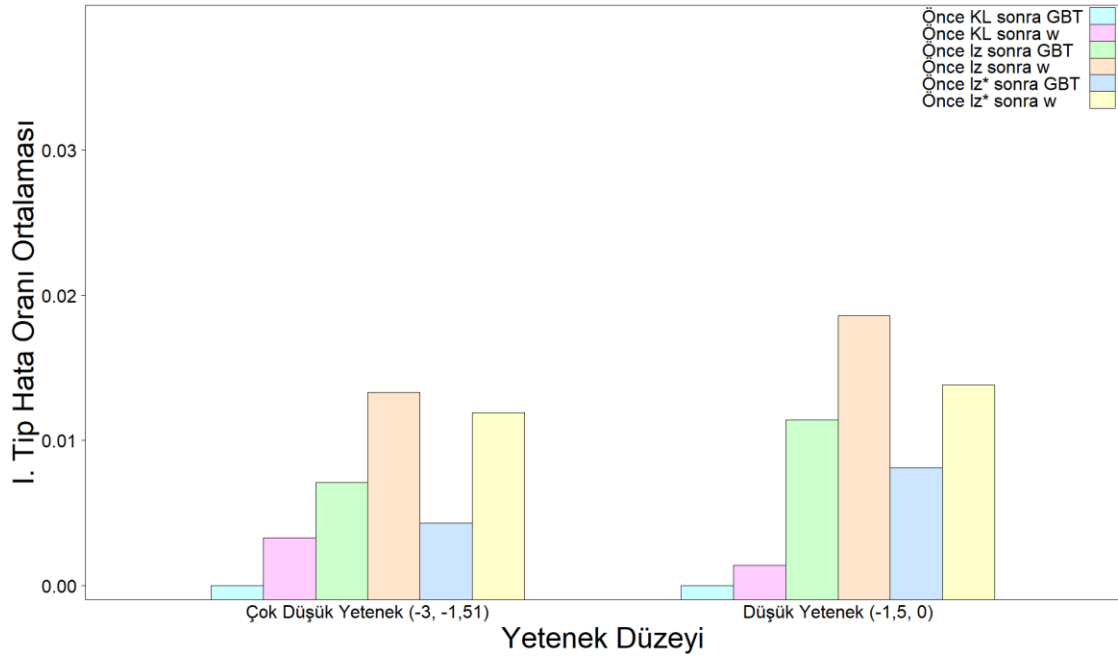
Çalışmada ω ve *GBT* indeksleri için her ortalama test güçlük düzeyinde hile yapanları belirleme gücü performansının yüksek olduğu görülmüştür. Başka bir ifade ile ω ve *GBT* indekslerine ait elde edilen bulguların, Sünbül ve Yormaz'ın (2018a, 2018b) çalışmalarında $\alpha = 0.05$ düzeyi için elde ettiği sonuçlara benzer olduğu görülmüştür.

l_z^* istatistiğinin gücü, kolay testte ve düşük yetenek ile tüm kopya oranı düzeyleri dışında genel olarak ω indeksinden büyük olarak elde edilmiştir. Elde edilen sonuçlar Voncken (2014) çalışmasıyla benzerlik göstermiştir. Şekil IV, Şekil V ve Şekil VI'ya göre l_z^* ve l_z istatistiklerinin hile yapanları belirleme güçleri tüm koşulların temel etkisinde bir birine yakın değerler verdiği gözlenmiştir. Şekil VII'ye göre kolay testte ve düşük yeteneğin tüm kopya oranları düzeyleri dışında diğer tüm koşullarda temel etki grafiklerinde görülen sonuca benzer olarak birey uyum istatistiklerinin güçleri birbirine yakın değerler almıştır. Bununla birlikte l_z^* istatistiğinin gücü l_z istatistiğinin gücüne göre daha yüksek olduğu gözlenmiştir. l_z^* istatistiğinin gücünün yüksek olduğu sonucu Magis, Raîche ve Beland (2012) elde ettiği bulgularla örtüştüğü görülmüştür.

KL indeksinin test hilesi yapan bireyi belirleme gücüne bakıldığında performansının l_z istatistiğinden düşük olduğu gözlenmiştir. Elde edilen bulgu Belov (2015) çalışmasında kağıt-kalem ve bilgisayar tabanlı testlerde elde ettiği bulguya benzer olduğu görülmüştür. Düşük yetenek, kolay test ve kopya oranının 0.1, 0.4 ve 0.6 olarak değiştiği koşullarda kullanılan indekslerin performanslarındaki değişim diğer koşullara göre en düşükten yükseğe *KL* indeksi, l_z^* , l_z ve kopya benzerlik indeksleri şeklinde sıralanmıştır. Çok düşük yetenek ve kolay testin kullanıldığı şartlar dışında kopya oranı arttıkça, l_z^* , ve l_z istatistiklerinin gücünün arttığı görülmüştür. Bu sonuç Reise (1995) de ifade ettiği yetenek seviyeleri ile madde güçlük parametreleri arasındaki fark büyük olduğunda l_z istatistiğinin iyi belirleme performansı elde edildiği bulgusuyla paralellik göstermiştir.

İndekslerin İki Aşamalı Kullanımı Sonucunda Elde Edilen I. Tip Hata Oranlarına Yönelik Bulgular

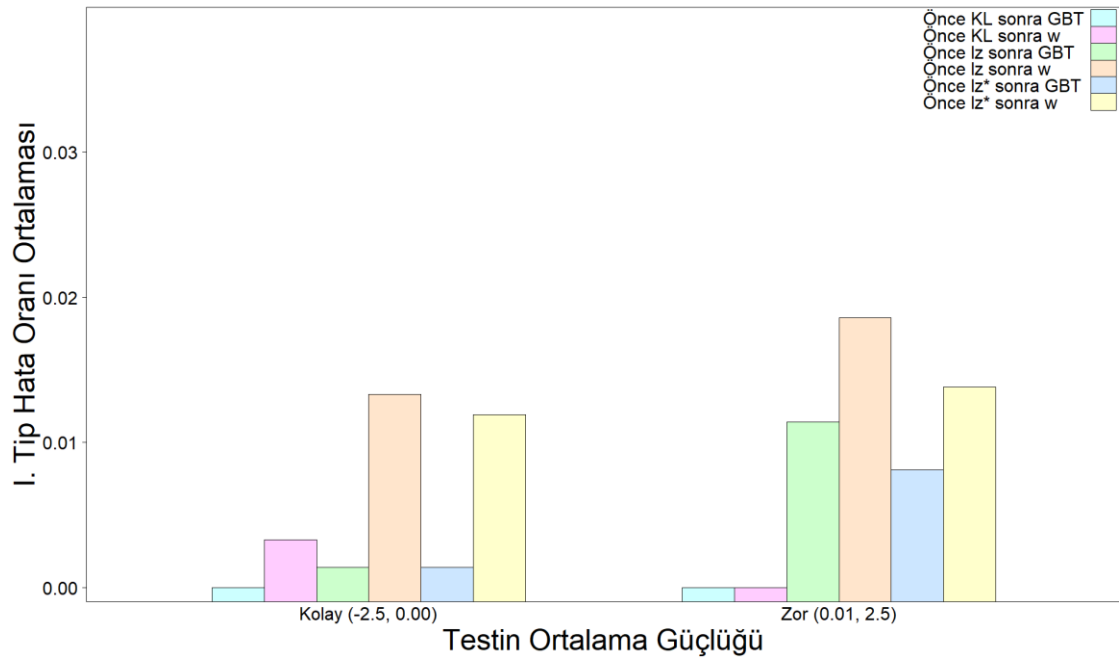
Testin ortalama güçlüğü, kopyacının yetenek düzeyinin; kopya ve benzerlik indeksleri, birey uyum istatistikleri ve KL indeksinin iki aşamalı olarak kullanıldığı durumda, kopya ve benzerlik indekslerinin I. tip hata oranına olan temel etkisi. Çalışmanın üçüncü alt problemde müdahale edilmemiş verilere yönelik test hilesini belirleme amacıyla iki aşamalı olarak gerçekleştirilen analizlerde kullanılan testin ortalama güçlüğü ve kopyacının yetenek düzeyi değişkenlerinin ω ve GBT indekslerinin I. tip hatasına olan temel etkilerine yönelik bulgular elde edilmiştir. Şekil VIII’de yetenek düzeyinin, Şekil IX’da ise testin ortalama güçlüğü değişkenlerinin indekslerin I. Tip hata oranlarına olan temel etkilerine yönelik elde edilen bulgular sunulmuştur.



Şekil VIII. Yetenek Düzeyinin I. Tip Hata Oranları Ortalamasına Olan Temel Etkisinin Grafiği. Farkların daha görülebilir olması için (-0.001,0.04) arasında ölçeklendirme yapılmıştır.

Şekil VIII’e göre her iki yetenek düzeyinde l_z birey uyum istatistiği uygulandıktan sonra kullanılan ω istatistiğinin gücünün diğer analizlere göre yüksek elde edilmiştir. Her iki yetenek düzeyinde, KL indeksi sonrasında kullanılan GBT indeksinin I. tip hata oranı ortalaması önce l_z sonra ω , önce l_z^* sonra ω , önce l_z sonra GBT ve önce l_z^* sonra GBT ’nin

kullanıldığı aşamalı analiz sonucunda GBT ve ω indeksinin I. tip hata oranı ortalamasına göre en düşük değere sahip olduğu görülmüştür. Genel olarak iki aşamalı analizde ω ve GBT indekslerinin I. tip hata oranları ortalaması düşük elde edilmiştir. Bununla birlikte yetenek düzeyi arttıkça I. tip hata oranları ortalaması değerlerinin arttığı gözlenmiştir. Yetenek düzeyi arttığında l_z istatistiği sonrasında kullanılan ω ve yine l_z istatistiği sonrasında kullanılan GBT indeksinin I. tip hata oranları ortalamasının arttığı görülmüştür.

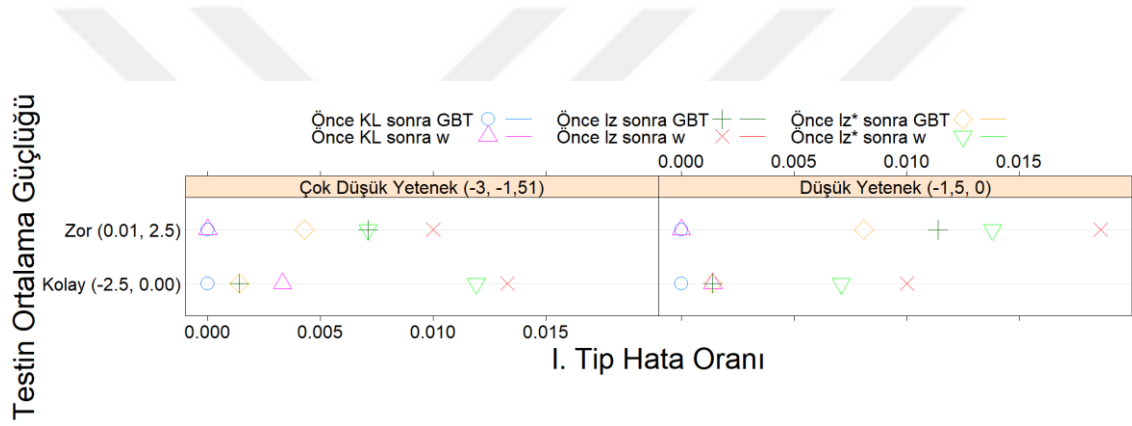


Şekil IX. Ortalama Test Güçlüğü'nün I. Tip Hata Oranları Ortalamasına Olan Temel Etkisinin Grafiği. Farkların daha görülebilir olması için (-0.001,0.04) arasında ölçeklendirme yapılmıştır.

Şekil IX incelendiğinde genel olarak iki aşamalı analiz sonucunda testin ortalama güçlüğü arttığında ω ve GBT indekslerinin I. tip hata oranları ortalamasının da arttığı gözlenmiştir. Bununla birlikte KL indeksi ile iki aşamalı kullanılan GBT indeksinin her iki test güçlüğü düzeyinde, yine KL indeksi ile iki aşamalı kullanılan ω indeksinin zor testte en düşük I. tip hata oranları ortalaması değerine sahip olduğu görülmüştür. En yüksek I. tip hata oranı ortalaması değeri l_z istatistiği ile iki aşamalı kullanılan ω indeksinde olduğu gözlenmiştir. Testin ortalama güçlüğü'nün her iki düzeyinde ω

indeksinin hem l_z hem de l_z^* istatistikleriyle birlikte iki aşamalı kullanıldığı analizlerde I. tip hata oranı ortalamalarının KL indeksi ile iki aşamalı kullanıldığı analizlere göre daha yüksek olduğu görülmüştür.

Testin ortalama güçlüğü ve kopyacının yetenek düzeyinin; kopya ve benzerlik indeksleri, birey uyum istatistikleri ve KL indeksinin iki aşamalı olarak kullanıldığı durumda, kopya ve benzerlik indekslerinin I. tip hata oranına olan ortak etkisi. Analizde kullanılan testin ortalama güçlüğü kopyacının yetenek düzeyi faktörleri altında ω ve GBT indekslerinin I. tip hatasına olan ortak etkilerine yönelik bulgular elde edilmiştir. Şekil X’da değişimlenen faktörlerin indekslerin I. tip hata oranlarına olan ortak etkilerine yönelik elde edilen bulgular sunulmuştur.



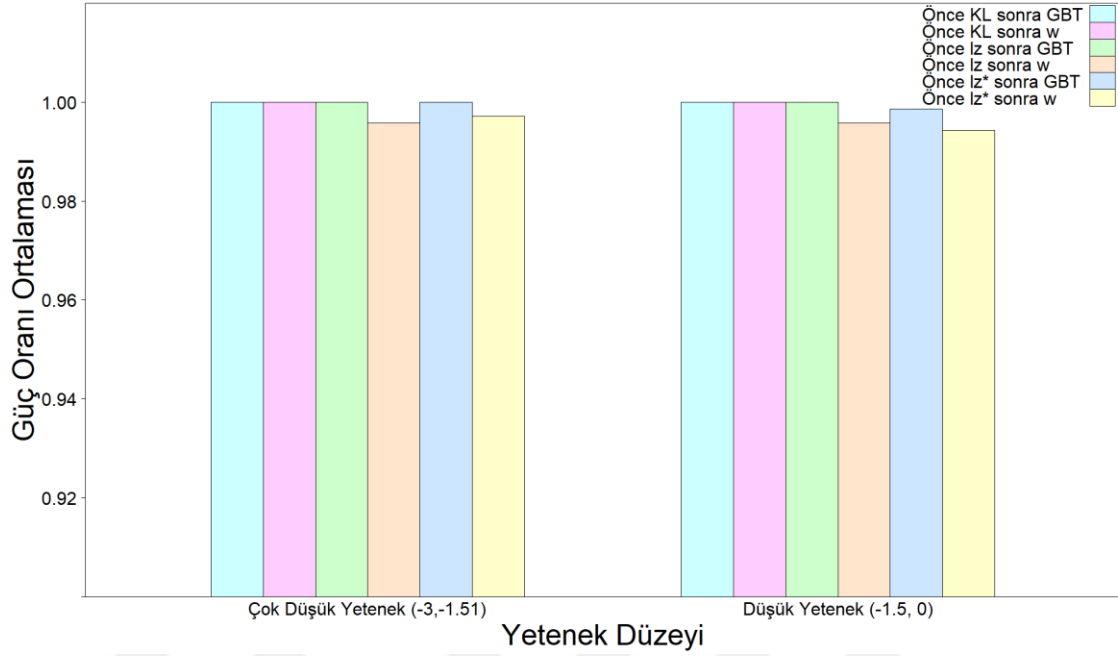
Şekil X. Değişimlenen Faktörlerin I. Tip Hata Oranları Olan Ortak Etkisinin Grafiği. Farkların daha görülebilir olması için (-0.001,0.02) arasında ölçeklendirme yapılmıştır.

Şekil X incelendiğinde düşük yetenek düzeyinde testin ortalama güçlüğü arttığında l_z istatistiği ile iki aşamalı kullanılan ω indeksinin I. tip hata oranlarının arttığı gözlenmiştir. Zor testte tüm yetenek düzeyleri altında KL indeksi ile iki aşamalı kullanılan ω ve KL indeksi ile iki aşamalı kullanılan GBT indekslerinin I. tip hata oranları en düşük değerleri almıştır. Zor testte ise tüm yetenek düzeylerinde en düşük I. tip hata oranları KL indeksi ile iki aşamalı kullanılan GBT indeksine ait olduğu görülmüştür. Bununla birlikte tüm şartlar altında genel olarak, ikinci en düşük I. tip hata oranı ortalamaları l_z^* ile iki aşamalı kullanılan GBT indeksinde gözlenmiştir. Tüm şartlar altında l_z ile iki aşamalı kullanılan ω indeksinin ise üçüncü en yüksek I. tip hata oranına sahip olduğu görülmüştür.

Seol ve Rubright (2013) K indeksinin, KL indeksi ile birlikte iki aşamalı kullanımı sonucunda I. Tip hata oranının tutarlı bir biçimde düşük elde edildiği ifade edilmiştir. Bu çalışmada ω ve GBT indeksinin aşamasız kullanımı ile elde edilen I. tip hata oranlarının KL , l_z ve l_z^* istatistikleri ile iki aşamalı kullanılan ω ve GBT indekslerinin sonucun da elde edilen I. tip hata oranlarına göre daha düşük olduğu gözlenmiştir.

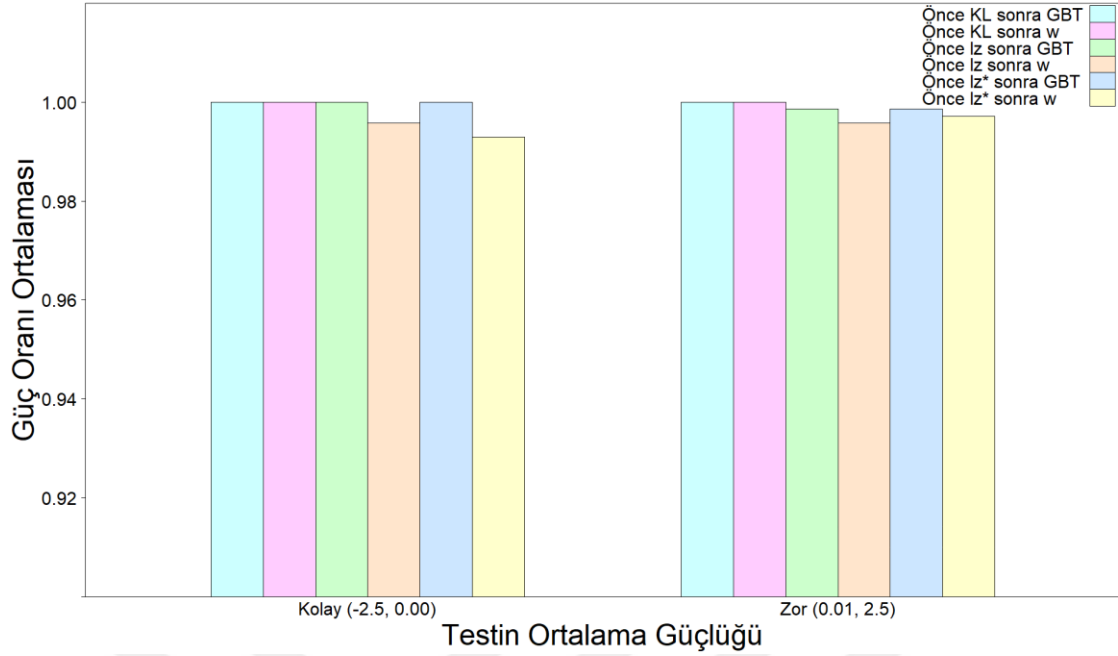
İndekslerin İki Aşamalı Kullanımı Sonucunda Elde Edilen Güç Oranlarına Yönelik Bulgular

Testin ortalama güçlüğü, kopyacının yetenek düzeyi ve kopya oranının; kopya ve benzerlik indeksleri, birey uyum istatistikleri ve KL indeksinin iki aşamalı olarak kullanıldığı durumda, kopya ve benzerlik indekslerinin güç oranına olan temel etkisi. Çalışmanın dördüncü alt probleminde müdahale edilmiş verilere yönelik test hilesini belirleme amacıyla iki aşamalı olarak gerçekleştirilen analizlerde testin ortalama güçlüğü, kopyacının yetenek düzeyi ve kopya oranı faktörleri altında ω ve GBT indekslerinin indekslerin gücüne olan temel etkilerine yönelik bulgular elde edilmiştir. Şekil XI'de kopyacının yetenek düzeyinin, Şekil XII'de testin ortalama güçlüğü'nün, Şekil XIII'de kopya oranı değişkenlerinin indekslerin gücüne olan temel etkilerine yönelik elde edilen bulgular sunulmuştur.



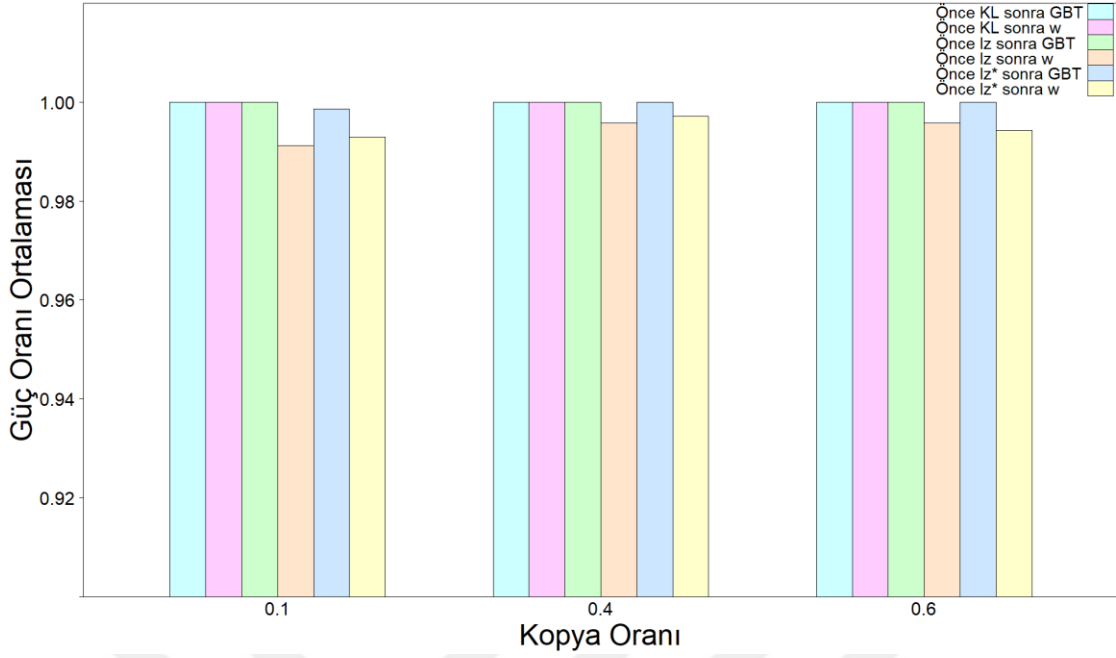
Şekil XI. Yetenek Düzeyinin Güç Oranı Ortalamasına Olan Temel Etkisinin Grafiği. Farkların daha görülebilir olması için (0.9,1.02) arasında ölçeklendirme yapılmıştır.

Şekil XI incelendiğinde yeteneğin tüm düzeylerinde KL , l_z ve l_z^* istatistikleri ile iki aşamalı kullanılan ω ve GBT indekslerinin test hilesi yapan bireyleri belirleme güç oranı ortalamaları genel olarak bir birine yakın değerler almıştır. Bununla birlikte KL indeksi ile iki aşamalı kullanılan ω ve GBT indekslerinin belirleme güç oranı ortalaması her iki yetenek düzeyinde de en yüksek değerleri aldığı gözlenmiştir. l_z ve l_z^* istatistikleri ile iki aşamalı kullanılan ω indeksinin hile yapan bireyleri belirleme güç oranı ortalamaları ise l_z ve l_z^* istatistikleri ile iki aşamalı kullanılan GBT indeksinin belirleme gücü ortalamalarına göre düşük olduğu görülmüştür.



Şekil XII. Ortalama Test Güçlüğü'nün Güç Oranı Ortalamasına Olan Temel Etkisinin Grafiği. Farkların daha görülebilir olması için (0.9,1.02) arasında ölçeklendirme yapılmıştır.

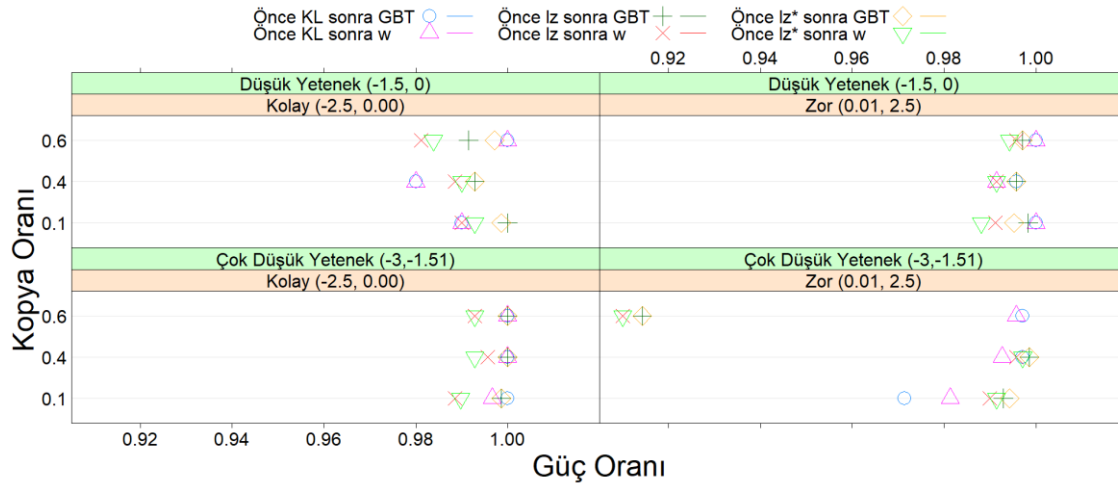
Şekil XII incelendiğinde testin ortalama güçlüğü arttıkça, l_z ve l_z^* istatistikleri ile iki aşamalı kullanılan ω ve GBT indekslerinin test hilesi yapan bireyleri belirleme güçleri ortalamalarının arttığı görülmüştür. Hem kolay testte hem de zor testte KL indeksi ile iki aşamalı kullanılan ω ve GBT indekslerinin belirleme güçleri en yüksek değerleri almıştır. Bununla birlikte testin ortalama güçlüğü'nün tüm düzeylerinde l_z istatistiği ile iki aşamalı kullanılan GBT indeksinin belirleme gücü hem kolay hem de zor testlerde ikinci en yüksek değere sahip olduğu görülmüştür. l_z ve l_z^* istatistikleri ile iki aşamalı kullanılan ω indeksinin hile yapan bireyleri belirleme gücü ortalamaları ise kolay ve zor testte benzer değerleri almıştır.



Şekil XIII. Kopya Oranının Güç Oranına Ortalaması Olan Temel Etkisinin Grafiği. Farkların daha görülebilir olması için (0.9,1.02) arasında ölçeklendirme yapılmıştır.

Şekil XIII incelendiğinde genel olarak kopya oranının tüm düzeylerinde KL , l_z ve l_z^* istatistikleri ile iki aşamalı kullanılan ω ve GBT indekslerinin test hilesi yapan bireyleri belirleme güç oranı ortalamaları genel olarak bir birine yakın değerler almıştır. Bununa birlikte KL ve l_z istatistikleri ile iki aşamalı kullanılan GBT indeksinin belirleme gücü ve KL indeksi ile iki aşamalı kullanılan ω indeksinin test hilesi yapan bireyleri belirleme gücü tüm kopya oranı düzeylerinde en yüksek performansı göstermiştir. Kopya oranının 0.40 olduğu koşullarda l_z^* istatistiği ile iki aşamalı kullanılan ω ve GBT indekslerinin belirleme gücünün en yüksek olduğu elde edilmiştir.

Çeşitli faktörlerin testin ortalama güçlüğü, kopyacının yetenek düzeyi ve kopya oranının; kopya ve benzerlik indeksleri, birey uyum istatistikleri ve KL indeksinin iki aşamalı olarak kullanıldığı durumda, kopya ve benzerlik indekslerinin I. tip hata oranına olan ortak etkisi. Analizde kullanılan testin ortalama güçlüğü, kopyacının yetenek düzeyi ve kopya oranı faktörleri altında ω ve GBT indekslerinin gücüne olan ortak etkilerine yönelik bulgular elde edilmiştir. Şekil XIV’de değişimlenen faktörlerin indekslerin gücüne olan ortak etkilerine yönelik elde edilen bulgular sunulmuştur.



Şekil XIV. Değişimlenen Faktörlerin Güç Oranına Olan Ortak Etkisinin Grafiği. Farkların daha görülebilir olması için (0.905,1.02) arasında ölçeklendirme yapılmıştır.

Şekil XIV' e göre tüm şartlar altında *KL* indeksi ile iki aşamalı kullanılan ω ve *GBT* indekslerinin test hilesi yapan bireyleri belirleme gücü oranının yüksek olduğu gözlenmiştir. Çok düşük yetenek düzeyi ve zor testte kopya oranı 0.60 iken l_z^* ve l_z istatistikleri ile aşamalı kullanılan ω ve *GBT* indekslerinin belirleme gücü oranı *KL* indeksi ile iki aşamalı kullanılan ω ve *GBT* indekslerinin belirleme gücü oranına göre düşük olduğu görülmüştür. Belov ve Armstrong (2010), *KL* indeksi ile *K* indeksini iki aşamalı kullanarak *K* indeksinin test hilesi yapan bireyi belirleme gücünde tutarlı bir biçimde yüksek performans gösterdiğini ifade etmişlerdir. Bu çalışmada da elde edilen bulgular ışığında l_z ve l_z^* istatistikleri ile iki aşamalı kullanılan ω ve *GBT* indekslerin kopya belirleme güçleri performanslarının arttığı ve *KL* indeksi ile iki aşamalı kullanıldığında ω ve *GBT* indekslerin kopya belirleme güçleri genel olarak en yüksek performansı gösterdiği sonucuna ulaşılmıştır.

Ayrıca bu çalışmada değişkenlerin temel ve ortak etki bulguları ışığında ω ve *GBT* indekslerinin aşamasız kullanıldığında elde edilen I. tip hata oranı değerleri, iki aşamalı olarak kullanıldığında elde edilen I. tip hata oranı değerlerine göre yüksek elde edilmiştir (Bu konuda daha ayrıntılı bilgi için bkz. EK7). ω ve *GBT* indekslerinin aşamasız ve iki aşamalı kullanıldığında genel olarak ω indeksinin güç performansları iki aşamalı kullanılan ω ve *GBT* indeksinin güç oranına ve aşamsız kullanılan *GBT* indeksinin güç oranına göre düşük değerler aldığı görülmüştür (Bu konuda daha ayrıntılı bilgi için bkz. EK8). Sonuç olarak çalışmada kullanılan kopya ve benzerlik indekslerinin iki aşamalı

analizleri ile elde edilen I. tip hata ve güç oranları performansları aşamasız kullanımları ile elde edilen performanslarına göre daha iyi sonuç vermiştir.



BÖLÜM 5

SONUÇLAR VE ÖNERİLER

Bu bölümde çalışmada elde edilen bulgulara dayalı olarak ulaşılan sonuçlara ve önerilere yer verilmiştir.

Sonuçlar

Bu çalışma, bilgisayar temelli ve kağıt-kalem testi uygulamalarında, sınav maddelerine ilişkin madde ön bilgisine sahip olmayan ve kopya çekme davranışı gösteren bireyleri tespit etmek için aşamasız ve iki aşamalı olarak kullanılan kopya ve benzerlik indeksleri, birey uyum istatistikleri ve *KL* indeksinin testin ortalama güçlüğü, kopyacının yetenek düzeyi ve kopya oranı değişkenleri altında I. tip hata oranları ve kopya belirleme güçlerinin nasıl etkilendiği incelenmiştir. Aşağıda çalışmanın nihayetinde elde edilen sonuçlar aşağıda özetlenmiştir.

1. Müdahalenin/manipülasyonun yapılmadığı aşamasız analizlerde çok düşük yetenek düzeyi koşulunda *KL* indeksinin I. tip hata oranı ortalaması ω ve *GBT*, l_z ve l_z^* istatistiklerine göre yüksek çıkmıştır. Dolayısıyla I. tip hata oranları ortalamalarına göre test hilesi yapmayan bireyleri de test hilesi yapan birey olarak yanlış sınıflandırdığı gözlenmiştir. Düşük yetenek düzeyinde ise *KL* indeksi ile birlikte l_z ve l_z^* istatistiklerinin I. tip hata oranları ortalamasının da bireyleri yanlış pozitif (test hilesi yapmayan bireyleri de test hilesi yapan birey olarak yanlış sınıflandırdığı) olarak belirlediği görülmüştür.
2. Kolay testlerdeki müdahalenin/manipülasyonun yapılmadığı aşamasız analizlerin sonuçlarına göre *KL*, l_z ve l_z^* istatistiklerinin I. tip hata oranı ortalamasının test hilesi yapmayan bireyleri test hilesi yapmış şeklinde sınıflandırma oranının yüksek olduğu görülmüştür. Zor testte ise *KL* indeksinin I. tip hata oranı ortalamasının yüksek olduğu gözlenmiştir.
3. Testin ortalama güçlük ve yetenek düzeyi değişkenlerinin çaprazlandığı tüm koşullarda müdahalenin/manipülasyonun yapılmadığı aşamasız analizlerin

- sonuçlarına göre *KL* indeksinin I. tip hata oranı ortalamasının yüksek olduğu fakat *GBT* indeksinin I. Tip hata oranı ortalamasının düşük olduğu sonucuna ulaşılmıştır.
4. Müdahalenin/manipülasyonun yapıldığı aşamasız analizlerde tüm yetenek düzeylerinde en yüksek güç performansını *GBT* indeksi göstermiştir. ω , l_z ve l_z^* istatistiklerinin de güç performansı birbirine benzer ve yüksek olarak elde edilmiştir.
 5. Ortalama test güçlüğü düzeylerinin hepsinde manipülasyonun yapıldığı aşamasız analiz sonuçlarına göre, tutarlı olarak en yüksek güce *GBT* indeksinin sahip olduğu görülmüştür. Yetenek düzeyi faktöründe olduğu gibi, benzer elde edilen bir diğer sonuç ise ω , l_z ve l_z^* istatistiklerinin güç ortalamalarının yüksek olmasıdır.
 6. Manipülasyonun yapıldığı aşamasız analizlerde diğer değişimlenen faktörlere benzer olarak *GBT* indeksinin gücü tutarlı olarak tüm kopya oranı düzeylerinde en yüksek performansı göstermiştir. En düşük performans ise *KL* indeksinde gözlenmiştir.
 7. Testin ortalama güçlüğü, yetenek düzeyi ve kopya oranı değişkenlerinin çaprazlandığı tüm koşullarda manipülasyonun yapıldığı aşamasız verilerin analiz sonuçlarına göre, tutarlı olarak en yüksek güç performansı *GBT* indeksine ait iken en düşük güç performansı *KL* indeksinde gözlenmiştir.
 8. Manipülasyonun yapılmadığı indekslerin iki aşamalı kullanıldığı verilerde tüm yetenek düzeylerinde en düşük I. tip hata oranı *KL* indeksi ile kullanılan *GBT* indeksinde görülürken, en yüksek I. tip hata değeri l_z indeksi ile kullanılan ω indeksinde gözlenmiştir.
 9. Zor ve kolay testin kullanıldığı manipülasyonun yapılmadığı iki aşamalı analizlerde en düşük yanlış kararın verilmesini sağlayan indeksin *GBT* indeksi olduğu görülmüştür. Bununla birlikte kolay testte birey uyum indeksleri ile aşamalı kullanılan ω indeksi dışında birey uyum indeksleri ile iki aşamalı kullanılan *GBT* indeksinin I. tip hata oranının yüksek elde edildiği gözlenmiştir.
 10. Manipülasyonun yapılmadığı ve analizlerin iki aşamalı gerçekleştirildiği verilerde zor testte ve tüm yetenek düzeylerinde yanlış kararın en az verilmesini sağlayan indeksin *KL* indeksi ile birlikte kullanılan *GBT* ve ω indeksi olurken, tüm şartlar altında ise en az yanlış kararın verilmesini sağlayan *KL* indeksi ile iki aşamalı olan *GBT* indeksine ait olduğu sonucuna ulaşılmıştır. Bununla birlikte en yüksek I. tip hata oranına sahip indeksinin ise l_z ile aşamalı kullanılan ω 'ya ait olduğu görülmüştür.
 11. İki aşamalı analizin yapıldığı verilerde, yetenek faktörünün tüm düzeylerinde test hilesi yapan bireyi tespit ederek doğru karar alınmasını sağlayan indeksin *KL* ile aşamalı kullanılan ω ve *GBT* indekslerine ait olduğu görülmüştür.

12. Testin ortalama güçlük düzeylerinin tüm seviyelerinde test hilesi yapan bireyi belirleme gücü tutarlı olarak en yüksek olan indeks *KL* ile aşamalı kullanılan ω ve *GBT* indeksine ait olarak elde edilmiştir.
13. Yetenek ve testin ortalama güçlüğü değişkenlerinin temel etkileri sonucuna benzer olarak, kopya oranının tüm düzeylerinde en yüksek güce sahip olan indeks, *KL* ile aşamalı kullanılan *GBT* ve ω indekslerinin olduğu görülmektedir.
14. Testin ortalama güçlüğü, yetenek düzeyi ve kopya oranı değişkenlerinin birlikte olan etkisinde tüm koşullarda *KL* indeksi ile aşamalı kullanılan *GBT*, *KL* indeksi ile aşamalı kullanılan ω indeksinin en yüksek test hilesi yapan bireyi belirleme performansına sahip olduğu görülmüştür.

Öneriler

Bu bölümde çalışma sonucunda elde edilen bulgulara dayalı olarak, uygulayıcılara ve araştırmacılara yönelik olarak önerilere yer verilmiştir.

Uygulayıcılara Öneriler

1. Ulusal çaptaki kurum ya da kuruluşların gerçekleştirdiği geniş ölçekli sınavlarda, sınav yönetim sürecinin ve test güvenliğinin bir parçası olarak, test hilesinde bulunan bireyleri belirlemeye yönelik yapılacak çalışmalar için ω ve *GBT* benzerlik indekslerinin yanlış pozitif oranlara düşmesini sağlayan iki aşamalı analizlerde, *KL* kullanılması önerilmektedir.
2. Sınıf içi ya da okul genelinde yapılan ortak sınavlarda, bireylerin ölçülen özellikle ilgili daha önce test hilesinin karışmadığı verileri kullanılarak *KL* indeksinin önce ω ve *GBT* benzerlik indekslerinin ise sonra uygulandığı analizler ile belirlenen şüpheli bireylere yönelik kanıtlar elde edilebilir.
3. Ortalama güçlüğü yüksek testlerin kullanıldığı sınavlarda, cevap kopyalamadan kaynaklı test hilesi yapan bireyleri belirlemek için iki aşamalı analizlerin kullanılması önerilmektedir.
4. Çalışmadaki cevap kopyalamadan kaynaklı test hilesinde bulunan bireyleri belirlemeye yönelik aşamasız ve iki aşamalı analizler 150 kişilik sınav salonları ve 30 maddeden oluşan testler baz alınarak gerçekleştirilmiştir. Geniş ölçekli sınavlarda

tüm verilerin KL indeksi değerleri elde edilerek, olası kopyacı olarak belirlenen bireylerin yer aldıkları sınav salonundaki tüm potansiyel kaynaklar kullanılarak ω ve *GBT* analizinin gerçekleştirilmesi önerilmektedir. Böylece sınav salonunda yer alan bireylerin tümüyle gerçekleştirilecek olan analizden daha kısa sürede analiz tamamlanmış olacaktır.

Araştırmacılara Öneriler

1. Bu çalışmada kopya oranı, olası kopyacının yetenek düzeyi, testin ortalama güçlüğü değişkenleri altında ω ve *GBT* indekslerinin aşamasız ve iki aşamalı olarak I. tip hata ve güçleri elde edilmiştir. İki aşamalı analizlerde kullanılan kopya ve benzerlik (ω ve *GBT*) indekslerinin güçlerinin yüksek, I. tip hataları düşük değerler almıştır. Bu indeks ve istatistiklerde aranan özelliklerdir. Dolayısıyla farklı kopya ve benzerlik indekslerinin de iki aşamalı analiz sonucunda hile yapanı belirleme performanslarına yönelik bilgiler sunmasına yönelik başka simülasyon çalışmaları gerçekleştirilebilir.
2. İndekslerin test hilesini belirleme performanslarını etkileyen örneklem büyüklüğü, test uzunluğu, test maddelerinin yapısı, testin boyutluluğu gibi birçok faktör mevcuttur. Bu çalışmada örneklem büyüklüğü ve test uzunluğu faktörlerinin düzeyleri ve olası kopyacı oranı sabitlenmiştir. Değişimlenen şartlar altında kullanılan ω ve *GBT* indekslerinin test hilesi yapan bireyi belirleme performanslarının daha iyi sonuçlar ürettiği gözlenmiştir. İki aşamalı analizlerin kullanılacağı çalışmalarda farklı örneklem büyüklükleri, test uzunlukları, kaynağın yetenek düzeyi, kopyacının yetenek düzeyi gibi değişkenler ve bu değişkenlerin düzeyleri farklılaştırılarak başka simülasyon çalışmaları yapılabilir.
3. Simülasyon çalışması olarak gerçekleştirilen bu çalışma gerçek veriler kullanılarak farklı alfa düzeylerinde tekrarlanabilir.
4. Bu çalışmada parametrik bireye uyum istatistikleri olan l_z ve l_z^* istatistiklerine yer verilmiştir. Çalışmada kullanılan istatistiklerin test hilesi yapan bireyi belirlemedeki performansları ω ve *GBT* indekslerinin performanslarına benzer sonuçlar genel olarak üretmemiştir. Parametrik olmayan birey uyum istatistiklerine de yer verilerek bu çalışmaya benzer simülatif çalışmalar düzenlenerek, analizlerin aşamasız ve iki aşamalı kullanıldıktan sonra performanslarına yönelik yeni bulguların üretileceği çalışmalar düzenlenebilir.

5. Bu çalışmada test hilesini belirlemeye yönelik kanıtın arttırılması için iki aşamalı analiz gerçekleştirilmiştir. İki aşamalı analiz sonucunda ω ve *GBT* indekslerinin test hilesi yapan bireyleri belirleme performanslarının arttığı görülmüştür. Test hilesi yapan bireyi belirlemeye yönelik kanıtın arttırılması için bilgisayar temelli testlerde bireylerin maddelere verdikleri tepki sürelerini de göz önünde bulundurarak iki aşamalı bir simülasyon çalışması ya da gerçek verilere dayalı bir araştırma planlanabilir.
6. İki aşamalı olarak gerçekleştirilen analizde cevap kopyalamadan kaynaklı test hilesini belirlemede kullanılan kopya ve benzerlik (ω ve *GBT*) indekslerinin I. tip hatası düşük, gücü ise yüksek elde edilmiştir. Benzer koşullar altında çalışmada kullanılan kopya ve benzerlik indekslerinin F ölçütü ve doğruluk ölçütlerine bakılarak indekslerin performanslarına yönelik yeni bilgiler sunulabilir.
7. Örneklem büyüklüğü ve test uzunluğu çalışmanın sınırlılığını oluşturmuştur. Çeşitli örneklem büyüklüklerinde, değişen test uzunluklarında, kopyacı sayılarının da değiştiği farklı koşullar altında, cevap kopyalamadan kaynaklı test hilesini belirlemek için iki aşamalı yapılan analiz yapılarak kopya ve benzerlik (ω ve *GBT*) indekslerinin I. tip ve II. tip hatalarının karşılaştırıldığı bir simülasyon çalışması planlanabilir.
8. Çalışmada *KL* indeksi değerlerinin hesaplanabilmesi için R programında yer alan bir paketi kullanılmıştır. Bu durum çalışmanın sınırlıklarından birini oluşturmuştur. Çalışmada değişimlenen şartlar altında, yetenek parametrelerinin kestirileceği farklı programlar ya da paket programları kullanılarak *KL* indeksinin cevap kopyalamadan kaynaklı test hilesini belirlemedeki gücü ve birinci tip hata oranına yönelik bilgilerinin karşılaştırılacağı bir çalışma planlanabilir.

KAYNAKLAR

- Angoff, W. H. (1972). *The development of statistical indices for detecting cheaters*. Berkeley, CA: Educational Testing Service. Retrieved from <https://www.scribd.com/document/242510828/ANGOFF-1974-The-development-of-statistical-indices-for-detecting-cheaters-pdf>
- Armstrong, R., & Shi, M. (2009). Model-free CUSUM methods for person fit. *Journal of Educational Measurement* 46, 408-428. doi: 10.1111/j.1745-3984.2009.00090.x
- Armstrong R. D., Stoumbos, Z. G., Kung, M. T., & Shi, M. (2007). On the performance of the lz person-fit statistic *Practical Assessment Research & Evaluation* 12. Retrieved from <http://pareonline.net/getvn.asp?v=12&n=16>
- Atılgan, H., Kan, A. ve Aydın, B. (2014). *Eğitimde Ölçme ve Değerlendirme*. Ankara: Anı Yayıncılık.
- Balta, E., Uçar, A. & Sahin, A.(June,2019).*Detection of cheating behavior in online unproctored CATs via a validation test*. Paper presented at International Association for Computerized Adaptive Testing (IACAT) Conference, Minneapolis, MN, USA. Abstract retrieved from <http://iacat.org/sites/default/files/IACAT-Program-7June.pdf>
- Baykul, Y. (2010). *Eğitimde ve Psikolojide Ölçme: Klasik Test Teorisi ve Uygulaması*. Ankara: Pegem Yayınları.
- Belov, D. I., & Armstrong, R. D. (2010). Automatic detection of answer copying via Kullback–Leibler divergence and K-index. *Applied Psychological Measurement* 34, 379–392. doi:10.1177/0146621610370453
- Belov, D. I. (2011). Detection of answer copying based on the structure of a high-stakes test. *Applied Psychological Measurement* 35, 495–517. Retrieved from <https://eric.ed.gov/?id=EJ942262>
- Belov, D. I. (2013). Detection of test collusion via Kullback–Leibler divergence. *Journal of Educational Measurement*, 50, 141–163. doi: 10.1111/jedm.12008
- Belov, D. (2014a). *Detection of aberrant answer changes via Kullback–Leibler divergence*. (Lsac Research Report Series14-04) Law School Admission Council Research Report. Retrieved from <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.732.6331&rep=rep1&type=pdf>
- Belov, D. I. (2014b). Detecting item preknowledge in computerized adaptive testing using information theory and combinatorial optimization. *Journal of Computerized Adaptive Testing* 2, 37–58. doi: 10.7333/1410-0203037

- Belov, D. (2015). Robust detection of examinees with aberrant answer changes. *Journal of Educational Measurement* 52(4), 437–456. doi:[10.1111/jedm.12094](https://doi.org/10.1111/jedm.12094)
- Belov, D. (2016). Comparing the performance of eight item preknowledge detection statistics. *Applied Psychological Measurement*, 40 (2), 83-9. doi: 10.1177/0146621615603327
- Belov, D. I., Pashley, P. J., Lewis, C., & Armstrong, R. D. (2007). Detecting aberrant responses with Kullback–Leibler distance. In K. Shigemasu, A. Okada, T. Imaizumi and T. Hoshino (Eds.), *New trends in psychometrics* (pp. 7-14). Tokyo:Universal Academy Press.
- Bird, C. (1927). The detection of cheating in objective examinations. *School and society*, 25, 261-262. Retrieved from <https://psycnet.apa.org/record/1927-01217-001>
- Bird, C. (1929). An improved method of detection cheating in objective examinations. *Journal of Educational Research*, 25, 261-262.
- Brussow, J. A., Skorupski, W.P. & Thompson, W. J.(2018). *A hierarchical IRT model for identifying group-level aberrant growth*. Paper presented at the National Conference on Measurement in Education annual meeting. New York. Abstract retrieved from <https://www.wjakethompson.com/publication/2018-ncme-aberrant-growth/>
- Chalmers, P. (2019). Multidimensional item response theory (mirt) [Computer software manual]. Retrieved from <https://cran.r-project.org/web/packages/mirt/mirt.pdf>
- Chang, H.-H., & Ying, Z. (1996). A global information approach to computerized adaptive testing. *Applied Psychological Measurement*, 20, 213–229. Retrieved from <https://journals.sagepub.com/doi/10.1177/014662169602000303>
- Cizek, G. & Wollack, J. (2017). Exploring cheating on tests – the context, the concern, and the challenges. In G. Cizek & J. Wollack (Eds.), *Handbook of Quantitative Methods for Detecting Cheating on Tests* (pp 3-19). Routledge: New York.
- Clark, J. M. (2010). *Aberrant response patterns as a multidimensional phenomenon: using factor-analytic model comparison to detect cheating* (Unpublished doctoral dissertation). University of Kansas, Lawrence, KS. Retrieved from ProQuest Dissertations and Theses database. (UMI No: 3423807)
- Conijn, J. M. (2013). *Detecting and explaining person misfit in non-cognitive measurement*. (Unpublished doctoral dissertation). Tilburg University, Netherlands, Holland.
- Cover, T. M., & Thomas, J. A. (1991). *Elements of information theory*. New York, NY: John Wiley.
- Çiçek, E.K. (2020). *Test Tahriyatının Belirlenmesinde Silme Belirleme İndeksinin I. Tip Hata ve İstatistiksel Güç Açısından İncelenmesi*.(Yayımlanmamış Doktora Tezi), Ankara Üniversitesi, Ankara, Türkiye.

- Drasgow, F., Levine, M. V., & Williams, E. A. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology*, 38(1), 67–86. Retrieved from <https://doi.org/10.1111/j.2044-8317.1985.tb00817.x>
- Drasgow, F., Levine, M.V., & McLaughlin, M. E. (1987). Detecting inappropriate test scores with optimal and practical appropriateness indices. *Applied Psychological Measurement*, 11, 59-79. Retrieved from <https://doi.org/10.1177/014662168701100105>
- Eggen, T. J. H. M. (1999). Item selection in adaptive testing with the sequential probability ratio test. *Applied Psychological Measurement*, 23, 249–261. Retrieved from <https://doi.org/10.1177/01466219922031365>
- Hanson, B. A., Harris, D. J., & Brennan, R. L. (1987). *A comparison of several statistical methods for examining allegations of copying*. (ACT Research Report No. 87–15). Iowa City, IA: American College Testing. Retrieved from http://www.act.org/content/dam/act/unsecured/documents/ACT_RR87-15.pdf
- Hauser, C., Kingsbury, G. G., & Houser, R. L. (2011, April). *Individual score validity: using the wariness index to identify test performance to treat with caution*. Paper presented at the Annual Meeting of the National Council on Measurement in Education. Abstract retrieved from <https://www.nwea.org/content/uploads/2014/08/Individual-Score-Validity-Using-e-Wariness-Index-to-Identify-Test.pdf>
- He, Q., Meadows, M., & Black, B. (2018). *Statistical techniques for studying anomaly in test results: a review of literature*. London: Office of Qualifications and Examinations Regulation. Retrieved from https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/690007/Statistical_techniques_for_studying_anomaly_in_test_result_s-_a_review_of_literature.pdf
- Man, K., Harring, J. R, Ouyang, Y., & Thomas, S. L. (2018): Response time based nonparametric Kullback-Leibler divergence measure for detecting aberrant test-taking behavior, *International Journal of Testing*. doi:10.1080/15305058.2018.1429446
- Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education*, 16(4), 277–298. Retrieved from https://www.tandfonline.com/doi/abs/10.1207/S15324818AME1604_2
- Krimpen-Stoop, E. M. L. A., & Meijer, R. R. (2001). CUSUM-based person-fit statistics for adaptive testing. *Journal of Educational and Behavioral Statistics*, 26(2), 199–217. Retrieved from https://ris.utwente.nl/ws/portalfiles/portal/5156759/RR_99-05.pdf

- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22, 79-86. Retrieved from <http://mathfaculty.fullerton.edu/sbehseta/Kullback.pdf>
- Levine, M. V., & Rubin, D. B. (1979). Measuring the appropriateness of multiple choice test scores. *Journal of Educational Statistics* 4, 269–290. doi: 10.2307/1164595
- Li, M. F., & Olejnik, S. (1997). The power of Rasch person-fit statistics in detecting unusual response patterns. *Applied Psychological Measurement* 21, 215-231. Retrieved from <https://doi.org/10.1177/01466216970213002>
- Magis, D., Raîche, G., & Béland, S. (2012). A didactic presentation of Snijders's $l(z)^*$ index of person fit with emphasis on response model selection and ability estimation. *Journal of Educational and Behavioral Statistics*, 37, 57–81. Retrieved from <https://doi.org/10.3102/1076998610396894>
- Man, K., Harring, J., Quayang, Y., & Thomas, S. (2018). Response time based nonparametric Kullback-Leibler divergence measure for detecting aberrant test-taking behavior. *International Journal of Testing*, 18(2), 155-177. doi:10.1080/15305058.2018.1429446
- Meijer, R. R. (1994). The number of Guttman errors as a simple and powerful person-fit statistic. *Applied Psychological Measurement*, 18(4), 311–314. doi:10.1177/014662169401800402
- Meijer, R. R. (1996). Person-Fit research: An introduction. *Applied Measurement in Education*, 9(1), 3-8. doi:10.1207/s15324818ame0901_2
- Meijer, R. R., & Sijtsma, K. (2001). Methodology review: evaluating person fit. *Applied Psychological Measurement* 25, 107-135. Retrieved from <https://doi.org/10.1177/01466210122031957>
- Meijer, R. R., & Sotaridona, L. S. (2006). *Detection of advance item knowledge using response times in computer adaptive testing* (Vol. 3, No. 3). Law School Admission Council. Retrieved from https://ris.utwente.nl/ws/portalfiles/portal/5129730/LSAC_CT-03-03.pdf
- Meijer, R. R., & Tendeiro, J. N. (2014). *The use of person-fit scores in high stakes educational testing: How to use them and what they tell us* (LSAC Research Report 14-03). Retrieved from <http://www.lsac.org/lisacresources/research/all/rr>
- Molenaar, I. W., & Hoijtink, H. (1990). The many null distributions of person fit indices. *Psychometrika*, 55, 75–106. doi:10.1007/bf02294745
- Nering, M. L. (1995). The distribution of person fit using true and estimated person parameters. *Applied Psychological Measurement*, 19, 121-129. Retrieved from <https://doi.org/10.1177/014662169501900201>

- Nering, M. L., & Meijer, R. R. (1998). A comparison of the person response function and the I_z person-fit statistic. *Applied Psychological Measurement*, 22, 1, 53-69. Retrieved from <https://doi.org/10.1177/01466216980221004>
- Pardo, L. (2006). *Statistical inference based on divergence measures*. Chapman & Hall/CRC: New York.
- Raton-Lopez, M., Rodriguez-Alvarez, X. M., Suarez- Cadarso, C., & Sampedro-Gude, F. (2014). Package 'OptimalCutpoints'. Retrieved from <https://cran.r-project.org/web/packages/OptimalCutpoints/OptimalCutpoints.pdf>
- Reise, S. P. (1995). Scoring method and the detection of person misfit in a personality assessment context. *Applied Psychological Measurement* 19, 213-229. Retrieved from <https://doi.org/10.1177/014662169501900301>
- Reise, S. P., & Due, A. M. (1991). The influence of test characteristics on the detection of aberrant response patterns. *Applied Psychological Measurement*, 15, 217-226. Retrieved from <https://doi.org/10.1177/014662169101500301>
- Partchev, I. (2016). A collection of functions related to item response theory (irtoys) [Computer software manual]. <https://cran.rproject.org/web/packages/irtoys/irtoys.pdf> adresinden alınmıştır.
- Schnipke, D. L., & Scrams, D. J. (1999). *Exploring issues of test taker behavior: insights gained from response-time analyses*. Law School Admission Council Computerized Testing Report. LSAC Research Report Series. Retrieved from https://archive.org/details/ERIC_ED467811/page/n5
- Seol, J., & Rubright, J.D. (2013). *The impact of test characteristics on Kullback-Leibler divergence index to identify examinees with aberrant responses*. Paper presented at the 2nd Annual Conference on Statistical Detection of Potential Test Fraud. Madison, WI. Abstract retrieved from <https://conferenceontestsecurity.org/cots-2013>
- Shu, Z. (2010). *Detecting test cheating using a deterministic, gated item response theory model*. (Unpublished doctoral dissertation). The University of North Carolina at Greensboro, Greensboro. (UMI No. 3434164).
- Sideridis, G.D., & Zopluoğlu, C. (2018). Validation of response similarity analysis for the detection of academic cheating: An experimental study. *Journal of Applied Measurement*, 19(1), 59-75. Retrieved from https://www.researchgate.net/publication/321342650_Validation_of_Response_Siilarity_Analysis_for_the_Detection_of_Academic_Cheating_An_Experimental_Study
- Singmann, H. (2020). Complete environment for Bayesian inference (LaplaceDemon) [Computer software manual]. Retrieved from <https://cran.r-project.org/web/packages/LaplacesDemon/LaplacesDemon.pdf>

- Sotaridona, L. S., & Meijer, R. R. (2002). Statistical properties of the K-index for detecting answer copying in a multiple-choice test. *Journal of Educational Measurement* 39, 115–132. doi: 10.1111/j.1745-3984.2002.tb01138.x
- Sotaridona, L. S., & Meijer, R. R. (2003). Two new statistics to detect answer copying. *Journal of Educational Measurement* 40, 53–70.
- Sotaridona, L.S., van der Linden, W.J., & Meijer, R.R. (2006). Detecting answer copying using the kappa statistic. *Applied Psychological Measurement*, 30, 412-431. Retrieved from <https://www.utwente.nl/nl/bms/omd/Medewerkers/artikelen/APM%202006%2C%20412-431.pdf>
- Steinkamp, S. (2017). Identifying aberrant responding: Use of multiple measures (Unpublished doctoral dissertation). University of Minnesota at Minneapolis, Minneapolis. Retrieved from https://conservancy.umn.edu/bitstream/handle/11299/188885/Steinkamp_umn_0130E_18212.pdf?sequence=1&isAllowed=y
- St-Onge C., Valois P., Abdous B., & Germain S. (2011). Accuracy of person-fit statistics: a monte carlo study of the influence of aberrance rates. *Applied Psychological Measurement*, 35(6), 419–432. doi:10.1177/0146621610391777
- Sunbul O., & Yormaz, S. (2018a). Effects of test level discrimination and difficulty on answer-copying indices. *International Journal of Evaluation and Research in Education*, 7(1), 32-38. doi: 10.11591/ijere.v7.i1.11488
- Sunbul, O., & Yormaz, S. (2018b). Investigating the performance of omega index according to item parameters and ability levels. *Eurasian Journal of Educational Research*, 74, 207-226. doi: 10.14689/ejer.2018.74.11
- Şahin, A. (2012). *Madde Tepki Kuramı'nda Test Uzunluğu Ve Örneklem Büyüklüğünün Model Veri Uyumu, Madde Parametreleri Ve Standart Hata Değerlerine Etkisinin İncelenmesi*. (Yayımlanmamış Doktora Tezi). Ankara Üniversitesi, Ankara, Türkiye.
- Thiessen, B. (2008). *Relationship between test security policies and test score manipulations*. (Unpublished doctoral thesis), University of Iowa, Iowa. Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3347249).
- Tendeiro, J. N., Meijer, R.R., & Niessen, A. S. (2016). PerFit: an R package for person fit in IRT. *Journal of Statistical Software* 74, issue 5. Retrieved from <https://www.jstatsoft.org/article/view/v074i05>
- Tendeiro, J. N. (2018). Person fit (PerFit) [Computer software manual]. Retrieved from <https://cran.r-project.org/web/packages/PerFit/PerFit.pdf>
- Tendeiro, N.J., & Meijer, R.R. (2012). A CUSUM to detect person misfit: a discussion and some alternatives for existing procedures, *Applied Psychological Measurement* 36(5), 420–442. doi:10.1177/0146621612446305
- Tendeiro, N.J., & Meijer, R.R. (2014). Detection of invalid test scores: the usefulness

- of simple nonparametric statistics. *Journal of Educational Measurement*, 51(3), 239–259. Retrieved from <https://www.semanticscholar.org/paper/Detection-of-Invalid-Test-Scores%3A-The-Usefulness-of-TendeiroMeijer/3c46895d2bb6fb763af4b78ad00c532eaa1fba26>
- Turgut, M., ve Baykul, Y. (2010). *Eğitimde Ölçme ve Değerlendirme*. Ankara:Pegem Yayınevi.
- van der Linden, W. J. (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics*, 31(2), 181-204. doi:10.2307/3701364
- van der Linden, W. J. (2007). A hierarchical framework for modeling speed and accuracy on test items. *Psychometrika*, 72, 287–308. doi:10.1007/s11336-006-1478-z
- van der Linden, W. J., & Guo, F. M. (2008). Bayesian procedures for identifying aberrant response time patterns in adaptive testing. *Psychometrika*, 73(3), 365–384. doi:[10.1007/s11336-007-9046-8](https://doi.org/10.1007/s11336-007-9046-8)
- van der Linden, W. J. (2009). A bivariate log-normal response-time model for the detection of collusion between test takers. *Journal of Educational and Behavioral Statistics*, 34, 378–394. Retrieved from <https://journals.sagepub.com/doi/abs/10.3102/1076998609332107?journalCode=jeb>
- van der Ark, L. A., Emons, W. H. M., & Sijtsma, K. (2008). Detecting answer copying using alternate test forms and seat locations in small-scale examinations. *Journal of Educational Measurement*, 45, 99-117. doi: 10.1111/j.1745-3984.2007.00055.x
- van der Linden, W. J., & van Krimpen-Stoop, E. M. (2003). Using response times to detect aberrant responses in computerized adaptive testing. *Psychometrika*, 68(2), 251-265. Retrieved from <https://link.springer.com/article/10.1007/BF02294800>
- van Krimpen-Stoop, E., & Meijer, R. (1999). The null distribution of person-fit statistics for conventional and adaptive tests. *Applied Psychological Measurement*, 23, 327–344. Retrieved from <https://journals.sagepub.com/doi/10.1177/01466219922031446>
- Voncken, L. (2014). *Comparison of the lz* person-fit index and ω copying-index in copying detection*. (First Year Paper). Universiteit van Tilburg. Retrieved from <http://arno.uvt.nl/show.cgi?fid=135361>
- Wang, C., & Xu, G. (2015). A mixture hierarchical model for response times and response accuracy. *British Journal of Mathematical and Statistical Psychology*, 68(3), 456-477. doi:[10.1111/bmsp.12054](https://doi.org/10.1111/bmsp.12054). Retrieved from <https://onlinelibrary.wiley.com/doi/abs/10.1111/bmsp.12054>

- Wise, S. L., & DeMars, C. E. (2006). An application of item response time: the effort-moderated IRT model. *Journal of Educational Measurement*, 43(1), 19-38. doi: 10.1111/j.1745-3984.2006.00002.x. Retrieved from https://www.researchgate.net/publication/229985871_An_Application_of_Item_Response_Time_The_Effort-Moderated_IRT_Model
- Wise, S. L., Ma, L., Kingsbury, G. G., & Hauser, C. (2010). An investigation of the relationship between time of testing and test-taking effort. *Northwest Evaluation Association*. Retrieved from <https://www.nwea.org/content/uploads/2010/04/An-Investigation-of-the-Relationship-between-Time-of-Testing-and-Test-Taking-Effort.pdf>
- Wise, S. L., & Kong, X. (2005). Response time effort: a new measure of examinee motivation in computer-based tests. *Applied Measurement in Education*, 18(2), 163-183. Retrieved from <https://files.eric.ed.gov/fulltext/ED490203.pdf>
- Voncken, L. (2014). Comparison of the Iz^* person-fit index and ω copying-index in copying detection. (First Year Paper). Universiteit van Tilburg. Retrieved from <http://arno.uvt.nl/show.cgi?fid=135361>
- Wollack, J. A. (1997). A nominal response model approach for detecting answer copying. *Applied Psychological Measurement*, 21(4), 307–320. doi:10.1177/01466216970214002. Retrieved from <https://journals.sagepub.com/doi/10.1177/01466216970214002>
- Wollack, J. A. (2003). Comparison of answer copying indices with real data. *Journal of Educational Measurement*, 40(3), 189–205. doi: 10.1111/j.1745-3984.2003.tb01104.x
- Wollack, J. A. (2006). Simultaneous use of multiple answer copying indexes to improve detection rates. *Applied Measurement in Education*, 19(4), 265–288. Retrieved from https://www.tandfonline.com/doi/abs/10.1207/s15324818ame1904_3
- Wollack, J. A., & Cohen, A. S. (1998). Detection of answer copying with unknown item and trait parameters. *Applied Psychological Measurement*, 22(2), 144–152. doi: 10.1177/01466216980222004. Retrieved from <https://journals.sagepub.com/doi/abs/10.1177/01466216980222004>
- Wollack, J. A., & Maynes, D. (2011, April). *Detection of test collusion using item response data*. Paper presented at the Annual Meeting of the National Council on Measurement in Education. New Orleans, LA.
- Wollack, J. A., & Maynes, D. D. (2017). Detection of test collusion using cluster analysis. In G. Cizek & J. Wollack (Eds.), *Handbook of quantitative methods for detecting cheating on tests* (pp. 124-150). New York:Routledge
- Yormaz, S., & Sünbül, Ö. (2017). Determination of type I error rates and power of answer copying indices under various conditions. *Educational Sciences: Theory & Practice*, 17(1), 5-26. doi: 10.12738/estp.2017.1.0105

- Yormaz, S. (2019). *Test Güvenliđi Açısından Bireyler Arasındaki Olası İş Birliđinin İncelenmesi*. (Yayımlanmamış Doktora Tezi). Mersin Üniversitesi, Mersin, Türkiye.
- Zopluođlu, C., & Davenport, E. C., Jr. (2012). The empirical power and type I error rates of the GBT and ω indices in detecting answer copying on multiple-choice tests. *Educational and Psychological Measurement*, 72(6), 975–1000. Retrieved from https://www.researchgate.net/publication/235247021_The_Empirical_Power_and_Type_I_Error_Rates_of_the_GBT_and_Indices_in_Detecting_Answer_Copying_on_Multiple-Choice_Tests
- Zopluođlu, C. (2016). Classification performance of answer-copying indices under different types of IRT models. *Applied Psychological Measurement* 40, 592–607. Retrieved from <https://doi.org/10.1177/0146621616664724>
- Zopluođlu, C. (2017). Similarity, answer copying, and aberrance: Understanding the status quo. In G. Cizek & J. Wollack (Eds.), *Handbook of Detecting Cheating on Tests* (pp. 44-72). New York:Routledge.
- Zopluođlu, C. (2018). Computing response similarity indices for multiple-choice tests (CopyDetect) [Computer software manual]. Retrieved from <https://cran.r-project.org/web/packages/CopyDetect/CopyDetect.pdf>



EK 1. Etik Kurul Onayı

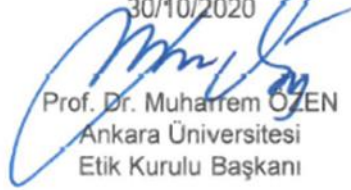
ANKARA ÜNİVERSİTESİ SOSYAL BİLİMLER ALT ETİK KURULU KARAR ÖRNEĞİ

Karar Tarihi : 30/10/2020
Toplantı Sayısı : 7
Karar Sayısı : 134

134-Eğitim Bilimleri Enstitüsü doktora öğrencisi **Arzu Uçar**'ın "Kopya belirlemede benzerlik indekslerinin birey-uyum istatistikleri aracılığıyla aşamalı kullanımının I. tip hataların ve gücünün incelenmesi" başlıklı tezi ile ilgili "İnsan Üzerinde Yapılan Klinik Dışı Araştırmalar Başvuru Formu" Etik Kurulumuzca incelendi.

Eğitim Bilimleri Enstitüsü doktora öğrencisi **Arzu Uçar**'ın "Kopya belirlemede benzerlik indekslerinin birey-uyum istatistikleri aracılığıyla aşamalı kullanımının I. tip hataların ve gücünün incelenmesi" başlıklı tezi ile ilgili, 2019 yılı Mayıs ayında Enstitüye sunulan bu doktora tezinin araştırma planı, tamamlanmış bir tez olduğu izlenimine yol açmaktadır. Dolayısıyla etik bir değerlendirme olanağı bulunmamaktadır.

ASLININ AYNIDIR
30/10/2020


Prof. Dr. Muharrem ÖZEN
Ankara Üniversitesi
Etik Kurulu Başkanı

EK 2. Aşamasız ve İki aşamalı Analizlere Ait I. Tip Hata ve Güç Oranları

Tablo 4

Yetenek Düzeyi Altındaki Yöntemlerin I. Tip Hata Oranı Ortalaması

Yetenek	Yöntem	Belirlenen Birey (Çift) Oranı %f	I. Tip Hata Oranı
Düşük Yetenek	ω	2.21	0.0279
	<i>GBT</i>	0.36	0.0107
	l_z	6.29	0.0607
	l_z^*	8.00	0.0814
	<i>KL</i>	16.36	0.1829
Çok Düşük Yetenek	ω	2.50	0.0193
	<i>GBT</i>	1.93	0.0071
	l_z	2.14	0.0186
	l_z^*	10.43	0.0171
	<i>KL</i>	10.21	0.1550

Tablo 5

Test Güçlüğü Düzeyleri Altındaki Yöntemlerin I. Tip Hata Oranı Ortalaması

Test Güçlüğü	Yöntem	Belirlenen Birey (Çift) Oranı %f	I. Tip Hata Oranı
Kolay	ω	2.2	0.0221
	<i>GBT</i>	0.4	0.0035
	l_z	6.3	0.0628
	l_z^*	8.0	0.08
	<i>KL</i>	16.4	0.1635
Zor	ω	2.5	0.025
	<i>GBT</i>	1.4	0.0142
	l_z	1.6	0.0164
	l_z^*	1.9	0.0185
	<i>KL</i>	17.4	0.1742

Tablo 6

Değişimlenen Faktörler Altındaki I. Tip Hata Oranları

Test Güçlüğü	Yetenek	Yönte m	Belirlenen Birey (Çift) Oranı %f	I. Tip Hata Oranı
Kolay	Düşük Yetenek	ω	2.42	0.0242
		<i>GBT</i>	0.28	0.0028
		l_z	10.14	0.1014
		l_z^*	14	0.14
		<i>KL</i>	18	0.18
	Çok Düşük Yetenek	ω	2	0.02
		<i>GBT</i>	0.42	0.0042
		l_z	2.42	0.0242
		l_z^*	2	0.02
		<i>KL</i>	14.71	0.1471
Zor	Düşük Yetenek	ω	3.14	0.0314
		<i>GBT</i>	1.85	0.0185
		l_z	2	0.02
		l_z^*	2.28	0.0228
		<i>KL</i>	18.57	0.1857
	Çok Düşük Yetenek	ω	1.85	0.0185
		<i>GBT</i>	1	0.01
		l_z	1.28	0.0128
		l_z^*	1.42	0.0142
		<i>KL</i>	16.28	0.1628

Tablo 7

Yetenek Düzeyi Altındaki Güç Oranı Ortalaması

Yetenek	Yöntem	Belirlenen Birey (Çift) Oranı %f	Güç Oranı
Düşük Yetenek	ω	98	0.9788
	<i>GBT</i>	99	0.9912
	l_z	94	0.9436
	l_z^*	93	0.9288
	<i>KL</i>	83	0.8260
Çok Düşük Yetenek	ω	98	0.9845
	<i>GBT</i>	100	0.9952
	l_z	98	0.9760
	l_z^*	98	0.9817
	<i>KL</i>	85	0.8450

Tablo 8

Ortalama Test Güçlüğü Düzeyleri Altındaki Güç Oranı Ortalaması

Test Güçlüğü	Yöntem	Belirlenen Birey (Çift) Oranı %f	Güç Oranı
Kolay	ω	98	0.9800
	<i>GBT</i>	99	0.9957
	l_z	92	0.9360
	l_z^*	90	0.9250
	<i>KL</i>	84	0.8417
Zor	ω	98	0.9833
	<i>GBT</i>	99	0.9907
	l_z	98	0.9836
	l_z^*	99	0.9855
	<i>KL</i>	83	0.8293

Tablo 9 *Kopya Oranı Düzeyleri Altındaki Güç Oranı Ortalaması**Kopya Oranı Düzeyleri Altındaki Güç Oranı Ortalaması*

Kopya Oranı	Yöntem	Belirlenen Birey (Çift) Oranı %f	Güç Oranı
0.1	ω	98	0.9771
	<i>GBT</i>	99	0.9914
	l_z	96	0.9593
	l_z^*	95	0.9496
	<i>KL</i>	84	0.8357
0.4	ω	98	0.9843
	<i>GBT</i>	99	0.9943
	l_z	96	0.9568
	l_z^*	96	0.9579
	<i>KL</i>	84	0.8357
0.6	ω	98	0.9836
	<i>GBT</i>	99	0.9939
	l_z	96	0.9632
	l_z^*	96	0.9582
	<i>KL</i>	84	0.8350

Tablo 10

Değişimlenen Faktörler Altındaki Güç Ortalamasına Ortak Etkisi

Test Güçlüğü	Yetenek	Kopya Oranı	Yöntem	Belirlenen Kişi(Çift) Sayısı	Güç Oranı
Kolay	Düşük Yetenek	0.1	ω	97.6	0.9757
			<i>GBT</i>	99.7	0.9971
			l_z	89.6	0.8957
			l_z^*	85.7	0.8571
			<i>KL</i>	82.9	0.8286
		0.4	ω	97.6	0.9757
			<i>GBT</i>	98.8	0.9885
			l_z	90.3	0.9029
			l_z^*	88.3	0.8829
			<i>KL</i>	83.3	0.8329
	0.6	ω	97.6	0.9757	
		<i>GBT</i>	99.3	0.9929	
		l_z	92.0	0.9200	
		l_z^*	88.7	0.8871	
		<i>KL</i>	83.1	0.8314	
	Çok Düşük Yetenek	0.1	ω	98.1	0.9814
			<i>GBT</i>	99.7	0.9971
			l_z	97.7	0.9771
			l_z^*	98.0	0.9800
			<i>KL</i>	85.4	0.8543
0.4		ω	98.7	0.9871	
		<i>GBT</i>	100.0	1.0000	
		l_z	96.1	0.9614	
		l_z^*	97.7	0.9771	
		<i>KL</i>	84.9	0.8486	
0.6	ω	98.4	0.9843		
	<i>GBT</i>	99.9	0.9986		
	l_z	95.9	0.9586		
	l_z^*	96.6	0.9657		
	<i>KL</i>	85.4	0.8543		
Zor	Düşük Yetenek	0.1	ω	97.0	0.9700
			<i>GBT</i>	98.1	0.9814
			l_z	97.9	0.9786
			l_z^*	97.6	0.9757
			<i>KL</i>	82.1	0.8214
	0.4	ω	99.0	0.9900	
		<i>GBT</i>	99.6	0.9957	
		l_z	97.9	0.9786	
		l_z^*	97.9	0.9786	
		<i>KL</i>	97.9	0.9786	

(Devam ediyor)

Tablo 10 (Devam)

Değişimlenen Faktörler Altındaki Güç Ortalamasına Ortak Etkisi

		0.4	<i>KL</i>	82.1	0.8214
Zor	Düşük Yetenek	0.6	ω	98.6	0.9857
			<i>GBT</i>	99.1	0.9914
			l_z	98.6	0.9857
			l_z^*	99.1	0.9914
			<i>KL</i>	82.0	0.8200
	0.1	ω	98.1	0.9814	
		<i>GBT</i>	99.0	0.9900	
		l_z	98.6	0.9857	
		l_z^*	98.6	0.9857	
		<i>KL</i>	83.9	0.8386	
	Çok Düşük Yetenek	0.4	ω	98.4	0.9843
			<i>GBT</i>	99.3	0.9929
			l_z	98.4	0.9843
			l_z^*	99.3	0.9929
			<i>KL</i>	84.0	0.8400
0.6	ω	98.9	0.9886		
	<i>GBT</i>	99.3	0.9929		
	l_z	98.9	0.9886		
	l_z^*	98.9	0.9886		
	<i>KL</i>	83.4	0.8343		

Tablo 11

 *ω ve *GBT* İndekslerinin Yetenek Düzeyi Altındaki I. Tip Hata Oranı Ortalaması*

Yetenek	Yöntem	İlk Aşamada Belirlenen Birey Oranı %f	İkinci Aşamada Belirlenen Çift Oranı %f	I. Tip Hata Oranı
Düşük Yetenek	$l_z + \omega$	6.255	0.095	0.014286
	$l_z + GBT$	6.255	0.04	0.006429
	$l_z^* + \omega$	6.085	0.07	0.010476
	$l_z^* + GBT$	6.085	0.03	0.004762
	$KL + \omega$	4.435	0.005	0.000714
	<i>KL</i>			
	+ <i>GBT</i>	4.435	0	0
Çok Düşük Yetenek	$l_z + \omega$	6.25	0.08	0.011667
	$l_z + GBT$	6.25	0.03	0.004286
	$l_z^* + \omega$	6.135	0.065	0.009524
	$l_z^* + GBT$	6.135	0.02	0.002857
	$KL + \omega$	4.76	0.01	0.001667
	<i>KL</i>			
	+ <i>GBT</i>	4.76	0	0

Tablo 12

ω ve *GBT* İndekslerinin Test Güçlüğü Düzeyleri Altındaki I. Tip Hata Oranı Ortalaması

Test Güçlüğü	Yöntem	İlk Aşamada Belirlenen Birey Oranı %f	İkinci Aşamada Belirlenen Çift Oranı %f	I. Tip Hata Oranı
Kolay	$l_z + \omega$	6.285	0.08	0.011667
	$l_z + GBT$	6.285	0.01	0.001429
	$l_z^* + \omega$	6.175	0.065	0.009524
	$l_z^* + GBT$	6.175	0.01	0.001429
	$KL + \omega$	4.64	0.015	0.002381
	$KL + GBT$	4.64	0	0
Zor	$l_z + \omega$	6.22	0.095	0.014286
	$l_z + GBT$	6.22	0.06	0.009286
	$l_z^* + \omega$	6.045	0.07	0.010476
	$l_z^* + GBT$	6.045	0.04	0.00619
	$KL + \omega$	4.555	0	0
	$KL + GBT$	4.555	0	0

Tablo 13

 ω ve GBT İndekslerinin Değişimlenen Faktörler Altındaki I. Tip Hata Oranları

Test Güçlüğü	Yetenek	Yöntem	İlk Aşamada	İkinci Aşamada	I. Tip Hata Oranı
			Belirlenen Birey Oranı %f	Belirlenen Çift Oranı %f	
Kolay	Düşük Yetenek	$l_z + \omega$	0.07	0.01	0.0100
		$l_z + GBT$	0.01	0.00	0.0014
		$l_z^* + \omega$	0.05	0.01	0.0071
		$l_z^* + GBT$	0.01	0.00	0.0014
		$KL + \omega$	0.01	0.00	0.0014
		$KL + GBT$	0.00	0.00	0.0000
	Çok Düşük Yetenek	$l_z + \omega$	0.09	0.01	0.0133
		$l_z + GBT$	0.01	0.00	0.0014
		$l_z^* + \omega$	0.08	0.01	0.0119
		$l_z^* + GBT$	0.01	0.00	0.0014
		$KL + \omega$	0.02	0.00	0.0033
		$KL + GBT$	0.00	0.00	0.0000
Zor	Düşük Yetenek	$l_z + \omega$	0.12	0.02	0.0186
		$l_z + GBT$	0.07	0.01	0.0114
		$l_z^* + \omega$	0.09	0.01	0.0138
		$l_z^* + GBT$	0.05	0.01	0.0081
		$KL + \omega$	0.00	0.00	0.0000
		$KL + GBT$	0.00	0.00	0.0000
	Çok Düşük Yetenek	$l_z + \omega$	0.07	0.01	0.0100
		$l_z + GBT$	0.05	0.01	0.0071
		$l_z^* + \omega$	0.05	0.01	0.0071
		$l_z^* + GBT$	0.03	0.00	0.0043
		$KL + \omega$	0.00	0.00	0.0000
		$KL + GBT$	0.00	0.00	0.0000

Tablo 14

 ω ve GBT İndekslerinin Yetenek Düzeyi Altında Güç Oranı Ortalaması

Yetenek	Yöntem	İlk Aşamada Belirlenen Birey (Çift) Oranı %f	İkinci Aşamada Belirlenen Çift Oranı %f	Güç Oranı
Düşük Yetenek	$l_z + \omega$	6.225	6.1533	0.9896
	$l_z + GBT$	6.225	6.1966	0.9959
	$l_z^* + \omega$	6.1116	6.0433	0.99
	$l_z^* + GBT$	6.1116	6.085	0.9961
	$KL + \omega$	4.315	4.3066	0.9935
	$KL + GBT$	4.315	4.3066	0.9942
Çok Düşük Yetenek	$l_z + \omega$	6.235	6.1816	0.9788
	$l_z + GBT$	6.235	6.21833	0.9840
	$l_z^* + \omega$	6.1516	6.1	0.979
	$l_z^* + GBT$	6.1516	6.1366	0.9842
	$KL + \omega$	5.585	5.5366	0.9943
	$KL + GBT$	5.585	5.5366	0.9942

Tablo 15

 ω ve GBT İndekslerinin Ortalama Test Güçlüğü Düzeyleri Altındaki Güç Oranı Ortalaması

Test Güçlüğü	Yöntem	İlk Aşamada Belirlenen Birey Oranı %f	İkinci Aşamada Belirlenen Çift Oranı %f	Güç Oranı
Kolay	$l_z + \omega$	6.225	6.1516	0.9894
	$l_z + GBT$	6.225	6.205	0.9971
	$l_z^* + \omega$	6.0933	6.0266	0.9903
	$l_z^* + GBT$	6.0933	6.0783	0.9978
	$KL + \omega$	4.89	4.8783	0.9944
	$KL + GBT$	4.89	4.8816	0.995
Zor	$l_z + \omega$	6.235	6.1833	0.979
	$l_z + GBT$	6.235	6.21	0.9828
	$l_z^* + \omega$	6.17	6.1166	0.9787
	$l_z^* + GBT$	6.17	6.1433	0.9825
	$KL + \omega$	5.01	4.965	0.9935
	$KL + GBT$	5.01	4.9616	0.9935

Tablo 16

 ω ve GBT İndekslerinin Kopya Oranı Düzeyleri Altındaki Güç Oranı Ortalaması

Kopya Oranı	Yöntem	İlk Aşamada Belirlenen Birey Oranı %f	İkinci Aşamada Belirlenen Çift Oranı %f	Güç Oranı
0.1	$l_z + \omega$	6.225	6.155	0.9899
	$l_z + GBT$	6.225	6.2075	0.9974
	$l_z^* + \omega$	6.105	6.04	0.9905
	$l_z^* + GBT$	6.105	6.0825	0.9966
	$KL + \omega$	4.525	4.4725	0.9920
	$KL + GBT$	4.525	4.4625	0.9903
0.4	$l_z + \omega$	6.185	6.135	0.9928
	$l_z + GBT$	6.185	6.1625	0.9967
	$l_z^* + \omega$	6.11	6.06	0.9928
	$l_z^* + GBT$	6.11	6.0875	0.9967
	$KL + \omega$	5.1025	5.0775	0.9910
	$KL + GBT$	5.1025	5.085	0.9932
0.6	$l_z + \omega$	6.28	6.2125	0.9699
	$l_z + GBT$	6.28	6.2525	0.9757
	$l_z^* + \omega$	6.18	6.115	0.9702
	$l_z^* + GBT$	6.18	6.1625	0.9771
	$KL + \omega$	5.2225	5.215	0.9989
	$KL + GBT$	5.2225	5.2175	0.9992

Tablo 17

 ω ve GBT İndekslerinin Değişimlenen Faktörler Altındaki Güç Oranı

Test Güçlüğü	Yetenek	Kopya Oranı	Yöntem	İlk Aşamada Belirlenen Birey Oranı %f	İkinci Aşamada Belirlenen Çift Oranı %f	Güç Oranı
Kolay	Düşük Yetenek	0.1	$l_z + \omega$	6.2	6.13	0.99
			$l_z + GBT$	6.2	6.2	1
			$l_z^* + \omega$	6.05	6	0.9928
			$l_z^* + GBT$	6.05	6.04	0.9985
			$KL + \omega$	4.23	4.23	0.99
			$KL + GBT$	4.23	4.23	0.99
		0.4	$l_z + \omega$	6.17	6.09	0.9885
			$l_z + GBT$	6.17	6.12	0.9928
			$l_z^* + \omega$	6.07	6	0.99
			$l_z^* + GBT$	6.07	6.02	0.9928
			$KL + \omega$	4.3	4.25	0.98
			$KL + GBT$	4.3	4.25	0.98
	0.6	$l_z + \omega$	6.39	6.26	0.9811	
		$l_z + GBT$	6.39	6.33	0.9914	
		$l_z^* + \omega$	6.28	6.17	0.9838	
		$l_z^* + GBT$	6.28	6.26	0.9971	
		$KL + \omega$	4.52	4.52	1	
		$KL + GBT$	4.52	4.52	1	
	Çok Düşük Yetenek	0.1	$l_z + \omega$	6.34	6.26	0.9885
			$l_z + GBT$	6.34	6.33	0.9985
			$l_z^* + \omega$	6.21	6.14	0.9897
			$l_z^* + GBT$	6.21	6.2	0.9985
			$KL + \omega$	5.15	5.13	0.9966
			$KL + GBT$	5.15	5.15	1
0.4		$l_z + \omega$	6.16	6.13	0.9957	
		$l_z + GBT$	6.16	6.16	1	
		$l_z^* + \omega$	6.04	5.99	0.9928	
		$l_z^* + GBT$	6.04	6.04	1	
		$KL + \omega$	5.6	5.6	1	
		$KL + GBT$	5.6	5.6	1	
0.6	$l_z + \omega$	6.09	6.04	0.9928		
	$l_z + GBT$	6.09	6.09	1		
	$l_z^* + \omega$	5.91	5.86	0.9928		
	$l_z^* + GBT$	5.91	5.91	1		
	$KL + \omega$	5.54	5.54	1		
	$KL + GBT$	5.54	5.54	1		

(Devam ediyor)

Tablo 17 (Devam)

 ω ve GBT İndekslerinin Değişimlenen Faktörler Altındaki Güç Oranı

Zor	Düşük Yetenek	0.1	$l_z + \omega$	6.22	6.16	0.9911
			$l_z + GBT$	6.22	6.21	0.9983
			$l_z^* + \omega$	6.09	6.01	0.9880
			$l_z^* + GBT$	6.09	6.06	0.9952
			$KL + \omega$	4.18	4.18	1
			$KL + GBT$	4.18	4.18	1
			$l_z + \omega$	6.16	6.1	0.9914
	$l_z + GBT$	6.16	6.13	0.9957		
	$l_z^* + \omega$	6.08	6.02	0.9914		
	$l_z^* + GBT$	6.08	6.05	0.9957		
	$KL + \omega$	4.26	4.26	0.9914		
	$KL + GBT$	4.26	4.26	0.9957		
	$l_z + \omega$	6.21	6.18	0.9957		
	$l_z + GBT$	6.21	6.19	0.9971		
	$l_z^* + \omega$	6.1	6.06	0.9942		
	$l_z^* + GBT$	6.1	6.08	0.9971		
	$KL + \omega$	4.4	4.4	1		
	$KL + GBT$	4.4	4.4	1		
	$l_z + \omega$	6.14	6.07	0.99		
	$l_z + GBT$	6.14	6.09	0.9928		
	$l_z^* + \omega$	6.07	6.01	0.9914		
	$l_z^* + GBT$	6.07	6.03	0.9942		
	$KL + \omega$	4.54	4.35	0.9813		
	$KL + GBT$	4.54	4.29	0.9713		
$l_z + \omega$	6.25	6.22	0.9957			
$l_z + GBT$	6.25	6.24	0.9985			
$l_z^* + \omega$	6.25	6.23	0.9971			
$l_z^* + GBT$	6.25	6.24	0.9985			
$KL + \omega$	6.25	6.2	0.9926			
$KL + GBT$	6.25	6.23	0.9971			
$l_z + \omega$	6.43	6.37	0.91			
$l_z + GBT$	6.43	6.4	0.9142			
$l_z^* + \omega$	6.43	6.37	0.91			
$l_z^* + GBT$	6.43	6.4	0.9142			
$KL + \omega$	6.43	6.4	0.9957			
$KL + GBT$	6.43	6.41	0.9971			

**EK 3. Rasch Modelde 150 Örneklem Büyüklüğü Ve 30 Test Uzunluğu
Kullanıldığında Elde Edilen Faktör Analizi Ve Paralel Analiz Sonuçları**

Tablo 18

Paralel Analizden Elde Edilen Öz Değerler

Madde No	Kolay		Zor	
	Gerçek Özdeğer	Seçkisiz Veri Özdeğeri	Gerçek Özdeğer	Seçkisiz Veri Özdeğeri
1	8.40	1.92	9.95	2.23
2	1.31	1.45	1.49	1.72
3	1.08	1.27	1.24	1.50
4	0.96	1.10	1.05	1.30
5	0.78	0.96	0.90	1.12
6	0.70	0.85	0.79	0.97
7	0.60	0.72	0.67	0.85
8	0.49	0.62	0.65	0.73
9	0.38	0.51	0.54	0.60
10	0.30	0.41	0.34	0.49

Tablo 19

Zor ve Kolay Test İçin Doğrulamalı Faktör Analizi Sonuçları

	χ^2	sd	(χ^2/sd)	RMSEA	SRMR
Zor Test	622.055	405	1.535	0.060	0.074
Kolay Test	487.918	405	1.204	0.037	0.066

EK 4. Kullback Leibler Uzaklık İndeksi İçin Elde Edilen Kesme Puanları

Tablo 20

KL Değerlerinin Kesme Puanları

Test Güçlüğü	Yetenek Düzeyi	Kopya Oranı	Hata			Güç		
			Youden İndeks	Kar-Yarar	Minimum Puan	Youden İndeks	Kar-Yarar	Minimum Puan
Kolay Test	Düşük Yetenek	0.1	0.05382	2.21767	0.04750	0.05360	2.22072	0.04749
		0.4	0.04908	2.45493	0.04329	0.04915	2.46357	0.04337
		0.6	0.05181	2.40022	0.04543	0.05142	2.40439	0.04540
	Çok Düşük Yetenek	0.1	0.00562	2.35819	0.04543	0.00562	2.34847	0.04540
		0.4	0.00579	2.40480	0.00524	0.00580	2.40401	0.00520
		0.6	0.00496	2.36224	0.00457	0.00497	2.36533	0.00462
Zor Test	Düşük Yetenek	0.1	0.03141	0.24848	0.02774	0.03106	0.24828	0.02774
		0.4	0.03108	0.23357	0.02773	0.03108	0.23360	0.02775
		0.6	0.03249	0.23394	0.02784	0.03251	0.23295	0.02787
	Çok Düşük Yetenek	0.1	0.00605	0.22745	0.00511	0.00605	0.22709	0.00511
		0.4	0.00551	0.22002	0.00430	0.00551	0.21743	0.00432
		0.6	0.00580	0.22891	0.00495	0.00580	0.23083	0.00495

EK 5. Farklı Kullback Leibler Uzaklık İndeksi Kesme Puanlarının I. Tip Hata ve Güç Oranları

Tablo 21

Güç ve I. Tip Hata Oranları

Test Güçlüğü	Yetenek Düzeyi	Hata Oranları				Güç Oranları		
		Kopya Oranı	Youden İndeksi	Kar-Yarar	Minimum Puan	Youden İndeksi	Kar-Yarar	Minimum Puan
Kolay Test	Düşük Yetenek	0.1	0.0315	0.0066	0.36	0.6863	0.0068	0.74
		0.4	0.0282	0.0067	0.357	0.7174	0.0069	0.7426
		0.6	0.0303	0.0066	0.368	0.6979	0.0068	0.7305
	Çok Düşük Yetenek	0.1	0.034	0.0064	0.371	0.9661	0.0066	0.7289
		0.4	0.0346	0.0064	0.135	0.965	0.0066	0.9648
		0.6	0.0299	0.0064	0.130	0.9702	0.0066	0.9695
Zor Test	Düşük Yetenek	0.1	0.0488	0.0064	0.551	0.5148	0.0066	0.5489
		0.4	0.049	0.0065	0.556	0.5094	0.0067	0.5436
		0.6	0.05	0.0066	0.551	0.4993	0.0068	0.5481
	Çok Düşük Yetenek	0.1	0.0151	0.0064	0.2428	0.848	0.0066	0.857
		0.4	0.0140	0.0064	0.2324	0.8591	0.0066	0.8678
		0.6	0.014	0.0065	0.2396	0.8588	0.0067	0.8601

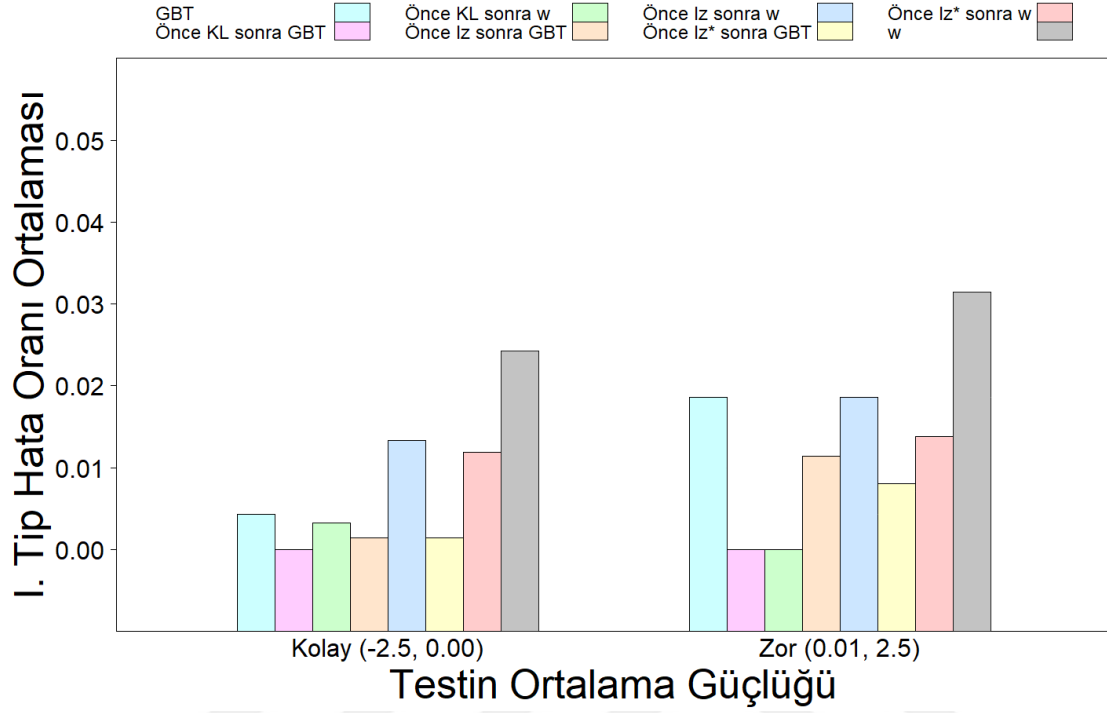
EK 6. Güçlük Parametresi Ortalaması

Tablo 22

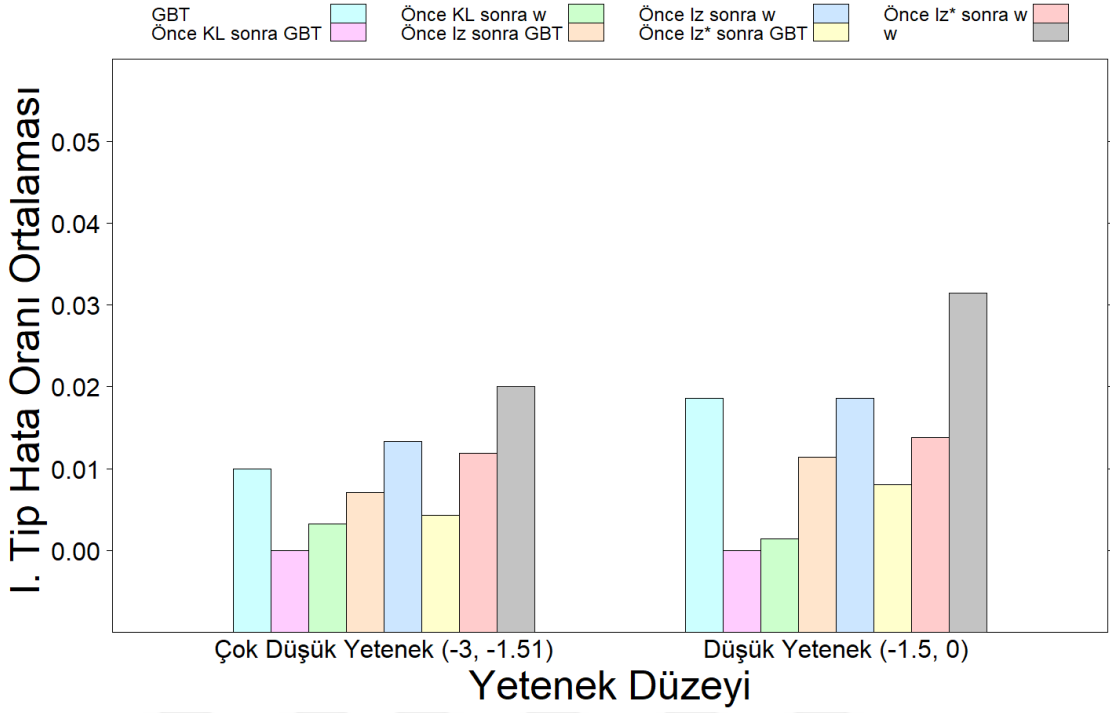
Güçlük Parametresi Ortalaması

Test Güçlüğü	Yetenek Düzeyi	Kopya Oranı	Üretilen Ortalama b Parametresi	Kestirilen Ortalama b Parametresi
Kolay	Çok Düşük Yetenek	0.1	-1.462	-1.779
		0.4	-0.912	-1.852
		0.6	-1.251	-1.903
	Düşük Yetenek	0.1	-1.501	-1.906
		0.4	-1.138	-1.987
		0.6	-1.087	-1.754
Zor	Çok Düşük Yetenek	0.1	1.494	0.850
		0.4	1.658	1.001
		0.6	1.230	1.994
	Düşük Yetenek	0.1	1.879	1.511
		0.4	1.332	1.099
		0.6	1.249	0.755

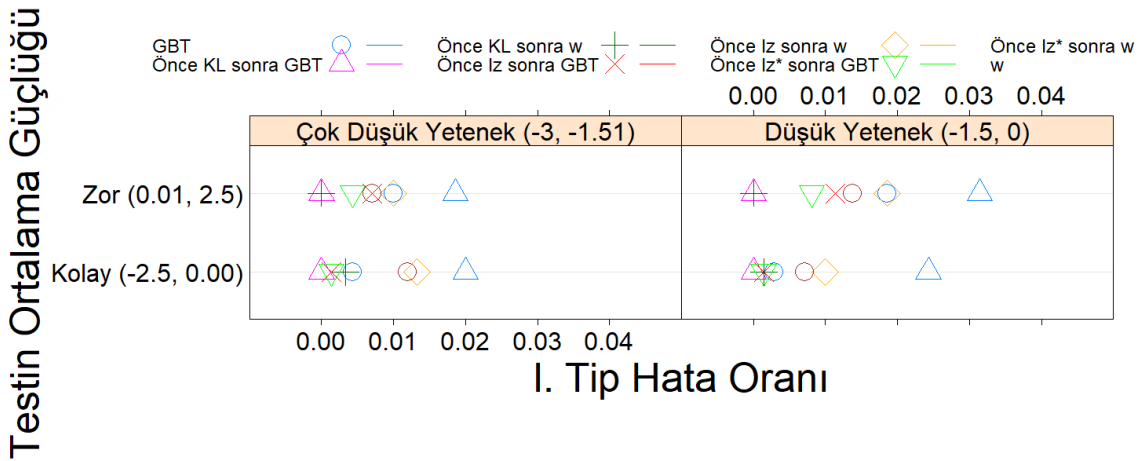
EK 7. ω ve GBT indekslerinin Aşamasız ve İki Aşamalı Analizleri I. Tip Hata Oranları Temel Etki ve Ortak Etki Grafikleri



Şekil XV. Testin Ortalama Güçlüğü'nün I. Tip Hata Oranı Ortalamasına Olan Temel Etkisinin Grafiği. Farkların daha görülebilir olması için (-0.01,0.05) arasında ölçeklendirme yapılmıştır.

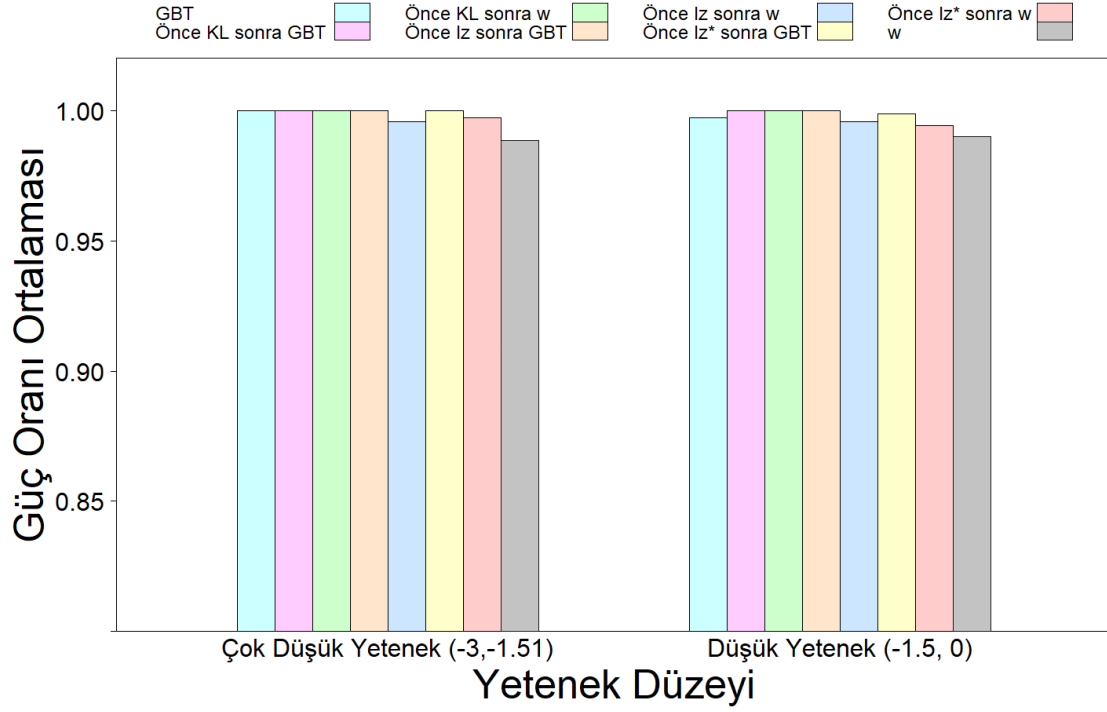


Şekil XVI. Yetenek Düzeyinin I. Tip Hata Oranı Ortalamasına Olan Temel Etkisinin Grafiği. Farkların daha görülebilir olması için (-0.01,0.05) arasında ölçeklendirme yapılmıştır.

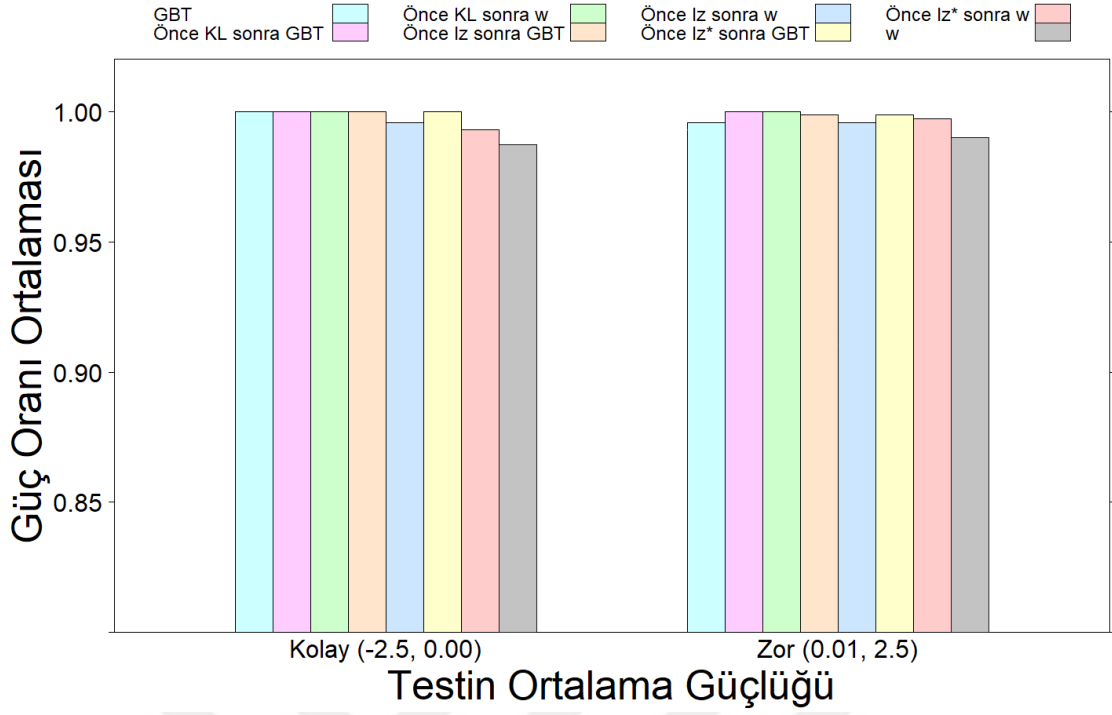


Şekil XVII. Değişimlenen Faktörlerin I. Tip Hata Oranları Olan Ortak Etkisinin Grafiği. Farkların daha görülebilir olması için (-0.001,0.02) arasında ölçeklendirme yapılmıştır.

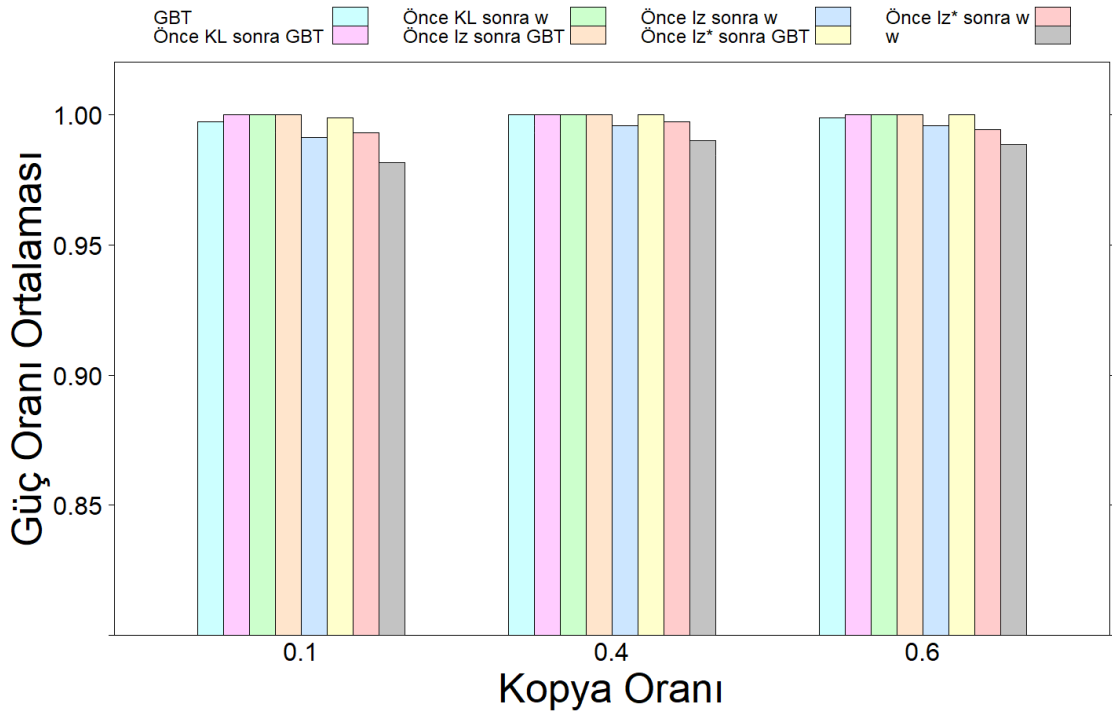
EK 8. ω ve GBT indekslerinin Aşamasız ve İki Aşamalı Analizleri Güç Oranları
Temel Etki ve Ortak Etki Grafikleri



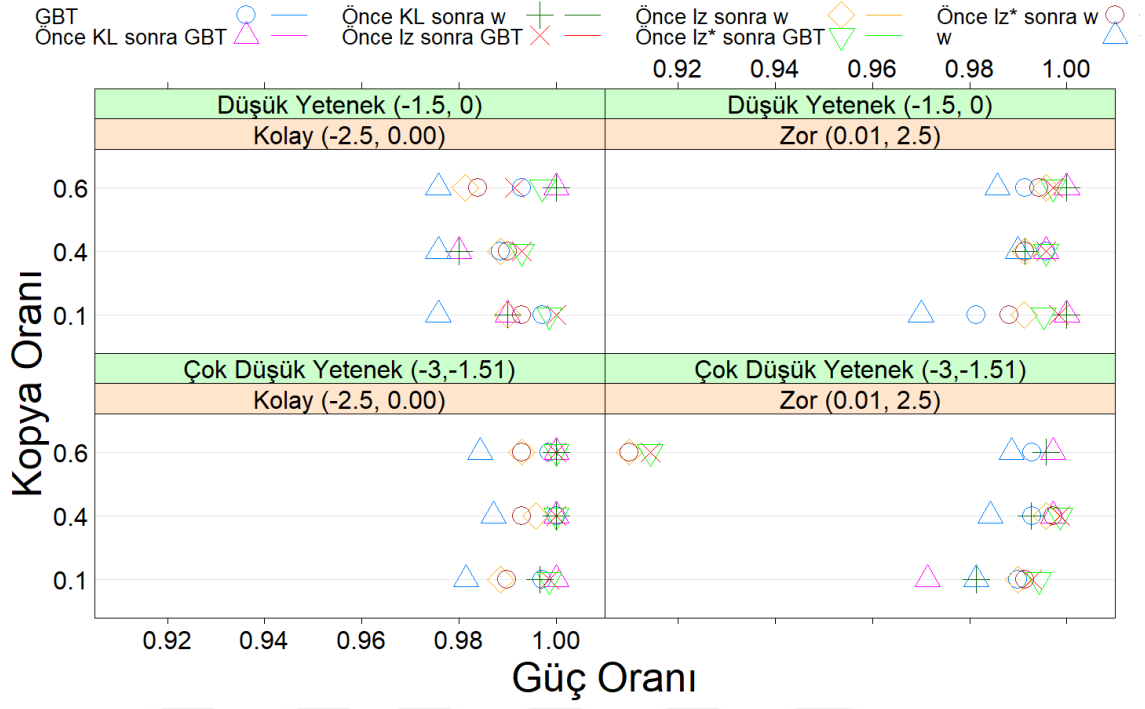
Şekil XVI. Yetenek Düzeyinin Güç Oranı Ortalamasına Olan Temel Etkisinin Grafiği. Farkların daha görülebilir olması için (0.8,1.02) arasında ölçeklendirme yapılmıştır.



Şekil XVIII. Testin Ortalama Güçlüğü'nün Güç Oranı Ortalamasına Olan Temel Etkisinin Grafiği. Farkların daha görülebilir olması için (0.8,1.02) arasında ölçeklendirme yapılmıştır.



Şekil XX. Testin Ortalama Güçlüğü'nün Güç Oranı Ortalamasına Olan Temel Etkisinin Grafiği. Farkların daha görülebilir olması için (0.8,1.02) arasında ölçeklendirme yapılmıştır.



Şekil XVIII VII. Değişimlenen Faktörlerin Güç Oranları Olan Ortak Etkisinin Grafiği. Farkların daha görülebilir olması için (0.9,1.02) arasında ölçeklendirme yapılmıştır.

BENZERLİK BİLDİRİMİ

“Kopya Belirlemede Benzerlik İndekslerinin Birey-Uyum İstatistikleri Aracılığıyla Aşamalı Kullanımının I. Tip Hatalarının Ve Gücünün İncelenmesi” 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) “Beş sözcükten daha az olan benzeşmeler” (2) “Kaynaklar” (3) “Doğrudan Alıntılar” dışarıda tutulmuştur. Benzerlik kontrolüne ilişkin rapordan elde edilen sonuçlar aşağıda sunulmuştur.

Rapor Tarihi	: 06.01.2021
Gönderim Numarası	: 1483660624
Sayfa Sayısı	:59
Sözcük Sayısı	: 15340
Karakter Sayısı	: 104945
Benzerlik Oranı	: %1
Savunma Tarihi	: 28 Ocak 2021

Yukarıda belirtilen sonuçları gösteren Turnitin İntihali Engelleme Programı’na ilişkin orijinal raporu, sonuçlarda herhangi bir değişiklik yapmaksızın bu beyanım ekinde Enstitüye teslim ettiğimi, tezimin %10’dan fazla benzerlik oranı içerdiğinin belirlenmesi durumunda, bundan doğabilecek tüm yasal sorumluluğu kabul ettiğimi bildirir, saygılarımı sunarım.

Öğrencinin Adı Soyadı: Arzu UÇAR

Tarih: 06.01.2021

İmza:

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı ve Soyadı :Arzu UÇAR

E-Posta Adresi :

İş Deneyimi :

Unvan	Görev Yeri	Yıl
Araştırma Görevlisi	Hakkari Üniversitesi Eğitim Fakültesi/Eğitim Bilimleri Bölümü/ Eğitimde Ölçme ve Değerlendirme Anabilim Dalı (ÖYP)	2020- Devam Ediyor
Araştırma Görevlisi	Ankara Üniversitesi Eğitim Bilimleri Enstitüsü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı	2017- 2020
Araştırma Görevlisi	Mersin Üniversitesi Eğitim Bilimleri Enstitüsü, Eğitimde Ölçme ve Değerlendirme Anabilim Dalı	2014-2016
Araştırma Görevlisi	Hakkari Üniversitesi Eğitim Fakültesi/Eğitim Bilimleri Bölümü/ Eğitimde Ölçme ve Değerlendirme Anabilim Dalı	2013- 2014

Akademik Bilgiler

Öğrenim Durumu:

Derece	Bölüm/Program	Üniversite	Yıl
Lisans	Matematik	Muğla Üniversitesi	2006-2010
Yüksek Lisans	Matematik	Muğla Sıtkı Koçman Üniversitesi	2010-2013
Yüksek Lisans	Eğitimde Ölçme ve Değerlendirme	Mersin Üniversitesi	2013-2016
Doktora	Eğitimde Ölçme ve Değerlendirme	Ankara Üniversitesi	2016-2021

Yayımlar:

Uluslararası - ESCI kapsamındaki dergilerde yayımlanan tam makaleler:

1. Uçar, A., & Doğan, C.D. (2021). Defining cut point for Kullback-Leibler divergence to detect answer copying. *International Journal of Assessment Tools in Education*, 8(1), 156-166. doi: <https://doi.org/10.21449/ijate.864078>

Uluslararası bilimsel toplantılarda sunulan ve bildiri kitaplarında (proceedings) basılan bildiriler:

1. Uçar A. & Balta E. (June,2019). *Comparison of different test termination rules in terms of measurement precision in computerized adaptive testing*. International Conference on Research in Teaching and Education (RTECONF),Vienna,Austria. (Oral Presentation)
2. Balta E., Uçar A. & Sahin A.(June,2019). *Detection of cheating behavior in online unproctored CATs via a validation test*. International Association for Computerized Adaptive Testing (IACAT) Conference, Minneapolis, MN, USA. (Oral Presentation)
3. Çokluk Bökeoglu Ö.,Uçar A. & Balta E. (2018). *Madde Tepki Kuramına Dayalı Test Eşitleme Yöntemlerinden Kestirilen Eşitleme Yöntemlerinin Eşitleme Hatalarının Çeşitli Faktörlere Göre Karşılaştırılması*. V th International Eurasian Educational Research Congress, Antalya, Turkey.
4. Sünbül Ö. &Uçar A. (2018). *Çeşitli Faktörler Altında Alt Testlerin Artı Değer Özelliğine Sahip Olma Durumlarının İncelenmesi*. V th International Eurasian Educational Research Congress, Antalya, Turkey.
5. Sünbül Ö. & Kapçık A. (2017). *Artı Değer Özelliğine Sahip Olan Ve Sahip Olmayan Alt Testlerde Eşitleme Yöntemlerinin Karşılaştırılması*. V th International Eurasian Educational Research Congress, Antalya, Turkey.

Ulusal bilimsel toplantılarda sunulan ve bildiri kitaplarında basılan bildiriler:

1. Uçar A.,Kozan S.,Balta E., Uzun N. B. &Odabası O. (2015). *Diş Hekimliği Öğrencilerinin İletişim Becerilerinin Değerlendirilmesinde Kullanılacak Uygun Koşul Sayılarının Genellenabilirlik Teorisiyle Belirlenmesi*. Hacettepe Üniversitesi Klinik ve İletişim Becerileri Kongresi, Ankara, Turkey.