

**ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

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

**MADDE ÖN-BİLGİSİ KAYNAKLI TEST
HİLESİNİN MADDE TEPKİ SÜRESİ MODELİ
KULLANILARAK FARKLI KOŞULLAR ALTINDA
İNCELENMESİ**

DOKTORA TEZİ

EBRU BALTA

**ANKARA
OCAK, 2021**



**ANKARA ÜNİVERSİTESİ
EĞİTİM BİLİMLERİ ENSTİTÜSÜ**

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

**MADDE ÖN-BİLGİSİ KAYNAKLI TEST HİLESİNİN
MADDE TEPKİ SÜRESİ MODELİ KULLANILARAK
FARKLI KOŞULLAR ALTINDA İNCELENMESİ**

DOKTORA TEZİ

EBRU BALTA

DANIŞMAN: DOÇ. DR. CELAL DEHA DOĞAN

**ANKARA
OCAK, 2021**

Ankara Üniversitesi Eğitim Bilimleri Enstitüsü Müdürlüğüne,

Ebru BALTA adlı öğrencinin hazırladığı “Madde Ön-Bilgisi Kaynaklı Test Hilesinin Madde Tepki Süresi Modeli Kullanılarak Farklı Koşullar Altında İncelenmesi” başlıklı bu çalışma Eğitim Bilimleri Anabilim Dalı / Eğitimde Ölçme ve Değerlendirme Programı’nda jüri üyelerince oy birliği ile **Doktora Tezi** olarak kabul edilmiştir.

	Jüri Üyeleri	İmza
Başkan	Prof.Dr.Ömay ÇOKLUK BÖKEOĞLU	
Üye (Danışman)	Doç.Dr.Celal Deha DOĞAN	
Üye	Doç.Dr.Önder SÜNBÜL	
Üye	Prof.Dr.Nuri DOĞAN	
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.

Ebru BALTA

ÖZET

MADDE ÖN-BİLGİSİ KAYNAKLI TEST HİLESİNİN MADDE TEPKİ SÜRESİ MODELİ KULLANILARAK FARKLI KOŞULLAR ALTINDA İNCELENMESİ

BALTA, Ebru

Doktora Tezi, Ölçme ve Değerlendirme Anabilim Dalı

Tez Danışmanı: Doç. Dr. Celal Deha DOĞAN

Ocak, 2021, xvi+171 sayfa

Bu araştırmada, madde tepki ve tepki süresi örüntülerinin birlikte hiyerarşik modellenmesi (HMTK-TS) aracılığıyla madde ön bilgisi kaynaklı test hile davranışlarını tespit etmede, tepki süresi tabanlı *Kullback-Leibler* yaklaşımının ve L birey uyum istatistiğinin, farklı koşullar altında I. tip hata ve test hilesi belirleme güç oranlarının ortaya konulması amaçlanmaktadır. Araştırmanın amacı doğrultusunda, yöntemlerin, I. Tip hata oranı ortalamalarının elde edilmesi için madde tepkilerinin üç parametrelili lojistik model ile tepki sürelerinin de bayesci log-normal tepki süresi modeli ile modellendiği, test uzunluğu (15, 50) ve teşhir olan maddelerin güçlük düzeyi (orta güçlük, zor) koşulları altında 50 yineleme ile 200 veri seti oluşturulmuştur. Yöntemlerin güç oranı ortalamalarının elde edilmesi için ise test uzunluğu, teşhir olan maddelerin güçlük düzeyi, teşhir olan maddelerin oranı (%20, %40 ve %60) ve teşhir olan maddelerin tepki süresi değişimlemesi (random olarak 10 ile 15 saniye arasındaki uniform dağılımdan oluşturma, 20 saniyeye sabitleme, 30 saniyeye sabitleme) koşulları altında 50 yineleme ile 1800 veri seti oluşturulmuştur. Her üretilen veri seti için HMTK-TS model parametreleri kestirimlerinde, Monte Carlo Markov Zinciri (MCMC) yaklaşımı olarak Gibbs örnekleme algoritması kullanılmıştır. Verilerin analizinde, hile senaryosunda, hile yapan bireyler düşük yetenek düzeyindeki bireylerden seçilerek oranı, 1000 örneklem büyüklüğünün %5 'i olacak şekilde hileli veriler oluşturularak yöntemlerin , test hilesi belirleme performansları incelenmiştir. Veri üretiminde ve analizinde R.4.0.1. programı kullanılmıştır. Çalışmanın sonucunda, değişimlenen tüm koşullar altında, uygun olmayan madde tepki ve tepki süresi örüntülerinin tespit edilmesinde, *Kullback-Leibler* uzaklığının L birey uyum istatistiğine göre daha güçlü olduğu; ancak yüksek I. Tip hata oranı barındırdığı görülmüştür.

Anahtar Sözcükler: test güvenliği, madde tepki süresi modeli, birey uyum istatistikleri, MCMC, *Kullback-Leibler* uzaklığı yaklaşımı



ABSTRACT

INVESTIGATION OF ITEM PRE-KNOWLEDGE CHEATING USING ITEM RESPONSE TIME MODEL UNDER DIFFERENT CONDITIONS

BALTA, Ebru

Dissertation, Department of Measurement and Evaluation

Supervisor: Assoc. Prof. Dr. Celal Deha DOĞAN

January, 2021, xvi+171 pages

In this study, it is aimed to reveal in the detection of preknowledge cheating behaviors the rate of Type I error and the rate of power of response time based Kullback-Leibler divergence measure and L person fit statistic under different conditions via modeling patterns of response accuracy and response times using a joint hierarchical model. In line with the purpose of the study, 200 data sets were generated with 50 replication that item responses were modeled with three parameter logistic model and response times were modeled with the Bayesian log-normal response time model under the conditions of test length (15, 50) and difficulty level of the compromised items (medium difficulty, difficult) for obtained means of Type I error rate of methods. In order to obtain the mean power rate of the methods, 1800 data sets were generated with 50 replication under the conditions of test length, the difficulty level of the compromised items, the ratio of compromised items (20%, 40%, 60%), changing in response time of the compromised items (from a uniform distribution in the range of 10 to 15 seconds, fix to 20 seconds, fix to 30 seconds). Gibbs sampling algorithm is used as Monte Carlo Markov Chain (MCMC) approach in estimating patterns of response accuracy and response times using a joint hierarchical model parameters for each data set. In the analysis of the data, in the cheating scenario, who have item preknowledge individuals were selected from individuals with low ability level, and the rate of fraudulent data was created as 5% of the 1000 sample size, were examined of the methods of performance of pre-knowledge cheating. R.4.0.1 software was used for data generation and analyses. As a result of the study, it was found that the *Kullback-Leibler* divergence measure was stronger than the L person fit statistics in detecting aberrant patterns of item response

and response times patterns under all changing conditions; however, it was found to have a high Type I error rate.

Key Words: test security, item response time model, person fit statistics, MCMC, *Kullback-Leibler* divergence measure



ÖNSÖZ

Küresel çapta görülen COVID 19 pandemi dönemi ile birlikte bireylerin hayatlarında önemli kararların alınmasında önemli yer tutan, geniş ölçekli ve yüksek riskli sınavların uygulanması ve uzaktan eğitim sürecinde, ölçme ve değerlendirme uygulamalarında kullanılan çevrim içi sınav uygulamalarının gerekliliği günümüzde giderek hız kazanmaktadır. Kullanımı yaygınlaşan çevrim içi bilgisayar temelli test uygulamalarında, test puanlarının geçerlik ve güvenilirliğine ilişkin daha fazla kanıt sunulmasında, test güvenliğinin sağlanmasına yönelik hile belirleme çalışmalarında, bireylerin test maddelerine ilişkin madde tepkilerinin yanında test maddesinin bilgisayar ekranında görülme anı, bireyin test maddesini çözme süresi ve maddeye tepki verme anı gibi daha ayrıntılı bilgilerin elde edilmesinin önemi artmaktadır. Bu nedenle, geniş ölçekli ve yüksek riskli sınav uygulamalarının koşulları da göz önünde bulundurularak madde ön bilgisinden kaynaklı test hilesini tespit etmede, bireylerin madde tepkileri ve tepki süreleri birlikte değerlendirilerek test hilesini belirlemeye yönelik ek kanıtlar ortaya koymayı amaçlayan çalışma gerçekleştirilmiştir. Çalışmadan elde edilen bulgular ışığında uygulamaya ve araştırmaya yönelik önerilerde bulunulmuştur.

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

Tezime görüş ve önerileriyle katkıda bulunan, Prof. Dr. Ömay ÇOKLUK BÖKEOĞLU'na, çalışmamın probleminin tespitinde ve çalışmam süresince görüş ve önerileriyle katkıda bulunan Doç. Dr. Önder SÜNBÜL'e ayrıca tez jürimde yer alarak görüş ve önerileriyle katkı sağlayan Prof. Dr. Nuri DOĞAN'a ve Doç. Dr. Seher YALÇIN'a teşekkürlerimi sunarım.

Yüksek lisans ve doktora eğitimim boyunca desteğini esirgemeyen ve tez çalışmam sürecine büyük katkıları olan değerli yol arkadaşım Arş. Gör. Arzu UÇAR'a , bu anlamlı süreçte bana destek olan değerli dostum Reyhan KARA'ya ve hayatım boyunca desteğini esirgemeyen farklı şehirlerde olsak ta her koşulda yanımda hissettiğim aileme sonsuz teşekkür ederim.

İÇİNDEKİLER

Sayfa

ETİK İLKELERE UYGUNLUK BİLDİRİMİ.....	iii
ÖZET	iv
ABSTRACT	vi
ÖNSÖZ.....	viii
İÇİNDEKİLER.....	ix
TABLolar DİZİNİ.....	xii
ŞEKİLLER DİZİNİ.....	xiii
KISALTMALAR	xvi
BÖLÜM 1.....	1
GİRİŞ.....	1
Problem.....	1
Amaç.....	12
Önem	12
Sınırlılıklar.....	15
BÖLÜM 2.....	17
KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR.....	17
Madde Ön Bilgisi Belirleme Yöntemleri.....	17
Teşhir Olan Maddelerin Ön Bilgisinden Yararlanan Bireylerin Belirlenmesini Sağlayan Yöntemler	18
DGIRTM (Deterministic Gated IRT Model).....	18
FLOR (Final Log-Odds Ratio) İndeksi.....	19
Değişen Birey Fonksiyonu (Differential Person Functioning).....	22
Birey Uyum İstatistikleri	24
Teşhir Olan Maddeleri Belirlemeye Yönelik Yöntemler	34
Hareketli Ortalamalar Yöntemi (Moving Averages).....	34
Log-Odd Oranlar İstatistiği (Log Odds Ratio Statistic)	35
Teşhir Olan Maddelerden Yararlanan Bireylerin Gruplarının Belirlendiği Yöntemler	37
Kümeleme Analizi	37
Kullback-Leibler Uzaklığı Yaklaşımı.....	38
Faktör Analizi	40
Teşhir Olan Maddelerin ve Bu Maddelerin Ön Bilgisinden Yararlanan Bireylerin Birlikte Belirlendiği Yöntemler.....	40

Loglineer Madde Tepki Süresi Modeli.....	41
Etkili Tepki Süresi Modeli.....	42
Bayesci Log-Normal Tepki Süresi Modeli.....	43
Madde Tepkilerinin ve Madde Tepki Sürelerinin Birlikte Hiyerarşik Modellenmesi (HMTK-TS).....	46
Madde Tepkilerinin ve Madde Tepki Sürelerinin Kullanıldığı Karışım Modelleri.....	68
DG-LNRTM (Deterministic Gated Lognormal Response Time Model)	71
İlgili Araştırmalar	73
BÖLÜM 3.....	81
YÖNTEM	81
Araştırmanın Modeli	81
Simülasyon Deseni	81
Araştırmanın Değişkenleri.....	81
Sabit Değişkenler.....	82
Manipule Edilen Değişkenler	83
Verilerin Üretilmesi	86
HMTK-TS model parametrelerinin MCMC ile kestirilmesi.....	87
Test Hilesi Senaryosu	89
Verilerin Analizi	90
Uygun Olmayan Madde Tepkileri ve Tepki Sürelerini Belirlemede Kullback-Leibler Uzaklığının I.Tip Hata ve Test Hilesi Belirleme Güç Oranlarının Belirlenmesi.....	91
Uygun Olmayan Madde Tepkileri ve Tepki Sürelerini Belirlemede Birey Uyum İstatistiklerinin I.Tip Hata ve Test Hilesi Belirleme Güç Oranlarının Belirlenmesi	93
BÖLÜM 4.....	95
BULGULAR VE YORUMLAR	95
Madde Ön Bilgisinden Kaynaklı Test Hilesini Belirlemede Yöntemlerin I. Tip Hata Oranlarına İlişkin Bulgular ve Yorumlar.....	95
Koşulların Test Hilesini Belirlemede Yöntemlerin I. Tip Hata Oranına Temel Etkisi.....	95
Test Uzunluğunun Test Hilesini Belirlemede Yöntemlerin I. Tip Hata Oranına Temel Etkisi.....	96
Teşhir Olan Maddelerin Güçlük Düzeyinin Yöntemlerin Test Hilesini Belirlemede I. Tip Hata Oranına Temel Etkisi.....	98
Koşulların Test Hilesini Belirlemede Yöntemlerin I. Tip Hata Oranına Ortak Etkisi.....	100
Madde Ön Bilgisinden Kaynaklı Test Hilesini Belirlemede Yöntemlerin Hile Belirleme Güç Oranlarına İlişkin Bulgular ve Yorumlar	103
Koşulların Test Hilesi Belirleme Güç Oranına Temel Etkisi	103

Test Uzunluğunun Test Hilesi Belirleme Güç Oranına Temel Etkisi	103
Teşhir Olan Maddelerin Güçlük Düzeyinin Test Hilesi Belirleme Güç Oranına Temel Etkisi	106
Teşhir Olan Maddelerin Oranının Test Hilesi Belirleme Güç Oranına Temel Etkisi	108
Teşhir Olan Maddelerin Tepki Süresi Değişimlesinin Test Hilesi Belirleme Güç Oranına Temel Etkisi	111
Koşulların Test Hilesi Belirleme Güç Oranına Ortak Etkisi	113
BÖLÜM 5	117
SONUÇLAR VE ÖNERİLER	117
Sonuçlar	117
Öneriler	122
Uygulayıcılara Öneriler	122
Araştırmacılara Öneriler	124
KAYNAKLAR	127
EKLER	147
EK 1. Etik Kurul Onayı	148
EK 2. Test Maddeleri İçin 10000 MCMC Yinelemesine İlişkin Madde Güçlük Parametresi İçin İz, Sonsal Olasılık Yoğunluk ve Otokorelasyon Grafikleri	149
EK 3. Test Maddeleri İçin 10000 MCMC Yinelemesine İlişkin Madde Ayırtecdilik Parametresi İçin İz, Sonsal Olasılık Yoğunluk ve Otokorelasyon Grafikleri	152
EK 4. Test Maddeleri İçin 10000 MCMC Yinelemesine İlişkin Zaman Ayırtecdilik Parametresi İçin İz, Sonsal Olasılık Yoğunluk ve Otokorelasyon Grafikleri	155
EK 5. Test Maddeleri İçin 10000 MCMC Yinelemesine İlişkin Zaman Yoğunluk Parametresi İçin İz, Sonsal Olasılık Yoğunluk ve Otokorelasyon Grafikleri	158
EK 6. Geweke Yakınsaklık Test İstatistik Değerleri	161
EK 7. Koşullara Göre Uygun Olmayan Madde Tepkilerini Belirlemede Kullanılan Yöntemlerin I. Tip Hata Oranları	162
EK 8. Koşullara Göre Uygun Olmayan Madde Tepki Sürelerini Belirlemede Kullanılan Yöntemlerin I. Tip Hata Oranları	163
EK 9. Koşullara Göre Uygun Olmayan Madde Tepkilerini Belirlemede Kullanılan Yöntemlerin Test Hilesi Belirleme Güç Oranları	164
EK 10. Koşullara Göre Uygun Olmayan Madde Tepki Sürelerini Belirlemede Kullanılan Yöntemlerin Test Hilesi Belirleme Güç Oranları	165
BENZERLİK BİLDİRİMİ	167
ÖZGEÇMİŞ	169

TABLÖLAR DİZİNİ

Tablo	Sayfa
Tablo 1. Lognormal Modeli Altında Tepki Süreleri İçin Birey Uyum İstatistikleri	62
Tablo 2. Simülasyon Deseni	85
Tablo 3. I.Tip Hata ve Madde Ön Bilgisi Kaynaklı Test Hilesi Belirleme Gücü	90
Tablo 4. HMTK-TS Model Parametrelerine İlişkin Geweke Yakınsaklık Test İstatistik Değerleri	161
Tablo 5. Koşullara Göre Madde Tepkileri İçin Kullback-Leibler Uzaklığı ve L Birey Uyum İstatistiğinin I.Tip Hata Oranları Ortalaması	162
Tablo 6. Koşullara Göre Madde Tepki Süreleri İçin Kullback-Leibler Uzaklığı ve L Birey Uyum İstatistiğinin I.Tip Hata Oranları Ortalaması	163
Tablo 7. Koşullara Göre Madde Tepkileri İçin Kullback-Leibler Uzaklığı ve L Birey Uyum İstatistiğinin Test Hilesi Belirleme Güç Oranları Ortalaması	164
Tablo 8. Koşullara Göre Madde Tepki Süreleri İçin Kullback-Leible Uzaklığı ve L Birey Uyum İstatistiğinin Test Hilesi Belirleme Güç Oranları Ortalaması	165

ŞEKİLLER DİZİNİ

Şekil	Sayfa
Şekil I. Test Maddelerindeki Hız ve Doğruluk İçin Hiyerarşik Modelleme	51
Şekil II. Test Uzunluğuna Göre Uygun Olmayan Madde Tepki Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin I. Tip Hata Oranları Ortalaması Grafiği	96
Şekil III. Test Uzunluğuna Göre Uygun Olmayan Madde Tepki Süresi Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin I. Tip Hata Oranları Ortalaması Grafiği	97
Şekil IV. Teşhir Olan Maddelerin Güçlük Düzeyine Göre Uygun Olmayan Madde Tepki Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin I. Tip Hata Oranları Ortalaması Grafiği	98
Şekil V. Teşhir Olan Maddelerin Güçlük Düzeyine Göre Uygun Olmayan Madde Tepki Süresi Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin I. Tip Hata Oranları Ortalaması Grafiği	99
Şekil VI. Değişimlenen Koşullara Göre Uygun Olmayan Madde Tepki Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin I. Tip Hata Oranları Grafiği	101
Şekil VII. Değişimlenen Koşullara Göre Uygun Olmayan Madde Tepki Süresi Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin I. Tip Hata Oranları Grafiği	102
Şekil VIII. Test Uzunluğuna Göre Uygun Olmayan Madde Tepki Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin Güç Oranı Ortalaması Grafiği	104
Şekil IX. Test Uzunluğuna Göre Uygun Olmayan Madde Tepki Süresi Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin Güç Oranı Ortalaması Grafiği	105
Şekil X. Teşhir Olan Maddelerin Güçlük Düzeyine Göre Uygun Olmayan Madde Tepki Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin Güç Oranı Ortalaması Grafiği	106
Şekil XI. Teşhir Olan Maddelerin Güçlük Düzeyine Göre Uygun Olmayan Madde Tepki Süresi Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin Güç Oranı Ortalaması Grafiği	107
Şekil XII. Teşhir Olan Maddelerin Oranına Göre Uygun Olmayan Madde Tepki Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin Güç Oranı Ortalaması Grafiği	109

Şekil XIII. Teşhir Olan Maddelerin Oranına Göre Uygun Olmayan Madde Tepki Süresi Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin Güç Oranı Ortalaması Grafiği	110
Şekil XIV. Teşhir Olan Maddelerin Tepki Süresi Değişimlemesine Göre Uygun Olmayan Madde Tepki Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin Güç Oranı Ortalaması Grafiği	111
Şekil XV. Teşhir Olan Maddelerin Tepki Süresi Değişimlemesine Göre Uygun Olmayan Madde Tepki Süresi Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin Güç Oranı Ortalaması Grafiği	112
Şekil XVI. Değişimlenen Koşullara Göre Uygun Olmayan Madde Tepki Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin Güç Oranları Grafiği	114
Şekil XVII. Değişimlenen Koşullara Göre Uygun Olmayan Madde Tepki Süresi Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin Güç Oranları Grafiği.....	115
Şekil XVIII. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Madde Güçlük Parametresi İçin İz Grafikleri	149
Şekil XIX. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Madde Güçlük Parametresi İçin Olasılık Yoğunluk Grafikleri.....	150
Şekil XX. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Madde Güçlük Parametresi İçin Otokorelasyon Grafikleri.....	151
Şekil XXI. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Madde Ayırtecilik Parametresi İçin İz Grafikleri.....	152
Şekil XXII. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Madde Ayırtecilik Parametresi İçin Olasılık Yoğunluk Grafikleri	153
Şekil XXIII. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Madde Ayırtecilik Parametresi İçin Otokorelasyon Grafikleri	154
Şekil XXIV. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Zaman Ayırtecilik Parametresi İçin İz Grafikleri.....	155
Şekil XXV. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Zaman Ayırtecilik Parametresi İçin Olasılık Yoğunluk Grafikleri	156
Şekil XXVI. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Zaman Ayırtecilik Parametresi İçin Otokorelasyon Grafikleri	157
Şekil XXVII. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Zaman Yoğunluk Parametresi İçin İz Grafikleri	158

Şekil XXVIII. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Zaman Yoğunluk Parametresi İçin Olasılık Yoğunluk Grafikleri.....	159
Şekil XXIX. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Zaman Yoğunluk Parametresi İçin Otokorelasyon Grafikleri.....	160



KISALTMALAR

MTK	Madde Tepki Kuramı
2PLM	İki Parametrelili Lojistik Model
3PLM	Üç Parametrelili Lojistik Model
STM	Sınıflamalı Tepki Modeli (Nominal Response Model, NRM)
HMTK-TS	Madde Tepkilerinin ve Madde Tepki Sürelerinin Hiyerarşik Olarak Birlikte Modellenmesi

BÖLÜM 1

GİRİŞ

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

Problem

Günümüzde eğitim, girdi, süreç, çıktı ve kontrol ögeleri olan ve insan davranışlarını geliştiren ve değiştiren bir sistem olarak görülmektedir. Eğitim sisteminin kontrolü, öğelerin iyi işleyip işlemediği, varsa işlemeyen yönlerinin ortaya konulması ve sistemin onarılmasının sağlanması ölçme ve değerlendirme ile gerçekleştirilmektedir. Değerlendirme, bir karar verme, yargıda bulunma süreci olmakla birlikte sistemin sağlıklı işlemesi açısından oldukça önemlidir (Baykul, 2010). Eğitimde ölçme konusu olan ve doğrudan gözlenemeyen, yetenek, başarı, tutum, zeka, ilgi vb. gibi değişkenler, operasyonel olarak tanımlanmasında güçlük çekilen psikolojik değişkenlerdir. Eğitimde ölçme ve değerlendirme çalışmalarının çoğunda, öğrencilerin, bu psikolojik özelliklerinin ortaya çıkarılması amaçlanır. Bu psikolojik özellikler ve yapılar doğrudan gözlenemeyip ancak dolaylı olarak bir ölçme aracıyla gözlenebilmektedirler. Ölçme araçları; bireyleri tanımak, bireyler arasındaki farklılıkları ortaya çıkarmak, bireyleri yetenek düzeylerine göre sıralamak ve ayırmak gibi çeşitli amaçlar için geliştirilebilmektedir. Ölçme araçlarından elde edilen sonuçlara göre, bireyler hakkında çeşitli kararlar alınmakta ve bu alınan kararlar kişisel, sosyal ve politik öneme sahip olabilmektedir. Bireyler hakkında bu derece ciddi kararların verildiği ölçme ve değerlendirme uygulamalarının da sağlıklı olması gerekmektedir. Bunun için, öncelikle, ölçme araçlarının nitelikli olması, bir başka ifadeyle ölçme araçlarının yeterli düzeyde güvenilir ve geçerli olduklarına dair kanıtların toplanmış olması gerekmektedir. Nitelikli bir ölçme aracında bulunması gereken özelliklerden biri güvenirliliktir. Güvenirlilik, bir nitelik ile ilgili olarak aynı bireyler üzerinden yapılan ölçmelerin benzer şartlar oluşturulduğunda tekrarlanabilir olmasıdır. Bir ölçme aracının güvenirliliği; test uzunluğu, grup homojenliği, puanlayıcılar arası tutarlılık, kopya çekme davranışı vb. değişkenlerden etkilenmektedir (Cizek ve Wollack,

2017). Bir ölçme aracında bulunması gereken diğer yapısal özellik ise geçerliktir. Geçerlik ise güvenilirliği de içinde barındıran daha geniş bir terim olup ölçülmek istenilen değişkenin yapısı, ölçme aracının ilgili konu alanını kapsama düzeyi, kullanım amacı, yanlılık vb. pek çok değişkenden etkilenmektedir (Crocker ve Algina, 1986). Güvenirliği ve geçerliği etkileyen değişkenler incelendiğinde, testin psikometrik özellikleri için tehdit oluşturan faktörlerden biri olan test güvenliğini bozan akademik hile davranışlarının (madde ön bilgisi, maddelerin teşhir olması, cevap kopyalama, test işbirliği vb.), bu değişkenler içinde önemli bir yer tuttuğu söylenebilir. Eğitim kurumlarına bireylerin yerleştirilmesinde uygulanan geniş ölçekli testler ve iş alanlarına bireyleri seçmede kullanılan performans değerlendirme ile bireyin yeterlik tahminleri elde edilmeye çalışılmaktadır. Bu yeterlik tahminlerinin bireyler için kişisel, sosyal ve politik bakımdan önemli bir yere sahip olmasından kaynaklı olarak bireylerin, akademik hile yapma motivasyonu artmaktadır. Cizek ve Wollack (2017), akademik hile davranışını, test bağlamında, test uygulama sırasında, öncesinde ya da sonrasında yapılan, haksız bir avantaj elde etmek veya yanlış sonuçlar üretmeye yönelik herhangi bir işlem olarak tanımlamıştır. Böylelikle, akademik hilenin, test geliştirme, testin uygulanma süreci ve test sonuçlarının puanlanması süreçleri ile ilişkili bir durum olduğu görülmektedir. Günümüzde geniş ölçekli sınavların uygulanmasında, test puanlarının bütünlüğünün korunması önemli bir konu olarak görülmekle birlikte test puanının geçerliliğinin ve güvenirliliğinin bir başka yönü haline gelmiştir. Testi alan bireyler arasındaki beklenmedik tepki benzerliği ya da uygun olmayan tepki örüntüleri test verilerindeki düzensizliğin tipik bir göstergesi olup test hilesine işaret etmektedir. Testte hile yapan bireyin test puanı sistematik bir şekilde artıyorsa bu durum puanların geçerliğini, sistematik olarak artmıyorsa bu durum da seçkisiz hata kaynağı olarak değerlendirilip, puanların güvenirliliğini tehdit edeceği söylenebilir. Böylelikle test hilesi, bir ölçme aracının güvenirliliğini, geçerliğini ve buna bağlı olarak da ölçme değerlendirme süreci sonunda bireyler hakkında verilen kararların isabetliliğini olumsuz yönde etkileyebilmektedir. Bu bağlamda, akademik hile davranışının önlenmesi adına; yapılan sınavlarda farklı kitapçık türleri oluşturma, sınav salonlarında birden fazla sınav görevlisi yetkilendirme, test kitapçıklarında madde ve seçeneklerin yerlerini değiştirme, vb. pek çok güvenlik tedbiri alınmaktadır. Ancak günümüz teknolojisinde gerçekleşen hızlı ilerleme sonucunda alınan bu tedbirlerin yetersiz kaldığı görülmektedir. Sınavlarda oluşan test puanlarındaki şüpheli değişiklikler, şüpheli cevapları silme davranışı ve olağandışı tepki benzerlikleri gibi farklı test sahtekarlıkları tartışılmaktadır (Cizek ve Wollack, 2017). Bu durum, araştırmacıları

testlerde yapılan hileleri belirleme yöntemleri üzerine araştırma yapmaya yönelmiş ve böylelikle bu yöntemler üzerine araştırmalar günümüzde hız kazanmıştır.

Cevap kopyalamayı belirlemede kullanılan benzerlik ve birey uyum istatistikleri, madde ön bilgisine sahip olan bireyleri ve teşhir olan maddeleri belirlemeye yönelik kullanılan yöntemler, testlerde oluşan beklenmedik puan artışını ve silme davranışını belirleme yöntemleri, testlerde yapılan hileleri belirlemede kullanılmaktadır. Testlerde yapılan hileleri belirleme yöntemleri deneysel (empirical) ve şans olasılığına dayalı olmak üzere iki şekilde sınıflandırılmıştır. Deneysel yöntemlerde, hile yaptığından şüphelenilen bireylere ait madde tepkilerinin dağılımı ile hile yaptığı bilinen bir gruba ait madde tepkilerinin dağılımları karşılaştırılmaktadır. Şans olasılığına dayalı yöntemlerde ise hile yaptığından şüphelenilen bireylere ait bilinen bir dağılım ile teorik bir dağılım (binom,poission,vb) karşılaştırılmaktadır (Cizek ve Wollack, 2017; Frary, Tideman, ve Watts, 1977; Wollack, 1997). Günümüzde kullanılan yöntemlerin birçoğu bireylerin birbirinden bağımsız, cevap kopyalama ya da ortaklaşma olmadan verilen tepkilerin benzerliğinin olasılığına (şans olasılığı) dayalı yöntemlerdir. Cevap kopyalamadan kaynaklı test hilesini belirleme çalışmalarında, kopyacı ve kaynak durumuna göre cevap kopyalama, benzerlik istatistikleri ve birey-uyum istatistikleri kullanılmaktadır. Cevap kopyalamayı belirlemede kullanılan istatistikler genel olarak bir bireyin potansiyel olarak kaynak olarak belirlenen başka bir bireyden kopya çekip çekmediğini belirlemek için kullanılmaktadır (Angoff,1974; Belov, 2011; Belov ve Armstrong,2010; Belov,Pashley ve Lewis,2007; Frary,1993; Harpp ve Hogan,1993; Holland,1996; Sotaridona, van der Linden ve Meijer, 2006; Wollack,1997). Klasik test kuramına ve madde tepki kuramına dayalı olarak farklı cevap kopyalama istatistikleri [G indeksi (Guttman,1950), *Angoff'un A değişimi* (Angoff,1974), g_1 ve g_2 indeksi (Frary,Tideman ve Watts,1977), K indeksi (Holland, 1996) , ω indeksi (Wollack, 1997), K_1 ve K_2 indeksleri (Sotaridona ve Meijer, 2002) , S_1 ve S_2 indeksi (Sotaridona ve Meijer, 2003), Shifted Binomial index (van der Linden ve Sotaridona,2004), Kappa İstatistiği (Sotaridona, van der Linden ve Meijer, 2006), VM – indeks (Belov, 2011) , τ İstatistikleri (van der Ark, Emons ve Sijtsma,2008)] geliştirilmiştir. Cevap kopyalama istatistiklerinden bazıları kopyacı ve kaynağı ayırt etmeden kopyacı ve kaynağın gözlenen benzer tepkileri ile şansa bağlı benzerliği karşılaştırarak kopya olasılığı üzerine odaklanırken (örneğin, Angoff'un A , Wollack'ın omega'sı) bazıları (örneğin, K indeksi) aynı testi alan iki bireyin, maddelerdeki yanlış cevaplarının olağan dışı olarak eşleşmesini ölçmek amacıyla, bazıları ise (örneğin, τ

İstatistikleri) bireylerin şüpheli cevaplarının belirlenmesi ve ardından şüpheli cevaplardan kopyacı ve kaynağın tespit edilmesi için kullanılabilir.

İşbirliği ile yapılan hile (test collusion), bir test uygulaması öncesinde ya da esnasında iki veya daha fazla bireyin kişisel test puanlarını artırabilmek için maddelere verdikleri tepkileri kopyalayarak ya da birbirleriyle günümüz teknolojisinin gelişmesiyle mikro kayıt cihazları, kameralar, çok programlı hesap makineleri, avuç içi bilgisayarlar ile paylaşmasını ifade etmektedir. Bireylerin, testteki maddelerin cevaplarını kopyalayarak ya da birbirleriyle paylaşarak cevaplayıp cevaplamadıklarını tespit etmek için iki tepki vektörü arasındaki uyum derecesini inceleyen tepki benzerliği istatistikleri ve bir tepki vektörünün belirli bir tepki modeli altındaki uyumunu inceleyen birey-uyum istatistikleri olmak üzere iki istatistiksel yöntem geliştirilmiştir. Bireylerin tepkilerinin birbirinden bağımsız olduğu, ortaklaşma olmadan verilen tepkilerin benzerliğinin olasılığına dayalı olarak cevap kopyalamayı belirlemede kullanılmak üzere benzerlik istatistikleri üretilmiştir. Tepki benzerliği istatistikleri, referans olarak alınan istatistiksel dağılıma veya şüpheli iki tepki vektörü arasındaki benzerlik olasılığına göre iki farklı sınıfta toplanmaktadır (Zopluoğlu,2017). Bu istatistikler, gözlenen uyum olasılığının hesaplanması için referans alınan istatistiksel dağılım (normal, binom, Poisson veya bileşik binom gibi) ve varsayımlarda farklılık göstermektedir. Ayrıca benzerlik istatistiklerinden bazıları, sadece iki tepki vektörü arasında eşleşen yanlış cevapları kullanırken, bazıları, hem doğru hem de yanlış cevapları uyumun kanıtı olarak kullanmaktadır (Zopluoğlu, 2017). Benzerlik istatistikleri ile ilgili araştırmalar incelendiğinde, bazı araştırmacıların (Angoff,1974; Bellezza ve Bellezza,1989; Holland, 1996), bazı benzerlik istatistiklerini (*PAIR1* ve *PAIR2*, *ESA* (error similarity analysis)), yanlış cevaplar gözönünde bulundurularak yanlış cevaplanan maddelerin istatistiklerin hesaplanmasına eşit derecede etkide bulunduğu varsayımı altında geliştirdikleri görülmektedir. Maynes (2013), testte yer alan bütün maddelerin eşit güçlük düzeyinde olduğu varsayımına dayanan benzerlik istatistikleri algoritmaları geliştirerek bu varsayımın geçersiz olduğunu belirtmiştir. Bununla birlikte, bazı araştırmacılar, doğru cevaplardan önemli bilgilerin elde edilebileceğini savunup sayısı çok olmasa da tüm maddeleri göz önünde bulunduran indeksler (Wesolowsky'nin *Z* benzerlik indeksi (Wesolowsky, 2000a), *GBT* (van der Linden ve Sotaridona, 2006) , *S-Check* (Wollack, 1997), *M₄* (Maynes,2005)) geliştirmişlerdir. Benzerlik istatistiklerinin, test cevaplarının olasılık dağılımlarının bağımsızlık varsayımına dayalı olması, matematiksel model kullanılarak hesaplanması ve yerel ya da şartlı bağımsız olması varsayımlarını sağlaması

gerekmektedir. Bağımsızlık varsayımı sağlanmadığı durumlar, öğretimsel yanlılığa (öğrencinin problemi çözerken yanlış düşünmesi ya da öğretimin yanlış olması), işbirlikli öğrenmeye (öğrencilerin aynı anda birbirlerine yardım etmesi), tekrar eden test uygulamalarına ve test hilesine (öğrencilerin madde ön bilgisine sahip olması, sınav esnasında birbirlerinden cevap kopyalamaları, sınav esnasında iletişim kurmaları, test uygulayıcılarının, testi alan bireylerin maddelere verdiği cevaplarda değişiklik yapması) işaret edebilmektedir. Madde tepkilerinin olasılıklarının uygun matematiksel model kullanılarak modellenmediği durumlar ise, testi alan bireylerin, testteki maddelerin bir kısmını yanıtlamamış olmasının, bazı maddelerin teşhir olmasının, negatif madde ayırtedici maddelerin olmasının ve bazı maddelerde değişen madde fonksiyonu gözlenmesinin göstergesi olabilmektedir. Benzerlik istatistiklerinin yeterli büyük örneklem gerektirmesi, kopyacı ve kopya kaynağı için bazı değerleri kendisinin üretmesi yani bir kopyacının belirlenmesine ihtiyaç duyulmaması ,bir grubun yapısının ortaya çıkartılmasında kümeleme tekniklerini kullanmaya izin vermesini sağlayan simetri özelliğinin varlığı, paylaşılan madde sayısına duyarlı olması, dağılımlarının testi alan bireylerin performanslarına bağlı olması ve istatistiklerin modellenirken hangi maddelerin gözönünde bulundurulması gerektiğinin belirlenmesi sınırlılıkları olarak gösterilebilmektedir (Cizek ve Wollack, 2017; Wollack ve Maynes, 2011). Sonuç olarak, tepki benzerliği indekslerinin iki tepki vektörü arasındaki uyum derecesine odaklandığı görülmektedir. Birey uyum istatistikleri ise, tek bir tepki vektörünün belirli bir tepki modeli altındaki uyumunu incelemektedir. Birey uyum istatistikleri ile bireyin yeteneği ile maddelere verdiği tepkiler arasındaki ilişkiler incelenerek bireyin tepki örüntüsünün uygun (normal) ya da uygun olmaması durumu (anormal) belirlenmektedir (de la Torre ve Deng, 2008; Deng, 2007; Dimitrov ve Smith, 2006; Karabatsos, 2003). Bireylerin, maddelerin cevaplarını kopyalama, madde ön bilgisine sahip olarak hile yapma, şans ile tahminde bulunma ve random tepki davranışlarının uygun olmayan tepki örüntülerine neden olduğu görülmektedir. (Cizek ve Wollack, 2017; Haberman ve Lee, 2017; Karabatsos, 2003; Kingston ve Clark, 2014; Lee ve Haberman, 2016; Sijtsma ve Meijer, 1992; Sinharay, 2017a). Böylelikle, birey uyum istatistikleri ile bireylerin uygun olmayan tepki örüntüleri belirlenebilmesine rağmen, uygun olmayan tepki örüntülerin varlığının nedeninin cevap kopyalama olduğu konusunda yetersiz kalmasından kaynaklı olarak bu istatistiklerin daha çok madde ön bilgisinden kaynaklı test hilelerini belirleme çalışmalarında kullanıldığı görülmektedir (Wollack, 1997).

Uygun olmayan tepki örüntülerinin ortaya çıkmasına sebep olan madde ön bilgisi, sınav güvenlik önlemlerinin ihlal edilerek sınav uygulaması öncesinde, madde havuzundaki bazı sınav maddelerine yasa dışı erişilmesi ve böylelikle maddelerin ezberlenmesi sonucunda oluşmaktadır. Madde ön bilgisine sahip bireylerin, teşhir olan maddelere doğru cevap verebilecek yeterli düzeyine sahip olmadıkları halde maddeyi doğru cevapladığı ve böylelikle test puanlarının geçerliğinin zedelendiği görülmektedir (Man, Harring, Quyang ve Thomas ,2018). Test puanlarının geçerliğinin tespiti için madde ön bilgisinden kaynaklı uygun olmayan tepki örüntülerinin tespit edilmesinde, bireyin yetenek düzeyinin madde güçlük düzeyine eşit olduğu ve teşhir olan maddelerin kesin olarak bilinmediği durumlarda birey uyum istatistiklerinin duysarsız olması nedeniyle birey uyum istatistikleri ile birlikte bazı istatistiksel yöntemlerin geliştirildiği görülmektedir. Madde ön bilgisinden kaynaklı test hilelerini belirleme yöntemlerini genel olarak, teşhir olan maddelerin ön bilgisinden birey düzeyinde yararlanmış olabilen bireylerin belirlenmesini sağlayan yöntemler, teşhir olan maddeleri belirlemeye yönelik yöntemler, teşhir olan maddelerden yararlanmış olabilen bireylerin gruplarının belirlendiği yöntemler ve teşhir olan maddelerin ön bilgisinden yararlanmış olabilen bireylerin ve teşhir olan maddelerin birlikte belirlendiği yöntemler ve olmak üzere dört kategoride ele almak mümkündür (Cizek ve Wollack,2017).

Teşhir olan maddelerin ön bilgisinden birey düzeyinde yararlanmış olabilen bireylerin belirlenmesini sağlayan yöntemler arasında *DGIRTM (Deterministic Gated IRT Model)*(Shu,2010; Shu, Henson ve Leucht,2013), *FLOR(Final Log-Odds Ratio) İndeksi* (McLeod ve Schnipke, 2006), *Değişen Birey Fonksiyonu (Differential Person Functioning)* (Smith ve Davis Becker, 2011) ve birey uyum istatistiklerinin yaygın kullanıldığı görülmektedir. Birey uyum istatistikleri genellikle parametrik ve parametrik olmayan istatistikler olarak iki kategoride incelenmekte birlikte (Karabatsos, 2003; Mousavi, Tendeiro ve Younesi, 2016) Klasik Test Kuramı ve Madde Tepki Kuramı modellerine göre (Meijer ve Sijtsma, 2001) sınıflandırılabilir. Parametrik ve parametrik olmayan birey uyum istatistikleri arasındaki temel farklılık, madde tepkilerinin Mokken Homojenlik Modeline uyum göstermesi, diğer bir deyişle tek boyutluluk, yerel bağımsızlık ve madde karakteristik eğrilerinin monotonluğu varsayımlarının sağlanması gerekliliğidir (Emons, 2008). Meijer ve Sijtsma (2001), Karabatsos (2003), Thiessen (2008) ve Meijer ve Tendeiro (2014) çalışmalarında, birey uyum istatistiklerinin farklı uygulamalarını tartışmışlardır. Çalışmalar incelendiğinde, madde tepki kuramına dayalı olarak geliştirilen parametrik olan ikili puanlanan maddeler

için U (Wright ve Stone,1979), l_0 (Levine ve Rubin, 1979), W (Wright ve Masters,1982), $D(\theta)$ (Weiss, 1973; Trabin ve Weiss, 1983) , ECI (Tatsuoka, 1984), UB ve UW (Smith,1985), l_z (Drasgow, Levine, ve Williams, 1985), JK , O/E (Drasgow, Levine ve McLaughlin, 1987), c (Levine ve Drasgow, 1988) , l_{zm} (Drasgow, Levine ve McLaughlin, 1991), M (Molenaar ve Hoijtink,1990), χ_{sc}^2 (Klauer ve Retting,1990), $T(X)$ (Klauer,1991) , l_z^* (Snijders,2001) ve çoklu puanlanan maddeler için l_z^p (Drasgow ve diğerleri,1985; Sinharay,2015b) birey uyum istatistiklerinin, klasik test kuramına dayalı olarak geliştirilen parametrik olmayan ikili puanlanan maddeler için G (Guttman, 1944, 1950), $rpbis$, $rbis$ (Donlon ve Fischer, 1968), G^* (van der Flier, 1977), C (Sato,1975), $U3$, (van der Flier, 1980; Meijer, 1994), A , D , E (Kane ve Brennan, 1980), C^* (Harnisch ve Linn, 1981), $ZU3$ (van der Flier, 1982), NCI , ICI (Tatsuoka ve Tatsuoka, 1983), H_i^T (Sijtsma,1986; Sijtsma ve Meijer,1992) ve çoklu puanlanan maddeler için G^p , G_N^p ve U_3^p (Emons, 2008) birey uyum istatistiklerinin geliştirildiği görülmektedir. Bununla birlikte, son yıllarda bireye uyarlanmış bilgisayar temelli test uygulamalarında, madde ön bilgisi belirleme çalışmalarının hız kazanması ile ikili ve çoklu puanlanan maddeler için bazı (Z_c (McLeod ve Lewis ,1999), *parametrik birikimli toplam istatistiği* (*parametric cumulative sum statistics-CUSUM*) (Bradlow Weiss ve Cho, 1998; van Krimpen-Stoop ve Meijer, 2000, 2001; Egberink, Meijer,Veldkamp,Schakel ve Smid.,2010; Meijer,2002), *birikimli toplam istatistiği (CUSUM)* (van Krimpen-Stoop ve Meijer,2002; Krimpen-Stoop ve diğerleri,2010), K (Bradlow,Weiss ve Cho,1998), T (van Krimpen-Stoop ve Meijer,2000)) birey uyum istatistiklerinin geliştirildiği görülmektedir. Teşhir olan maddeleri belirlemeye yönelik yöntemler olarak ise *Hareketli Ortalamalar yönteminin (Moving Averages)* (Han,2003) ve *Log-Odd Oranlar istatistiğinin (Log Odds Ratio Statistic)* (McLeod ve Schnipke ,2006) kullanıldığı görülmektedir.

Sınav uygulaması öncesinde veya esnasında bireyler arasında madde cevaplarının paylaşılması ya da bir kaynaktan kopya çekilmesi, sınava katılan bireyler arasında beklenmedik tepki benzerliğine neden olmaktadır. Böylelikle, son yıllarda araştırmacılar, hile davranışının bir nedeni olarak uygun olmayan tepki örüntülerinin varlığını incelemek için potansiyel test sahtekarlığını grup düzeyinde araştırmaya yönelmişlerdir. Teşhir olan maddelerden yararlanmış olabilen bireylerin gruplarının belirlenmesinde, benzerlik ve cevap kopyalama istatistikleri ile birey-uyum istatistiklerinin kullanıldığı yaklaşımlar ikiden daha fazla bireyin grup olarak, testin uygulanması öncesinde, uygulanma sürecinde ve sonrasında test güvenliğini ihlal etmesini bir başka deyişle hile yapmasını ya da test

maddelerinin cevaplarını paylaşmasını hesaba katmadığından sınırlı kalmaktadır. İlgili çalışmalar incelendiğinde, gruplardaki test puanlarında beklenmedik puan farklılaşmalarını ve beklenmedik tepki örüntülerinin benzerliğinin belirlenmesinde, hiyerarşik lineer modelleme yaklaşımının (Heck ve Thomas, 2015), yanıştan doğruya cevap değişimini belirlemeye yönelik yaklaşımların (van der Linden ve Jeon, 2012; Wollack, Cohen ve Eckerly, 2015; Wibowo, Sotaridona ve Hendrawan, 2013), benzerlik indekslerine dayalı analizlerin (en yakın komşu kümeleme teknikleri bağlamında benzerlik indekslerinin kullanımı (M_4 indeksi; (Maynes, 2005, 2013, 2014a, 2014b, 2017), benzerlik ölçülerinin grup ortalamasının kullanılması (Z_{nm}' ; Sotaridona, Wibowo ve Hendrawan, 2014), birey uyum istatistiklerine dayalı analizlerin (gruplardaki beklenmedik tepkileri olan bireyler için faktör analizinin kullanımı (Q tipi faktör analizi), Rasch model için ağırlıklandırılmamış U birey-uyum istatistiği, (Zhang, Searcy ve Horn, 2011), *Kullback-Leibler uzaklığının* kullanımı (Belov, 2013, 2014, 2016), birey uyum istatistikleri aracılığıyla beklenmedik madde tepki örüntülerine sahip olan belirlenen bireylerin grup oranlarının kullanılması), grupların test puanları dağılımına ve madde tepkilerine dayalı analizlerin (test puanları dağılımı ve madde tepkilerine dayalı faktör ve kümeleme analizi kullanımı (Quintano, Castellano ve Longobardi, 2009), farklı koşullar altında aynı testi alan grupların test puan dağılımlarının ve madde tepkilerinin karşılaştırılması (k -ortalama kümeleme yaklaşımı), zaman içerisindeki test uygulamalarında radikal puan farklılaşmasını belirlemeye yönelik ve tepki örüntülerinin benzerliğine dayalı analizlerin (test puanlarındaki beklenmedik değişiklikler için indeks (SC_{cbt} indeksi, Jacob ve Levitt, 2013), (ANS_{cbt} indeksi, Thiessen, 2007, 2008; Wollack ve Maynes, 2011), iki oranlı Z puanı ve çoklu lojistik regresyon yaklaşımı (Gaertner ve McBride, 2017), bayesli hiyerarşik lineer büyüme modeli yaklaşımı (Skorupski ve Egan, 2012, 2014); Skorupski, Fitzpatrick, Egan, 2017), bayesli polinom karışık- etki büyüme modeli (Liu, Liu ve Simon, 2014), birikimli lojit regresyon modeli (Skorupski ve Murphy, 2013; Clark, Skorupski ve Murphy 2017), regresyon ve kümeleme temelli yaklaşımlar (RedLOD, Simon, 2014; Li, Huang ve Harris, 2014), çoklu modelleme (See, Kim ve Lalancette, 2013; He ve Stockford, 2015), standartlaştırılmış puanlar kullanılarak artık analiz yaklaşımı (Hee ve Tymss, 2014; He ve Stockford, 2015) ve makine öğrenmesi (denetimli, denetimsiz ve pekiştirmeli öğrenme (Bishop, 2006; Murphy, 2012) gibi istatistiksel yöntemlerin kullanıldığı görülmektedir.

Teşhir olan maddelerin ve bu maddelerin ön bilgisinden yararlanan bireylerin birlikte belirlenmesinde ise bireylerin maddelere verdikleri tepki sürelerinin ölçülmesinin önem taşıdığı görülmektedir (Sinharay,2017b; Qian, Staniewska, Reckase ve Woo, 2016; van der Linden ve Guo,2008). Test teorisyenleri, hızı ve gücü ölçen testleri birbirinden ayırmaktadır (Gulliksen, 1950). Güç testlerinde, maddeler zor olmakla birlikte zaman sınırı bulunmamaktadır. Bu testler ile maddeleri alan bireylerin maddelere ne derece doğrulukta cevap verdiklerinin ölçülmesi amaçlanmaktadır. Hız testlerinde ise, maddeler kolay olmakla birlikte zaman sınırı bulunmakta ve bu testler ile bireylerin maddeleri ne kadar hızlı cevaplayabildiklerinin ortaya çıkartılması amaçlanmaktadır. Bazı testler ise hem hız hem de güç bileşenlerinin özelliklerini göstermesinden kaynaklı olarak hızlılık (speeded) testleri olarak adlandırılmaktadır. Hızlılık testleri genellikle güç testlerinin zamanla sınırlandırılarak belirli bir gruba uygulanmasını gerektirmektedir (Schnipke ve Scrams,1997). Son yıllarda kullanımı yaygınlaşan bilgisayar destekli geniş ölçekli ya da sınıf içi yapılan hız ve hızlılık (speeded) testi uygulamaları, bireylerin, test maddelerine verdikleri tepkiler ile daha ayrıntılı bilgiler olan test maddesinin bilgisayar ekranında görülme anı, bireyin test maddesini çözme süresi ve maddeye tepki verme anı gibi zaman bilgilerini de elde etmeyi sağlamaktadırlar. Bu bilgiler ışığında, uygun olmayan madde tepki davranışları (madde ön bilgisi, cevapları kopyalama, hızlı tahmin vb) ve uygun olmayan madde tepki örüntüleri tespit edildiğinde, bu tespiti desteklemek ya da tanımlamak için bireyin test maddeleri karşısındaki kayıtlı zaman bilgilerinin de kullanıldığı görülmektedir. “Tepki Süresi” terimi maddenin bireye sunulması ile bireyin maddeyi cevaplama arasındaki zaman farkı (saniye veya milisaniye) olarak tanımlanmaktadır (Kingsbury, Zara ve Houser, 1993; Schnipke ve Scrams, 1997; Wise ve Kong, 2005). Önceki çalışmalarda, tepki süresinin kullanımı, daha kesin yeterli tahminlerinin elde edilmesinde (Rasch, 1960; Tatsuoka ve Tatsuoka, 1980; Thissen, 1983), uygun süre sınırlarının belirlenmesinde (Bhola, Plake ve Roos, 1993; Gershon, Bergstrom ve Lunz, 1993; Reese, 1993), olağandışı test performanslarının belirlenmesinde, testteki maddeleri cevaplama süre sorunlarının incelenmesinde (Schnipke ve Scrams, 1997) ve bireye uyarlanmış test uygulamalarında madde seçimlerinde (van der Linden, Scrams ve Schnipke, 1998) araştırılmıştır. Ayrıca, tepki süresi bilgisinin, bazı çalışmalarda (Bhola, 1994; Schnipke, 1995, 1996, 1999; Schnipke ve Scrams, 1997; Wise ve DeMars, 2006), madde tepki kuramına dayalı madde parametrelerini iyileştirmek için kullanıldığı, bazı çalışmalarda da (Wise ve DeMars; 2006; Wise ve Kong, 2005) madde çözüm davranışının ve hızlı tahmin davranışının

tespitine uygulanabilirliğinin araştırıldığı görülmektedir. Son yıllarda yapılan çalışmalarda ise (Boughton, Smith ve Ren, 2017; Cizek ve Wollack, 2017; Eckerly, Smith ve Lee, 2018 ; Fox ve Marianti, 2017; Kasli ve Zopluoğlu, 2018; Kasli, Zopluoğlu ve Toton, 2020; Lee, 2018; Marianti, Avetisyan, Veldkamp ve Tijmstra, 2014; Man, Haring, Quayang ve Thomas, 2018; Qian vd., 2016; van der Linden ve Guo, 2008; van der Linden ve Krimpen-Stoop, 2003; Wang, Xu, Shang ve Kuncel, 2018), bireye uyarlanmış ve bilgisayar temelli test uygulamalarında, bireylerin, maddelere verdiği tepki sürelerinin de ölçülmesinin, madde ön bilgisine sahip, cevap kopyalayan ve tahmin ile cevaplayan bireylerin belirlenmesinde önemli bir değişken olarak ele alındığı görülmektedir. Böylelikle, hız-doğruluk ilişkilerini, hız testlerini, test alma stratejilerini (örneğin, ilerleme hızı ve tahmin etme davranışı) ve alt grup farklılıkları gibi konuları araştırmak için çeşitli madde tepki süresi modelleri geliştirilmiştir (Fox, Klein Entink ve van der Linden, 2007; Klein Entink, Fox ve van der Linden, 2009a; Maris, 1993; Meijer ve Sotaridona, 2006; Schnipke ve Scrams, 1997; Thissen, 1983; Wise ve DeMars, 2006; Wise, Ma, Kingsbury ve Hauser, 2010; Wise ve Kong, 2005; van der Linden, 2006, 2007; van der Linden ve Guo, 2008). Tepki sürelerinin nasıl modelleneceği konusu farklı şekillerde ele alınmıştır. Bir yaklaşım, tepki sürelerini, madde tepki kuramına eklenen zaman parametreleriyle modellemiştir (Roskam, 1997; Thissen, 1983). İkinci yaklaşım ise, loglineer tepki süresi modellemesi olan; madde tepki sürelerinin madde tepkilerinden ayrı olarak modellenmesi ile karakterize edilmesidir (Maris, 1993; Schnipke ve Scrams, 1997; van der Linden, Scrams ve Schnipke, 1999; van der Linden ve van Krimpen-Stoop, 2003). Bu yaklaşıma göre, tepki sürelerinin dağılımı, sadece testi alan bireylerin ve test maddelerinin karakteristik özelliklerinden oluşarak madde tepki süreleri, tepkinin doğruluğundan bağımsız olarak modellenmektedir. Üçüncü bir yaklaşım ise; bireyin maddeyi doğru cevaplama için gereken süre olarak tanımlanan Etkili Tepki Süresi (Meijer ve Sotaridona, 2006) modellemesidir. Dördüncü bir yaklaşım, bayesci log-normal tepki süresi modellemesi (van der Linden, 2006) ve madde tepkilerinin ve tepki sürelerinin hiyerarşik olarak birlikte modellenmesidir (van der Linden, 2007; van der Linden ve Guo, 2008). Bir diğer yaygın istatistiksel yaklaşım ise testi alan farklı tepki davranışları gösteren bireylerin, karışık model (mixture model) aracılığıyla belirlenmesidir (Rost, 1990; Meyer, 2010; Schnipke ve Scrams, 1997; von Davier ve Rost, 2007; Wang ve Xu, 2015).

Madde tepkilerinin ve madde tepki sürelerinin hiyerarşik olarak birlikte modellenmesinin (HMTK-TS), herhangi bir madde tepki modelinin ya da tepki süresi

modelinin kullanılmasına izin vermesi ve madde ve birey parametrelerinin parametreleştirilmesine olanak sağlaması açısından esnek bir modelleme olduğu görülmektedir (Lee, 2018). Madde ön bilgisi belirleme çalışmalarında, madde tepkileri ile birlikte madde tepki sürelerinin de değerlendirilmesinin, bireyin uygun olmayan tepkilerinin belirlenmesinde ve madde tepki davranışları hakkında daha ayrıntılı bilgi edinilmesinde önem teşkil ettiği belirtilmiştir (van der Linden, 2007; van der Linden, Klein Entink ve Fox, 2010; Lee, 2018). Bir testten elde edilen puanların uygun olup olmadığı, hile, tahmin ya da seçkisiz yanıtlama davranışları göz önünde bulundurularak doğrudan madde tepki örüntülerinden, madde tepki sürelerinden ya da hem madde tepkilerinden hem de madde tepki sürelerinden tespit edilerek belirlenebilmektedir (Meijer, 1996; van der Linden ve Lewis, 2015). Bu nedenle, yetenek ve hızı ölçmek için HMTK-TS model kullanıldığında, uygun olmayan madde tepki örüntüleri ya da madde tepki sürelerinin normal madde tepki örüntüleri ya da madde tepki sürelerinden ayırt edilmesi gerekmektedir (Fox ve Marianti, 2017). HMTK-TS modelin kullanıldığı çalışmalar incelendiğinde, madde ön bilgisi ya da cevap kopyalamadan kaynaklı uygun olmayan madde tepki örüntülerini belirlemede, beklenen madde tepki örüntülerinden gözlenen madde tepki örüntülerinin değerlendirildiği *Neyman-Person lemma yaklaşımının* (Levine ve Drasgow, 1988), $L(l_s^y)$ (Marianti, Fox, Avetisyan ve Tijmstra, 2014), *PSS (Posterior Shift Statistic)* (Belov, 2015), SBNPL (Belov, 2016), L_s birey-uyum istatistiklerinin (Sinharay, 2017a) ve artık analizlerinin (Qian vd., 2016), uygun olmayan madde tepki süresi örüntülerinin belirlenmesinde ise $L(l^t)$ (Marianti vd., 2014), χ_{pf} (Sinharay, 2018), Λ_s (Sinharay, 2020) birey uyum istatistiklerinin ve tepki süresi tabanlı *Kullback-Leibler uzaklığı* yaklaşımının yaygın olarak kullanıldığı görülmektedir.

Böylelikle günümüze kadar yapılan test güvenliği ile ilgili çalışmalar incelendiğinde, bireylerin, maddelere verdiği tepki sürelerinin, bir istatistiksel teste ve uç değerlerin belirlenmesine yardımcı olmasını sağlayan bir sürekli değişken olması, bireylerin maddeyi cevaplamak için harcadığı süre ile maddeyi yanıtlamak için gerekli olan süre arasında karşılaştırma yapmayı sağlaması, test dizaynı, test geliştirme sürecindeki madde kalibrasyonu, uygun olmayan tepki davranışlarının belirlenmesi (cevapları kopyalama, madde ön bilgisine sahip olma, hızlı tahmin vb.), bireyselleştirilmiş bilgisayar destekli testlerde madde seçim kurallarının etkililiği ve gizil yetenek kestirimi hakkında bilgi verebilmesi açısından önemli olduğu görülmektedir. Bu nedenle, bu çalışmada, geniş ölçekli yapılan bilgisayar temelli test uygulamalarında

madde ön bilgisine sahip olan bireylerin hile davranışları madde tepki süresi modeli kullanılarak incelenmiştir.

Amaç

Bu çalışmanın amacı, madde ön bilgisi kaynaklı test hilelerini tespit etmede, madde tepkilerinin ve madde tepki sürelerinin hiyerarşik olarak birlikte modellendiği durumda, birey uyum istatistiklerinin ve tepki süresi tabanlı *Kullback-Leibler* uzaklığı yaklaşımının, uygun olmayan madde tepkileri ve madde tepki süresi örüntülerini tespit etmede, performansının belirlenmesidir. Bu genel amaç doğrultusunda, aşağıdaki araştırma sorularına cevap aranmıştır:

1. Test uzunluğunun ve teşhir olan maddelerin güçlük düzeyinin, tepki süresi tabanlı *Kullback Leibler* uzaklığı yaklaşımı ve birey uyum istatistiklerinin I. tip hata oranları üzerindeki,
 - a. temel etkisi nedir?
 - b. ortak etkisi nedir?
2. Test uzunluğunun, teşhir olan maddelerin güçlük düzeyinin, teşhir olan maddelerin oranının ve teşhir olan maddelerin tepki süresi değişimlemesinin, tepki süresi tabanlı *Kullback Leibler* uzaklığı yaklaşımı ve birey uyum istatistiklerinin güç oranları üzerindeki,
 - c. temel etkisi nedir?
 - d. ortak etkisi nedir?

Önem

Eğitimde ve psikolojide bireyler hakkında ciddi kararların verildiği geniş ölçekli ve yüksek riskli sınavlarda, puanlama objektifliği nedeniyle güvenilirliği yüksek ölçümler sağlaması yanında birçok avantajından dolayı çoktan seçmeli maddelerden oluşan testler kullanılmaktadır. Bu çoktan seçmeli maddelerin kullanıldığı testlerin güvenilirlik ve geçerliğinin sağlanması, bireyler hakkında kararlar alınırken önemli bir yer tutmaktadır. Gelişen teknoloji ile birlikte bu sınavların uygulanma süreçlerinde, bireylerde gözlenen

çeşitli test hilesi davranışlarından kaynaklı olarak sınavların geçerliği zedelenmektedir. Bireylerin maddelere verdikleri tepkilerin benzerliği, test verilerinde ortaya çıkabilecek bir usulsüzlük türü olarak değerlendirilip potansiyel test hilesinin göstergesi olarak görülmektedir (Cizek ve Wollack, 2017). Böylelikle, kağıt-kalem, bilgisayar tabanlı ve bireye uyarlanmış test uygulamalarında, test puanlarının geçerliğinin tespiti için bireylerin madde tepki örüntülerinin benzerliğinin ve uygun olmayan madde tepki örüntülerinin tespit edilmesinin de önem kazandığı görülmektedir.

Literatürdeki test hilesi ile ilgili araştırmalar incelendiğinde, son yıllarda araştırmacıların (Armstrong ve Shi, 2009; Balta, Uçar ve Şahin, 2019; Belov ve Armstrong, 2010; Belov, 2013, 2014, 2016; Belov, Pashley, Lewis ve Armstrong, 2007; Boughton, Smith ve Ren, 2017; Karabatsos, 2003; Krimpen-Stoop ve Meijer, 2001; Jacob ve Levitt, 2003; Kingston ve Clark, 2014; Lee ve Haberman, 2016; Maynes, 2017; Uçar ve Doğan, 2020; Shu, 2011; Sotaridona ve Meijer, 2002, 2003; Sijtsma ve Meijer, 1992; Sinharay, 2017a, 2017b; Sünbül ve Yormaz, 2018; Qian vd., 2016; Wollack, 1997; Wollack ve Cohen, 1998; Wollack ve Maynes, 2011, 2017; Wollack, 2003, 2006; Yormaz, 2014; Yormaz ve Sünbül, 2017; Yormaz, 2019; Zhang, Searcy, ve Horn, 2011; Zopluoğlu ve Davenport, 2012; Zopluoğlu, 2016) genel olarak, çoktan seçmeli maddelerden oluşan çeşitli sınav uygulamalarında (kağıt-kalem, bilgisayar temelli ve bireye uyarlanmış test), bireylerin, maddelere verdikleri tepkilerin benzerliğine odaklandığı görülmektedir. Çalışmalar incelendiğinde, testlerde yapılan hile davranışını belirlemede birçok yöntemin kullanılmakta olduğu ve bu yöntemler içinde madde tepki kuramı ve klasik test kuramına dayalı çeşitli cevap kopyalama, benzerlik ve birey uyum istatistiklerinin geliştirildiği görülmektedir. Bu istatistikler ile genel olarak, bir bireyin potansiyel kaynak olarak belirlenen başka bir bireyden kopya çekip çekmediğini belirlemek ya da bireyin yeteneği ile maddelere verdiği tepkiler arasındaki ilişkiler incelenerek bireyin tepki örüntüsünün beklenen ya da beklenen dışı olması durumu belirlenmektedir. Ancak bu yaklaşımlar iki veya daha fazla bireyin, testin uygulanması öncesinde, sürecinde ve sonrasında test güvenliğini ihlal etmesini hesaba katmamasından kaynaklı sınırlı kalmaktadır. Sınav öncesinde ya da uygulanma sürecinde, bireyler arasında madde cevaplarının paylaşılması, sınava katılan bireyler arasında beklenmeyen madde tepki örüntü benzerliğine neden olmaktadır. Son yıllarda bilgisayar temelli ve bireye uyarlanmış testlerin kullanımının yaygınlaşması ile birlikte sınav uygulamaları öncesinde maddelerin teşhir olmasından kaynaklı birey ve grup düzeyinde hile davranışlarının belirlenmesinin öneminin arttığı görülmektedir.

Madde ön bilgisine sahip olan bireylerin ve teşhir olan maddelerin belirlenmesine yönelik çalışmalar incelendiğinde, çeşitli yöntemlerin geliştirilmiş olduğu görülmektedir. Bu yöntemlerin yer aldığı çalışmalar (Fox, Klein Entink ve van der Linden, 2007; Kasli ve Zopluoğlu, 2019; Kasli, Zopluoğlu ve Toton, 2020; Klein Entink, Fox ve van der Linden, 2009; Lee, 2018; Meijer ve Sotaridona, 2006; van der Linden, 2007; van der Linden, 2006, 2007, 2009, 2010; van der Linden ve Guo, 2008; van der Linden, Scrams ve Schnipke, 1999; van der Linden ve Krimpen-Stoop, 2003) incelendiğinde, bireylerin madde tepkileri ile birlikte maddelere verdiği tepki sürelerinin de önemli bir değişken olarak ele alındığı görülmektedir. Bu çalışmalarda, cevap kopyalama ve madde ön bilgisinden kaynaklı test hilelerinin belirlenmesinde, madde tepki süresi modellerinin kullanıldığı görülmektedir. Ayrıca, bilgisayar temelli test uygulamalarında, bireylerin madde tepki sürelerinin maddeyi cevaplamadaki çalışma hızı ve zaman yoğunluğu hakkında, madde tepkilerinin ise bireyin yeteneği hakkında bilgi vermesinden kaynaklı olarak madde tepkilerinin de madde tepki süreleri ile birlikte uygun olmayan tepki davranışlarının belirlenmesinde değerlendirilmesinin önerildiği görülmektedir. Bununla birlikte, madde tepkilerinin ve madde tepki sürelerinin birlikte hiyerarşik olarak modellenmesi ile başka bir deyişle uygun olmayan madde tepki davranışlarının belirlenmesinde, madde tepkilerinin ve madde tepki sürelerinin birlikte değerlendirilerek birey hız ve yetenek parametrelerinin birlikte modellenmesi ile test hilesi davranışlarının araştırıldığı az çalışmaya (Fox ve Marianti, 2017; Man, Harring ve Thomas, 2018; Marianti vd., 2014) rastlanmaktadır. Küresel çapta görülen COVID 19 pandemi dönemi ile birlikte bireylerin hayatlarında önemli kararların alınmasında önemli yer tutan, Türkiye’de de, Ölçme, Seçme ve Yerleştirme Merkezi (ÖSYM) ve Milli Eğitim Bakanlığı (MEB) tarafından gerçekleştirilen geniş ölçekli ve yüksek riskli sınavların uygulanması ve uzaktan eğitim sürecinde, ölçme ve değerlendirme uygulamalarında kullanılan sınıf içi ya da okul geneli uygulanan ortak sınavların, çevrim içi uygulanması gerekliliği günümüzde giderek hız kazanmaktadır. Kullanımı yaygınlaşan çevrim içi bilgisayar temelli test uygulamalarında, test puanlarının geçerlik ve güvenilirliğine ilişkin daha fazla kanıt sunulmasında, test güvenliğinin sağlanmasına yönelik hile belirleme çalışmalarında, bireylerin test maddelerine ilişkin madde tepkilerinin yanında test maddesinin bilgisayar ekranında görülme anı, bireyin test maddesini çözme süresi ve maddeye tepki verme anı, bireyler hakkında biyometrik bilgiler (göz izleme ve hareket analizinden elde edilen veriler) vb. gibi daha ayrıntılı bilgilerin elde edilmesinin önemi artmaktadır. Bu nedenle, bu çalışmanın, geniş ölçekli bilgisayar temelli test

uygulamalarında, bireylerin maddelere verdiği tepkiler ve madde tepki sürelerinin birlikte hiyerarşik olarak modellenmesi kullanılarak madde ön bilgisinden kaynaklı test hilesi davranışlarının belirlenmesi üzerine çalışma olması bakımından araştırmanın sonuçlarının literatüre katkıda bulunacağı düşünülmektedir.

Sınırlılıklar

1. Araştırmada, madde tepkilerinin 3PLM, tepki sürelerinin ise bayesci log-normal tepki süresi modeli kullanılarak modellenmesi, araştırmanın, sınırlılığıdır.
2. Araştırmada, *Kullback-Leibler* uzaklığının sıfır hipotezi altında örneklem dağılımının bilinmemesinden kaynaklı olarak uygun olmayan madde tepkilerinin ve madde tepki sürelerinin belirlenmesinde kullanılan kesme puanının elde edilmesinde, ROC analiz yöntemlerinden *Youden İndeksi* kullanılması, araştırmanın, sınırlılığıdır.
3. Uygun olmayan madde tepkilerinin ve madde tepki sürelerinin belirlenmesinde kullanılan *Kullback-Leibler* uzaklık değerlerinin hesaplanmasında, birey yetenek parametrelerinin kestirimi işlemi “*irtos*” paket programıyla sınırlandırılmıştır.
4. Araştırmada, birey uyum istatistikleri ile uygun olmayan madde tepkilerinin ve madde tepki sürelerinin belirlenmesinde, kesme puanının $\alpha=0.05$ bayesci anlamlılık düzeyinde elde edilmesi, araştırmanın, sınırlılığıdır.
5. Araştırmada kullanılan bilgisayar programı ve paket programlarıyla üretilen simülasyon verileri, araştırmaya dahil edilen madde ön bilgisi belirleme yöntemleri, seçilen koşullar (test uzunluğu, teşhir olan maddelerin oranı, teşhir olan maddelerin güçlük düzeyi, teşhir olan maddelerin tepki süresi değişimlemesi ve madde ön bilgisine sahip bireylerin oranı) ve düzeyleri, araştırmanın, sınırlılığıdır.

BÖLÜM 2

KURAMSAL ÇERÇEVE VE İLGİLİ ARAŞTIRMALAR

Bu bölümde, madde ön bilgisi kaynaklı test hilesini belirlemeye yönelik kullanılan yöntemlere yönelik kuramsal bilgilere ve ilgili araştırmalara yer verilmiştir.

Madde Ön Bilgisi Belirleme Yöntemleri

Aşağıda, madde ön bilgisi belirleme yöntemleri;

- Teşhir Olan Maddelerin Ön Bilgisinden Yararlanan Bireylerin Belirlenmesini Sağlayan Yöntemler
- Teşhir Olan Maddeleri Belirlemeye Yönelik Yöntemler
- Teşhir Olan Maddelerden Yararlanan Bireylerin Gruplarının Belirlendiği Yöntemler
- Teşhir Olan Maddelerin ve Bu Maddelerin Ön Bilgisinden Yararlanan Bireylerin Birlikte Belirlendiği Yöntemler

olmak üzere dört ana başlık altında sunulmuştur. Mevcut araştırma kapsamında kullanılan yöntemlerin diğer yöntemler ile olan ilişkisinin anlaşılabilmesi adına sunum sürecinde yukarıda belirtilen sistematik sıra dikkate alınmıştır. Bu bağlamda, mevcut araştırmada kullanılan *Kullback-Leibler* uzaklığına ve madde tepki süresi modeline ilişkin kuramsal bilgiler “teşhir olan maddelerden yararlanan bireylerin gruplarının belirlendiği yöntemler” ve “teşhir olan maddelerin ve bu maddelerin ön bilgisinden yararlanan bireylerin birlikte belirlendiği yöntemler” başlıkları altında sunulmuştur.

Teşhir Olan Maddelerin Ön Bilgisinden Yararlanan Bireylerin Belirlenmesini Sağlayan Yöntemler

DGIRTM (Deterministic Gated IRT Model)

Shu (2010) ve Shu, Henson ve Leucht (2013) tarafından önerilen DGIRTM, potansiyel kopyacıları, onların teşhir olmayan maddelerden elde edilen puanlarından teşhir olan maddeler üzerindeki puan kazancının hesaplanarak belirlenmesini sağlayan yöntemdir. *DGIRTM*, testi alan bireyleri, teşhir olan ve olmayan maddeler aracılığıyla kopyacı ve kopyacı olmayan olarak iki gruba ayırarak her birey için çok boyutlu Rasch modeli kullanılarak kestirilen gerçek yetenek ve hile yeteneği (testi alan bireyin güvenliği ihlal edilmiş olarak belirtilen maddelerdeki (teşhir olan) performansına göre tahmin edilen) olmak üzere iki yetenek düzeyi belirlemektedir. *DGIRTM*, Eşitlik 1 ile tanımlanmaktadır.

$$\begin{aligned} P(U_{ij} = 1 | \theta_{tj}, \theta_{cj}, b_i, T_j, I_i) \\ = P(U_{ij} = 1 | \theta_{tj})^{1-T_j} * [(1 - I_i) \\ * P(U_{ij} = 1 | [(1 - I_i) * P(U_{ij} = 1 | \theta_{tj}) + I_i * P(U_{ij} = 1 | \theta_{cj})]^{T_j} , \end{aligned}$$

$$P(U_{ij} = 1 | \theta_{tj}, b_i) = \frac{e^{(\theta_{tj} - b_i)}}{1 + e^{(\theta_{tj} - b_i)}}$$

$$P(U_{ij} = 1 | \theta_{cj}, b_i) = \frac{e^{(\theta_{cj} - b_i)}}{1 + e^{(\theta_{cj} - b_i)}} , \sum b_i = 0 ,$$

$$\theta_{tj} < \theta_{cj} \text{ ise, } T_j = 1 \quad (1)$$

Eşitlik 1 ile tanımlanan modelde, θ_{tj} ; testi alan j.bireyin gerçek yeteneğini, θ_{cj} ; j.bireyin hile yeteneğini, b_i ; i maddesi için madde güçlük düzeyini ve I_i ise kazanç mekanizmasının göstergesini (i maddesinin kullanıcı tarafından teşhir olmayan ($I_i = 0$) veya teşhir olan için ($I_i = 1$) olduğu varsayılır) ifade etmektedir. Eşitlik 1 incelendiğinde, teşhir olmayan madde için, madde ön bilgisine sahip olan ve olmayan her iki bireyin,

madde üzerindeki tepkilerinin gerçek yetenek düzeyinden kaynaklandığı, teşhir olan madde için ise madde ön bilgisine sahip olmayan bireylerin, madde üzerindeki tepkilerinin gerçek yetenek düzeylerinden, madde ön bilgisine sahip olan bireylerin tepkilerinin ise hile yetenek düzeylerinden kaynaklandığı görülmektedir (Shu,2010).

Shu (2010) çalışmasında, *DGIRTM*'nin duyarlılığını (doğru olarak belirlenen hile yapan bireylerin oranı) ve belirleyiciliğini (doğru olarak belirlenen hile yapmayanların oranı) l_z birey uyum istatistiği ile karşılaştırarak araştırmıştır. Çalışmada, model parametrelerinin kestiriminde, Markov Chain Monte Carlo (MCMC) yöntemi kullanarak teşhir olan ve olmayan maddelerden oluşan testlerde, test uzunluğunu, teşhir olan maddelerin oranını, madde güçlüğü, madde güçlüğü, standart hatasını ve hile karakteristiklerini (hile yapanların sayısı, madde başına düşen hile yapanların sayısı (hile derecesi), hileden elde edilen puan kazançları (hilenin etkililiği) değişimlemiştir. Çalışmanın sonucunda, *DGIRTM*'nin l_z birey uyum istatistiğine göre daha duyarlı olduğu ve birbirine yakın düzeyde belirleyicilikte olduğu görülmüştür. Ayrıca *DGIRTM*'nin hileden elde edilen puan kazancı (yüksek düzeyde madde ön bilgisine sahip olan bireyler), teşhir olan maddelerin oranı ve hilenin derecesi arttıkça duyarlılığının arttığı ve hile yapan bireylerin artışıyla hileyi belirleme oranının azaldığı görülmüştür. Shu ,Henson ve Leucht (2013) çalışmalarında, teşhir olan maddelerin oranını, madde ön bilgisine sahip hile yapan bireylerin oranını ve hile etkililiğini değişimleyerek *DGIRTM*'nin duyarlılığını araştırmışlardır. Çalışmanın sonucunda, yöntemin belirleyiciliğinin teşhir olan maddelerin oranı ve madde ön bilgisine sahip olan bireylerin oranı arttıkça artış gösterdiği, duyarlılığının ise teşhir olan ve olmayan madde oranlarının eşit olduğu durumda ve madde ön bilgisine sahip hile yapan bireylerin oranı yüksek olduğunda arttığı görülmüştür.

FLOR (Final Log-Odds Ratio) İndeksi

McLeod ve Schnipke (2006), bireye uyarlanmış testlerde, çeşitli tür ve miktarlarda, çeşitli kaynak türleri aracılığıyla (yüksek yetenek düzeyine sahip "profesyonel" ve en zor maddeleri ezberlemeyi deneyen "daimi" ve "ortalama yetenek düzeyine sahip daimi kaynaklar"), kaynak sayısını ve teşhir olan maddelerden yararlanan bireylerin yüzdesini değişimleyerek öncelikle hile yapan şüpheli bireylerin belirlendiği

ardından bu sınıflandırma ile teşhir olması muhtemel şüpheli maddelerin belirlendiği bir yaklaşım önermişlerdir. İlk aşamada, MTK modeline dayalı Bayesci sonsal log-odds oranı olan FLOR indeksini, bireyleri belirlemek için hesaplamışlardır. Bayesci sonsal log-odds FLOR indeksi ilk olarak birey düzeyinde hesaplanmakta ve madde tepkileri için bireyin dürüst olduğu ($\bar{s}$) ve bireyin teşhir olan maddelerden yararlandığı (s) model olmak üzere iki modele dayanmaktadır. Madde ön bilgisi olmadığında maddeye verilen tepkiler için 3PLM Eşitlik 2 ile gösterilmiştir.

$$p(u_i = 1|\bar{s}, \theta) = c_i + \frac{(1 - c_i)}{1 + e^{[-1.7a_i (\theta - b_i)]}} \quad (2)$$

Madde ön bilgisi olduğu durumda, her bir maddenin teşhir olma olasılığını da modelleyen maddeye verilen tepkiler için modifiye edilmiş 3PLM ise Eşitlik 3 ile gösterilmiştir (Birnbaum, 1968).

$$\begin{aligned} p(u_i = 1 | s, \theta) &= (1)p(m_i) + (1 - p(m_i))p(u_i = 1|\bar{m}, \theta) \\ &= p(m_i) + \left(c_i + \frac{(1 - c_i)}{1 + e^{[-1.7a_i (\theta - b_i)]}} \right) \\ &\quad - p(m_i) \left(c_i + \frac{(1 - c_i)}{1 + e^{[-1.7a_i (\theta - b_i)]}} \right) \end{aligned} \quad (3)$$

Eşitlik 3’de, $p(m_i)$ fonksiyonu, her bir maddenin teşhir olma olasılığını göstermektedir. Eşitlik 3’deki madde tepki modeli incelendiğinde, madde ön bilgisine sahip olan bireyin, maddeyi yanıtlama olasılığının 1 olarak ele alınmasıyla olasılık fonksiyonu Eşitlik 4 ile, ilk maddeden sonra bireyin teşhir olan maddelerin ön bilgisine sahip olup olmama durumu ile yeteneklerinin bağımsız olması varsayımı altında $p(s)$ Eşitlik 5 ile ifade edilmektedir.

$$p(s|u_1) = \frac{\int p(u_1|s, \theta)p(s)p(\theta)d(\theta)}{\int p(u_1 | s, \theta)p(s)p(\theta)d(\theta) + \int p(u_1|\bar{s}, \theta)p(\bar{s})p(\theta)d(\theta)} \quad (4)$$

$$p(s|u_1, u_2) = \frac{\int p(u_2|s, \theta)p(s, \theta|u_1)d(\theta)}{\int p(u_2 | s, \theta)p(s, \theta|u_1)d(\theta) + \int p(u_2|\bar{s}, \theta)p(\bar{s}, \theta|u_1)d(\theta)} \quad (5)$$

Böylelikle, n tane madde için teşhir olan maddelerden yararlanan bireylerin sonsal olasılık dağılımları Eşitlik 6 ile gösterilmektedir.

$$p(s|u_1, \dots, u_n) \propto \int p(u_n | s, \theta) p(s, \theta | u_1, \dots, u_{n-1}) d\theta \quad (6)$$

Daha yorumlanabilir bir indeks hesaplamak için Eşitlik 6 ile ifade edilen sonsal olasılık miktarları, testi alan her bir birey için log odds oranlarını hesaplamak için kullanılabilir. Eşitlik 7 ile gösterilen l_0 istatistiği, mevcut oranların (n tane madde tepkisi için), teşhir olan maddelerden yararlanan bireylerin önceki oranlara ($p(s)$) oranının 10 tabanındaki logaritmasını ifade etmektedir.

$$l_0 = \log_{10} \left[\frac{p(s|u_1, \dots, u_n)/[1 - p(s|u_1, \dots, u_n)]}{p(s)/[1 - p(s)]} \right] \quad (7)$$

Log odds oranının değerleri önsel olasılık değerleriyle karşılaştırıldığında, oranın negatif değerlerinin, madde ön bilgisinden faydalanma olasılığının düşük olduğunu, pozitif değerlerinin ise daha yüksek olduğunu gösterdiği görülmektedir (McLeod ve Schnipke, 2006). McLeod ve Schnipke (2006), çalışmaları sonucunda, teşhir olan maddeleri elde eden kaynakların yeterlik düzeyinin ve sayısının arttıkça bayesci sonsal log-odds *FLOR* indeksinin gücünün arttığını göstermiştir. McLeod, Lewis ve Thissen (2003) çalışmalarında, bireye uyarlanmış test uygulamalarında, teşhir olan maddelerin olasılığının nasıl modellendiğine (örneğin; teşhir olan maddelerin olasılığının sabit olarak ele alınması, güçlük düzeyine göre ağırlıklandırılması (zor maddelerin teşhir olma olasılığının yüksek olması vb.), madde seçim algoritması kullanılarak deneysel bir modelin uygulanması vb.) ve değişen gerçek puanlardaki farklılıkları belirlemek için kaynaklardan bir kaç sayıda ezberlemeye göre birkaç farklı *FLOR* indeksi önermekle birlikte hem şüpheli teşhir olan maddeleri hem de şüpheli teşhir olan maddelerden faydalanan bireyleri belirleyen log odds *FLOR*₃ indeksini kullanmışlardır. Testi alan her birey için *FLOR*₃ değerleri hesaplandıktan sonra, *FLOR*_i istatistiği, hangi maddelerin teşhir olduğunu belirlemek için madde düzeyinde hesaplanmıştır. Sınava giren bireyler, *FLOR*₃ değerlerinin sıfırdan büyük ve küçük olma durumlarında göre şüpheli ve şüpheli

olmayan olmak üzere iki farklı grupta sınıflandırılmıştır. $FLOR_i$ endeksi Eşitlik 8 ile gösterildiği şekilde hesaplanmaktadır.

$$FLOR_i = \log_{10} \left[\frac{p(u = 1|s)/[1 - p(u = 1|s)]}{p(u = 1|n)/[1 - p(u = 1|n)]} \right] \quad (8)$$

Eşitlik 8'de, $p(u = 1 | s)$; şüpheli gruptaki maddeyi doğru cevaplayan bireylerin oranını, $p(u = 1 | n)$; şüpheli olmayan gruptaki maddeyi doğru cevaplayan bireylerin oranını ifade etmektedir. $FLOR_i$ indeksinin sıfıra eşit olması şüpheli ve şüpheli olmayan gruplarda maddenin doğru cevaplanma oranının aynı olduğunu, indeks değerlerinin sıfırdan büyük değerleri ise şüpheli gruptaki bireylerin maddeyi cevaplama oranlarının şüpheli olmayan gruptakilere göre daha yüksek olduğunu göstermektedir.

Hui (2008) çalışmasında, bireye uyarlanmış testlerde beklenmedik tepki örüntülerinin belirlenmesinde, bireyleri, madde ön bilgisi olup olmama durumuna göre sınıflandırıp güvenli maddelerin sayısını, teşhir olan maddelerin sayısını ve test uzunluğunu değişimleyerek $FLOR$ indeksinin madde ön bilgisini belirlemede duyarlılığını, birey uyum indekslerinden I_z (Drasgow, Levine ve Williams, 1985) ve $CUSUM$ (Armstrong ve Shi, 2009) ile karşılaştırarak ortaya koymuştur. Çalışmanın sonucunda, $FLOR$ indeksinin, bireylerin teşhir olan bütün maddeleri doğru, güvenli olan bütün maddeleri yanlış cevapladıklarında en etkili teknik olduğu ve diğer indekslerle karşılaştırıldığında bireye uyarlanmış testlerde sabit uzunluk ve sabit maddeli testlerde $FLOR$ 'un daha duyarlı olduğu ortaya konmuştur.

Değişen Birey Fonksiyonu (Differential Person Functioning)

Smith ve Davis-Becker (2011), test uygulamasından önce teşhir olan maddelerden kaynaklı test hilelerini tespit etmek için değişen birey fonksiyonu analizini önermişlerdir. Değişen birey fonksiyonu, teste uygulamasına katılan bireyler ile testte yer alan sınıflandırılmış maddeler arasında etkileşim olduğu zaman gözlenmektedir. Böylelikle, değişen birey fonksiyonu analizleri, testi alan bireylerin tepki davranışlarının test edilmesinin bir yolu olarak görülmektedir. Değişen madde fonksiyonunun varlığı, geçerliğin önemli bir yönünü oluşturan ölçme değişmezliğinin ihlali olarak

görülmektedir. Değişen birey fonksiyonu analizini yapmanın farklı yolları vardır. Bunlardan bir tanesi madde ve birey parametrelerinin kestiriminde tüm birey ve maddelerin birlikte analiz edilmesidir. Böylelikle, madde parametreleri kestirilip değerleri sabitlenerek bireylerin yetenek düzeyleri her iki madde seti de kullanılarak tekrardan kestirilir. Değişen birey fonksiyonu analizlerinin hile yapan bireyleri belirleyebilmesi için temel varsayımı, testte yer alan maddelerin iki alt kümeye ayrılıp bir bireyin testteki maddelerin bir alt kümesi kullanılarak kestirilen yeteneğinin, diğer alt kümede yer alan maddeler kullanılarak kestirilen yeteneğine benzer olmasıdır. Bu yetenek düzeyleri arasında anlamlı düzeyde farklılık görüldüğü durumlarda bireyin potansiyel olarak uygun olmayan test yanıtlama davranışı gösterdiği yani hile yaptığı varsayılmaktadır. Smith ve Davis-Becker (2011) çalışmalarında, madde tepki modellerinden Rasch model kullanarak testteki maddeleri ön bilgi veya diğer uygun olmayan test yanıtlama davranışlarına eğilimli olup olmama durumuna göre iki şekilde sınıflandırarak değişen birey fonksiyonu analiz yöntemlerinin duyarlılığını ve kararlılığını incelemişlerdir. Çalışmada, sekiz güvenli madde ile değişen birey fonksiyonun 3'ten daha büyük değerleri için farklılaştığı ve işaretleme olasılıklarının 0.005'ten düşük olduğu değerler ile değişen birey fonksiyonu yaklaşımının, % 91 karar tutarlılığı, % 1.1 Tip I hata oranı ve % 7.9 Tip II hata oranı ile sonuçlanacağını belirtmişlerdir.

O'Leary ve Smith (2013), çalışmalarında, öncelikle madde ön bilgisinden yararlanan bireylerin belirlendiği ardından bu sınıflandırmayı kullanarak teşhir olma olasılığı olan maddelerin belirlendiği iki aşamalı bir yaklaşım önermişlerdir. Bu yöntemin kullanılabilmesi için test uygulayıcıları tarafından teşhir olmamış olan güvenli maddelerin testin içerisine yerleştirilmiş olması gerekmektedir. Ardından bireyler, değişen birey fonksiyonu analizlerine göre madde ön bilgisi durumlarına göre sınıflandırılıp her bir grup için teşhir olan maddeleri belirlemek için değişen madde fonksiyonu analizleri yapılır. Araştırmacılar, değişen birey ve madde fonksiyonu analizlerinin birlikte kullanılmasının madde ön bilgisi durumu aracılığıyla beklenmedik test puanlarından şüpheli hileci bireylerin belirlenmesi ve madde ön bilgisi durumuna göre sınıflandırılan bireylerin güvenli ve teşhir olan maddeleri belirlemesi aracılığıyla test programlarının güvenli bir şekilde sürdürülebilirliğinin sağlanacağını vurgulamışlardır.

Birey Uyum İstatistikleri

Birey uyum analizleri, bir bireyin, gözlenen tepki örüntüleri ile beklenen tepki örüntülerinin karşılaştırılmasını ve birey uyum istatistikleri ise bu gözlenen ve beklenen örüntüler arasındaki benzerliğin karakterize edilerek üretilmesini ifade etmektedir (Karabatsos, 2003; Meijer ve Tenderio, 2014). Birey uyum istatistiklerinin üretilmesindeki temel düşünce madde tepki örüntülerinin maddelerin güçlük dağılımını (örneğin; bireyin kolay maddeyi doğru yanıtlama olasılığının zor maddeden yüksek olması) yansıtmasıdır (Bishop ve Stephens, 2013). Birey uyum analizlerinin amacı, önerilen ölçme modeline göre bireysel test puanlarının uyum gösterip göstermediğini belirlemek ve bireysel test puan vektörlerini tanımlamaktır. Böylelikle, birey uyum istatistikleri, bireylerin test maddelerine verdikleri tepkilerden beklenmedik test performansını ortaya çıkararak bireyler hakkında önemli kararlar vermede, geçersiz test puanlarının belirlenip daha geçerli sonuçlara ulaşılmasında önemli rol oynamaktadır. (Emons, 2008; Meijer ve Sijtsma, 2001). Aşağıda ilgili literatürde yer alan ikili puanlanan maddeler için güncel olarak kullanılan parametrik ve parametrik olmayan birey uyum istatistiklerinin açıklamalarına yer verilecektir.

Parametrik Olmayan Birey Uyum İstatistikleri

Parametrik olmayan birey-uyum istatistikleri ile yapılan analizlerde, bireyin madde tepki örüntüsün aynı örneklemdaki başka bir bireyin madde tepki örüntüsü ile karşılaştırılmaktadır (Meijer ve Sijtsma, 2001).

Guttman'ın G indeksi

Parametrik olmayan ilk birey-uyum istatistiği olarak *G* istatistiği (Guttman, 1944) üretilmiştir. Guttman tepki örüntüleri, maddeler güçlük düzeylerine göre (maddeyi doğru cevaplama olasılıkları (doğru-oran (p) değerleri)) sıralandıktan sonra oluşturulmaktadır. Bu tepki örüntüleri, zor olan maddelerin doğru, kolay olan maddelerin yanlış cevaplandırılmaya meyilli olduğu durumlarda beklenmedik tepki olarak nitelendirilebilmektedir (Meijer ve Sijtsma, 2001). Guttman modelinde, bireyin tepki örüntüsünün “Guttman Mükemmel Örüntü”sü olarak adlandırılabilmesi için bireyin

sadece kolay maddeleri doğru cevaplandırması gerekmekte ve bu örüntüden sapmalar Eşitlik 9’da gösterildiği gibi G istatistiği olarak adlandırılmaktadır (Meijer, 1994).

$$G_n = \sum_{i=1}^{J-1} \sum_{j=i+1}^J X_{ni}(1 - X_{nj}) \quad (9)$$

Eşitlik 9’da, J ; madde sayısını, X_{ni} ve X_{nj} ise n bireyinin i . ve j . madde üzerindeki test puanlarını ifade etmektedir. Guttman Mükemmel Örüntüsü, G değeri 0 olduğunda elde edilir. Test uzunluğu arttıkça G değeri de artma eğiliminde olduğundan Eşitlik 10 ile gösterilen G^* istatistiği (van der Flier, 1977) önerilmiştir.

$$G_n^* = \frac{G_n}{r_n(J - r_n)} \quad (10)$$

Eşitlik 10’da, r_n ; n .bireyin J tane madde üzerindeki toplam test puanını göstermektedir. G^* istatistiği G istatistiğinin aralığını $[0,1]$ olarak normalleştirmektedir (Karabatsos, 2003).

Uyarı İndeksi (Caution Index, C, Modified Caution Index (MCI))

Korelasyon ya da kovaryansa dayalı parametrik olmayan indeksler, bireylerin maddelere ilişkin tepki vektörleri ile bireyin her bir maddeye ilişkin doğru cevaplarının oranı arasındaki ilişkiye dayanmaktadır. Sato (1975) tarafından önerilen Uyarı İndeksi C , bireyin madde puanları ile madde oran puanları ve bireyin en kolay maddelerdeki madde puanları ile madde oran puanları arasındaki kovaryans oranını inceleyerek beklenmedik tepki örüntüsüne sahip bireyleri Eşitlik 11 ile gösterildiği şekilde belirlemektedir.

$$C = 1 - \frac{Cov(X_n, p)}{Cov(X_n^*, p)} \quad (11)$$

Eşitlik 11’de, $X_n = (X_{n1}, X_{n2}, \dots, X_{Nj})$; n. bireyin madde tepki vektörünü, $p = (p_1, p_2, \dots, p_j)$; madde doğru cevap oran vektörünü, X_n^* ise n. bireyin en kolay maddelerdeki doğru cevaplarını içeren madde tepki vektörünü ifade etmektedir. İndeksin 0 değerinin Guttman Mükemmel Örüntüsünü gösterdiği böylelikle indeksin Guttman Mükemmel Örüntüsünden sapan tepki örüntülerinin derecesini ortaya koymayı amaçladığı görülmektedir (Meijer ve Sijtsma, 2001). Sato (1975) ve Huang (2012) çalışmalarında, C değerinin 0.50 üzeri olması durumunda, Karabatsos (2013) ise 0.53 olması durumunda tepki örüntülerinin beklenmeyen tepki örüntüsü olarak değerlendirilebileceğini belirtmişlerdir.

Harnisch ve Linn (1981) ise uyarı indeksini normalleştirerek bireylerin tepki örüntülerinin Guttman Mükemmel örüntüsünden sapma derecesini belirlemek için Eşitlik 12 ile gösterilen MCI (Modified Caution Index) üretmişlerdir.

$$MCI = \frac{Cov(X_n^*, p) - Cov(X_n, p)}{Cov((X_n^*, p) - Cov(X_n', p))} \quad (12)$$

Eşitlik 12’de, X_n' ; n. bireyin en zor maddelerdeki doğru cevaplarını içeren madde tepki vektörünü ifade etmektedir. MCI , Guttman Mükemmel Örüntü modelinin tepki vektörlerinin kovaryansı ile en zor maddeler için Guttman tepkilerine mükemmel zıtlıkta olan tepki vektörünün kovaryansı arasındaki oranı ifade etmektedir. Karabatsos (2003), uygun olmayan tepki örüntülerini belirlemede $MCI \geq 0.26$, Meijer ve Sijtsma (2001) ve Meijer ve Tendeiro (2014) ise $MCI \geq 0.30$ kritik değerini kullanmayı önermişlerdir.

U3 ve ZU3 İndeksi

van der Flier (1980), MCI için kovaryans yerine logaritma kullanan Eşitlik 13 ile gösterilen $U3$ ve bu indeksin standardize edildiği Eşitlik 14 ile gösterilen $ZU3$ indeksini önermiştir.

$$U3 = \frac{\ln(X_n^*) - \ln(X_n)}{\ln(X_n^*) - \ln(X_n')} , \quad (13)$$

$$\ln(X_n) = \sum_{j=1}^J \left[X_{nj} \log\left(\frac{p_j}{1-p_j}\right) \right]$$

$$\ln(X_n^*) = \sum_{j=1}^{r_n} \left[\log\left(\frac{p_j}{1-p_j}\right) \right]$$

$$\ln(X_n') = \sum_{j=r_n+1}^J \left[\log\left(\frac{p_j}{1-p_j}\right) \right]$$

$$ZU3 = \frac{U3 - E(U3)}{\sqrt{Var(U3)}} \quad (14)$$

Eşitlik 13’de, p_j ; n.bireyin j maddesi için doğru cevaplama olasılığını, Eşitlik 14’de $E(U3)$; $U3$ için beklenen değeri ve $Var(U3)$; $U3$ için varyans değerini ifade etmektedir.

Sijtsma’nun H^T İndeksi

Sijtsma (1986) ve Sijtsma ve Meijer (1992) tarafından bireyin tepki vektörü ile diğer bireylerin tepki vektörleri arasındaki uyuşmayı inceleyen Eşitlik 15 ile gösterilen H_n^T istatistiği önerilmiştir.

$$H_n^T = \frac{\sum_{m \neq n} (\beta_{nm} - \beta_n \beta_m)}{\sum_{m \neq n} \min[\beta_n(1 - \beta_m), (1 - \beta_n)\beta_m]} \quad (15)$$

Eşitlik 15’de, β_n ; yerel bağımsız olan tekrarlı ölçümlerde J sabit maddeler dizisi için n. bireyinin maddelere verdiği doğru cevapların beklenen oranını, β_m ; J sabit maddeler dizisi için m. bireyinin maddelere verdiği doğru cevapların beklenen oranını, β_{nm} ise m ve n bireylerinin puanları arasındaki kovaryansı ifade etmektedir. m ve n bireyine ait tepki örüntüleri arasındaki kovaryansların her biri maksimum değerini aldığı anda H_n^T istatistiği de maksimum değeri olan 1’i almaktadır (Meijer ve Sijtsma,

2001). Karabatsos (2003) çalışmasında, $H_n^T \leq 0.22$ kritik değerinde, uygun olmayan tepki örüntüleri sergileyen bireylerin en iyi şekilde belirlendiğini belirtmiştir.

Parametrik Birey Uyum İstatistikleri

Madde Tepki Kuramına dayalı birey-uyum istatistikleri; Rasch modeli, 2PLM ve 3PLM'ye göre farklılaşmakta olup artıklara dayalı ve olabilirliğe dayalı olmak üzere sınıflandırılmıştır. Artıklara dayalı birey uyum istatistikleri genellikle gözlenen ve madde tepki kuramı modellerinden yordanan madde tepki örüntüleri arasındaki farklar üzerine dayalı iken, olabilirliğe dayalı birey uyum istatistikleri beklenen tepkileri olabilirlik fonksiyonunun logaritmasını (log-olabilirlik fonksiyonu) maksimum yapan yetenek düzeyinin kestirimi ile belirleyerek uygun olmayan tepki örüntülerini ortaya koymaya çalışmaktadır (Meijer ve Sijtsma ,2001).

Wright'ın Ağırlıklandırılmış (W) ve Ağırlıklandırılmamış (U) Birey Uyum İstatistikleri

Wright ve Stone (1979) tarafından önerilen U istatistiği (uyum dışı ortalama kare (Outfit Mean Square)) standardize edilmiş artıkların karelerine dayanırken, Wright (1980) tarafından önerilen W istatistiği (uyum içi ortalama kare (Infit Mean Square, IMS) varyans-artıkların karelerinin ağırlıklandırılmış ortalaması olarak Rasch modele dayalı olarak Eşitlik 16 ve Eşitlik 17'de gösterildiği gibi hesaplanmaktadır.

$$U = \frac{1}{J} \sum_{j=1}^J \frac{(X_{nj} - P_{nj1})^2}{P_{nj1}P_{nj0}} \quad (16)$$

$$W = \frac{\sum_{j=1}^J ((X_{nj} - P_{nj1})^2)}{\sum_{j=1}^J P_{nj1}P_{nj0}} \quad (17)$$

U istatistiği, J tane maddeye dayalı standardize edilmiş artıkların karelerinin ortalaması şeklinde yorumlanmaktadır (Meijer ve Sijtsma, 2001). W istatistiğinin, bireyin yeteneğinden farklı güçlüğüne sahip bir maddeye verilen beklenmeyen tepkilere U

istatistiğine göre daha az duyarlı olduğu görülmektedir (Wright ve Stone,1979). Wright ve Panchapakesan (1969) ve Wu ve Adam (2007) çalışmalarında, her iki istatistiğin 1’den büyük değerleri için tepki örüntülerinin modelle daha iyi uyum gösterdiği, 1’den küçük değerler için tepki örüntülerinin modelden kestirilen tepki örüntülerinden farklılaştığını belirtmişlerdir. Keeves ve Alagumalai (1999) ve Linacre (2002) çalışmalarında, U ve W birey uyum istatistiklerinin 0.7 ile 1.30 aralığındaki değerlerinde tepki örüntüleri ile Rasch modelden kestirilen tepki örüntülerinin iyi uyum sağladığını, Karabatsos (2003) ise çalışmasında, bu istatistikler için 1.3 değerinin minimum kritik değer olduğunu vurgulamıştır.

Trabin ve Weiss’ ın $D(\theta)$ Birey Uyum İstatistiği

Trabin ve Weiss (1983), uygun olmayan madde tepki örüntülerini belirleyebilmek için sabit yetenek değeri için maddeyi doğru cevaplama olasılığını, madde konumunun fonksiyonu olarak tanımlandığı birey tepki fonksiyonunu önermiştir. Gözlenen birey tepki fonksiyonunu belirlemek için testteki maddeler yer parametre δ değerlerine göre gruplandırılarak maddelerin alt kümeleri oluşturularak her alt kümede yer alan maddelere ilişkin doğru cevap oranları belirlenerek yetenek parametresi kestirililmektedir. 3PLM’ye göre doğru cevap olasılıklarının ortalaması kestirilerek beklenen birey tepki fonksiyonu oluşturulur. Böylelikle, gözlenen ve beklenen birey tepki fonksiyonu arasındaki fark hesaplanarak elde edilen madde tepkilerinin model ile uyuşma derecesini veren Wright ve Stone (1979) tarafından önerilen ağırlıklandırılmamış $D(\theta)$ birey uyum istatistiği Eşitlik 18 ile gösterildiği şekilde hesaplanmaktadır.

$$D(\theta) = \sum_{s=1}^S \frac{1}{J_s} \left(\sum_{j \in s} [X_{nj} - P_{nj1}]^2 \right) \quad (18)$$

Eşitlik 18’de, S ; testteki toplam madde alt küme sayısını, J_s ; s alt kümesinde yer alan madde sayısını ifade etmektedir. $D(\theta)$, belirli bir θ değeri için her bir alt testte beklenen ve gözlenen birey tepki fonksiyonu arasındaki farkın, madde sayısına oranının her alt test için toplamını ifade etmektedir (Meijer ve Sijtsma, 2001).

l_0 , l_z , l_z^* ve l_{zm} Birey Uyum İstatistikleri

Madde Tepki Kuramı modellemesinde, bir grup bireyden ve madde tepki modelinden elde edilen tepki örüntüleri göz önüne alındığında, hem birey hem de madde parametrelerinin tahmin edilmesi gerekmekte ve bu süreç olabilirlik fonksiyonunun kullanımını içermektedir. Olabilirliğe dayalı birey-uyum indekslerinin tümü log-olabilirlik fonksiyonuna bağlıdır. Olabilirlik fonksiyonunun logaritmasını maksimum yapan yetenek düzeyinin kestirimi (MLE) ile beklenen tepkiler belirlenerek beklenmeyen madde tepki örüntüleri ortaya konulmaya çalışılmaktadır. Levine ve Rubin (1979) tarafından birey-uyum indeksi olarak önerilen l_0 log-olabilirlik fonksiyonu Eşitlik 19 ile gösterilmiştir.

$$l_0 = \ln[L_n(\theta_n)] = \sum_{j=1}^J [X_{nj} \ln P_{nj1}(\theta_n) + (1 - X_{nj}) \ln P_{nj0}(\theta_n)] \quad (19)$$

Molenaar ve Hoijtink (1990), 1PL model ve Rasch modeli altında l_0 istatistiği yerine hesaplama kolaylığı açısından Eşitlik 20'deki gibi ifade edilen M istatistiğinin kullanımını önermiştir

$$l_0 = \sum_{j=1}^J \ln[1 + \exp(\theta_n - \delta_j)] + r\theta_n - \sum_{j=1}^J X_{nj} \delta_j ,$$

$$M = - \sum_{j=1}^J X_{nj} \delta_j \quad (20)$$

Bununla birlikte, l_0 birey uyum istatistiğinin, standardize olmaması sınırlılığından dolayı Drasgow, Levine ve Williams (1985), standartlaştırılmış hali olan gerçek yetenek düzeyleri kullanıldığında asimptotik standart normal dağılıma sahip olan Eşitlik 21 ile ifade edilen l_z istatistiğini önermiştir.

$$l_z = \frac{l_0 - E(l_0)}{\sqrt{Var(l_0)}} = \frac{l_0 - \sum_{j=1}^J [P_{nj1} \ln P_{nj1} + (1 - P_{nj1}) \ln(1 - P_{nj1})]}{\sqrt{\sum_{j=1}^J P_{nj1} (1 - P_{nj1}) [\ln(P_{nj1}/(1 - P_{nj1}))]^2}} \quad (21)$$

Eğitimsel ölçmelerde, birey uyum istatistikleri ile yapılan analizlerde sıklıkla kullanılan ve uygun olmayan madde tepki örüntülerini belirleme gücü yüksek olan l_z birey-uyum istatistiği, l_0 birey-uyum istatistiğinin standardize edilmiş log olasılığıdır. Standart normal dağılım göstermesinden kaynaklı l_z istatistiğinin yüksek negatif değerleri (-2'den küçük) bireyin tepki örüntülerinin uygun olmadığı yönünde yorumlanırken, yüksek pozitif değerleri ise (+2'den büyük) tepkilerin model ile oldukça iyi bir şekilde uyum gösterdiği yönünde yorumlanabilmektedir (Karabatsos, 2003; Dimitrov ve Smith, 2006). Meijer ve Sijtsma (2001), l_z istatistiğinin -1.65 değerinden küçük değerler için birey tepkilerinin uygun olmadığı yönünde değerlendirilebileceğini belirtmiştir. Karabatsos (2003) ve Wollack (1997) çalışmalarında l_z istatistiğinin performansının iyileştirilmesi gerektiği sonucuna ulaşırken de la Torre ve Deng (2008) çalışmalarında, bu istatistiğin iyi performans gösterdiği sonucuna ulaşmışlardır.

Snijders (2001), kestirilen yetenek düzeyi kullanıldığında l_z istatistiği yerine asimtotik standart normal dağılıma sahip olan Eşitlik 22 ile ifade edilen l_z^* istatistiğini önermiştir.

$$l_z^* = \frac{l_0(\hat{\theta}) - E(l_0(\hat{\theta})) + c_J(\hat{\theta})r_0(\hat{\theta})}{\sqrt{\tilde{V}(l_0(\hat{\theta}))}} \quad (22)$$

Sinharay (2015a) çalışmasında, l_z^* birey uyum istatistiğinin performansının geliştirilmesine yönelik yöntemler önermiştir. Drasgow, Levine ve McLaughlin (1991) ise çalışmalarında, alt testlerden oluşan tek boyutlu testler için alt testlerin varyanslarını hesaplamaya dahil eden l_z 'nin genelleştirilmiş hali olan l_{zm} istatistiğini önermişlerdir.

Modifiye Edilmiş Uyarı İndeksleri

Tatsuoka (1984), parametrik olmayan uyarı indeksini (C), MTK modellerine genişleterek $ECI1$, $ECI2$, $ECI3$, $ECI4$, $ECI5$ ve $ECI6$ indekslerini geliştirmiştir. Bu indeksler bireyin tepki vektörü (X), maddenin p değerlerinin vektörü (P) ya da bireyler boyunca bir maddenin doğru cevabın ortalama olasılığına ait vektörün (G) kovaryansına

ya da korelasyonuna bağılı olarak türetilmişlerdir. Genişletilmiş ilk indeks *ECI1* indeksi Eşitlik 23 ile gösterilmiştir.

$$ECI1 = 1 - \frac{Cov(X_n, p)}{Cov(P_n, p)} \quad (23)$$

ECI1 indeksi, bireyin puan vektörü ile bireyler tarafından doğru cevaplanan madde puan vektörü arasındaki kovaryansın, MTK modelleri altındaki beklenen doğru cevaplama olasılık vektörü ile bireyler tarafından doğru cevaplanan madde puan vektörü arasındaki kovaryansa oranını göstermektedir. *ECI2* indeksi ise bireyler tarafından bir maddeyi doğru cevaplama olasılığının ortalama vektörü ile madde-puan vektörü arasındaki kovaryansın, MTK modelleri altındaki beklenen doğru cevaplama olasılık vektörü ile bireyler tarafından bir maddeyi doğru cevaplama olasılığının ortalama vektörü arasındaki kovaryansa oranıdır (Meijer ve Sijtsma, 2001). *ECI2* indeksi Eşitlik 24 ile gösterilmiştir.

$$ECI2 = 1 - \frac{Cov(X_n, \bar{G})}{Cov(P_n, \bar{G})} \quad (24)$$

ECI3 indeksinde ise *ECI2* indeksinden farklı olarak hesaplamada, kovaryans yerine korelasyon kullanılmaktadır. *ECI3* indeksi, bireyler boyunca bir maddenin doğru cevaplanma olasılığının ortalaması ile madde-puan vektörü arasındaki korelasyona eşittir (Meijer ve Sijtsma, 2001). *ECI3* indeksi Eşitlik 25 ile gösterilmiştir.

$$ECI3 = 1 - \frac{Corr(X_n, G)}{Corr[P(\theta), G]} \quad (25)$$

ECI4, *ECI5* ve *ECI6* indeksleri, bireyin tepki vektörü (*X*), maddenin *p* değerlerinin vektörü (*P*) arasındaki kovaryans ya da korelasyon hesaplanarak üretilmişlerdir. Bu indeksler, Eşitlik 26, 27 ve 28 ile gösterilmiştir.

$$ECI4 = 1 - \frac{Cov(X_n, P_n)}{Cov(P_n, \bar{G})} \quad (26)$$

$$ECI5 = 1 - \frac{Corr[X_n, P(\theta)]}{Corr[G, P(\theta)]} \quad (27)$$

$$ECI6 = 1 - \frac{Cov(X_n, P_n)}{Cov(P_n, P_n)} \quad (28)$$

Tatsuoka (1984), *ECI6* indeksinin birey yeteneğinin bir fonksiyonu olmasından dolayı diğer indekslerin beklenen değerlerinin 0' a eşit olması durumunun oluştuğunu, buna karşılık uygun olmayan tepki örüntülerinin daha yorumlanabilir olması için indeklerin genel bir ifade ile standart normal dağılıma dönüştürülmesini önermiştir.

Parametrik Birikimli Toplam İşlemi (cumulative sum procedure-CUSUM)

Bradlow Weiss ve Cho (1998), van Krimpen-Stoop ve Meijer (2000, 2001), Egberink ve diğerleri (2010) ve Meijer (2002), bireye uyarlanmış testlerdeki uygun olmayan tepki örüntülerini belirlemek üzere birikimli toplam prosedürünü (CUSUM) önermişlerdir. Bu istatistik, testin sonlarında kestirilen yetenek düzeyinin bireyin gerçek yetenek düzeyine yaklaştığı durumlarda, maddelere verilen ardışık doğru ve yanlış tepkilerden oluşan madde puan örüntüsünün madde tepki modeli ile uyumunu ortaya koymaktadır (Meijer, 2002). Testteki her i maddesi için pozitif veya negatif artıkların toplamına birikimli toplam prosedürünün uygulanmasında öncelikle uyumsuzluğun ölçüsü olarak artıkların ağırlıklandırılmış değeri olan Eşitlik 29 ile ifade edilen T istatistiği hesaplanmaktadır.

$$T_i = \frac{1}{J} [x_i - p_i(\theta)] \quad (29)$$

Eşitlik 29'da, J ; testteki toplam madde sayısını, $x_i - P_i(\theta)$; pozitif veya negatif artıkları ifade etmektedir. T istatistiği olarak Eşitlik 30 ve Eşitlik 31'de ifade edildiği gibi C_i^+ pozitif ardışık artıkların toplamı ile C_i^- negatif ardışık artıkların toplamı olmak üzere iki tane birikimli istatistik tanımlanmaktadır (Meijer, 2002, Meijer ve Sijtsma, 2001).

$$C_i^+ = \max[0, T_i + C_{i-1}^+] \quad (30)$$

$$C_i^- = \min[0, T_i + C_{i-1}^-] \quad (31)$$

Pozitif ardışık artıkların toplamı üst sınırdan büyük, negatif ardışık artıkların toplamı ise alt sınırdan küçük olduğunda bireyin tepki örüntüsünün uygun olmayan tepki örüntüsü olduğu söylenebilmektedir (Meijer ve Sijtsma, 2001). Uygun olmayan tepki örüntülerinin belirlenmesinde, Meijer (2002)'in çalışmasında veri setinden teorik olarak elde edilen alt ve üst sınırlar kullanılabileceği gibi bu sınırları belirlemek için simülasyon çalışmaları da uygulanabilir (van Krimpen-Stoop ve Meijer, 2002). Genel olarak, birey uyum istatistikleri ile uygun olmayan tepki örüntüleri tespit edilebilse de madde sıralamalarının dikkate alınmadığı görülmektedir. Böylelikle, CUSUM istatistiğinin, bireyin tepkisinin hangi madde ile değiştiğine dair bilgi elde edilmesinde önemli rol oynadığı görülmektedir (Armstrong ve Shi, 2009; Kim, Woo ve Dickison, 2017).

Teşhir Olan Maddeleri Belirlemeye Yönelik Yöntemler

Hareketli Ortalamalar Yöntemi (Moving Averages)

Han (2003), teşhir olan maddelerin belirlenmesine yönelik olarak madde güçlük tahminlerinin kullanılması yerine "p değerlerinin" (her madde için doğru cevap oranı) dikkate alındığı hareketli ortalamalar yöntemini kullanmayı önermiştir. Bu yöntemle, madde güçlüklerindeki zaman içerisindeki değişimler incelenerek madde ön bilgisine sahip olma durumundaki tepki örüntüleri belirlenmeye çalışılmaktadır. Hareketli ortalamalar, n tane bireyin madde tepki verilerinin analizi için bir pencere boyutu (k) seçilerek hesaplanır. k=100 için i maddesinin hareketli p değerlerinin dizisi Eşitlik 32 ile gösterilmiştir.

$$(p_{100}, p_{101}, \dots, p_{n-100}),$$

$$p_{100} = \frac{1}{k}(u_{i,1} + u_{i,2} + \dots + u_{i,100})$$

$$p_{101} = \frac{1}{k}(u_{i,2} + u_{i,3} + \dots + u_{i,101})$$

$$p_{102} = \frac{1}{k}(u_{i,3} + u_{i,4} + \dots + u_{i,102})$$

...

$$p_{n-100} = \frac{1}{k}(u_{i,n-99} + u_{i,n-98} + \dots + u_{i,n}) \quad (32)$$

Ortalama p-değerini izlemek yerine p-değerlerinin hareketli ortalamalarını izlemenin amacı, ortalama p değerinin değişmeye başladığı noktanın belirlenmesi ve böylelikle, testte, teşhir olmuş maddelerin işaretlenmesini sağlamaktır (Han ve Hambleton, 2004). Han ve Hambleton (2004) çalışmalarında, hareketli ortalamaların kullanımını, hareketli p değerleri, madde artıklarının hareketli ortalamaları ve standartlaştırılmış madde artıklarının hareketli ortalamaları olmak üzere üç istatistiğin performanslarını çeşitli koşullar altında karşılaştırarak incelemişlerdir. Çalışmanın sonucunda, madde artıklarının ve standartlaştırılmış madde artıklarının hareketli ortalamalarının, teşhir olan maddeler çok zor maddeler olmadığında ve teşhir olan madde sayısı az olduğunda iyi performansa sahip olduğu belirtilmiştir.

Log-Odd Oranlar İstatistiği (Log Odds Ratio Statistic)

Obregon (2013), McLeod ve Schnipke (2006)'nin $FLOR_i$ indeksine benzer şekilde, Rasch model çerçevesinde, bireylerin madde düzeyinde log odd olasılık değerlerini kullanarak teşhir olan maddeleri belirlemeye yönelik log odd oranlar indeksini önermiştir. *Log-odd oranlar* istatistiğinin hesaplanmasında, madde tepki vektörü X_i olan teşhir olan i maddesi için sonsal olasılığın hesaplanmasında Eşitlik 33 ile gösterilen Bayes teoremi kullanılır.

$$p(c|X_i) = \frac{p(c)p(X_i|c)}{p(c)p(X_i|c) + p(\bar{c})p(X_i|\bar{c})} \quad (33)$$

Eşitlik 33’de, $p(c)$; teşhir olan i maddesinin önsel olasılık değerini, $p(\bar{c})$; i maddesinin teşhir olmadığı durumda önsel olasılık değerini ($1 - p(c)$), $p(X_i|c)$; madde tepki vektörü X_i olan teşhir olan i maddesi için olabilirlik değerini, $p(X_i|\bar{c})$ ise madde tepki vektörü X_i olan teşhir olmayan i maddesi için olabilirlik değerini ifade etmektedir. Birey ve madde parametrelerinin kestiriminde Rasch model kullanıldığı durumda teşhir olan madde için sonsal olasılık değeri Eşitlik 34’de gösterildiği gibi ifade edilmektedir.

$$p(c|X_i) = \frac{p(c) \prod_{j=1}^n p(x_{ji}|c, \hat{\theta}_j)}{p(c) \prod_{j=1}^n p(x_{ji}|c, \hat{\theta}_j) + p(\bar{c}) \prod_{j=1}^n p(x_{ji}|\bar{c}, \hat{\theta}_j)} \quad (34)$$

Eşitlik 34’de, $p(x_{ji}|c, \hat{\theta}_j)$; j.bireyin teşhir olan i maddesine verdiği x_{ji} tepkisinin olasılık değerini, $p(x_{ji}|\bar{c}, \hat{\theta}_j)$ ise j.bireyin teşhir olmayan i maddesine verdiği x_{ji} tepkisinin olasılık değerini ifade etmektedir. Teşhir olmayan madde için sonsal olasılık değeri, Eşitlik 35, teşhir olan madde için sonsal olasılık değeri, koşullu madde tepki fonksiyonu kullanılarak Eşitlik 36 ile gösterilmektedir.

$$p(x_{ji}|\bar{c}, \hat{\theta}_j) = \frac{\exp[x_{ji}(\hat{\theta}_j - b_i)]}{1 + \exp(\hat{\theta}_j - b_i)} \quad (35)$$

$$p(x_{ji}|c, \hat{\theta}_j) = p(E_j) \left(\frac{\exp[x_{ji}(\hat{\theta}_j + \hat{D} - b_i)]}{1 + \exp(\hat{\theta}_j + \hat{D} - b_i)} \right) + [1 - \hat{p}(E_j)] \left(\frac{\exp[x_{ji}(\hat{\theta}_j - b_i)]}{1 + \exp(\hat{\theta}_j - b_i)} \right) \quad (36)$$

Eşitlik 36’da, $\hat{p}(E_j)$; madde ön bilgisine sahip olan j.bireyin i maddesi için kestirilen olasılık değerini ve maddenin teşhir olma durumunda madde güçlük düzeyindeki beklenen düşüşü (beklenen logit kayma) ifade etmektedir. İstatistiğin log odds oranı ise Eşitlik 37 ile ifade edilmektedir.

$$l_0 = \log_{10} \left[\frac{p(c|X_i)/[1 - p(c|X_i)]}{p(c)/[1 - p(c)]} \right] \quad (37)$$

Obregon (2013) çalışmasında, madde ön bilgisine sahip bireylerin oranını (%5, %10,%20) ve beklenen logit kayma oranını (1.0,2.0,3.0) değişimleyerek l_0 istatistiğinin performansını araştırmıştır. Çalışmanın sonucunda, beklenen logit kayma oranının 1.0 olduğu durumda istatistiğinin performansının iyi olduğunu belirtmiştir. Eckerly ve diğerleri (2015) ise, Obregon (2013)'un çalışmasını, örneklem büyüklüğü ve bireylerin yetenek dağılımı faktörlerini de değişimleyerek bu faktörlerin madde ön bilgisi varlığı durumunda l_0 istatistiğinin dağılımını etkileyebileceği sonucuna ulaşmışlardır.

Teşhir Olan Maddelerden Yararlanan Bireylerin Gruplarının Belirlendiği Yöntemler

Madde ön bilgisine dayalı hile belirleme çalışmalarında, teşhir olan maddelerden yararlanan bireylerin gruplarının belirlenmesinde genel olarak birey uyum istatistiklerine dayalı olan kümeleme analizinin , Kullback-Leibler uzaklığının ve faktör analizinin yaygın olarak kullanıldığı görülmektedir. Aşağıda güncel olarak kullanılan bu yöntemlerin açıklamalarına kısaca yer verilmiştir.

Kümeleme Analizi

Kümeleme analizi, veri kümesindeki bilgileri, belirli uzaklık ölçüleri kullanarak kümelere ayırma işlemidir. Küme sayısının bilinmediği hiyerarşik kümeleme analizinde, iş birliği yapan grupları belirlemede, veri kümesindeki bilgiler, belirli yakınlık derecelerine (tek bağlantı (en yakın komşuluk), tam bağlantı (en uzak komşuluk), ortalama bağlantı, merkezi ve Ward yöntemleri, bölünmüş ortalamalar ve otomatik etkileşim belirleme) göre kümelere ayrılırlar (Çokluk, Şekercioğlu ve Büyüköztürk, 2014). Wollack ve Maynes (2011, 2017) çalışmalarında, hile yapan bireylerin gruplarının belirlenmesinde, testi alan bireylerin madde tepkilerinin benzerliğine dayalı olarak kümelerin incelendiği yaklaşım kullanmışlardır. Bu yöntem, en yakın komşu kümelemeyi, testi alan iki bireyin tepkilerinin benzerliğinin derecesini karakterize eden benzerlik indeksi ile oluşturmaktadır. Çalışmada kullanılan yaklaşımda, öncelikle, Maynes (2005) tarafından önerilen M_4 indeksi kullanılarak testi alan bireylerin madde tepkileri arasında beklenmeyen benzerlik olduğu için işaretlenen bireyler için eşik değeri belirlenerek işaretlenen bu bireyler, en yakın komşu kümeleme yöntemine göre gruplara

atanmışlardır. En yakın komşu kümeleme yönteminde öncelikle M_4 indeksi kullanarak oluşturulan uzaklık matrisindeki en yakın iki gözlem alınarak ilk küme oluşturulmaktadır. Ardından, aşamalı olarak kümeler oluşturulmaya devam edilerek süreç sonunda tüm gözlemler bir kümeye dahil edilmektedir. Böylelikle, en yakın komşu kümeleme yöntemi adımları şu şekilde sıralanabilir:

- 1) Her bir grup için etkileşim miktarı (işbirliği parametresi; $\emptyset_g$) ve teşhir olan madde sayısı (m) random olarak belirlenir.
- 2) G grup büyüklüğünde, ortalaması 0 standart sapması $\frac{1}{G}$ olan dağılımdan $N(0, \frac{1}{G})$ ortalama yetenek Θ_g bağımsız örneklenir. Her madde için, kaynak olan s bireyinin her madde için, maddedeki her bir cevap alternatifini seçme olasılığı (π_{si}) ayrı hesaplanır.
- 3) g grubundaki j bireyinin yetenek parametresi $\theta_{j(g)}$, $N(\Theta_g, 1)$ dağılımından elde edilir. Her madde için, j bireyinin, maddedeki her bir cevap alternatifini seçme olasılığı (π_{ji}) ayrı hesaplanır.
- 4) π_{ji} , π_{si} , $\emptyset_g$ kullanılarak j ve s bireylerinin her bir madde için aynı cevabı verme olasılığı π_{jsi} hesaplanarak AxA tablosu elde edilir.
- 5) $\theta_s = \emptyset_g$ ve i. maddedeki her bir seçeneği seçme olasılığı (π_{si}) kullanılarak grup için amaçlanan madde tepki vektörü ($U_g = (u_{g1}, \dots, u_{gi}, \dots, u_{gn})$) oluşturulur. Bu vektör, g grubundaki genel bilgiyi verir.
- 6) U_g vektörü kullanılarak, $U_{si} = u_{gi}$ (i. madde için amaçlanan cevap) için AxA tablosundan koşullu olasılıklar seçilir.

Kullback-Leibler Uzaklığı Yaklaşımı

Bireye uyarlanmış test uygulamalarında, madde seçim yöntemi olarak kullanılan türev almayı gerektirmeyen bilgi fonksiyonu Kullback-Leibler uzaklığının, Belov, Pashley, Lewis ve Armstrong (2007), birey uyum istatistiği olarak, Belov ve Armstrong (2010) ise cevap kopyalamadan kaynaklı test hilesini belirlemek için çeşitli uygulamalar ile kullanımını geliştirmişlerdir. Kullback Leibler uzaklığı (KLD) iki dağılım arasındaki benzerliği ölçmek için yaygın olarak kullanılmaktadır. Belov ve Armstrong (2011) çalışmasında, Kullback-Leibler uzaklığının asimtotik dağılımını kullanarak test

güvenliğine ilişkin uygulamalara yer vermiştir. Belov (2013, 2014, 2016), geniş ölçekli sınavlarda, maddelerin cevaplarını içeren test materyallerinin paylaşılmasından kaynaklı olarak gözlenen uygun olmayan tepki örüntülerini belirlemek için Kullback-Leibler ayrışma indeksinin kullanılmasını önermiştir. Bu yaklaşım iki aşamalı uygulanmaktadır:

1.Aşama: C_s bütün test merkezlerini, c ise bir test merkezini ifade etsin. ($c \in C_s$). Olağandışı dağılıma sahip test merkezleri, g_c birey-uyum istatistiği kullanılarak belirlenir. Eşitlik 38 ile ifade edilen g_c birey-uyum istatistiği, c test merkezinin dağılımının, birey-uyum istatistiğinin dağılımı ile diğer test merkezlerinin dağılımı arasındaki farklılığın ölçüsünü göstermektedir.

$$g_c = \sum_{x \in C} [D(H_c || H_x) + D(H_x || H_c)] \quad (38)$$

Eşitlik 38’de, H_c ; birey-uyum istatistiğinin deneysel dağılımını, $D(H_c || H_x)$ ise; k değerli sonlu kümede $(d1, d2, d3, \dots, dk)$ tanımlanmış iki ayrı H_c ve H_x dağılımı arasındaki Eşitlik 39 ile hesaplanan Kullback Leibler uzaklığını ifade etmektedir.

$$D(H_c || H_x) = \sum_{k=1}^k H_c(dk) \ln \frac{H_c(dk)}{H_x(dk)} \quad (39)$$

Ayrık ve sürekli dağılımlar için geçerli olan Kullback Leibler yaklaşımına göre ayrışma ne kadar büyükse dağılımlar arasındaki farklılık o kadar fazladır. $D(H_c || H_x) \geq 0$ değerleri, iki dağılımın özdeş olduğunu göstermektedir. Ayrıca Kullback Leibler yaklaşımı asimetriktir; ($D(H_c || H_x) \neq D(H_x || H_c)$).

2.Aşama: Belirlenen test merkezlerindeki beklenmeyen madde tepki örüntüsüne sahip olan bireyle, benzerlik ve birey-uyum istatistikleri kullanılarak belirlenir (Belov, 2013; Kullback ve Leibler, 1951; Cizek ve Wollack, 2017). Belov (2012, 2013, 2015), sınav merkezi ve eğitim programları gibi grup mekanizmaları dahilinde madde ön bilgisine sahip bireyleri belirlemek için çok aşamalı yaklaşım önermiştir. Belov (2013)’un yönteminin farklı varyasyonları üç farklı durum için kullanılabilir:

- 1) Madde ön bilgisine sahip olan her grup, teşhir olan aynı madde setlerini almıştır ve bu gruplar bilinmektedir.

- 2) Madde ön bilgisine sahip olan her grup, teşhir olan farklı madde setlerini almıştır ve bu gruplar bilinmektedir.
- 3) Teşhir olan maddeler bilinmemektedir.

Kullback-Leibler uzaklığı, grupların önsel dağılımının bilindiği kümeleme analizi ile benzer süreçleri içermektedir.

Faktör Analizi

Faktör analizi, aralarında ilişki bulunan p değişkenle açıklanan yapıyı, kümeleme analizinde kullanılan veriler arası uzaklık ölçüleri yerine varyans-kovaryans bilgilerini kullanarak, aralarında ilişki bulunmayan fakat kendi içinde ilişkili daha az sayıdaki ($k < p$) yeni değişkenle (faktör) açıklamaktır (Alpar, 2013). Faktör analizinin, açıklayıcı faktör analizi (AFA), doğrulayıcı faktör analizi (DFA), R tipi, Q tipi, O tipi, T tipi ve S tipi faktör analizi olmak üzere çeşitli türler vardır. İş birliği yapan bireyler arasındaki ilişkiyi ortaya koymada bu analizden faydalanıldığı görülmektedir.

Zhang, Searcy ve Horn (2011) çalışmalarında, farklı gruplarda yer alan uygun olmayan madde tepki örüntülerine sahip olan bireyleri, Rasch model için geliştirilen U birey uyum istatistiği kullanarak belirleyerek bu bireyleri kümelemek için Q tipi faktör analizini kullanmışlardır. Bu yaklaşım, madde ön bilgisine sahip olma durumundan kaynaklı uygun olmayan madde tepki örüntülerine sahip bireylerin gruplar arasında kümelendiği süreçleri içermektedir. Çalışmanın sonucunda, teşhir olan maddelerin ve bu maddelerden faydalanan bireylerin sayısının yöntemin performansını etkilediği görülmüştür.

Teşhir Olan Maddelerin ve Bu Maddelerin Ön Bilgisinden Yararlanan Bireylerin Birlikte Belirlendiği Yöntemler

Aşağıda madde ön bilgisine sahip olan bireyleri belirlemede kullanılan madde tepki süresi modelleri kısaca açıklanacaktır.

Loglineer Madde Tepki Süresi Modeli

van der Linden, Scrams ve Schnipke (1999), bireye uyarlanmış test uygulamalarında, zaman limitlerinin farklı etkisini azaltarak madde seçimini kontrol altında tutmak için loglineer tepki süresini önermişlerdir. van der Linden ve van Krimpen-Stoop (2003) ise çalışmasında, bu modelin, bireye uyarlanmış testlerde madde ön bilgisi durumunu ve testlerdeki hız-doğruluk ilişkilerini belirlemede yardımcı olup olmadığını araştırmışlardır. Loglineer madde tepki süresi, gözlemlenen madde tepki sürelerinin büyük ortalaması, tepki süresi üzerindeki madde etkisi ve birey etkisine ayrıştırılarak Eşitlik 40 ile ifade edilmektedir.

$$\ln T_{ij} = \mu + \delta_i + \tau_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \sigma^2) \quad (40)$$

Eşitlik 40'da, $\ln(T_{ij})$; j . bireyin i . maddeye ilişkin gözlenen tepki süresinin doğal logaritmasını, μ ; tüm bireylerin bütün maddelere verdikleri tepki süresinin büyük ortalamasını, τ_j ; j . bireyin i . maddedeki tepki hızı üzerindeki bireyin yavaşlık etkisi parametresini, δ_i ; i . maddenin tepki süresi üzerindeki madde etkisi parametresini ifade etmektedir. Bu model, tepki sürelerinin random dağıldığı varsayımına dayanmaktadır. Her parametre değeri, Eşitlik 41, Eşitlik 42 ve Eşitlik 43' de gösterildiği gibi ifade edilmektedir.

$$\mu = E_{ij}(\ln T_{ij}) \quad (41)$$

$$\delta_i = E_j(\ln T_{ij}) - \mu \quad (42)$$

$$\tau_j = E_i(\ln T_{ij}) - \mu \quad (43)$$

van der Linden ve van Krimpen-Stoop (2003) çalışmalarında, bireye uyarlanmış testlerde, madde ön bilgisinden ve hız-doğruluk ilişkilerinden kaynaklı olan uygun olmayan madde tepki örüntülerini belirlemek için loglineer tepki süresi modelini kullanarak, bireysel tepki süresi vektörlerinin madde tepki kuramı uyum analizlerini, bayesci sonsal yordayıcı ve klasik prosedürler kullanarak değerlendirmişlerdir. Beklenmeyen tepki davranışlarının tespiti, simülasyon çalışmaları ile beklenen ve yordanan loglineer madde tepki süresinin artık analizler kullanılarak karşılaştırılması ile

gerçekleştirilmiştir. Araştırmanın sonucunda, loglineer madde tepki süresi kullanımının uygun olmayan madde tepki örüntülerini belirlemede etkili olmadığı ve loglineer madde tepki süresi modelinin bireysel tepki süresi vektörlerine uygunluğunu belirlemede bayesci sonsal yordayıcı kontrollerin klasik prosedürlere göre daha etkili olduğu gözlenmiştir.

Etkili Tepki Süresi Modeli

Meijer ve Sotaridona (2006), madde ön bilgisine sahip bireyleri tespit etmek için etkili tepki süresi modelini geliştirmişlerdir. Etkili Tepki Süresi (ERT) adı verilen ölçü, bireyin maddeyi doğru cevaplaması için gereken süre olarak tanımlanmıştır. Beklenen tepki süresine göre daha kısa sürede verilen tepkiler madde ön bilgisinin göstergesi olabilmektedir. Etkili Tepki Süresi Modeli Eşitlik 44 ile ifade edilmektedir.

$$\ln T_{ij} = \beta_0 + \beta_1 \theta_j + \beta_2 \tau_j + \varepsilon_j, \quad \varepsilon_j \sim N(0, \sigma^2) \quad (44)$$

Eşitlik 44’de, $\ln(T_{ij})$; j.bireyin i. maddeye ilişkin gözlenen tepki süresinin doğal logaritmasını, θ_j ; j.bireyin yetenek parametresini, τ_j ; j. bireyin i.maddedeki tepki hızı üzerindeki bireyin yavaşlık etkisi parametresini, β ’lar ise regresyon katsayılarını ifade etmektedir. Böylelikle etkili tepki süresi modellemesinde, madde karakteristiklerinin yer almadığı, madde tepki süresi kestirimlerinin birey parametreleri ile yapıldığı görülmektedir. Meijer ve Sotaridona (2006), bireyin yetenek düzeyinin kestirilmesi için 3PLM uygulamalarını, yavaşlık etkisi parametrelerinin kestirimi için loglineer tepki süresi modelini (van der Linden, Scrams ve Schinipke, 1999) önermişlerdir. Etkili tepki süresi modelinin bir özelliği, bireyin maddeyi doğru yanıtlaması için yeterince yüksek yetenek düzeyinde olmasını gerektirmesi ve düşük yetenekli bireylerin madde tepki süresi kestirimlerinin bilgilendirici olmaması ya da yanlış bilgilendirme (hızlı tahmin vs.) yapabilmesidir. Bu nedenle, Meijer ve Sotaridona (2006), etkili tepki süresi modeline, testte yer alan maddeleri doğru cevaplama olasılıkları yüksek olan bireyleri dahil etmeyi önermişlerdir. Çalışmada, madde ön bilgisinin değerlendirilmesi, testte yer alan her maddenin gözlenen tepki süresi ile etkili tepki süresinden tahmin edilen tepki süresinin artık analizi ile karşılaştırılması ile yapılmıştır. Araştırmanın sonucunda, etkili tepki

modelinin düşük yetenekli bireylerin madde tepki sürelerinin kestiriminde yanlış sonuçlar oluşturduğu görülmüştür.

Bayesci Log-Normal Tepki Süresi Modeli

Tepki süreleri, testi alan bireylerin tepki davranışı hakkında önemli bilgiler sağlamasına rağmen bireylerin yetenekleri hakkında bilgi sahibi olunamaması durumu madde ön bilgisine sahip olan bireyleri belirlemede yetersiz kalabilmektedir. Böylelikle, madde tepki sürelerinin madde tepkileri ile entegrasyonu tepki davranışlarının araştırılmasında önemli hale gelmektedir. Loglineer madde tepki süresi ve etkili tepki süresi modelleri sadece tepki sürelerini dikkate alırken van der Linden (2006), tepki süresi modellenmesinde, madde tepkileri ile uyum sağlayabilecek yani tepki değişkenlerinin maddelere verilen tepki sürelerinden bağımsız olarak modellendiği lognormal tepki süresi modelini önermiştir. Maddeyi yanıtlamak için gerekli olan süreyi ifade eden zaman yoğunluğu ve bireyin maddeyi yanıtlamak için harcadığı süreyi yani hız parametresini içeren lognormal madde tepki süresinin modelinin olasılık yoğunluk fonksiyonu Eşitlik 45 ile ifade edilmektedir.

$$f(t_{ij}|\alpha_i, \beta_i) = \frac{\alpha_i}{t_{ij}\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} [\alpha_i (\ln t_{ij} - (\beta_i - \tau_j))]^2 \right\} \quad (45)$$

Eşitlik 45’de, $\ln t_{ij}$; j.bireyin i. maddeye ilişkin gözlenen tepki süresinin doğal logaritmasını, τ_j ; j. bireyin i.maddedeki örtük hız parametresini, β_i ; i. madde için zaman yoğunluğu parametresini, α_i ; i.madde için zaman ayıricılık parametresini ifade etmektedir. van der Linden (2009) , çalışmasında, lognormal madde tepki süresi modeli ile bir maddenin gözlenen tepki süresini, hız ve zaman parametrelerine ayrıştırarak gösterilebileceğini Eşitlik 46 ile ifade etmiştir.

$$t_{ij} = \frac{\beta_i}{\tau_j} \quad (46)$$

Modelde, tepki sürelerinin logaritması modellenmiş ve dağılımları normal varsayılmıştır. Thissen (1983), Schnipke ve Scrams (1997) ve van der Linden ve diğerleri (1999) çalışmalarında, tepki süresinin negatif olamamasından kaynaklı olarak dağılımlarının sağa çarpık olma eğiliminde olduğuna dikkat çekip dağılımı simetrik hale getirmek için logdönüşümü önermişler ve lognormal dağılımın tepki süresi ile iyi uyum gösterdiğini ortaya koymuşlardır. van der Linden (2009, 2011), bireylerin aynı maddeye verdiği tepki süresinin farklı ölçüm zamanlarında farklılık gösterebilmesinden kaynaklı olarak gözlenen madde tepki sürelerinin random olarak ele alınması gerektiğini belirtmiştir. Böylelikle 2PLM için zaman ayıricılık parametresi, hız parametresi ve madde hata varyanslarını içeren log-normal madde tepki süresi modeli Eşitlik 47 ile gösterilmiştir.

$$\ln T_{ij} = \beta_i - \tau_j + \varepsilon_{ij}, \quad \varepsilon_{ij} \sim N(0, \alpha_i^{-2}) \quad (47)$$

Eşitlik 47’de, β_i ; zaman yoğunluk parametresini (logaritmik ölçekte i.maddeyi yanıtlamak için gerekli olan ortalama süre), τ_j ; j.birey için hız parametresini, α_i ; i.madde için zaman ayıricılık parametresini ε_{ij} ; normal dağılımın artık hata terimini, α_i^{-2} ; hata teriminin standart sapmasının tersi olan zaman-basıklık parametresini ifade etmektedir. Lognormal madde tepki süresi modelindeki tepki süresi parametrelerinin her ikisinin de bilinmediği ve kestirilmesi gerektiği görülmektedir. Bu durum, birçok parametre setinin aynı tepki süresinin kestirilmesine neden olabilmektedir. Bu sorunun çözümüne yönelik olarak van der Linden (2010), parametrelerin metriğinin düzeltilmesine yönelik olarak hız ve zaman yoğunluk parametresinin kalibrasyonu esnasında ortalamasının sıfıra eşitlenmesi, bir maddenin zaman yoğunluk ya da hız parametresinin keyfi bir değere eşitlenmesi gibi öneriler geliştirmiştir. van der Linden (2006) çalışmasında, gerçek veri setinin model parametre kestirimlerinin bayesci yaklaşım kullanılarak model parametrelerinin sonsal dağılımlarından belirlendiği Gibbs örneklemeyi referans alan MCMC (Markov Chain Monte Carlo) metodu kullanarak lognormal madde tepki modeli ile uyumunu incelemiştir. Çalışmada, modelin uyumunu değerlendirmek için sonsal yordayıcı dağılım altındaki lognormal gözlenen tepki sürelerinin olasılığını bayesci yordayıcı kontroller kullanarak, bireyler ve maddeler açısından lognormal tepki süresi modeli ile model veri uyumunun sağlandığı sonucuna ulaşmıştır. Bu bulgular, lognormal madde tepki süresi modelinin, tepki sürelerini analiz etmede iyi performans sergilediğini

ve tepki süresi parametreleri aracılığıyla test maddelerinin kalitesinin kontrol edilmesinde kullanılabileceğini göstermektedir.

Beklenmeyen madde tepki davranışını tespit edebilmek için ilk önce tipik tepki davranışını iyi tanımlama gerekliliğinden dolayı dürüst (hile yapmayan) bireyler için iyi bir madde tepki modeli geliştirilmesi, madde ön bilgisinin belirlenmesi için önemli bir adımdır. Tepki süresini sadece beklenmeyen madde tepki davranışını belirlemede kullanmak, bu tepki davranışlarının çeşitli türlerini ayırt etme olanağını sınırlamaktadır. Örneğin, testi alan bireyin, bir maddeye birkaç saniye içinde tepki vermesi hem madde ön bilgisinin göstergesi hem de maddeyi okumadan meydana gelen dikkatsiz bir tepkinin göstergesi de olabilmektedir. Bu durumda, madde tepki sürelerinin madde tepkileri ile entegrasyonu, beklenmeyen madde tepki davranışın türünü belirlemede belirleyici olabilmektedir. Eğitimsel ve psikolojik testlerle ilgili çalışmalar (Roksam, 1997; Verhelst, Verstralen ve Jansen, 1997) , madde tepki süresi ile hız-tepki doğruluğu gibi performans görevleri arasında negatif bir ilişki olduğunu belirtmişlerdir. Bu ilişki, uygulamada zaman sınırı olan testlerin doğası göz önüne alındığında, bireyin maddeye doğru tepkiyi verebilmesi için düşünme, akıl yürütme ve çözme için yeterli süreye sahip olması gerektiğinden dolayı doğrulanabilir özelliktedir. Bireye, maddeye doğru tepkiyi verebilmesi için yeterli sürenin verilmemesi maddenin yanlış cevaplandırmasına sebep olabilmektedir (Lee,2018).

Hız-doğruluk ilişkisine dayanarak madde tepkileri ile madde tepki sürelerini entegre etmek için çeşitli modeller geliştirilmiştir. Wang ve Hanson (2005), tepki süresini 3PLM ile birleştirildiği Eşitlik 48 ile tanımlanan modeli önermişlerdir.

$$P(U_{ij} = 1|\theta_j, a_i, b_i, c_i, d_i, \rho_j, t_{ij}) = c_i + \frac{1 - c_i}{1 + \exp\left(-a_i \left(\theta_j - \left(\frac{\rho_i d_i}{t_{ij}}\right) - b_i\right)\right)} \quad (48)$$

Eşitlik 48’de, θ_j ; testi alan j.bireyin yeteneği, a_i , b_i , c_i sırasıyla i.maddenin ayırt ediciliği, güçlük düzeyi ve şans parametresini, d_i ; i.maddenin hız parametresini, ρ_j ; testi alan j.birey için hız parametresini ve t_{ij} ; j.bireyin i.madde için gözlenen tepki süresini ifade etmektedir. Bu modele göre maddenin gözlenen tepki süresi arttıkça maddeye doğru tepki verme olasılığının arttığı ve bu durumun da hız-doğruluk ilişkisi ile tutarlı olduğu görülmektedir. Bununla birlikte, van der Linden (2007) çalışmasında, hız-doğruluk değişiminin tek bir birey için gözlemlenen bir olgu olduğunu ve daha yüksek yetenek

düzeyine sahip olan bireylerin bir maddeye daha hızlı tepki verebilmesi için tersine bir ilişkinin açıkça mümkün olduğunu belirtmiştir. Bununla birlikte, Wang ve Hanson (2005)' in tepki süresi modelinde de , gözlenen tepki sürelerinin hız faktörü ile eş değer olmadığı ancak maddenin karakteristik özellikleri aracılığıyla belirlenen tepki süreleri ve bireyin maddeye tepki vermek için harcadığı süre olmak üzere iki faktör tarafından belirlendiği görülmektedir. Madde Tepki Kuramının yerel bağımsızlık varsayımına benzer şekilde, testi alan bireyin tepki süresinin bireyin hız seviyesine şartlı olarak bağımsız olduğu varsayılmaktadır. Böylelikle, madde tepkileri ve tepki süreleri bireyler ve maddeler çerçevesinde bütünleştirildiği zaman yetenek ile hız parametresi ve madde güçlüğü ile madde zaman yoğunluk parametresi arasındaki ilişkinin gözlemlenebilir olduğu ve böylece modellemeye ihtiyaç duyulduğu görülmektedir (van der Linden ve Glas, 2010). van der Linden (2007), madde tepki süreleri için bir model ve gizil değişkenlerin (birey yetenek düzeyi ve madde düzeyi) gözlemlerindeki değişimin açıklandığı tepki doğruluğu için ayrı bir model tanımlayıp bu iki süreç arasında ve içindeki koşullu bağımsızlığı tanımlamıştır. Böylelikle, van der Linden (2006)'in tepki süresi modellemesinin, madde tepkilerinin ve madde tepki sürelerinin dağılımlarının ikinci düzey bir modelleme aracılığıyla ayrı parametrelerle belirlendiği varsayımına dayandığı üçüncü bir yaklaşımı temel aldığı ve bu modelin hiyerarşik bir çerçeveye genişletildiği görülmektedir (van der Linden,2007).

Madde Tepkilerinin ve Madde Tepki Sürelerinin Birlikte Hiyerarşik Modellenmesi (HMTK-TS)

Aşağıda madde tepkilerinin ve madde tepki sürelerinin hiyerarşik bir çerçevede nasıl modellendiği ve madde ön bilgisini belirlemede nasıl uygulandığına yönelik açıklamalar yer verilecektir:

Modellemenin 1.aşaması:

Hiyerarşik ortak modelleme yaklaşımında, yetenek, gizil değişken olarak ele alınmakta ve maddelere verilen tepkiler bireyin yeteneğinin bir göstergesi olarak kabul edilmektedir. Hız da zaman yoğunlukları ile ilişkili olarak ölçülen, doğrudan gözlenemeyen gizil bir değişken olarak ele alınmaktadır. İlk aşamada, madde tepkilerinin

ve tepki sürelerinin dağılımlarının farklı modellerden (farklı birey ve madde parametreleri) geldiği varsayılmaktadır. Birey parametreleri, bireyin madde üzerindeki hız ve madde tepki doğruluğunu (yetenek) ifade etmektedir. Bu modellemenin ilk aşamasında, madde tepki süresi ve madde tepkilerinin hız ve doğruluk parametrelerine koşullu olarak bağımsız olduğu varsayılmaktadır (Fox ve Mariani, 2017). Tepki süresi modelleri için hiyerarşik modelleme çerçevesine çeşitli madde tepki modelleri uygulanabilmektedir. Fox, Klein Entink ve van der Linden (2007), Klein Entink, Fox ve diğerleri (2009a) ve van der Linden (2007), birinci aşama gözlemleri modellemek için lognormal madde tepki süresi modeli ve MTK modelinin kullanıldığı bir Bayesci modelleme çerçevesi geliştirmişlerdir. van der Linden (2007) çalışmasında, Rouder, Sun, Speckman, Lu ve Zhou (2003) ve Tatsuoka ve Tatsuoka (1980) tarafından önerilen kayma ya da yer parametresini içeren ,Weibull dağılımını temel alan tepki süresi modelini, çerçeveye uygun bir model oluşturmak için Eşitlik 49’da ifade edilen yoğunluk fonksiyonu ile parametreleştirmiştir.

$$f(t_{ij}) = \pi \alpha_i^\pi \left(t_{ij} - (\beta_i - \tau_j) \right)^{\pi-1} \exp \left(- \left(\alpha_i \left(t_{ij} - (\beta_i - \tau_j) \right) \right)^\pi \right),$$

$$t_{ij} > \beta_i - \tau_j \quad (49)$$

Eşitlik 49’da , t_{ij} ; j.bireyin i.maddesi üzerindeki tepki süresi, τ_j ; j.birey için hız parametresi, β_i ; j.birey için zaman yoğunluk parametresini, α_i ; i.madde için zaman ayıricılık parametresini, $\beta_i - \tau_j = \psi_j$; kaymayı, σ_j ; ölçek parametresini ve π_j ise genel biçim parametresini ifade etmektedir. Üç parametrelili normal ogive madde tepki modeli temel alınarak her bir tepki değişkeninin dağılımı Eşitlik 50, maddeyi doğru cevaplama Bernoulli olasılığı Eşitlik 51, tepki sürelerinin lognormal dağılımı Eşitlik 52 ve lognormal tepki süresinin olasılık yoğunluk fonksiyonu Eşitlik 53 ile gösterilmiştir.

$$U_{ij} \sim f(u_{ij}; \theta_j, a_i, b_i, c_i), \quad (50)$$

$$p_i(\theta_j) = c_i + (1 - c_i) \Phi \left(a_i (\theta_j - b_i) \right), \quad (51)$$

$$T_{ij} \sim f(t_{ij}; \tau_j, \alpha_i, \beta_i), \quad (52)$$

$$f(t_{ij}|\alpha_i, \beta_i) = \frac{\alpha_i}{t_{ij}\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} [\alpha_i (\ln t_{ij} - (\beta_i - \tau_j))]^2 \right\} \quad (53)$$

Eşitlik 50’de, $i = 1, \dots, I$; madde dizisini, $j = 1, \dots, J$; birey dizisini, $U_j = (U_{1j}, \dots, U_{Ij})$; j.bireyin beklenen madde tepki vektörünü, $T_j = (T_{1j}, \dots, T_{Ij})$; j.bireyin beklenen madde tepki süresi vektörünü, $u_j = (u_{1j}, \dots, u_{Ij})$; j.bireyin gözlenen madde tepki vektörünü, $t_j = (t_{1j}, \dots, t_{Ij})$; j.bireyin gözlenen madde tepki süresi vektörünü, Eşitlik 51’de, θ_j ; birey yeteneğini, a_i , b_i ve c_i sırasıyla madde ayırt edicilik, madde güçlük düzeyini ve şans parametresini, ϕ ise normal dağılım fonksiyonunu ifade etmektedir. Bireyin beklenen madde tepki ve madde tepki süresi vektörü (U_j, T_j) ’nin koşullu bağımsız olmasından dolayı birey parametreleri vektörünün, $\xi_j = (\theta_j, \tau_j)$ ile madde parametreleri vektörünün, $\Psi = \Psi_i = (a_i, b_i, c_i, \alpha_i, \beta_i)$ ile gösterildiği örneklem dağılımı Eşitlik 54 ile ifade edilmektedir.

$$f(u_j, t_j; \xi_j, \psi) = \prod_{i=1}^I f(u_{ij}; \theta_j, a_i, b_i, c_i) f(t_{ij}; \xi_j, \psi) \quad (54)$$

Eşitlik 54 ile gösterilen lognormal tepki süresinin olasılık yoğunluk fonksiyonu, tek boyutlu ikili puanlanan maddeler için genel olarak kullanılan parametreleştirmedir. Modelde, zaman yoğunluk parametresinin (β_i) bağımlı olmasından dolayı, farklı maddeler üzerindeki sürelerin logaritmasının varyansları arasındaki farkların esnek olmasına izin verilmektedir. Zaman yoğunluk parametresinin yüksek olması, madde üzerinde zaman dağılımının konumunun pozitif kaymasına yol açmaktadır. Fox ve diğerleri (2007), Klein Entink ve diğerleri (2009a), zaman ayıricılık parametresini, testi alan bireyler için maddenin farklı hız seviyelerindeki duyarlılığının modellendiği bir eğim parametresi olarak tanımlamışlardır. Bu tanımlama, van der Linden (2009)’in zaman ayıricılık parametresi tanımından farklılık göstermektedir. van der Linden (2009)’in yaklaşımında zaman ayıricılık parametresi, ölçmenin hatasının standart sapmasının tersi olarak adlandırılmaktadır. Bu durum, maddeye özgü farklılıklar sağlamaktadır. Bununla birlikte, van der Linden (2006, 2007), Thissen (1983), Schnipke ve Scrams (1997) ve van der Linden ve diğerleri (1999) çalışmalarında, madde tepki süresi verilerinin birkaç modele (örneğin üstel, gama, lognormal ve Weibull dağılımları) uyumunu değerlendirip lognormal dağılımın tepki süresinin en iyi uyum sağlayan dağılım olduğu sonucuna ulaşmışlardır.

Modellemenin 2.aşaması:

Hiyerarşik modellemenin ikinci aşamasında, tepki sürelerinin yanında madde tepkileri de modellenerek birey parametrelerinden olan hız ve yeteneğin ortak dağılımları Eşitlik 55’de gösterildiği gibi iki değişkenli normal dağılım olarak tanımlanmaktadır. Eşitlik 56’da olasılık yoğunluk fonksiyonu gösterilmiştir.

$$\xi_j \sim f(\xi_j; \mu_p, \Sigma_p) \sim MVN(\mu_p, \Sigma_p), \mu_p = (\mu_\theta, \mu_\tau), \Sigma_p = \begin{pmatrix} \sigma_\theta^2 & \sigma_{\theta\tau} \\ \sigma_{\theta\tau} & \sigma_\tau^2 \end{pmatrix} \quad (55)$$

$$f(\xi_j; \mu_p, \Sigma_p) = \frac{|\Sigma_p^{-1}|^{1/2}}{2\pi} \exp \left[-\frac{1}{2} (\xi_j - \mu_p)^T \Sigma_p^{-1} (\xi_j - \mu_p) \right] \quad (56)$$

Eşitlik 55’de, μ_p ; birey parametrelerinin örneklemine, μ_p ; ortalama vektörünü, Eşitlik 56’da, Σ_p ; kovaryans matrisini, p parametresi, birey parametreleri arasındaki kovaryansı göstermektedir. Hız ve yeteneğin, modellemenin ikinci aşamasında, testi alan bireylerin örneklemine temsil etmesi açısından bu modelleme, örneklem modellemesi olarak adlandırılır. Bayesci yaklaşımda bilinmeyen parametre Θ , sabit bir parametre değil, bir random değişkeni olarak kabul edilmektedir. Bu sayede, klasik yaklaşımın aksine parametreler hakkında olasılıklar dikkate alınarak yorumlanabilmektedir. Bayesci yaklaşımın temel amacı, veri elde edildikten sonra bilinmeyen parametre dağılımlarını elde etmek için önceden elde edilmiş bilgilerden yararlanarak önsel bilgilerle sonsal parametrelere ait kesitirimlerde bulunmaktır. Bayesci bakış açısına göre, testi alan bireylerin değiştirilemez olduğu ve dağılımın birey parametreleri için öncekini temsil ettiği savunulur. Σ_P kovaryans matrisi için hiperönsel (hyperprior) vP serbestlik dereceli ters Wishart dağılımı (gama dağılımının çok boyutlara bir genellemesi) ve ölçek parametresi V_1 seçilir. Modellemenin ikinci aşamasında, madde parametreleri arasındaki ilişkiler de belirlenerek birey parametrelerine benzer şekilde, madde parametrelerinin dağılımı da çok değişkenli normal dağılım olarak tanımlanmaktadır. Madde parametreleri arasındaki ilişkiler, madde alanı (I) içindeki madde parametrelerinin ortak olasılığının modellenmesiyle belirlenmektedir. Böylelikle, bu model ,madde alanı modeli olarak adlandırılmaktadır. Eşitlik 57, madde vektörünün (ψ_i) çok değişkenli normal dağılımını,

Eşitlik 58, olasılık yoğunluk fonksiyonunu, Eşitlik 59, ortalama vektörünü, Eşitlik 60 ise kovaryans matrisini göstermektedir.

$$\psi_i \sim f(\psi_i; \mu_I, \Sigma_I) \quad (57)$$

$$f(\psi_i; \mu_I, \Sigma_I) = \frac{|\Sigma_I^{-1}|^{1/2}}{(2\pi)^{5/2}} \exp \left[-\frac{1}{2} (\psi_i - \mu_I)^T \Sigma_I^{-1} (\psi_i - \mu_I) \right] \quad (58)$$

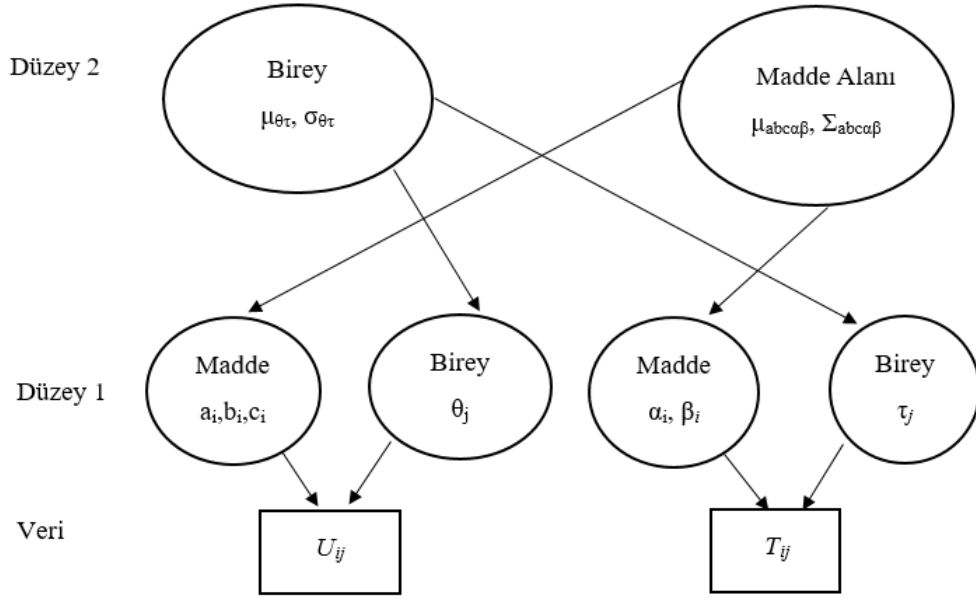
$$\mu_I = (\mu_a, \mu_b, \mu_c, \mu_\alpha, \mu_\beta) \quad (59)$$

$$\Sigma_I = \begin{pmatrix} \sigma_a^2 & \sigma_{ab} & \sigma_{ac} & \sigma_{a\alpha} & \sigma_{a\beta} \\ \sigma_{ab} & \sigma_b^2 & \sigma_{bc} & \sigma_{b\alpha} & \sigma_{b\beta} \\ \sigma_{ac} & \sigma_{bc} & \sigma_c^2 & \sigma_{c\alpha} & \sigma_{c\beta} \\ \sigma_{a\alpha} & \sigma_{b\alpha} & \sigma_{c\alpha} & \sigma_\alpha^2 & \sigma_{\alpha\beta} \\ \sigma_{a\beta} & \sigma_{b\beta} & \sigma_{c\beta} & \sigma_{\alpha\beta} & \sigma_\beta^2 \end{pmatrix} \quad (60)$$

Hiyerarşik modelleme için iki aşamayı içine alan tüm modeli betimleyen örneklemin olasılık yoğunluk fonksiyonu Eşitlik 61 ile gösterilmiştir.

$$f(u, t; \xi, \psi) = \prod_{j=1}^J \prod_{i=1}^I f(u_j, t_j; \xi_j, \psi_i) f(\xi_j; \mu_\rho, \Sigma_\rho) f(\psi_i; \mu_I, \Sigma_I) \quad (61)$$

Madde parametreleri için kovaryans matrisi, modellemenin birinci aşamasındaki parametreler ile madde parametreleri arasındaki korelasyonu tanımlamaktadır. Böylelikle, modellemenin iki aşamasında da tanımlanan kovaryans yapıları, madde tepkileri ve madde tepki süreleri arasındaki ilişkiyi ortaya koymaktadır. Hiyerarşik modellemenin grafiksel olarak gösterimi Şekil I ile gösterilmektedir.



Şekil 1. Test Maddelerindeki Hız ve Doğruluk İçin Hiyerarşik Modelleme

Kaynak. van der Linden, 2007, 295.

van der Linden (2007) çalışmasında, madde tepkilerinin 3PLM ile madde tepki sürelerinin de lognormal madde tepki süresi ile modellendiği hiyerarşik modellemenin bilgisayar temelli test uygulamasından elde edilen veriler ile uyumunun sağlanması aracılığıyla birey ve madde parametreleri arasındaki ilişkiyi araştırmıştır. Çalışmanın sonucunda, yetenek ve hız parametreleri ile madde güçlüğü ve zaman yoğunluk parametreleri arasında pozitif bir korelasyon bulunmuştur. Ayrıca, van der Linden, Klein Entink ve Fox (2010) çalışmalarında, 3PLM için parametre kestirimlerinin doğruluğunun madde tepkilerinin ve madde tepki sürelerinin hiyerarşik olarak modellenmesiyle artış gösterdiği sonucuna ulaşmışlardır. Bulgular, madde tepki sürelerinin hesaplanmasının madde tepkilerinin modellenmesinde yardımcı olduğunu göstermektedir. Böylelikle, madde tepkilerinin yanında madde tepki sürelerinin de analiz edilmesinin ve bununla birlikte parametre kestirimlerindeki doğruluğun artış göstermesinin bireyin madde tepki davranışı hakkında daha fazla bilgi sağladığı ve beklenmedik madde tepki davranışlarını ortaya çıkardığı görülmektedir (Lee, 2018). Zhan, Jiao, Wang ve Man (2018) çalışmalarında, çok boyutlu testlerde de, gizil hız ve gizil yetenek parametrelerinin çok boyutluluğu göz önünde bulundurularak eş zamanlı olarak madde tepkilerinin ve madde

tepki sürelerinin çok boyutlu log-normal model birlikte hiyerarşik modellemenin uygun olduğunu göstermişlerdir.

Aşağıda madde tepkilerinin ve madde tepki sürelerinin hiyerarşik olarak modellenmesinde uygun olmayan madde tepki örüntüleri ve madde tepki sürelerinin belirlenmesinde kullanılan birey uyum istatistiklerinden kısaca bahsedilecektir.

HMTK-TS Modellemesinde Madde Ön Bilgisinden Kaynaklı Test Hilesinde Uygun Olmayan Madde Tepki Örüntülerini Belirlemek İçin Kullanılan Birey-Uyum İstatistikleri

Neyman–Pearson lemma Yaklaşımı

Levine ve Drasgow (1988), olabilirlik oran λ değeri Eşitlik 62 ile gösterilen yaklaşımı önermiştir.

$$\lambda = \frac{P_U(X = x)}{P_N(X = x)} \quad (62)$$

Eşitlik 62’de, x ; bireyin bütün maddeleri verdiği cevaplardan elde edilen test puanını, $P_U(X = x)$; uygun olmayan tepki örüntülerine sahip olan bireylerin madde-puan olasılık modelini, $P_N(X = x)$; normal tepki örüntülerine sahip olan bireylerin madde-puan olasılık modelini ifade etmektedir. k , normal tepki örüntülerine sahip olan bireylerin simülasyon çalışmaları aracılığıyla elde edilen kritik değeri göstermek üzere, $\lambda > k$ durumunda birey, uygun olmayan tepki örüntüsüne sahip olan birey olarak değerlendirilmektedir. $P_N(X = x)$, Eşitlik 63’de gösterildiği gibi, bir madde tepki modeli altında x değerinin hesaplanan marjinal olasılık değerini ifade etmektedir.

$$P_N(X = x) = \int_{\theta} f(\theta) \prod_{i=1}^I P_i(\theta)^{x_i} [1 - P_i(\theta)]^{(1-x_i)} d\theta, \quad f(\theta) = \frac{1}{\sqrt{2\pi}} e^{-\frac{\theta^2}{2}} \quad (63)$$

$P_U(X = x)$, değerinin hesaplanması ise testte yapılan hile türüne bağlı olup daha zordur. Drasgow ve diğerleri (1996), teşhir olmayan maddelerden hesaplanan test puanlarını göz önünde bulundurarak $P_U(X = x)$ için Eşitlik 64’de gösterildiği gibi bir hesaplama önermişlerdir.

$$P_U(X = x) = p^{n_{1c}} (1 - p)^{n_{1w}} \int_{\theta} f(\theta) \prod_{i \in u} P_i(\theta)^{x_i} [1 - P_i(\theta)]^{(1-x_i)} d\theta, n_{1c}$$

$$= \sum_{i \in c} x_i, \quad n_{1w} = \sum_{i \in c} (1 - x_i) \quad (64)$$

Eşitlik 64’de, u ; bireyin teşhir olmayan maddelerden elde edilen toplam test puanını, n_{1c} ; teşhir olan maddeleri cevaplayan bireylerin doğru cevap sayısını, n_{1w} , teşhir olan maddeleri cevaplayan bireylerin yanlış cevap sayısını ifade etmektedir. Böylelikle, Eşitlik 63 ve Eşitlik 64 kullanılarak teşhir olan madde alt setlerinin bilindiği durumlarda madde ön bilgisinden kaynaklı test hilesini belirlemek için olabilirlik oranı hesaplanabilmektedir (Drasgow vd,1996; Zickar ve Drasgow, 1996). Belov (2016) çalışmasında, yetenek düzeyinin $[-5,5]$ aralığında 101 eşit yetenek düzeyi aralığı için Eşitlik 65 ile ifade edilen *SBNPL*’yi önermiştir.

$$SBNPL = \frac{\sum_{j=1}^{101} f(\theta_j) \prod_{i \in u} P_i(\theta)^{x_i} [1 - P_i(\theta)]^{(1-x_i)} \prod_{i \in c} 0.95^{x_i} (1 - 0.95)^{(1-x_i)}}{\sum_{j=1}^{101} f(\theta) \prod_{i=1}^I P_i(\theta)^{x_i} [1 - P_i(\theta)]^{(1-x_i)}}$$

$$= \frac{0.95^{n_{1c}} (1 - 0.95)^{n_{1w}} \sum_{j=1}^{101} f(\theta) \prod_{i \in u} P_i(\theta)^{x_i} [1 - P_i(\theta)]^{(1-x_i)}}{\sum_{j=1}^{101} f(\theta) \prod_{i=1}^I P_i(\theta)^{x_i} [1 - P_i(\theta)]^{(1-x_i)}} \quad (65)$$

PSS (Posterior Shift Statistic)

Belov (2015), madde ön bilgisi kaynaklı test hilesini belirlemede, teşhir olan maddelerin bilindiği durumda, uygun olmayan madde tepkilerini belirlemek için *PSS (posterior shift statistic)* istatistiğini önermiştir. Bu istatistik, madde ön bilgisine sahip olan birey için teşhir olan maddelere verilen cevaplardan hesaplanan sonsal olasılık değerinin, teşhir olmayan maddelere verilen cevaplardan hesaplanan sonsal olasılık değerinin daha yüksek yetenek seviyesine kaydırılması gerekliliği varsayımına

dayanmaktadır. Bu istatistik ölçüsü, Eşitlik 66'da ifade edildiği gibi, ilk sonsal olasılık değerinin, ikinci sonsal olasılık değerinden daha düşük olmadığı en büyük sağ sınır bölgesindeki iki sonsal olasılık değeri arasındaki farkı göstermektedir.

$$S(F_C||F_U) = h \sum_{j=1}^J F_C(\theta_j) - F_U(\theta_j) \quad (66)$$

Eşitlik 66'da, $j=(1,...J)$; birey sayısını, $h=\theta_j - \theta_{j+1}$; birey yetenekleri arasındaki farkı, F_C ; bireylerin şüpheli teşhir olan maddelerden kestirilen yeteneklerinin sonsal dağılımını, F_U ; bireylerin şüpheli teşhir olmayan maddelerden kestirilen yeteneklerinin sonsal dağılımını, $S(F_C||F_U)$ ise bireylerin şüpheli teşhir olan ve teşhir olmayan maddelerden kestirilen yeteneklerinin sonsal dağılımları arasındaki farkı ifade etmektedir.

L_s İstatistiği

Guo ve Drasgow (2010) tarafından önerilen *LRT* istatistiğinin, bireyin teşhir olan ve teşhir olmayan maddelerden elde edilen yetenek düzeylerinin eşitliğinin test edildiği sıfır hipotezi Eşitlik 67 ile ifade edilmiştir.

$$\Gamma = 2[\ell(\hat{\theta}_c; x_i, i \in c) + \ell(\hat{\theta}_{\bar{c}}; x_i, i \in \bar{c}) - \ell(\hat{\theta}; x_i, i = 1, 2, \dots, I)],$$

$$\ell(\hat{\theta}_c; x_i, i \in c) = \sum_{i \in c} \log P_i(x_i | \hat{\theta}_c) \quad (67)$$

Eşitlik 67'de, c ; bireye uygulanan teşhir olan madde setini, $\bar{c}$; bireye uygulanan teşhir olmayan madde setini, $\hat{\theta}_c$; bireyin, teşhir olan maddelere verdiği tepkilerden kestirilen yeteneğini, $\hat{\theta}_{\bar{c}}$; bireyin, teşhir olmayan maddelere verdiği tepkilerden kestirilen yeteneğini, $\hat{\theta}$; bireyin, testteki tüm maddelere verdiği tepkilerden kestirilen yeteneğini, $\ell(\hat{\theta}_c; x_i, i \in c)$; $\hat{\theta}_c$ yetenek düzeyinde, teşhir olan madde setinden elde edilen test puanlarının log-olabilirlik değerini, $\ell(\hat{\theta}_{\bar{c}}; x_i, i \in \bar{c})$; $\hat{\theta}_{\bar{c}}$ yetenek düzeyinde, teşhir olmayan madde setinden elde edilen test puanlarının log-olabilirlik değerini ve

$\ell(\hat{\theta} ; x_i, i = 1, 2, \dots, I)$ ise $\hat{\theta}$ yetenek düzeyinde, testteki tüm maddelerden elde edilen test puanlarının log-olabilirlik değerini ifade etmektedir. Böylelikle, LRT istatistiği Eşitlik 68’de gösterildiği gibi ifade edilmektedir.

$$\Gamma = 2 \left\{ \sum_{i \in c} \log P_i(x_i | \hat{\theta}_c) + \sum_{i \in \bar{c}} \log P_i(x_i | \hat{\theta}_{\bar{c}}) - \sum_{i=1}^I \log P_i(x_i | \hat{\theta}) \right\} \quad (68)$$

Sinharay (2017a), madde ön bilgisine dayalı test hilesini belirlemek için Eşitlik 69’da gösterilen olabilirlik oran testine (LRT) dayalı L_s istatistiğini önermiştir.

$$L_s = \begin{cases} \sqrt{\Gamma} & , \quad \hat{\theta}_c \geq \hat{\theta}_{\bar{c}} \\ -\sqrt{\Gamma} & , \quad \hat{\theta}_c < \hat{\theta}_{\bar{c}} \end{cases} \quad (69)$$

L_s istatistiği, madde ön bilgisi olmadığı durumda sıfır hipotezi altında asimptotik standart normal dağılım sergilemektedir (Sinharay,2017a). Sinharay (2017b) çalışmasında, madde ön bilgisini belirlemeye yönelik PSS ve L_s istatistiğinin performansını, simülasyon çalışması ve bireye uyarlanmış test uygulamasından elde edilen gerçek veri seti ile karşılaştırarak, L_s istatistiğinin, uygun olmayan tepki örüntülerini belirlemeye yönelik performansının daha iyi olduğu sonucuna ulaşmıştır.

$L(l^y)$ Birey Uyum İstatistiği

Meijer ve Sijtsma (2001), Karabatsos (2003) ve Rupp (2013) çalışmalarında, uygun olmayan madde tepki örüntülerini tespit etmek için çok sayıda birey-uyum istatistiğine genel bir bakış sunmaktadır. Parametrik testler, bir dizi maddeye verilen tepkilerin uyumunun MTK modeli altında değerlendirilmesi ilkesine dayanmaktadır. Eğitimsel ölçmelerde, birey uyum analizlerinde sıklıkla kullanılan ve uygun olmayan madde tepki örüntülerini belirleme gücü yüksek olan l_z birey-uyum istatistiği (Drasgow, Levine ve Williams, 1985), l_0 birey-uyum istatistiğinin (Levine ve Rubin,1979) standardize edilmiş log olasılığıdır (Karabatsos, 2003; Dimitrov ve Smith, 2006). Marianti ve diğerleri (2014) ve Fox ve Marianti (2017), madde ön bilgisinden kaynaklı test hilesini belirlemeye yönelik, madde tepkilerinin log olasılığı olan l^y birey uyum

istatistiğini, tepki örüntülerinin uyumunun değerlendirilmesi için önermişlerdir. Madde tepkilerinin 2PLM ile modellendiği durumda tepki doğruluğunu tanımlayan maddeyi doğru yanıtlama olasılığı Eşitlik 70 ile ifade edilmiştir.

$$P(Y_{ji} = 1|\theta_j, a_i, b_i) = \Phi(a_i\theta_j - b_i) \quad (70)$$

Eşitlik 70’de, Φ , normal yığılmalı dağılım fonksiyonunu, a_i , madde ayırtedicilik parametresini, b_i , madde güçlük parametresini ifade etmektedir. Ortalaması $a_i\theta_j - b_i$ varyansı 1 olan normal dağılıma sahip $(N(a_i\theta_j - b_i, 1))$ gizil tepki değişkeni Z_{ji} tanımlandığında, böylelikle doğru cevabın olasılığı, gizil tepki değişkeninin değeri, 0’dan büyük olduğundaki olasılık değerine eşit olmaktadır. Z_{ji} , sıklıkla gözlenen kategorik tepkinin, gizil sürekli tepki olarak sunulmasında kullanılmaktadır (Fox, 2010). Gizil tepki değişkenlerinin, Markov Chain Monte Carlo (MCMC) simülasyon prosedürlerinde, model parametrelerinin kestirilmesinde fayda sağladığı görülmektedir (Albert ve Chib, 1993; Fox, 2010; Klein Entink vd., 2009a; Meng, Tao ve Chang, 2015). HMTK-TS Modellemesi için gizil sürekli tepkiler Eşitlik 71 ile ifade edilmektedir.

$$Z_{ik} = a_i\theta_j - b_i + e_{ji}, e_{ji} \sim N(0,1) \quad (71)$$

Birey uyum istatistiğinin, madde tepkilerinin 2PLM ile modellendiği duruma göre log olabilirlik değeri, beklenen değeri, varyansı ve normal dağılıma uygun olarak standardize edilmiş değeri sırasıyla Eşitlik 72, 73, 74 ve 75’de gösterilmiştir.

$$\begin{aligned} l^y(\theta_j, a, b; y_j) &= \log p(y_j|\theta_j, a, b) \\ &= \sum_{i=1}^I \log p(y_{ji}|\theta_j, a_i, b_i) \\ &= \sum_{i=1}^I \left[y_{ji} \log p(y_{ji}|\theta_j, a_i, b_i) \right. \\ &\quad \left. + (1 - y_{ji}) \log (1 - p(y_{ji}|\theta_j, a_i, b_i)) \right] \end{aligned} \quad (72)$$

$$\begin{aligned}
E(l^y(\theta_j, a, b; y_j)) &= \sum_{i=1}^I p(y_{ji} = 1 | \theta_j, a_i, b_i) \log p(y_{ji} = 1 | \theta_j, a_i, b_i) \\
&+ (1 - p(y_{ji} = 1 | \theta_j, a_i, b_i)) \log (1 - p(y_{ji} = 1 | \theta_j, a_i, b_i))
\end{aligned} \tag{73}$$

$$\begin{aligned}
Var(1 - p(y_{ji} | \theta_j, a_i, b_i)) &= \sum_{i=1}^I \left[p(y_{ji} = 1 | \theta_j, a_i, b_i) (1 - p(y_{ji} = 1 | \theta_j, a_i, b_i)) \log \left(\frac{p(y_{ji} = 1 | \theta_j, a_i, b_i)}{1 - p(y_{ji} = 1 | \theta_j, a_i, b_i)} \right)^2 \right]
\end{aligned} \tag{74}$$

$$l_s^y(\theta_j, a_i, b_i; y_j) = \frac{l^y(\theta_j, a_i, b_i; y_j) - E(l^y(\theta_j, a_i, b_i; y_j))}{\left(Var(l^y(\theta_j, a_i, b_i; y_j)) \right)^{\frac{1}{2}}} \tag{75}$$

Madde tepki örüntülerinin uç değer gösterip göstermediği bayesyan anlamlılık testleri ile belirlenebilmektedir. Eşitlik 71 incelendiğinde, madde tepkilerinin olabilirliği arttığında birey-uyum istatistik değerinin de yükseldiği görülmektedir. Tepki olasılıklarının logaritmasının negatif değerler almasından dolayı daha olası cevap olasılıklarının logaritması, daha az olası cevap olasılıklarının logaritmasından daha yüksek değerler almaktadır. Böylelikle, bireyin maddelere verdiği tepkilerden hesaplanan ve bireye özgü negatif istatistiklerin artan değerleri, uyumsuzluğa karşılık gelmektedir. Sonuç olarak, l_s^y birey-uyum istatistiğinin değerini artırmak için uyumsuzluğun sonsal olasılık değerinin arttırılması gerektiği görülmektedir. Sonsal olasılık değeri, birey uyum istatistiğinin değeri, istatistik için belirlenen eşik değerinden daha yüksek ise Eşitlik 76 ile gösterilmektedir (Fox ve Mariani, 2017).

$$P(l_s^y(\theta_j, a_i, b_i; y_j) > C) = \Phi(l_s^y(\theta_j, a_i, b_i; y_j) > C) = p_{ly} \tag{76}$$

HMTK-TS modellemesinde, birey-uyum istatistiği, hız ve yetenek gibi model parametreleri arasındaki ilişkilerin de göz önünde bulundurulması gerekliliğinden dolayı daha zor hesaplanmaktadır. Modelde, model parametreleri ve test istatistikleri bayes kestirim yöntemi kullanılarak hesaplanmaktadır. Bayesci yaklaşım, bilinmeyen parametre dağılımlarını elde etmek için önsel bilgilerden yararlanarak örneklem bilgilerinin bir araya getirilip, parametrelere dair kestirimlerin yapıldığı bir yöntemdir (DeMars, 2010). Bayesci yaklaşıma dayalı çalışmaların büyük bir kısmında MCMC yöntemleri, sonsal dağılımların benzetiminde kullanılmaktadır. Bu yöntemler ile karışık sonsal dağılımlardan örneklem çekme ve sonsal momentleri hesaplamak mümkün olmaktadır. MCMC yöntemleri, modelin karmaşıklığının arttığı durumlarda dağılımdan random çekimlerin simüle edilmesini ile eşzamanlı olarak tüm modelin parametrelerinin kestirimine olanak sağlamaktadır. MCMC'nin amacı, sonsal dağılımı kullanarak örneklemelerin elde edilmesidir. Monte Carlo tekniği, ilgilenilen teorik dağılımı özetleyen formüllerden hesaplanan değerler yerine farklı dağılımlardan elde edilen random örneklemeler kullanılmaktadır. Genel olarak MCMC yöntemleri, yöntemde her değer zincirde bulunan bir önceki değere bağlanarak bir zincir oluşturulmakta ve bu zincir istenilen sonsal dağılıma yaklaştırılmaktadır. Gibbs Örneklemesi ve Metropolis Hastings algoritması en yaygın kullanılan MCMC algoritmalarıdır. Gibbs örnekleme, her bir yinelemede, model parametrelerinin tam koşullu dağılımlarından bir örneklemin elde edildiği yinelemeli bir kestirim yöntemidir. Gibbs örnekleme olarak adlandırılan MCMC yöntemi ile, örneklem, model parametreleri arasındaki tam kovaryans yapı hesaplanarak model parametrelerinin beklenen sonsal dağılımlarından elde edilir (Klein Entink, Fox ve van der Linden, 2009; van der Linden, 2007; Fox, Klein Entink ve van der Linden, 2007; von Davier ve Rost, 2007). MCMC algoritması, model parametrelerinin tam kovaryans yapısı kullanılarak Eşitlik 76 ile gösterilen birey-uyum istatistiğinin hesaplanmasında kullanılmaktadır. Böylelikle, ikili sınıflandırma değişkeni olan F_j^y 'nin testi alan j.birey için verilen birey ve madde parametreleri değerlerinde aldığı değerler Eşitlik 77 ile gösterilmiştir.

$$F_j^y = \begin{cases} 1 & , P(l_s^y(\theta_j, a_i, b_i; y_j)) > C \\ 0 & , P(l_s^y(\theta_j, a_i, b_i; y_j)) \leq C \end{cases} \quad (77)$$

F_j^y 'nin aldığı değer, her MCMC yinelemesinde hesaplanmakta ve ortalamanın üzerindeki MCMC yinelemeleri, uygun olmayan madde tepki örüntülerinin sonsal olasılıklarının kestirilmesinde kullanılabilmektedir. Bu metod, sonsal olasılığı C ile belirlenen eşik değerden büyük olan birey uyum istatistik değerleri hesaplanarak uygun olmayan madde tepki örüntülerinin belirlenmesine olanak sağlamaktadır (Fox ve Mariani,2017).

HMTK-TS Modellemesinde Uygun Olmayan Madde Tepki Süresi Örüntülerini Belirlemek İçin Kullanılan Birey-Uyum İstatistikleri

$L(l^t)$ Birey Uyum İstatistiği

Madde ön bilgisinden kaynaklı test hilesini belirlemeye yönelik çalışmalarda, tepki sürelerinin log-olabilirlik değerleri, tepki sürelerini içeren tepki örüntülerinin uyumunun değerlendirilmesinde kullanılmaktadır. Tepki süreleri için olabilirlik tabanlı birey-uyum istatistiği, tepki örüntülerinin yoğunluk bilgisini gerektirmektedir. Bu istatistik, tepki sürelerinin logaritması için normal bir modelden elde edilmektedir. Eşitlik 45 ile gösterilen lognormal madde tepki süresi modelinde, madde tepki örüntülerinin olasılık yoğunluk fonksiyonu bireysel tepki sürelerinin çarpımı ile ifade edilmektedir. Bu tepki süresi örüntülerinin $(t_j^* = (t_{1j}^*, \dots, t_{Ij}^*))$ olasılık yoğunlukları, Eşitlik 78'de gösterilmektedir.

$$\begin{aligned} l_0(\tau_j, \beta, \alpha, \sigma^2; t_j^*) &= -2 \log p(t_j^* | \tau_j, \beta, \alpha, \sigma^2) = -2 \sum_{i=1}^I \log p(t_{ij}^* | \tau_j, \beta_i, \alpha_i, \sigma_i^2) \\ &= \sum_{i=1}^I \left(\left(\frac{t_{ji}^* - \mu_{ji}}{\sigma_i} \right)^2 + \log(2\pi\sigma_i^2) \right) = \sum_{i=1}^I (Z_{ji}^2 + \log(2\pi\sigma_i^2)) \\ &= \sum_{i=1}^I l_{0i} \end{aligned} \quad (78)$$

Eşitlik 78'de, t_j^* ; j.bireyin i.madde üzerindeki tepki süresinin logaritmasını $(t_{ij}^* = \ln t_{ij})$ ifade etmektedir. Z_{ji} , tepki sürelerinin logaritmasının normal dağılımının

standardize edilmiş hata değerlerini göstermesinden kaynaklı normal dağılım göstermektedir. l_0 birey istatistiği, testte yer alan bütün maddeler üzerinden hesaplanabileceği gibi testin alt test oluşturan maddeler çerçevesinde de değerlendirilebilmekte ve test istatistiği, çeşitli model parametrelerine bağlı olup birey uyum istatistik değeri model parametrelerinin sonsal dağılımlarından hesaplanmaktadır (Marianti vd., 2014). Birey uyum istatistikleri ile ilgili çalışmalar incelendiğinde, standardize edilmiş birey-uyum istatistiklerinden olan l_z birey-uyum istatistiğinin asimptotik normal dağılıma sahip olduğu için dikkat çektiği görülmektedir (Drasgow vd., 1985; Meijer ve Sijsma, 1995; Molenaar ve Hoijsink, 1990; Snijders, 2001). Madde tepki süreleri için l_0^t birey uyum istatistiği ve standardize edilmiş versiyonu olan l_z^t birey uyum istatistiğinin koşullu beklenen ve varyans değerleri Eşitlik 79 ve Eşitlik 80 ile gösterilmiştir.

$$E[l_0(\tau_j, \beta, \alpha, \sigma^2) | t_j^*, \tau_j, \beta, \alpha, \sigma^2] = \sum_i (1 + \ln(2\pi\sigma_i^2)) \quad (79)$$

$$Var[l_0(\tau_j, \beta, \alpha, \sigma^2) | t_j^*, \tau_j, \beta, \alpha, \sigma^2] = 2I \quad (80)$$

Böylelikle, l_z^t birey uyum istatistiği, Eşitlik 88’de yer alan istatistikler kullanılarak Eşitlik 79 ve Eşitlik 80 ışığında standartlaştırma kullanılarak Eşitlik 81’ de gösterildiği şekilde elde edilmektedir.

$$\begin{aligned} l_z^t(\tau_j, \beta, \alpha, \sigma^2; t_j^*) &= \frac{(\sum_{i=1}^I Z_{ji}^2 + \log(2\pi\sigma_i^2)) - (\sum_{i=1}^I 1 + \log(2\pi\sigma_i^2))}{\sqrt{2I}} \\ &= \frac{\sum_{i=1}^I Z_{ji}^2 - 1}{\sqrt{2I}} \end{aligned} \quad (81)$$

Birey uyum istatistiğinin elde edilebilmesi için l_z^t istatistiğinin sıfır dağılımının elde edilmesi gereklidir. Testi alan bir bireyin lognormal tepki süresi modeli altındaki madde tepki süresi örüntüsü için birey uyum istatistiğinin sıfır dağılımının $l_z^t(T_j^*)$ oluşturulması üç farklı şekilde gerçekleştirilebilmektedir. İlk olarak $l_z^t(T_j^*)$ ’nin sıfır dağılımı, standartlaştırılmış hataların karelerinin toplamınının (Z_{ji}) I serbestlik dereceli ki kare dağılımının standart normal dağılıma sahip olmasından dolayı yaklaşık olarak standart normal dağılıma sahiptir. Box, Hunter ve Hunter (1978) çalışmalarında, T

değişkenli I serbestlik dereceli Ki- kare dağılımının $(T - I)\sqrt{2I}$ dağılımında normal dağılım sergilediğini göstermişlerdir. İkinci olarak; Eşitlik 81’de gösterildiği gibi $l_z^t(T_j^*)$ ’nin standardize edilmemiş versiyonu aracılığıyla tam sıfır dağılımı elde edilebilmektedir. Üçüncü olarak, $l_z^t(T_j^*)$ birey uyum istatistiğini standardize etmek için Eşitlik 82’de gösterildiği gibi Wilson-Hilferty dönüşümü ile yaklaşık olarak normal dağılım elde edilebilmektedir.

$$l^t(T_j^*) = \sum_{i=1}^I Z_{ji}^2, \quad l_s^t(T_j^*) = \frac{(\sum_{i=1}^I Z_{ji}^2 / I)^{1/3} - (1 - 2/(9I))}{\sqrt{2/(9I)}} \quad (82)$$

Birey uyum istatistiğinin sonsal dağılımı, tepki sürelerinin model altında beklenen tepki sürelerinden ayırt edilmesini yani uygun olmayan tepki süresi örüntüsüne sahip olan bireylerin belirlenmesini sağlamaktadır. Tepki süreleri örüntüsünün uç değer gösterip göstermediğini değerlendirebilmek için, sonsal dağılımın olasılık değeri olan gözlenen istatistik değeri ile gözlenen istatistik değeri için sıfır hipotezinin reddedildiği değerler kümesini içeren kritik bölgenin sınırını ifade eden ve ki- kare dağılımından elde edilen C eşik değeri Eşitlik 83’de gösterildiği gibi karşılaştırılmaktadır.

$$P(l^t(T_j^*) > C) = P(\chi_I^2 > C) = \alpha \quad (83)$$

Eşitlik 83 ’de α ; model parametrelerinin I serbestlik dereceli ki-kare değerlerini içeren test istatistiğinin sonsal dağılımın olasılığını ifade etmektedir. Test istatistiğinin sonsal olasılık değeri gözlenen istatistik değeri olan C kritik değerinden daha büyükse uç değer olarak değerlendirilmektedir. Tepki süreleri için sıfır dağılımlarının farklı yöntemler ile belirlendiği üç farklı birey uyum istatistiği Tablo 1’de verilmiştir (Marianti vd., 2014; Fox ve Marianti, 2017)

Tablo 1

Lognormal Modeli Altında Tepki Süreleri İçin Birey Uyum İstatistikleri

İstatistik	Sıfır Dağılımının Çeşidi	Tam Ya Da Yaklaşıklığı	Anlamlılık Düzeyinin Olasılık Değeri
l_z^t	Normal	Yaklaşık	$P(l_z^t(T_j^*) > C) \sim \Phi(l_z^t(T_j^*) > C)$
l^t	χ^2	Tam	$P(l^t(T_j^*) > C) = P(\chi_l > C)$
l_s^t	Normal	Yaklaşık	$P(l_s^t(T_j^*) > C) \sim \Phi(l_s^t(T_j^*) > C)$

Kaynak. Marianti ve diğerleri, 2014, 435.

Her MCMC yinelemesinde, örneklenen parametre değerleri göz önüne alındığında, model parametrelerinin bir fonksiyonu olan olasılık dağılımları da hesaplanabilmektedir. Model parametrelerinin sonsal dağılımından elde edilen örneklem değerlerinin sonsal olasılık değeri, Eşitlik 84’de gösterildiği gibi hesaplanmaktadır.

$$\begin{aligned}
 P(l^t(T_j^*) > C) &\approx \sum_{m=1}^M P(l^t(T_j^*) > C) p(\tau_j^{(m)}, \beta^{(m)} | t_j^*) \\
 &= \sum_{m=1}^M \Phi(l^t(T_j^*) > C) p(\tau_j^{(m)}, \beta^{(m)} | t_j^*)
 \end{aligned} \tag{84}$$

Eşitlik 84’de, m; MCMC yineleme sayısını göstermektedir. Her yinelemede örneklem model parametrelerinden test istatistiği hesaplanmaktadır. Ortalama sonsal olasılık, C kritik değerinden daha büyük belirlenen test istatistiklerinin marjinal sonsal olasılık değerlerine yakındır. Bir tepki süresi örüntüsünün bir eşik değeri (C) ile tanımlanan kritik bir alanda olup olmadığının test edilmesinin yanı sıra, model altında sağ kuyruk alanı olasılığını hesaplayarak gözlemlenen tepki süresi örüntüsünün uygunluk derecesinin ölçülmesi de mümkündür. Bu sağ kuyruk olasılığı, model altında gözlemlenen daha uç bir istatistik değerin sonsal olasılığını temsil etmektedir. Tahmini istatistik değeri, hataların karelerinin toplamından oluşturulmakta ve uç istatistik değeri, tepki süresi örüntülerinin lognormal model altında üretilmesinin mümkün olmadığını göstermektedir. Tepki süresi örüntülerinin uygun olmadığına dair karar hem test istatistiğinin değerinden hem de sonsal dağılımın belirsizliğinden etkilenmektedir. Test istatistiğinin dağılımı oldukça düz olduğunda, yüksek sonsal dağılımlı tepki örüntülerinin

yüksek tepeli dağılıma kıyasla uç değer olduğu sonucuna ulaşmak daha düşük olasılıklıdır. Test istatistiği doğrudan Z_{ji} 'nin tanımından gelen model parametrelerine bağlıdır. Gözlenen tepki sürelerinden daha uç değer olarak gözlemlenen tepki sürelerinin marjinal sonsal olasılık değerlerini hesaplamak için tüm model parametreleri üzerinde Eşitlik 85'de gösterildiği gibi bir entegrasyon yapılması gerekmektedir.

$$P(l^t(T_j^*) > C) = \int_{\beta} \int_{\tau_j} P(l^t(T_j^*) > C | \tau_j, \beta) p(\tau_j, \beta) d\tau_j d\beta \quad (85)$$

Marjinal sonsal olasılık, tüm model parametrelerinin entegre edilmesiyle elde edilmektedir. MCMC, model parametrelerinin sonsal dağılımından çekimler elde etmek için kullanılabilen ve MCMC yinelemesinde, model parametrelerinin sonsal dağılımından her çekimde hesaplanan istatistik değerinin eşik değeri olan C' nin üstünde olup olmadığı hesaplanabilmektedir. MCMC yenilemeleri üzerindeki ortalama sonsal olasılık değeri, Eşitlik 83'de belirtilen marjinal sonsal olasılık değerinin kestirimini ifade etmektedir. Eşitlik 85'de, test istatistiğinin dağılımı kestirilerek bu test istatistiği, “önsel yordayıcı test” olarak değerlendirilmektedir. Model parametreleri için önsel dağılımların göz önünde bulundurulması, gözlenen istatistik değerinin nasıl uç değer olarak ele alınacağı noktasında yardımcı olmaktadır. Önsel yordayıcı testi genellikle sonsal kestirim değerlendirmesinde, verinin iki defa kullanımının kestirilen kuyruk alanı olasılıklarında yanlışlık oluşturmamasından kaynaklı tercih edilmektedir. Veriler, model parametrelerini tahmin etmek ve test istatistiğinin dağılımını değerlendirmek için kullanıldığında, kuyruk alanı olasılıkları genellikle eşit olarak dağılmamaktadır. Bu durum, kestirilen olasılıkları yorumlamayı zorlaştırmaktadır. Önsel yordayıcı değerlendirme yaklaşımında, Eşitlik 84 ve 85'de belirtildiği gibi verilerin birden fazla kullanımından kaçınıldığı durumlarda kuyruk alanı olasılık kestirimleri doğru bir şekilde yorumlanabilmektedir. Bir tepki süresi örüntüsünün, uç değer olup olmadığını değerlendirmek için, test istatistiğinin değerine göre bir sınıflandırma yapılmaktadır. İstatistiğin sıfır dağılımı, bilinmeyen model parametrelerine bağlı olup bu dağılımın kesin veya doğru bir yaklaşımı bilinmektedir. Test istatistiğinin, parametre kestirimleri hesaba katıldığında, karşılık gelen kuyruk alanı olasılığı yanlış olabilmektedir. Bu nedenle, her bir MCMC yinelemesinde bir madde tepki süresi örüntüsünün uç değer olarak işaretlenme olasılığı değerlendirilmektedir. Böylelikle, ikili sınıflandırma değişkeni olan F_j^t 'nin, testi alan j.birey için verilen birey ve madde parametreleri değerlerinde aldığı değerler, Eşitlik 86 ile gösterilmiştir.

$$F_j^t = \begin{cases} 1 & , \quad P(l^t(T_j^*) > l^t(t_p^*)) < \alpha \\ 0 & , \quad P(l^t(T_j^*) > l^t(t_p^*)) \geq \alpha \end{cases} \quad (86)$$

Testi alan J tane birey için uç değer olarak işaretlenecek madde tepki süresi örüntülerinin marjinal sonsal olasılık dağılımı, Eşitlik 87 ile gösterilmiştir.

$$\begin{aligned} P(F_j^t = 1|t_j^*) &= \int_{\beta} \int_{\tau_j} I(F_j^t \\ &= 1|t_j^*, \tau_j, \beta) p(\tau_j, \lambda) d\tau_j d\beta \approx \sum_{m=1}^M I(F_j^{t^{(m)}} = 1|\tau_j^{(m)}, \beta^{(m)}) \\ &/M \end{aligned} \quad (87)$$

Eşitlik 87’de, MCMC için m yineleme sayısında $F_j^{t^{(m)}} = 1$ değerini $P(\chi^2 > l^t(t_j^*)|\tau_j^{(m)}, \beta^{(m)}) < \alpha$ olduğu durumda almaktadır. Bu nedenle, bir tepki süresi örüntüsünün uç değer olarak işaretlenme olasılığı, her yinelemede değerlendirilmektedir. Yinelemelere göre ortalama olasılık, işaretli bir tepki süresi örüntüsünün marjinal olasılığına yaklaşmaktadır. Eşitlik 87 ile hesaplanan lognormal model altında işaretlenen tepki süresi örüntülerinin sonsal olasılık değerleri ile uç değer dereceleri belirlenmektedir. Sonsal olasılık değerleri 0.95 ve üzeri olan madde tepki süresi örüntülerinin uç değer olarak işaretlenmesi, tepki süresi örüntülerini seçilen anlamlılık düzeyine (α) göre sınıflandırmak yerine her bir tepki süresi örüntüsünün sonsal olasılığının uç değer açısından değerlendirilmesine olanak sağladığından Tip 1 hata oranını azaltmaktadır (Marianti vd., 2014; Fox ve Marianti, 2017). Eşitlik 84 ile tanımlanan, uç değer olup olmadığı değerlendirilen tepki süresi örüntülerinin sonsal olasılıklarının hesaplanması, sonsal yordayıcı yaklaşım kullanılarak tanımlanabilmektedir. Model parametreleri göz önüne alındığında, test istatistiğinin sonsal olasılık değeri örneklem dağılımına göre değerlendirilmektedir. İstatistiğin dağılımının bilinmediği durumlarda, verilerin sonsal yordayıcı dağılımı test istatistiğinin dağılımını değerlendirmek için kullanılabilir. Bu durumda, kestirilen test istatistiğinin uç değer olup olmadığının değerlendirilmesi, Eşitlik 88’de gösterildiği gibi verilerin sonsal yordayıcı dağılımları kullanılarak yapılmaktadır.

$$\begin{aligned}
P\left(l^t(T_j^{*rep}) > l^t(t_j^*)\right) \\
= \int_{t_j^{*rep}} P\left(l^t(T_j^{*rep}) > l^t(t_j^*)\right) p(T_j^{*rep} | \tau_j, \beta) dT_j^{*rep}
\end{aligned} \quad (88)$$

Eşitlik 88’de, T_j^{*rep} , model altında yinelenen veriyi, eşitliğin sol tarafı ise gözlenen verilerden elde edilen istatistik değerinden daha büyük değer olan gözlenen istatistik değerinin sonsal yordanan olasılığını ifade etmektedir. Modellerin uyumunu değerlendirmek için birçok farklı uygulamada arka tahmin testleri önerilmiştir. Rubin (1984), diğerleri arasında, modelin verilerle uyumluluğunu değerlendirmek için posterior öngörücü değerlendirmenin kullanılmasını savunmuştur. Box (1980), önsel yordayıcı değerlendirmeleri bilinen modelin uyumunu değerlendirmek için verilerin marjinal yordayıcı dağılımının kullanılmasını, van der Linden ve Guo (2008), Boughton ve diğerleri (2017) ve Qian ve diğerleri (2016) çalışmalarında, tepki sürelerinin değerlendirilmesinde sonsal yordayıcı dağılımın kullanılmasını önermişlerdir.

χ_{pf} Birey Uyum İstatistiği

van der Linden (2006)’in lognormal tepki süresi Eşitlik 89 ile ifade edilmektedir.

$$\ln y_{ij} \sim N\left(\beta_i - \tau_j, \frac{1}{\alpha_i^2}\right) \quad (89)$$

Eşitlik 89’da, $\ln(y_{ij})$; j.bireyin i.madde için gözlenen tepki süresinin doğal logaritmasını ifade etmektedir. van der Linden (2006), gizil hız parametresinin maximum olabilirlik kestirimini (MLE), Eşitlik 90’da gösterildiği şekilde ifade etmektedir.

$$\hat{\tau}_j = \frac{\sum_i \alpha_i^2 (\beta_i - \ln(y_{ij}))}{\sum_i \alpha_i^2} \quad (90)$$

Sinharay (2018), madde ön bilgisi kaynaklı test hilesi çalışmalarında uygun olmayan tepki sürelerini belirlemek için, I tane madde için I-1 serbestlik derecesinde χ^2

dağılımlı madde tepkileri ile log-normal madde tepki süresi modelinin uyum sağladığı durumda, Eşitlik 91 ile ifade edilen χ_{pf} birey uyum istatistiğini önermiştir.

$$\chi_{pf} = \sum_i \alpha_i^2 (\ln y_{ij} - \beta_i + \hat{\tau}_j)^2 \quad (91)$$

Marianti ve diğerleri (2014) ve Fox ve Marianti (2017)'nin uygun olmayan madde tepki süresi örüntülerini belirlemede önerdikleri Bayesci yaklaşım ile χ_{pf} birey uyum istatistiğinin performansının benzer olduğu görülmektedir.

Λ_s İstatistiği

Toton ve Maynes (2019), madde tepkilerinin herhangi bir tepki modeline uyumunun gerektirmediği ve bireylerin tepki sürelerinin karşılaştırılmasının yani uygun olmayan tepki sürelerinin belirlenmesinin, madde ön bilgisine sahip olmayan bütün bireylerin (birey grupları) maddeyi doğru veya yanlış cevaplamasından bağımsız olarak madde üzerindeki ortalama tepki süreleri üzerinden yapıldığı bir yaklaşım önermişlerdir. Çalışmalarında, gerçek veri seti kullanarak madde ön bilgisine sahip olan bireylerin teşhir olan maddeye daha hızlı tepki verdiğini göstermişlerdir. Sonuç olarak, madde ön bilgisine sahip olan bireylerin hız parametreleri (τ_j), orjinal hız parametrelerine eşit olmamakla birlikte teşhir olan maddelerden kestirilen yetenek düzeyinin, teşhir olmayan maddelerden kestirilen yetenek düzeyinden daha yüksek olduğu görülmektedir (Sinharay, 2017a). Böylelikle, bireylerin, madde ön bilgisine sahip olup olmama durumunu ortaya çıkarmak için, teşhir olan ve olmayan maddeler üzerindeki hız parametresini ayrı incelemek faydalı olabilmektedir (Sinharay, 2020). Tepki süresi modeli kullanılarak madde ön bilgisini belirlemenin bir yolu, olabilirlik oran testi (*LTR*) kullanarak $H_0: \tau_c = \tau_{\bar{c}}$ sıfır hipotezini, $H_1: \tau_c > \tau_{\bar{c}}$ alternatif hipotezine karşı test etmektir (Casella ve Berger, 2002; Cox ve Hinkley, 1974; Lehmann ve Romano, 2005; Rao, 1973). Hipotezleri test etmede kullanılan *LRT* istatistiği Eşitlik 92'de ifade edilmiştir.

$$\Lambda = 2[\ell(\hat{\tau}_c; y_i, i \in c) + \ell(\hat{\tau}_{\bar{c}}; y_i, i \in \bar{c}) - \ell(\hat{\tau}; y_i, i = 1, 2, \dots, I)] \quad (92)$$

Eşitlik 92’de, $\hat{t}_c$; bireyin, teşhir olan maddelerden elde edilen gerçek hız parametresini, $\hat{t}_{\bar{c}}$; bireyin, teşhir olmayan maddelerden elde edilen gerçek hız parametresini ve $\ell(\hat{t}_c; y_i, i \in c)$ bireyin, $\hat{t}_c$ ’de hesaplanan, teşhir olan maddeler üzerindeki log-tepki sürelerinin log olabilirliğini ifade etmektedir. Log-normal madde tepki süresi modeli (van der Linden, 2006) için $\ell(\hat{t}_c; y_i, i \in c)$, Eşitlik 93’de gösterildiği gibi hesaplanmaktadır.

$$\begin{aligned}
\ell(\hat{t}_c; y_i, i \in c) &= \sum_{i \in c} \left[-\frac{1}{2} \log(2\pi) + \log(\alpha_i) - \frac{\alpha_i^2}{2} (y_i - \beta_i + \hat{t}_c)^2 \right] \\
&= \sum_{i \in c} \left[-\frac{1}{2} \log[2\pi] + \log(\alpha_i) - \frac{\alpha_i^2}{2} \hat{t}_c^2 - \frac{\alpha_i^2}{2} (y_i - \beta_i)^2 + \hat{t}_c \alpha_i^2 (y_i - \beta_i) \right] \\
&= \sum_{i \in c} \left[-\frac{1}{2} \log[2\pi] + \log(\alpha_i) \right] - \hat{t}_c^2 \sum_{i \in c} \frac{\alpha_i^2}{2} - \sum_{i \in c} \frac{\alpha_i^2}{2} (y_i - \beta_i)^2 \\
&\quad + \hat{t}_c \sum_{i \in c} \alpha_i^2 (y_i - \beta_i) \\
&= \sum_{i \in c} \left[-\frac{1}{2} \log[2\pi] + \log(\alpha_i) \right] - \hat{t}_c^2 \sum_{i \in c} \frac{\alpha_i^2}{2} - \sum_{i \in c} \frac{\alpha_i^2}{2} (y_i - \beta_i)^2 \\
&\quad + \hat{t}_c^2 \sum_{i \in c} \alpha_i^2 \\
&= \sum_{i \in c} \left[-\frac{1}{2} \log[2\pi] + \log(\alpha_i) \right] + \hat{t}_c^2 \sum_{i \in c} \frac{\alpha_i^2}{2} - \sum_{i \in c} \frac{\alpha_i^2}{2} (y_i - \beta_i)^2 \quad (93)
\end{aligned}$$

Eşitlik 92’de, Eşitlik 93 ile ifade edilen $\hat{t}_c$ ’de hesaplanan, teşhir olan maddeler üzerindeki log-tepki sürelerinin log olabilirliği yerine konulduğunda, *LRT* istatistiği için Eşitlik 94 elde edilmektedir.

$$\Lambda = \hat{t}_c^2 \sum_{i \in c} \alpha_i^2 + \hat{t}_{\bar{c}}^2 \sum_{i \in \bar{c}} \alpha_i^2 - \hat{t}^2 \sum_{i=1}^I \alpha_i^2 \quad (94)$$

Sinharay (2020), teşhir olan maddelerin bilindiği durumda, bireylerin, teşhir olan maddeler ve teşhir olmayan maddeler arasındaki hız farkından Eşitlik 93’de gösterildiği şekilde hesaplanan sıfır hipotez ve alternatif hipotezlerin test edildiği durumda kullanılmak üzere Eşitlik 95 ile ifade edilen Λ_s istatistiğini önermiştir.

$$\Lambda_s = \begin{cases} \sqrt{\Lambda}, & \tau_c \geq \tau_{\bar{c}} \\ -\sqrt{\Lambda}, & \tau_c < \tau_{\bar{c}} \end{cases} \quad (95)$$

Sinharay(2020), Λ_s istatistiğinin, testteki maddelerin hepsinin ya da tamamına yakın oranda teşhir olduğu durumda uygun olmayan madde tepki sürelerinin belirlenmesinde yetersiz olduğunu ve bu istatistiğin performansının, teşhir olan maddelerden kestirilen birey hız parametresinin varyansının büyük olmasından kaynaklı olarak teşhir olan madde sayısının az olduğu durumlarda yüksek olduğunu belirtmiştir.

Madde Tepkilerinin ve Madde Tepki Sürelerinin Kullanıldığı Karışım Modelleri

Beklenmeyen tepki davranışlarını belirlemek için kullanılan bir başka yaklaşım da madde ön bilgisine sahip olan ve olmayan bireylerin ayırt edilmesini sağlayan bilgilerin kullanıldığı karışım modellemeleridir (Lee, 2018). Madde tepki verileri için bir karışım modeli yaklaşımı, evrenin gizil gruplardan oluştuğu ve aynı tepki modeli altındaki her bir gizil grubun farklılığının, model parametrelerinin farklı değerler almasına (Rost, 1990; von Davier ve Rost, 2007) ya da her bir gizil grubun farklı tepki modellerine sahip olmasına (von Davier ve Yamamoto, 2006; Yamamoto,1989) izin verilerek oluşturulmasıdır. Karışım modellerinin kullanımı, testi alan bireylerin farklı tepki davranışlarını yansıtan bir modelleme olmasından dolayı eğitim ve psikoloji test uygulamalarında gittikçe artmaktadır. Farklı tepki davranışları sergileyen bireylerin (hızlı tahmin, test hızı, test hilesi (cevap kopyalama, madde ön bilgisi, vb.)) yer aldığı gruplarda, normal tepki davranışını üstlenen geleneksel psikometrik modeller, uygun olmayan tepki davranışlarını açıklamada yetersiz kalabilmektedir. Böylelikle, karışım modeli gibi psikometrik yaklaşımların tanımlanması ve farklılaştırılmasına gerek duyulmaktadır (Klein Entink, Fox ve van der Linden, 2009; Lee,2018).

Rost (1990) tarafından madde tepkileri için geliştirilen karışım Rasch modeli, karışım modeline genişletilen Rasch modelini ifade etmektedir. Karışım Rasch modelde, her bir gizil gruptaki madde tepkileri, farklı bir Rasch modeli altında oluşturularak model parametrelerinin gizil gruplarda farklı olmasına izin verilmektedir. Meyer (2010)

çalışmasında, madde tepkilerinin karışım Rasch modeli ile modellendiği, madde tepki sürelerinin ise karışım lognormal modeli altında modellendiği bir yaklaşım önererek bireyleri, madde çözüm ve hızlı tahmin davranışlarına göre sınıflandırmıştır. Karışım Rasch modelinde, madde tepkileri, Bernoulli dağılımı kullanılarak oluşturulmaktadır. Modelde, maddenin doğru yanıtlanma olasılığı Eşitlik 96'da gösterildiği gibi hesaplanmaktadır.

$$P(U_{ij} = 1 | \theta_{jg}, b_{ig}, g_j) = \frac{\exp(\theta_{jg} - b_{ig})}{1 + \exp(\theta_{jg} - b_{ig})} \quad (96)$$

Eşitlik 96'da, b_{ig} ; g.gizil gruptaki i maddesi için madde güçlük parametresini, θ_{jg} ise, g. gizil gruptaki j bireyinin yetenek parametresini göstermektedir. Modelde, ölçek tanımlama problemi her bir gizil grupta madde güçlük parametrelerinin toplamı 0'a eşitlenerek çözülebilmektedir. Tepki süreleri lognormal dağılımdan oluşturularak tepki süreleri için karışım lognormal modeli ise Eşitlik 97'de gösterildiği şekilde ifade edilmektedir.

$$f(t_{ij} | \lambda_{ig}, \tau_{ig}, g_j) = \sqrt{\frac{\tau_{ig}}{2\pi}} \frac{1}{t_{ij}} \exp\left(-\frac{\tau_{ig}}{2} (\ln t_{ij} - \lambda_{ig})^2\right) \quad (97)$$

Eşitlik 97'de, t_{ij} ; j. bireyin i.maddedeki tepki süresini, λ_{ig} ; g. gizil grubun i. madde için ortalama lognormal tepki süresini ve τ_{ig} ise g. gizil grubun i.madde için kestirilen lognormal tepki süresini ifade etmektedir. Karışım modellemesinde, madde tepkileri ile tepki süreleri arasındaki ilişki dikkate alınmamakla birlikte her bir gizil grup için madde tepkilerinin ve madde tepki sürelerinin koşullu bağımsız olduğu varsayılmakta ve madde tepkileri, Rasch modeldeki parametreleri, tepki süreleri ise sadece lognormal tepki süresi modelinin parametrelerini kestirmek için kullanılmaktadır (Lee,2018).

Klein Entink ve diğerleri (2009), çalışmalarında madde tepkileri ve madde tepki süresi bilgilerini kullanarak testi alan bireyler ve gruplar arasındaki hız ve doğruluk arasındaki ilişkiyi incelemek için modele, ortak değişkenlerin çok değişkenli çok seviyeli regresyon yapısını (çok değişkenli karışım modeli) genişletmişlerdir. Model parametrelerinin kestirimi, Markov zinciri Monte Carlo (MCMC) hesaplaması ile

bayesyan yaklaşımı ile gerçekleştirilmiştir. Çalışmada, MCMC'nin yan ürünleri olarak kolayca hesaplanabilen bayes faktörünün modele özgü uygulamaları (BF) ve model seçimi için sapma bilgisi kriteri (DIC) önerilerek hem simülasyon çalışmalarından elde edilen sonuçlar hem de gerçek veri uygulamaları ile bu modellemenin uygulanabilir olduğu gösterilmiştir. Ayrıca araştırmacılar, testi alan bireylerin, farklı madde çözüm stratejilerine göre farklı gizil sınıflara ayrıldığı ya da maddelerin farklı şekillerde sıralandığı (güçlük düzeylerine göre kolaydan zora, zordan kolay ve random olarak veya konularına göre vb.) ve kitapçık halinde sunulduğu durumlarda, hız ve doğruluk ilişkisinin farklılaşmasının beklenmesinden kaynaklı çok değişkenli çoklu karışım modelinin uygulanabilir olduğunu belirtmişlerdir. Wang ve Xu (2015) çalışmalarında, bireylerin madde çözüm davranışını araştırmak için madde tepkileri için 3PLM ve madde tepki süreleri için lognormal madde tepki süresi modelinin birlikte ele alındığını karışım modelini (WX) kullanmışlardır. Wang, Xu ve Shang (2018) çalışmalarında ise WX modelini, ilk aşamada, hızlı tahmin ve test hilesi davranışını madde çözüm davranışından ayırmak için WX modelini kullandıkları, ikinci aşamada ise madde tepkilerinin kesme puanı aracılığıyla sınıflandırılarak test hilesinin hangi tepki davranışına dahil olduğunun kararının verildiği iki aşamalı modele genişletmişlerdir.

Lee (2018) çalışmasında, testi alan bireyler arasında farklı tepki davranışları gösteren iki veya daha fazla alt grup bulunduğu durumda tepki davranışlarını ayırt edebilmek için van der Linden'in (2007) madde tepkilerinin ve madde tepki sürelerinin hiyerarşik olarak modellendiği katışık bir istatistiksel yaklaşımı, karışım Rasch lognormal tepki süresi modeline genişleterek madde ön bilgisine sahip olan bireyleri belirlemek için kullanmıştır. Çalışmada, simülasyon verileri kullanılarak madde ön bilgisinden kaynaklı test hilesini belirlemek için çeşitli koşullar altında (test uzunluğu (25 ve 50 madde), örneklem büyüklüğü (500, 1000), teşhir olan maddelerin oranı (%20, %40, %60 ve %80), teşhir olan maddelerden faydalanan bireylerin oranı (%10, %20, %50 ve %70)) karışım Rasch lognormal tepki süresinin performansı araştırılarak yöntemin performansı, diğer karışım modelleri (karışım Rasch model(MRM), karışım lognormal tepki süresi modeli (MixLn-RT), karışım Rasch model-karışım lognormal tepki süresi modeli (Meyer, 2010; MRM-RT) ile karşılaştırılmıştır. Ayrıca, yöntemin madde ön bilgisinden kaynaklı test hilesini belirlemeye yönelik performansı gerçek veri ile uygulama yapılarak gösterilmiştir. Çalışmada, bütün koşullarda, MixRL and MixLn-RT modellerinin madde ön bilgisi kaynaklı test hilesini belirleme performansının iyi olduğu ve madde ön bilgisine sahip olan ve dürüst bireyleri birbirinden ayırmada madde tepkilerinin ve madde tepki

sürelerinin birlikte kullanılmasının daha etkili olduğu sonucuna ulaşılmıştır. Ayrıca araştırmacılar, madde tepkilerinin ve madde tepki sürelerinin eşzamanlı olarak birlikte ele alındığı karışım modellemelerinin genellikle test hızından, madde çözüm davranışından ve hızlı tahminden kaynaklı beklenmeyen madde tepki davranışlarının araştırılmasında (Wise ve DeMars, 2006; Wang ve Xu, 2015; Wang, Xu ve Shang, 2018) kullanıldığını bu nedenle, madde ön bilgisinden kaynaklı beklenmeyen madde tepki davranışlarını ortaya çıkarmak için bu karışım modelleri üzerine daha fazla araştırma yapılması gerektiğini belirtmişlerdir.

DG-LNRTM (Deterministic Gated Lognormal Response Time Model)

Kasli ve Zopluoğlu (2018) , çalışmalarında log-normal madde tepki süresi modelini kullanarak madde ön bilgisine sahip olan ve olmayan bireylerin uygun olmayan tepki örüntülerinin benzerliğini belirlemeye çalışmışlardır. Benzer bir çalışmayı Kasli, Zopluoğlu ve Toton (2020), DG-LNRT (deterministic gated log-normal tepki süresi modeli) kullanarak yapmışlardır. DG-LNRT, Deterministic Gated Madde Tepki Teorisi Modelinin (DGIRTM; Shu, Henson ve Luecht, 2013) ve Lognormal Tepki Süresi Modelinin (LNRT; van der Linden, 2007) kombine edilmiş bir versiyonu olarak madde ön bilgisine sahip olan bireyleri belirlemek için önerilen tepki süresi modelidir. DGIRTM, testi alan bireyleri teşhir olan ve olmayan maddelere göre iki gruba ayırarak her birey için gerçek ve hile yeteneği olmak üzere iki yetenek düzeyi belirlemektedir. DG-LNRTM, yapısal olarak küçük fark ile DGIRTM'ye benzemekle birlikte gözlemlenen madde tepkilerinden elde edilen gizil yetenek parametresi yerine gözlemlenen madde tepki sürelerinden elde edilen gizil hız parametresini kullanmaktadır. DG-LNRTM, testi alan bireyleri teşhir olan ve olmayan maddelere göre iki gruba ayırarak her birey için iki gizil hız parametresi belirlemektedir. DG-LNRT model Eşitlik 98'de gösterildiği gibi tanımlanmaktadır.

$$\ln(y_{ij}) | \tau_t, \tau_c, \beta, \alpha, T, C \sim N(\mu_{ij}, \sigma_i^2) ,$$

$$\mu_{ij} = (\beta_i - \tau_{tj})^{1-T_j} \times [(1 - C_i) \times (\beta_i - \tau_{tj}) + C_i \times (\beta_i - \tau_{cj})]^{T_j}, \quad \sigma_i^2 = \frac{1}{\sigma_i^2} \quad (98)$$

Eşitlik 98 ile gösterilen DG-LNRT modelde, $\ln(y_{ij})$; j.bireyin i.madde üzerinde gözlenen tepki süresinin doğal logaritmasını, $\beta = \{\beta_1, \beta_2, \dots, \beta_I\}$; I tane madde için zaman yoğunluk parametresi vektörünü, $\alpha = \{\alpha_1, \alpha_1, \dots, \alpha_I\}$; I tane madde için zaman ayıricılık parametresi vektörünü, $\tau_t = \{\tau_{t1}, \tau_{t1}, \dots, \tau_{tJ}\}$ teşhir olmayan maddeleri cevapladığında J tane birey için gizil hız parametresi vektörünü, $\tau_c = \{\tau_{c1}, \tau_{c2}, \dots, \tau_{cJ}\}$ teşhir olan maddeleri cevapladığında J tane birey için gizil hız parametresi vektörünü, $T = \{T_1, T_2, \dots, T_J\}$; j.birey için madde ön bilgisine sahip olma durumunu gösterge değişkenini (0:dürüst birey, 1:hile yapan birey) ve $C = \{C_1, C_2, \dots, C_I\}$; i.madde için teşhir olma durumunu gösterge değişkenini (0:teşhir olmayan madde, 1:teşhir olan madde) göstermektedir. DG-LNRT model, Eşitlik 99’da gösterildiği gibi 4 koşullu modele ayrılabilir.

$$\ln(y_{ij}) | \tau_t, \tau_c, \beta, \alpha, T, C \sim \begin{cases} N\left(\beta_i - \tau_{tj}, \frac{1}{\sigma_i^2}\right), & (T_j = 0, C_i = 0) \\ N\left(\beta_i - \tau_{tj}, \frac{1}{\sigma_i^2}\right), & (T_j = 0, C_i = 1) \\ N\left(\beta_i - \tau_{tj}, \frac{1}{\sigma_i^2}\right), & (T_j = 1, C_i = 0) \\ N\left(\beta_i - \tau_{cj}, \frac{1}{\sigma_i^2}\right), & (T_j = 1, C_i = 1) \end{cases} \quad (99)$$

Eşitlik 99’da, τ_{tj} ; teşhir olmayan maddeyi cevaplayan j.bireyin gizil hız parametresini, τ_{cj} ; teşhir olan maddeyi cevaplayan j.bireyin gizil hız parametresini, α_i ve β_i , i.madde için zaman ayıricılık ve zaman yoğunluk parametresini göstermektedir. C_i , i. maddenin teşhir olma durumunun göstergesi olup $C_i = 1$ değeri i.maddenin teşhir olduğu durumu, T_j ; j.birey için madde ön bilgisine sahip olma durumunun göstergesi olup $\tau_{cj} > \tau_{tj}$ olduğu durumda $T_j = 1$ değeri bireyin madde ön bilgisine sahip olduğu, $\tau_{tj} > \tau_{cj}$ olduğu durumda $T_j = 0$ değeri bireyin madde ön bilgisine sahip olmadığı durumu göstermektedir.

Kasli, Zopluoğlu ve Toton (2020) çalışmalarında, madde ön bilgisi durumu manipule edilmiş gerçek veri ile simülasyon veri kullanarak DG-LNRT modelinin çeşitli koşullar altında performansını araştırmışlardır. Çalışmanın sonucunda, önerilen modelin uygulanabilir olduğu ve teşhir olan ve olmayan maddeler arasındaki tepki süresi farkının kullanılmasının test hilesinin araştırılmasında faydalı olduğuna ulaşılmıştır.

İlgili Araştırmalar

İlgili literatürde yer alan, HMTK-TS modelleme aracılığıyla, test hilesini belirlemede madde tepkilerinin ve madde tepki süresi örüntülerinin uygunluğunun incelenmesinde kullanılan birey uyum istatistiklerinin ve Kullback-Leibler yaklaşımının yer aldığı çalışmalar kronolojik sıraya göre kısaca şöyle özetlenebilir:

van der Linden ve van Krimpen-Stoop (2003) çalışmalarında, bireye uyarlanmış test uygulamalarında, madde ön bilgisi ve hızlılık (speededness) olmak üzere iki çeşit beklenmeyen madde tepki davranışından kaynaklı uygun olmayan madde tepki örüntülerini belirlemek için tepki süresinin log-normal dağılıma sahip bir değişken olarak ele alındığı loglineer tepki süresi geliştirip, bireysel tepki süresi vektörlerinin uyumunu bayesci yaklaşımlar ile değerlendirmişlerdir. van der Linden ve Guo (2008), van der Linden ve van Krimpen-Stoop (2003)'un çalışmasını, bireysel tepki süresi vektörlerinin uyumunun değerlendirilmesi diğer bir deyişle uygun olmayan madde tepki süresi örüntülerinin belirlenmesi için gözlenen tepki süreleri ile beklenen tepki süreleri karşılaştırmalarını lognormal tepki süresi modellemesi ve HMTK-TS modellemesi altında, bayesci artık kontrol kullanarak genişletmişlerdir. Çalışmada, öncelikle, HMTK-TS modellemesini, bireyselleştirilmiş test olarak uygulanan mezuniyet kabul testinin (GMAT) nicel kısmından elde edilen verilere uygulamışlardır. Çalışmanın sonucunda, gözlemlenen log- tepki sürelerinin sol-kuyruk sonsal yordayıcı olasılıklarının kümülatif dağılımının beklenenlerle tutarlı olduğu bulunmuştur. Ancak, hız parametresinin dağılımı, yetenek parametresinin dağılımı normal dağıldığında, bimodal olma eğilimindeyken, yetenek ve hız parametreleri arasındaki ilişkiyi modellemek için iki değişkenli normal dağılım yaklaşımının uygun olmayacağı ve hız parametresi için karışık bir dağılım yaklaşımının daha uygun olacağı sonucuna ulaşılmıştır. Bu nedenle, araştırmacılar, çalışmalarında, HMTK-TS modelleme yerine lognormal tepki süresi modelini uygulamayı tercih etmişlerdir. Lognormal tepki süresi modelinin verilere uyum sağlamasıyla log tepki sürelerinin sonsal kestirici yoğunlukları üretilmiş ve tepki sürelerinin standardize edilmiş artıkları kullanılarak uygun olmayan tepki süreleri işaretlenmiştir. Sonuçlar, bu yöntemin, gözlenen tepki sürelerinin benzer olduğu durumlarda bile uygun olmayan tepkileri diğerlerinden ayırmada etkili olduğunu göstermiştir. Bununla birlikte, araştırmacılar, işaretli tepki sürelerinin, sadece gözlemlenen tepki sürelerinin beklenenlerden saptığını göstermesinden dolayı hangi tür beklenmeyen madde tepki davranışlarının hile davranışının kaynağı olduğunu araştırmak

için işaretli tepki süresi örüntülerinin ve madde tepkilerinin daha fazla incelenmesi gerektiğini önermişlerdir. Ayrıca sonuçlar, beklenmeyen madde tepki davranışının tespit oranının, test uzunluğu arttıkça ve madde tepki süresi azaldıkça arttığını göstermektedir.

Belov (2013) çalışmasında, grup düzeyinde madde ön bilgisinden kaynaklı test hilesini belirlemek için *Kullback Leibler* yaklaşımını kullanmıştır. İki aşamalı istatistiksel test süreci kullanarak ilk aşamada test hilesi yapan grubu, ikinci aşamada ise hile yapan gruplardaki bireyleri belirlemiştir. Çalışmada, grup büyüklüğü (5,10,15,20,30), gruptaki bireylerin sayısı (5,10,15,20,30) faktörleri değişimlenerek G ve G^* birey uyum istatistiklerinin hile belirleme performansları araştırılmıştır. Araştırmanın sonucunda, iki aşamalı istatistiksel test süreci kullanmanın kullanılan yöntemlerin performansını arttırdığı, grup büyüklüğünün artmasının azalttığı görülmüştür.

Marianti, Fox, Avetisyan, Veldkamp ve Tijmstra (2014) çalışmalarında, uygun olmayan madde tepki süresi örüntülerini belirlemede, l^t , l_s^t ve l_z^t birey uyum istatistiklerinin, örneklem büyüklüğü (500,1000), test uzunluğu (10,20) ve uygun olmayan madde tepki oranı (%5,%10 ve %20) faktörlerinin çeşitli koşullarında, madde alt setlerinde, random tepki sürelerinin oluşturulduğu (random tepki davranışı), test hızlılığı ya da değişen çalışma hızının olduğu (madde ön bilgisi,motivasyon eksikliği vb.) ve tepki süresi örüntülerinde tek uygun olmayan tepki süresinin oluşturulduğu durumlarda, beklenmeyen madde tepki davranışlarını belirlemedeki performanslarını araştırmışlardır. Çalışmada uygun olmayan madde tepki süreleri, beklenmeyen madde tepki davranışlarından random madde tepki davranışında, maddelerin zaman yoğunluklarından daha yavaş veya daha hızlı olacak şekilde testteki maddelerin yarısı için tepki süresi, ortalaması, ortalama madde tepki süresine ve standart sapması tepki sürelerinin ortalamasının 3 katı olacak şekilde, değişen çalışma hızı beklenmeyen madde tepki davranışında, test maddelerinin yarısının tepki süresi, lognormal model altında beklenen tepki sürelerinden daha hızlı olacak şekilde, maddelerin diğer yarısı da değişen çalışma hızına sahip beklenmeyen tepki davranışı gösteren bireylerin o maddeler üzerindeki çalışma hızının örneklemin ortalama çalışma hızının 1.5 standart sapma daha hızlı olacak şekilde, tepki süresi örüntülerinde tek uygun olmayan tepki süresinin oluşturulduğu beklenmeyen tepki davranışında ise, ortalaması testteki maddelerin maksimum zaman yoğunluğunun en az iki katı olan lognormal dağılımdan üretilmiştir. Çalışmanın sonucunda, birey uyum istatistiklerinin performansları karşılaştırıldığında, l^t , l_s^t birey uyum istatistiklerinin performanslarının benzer, l_z^t birey uyum istatistiğinin ise

sonsal olasılıkları sınıflandırmada çok konservatif olduğu ve böylelikle uygun olmayan tepki örüntülerini belirlemede kullanışlı olmadığı sonucuna ulaşılmıştır. Bununla birlikte, l^t birey uyum istatistiğinin performansının, random madde tepki davranışında, örneklem büyüklüğü arttıkça, değişen çalışma hızı madde tepki davranışında ise test uzunluğu ve uygun olmayan madde tepki örüntü oranı arttıkça; tepki süresi örüntülerinde tek uygun olmayan tepki süresinin oluşturulduğu beklenmeyen madde tepki davranışında ise test uzunluğu arttıkça ve örneklem büyüklüğü azaldıkça azaldığı görülmüştür.

Belov (2016) çalışmasında, madde ön bilgisinden kaynaklı test hilesini belirlemede kullanılan *Kullbak Leibler* yaklaşımının, *Neyman-Pearson lemma* yaklaşımının, modifiye edilmiş l_z birey uyum istatistiğinin, *DGRTM* ve *PSS* istatistiğinin, uygun olmayan tepki örüntüsüne sahip bireylerin yetenek dağılımının tipi ($N(0,1)$, $U(-3,0)$), uygun olmayan tepki örüntüsüne sahip bireylerin oranı (%5,%10,%20), alt testte yer alan şüpheli teşhir olan maddelerin belirsizlik düzeyi (%0,%25,%50), faktörlerini değişimleyerek bireye uyarlanmış ve kâğıt-kalem test uygulaması olmak üzere iki farklı test türünde performanslarını araştırmıştır. Çalışmada, performans ölçüsü olarak ROC eğrisini kullanmıştır. Çalışmanın sonucunda *DGRTM* ve *PSS* istatistiğinin bütün koşullarda madde ön bilgisinden kaynaklı test hilesini belirlemede iyi performans gösterdiği sonucuna ulaşılmıştır.

Qian, Staniewska, Reckase ve Woo (2016) çalışmalarında, bilgisayar temelli ve bireye uyarlanmış test uygulamalarında, madde ön bilgisinden kaynaklı test hilelerinde teşhir olan maddeleri ve teşhir olan maddelerden yararlanan bireyleri HMTK-TS modelini kullanarak belirlemeye çalışmışlardır. Çalışmada, madde tepkileri, bilgisayar temelli test uygulamasında 2PLM ile bireye uyarlanmış test uygulamasında ise Rasch modele uyum sağlayacak şekilde, madde tepki süreleri de lognormal tepki süresi modeli kullanılarak üretilmiştir. Ayrıca çalışmada, beklenmeyen madde tepki davranışı tespitinde ilk aşamada madde kalibrasyonu için kullanılan bir örneklem ve madde ön bilgisine sahip olan bireylerin özelliklerini gösteren diğer örneklem olmak üzere iki farklı örneklem kullanılmıştır. Her iki örneklemde de madde ön bilgisine sahip olan bireylerin ve bu bireylerin faydalandığı teşhir olan maddelerin parametre kestirimleri HMTK-TS modelleme kullanılarak belirlenmiştir. Teşhir olan maddeleri ve bu maddelerden faydalanan bireyleri belirlemek için artıklar ($e_i = \alpha_i(\ln t_{ij} - (\beta_i - \tau_j))$, $e_i \sim N(0,1)$) hesaplanmıştır. Çalışmada, artık analizler kullanılarak bilgisayar temelli test için 111 test maddesinden 2 maddenin teşhir olduğu ve 1172 bireyden 2 bireyin madde ön bilgisine

sahip olduđu, bireye uyarlanmış test için teşhir olan madde ve madde ön bilgisine sahip bireyin belirlenmediđi sonucuna varılmıştır. Böylelikle, artık analizlerinin, teşhir olan maddeleri ve bu maddelerin ön bilgisine sahip olan bireyleri belirlemede yetersiz kaldığı ve daha fazla araştırmanın yapılması gerektiđi sonucuna ulaşılmıştır.

Boughton, Smith ve Ren (2017) çalışmalarında, teşhir olan maddeleri ve madde ön bilgisine sahip olan bireyleri belirlemek için lognormal tepki süresi ve HMTK-TS modelinin uygulamalarını göstermişlerdir. Araştırmacılar, çalışmalarında, van der Linden (2008)'in çalışmasına benzer şekilde Bayesci yöntemleri kullanarak ve ek ölçüt olarak her birey için uygun olmayan madde tepki süresi gösteren maddeler sayılarak işaretlenen maddelerin dağılımları yapılandırılmıştır. Dağılımın altında, 95. yüzdelerde veya daha üzerindeki bireyler madde ön bilgisine sahip bireyler olarak değerlendirilmiştir. Benzer şekilde, teşhir olan maddeleri belirleyebilmek için her madde için işaretlenen tepki sürelerine sahip olan bireylerin sayısı hesaplanmıştır. Böylelikle, maddeler boyunca işaretlenen bireylerin dağılımı yapılandırılmıştır. Teşhir olan maddeler, dağılımın 90. yüzdeler dilimin üzerine düşenler olarak belirlenmiştir. Çalışmanın sonucunda, teşhir olma durumuna göre işaretlenen madde ve bu maddelerden faydalanan bireylerin her iki tepki süresi modeli tarafından da tespit edildiđi ve işaretlenen bireyler ve maddelerdeki farklılıkların tepki süresi modellerinin doğasındaki farklılıklardan kaynaklı olduđu belirtilmiştir. Ayrıca, bayesci yaklaşımların, van der Linden (2008)'in çalışmasında olduđu gibi madde ön bilgisinden kaynaklı beklenmeyen madde tepki davranışını belirleme dışında hızlılık ve hızlı tahmin madde tepki davranışlarını belirlemede kullanılabileceđi gösterilmiş olup maddelerin gerçek teşhir olma durumunun ve testi alan bireylerin madde ön bilgisine sahip olma durumunun bilinmesi, bayes kalıntı kontrollerinin test şirketi tarafından sağlanan bilgilerle ne ölçüde uyumlu olduđunun karşılaştırılmasını sağlamıştır.

Fox ve Mariani (2017) çalışmalarında, madde tepkileri ve madde tepki sürelerinin katışık modellemesinde, yetenek ve hız madde karakteristik özellikleri arasındaki korelasyonu dikkate alarak, sonsal olasılıklar aracılığıyla Bayesci anlamlılık test istatistik değerlerindeki uç değerlerin belirlendiđi uygun olmayan tepki ve tepki süresi örüntülerini belirlemek için birey uyum istatistiklerinin performanslarını simülasyon çalışmaları ile çeşitli faktörler altında değerlendirip gerçek veri uygulaması ile karşılaştırmışlardır. Çalışmada her bir veri setinin madde tepkileri ve tepki sürelerini içeren, madde sayısının 20, birey sayısının ise 1000 olduđu toplamda 20 veri seti üretilmiştir. Madde parametreleri, madde ayırtedicilik parametresinin ortalaması 1, madde güçlüđü

parametresinin ortalaması 0, tepki süresi üzerindeki çalışma hızının madde özel etkisinin ortalaması 1, zaman yoğunluk parametresinin ortalaması 0 ve kovaryans matrisinin köşegen elemanları olan madde ayırtecilik parametresinin varyansı 0.05, madde güçlüğü parametresinin varyansı 1, tepki süresi üzerindeki çalışma hızının madde özel etkisinin varyansı 0.05, zaman yoğunluk parametresinin varyansı 1 ve şans parametresi ise 0.10 olacak şekilde üretilmiştir. Birey parametreleri ise, ortalamanın 0, varyansın 1 olduğu çok değişkenli normal dağılımdan elde edilerek hız ve yetenek parametreleri arasındaki kovaryans 0.75'e eşitlenmiştir. Ölçme hata varyansı ise ortalaması 0 ve standart sapması 0.20 olan lognormal dağılımdan üretilmiştir. Madde tepkileri için 3PLM'nin kullanıldığı HMTK-TS modellemenin parametre kestirimlerinde MCMC algoritması uygulanmıştır. MCMC algoritmasında, her veri seti için 10.000 yinelemede başlangıç periyodu olarak 1000 yineleme kullanılmıştır. Araştırmada, HMTK-TS modellemenin parametre kestirimlerinde beklenen ve kestirilen değerler arasındaki sapmaların yani kestirilen sonsal standart sapmaların verilerin 20 yenilemesinde 0.022 ile 0.127 arasında değiştiği gözlenmiştir. Böylelikle veriler için 20 yenileme kullanılmıştır. Araştırmacılar, çalışmada, l^t , l_s^t birey uyum istatistiklerinin, uygun olmayan tepki örüntülerinin değişimlenen koşullarında (%10,%20), düşük yetenekli bireylerin zor maddelere doğru tepkiler verdiği (cevap kopyalama), cevap kopyalayan ya da bir kaynaktan kopya çeken bireylerin maddeleri cevaplamak için yeterli süreyi kullanmadığı (random tepki sürelerinin üretildiği) ve hem cevap kopyalamanın hem de random tepki sürelerinin üretildiği üç farklı hile davranışını belirlemedeki performanslarını araştırmışlardır. Uygun olmayan tepki süreleri, ortalaması, ortalama tepki sürelerine, standart sapması da tepki sürelerinin ortalama standart sapmasının üç katı olan lognormal dağılımdan üretilmiştir. Gerçek veri uygulamasında, Cizek ve Wollack (2016)'ın tartıştığı 723 bireyin madde tepkilerinin ve madde tepki sürelerinin kaydedildiği veri seti, HMTK-TS modelleme kullanılarak analiz edilmiştir. Çalışmanın sonucunda, hem cevap kopyalamanın hem de random tepki sürelerinin üretildiği durumda, birey uyum istatistiklerinin, uygun olmayan tepki örüntülerinin yüksek oranında daha yüksek performans gösterdikleri görülmüştür.

Man, Haring, Quyang ve Thomas (2018) çalışmalarında, HMTK-TS modelleme kullanarak Mariani, Fox, Avetisyan, Velkamp ve Tijmstra (2014) tarafından önerilen parametrik $L(l^t)$ birey uyum istatistiğinin ve parametrik olmayan tepki süresi tabanlı *Kullback-Leibler* uzaklığı yaklaşımının, madde ön bilgisinden ve cevap kopyalamadan kaynaklı test hilelerini belirlemedeki etkililiğini incelemişlerdir. Çalışmada, yöntemlerin

performansı, geniş ölçekli yapılan testlerin koşullarına benzer şekilde test uzunluğu (15 ve 30 madde) ve örneklem büyüklüğü (500 ve 50.000) ve uygun olmayan madde tepki davranışı gösteren bireylerin oranı (%5,%10 ve %20) faktörleri değişimlenerek madde ön bilgisinden kaynaklı test hilesinin belirlenmesinde, teşhir olan maddelerin madde tepki süreleri, 10 ve 15 sn aralığındaki uniform dağılımdan elde edilerek, cevap kopyalamadan kaynaklı test hilesinin belirlenmesinde ise orjinal madde tepki sürelerine 30 ile 120 sn arasındaki normal dağılımdan elde edilen süreler eklenip manipüle edilerek oluşturulmasıyla farklı üç durum (uç değer gösteren 1 madde (en uzun tepki süresine sahip olan en zor madde), random 1 madde ve random olarak işaretlenen 5 madde) için Monte Carlo simülasyonu kullanılarak araştırılmıştır. Lognormal madde tepki süresi modelinin parametreleri, GRE sınavının nicel kısmının parametrelerine benzer olacak şekilde zaman ayıricılık parametresi 1 ile 3 arasındaki uniform dağılım, zaman yoğunluk parametresi ortalaması 0.45, standart sapması 0.97 olan dağılım, gizil yetenek ve gizil hız parametreleri aralarındaki korelasyon 0.3 olacak şekilde ortalaması 0 olan iki değişkenli normal dağılım kullanılarak elde edilmiştir. Çalışmada önerilen tepki süresi tabanlı parametrik olmayan *Kullback-Leibler* uzaklığı yaklaşımının, tüm koşullarda performansının daha yüksek olduğu ve bu yaklaşımın, madde ön bilgisinden kaynaklı test hilelerini belirlemede cevap kopyalamadan kaynaklı test hilesini belirlemeden daha duyarlı olduğu sonucuna ulaşılmıştır.

Balta, Uçar ve Şahin (2019) çalışmalarında, *Kullback Leibler* yaklaşımının, l_z birey uyum istatistiğinin ve önce *Kullback Leibler* ayrışım fonksiyonunun sonra l_z birey uyum istatistiğinin $(KL+l_z)$ iki aşamalı olarak kullanıldığı durumun, gözetimsiz gerçekleştirilen bireye uyarlanmış test uygulamasında çevrim içi doğrulama testi aracılığıyla madde ön bilgisinden kaynaklı test hilesini tespit etmede çeşitli koşullar altındaki performanslarını incelemişlerdir. Bu amaçla çalışmada, bireylerin gerçek yetenek düzeylerini belirlemek için bilgisayar ortamında bireye uyarlanmış test ve madde ön bilgisine sahip olan bireyleri belirlemek için kağıt-kalem testi olmak üzere iki tane paralel test kullanılmıştır. Bireye uyarlanmış test simüle edilirken, Hukuk Fakültesi Kabul Testinin (LSAT) mantıksal akıl yürütme bölümünün (LR) madde parametreleri referans alınmıştır. Böylelikle, madde tepki kuramı modellerinden 3PLM kullanılarak yeteneklerinin standart normal dağıldığı $N(0,1)$ 1000 birey için madde güçlük parametresi, ortalaması 0.49, varyansı 0.24, madde ayırtedicilik parametresi, ortalaması 0.75, varyansı 0.24, şans parametresi ortalaması 0.17, varyansı 0.25 olan dağılımdan elde edilen 500 test maddesi üretilmiştir. Test güvenliğini kontrol altına almak için bireylerin

gerçek yetenek düzeylerinin kestirilmesinde, ölçme kesinliğini yüksek düzeyde tutan $\Theta=0$ başlangıç koşulu, MFI madde seçimi yöntemi, EAP yetenek kestirimi yöntemi ve 50 sabit test uzunluklu sonlandırma kuralı gözönünde bulundurularak ve maddelerin teşhir olma durumu kontrol altına alınmayarak bireylere, bireyselleştirilmiş test uygulaması yapılmıştır. 50 maddeden oluşan kağıt- kalem testi verileri de bireye uyarlanmış test verilerindeki madde parametreleri ve yetenek kestirimleri göz önünde bulundurularak üretilmiştir. Çalışmanın ilk aşamasında, doğrulama çevrim içi testi kullanılarak madde ön bilgisinden kaynaklı test hilesi yapan bireylerin yetenek düzeyi (düşük,orta), test hilesi yapılan madde oranı(%60, %70) koşulları değiştirilerek yapılan simülasyon çalışmasında, aşamasız olarak l_z birey uyum istatistiği kullanılarak test hilesi yapan bireyler belirlenmiştir. Çalışmanın ikinci aşamasında, aynı verilerle test hilesi yapan bireyleri belirlemek için Kullback-Leibler ayrışım fonksiyon değerleri hesaplanarak, hile yaptığı belirlenen bu bireyler için tekrardan l_z birey uyum istatistik değerleri hesaplanmıştır. Çalışmanın sonucunda, l_z birey uyum istatistiğinin aşamalı olarak kullanımının madde ön bilgisinden kaynaklı test hilesini belirlemedeki performansını arttırdığı sonucuna ulaşılmıştır.

Uçar ve Doğan (2021), çalışmalarında, ROC analizinin sınıflama kuralının performansına ilişkin değerlendirmeleri, *Youden İndeksi*, *Kar-Yarar indeksi* ve *minimum puan* yaklaşımına göre yaparak elde edilen kesme puanlarını kullanarak *Kullback-Leibler* uzaklığının cevap kopyalamadan kaynaklı test hilesini belirleme performansını değerlendirmişlerdir. Çalışmada, *Youden İndeksi* ve *minimum puan* yaklaşımı ile elde edilen kesme puanlarının kullanıldığı durumda *Kullback-Leibler* ayrışım fonksiyonunun daha yüksek performans gösterdiği sonucuna ulaşılmıştır.

İlgili araştırmalar incelendiğinde, HMTK-TS modelleme aracılığıyla, madde ön bilgisinden kaynaklı test hilesini belirlemede, madde tepki ve tepki süresi örüntülerinin uygunluğunun değerlendirildiği az çalışma olduğu görülmektedir. Bununla birlikte, *Kullback-Leibler* uzaklığının performansının, çeşitli koşullar altında, madde tepki örüntülerinin uygunluğunun değerlendirilmesinde araştırıldığı ancak tepki süresi örüntülerinin uygunluğunun değerlendirilmesine yönelik yeterli düzeyde araştırılmadığı görülmektedir. Ayrıca, madde ön bilgisinden kaynaklı test hilesinin araştırılmasında, madde tepki süresi örüntüleri için $L(l^t)$ ve madde tepki örüntüleri için $L(l_s^y)$ birey uyum istatistiklerinin performansının teşhir olan maddelerin güçlük düzeyi ve tepki süresi değişimlemesi değişkenleri ele alınarak araştırılmadığı görülmektedir. Böylelikle, bu çalışmada, *Kullback-Leibler* uzaklığının, $L(l^t)$ ve $L(l_s^y)$ birey uyum istatistiklerinin

madde 6n bilgisinden kaynaklı test hilesini belirlemede, eřitli kořullar altında madde tepkileri ile birlikte tepki süresi örüntülerinin uygunluęunu tespit etmedeki performanslarının araştırılmasına karar verilmiştir.



BÖLÜM 3

YÖNTEM

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

Araştırmanın Modeli

Bu araştırmada, simülasyon verileri kullanılarak çeşitli koşullar altında, çoktan seçmeli maddelere ilişkin ön bilgiye sahip olan ve hile davranışı gösteren bireyleri tespit etmede, madde tepkilerinin ve madde tepki sürelerinin birlikte hiyerarşik olarak modellenmesi kullanılarak tepki süresi tabanlı *Kullback-Leibler* uzaklığı yaklaşımının, madde tepki süresi örüntüleri için $L(l^t)$ ve madde tepki örüntüleri için $L(l_s^y)$ birey uyum istatistiklerinin performansları hakkında bilgi elde edinilmeye çalışılmıştır. Bu çerçevede, mevcut araştırma, 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 araştırmada incelenen bağımlı değişkenler; madde ön bilgisinden kaynaklı test hile davranışını belirlemede kullanılan yöntemlerin 1.Tip hata oranı ve bu yöntemlerin güç oranıdır. Araştırmanın bağımsız değişkenleri; madde ön bilgisinden kaynaklı test hile

davranışını belirleme yöntemleri, test uzunluğu, teşhir olan madde oranı, teşhir olan maddelerin güçlük düzeyleri ve teşhir olan maddelerin tepki süresi değişimleridir. Araştırmanın sabit değişkenleri ise örneklem büyüklüğü ve madde ön bilgisi kaynaklı hile yapan birey oranıdır. İfade edilen değişkenler ve düzeyleri, ilgili literatürde yer alan benzer çalışmalardan elde edilen bulgular ile geniş ölçekli sınav uygulamaları dikkate alınarak belirlenmiştir.

Sabit Değişkenler

Örneklem Büyüklüğü. Bilgisayar temelli ve bireye uyarlanmış test uygulamalarında, madde ön bilgisinden kaynaklı test hilelerinin, madde tepki süresi modellemeleri aracılığıyla belirlendiği ilgili çalışmalar (Fox ve Marianti, 2017; Marianti ve diğerleri, 2014; Lee, 2018; Shu, 2011; van der Linden ve Guo, 2008; van der Linden ve Krimpen-Stoop, 2003) incelendiğinde, örneklem büyüklüğünün, 500, 1000, 10.000 ve 50.000 olarak değişimlendiği görülmektedir. Bu çalışmada, HMTK-TS model parametrelerinin bayesci kestirimlerinin yapıldığı MCMC algoritma yinelenmesinin uzun sürmesi nedeniyle örneklem büyüklüğü faktörü değişimlenmeyerek örneklem büyüklüğü, 1000 olacak şekilde sabitlenmiştir.

Madde Ön Bilgisi Kaynaklı Hile Davranışında Bulunan Birey Oranı. Madde ön bilgisi kaynaklı test hilesi ile ilgili çalışmalar incelendiğinde, Belov (2016) ve Fox ve Marianti (2017)'nin çalışmalarında, madde ön bilgisine sahip bireylerin tepkilerini “uygun olmayan tepki” şeklinde adlandırarak uygun olmayan tepki oranının %10 ve %20, Marianti ve diğerleri (2014)'nin çalışmalarında, %5, %10 ve %20, Karabatsos (2003)'un ve Man ve diğerleri (2018)'nin çalışmalarında ise %5, %10, %25 ve %50 olacak şekilde değişimlendiği görülmektedir. Bu çalışmada, 1000 bireyden oluşan örneklemde %5 oranında düşük yetenekli bireylerden madde ön bilgisine sahip bireyler seçilmiştir. Test hilesi davranışında bulunan bireylerin yetenek düzeyi, Sünbül ve Yormaz (2018)'in çalışmasında olduğu gibi (-3.00 -1.50) aralığında ele alınmıştır.

Manipule Edilen Değişkenler

Test Uzunluğu. Bilgisayar temelli ve bireye uyarlanmış test uygulamalarında, madde ön bilgisinden kaynaklı test hilelerinin, madde tepki süresi modellemeleri aracılığıyla belirlendiği ilgili çalışmalar (Fox ve Marianti, 2017; Marianti ve diğerleri, 2014; Lee, 2018; Shu, 2011; van der Linden ve Guo, 2008; van der Linden ve Krimpen-Stoop, 2003) incelendiğinde, test uzunluğunun, 15, 20, 25, 30 ve 50 olacak şekilde değişimlendiği görülmektedir. İlgili literatürde, test uzunluğunun, madde ön bilgisinden kaynaklı test hilesi davranışlarını belirlemede önemli bir faktör olarak görülmesi, çalışmalarda elde edilen, testteki madde sayısı arttıkça hile davranışının gözlenme olasılığının artması sonucu ve Türkiye'de, geniş ölçekli ve yüksek riskli sınav uygulamalarındaki test uzunlukları göz önünde bulundurularak test uzunluğu değişkeni, kısa ve orta test uzunluğu olacak şekilde 15 ve 50 olarak değişimlenmiştir.

Teşhir Olan Maddelerin Güçlük Düzeyi. Cevap kopyalamadan kaynaklı test hilesi ile ilgili yapılan araştırmalar incelendiğinde, teşhir olan maddelerin güçlük düzeyinin önemli bir faktör olarak ele alındığı görülmekle birlikte madde tepkilerinin ve madde tepki sürelerinin birlikte hiyerarşik modellenmesi ile madde ön bilgisinden kaynaklı test hilesini belirlemede teşhir olan maddelerin güçlük düzeylerinin değişimlenmediği görülmektedir. Altuner (2019) madde istatistikleri ile madde tepki süreleri arasındaki ilişkiyi incelediği çalışmasında, madde tepki sürelerinin maddelerin güçlük düzeyine göre değişiklik gösterdiğini belirtmektedir. Böylelikle, madde ön bilgisinden kaynaklı test hilesini belirlemeye yönelik bu çalışmada, teşhir olan maddelerin güçlük düzeyi orta güçlük ve zor olacak şekilde değişimlenmiştir. Sünbül ve Yormaz (2018) çalışmalarında, testteki maddelerin güçlük düzeylerini değişimleyerek b parametresinde kolay maddeler için (-2.50 – 0.00) aralıklarını, zor maddeler için (0.01 – 2.50) aralıklarını ele almışlardır. Bu çalışmada, b parametresinde orta güçlükteki maddeler için (-1.50 –1.50) aralıkları, zor maddeler için (1.51– 3.00) aralıkları ele alınmıştır.

Teşhir Olan Maddelerin Oranı. Zopluoğlu (2016) çalışmasında, hile yapılan teşhir olan madde oranlarını %20 (düşük), %40 (orta) ve %60 (yüksek) olacak şekilde, Wollack (2006) ve Sotaridona ve Meijer (2002) çalışmalarında, %10, %20, %30 ve %40 olarak , Lee (2018) çalışmasında, teşhir olan maddelerin oranını %20, %40, %60 ve %80,

Balta,Uçar ve Şahin (2019) çalışmalarında %60 ve %70 ve Maynes (2017) çalışmasında, %0 ve %55 aralığında %5 aralıkla artan 12 farklı oran olacak şekilde ele alınmıştır. Bu çalışmada, maddelerin %20, %40 ve %60'ının teşhir olduğu hile durumları oluşturulmuştur.

Teşhir Olan Maddelerin Tepki Süresi Değişimlemesi. Teşhir olan maddelerin zaman değişimlemesi ile ilgili olarak madde ön bilgisinden kaynaklı hile davranışlarının lognormal tepki süresi modeli kullanılarak belirlendiği simülasyon çalışmaları incelendiğinde, madde ön bilgisine sahip olmayan hipotetik dürüst bireyler ile madde ön bilgisine sahip olan hile yapan bireylerin teşhir olan maddeye verilen tepki sürelerinin nasıl simüle edileceğinin karar verilmesinin önem teşkil ettiği görülmektedir. Meijer ve Sotaridona (2006) çalışmalarında, madde ön bilgisine sahip bireyin, teşhir olan maddeye verdiği tepki süresini, orijinal tepki süresinin yarısı ve çeyreği kadar azaltarak, van der Linden ve Guo (2008) çalışmalarında, teşhir olan maddelerin tepki sürelerini 10, 20 ve 30 sn olacak şekilde sabitleyerek, Lee (2018) çalışmasında, teşhir olan maddelerin tepki sürelerini, random olacak şekilde 20 ve 30 saniye arasında uniform dağılımdan oluşturarak, Fox ve Marianti (2017) çalışmalarında, uygun olmayan madde tepki sürelerini, ortalamaları, ortalama tepki sürelerinin ortalaması, standart sapmalarının ise tepki sürelerinin ortalama standart sapmasının üç katı olan log-normal dağılımdan oluşturarak, Man ve diğerleri (2018)'nin çalışmalarında, madde ön bilgisinden kaynaklı test hilelerinin belirlenmesinde, teşhir olan maddelerin tepki sürelerini random olacak şekilde 10 ve 15 saniye arasında uniform dağılımdan, cevap kopyalamadan kaynaklı hilelerin belirlenmesinde, teşhir olan maddelerin orijinal tepki sürelerine, random olarak 30 ve 120 saniye arasında uniform dağılımdan elde edilen tepki süreleri eklenerek simüle edildiği görülmektedir. Kasli ve Zopluoğlu (2018), gerçek veri seti ile yaptığı çalışmada ise madde ön bilgisinin, bireyin örtük hız parametresini, ortalama 0.45 standart sapma yükselttiği sonucuna ulaşmıştır. Bu çalışmada, teşhir olan maddelerin tepki süreleri, random olarak 10 ve 15 saniye arasındaki uniform dağılımdan elde edilecek şekilde ve 20 saniyeye ve 30 saniyeye sabitlenenecek şekilde değişimlenmiştir. Tepki süresi değişimlemesinde bu zaman aralıklarının seçilmesi, Qian ve diğerleri (2016)'nin ve van der Linden ve Guo (2008)'nin çalışmalarında belirttiği gibi uygulamada olan geniş ölçekli sınavlardaki değerlendirmelerde, madde ön bilgisine sahip bireylerde gözlenen tepki sürelerinin göz önünde bulundurulması ve bu kısa zaman aralıklarının bireyin teşhir

olan maddeyi hatırlayıp tepkide bulunması arasında geçen süreyi ifade etmesinden kaynaklıdır.

Tablo 2’de ilgili literatürde yer alan çalışmalar göz önünde bulundurularak araştırmanın değişkenleri ve değişimlenen düzeylerine yer verilmiştir.

Tablo 2

Simülasyon Deseni

	Değişken	Düzeyler	Düzey Sayısı
Manipule Edilen	Test Uzunluğu	15/50	2
	Teşhir Olan Maddelerin Güçlük Düzeyi	Orta Güçlük/Zor	2
	Teşhir Olan Maddelerin Oranı	%20/%40/%60	3
	Teşhir Olan Maddelerin Tepki Süresi Değişimlemesi	Random Olarak 10 ile 15 saniye arasındaki uniform dağılım/ 20 saniyeye sabitleme/ 30 saniyeye sabitleme	3
Sabit	Örneklem Büyüklüğü	1000	1
	Madde Ön Bilgisi Kaynaklı Hile Davranışında Bulunan Birey Oranı	%5	1
Toplam Koşul Sayısı		2x2x3x3x1x1 =36	

Bu bilgiler ışığında, aşağıda, çalışmanın veri üretim sürecine dair ayrıntılı bilgilere yer almaktadır.

Verilerin Üretilmesi

Çalışmada değişimlenen faktörler ve düzeyleri için gerçekleştirilen veri üretimi ve verilerin manipülasyonu ile ilgili bilgilere aşağıda yer verilmiştir.

Çalışmada, yöntemlerin performansının belirlenmesine yönelik I.Tip hata oranlarını ortaya koymak için 2 (teşhir olan maddelerin güçlük düzeyi) x 2 (test uzunluğu) = 4 farklı, yöntemlerin güç oranlarını ortaya koymak için 2 (teşhir olan maddelerin güçlük düzeyi) x 3 (teşhir olan maddelerin zaman değişimlemesi) x 2 (test uzunluğu) x 3 (teşhir olan madde oranı) =36 farklı koşul düzeyi oluşturulmuştur. Çalışmada her bir koşulun düzeyi için 50 yinleme yapılarak veriler üretilmiştir. Çalışmada kullanılan koşulların düzeyleri için 50 yinleme sayısı, çalışmanın sonuçlarını Marianti ve diğerleri (2014), Fox ve Marianti (2017) ve Man ve diğerleri (2018)'nin çalışmaları ile karşılaştırma yapmak amacıyla seçilmiştir. Bu durumda, yöntemlerin I. Tip hata oranlarını ortaya koymak için $4 \times 50 = 200$ adet, hile belirleme güç oranlarını ortaya koymak için $36 \times 50 = 1800$ adet veri seti üretilmiştir. Veri setleri, R.4.0.1. programında yer alan “*LNIRT*” paketi (Fox, Klotzke ve Klein Entink, 2019) yardımıyla, madde tepkilerinin 3PLM ile madde tepki sürelerinin de bayesci log-normal madde tepki süresi ile modellendiği 1000 bireyin 15 ve 50 maddeye verdikleri tepkiler ve log-normal tepki sürelerinden oluşacak şekilde üretilmiştir. Her üretilen veri seti için HMTK-TS model parametreleri, veri setlerinin önsel dağılımından simüle edilmiştir. HMTK-TS model parametrelerinden madde parametreleri, çok değişkenli önsel dağılımları, kovaryans matrisinin köşegen elemanları $\sigma_a^2 = 0.05$, $\sigma_b^2 = 1$, $\sigma_\alpha^2 = 0.05$, $\sigma_\beta^2 = 1$ ve $\mu_a = 1$, $\mu_b = 0$, $\mu_\alpha = 1$, $\mu_\beta = 0$ olacak şekilde simüle edilmiştir. MCMC algoritmasının performansını, şans ile yanıtlama davranışının etkisini en aza indirerek araştırmak için şans parametresi $c=0.10$ olacak şekilde sabitlenmiştir. Modelde, birey parametreleri olan hız ve yetenek, ortalaması 0, varyansı 1 olan çok değişkenli normal dağılımdan elde edilmiştir. Madde tepkilerinin ve madde tepki sürelerinin birlikte hiyerarşik olarak modellenmesinde, model parametrelerinden birey hız ve yetenek parametreleri arasındaki korelasyonun, Fox, Klein Entink ve van der Linden (2007)'in ve Fox ve Marianti (2017)'nin çalışmalarında 0.3, 0.5 ve 0.75 olacak şekilde değişimlendiği ve modelde, birey parametreleri arasındaki korelasyonun yüksek olduğu durumlarda, yöntemlerin, test hilesini belirlemeye yönelik performansının arttığının gözlemlendiği görülmektedir. Bu nedenle, bu çalışmada, birey parametreleri arasındaki korelasyon değeri 0.75'e

sabitlenmiştir. Ölçmenin hata varyansı ise ortalaması 0, varyansı 0.30 olan log-normal dağılımdan elde edilmiştir.

HMTK-TS model parametrelerinin MCMC ile kestirilmesi

MCMC simülasyon yöntemi, örneklemden alınan bir ya da birden fazla önsel dağılımın boyutunu ve önsel dağılımın hareketini inceleyen deneysel bir yöntemdir. MCMC yaklaşımında, bilinmeyen her bir parametre değeri için yinleme yapılarak kestirim tahminleri üretilmektedir. HMTK-TS model parametrelerinin bayes kestirimleri, MCMC yaklaşımı kullanılarak R.4.0.1. programında yer alan “*LNIRT*” paketi ile gerçekleştirilmiştir. MCMC yaklaşımı olarak *LNIRT* paket programda yer alan Fox, Klein Entink ve van der Linden(2007) ve Klein Entink ve diğerleri (2009a)’nin çalışmalarında geliştirdikleri Gibbs örnekleme algoritması kullanılmıştır. Markov Zincirini kurmadan ve çalıştırmadan önce zincirin başlangıç değerinin ve yakınsamanın sağlandığı noktanın tespit edilerek hazırlık (yakma) periyodunun (Burn-in periyot) belirlenmesi ve modelde otokorelasyonun incelenmesi önem teşkil etmektedir. Markov zincirinin başlangıç değerlerinin (initial values) belirlenmesinde, araştırmacı, deneyimlerini ve ilgili literatürde yer alan çalışmaları göz önünde bulundurarak bir başlangıç değeri atayabileceği gibi, ML yönteminden elde edilen dağılımın mod değerine yakın bir değer de kullanılabilmektedir (Gill, 2015). Hazırlık periyodunun hesaplanmasında otokorelasyon önem teşkil etmektedir. Modele ilişkin otokorelasyonun ihmal edilebilecek seviyede olduğu durumlarda durağan ve dengeli bir zincire ulaşmak için çok sayıda yinlemenin atılmasına gerek duyulmamaktadır. Yeterince uzun bir zincirde, başlangıç değerinin %1 veya %2’sinin örneklemden atılması yeterli görülmektedir (Geyer, 1992; Burke, 2012).

Çalışmada, bayesci yöntemde Gibbs örnekleme algoritması ile elde edilen yakınsaklık sonuçlarını yani MCMC algoritmasının performansını araştırmak için 1000 bireyin 15 maddeye ilişkin, madde tepkilerinin 3PLM ile, madde tepki sürelerinin de log-normal model ile birlikte modellendiği HMTK-TS modeline uygun olarak 50 veri seti, her veri seti için MCMC algoritması 10000 yinleme için çalıştırılarak her zincir için 1000 yinlemelik hazırlık periyodu kullanılarak üretilmiştir. Başlangıç değeri olan 10000 yinlemenin belirlenmesinde, HMTK-TS model parametrelerinin kestirilmesinde

MCMC yöntemlerinden Gibbs örneklemesini kullanan çalışmalar (Fox ve Marianti,2017; Marianti vd.,2014; Qian vd.,2016) dikkate alınmıştır. Yakınsaklık incelemelerine başlamadan önce bireylerin yetenekleri göz önünde bulundurularak 15 maddeden oluşan 3PLM'ye uygun üretilen 1000 bireye ait verilerin tek boyutlu olup olmadığını kontrol etmek için 10 tane veri dosyası random olarak seçilmiştir. Seçilen dosyalardaki verilerin boyutluluk analizleri R.4.0.1 programında yer alan “*sirt*” (Robitzsch, 2020) ve “*paran*” (Dinno, 2018) paketleri kullanılarak incelenmiştir. Yakınsamanın sağlandığı noktayı bulmadaki temel amaç, hazırlık periyodunun belirlenip bu periyoda kadar olan kestirim tahminlerinin dikkate alınmamasını sağlamaktır. Öncelikli olarak, yakınsamanın sağlandığı nokta tespit edilerek bu nokta hazırlık periyodu olarak kullanılmıştır. Uygulamada, yakınsama oranının ne olacağı hakkında bir bilgi olmadığı için Markov zincirinin istenen sonsal dağılıma ulaşip ulaşmadığını incelemede, model parametrelerine ilişkin iz ve otokorelasyon grafiklerinin, parametrelerin hata değerlerinin ve yoğunlukların incelenmesi ve diagnostik testler önerilmiştir (Burke, 2012). Yakınsaklık incelenmelerinde R.4.0.1. programında yer alan “*R2OpenBUGS*” (Thomas,2020) ve “*coda*” (Plummer, Best, Cowles ve Vines, 2020) paketleri kullanılarak model parametrelerine ilişkin iz ve sonsal olasılık yoğunluk grafikleri, Geweke tanı testi sonuçları ve parametre zincirlerine bağlı olarak otokorelasyon grafikleri değerlendirilmiştir. EK 2, EK 3, EK 4 ve EK 5’ de, 15 test maddesi için 10000 MCMC yinelemesine ilişkin madde güçlük, madde ayırtedicilik, zaman ayırıcılık ve zaman yoğunluk parametreleri için iz, sonsal olasılık yoğunluk ve otokorelasyon grafikleri verilmiştir. İz grafikleri ile MCMC uygulaması sonucunda, hata terimlerine ait varyans kestirimlerinin yineleme boyunca oluşturdukları zincirler görülmektedir. İz grafikleri incelenerek zincirde yakınsama sağlanıp sağlanmadığına, hazırlık periyodunu artırmaya gerek olup olmadığına karar verilmektedir (Hosmer ve Lemeshow, 2013). Model parametrelerine ilişkin iz grafikleri incelendiğinde, aşırı dalgalanmaların olmadığı yani zincirin hızlı bir şekilde sonsal dağılıma ulaştığı, yinelemeler boyunca zincirlerin kuvvetli olduğu, her bir parametrenin yakınsadığı ve parametrelerdeki kestirimlerin tahmin değerlerinin giderek durağanlaştığı, aşırı uç değerler almamaya başladığı görülmektedir. Bayesci yaklaşımda, sonsal olasılık yoğunluk grafiği normal dağılım sergilediğinde sonsal dağılıma ulaşıldığı kabul edilmektedir (Burke, 2012). Model parametrelerine ilişkin sonsal olasılık yoğunluk grafikleri incelendiğinde, dağılımların normal dağılım sergilediği, böylelikle, zincirlerin yakınsama gösterdiği ve sonsal dağılıma ulaşıldığı görülmektedir. Markov zincirinde, simülasyonla elde edilen değerlerin birbirinden

bağımsız olmadığı ve bu değerler arasındaki ilişkinin otokorelasyon ile ölçüldüğü görülmektedir (Hosmer ve Lemeshow, 2013). Böylelikle, MCMC örnekleminde, yakınsamanın hızlandırılması için otokorelasyonun azaltılmasının önem teşkil ettiği görülmektedir (Cowles ve Carlin, 1996). Model parametrelerine ilişkin otokorelasyon grafikleri incelendiğinde, otokorelasyon değerlerinin genel olarak 0'a yaklaştığı yani parametre değerleri arasında ilişki olmadığı böylelikle yakınsamanın sağlandığı görülmektedir. Bu yöntemlere ek olarak, Geweke tanı testi uygulanmıştır. Geweke tanı testi, Markov zincirinin önceki %10 ve sonraki %50 kısımlarının ortalamalarını karşılaştırarak ortalama tahminlerinin yakınsamaya sahip olup olmadığını belirleyen Z test istatistiğine dayalı iki yönlü Markov zincirlerinde yakınsaklığı inceleyen testtir. Model parametrelerine ilişkin Geweke yakınsaklık test istatistik değerleri tablosu EK 6'da verilmiştir. Modelde, hipotez testleri $\alpha=0.05$ düzeyinde kurulduğundan dolayı EK 6'da yer alan tüm parametrelerin test istatistiği değerlerinin 1.96 değerinden küçük olduğu ve böylelikle yakınsamanın sağlandığı görülmektedir (Santos, Moala, Tachibana, 2009).

Test Hilesi Senaryosu

Çalışmada, değişimlenen faktörler ve düzeylerinin, tepki süresi tabanlı Kullback-Leibler yaklaşımının ve birey uyum istatistiklerinin I. tip hata ve güç oranlarına olan etkisini gözlemlemek için, R 4.0.1 bilgisayar programında, araştırmacı tarafından yazılan simülasyon kodları ile hileli veriler oluşturulmuştur. Hile senaryosunda, öncelikle, üretilen her bir veri dosyası için çok değişkenli normal dağılımdan elde edilen 1000 bireye ait yetenekler çekilmiştir. Bireylerin yetenekleri göz önünde bulundurularak 1000 bireyden oluşan örneklemden düşük yetenekli bireylerden %5 oranında (50 birey) madde ön bilgisine sahip bireyler belirlenmiştir. Bu işlemin ardından, hile yapan bireyler olarak seçilen bireylerin sırasıyla zor ve orta güçlük düzeyindeki teşhir olan maddelerinin log-normal tepki süreleri ve değişimlenen maddelere verdiği tepkiler doğru (1) olacak şekilde değişimlenmiştir.

Verilerin Analizi

Madde ön bilgisinden kaynaklı test hilesini belirlemeye yönelik ilgili literatür incelendiğinde, kullanılan yöntemlerin performansını belirlemede kullanılan ölçütlerin I.Tip hata ve güç olduğu görülmektedir. Hile belirleme çalışmalarında, hipotez testi sonucunda verilen karar ile gerçek durum arasındaki ilişki Tablo 3’de verilmiştir.

Tablo 3

I.Tip Hata ve Madde Ön Bilgisi Kaynaklı Test Hilesi Belirleme Gücü

		Hipotez Testi Sonucunda Verilen Karar	
		H_0 reddedildi “Madde ön bilgisi kaynaklı test hilesi var”	H_0 kabul edildi “Madde ön bilgisi kaynaklı test hilesi yok”
Gerçek Durum	H_0 doğru “Madde ön bilgisi kaynaklı test hilesi yok”	I.Tip hata	Doğru karar (Madde Ön Bilgisi Kaynaklı Test Hilesi Belirleme Gücü)
	H_0 yanlış “Madde ön bilgisi kaynaklı test hilesi var”	Doğru karar (Güven Düzeyi)	II.Tip hata

Tablo 3 incelendiğinde, madde ön bilgisi kaynaklı test hilesini belirleme yöntemlerinden beklenenin, testte hile yapıldıysa, hile yapan bireylerin doğru olarak tespit edilmesi olduğundan hile belirleme gücünün, madde ön bilgisine sahip olan bireylerin doğru olarak belirlenmesi, I.Tip hata oranının ise madde ön bilgisine sahip olmadığı halde yanlışlıkla birey için “ madde ön bilgisi kaynaklı test hilesi yaptı” kararının verilmesi olduğu görülmektedir.

Uygun Olmayan Madde Tepkileri ve Tepki Sürelerini Belirlemede *Kullback-Leibler* Uzaklığının I.Tip Hata ve Test Hilesi Belirleme Güç Oranlarının Belirlenmesi

Çalışmada, R 4.0.1 bilgisayar programında araştırmacı tarafından yazılan simülasyon kodları ile hileli veriler oluşturularak analizler yapılmıştır. Tepki süresi tabanlı *Kullback Leibler* yaklaşımında, çalışmada üretilen her veri seti için, madde tepkilerinden ve madde tepki sürelerinden madde ön bilgisine sahip bireyleri tespit etmede, Kullback Leibler ayrışım fonksiyonunun değerini elde edebilmek için hile yapan bireyler olarak seçilen bireylerin tepkilerinin, manipülasyon sonrası madde tepki matrisleri (1-0) kullanılarak yetenekleri R 4.0.1 bilgisayar programında yer alan “*LNIRT*” paketi yardımıyla yeniden kestirilmiştir. Yetenekleri yeniden kestirilen tüm bireyler için Kullback Leibler ayrışım fonksiyonunun değerleri, R 4.0.1 bilgisayar programında yer alan “*LaplaceDemon*” paketi (Singmann, 2020) kullanılarak hesaplanmıştır. Kullback Leibler ayrışım fonksiyonunun değerleri için kesme puanı belirlemede, ROC Eğrisi (Receiver Operating Characteristics Curve) yöntemi kullanılmıştır. Standart belirleme sürecinde, ROC Eğrisi yönteminin sınıflama kararlarını verirken, kesme puanını da belirleyebildiği görülmektedir (Kaftandjieva, 2010). ROC analizi, test edilen değişkenin kestirilen değerlerinden hareketle farklı eşik değerler için kestirilen duyarlılık (sensitivity) ve 1-özgüllük (spesifity) çiftlerinin kesişim noktalarının birleştirilmesiyle elde edilen ROC eğrisi yardımıyla yapılmaktadır. Böylelikle, ROC analizi aracılığıyla doğru olarak sınıflandırılmış pozitif oran (duyarlılık) ve doğru olarak sınıflandırılmış negatif oran (özgüllük) arasındaki ilişkinin incelendiği görülmektedir (Flach,2010). ROC Eğrisi, t kesme (eşik) puanına göre sınıflandırılan, dikey ekseninde doğru pozitif oranı (true positive rate), yatay ekseninde yanlış pozitif oranı (false positive rate) gösteren ve bu iki sınıfın puanlarının dağılımlarının kümülatif dağılım fonksiyonu bilgilerini özetleyen bir grafikdir (Krzanowski ve Hand, 2009). Sınıflama kuralının performansına ilişkin değerlendirmeler, *ROC eğrisi altında kalan alan (AUCROC)*, *Youden İndeksi (YI)*, *Kar-Yarar İndeksi* ve *maksimum dikey uzaklık (MDU)* gibi göstergeler üzerinden yapılmaktadır. ROC eğrisi üzerindeki t kesme puanına karşılık gelen, doğru pozitif (duyarlılık) ve yanlış pozitif oranlar arasındaki maksimum fark olan ($YI = \max(\text{Duyarlılık} - (1 - \text{Özgüllük}))$) *Youden İndeksi*, genellikle en uygun sınıflandırma eşik noktasıdır (Krzanowski ve Hand, 2009). Uçar ve Doğan (2021)’in ROC analizinin sınıflama kuralının performansına ilişkin değerlendirmelerinin çeşitli göstergeler ile ele

alındığı çalışmada, *Youden İndeksinin*, duyarlık ve özgüllük miktarlarını dengeli kullanarak sınıflandırmayı daha uygun yapma eğiliminde olduğu sonucuna ulaşılmıştır. Böylelikle, bu çalışmada, tepki süresi tabanlı *Kullback Leibler* yaklaşımında, madde tepkilerinden ve madde tepki sürelerinden madde ön bilgisine sahip bireyleri tespit etmede Kullback Leibler ayrışım fonksiyonunun değerleri için kesme puanları *Youden İndeks* kullanılarak hesaplanmıştır. *Youden İndeksi* kullanılarak $\alpha=0.05$ bayesci anlamlılık düzeyinde kesme puanlarını (kesme bayesci sonsal olasılık değerleri) elde etmek için R 4.0.1 programında yer alan '*OptimalCutpoints*' paketinden (Raton-Lopez, Rodriquez-Alvarez, Suarez- Cadarso ve Sampedro-Gude, 2014) yararlanılmıştır.

Madde tepkilerinden elde edilen kesme puanları kullanılarak, bu tepkilerden, madde ön bilgisine sahip olan bireyleri tespit etmede, Kullback Leibler uzaklığı yaklaşımının, değişimlenen faktörler altında madde ön bilgisinden kaynaklı hile belirleme gücü ve I. Tip hata oranı elde edilmiştir. Kullback Leibler ayrışım fonksiyonu ile madde ön bilgisinden kaynaklı hileyi belirlemede, I. Tip hata oranının belirlenmesi için hile yapılmadığı bilinen madde tepki verileri ile tekrarlanan her yinelemede, madde ön bilgisine sahip olduğu bilinen bireyler için Kullback Leibler ayrışım fonksiyonundan elde edilen sonsal olasılık değerleri hesaplanmıştır. Hesaplanan sonsal olasılıklardan, madde tepkileri için hesaplanan kesme puanından büyük olanlara (hile yapmadığı bilinen ve yanlışlıkla "hile yaptı" kararı verilen bireyler için) 1, diğerlerine ise 0 verilerek 1- 0 matrisi oluşturulmuştur. Matrisde "1" verilen değerlerin toplamının hile yaptığı bilinen birey sayısına oranı hesaplanmıştır. Bu şekilde, hile yaptığı bilinen bireyler için verilen kararın yanlış olma olasılığı (I. Tip hata oranı) ortaya konmuştur. Güç çalışmasında ise, teşhir olan maddelerin oranları da değişimlenerek her yinelemede hile yaptığı bilinen bireyler için Kullback Leibler ayrışım fonksiyonunun sonsal olasılık değerleri hesaplanmıştır. I. Tip hata oranını belirlerken yapıldığı gibi hesaplanan sonsal olasılıklardan, madde tepkileri için hesaplanan kesme puanından büyük olanlara (hile yaptığı bilinen bireyler için verilen "hile yaptı" kararının doğru olma olasılığı) 1, diğerlerine ise 0 verilerek 1- 0 matrisi oluşturulmuştur. Matrisde "1" verilen değerlerin toplamının hile yaptığı bilinen birey sayısına oranı hesaplanmıştır. Bu şekilde hile yaptığı bilinen bireyler için verilen "hile yaptı" kararının doğru olma olasılığı (madde ön bilgisinden kaynaklı hile belirleme güç oranı) ortaya koyulmuştur. Benzer analiz süreçleri, madde tepki sürelerinden, madde ön bilgisine sahip olan bireyleri tespit etmede, bu tepki sürelerinden elde edilen kesme puanları kullanılarak, Kullback Leibler uzaklığı

yaklaşımının, değişimlenen faktörler altında madde ön bilgisinden kaynaklı hile belirleme I. Tip hata ve güç oranlarının elde edilmesinde tekrarlanmıştır.

Uygun Olmayan Madde Tepkileri ve Tepki Sürelerini Belirlemede Birey Uyum İstatistiklerinin I.Tip Hata ve Test Hilesi Belirleme Güç Oranlarının Belirlenmesi

Madde tepkilerinden ve madde tepki sürelerinden , madde ön bilgisine sahip olan bireyleri tespit etmede, madde tepkileri için $L(l_s^y)$ ve madde tepki süreleri için $L(l^t)$ birey uyum istatistiklerinin değişimlenen faktörler altında I.tip hata oranları ve madde ön bilgisinden kaynaklı hile davranışını belirleme güçleri elde edilmiştir. Madde tepkilerinden madde ön bilgisine sahip bireyleri tespit etmede, $L(l_s^y)$ birey uyum istatistiğinin sonsal olasılık değerlerini elde edebilmek için çalışmada üretilen her veri seti için, hile yapan bireyler olarak seçilen bireylerin tepkilerinin, manipülasyon sonrası madde tepki matrisleri (1-0) kullanılarak yetenekleri R 4.0.1 bilgisayar programında yer alan “*LNIRT*” (Fox, Klotzke ve Klein Entink, 2019) paketi yardımıyla yeniden kestirilmiştir. Yetenekleri yeniden kestirilen tüm bireyler için $L(l_s^y)$ birey uyum istatistiğinin sonsal olasılık değerleri, R 4.0.1 bilgisayar programında yer alan “*LNIRT*” paketi (Fox, Klotzke ve Klein Entink, 2019) kullanılarak hesaplanmıştır. $L(l_s^y)$ birey uyum istatistiği ile madde ön bilgisinden kaynaklı hileyi belirlemede, I. Tip hata oranının belirlenmesi için, hile yapılmadığı bilinen madde tepki verileri ile tekrarlanan her yinelemede, madde ön bilgisine sahip olduğu bilinen bireyler için birey uyum istatistiğinin bayesci sonsal olasılık değerleri hesaplanmıştır. Hesaplanan sonsal olasılıklardan, madde tepkileri için belirlenen $\alpha=0.05$ anlamlılık düzeyinden küçük ve eşit olanlara (hile yapmadığı bilinen ve yanlışlıkla “hile yaptı” kararı verilen bireyler için) 1, diğerlerine ise 0 verilerek 1- 0 matrisi oluşturulmuştur. Matrisde “1” verilen değerlerin toplamının hile yaptığı bilinen birey sayısına oranı hesaplanmıştır. α düzeyinin belirlenmesinde literatürde yer alan ilgili araştırmalar (Fox ve Marianti,2017; Marianti ve diğerleri,2014) dikkate alınmıştır. Bu şekilde, hile yapmadığı bilinen bireyler için verilen kararın yanlış olma olasılığı (I. Tip hata oranı) ortaya konmuştur. Güç çalışmasında ise, teşhir olan maddelerin oranları da değişimlenerek her yinelemede hile yaptığı bilinen bireyler için $L(l_s^y)$ birey uyum istatistiğinin sonsal olasılık değerleri hesaplanmıştır. I. Tip hata oranını belirlerken yapıldığı gibi hesaplanan sonsal olasılıklardan, madde tepkileri için belirlenen $\alpha=0.05$ düzeyinden küçük ve eşit olanlara (hile yapmadığı bilinen

ve yanlışlıkla “hile yaptı” kararı verilen bireyler için) 1, diğerlerine ise 0 verilerek 1- 0 matrisi oluşturulmuştur. Matrisde “1” verilen değerlerin toplamının hile yaptığı bilinen birey sayısına oranı hesaplanmıştır. Bu şekilde hile yaptığı bilinen bireyler için verilen “hile yaptı” kararının doğru olma olasılığı (madde ön bilgisinden kaynaklı hile belirleme gücü) ortaya koyulmuştur. Benzer analiz süreçleri, madde tepki sürelerinden, madde ön bilgisine sahip olan bireyleri tespit etmede, $L(l^t)$ birey uyum istatistiğinin değişimlenen faktörler altında madde ön bilgisinden kaynaklı hile belirleme güç oranının ve I. Tip hata oranının elde edilmesinde tekrarlanmıştır. Manipülasyon sonrası yetenekleri yeniden yeniden kestirilen tüm bireyler için $L(l^t)$ birey uyum istatistiğinin sonsal olasılık değerleri, R 4.0.1 bilgisayar programında yer alan “*LNIRT*” paketi (Fox, Klotzke ve Klein Entink, 2019) kullanılarak hesaplanmıştır.



BÖLÜM 4

BULGULAR VE YORUMLAR

Bu bölümde, tepki süresi tabanlı *Kullback-Leibler* yaklaşımının, madde tepki süreleri için $L(l^t)$ ve madde tepkileri için $L(l_s^y)$ birey uyum istatistiklerinin, çeşitli koşullar altında, madde ön bilgisinden kaynaklı test hilesini belirlemede I. Tip hata oranı ve hile belirleme güç oranlarına ait bulgulara yer verilmiştir. Bulguların ortaya konmasında, tek boyutlu ve çok boyutlu grafiklerden yararlanılarak, çalışmada ele alınan koşulların I. Tip hata oranına ve hile belirleme güç oranlarına etkileri, ortak ve temel etki olmak üzere iki ayrı şekilde incelenmiştir. Ayrıca elde edilen bulgulara dayalı olarak yapılan yorumlar raporlanmıştır.

Madde Ön Bilgisinden Kaynaklı Test Hilesini Belirlemede Yöntemlerin I. Tip Hata Oranlarına İlişkin Bulgular ve Yorumlar

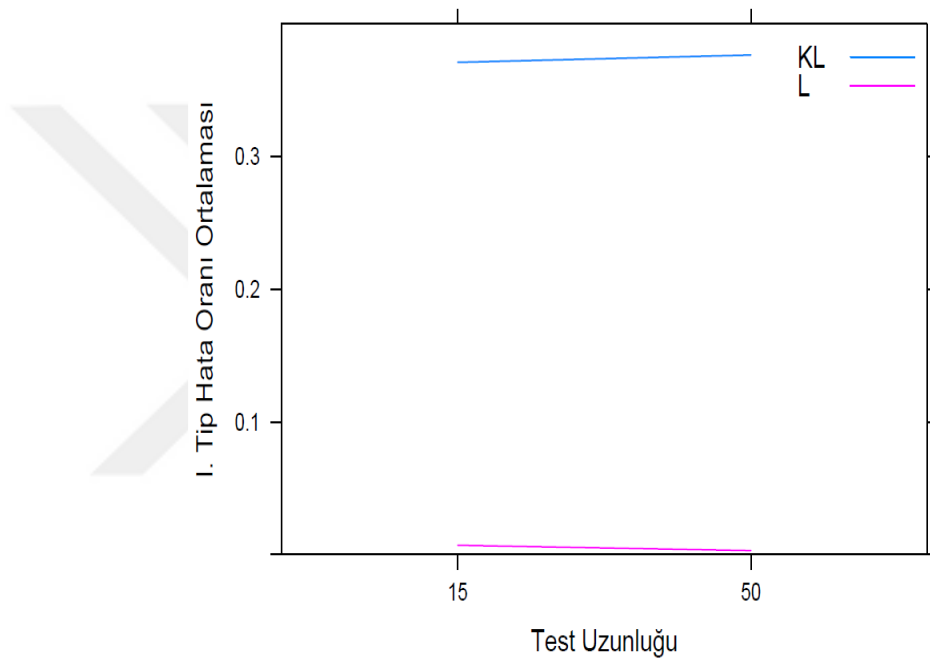
Test uzunluğu ve teşhir olan maddelerin güçlük düzeyi değişimlerinin, yöntemlerin, madde ön bilgisinden kaynaklı test hilesini belirlemede, I. Tip hata oranlarına ortak ve temel etkileri grafikler kullanılarak gösterilmiştir. Madde tepkileri için *Kullback-Leibler* uzaklığının ve $L(l_s^y)$ birey uyum istatistiğinin I. Tip hata oranları ortalaması tablosu EK 7’de, Madde tepki süreleri için *Kullback-Leibler* uzaklığının ve $L(l^t)$ birey uyum istatistiğinin I. Tip hata oranları ortalamaları tablosu EK 8’de verilmiştir.

Koşulların Test Hilesini Belirlemede Yöntemlerin I. Tip Hata Oranına Temel Etkisi

Çalışmanın birinci alt probleminde, manipüle edilmeyen test verileri ile madde tepkilerinin ve madde tepki sürelerinin birlikte hiyerarşik olarak modellenmesi (HMTK-TS) ile madde ön bilgisinden kaynaklı test hilesi belirlenmeye çalışılmıştır.

Test Uzunluğunun Test Hilesini Belirlemede Yöntemlerin I. Tip Hata Oranına Temel Etkisi

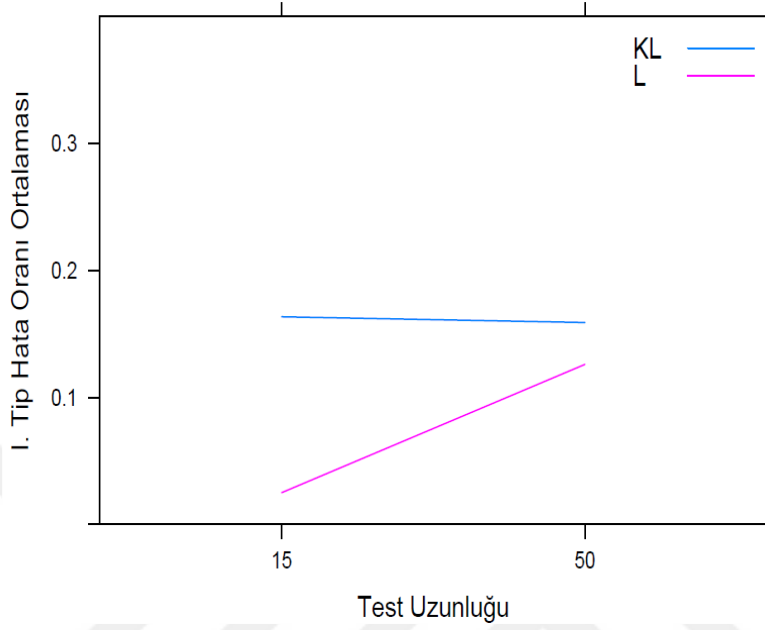
Test uzunluğu değişkeninin, madde tepkileri için *Kullback-Leibler* uzaklığının ve $L(l_s^y)$ birey uyum istatistiğinin I. Tip hata oranlarına oranlarına olan temel etkilerine yönelik elde edilen bulgular Şekil II’de, madde tepki süreleri için *Kullback-Leibler* uzaklığının ve $L(l^t)$ birey uyum istatistiğinin I. Tip hata oranlarına oranlarına olan temel etkilerine yönelik elde edilen bulgular Şekil III’de gösterilmektedir.



Şekil II. Test Uzunluğuna Göre Uygun Olmayan Madde Tepki Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin I. Tip Hata Oranları Ortalaması Grafiği

Şekil II incelendiğinde, $L(l_s^y)$ birey uyum istatistiğinin I. Tip hata oranı ortalama değerlerinin, kısa ve orta test uzunluğunda 0’a yakın değerler aldığı görülmekle birlikte EK 7 incelendiğinde, test uzunluğu arttıkça $L(l_s^y)$ birey uyum istatistiğinin hata oranlarında azalma olduğu gözlemlenmektedir. Madde tepkileri için *Kullback-Leibler* uzaklığının, kısa ve orta test uzunluğunda yakın I. Tip hata oranı ortalama değerleri gösterdiği görülmekle birlikte EK 7 incelendiğinde, test uzunluğu arttıkça madde tepkileri için *Kullback-Leibler* uzaklığının hata oranlarında az farklar ile artış olduğu gözlemlenmektedir. Bununla birlikte, HMTK-TS modelleme ile madde ön bilgisinden kaynaklı test hilesinin belirlenmesinde, uygun olmayan madde tepki örüntülerinin tespit

edilmesinde, test uzunluğu koşulu altında, *Kullback-Leibler* uzaklığının, $L(l_s^y)$ birey uyum istatistiğine göre daha yüksek I. Tip hata oranı ortalama değerlerine sahip olduğu görülmektedir.



Şekil III. Test Uzunluğuna Göre Uygun Olmayan Madde Tepki Süresi Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin I. Tip Hata Oranları Ortalaması Grafiği

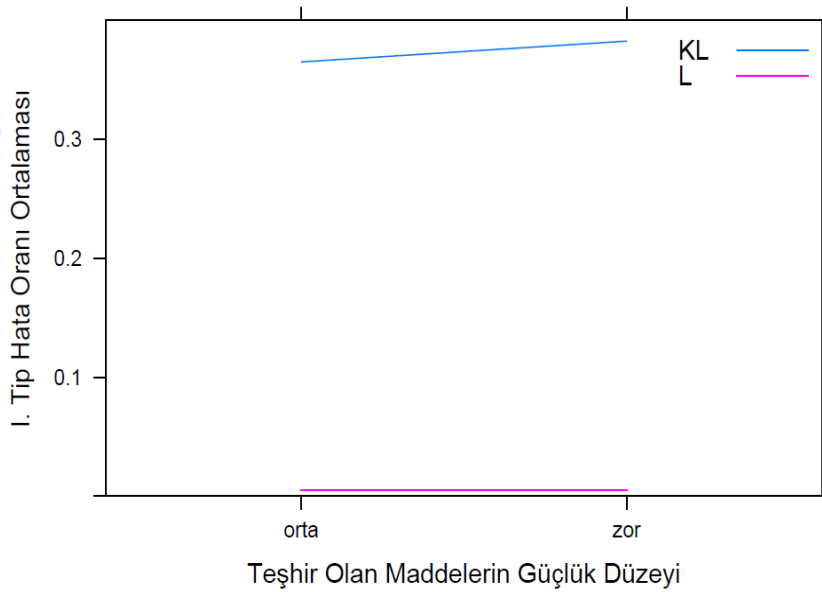
Şekil III incelendiğinde, $L(l^t)$ birey uyum istatistiğinin I. Tip hata oranı ortalama değerlerinin, test uzunluğu arttıkça artış gösterdiği, madde tepki süreleri için *Kullback-Leibler* uzaklığının I. Tip hata oranı ortalama değerlerinin ise test uzunluğu arttıkça küçük bir miktar azaldığı görülmektedir. $L(l^t)$ birey uyum istatistiğinin, kısa test uzunluğunda, I. Tip hata oranı ortalama değerlerinin 0'a yakın değerler aldığı görülmekle birlikte, test uzunluğu azaldıkça *Kullback-Leibler* uzaklığına göre daha düşük I. Tip hata oranı ortalama değerlerine sahip olduğu görülmektedir.

Böylelikle, HMTK-TS modelleme ile madde ön bilgisinden kaynaklı test hilesini belirlemeye yönelik çalışmalar (Man vd.,2018 ; Marianti vd.,2014) doğrultusunda, uygun olmayan madde tepki örüntülerinin belirlenmesinde, kısa ve orta test uzunluğunda, $L(l_s^y)$ birey uyum istatistiğinin, uygun olmayan madde tepki süresi örüntülerinin belirlenmesinde ise kısa test uzunluğunda $L(l^t)$ birey uyum istatistiğinin *Kullback-Leibler* uzaklığına göre I. Tip hata oranı ortalama değerinin oldukça düşük olduğu görülmektedir. Ayrıca, test uzunluğu koşulu altında, *Kullback-Leibler* uzaklığının ve L

birey uyum istatistiğinin, uygun olmayan madde tepki örüntülerinin belirlenmesinde uygun olmayan tepki süresi örüntülerinin belirlenmesine göre daha yüksek 1.Tip hata oranı ortalama değerlerine sahip olduğu görülmektedir.

Teşhir Olan Maddelerin Güçlük Düzeyinin Yöntemlerin Test Hilesini Belirlemede I. Tip Hata Oranına Temel Etkisi

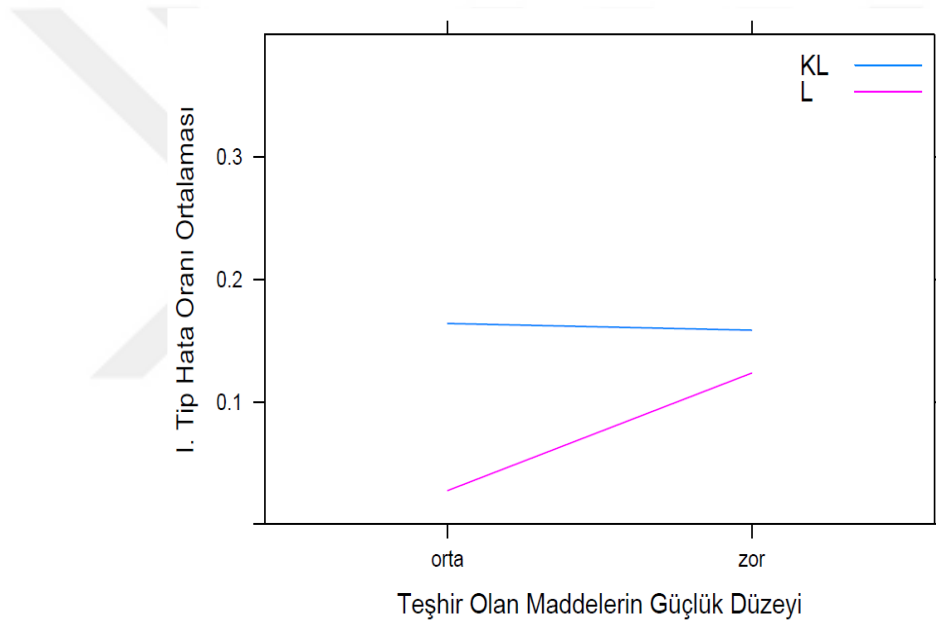
Teşhir olan maddelerin güçlük düzeyi değişkeninin, madde tepkileri için *Kullback-Leibler* uzaklığının ve $L(l_s^y)$ birey uyum istatistiğinin I. .Tip hata oranlarına oranlarına olan temel etkilerine yönelik elde edilen bulgular Şekil IV’de, madde tepki süreleri için *Kullback-Leibler* uzaklığının ve $L(l^t)$ birey uyum istatistiğinin I. .Tip hata oranlarına oranlarına olan temel etkilerine yönelik elde edilen bulgular Şekil V’de gösterilmektedir.



Şekil IV. Teşhir Olan Maddelerin Güçlük Düzeyine Göre Uygun Olmayan Madde Tepki Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin I. Tip Hata Oranları Ortalaması Grafiği

Şekil IV incelendiğinde, $L(l_s^y)$ birey uyum istatistiğinin I. Tip hata oranı ortalama değerlerinin, zor ve orta güçlükteki maddelerde 0’a yakın değerler aldığı görülmekle

birlikte EK 7 incelendiğinde, madde güçlük düzeyi arttıkça $L(l_s^y)$ birey uyum istatistiğinin hata oranlarının da artma eğiliminde olduğu gözlemlenmektedir. Madde tepkileri için *Kullback-Leibler* uzaklığının, zor ve orta güçlükteki maddelerde yakın I. Tip hata oranı ortalama değerleri gösterdiği görülmekle birlikte EK 7 incelendiğinde, madde güçlük düzeyi arttıkça madde tepkileri için *Kullback-Leibler* uzaklığının hata oranlarında az farklar ile artış olduğu gözlemlenmektedir. Bununla birlikte, HMTK-TS modelleme ile madde ön bilgisinden kaynaklı test hilesinin belirlenmesinde, uygun olmayan madde tepki örüntülerinin tespit edilmesinde, teşhir olan maddelerin güçlük düzeyi koşulu altında *Kullback-Leibler* uzaklığının, $L(l_s^y)$ birey uyum istatistiğine göre daha yüksek I. Tip hata oranı ortalama değerlerine sahip olduğu görülmektedir.



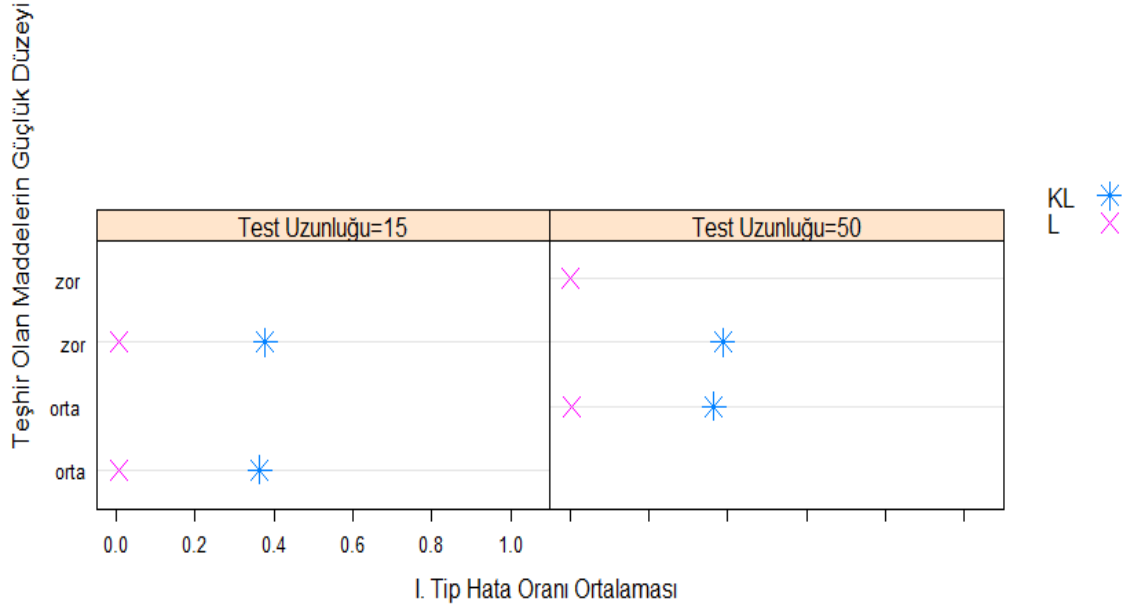
Şekil V. Teşhir Olan Maddelerin Güçlük Düzeyine Göre Uygun Olmayan Madde Tepki Süresi Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin I. Tip Hata Oranları Ortalaması Grafiği

Şekil V incelendiğinde, $L(l^t)$ birey uyum istatistiğinin I. Tip hata oranı ortalama değerlerinin, madde güçlük düzeyi arttıkça artış gösterdiği, madde tepki süreleri için *Kullback-Leibler* uzaklığının I. Tip hata oranı ortalama değerlerinin ise madde güçlük düzeyi arttıkça küçük bir miktar azalma eğiliminde olduğu görülmektedir. $L(l^t)$ birey uyum istatistiğinin, madde güçlük düzeyi koşulunun bütün düzeylerinde, I. Tip hata oranı

ortalama değerlerinin 0'a yakın değerler aldığı görülmekle birlikte *Kullback-Leibler* uzaklığına göre daha düşük I. Tip hata oranı ortalama değerlerine sahip olduğu gözlenmektedir. Man ve diğerleri (2018)'nin *Kullback-Leibler* uzaklığının ve $L(l^t)$ birey uyum istatistiğinin madde ön bilgisi ve cevap kopyalamadan kaynaklı test hilesini belirlemeye yönelik çalışmalarında, küçük örneklem büyüklüğünde (500), yöntemlerin, en uzun tepki süresine sahip olan en zor maddede, madde ön bilgisinden kaynaklı hile durumunu belirlemede, I. Tip hata oranı ortalama değerlerinin 0'a yakın değerler aldığı görülmekle birlikte *Kullback-Leibler* uzaklığının I. Tip hata oranı ortalama değerlerinin $L(l^t)$ birey uyum istatistiğine göre daha yüksek değerler gösterdiği sonucuna ulaşılmıştır. Böylelikle bu bulguların, Man ve diğerleri (2018)'nin çalışmalarında elde ettiği sonuçlar ile benzerlik gösterdiği görülmektedir. Ayrıca, teşhir olan maddelerin güçlük düzeyi koşulu altında, *Kullback-Leibler* uzaklığının ve L birey uyum istatistiğinin, uygun olmayan madde tepki örüntülerinin belirlenmesinde uygun olmayan madde tepki süresi örüntülerinin belirlenmesine göre daha yüksek I. Tip hata oranı ortalama değerlerine sahip olduğu görülmektedir.

Koşulların Test Hilesini Belirlemede Yöntemlerin I. Tip Hata Oranına Ortak Etkisi

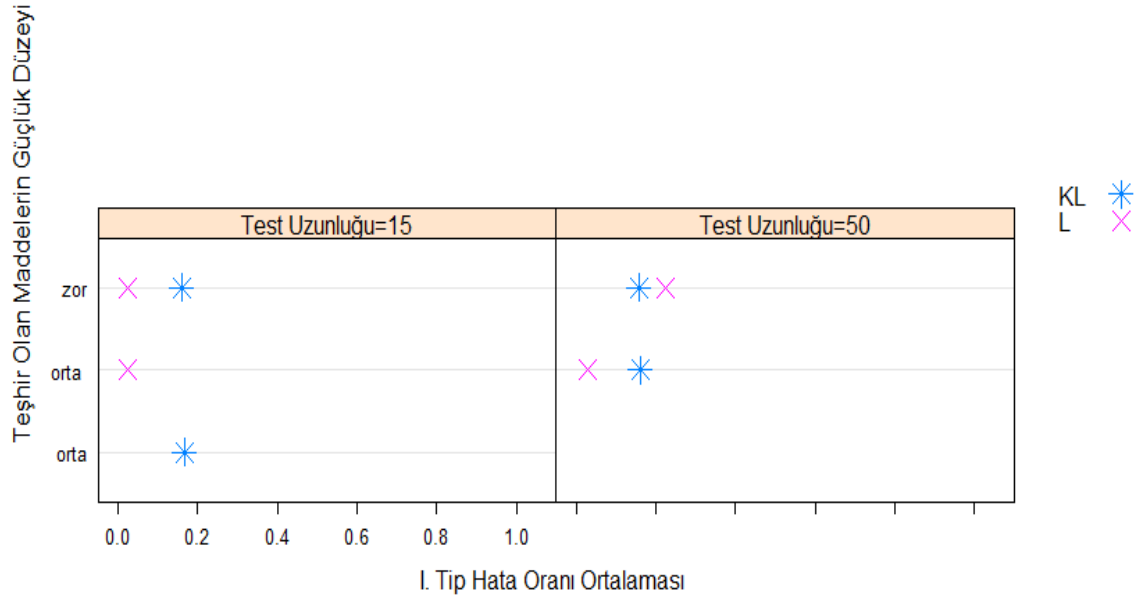
Test uzunluğu ve teşhir olan maddelerin güçlük düzeyi değişimlesinin, yöntemlerin, madde ön bilgisinden kaynaklı test hilesini belirlemede I. Tip hata oranlarına ortak etkileri grafikler kullanılarak gösterilmiştir. Değişimlenen koşulların, madde tepkileri için *Kullback-Leibler* uzaklığının ve $L(l_s^y)$ birey uyum istatistiğinin I. Tip hata oranlarına oranlarına olan ortak etkilerine yönelik elde edilen bulgular Şekil VI'da, madde tepki süreleri için *Kullback-Leibler* uzaklığının ve $L(l^t)$ birey uyum istatistiğinin I. Tip hata oranlarına olan ortak etkilerine yönelik elde edilen bulgular Şekil VII'da gösterilmektedir.



Şekil VI. Değişimlenen Koşullara Göre Uygun Olmayan Madde Tepki Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin I. Tip Hata Oranları Grafiği

Değişimlenen koşulların, madde tepkileri için *Kullback-Leibler* uzaklığının ve $L(l_s^y)$ birey uyum istatistiğinin I. Tip Hata oranına ortak etkisini gösteren Şekil VI incelendiğinde, *Kullback-Leibler* uzaklığının, tüm koşullar altında, $L(l_s^y)$ birey uyum istatistiğine göre yüksek I. Tip hata oranına sahip olduğu görülmektedir. Ayrıca *Kullback-Leibler* uzaklığının, orta test uzunluğu ve zor maddelerde en yüksek I. Tip hata oranı ortalama değeri gösterdiği görülmektedir. Bu durum, bireylerin yetenek parametrelerinin önsel ve sonsal olasılık dağılımları arasındaki fark olarak hesaplanan *Kullback-Leibler* uzaklığının, yetenek parametreleri kestiriminden kaynaklanan farklılıklara oldukça duyarlı olduğunu ve sıfır hipotezi altında dağılımının belirsizliğinden kaynaklı olarak uygun olmayan madde tepki örüntülerinin belirlenmesinde kullanılan kesme puanı yönteminin oldukça konservatif olduğunu göstermektedir. Uçar ve Doğan (2021)'in çalışmasında da belirttiği üzere kesme puanının belirlenmesinde kullanılan *Youden İndeksinin*, duyarlılık ve özgüllük miktarlarını dengeli kullanarak sınıflandırmayı daha uygun yapma eğiliminde olmasından kaynaklı olarak *Kullback-Leibler* uzaklığının, yüksek I. Tip hata oranı gösterdiği söylenebilir. $L(l_s^y)$ birey uyum istatistiğinin ise değişimlenen tüm koşullarda, I. Tip hata oranı değerlerinin 0'a yakın değerler aldığı görülmekle birlikte en yüksek değerlerin kısa test uzunluğu ve zor maddelerde gözlemlendiği

görülmektedir. Böylelikle, *Kullback-Leibler* uzaklığının ve L birey uyum istatistiğinin, zor maddelerde, uygun olmayan madde tepki örüntülerinin belirlenmesinde daha yüksek I. Tip hata oranı değerlerine sahip olduğu görülmektedir.



Şekil VII. Değişimlenen Koşullara Göre Uygun Olmayan Madde Tepki Süresi Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin I. Tip Hata Oranları Grafiği

Değişimlenen koşulların, madde tepki süreleri için *Kullback-Leibler* uzaklığının ve $L(l^t)$ birey uyum istatistiğinin I. Tip Hata oranına ortak etkisini gösteren Şekil VII incelendiğinde, $L(l^t)$ birey uyum istatistiğinin orta test uzunluğu ve zor maddeler koşulu dışında değişimlenen tüm koşullarda I. Tip hata oranı değerlerinin, 0'a yakın değerler aldığı görülmekle birlikte *Kullback-Leibler* uzaklığına göre daha düşük I. Tip hata oranı değerlerine sahip olduğu görülmektedir. Değişimlenen tüm koşullarda, en yüksek I. Tip hata oranını, uygun olmayan madde tepki süresi örüntülerinin tespit edilmesinde kullanılan yöntemlerden, $L(l^t)$ birey uyum istatistiğinin, orta test uzunluğu ve zor maddeler koşulunda gösterdiği görülmektedir. Böylelikle, çalışmadan elde edilen bulguların, Fox ve Marianti (2017)'nin HMTK-TS modelleme ile madde ön bilgisinden kaynaklı test hilesini belirlemede, $L(l^t)$ ve $L(l^t)$ birey uyum istatistiklerinin performanslarını belirlemeye yönelik çalışmalarında elde edilen düşük I. Tip hata oranı bulguları ile benzerlik gösterdiği görülmektedir.

Madde Ön Bilgisinden Kaynaklı Test Hilesini Belirlemede Yöntemlerin Hile Belirleme Güç Oranlarına İlişkin Bulgular ve Yorumlar

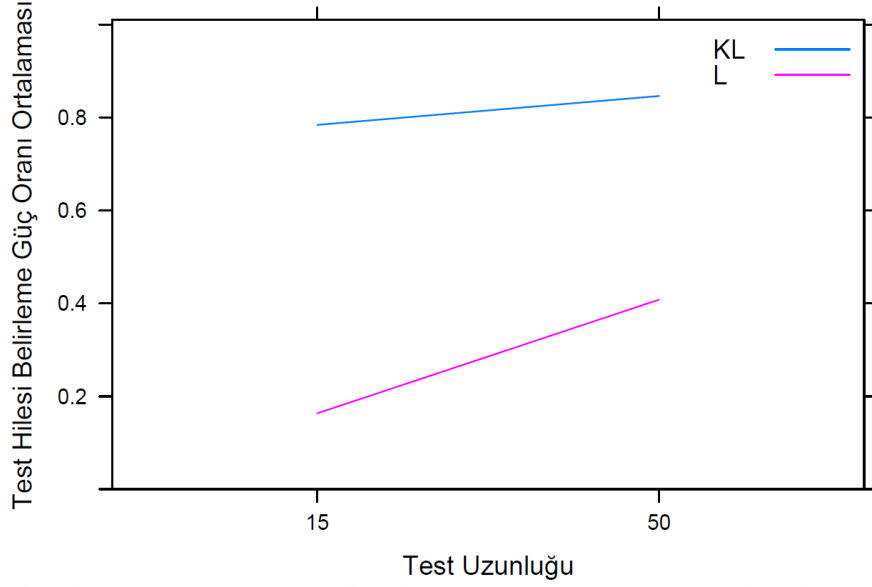
Test uzunluğu, teşhir olan maddelerin güçlük düzeyi, teşhir olan maddelerin oranı, teşhir olan maddelerin tepki süresi değişimlemesi değişkenlerinin, yöntemlerin, madde ön bilgisinden kaynaklı test hilesini belirlemedeki güç oranlarına olan ortak ve temel etkileri şekiller kullanılarak gösterilmiştir. Madde tepkileri için *Kullback-Leibler* uzaklığının ve $L(l_s^y)$ birey uyum istatistiğinin hile belirleme güç oranları ortalamaları tablosu EK 9’da, madde tepki süreleri için *Kullback-Leibler* uzaklığının ve $L(l^t)$ birey uyum istatistiğinin hile belirleme güç oranları ortalamaları tablosu EK 10’da verilmiştir.

Koşulların Test Hilesi Belirleme Güç Oranına Temel Etkisi

Çalışmanın ikinci alt probleminde, manipüle edilen test verileri ile HMTK-TS modelleme ile madde ön bilgisinden kaynaklı test hilesi belirlenmeye çalışılmıştır.

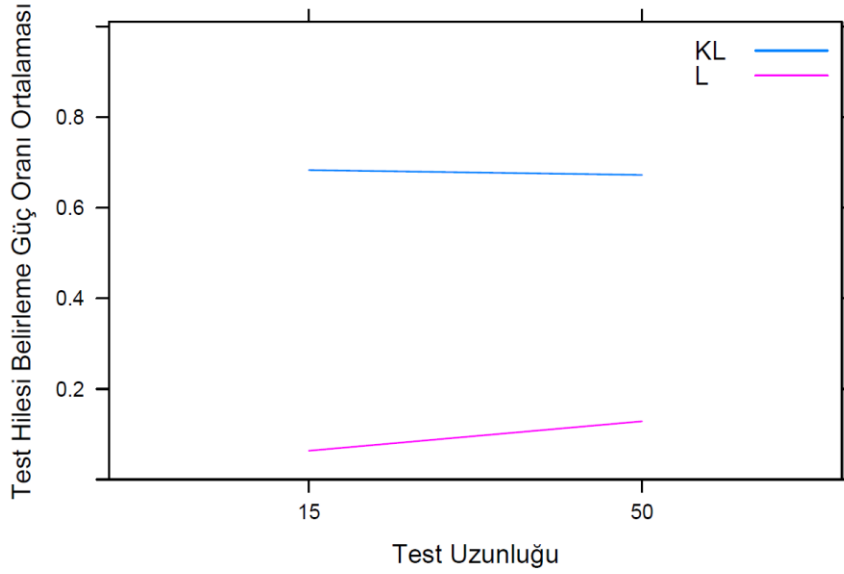
Test Uzunluğunun Test Hilesi Belirleme Güç Oranına Temel Etkisi

Test uzunluğu değişkeninin, madde tepkileri için *Kullback-Leibler* uzaklığının ve $L(l_s^y)$ birey uyum istatistiğinin test hilesi belirleme güç oranı ortalamalarının temel etkilerine yönelik elde edilen bulgular Şekil VIII’da, madde tepki süreleri için *Kullback-Leibler* uzaklığının ve $L(l^t)$ birey uyum istatistiğinin test hilesi belirleme güç oranı ortalamalarının temel etkilerine yönelik elde edilen bulgular Şekil IX’de gösterilmektedir.



Şekil VIII. Test Uzunluğuna Göre Uygun Olmayan Madde Tepki Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin Güç Oranı Ortalaması Grafiği

Şekil VIII incelendiğinde, $L(l_s^y)$ birey uyum istatistiğinin test hilesi belirleme güç oranlarının ortalama değerlerinin düşük değerler aldığı ve test uzunluğu arttıkça artış gösterdiği görülmektedir. Madde tepkileri için *Kullback-Leibler* uzaklığının, kısa ve orta test uzunluğunda yüksek ve birbirine yakın test hilesi belirleme güç oranları ortalama değerleri gösterdiği görülmekle birlikte EK 9 incelendiğinde, test uzunluğu arttıkça madde tepkileri için *Kullback-Leibler* uzaklığının güç oranları ortalama değerlerinde az farklar ile artış olduğu gözlemlenmektedir. Bununla birlikte, HMTK-TS modelleme ile madde ön bilgisinden kaynaklı test hilesi belirlenmesinde, uygun olmayan tepki örüntülerinin tespit edilmesinde, test uzunluğu koşulu altında, *Kullback-Leibler* uzaklığının, $L(l_s^y)$ birey uyum istatistiğine göre daha yüksek test hilesi belirleme güç oranları ortalama değerlerine sahip olduğu görülmektedir.



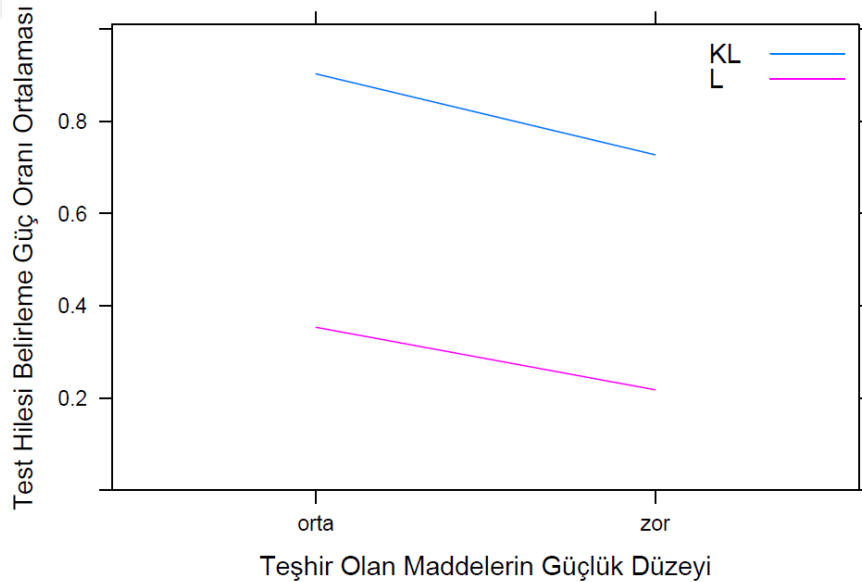
Şekil IX. Test Uzunluğuna Göre Uygun Olmayan Madde Tepki Süresi Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin Güç Oranı Ortalaması Grafiği

Şekil IX incelendiğinde, $L(l^t)$ birey uyum istatistiğinin test hilesi belirleme güç oranlarının ortalama değerlerinin düşük değerler aldığı ve test uzunluğu arttıkça artış gösterdiği görülmektedir. Madde tepki süreleri için *Kullback-Leibler* uzaklığının, kısa ve orta test uzunluğunda yüksek ve birbirine yakın test hilesi belirleme güç oranları ortalama değerleri gösterdiği görülmekle birlikte EK 10 incelendiğinde, test uzunluğu arttıkça madde tepkileri için *Kullback-Leibler* uzaklığının güç oranları ortalama değerlerinde küçük bir miktar artış olduğu gözlemlenmektedir. Bununla birlikte, HMTK-TS modelleme ile madde ön bilgisinden kaynaklı test hilesinin belirlenmesinde uygun olmayan tepki süresi örüntülerinin tespit edilmesinde, test uzunluğu koşulu altında, *Kullback-Leibler* uzaklığının, $L(l^t)$ birey uyum istatistiğine göre daha yüksek test hilesi belirleme güç oranı ortalama değerlerine sahip olduğu görülmektedir. Bu bulguların, Man ve diğerleri (2018)'nin çalışmalarında elde ettiği bulgular ile benzerlik gösterdiği görülmektedir. Ayrıca, test uzunluğu koşulu altında, *Kullback-Leibler* uzaklığının ve L birey uyum istatistiğinin, uygun olmayan madde tepki örüntülerinin belirlenmesinde uygun olmayan madde tepki süresi örüntülerinin belirlenmesine göre daha yüksek güç oranı ortalama değerlerine sahip olduğu görülmektedir. Böylelikle, uygun olmayan madde tepki ve tepki süresi örüntülerinin belirlenmesinde, test uzunluğu koşulu altında, bireylerin yetenek parametrelerinin önsel ve sonsal dağılımları arasındaki farka dayalı

yöntemlerin madde tepkilerinin ve tepki sürelerinin olabilirliğine dayalı yöntemlerin daha yüksek performansa sahip olduğu ve test uzunluğundaki artış ile performansın küçük bir miktar da olsa artış gösterdiği söylenebilir. Bunun nedeni olarak madde sayısının fazla olmasının bireylerin yetenek parametrelerinin kestiriminde daha fazla bilgi sağlaması gösterilebilir.

Teşhir Olan Maddelerin Güçlük Düzeyinin Test Hilesi Belirleme Güç Oranına Temel Etkisi

Teşhir olan maddelerin güçlük düzeyi değişkeninin, madde tepkileri için *Kullback-Leibler* uzaklığının ve $L(l_s^y)$ birey uyum istatistiğinin test hilesi belirleme güç oranı ortalamalarının temel etkilerine yönelik elde edilen bulgular Şekil X’de, madde tepki süreleri için *Kullback-Leibler* uzaklığının ve $L(l^t)$ birey uyum istatistiğinin test hilesi belirleme güç oranı ortalamalarının temel etkilerine yönelik elde edilen bulgular Şekil XI’da gösterilmektedir.



Şekil X. Teşhir Olan Maddelerin Güçlük Düzeyine Göre Uygun Olmayan Madde Tepki Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin Güç Oranı Ortalaması Grafiği

Şekil X incelendiğinde, $L(l_s^y)$ birey uyum istatistiğinin test hilesi belirleme güç oranlarının ortalama değerlerinin düşük değerler aldığı ve madde güçlük düzeyi arttıkça azaldığı görülmektedir. Madde güçlük düzeyi arttıkça yöntemin güç oranı ortalamasının azaldığına ilişkin bulgu, Reise (1995)'nin, madde tepkilerinin olabilirliğine dayalı birey uyum istatistiklerinin madde tepki modeline uygun olmayan tepkileri belirlemedeki performanslarının bireylerin yetenek parametreleri ile madde güçlük parametreleri arasındaki fark arttıkça artış göstermesi bulgusu ile farklılık göstermektedir. Bu durum, HMTK-TS, modelinde birey ve madde parametrelerinin kestiriminin madde tepkilerinin ve madde tepki sürelerinin birlikte ele alınarak yapılması ile açıklanabilir. Madde tepkileri için *Kullback-Leibler* uzaklığının, zor ve orta güçlük düzeyindeki maddelerde yüksek test hilesi belirleme güç oranı ortalama değerleri gösterdiği görülmekle birlikte madde güçlük düzeyi arttıkça madde tepkileri için *Kullback-Leibler* uzaklığının güç oranları ortalama değerlerinde azalma olduğu gözlemlenmektedir. Bununla birlikte, HMTK-TS modelleme ile madde ön bilgisinden kaynaklı test hilesinin belirlenmesinde, uygun olmayan tepki örüntülerinin tespit edilmesinde, teşhir olan maddelerin güçlük düzeyi koşulu altında, *Kullback-Leibler* uzaklığının, $L(l_s^y)$ birey uyum istatistiğine göre daha yüksek test hilesi belirleme güç oranı ortalama değerlerine sahip olduğu görülmektedir.

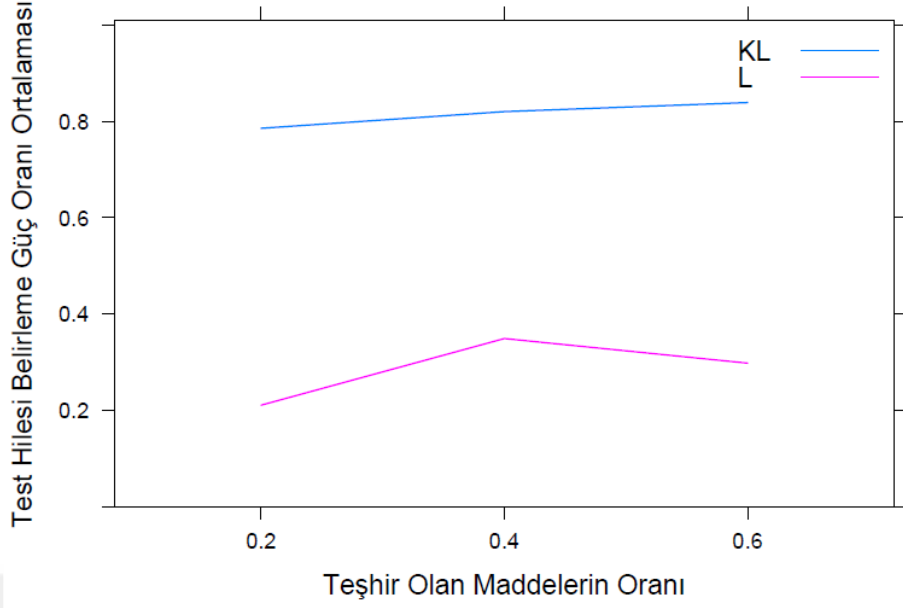


Şekil XI. Teşhir Olan Maddelerin Güçlük Düzeyine Göre Uygun Olmayan Madde Tepki Süresi Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin Güç Oranı Ortalaması Grafiği

Şekil XI incelendiğinde, $L(l^t)$ birey uyum istatistiğinin test hilesi belirleme güç oranlarının ortalama değerlerinin, düşük değerler aldığı ve madde güçlük düzeyi arttıkça azaldığı görülmektedir. Madde tepki süreleri için *Kullback-Leibler* uzaklığının, orta güçlük düzeyindeki maddelerde zor maddelere göre daha yüksek test hilesi belirleme güç ortalama değerleri gösterdiği görülmektedir. Bununla birlikte, HMTK-TS modelleme ile madde ön bilgisinden kaynaklı test hilesinin belirlenmesinde, uygun olmayan tepki süresi örüntülerinin tespit edilmesinde, teşhir olan maddelerin güçlük düzeyi koşulu altında, *Kullback-Leibler* uzaklığının, $L(l^t)$ birey uyum istatistiğine göre daha yüksek test hilesi belirleme güç oranları ortalama değerlerine sahip olduğu görülmektedir. Böylelikle, uygun olmayan madde tepki ve tepki süresi örüntülerinin belirlenmesinde, madde güçlük düzeyi koşulu altında, bireylerin yetenek parametrelerinin önsel ve sonsal dağılımları arasındaki farka dayalı yöntemlerin madde tepkilerinin ve tepki sürelerinin olabilirliğine dayalı yöntemlere göre daha yüksek performans gösterdiği söylenebilir.

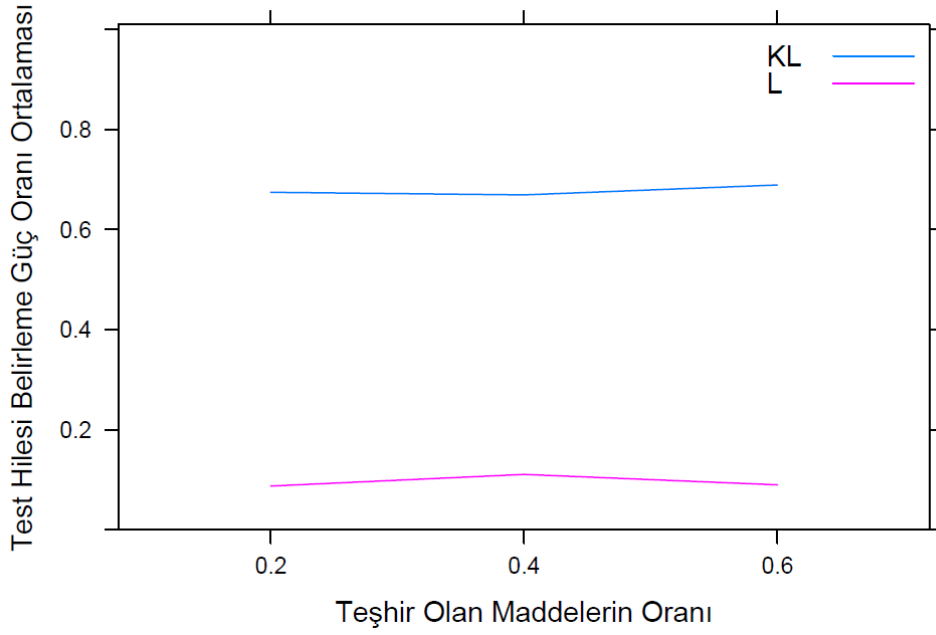
Teşhir Olan Maddelerin Oranının Test Hilesi Belirleme Güç Oranına Temel Etkisi

Teşhir olan maddelerin oranı değişkeninin, madde tepkileri için *Kullback-Leibler* uzaklığının ve $L(l_s^y)$ birey uyum istatistiğinin test hilesi belirleme güç oranı ortalamalarının temel etkilerine yönelik elde edilen bulgular Şekil XII’da, madde tepki süreleri için *Kullback-Leibler* uzaklığının ve $L(l^t)$ birey uyum istatistiğinin test hilesi belirleme güç oranı ortalamalarına temel etkilerine yönelik elde edilen bulgular Şekil XIII’da gösterilmektedir.



Şekil XII. Teşhir Olan Maddelerin Oranına Göre Uygun Olmayan Madde Tepki Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin Güç Oranı Ortalaması Grafiği

Şekil XII incelendiğinde, $L(l_s^y)$ birey uyum istatistiğinin test hilesi belirleme güç oranlarının ortalama değerlerinin düşük değerler aldığı ve teşhir olan maddelerin oranının orta düzeyde olduğu durumda en yüksek değerini, düşük olduğu düzeyde ise en düşük değerini aldığı görülmektedir. Madde tepkileri için *Kullback-Leibler* uzaklığının, maddelerin düşük, orta ve yüksek oranda teşhir olduğu durumlarda yüksek ve birbirine yakın test hilesi belirleme güç oranı ortalama değerleri gösterdiği görülmekle birlikte EK 9 incelendiğinde, teşhir olan maddelerin oranı arttıkça madde tepkileri için *Kullback-Leibler* uzaklığının güç oranı ortalama değerlerinde artış olduğu gözlemlenmektedir. Bununla birlikte, HMTK-TS modelleme ile madde ön bilgisinden kaynaklı test hilesi belirlenmesinde uygun olmayan tepki örüntülerinin tespit edilmesinde, teşhir olan maddelerin oranı koşulu altında, *Kullback-Leibler* uzaklığının, $L(l_s^y)$ birey uyum istatistiğine göre daha yüksek test hilesi belirleme güç oranı ortalama değerlerine sahip olduğu görülmektedir.

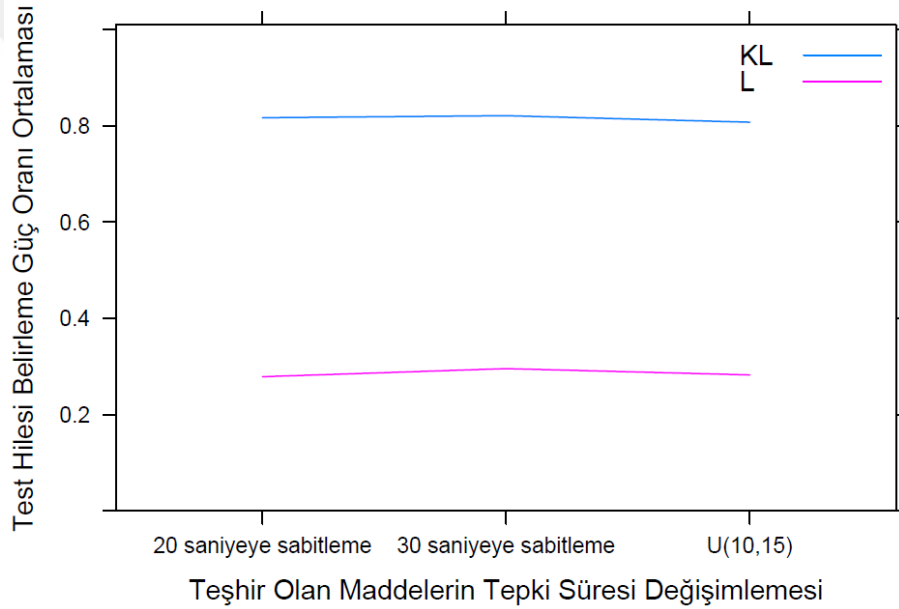


Şekil XIII. Teşhir Olan Maddelerin Oranına Göre Uygun Olmayan Madde Tepki Süresi Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin Güç Oranı Ortalaması Grafiği

Şekil XIII incelendiğinde, $L(l^t)$ birey uyum istatistiğinin test hilesi belirleme güç oranlarının ortalama değerlerinin düşük değerler aldığı ve teşhir olan maddelerin oranının orta düzeyde olduğu durumda en yüksek değeri, düşük ve yüksek olduğu düzeyde ise yakın değerler gösterdiği görülmektedir. Madde tepki süreleri için *Kullback-Leibler* uzaklığının, maddelerin düşük, orta ve yüksek oranda teşhir olduğu durumlarda yüksek ve birbirine yakın test hilesi belirleme güç oranı ortalama değerleri gösterdiği görülmekle birlikte EK 10 incelendiğinde teşhir olan maddelerin oranı arttıkça madde tepkileri için *Kullback-Leibler* uzaklığının güç ortalama değerlerinde artış olduğu gözlemlenmektedir. Bununla birlikte, HMTK-TS modelleme ile madde ön bilgisinden kaynaklı test hilesi belirlenmesinde, uygun olmayan tepki süresi örüntülerinin tespit edilmesinde, teşhir olan maddelerin oranı koşulu altında, *Kullback-Leibler* uzaklığının, $L(l^t)$ birey uyum istatistiğine göre daha yüksek test hilesi belirleme güç oranı ortalama değerlerine sahip olduğu görülmektedir. Böylelikle, uygun olmayan madde tepki ve tepki sürelerinin belirlenmesinde, teşhir olan maddelerin koşulu altında, bireylerin yetenek parametrelerinin önsel ve sonsal dağılımları arasındaki farka dayalı yöntemlerin madde tepkilerinin ve tepki sürelerinin olabilirliğine dayalı yöntemlere göre daha yüksek performans gösterdiği söylenebilir.

Teşhir Olan Maddelerin Tepki Süresi Değişimlesinin Test Hilesi Belirleme Güç Oranına Temel Etkisi

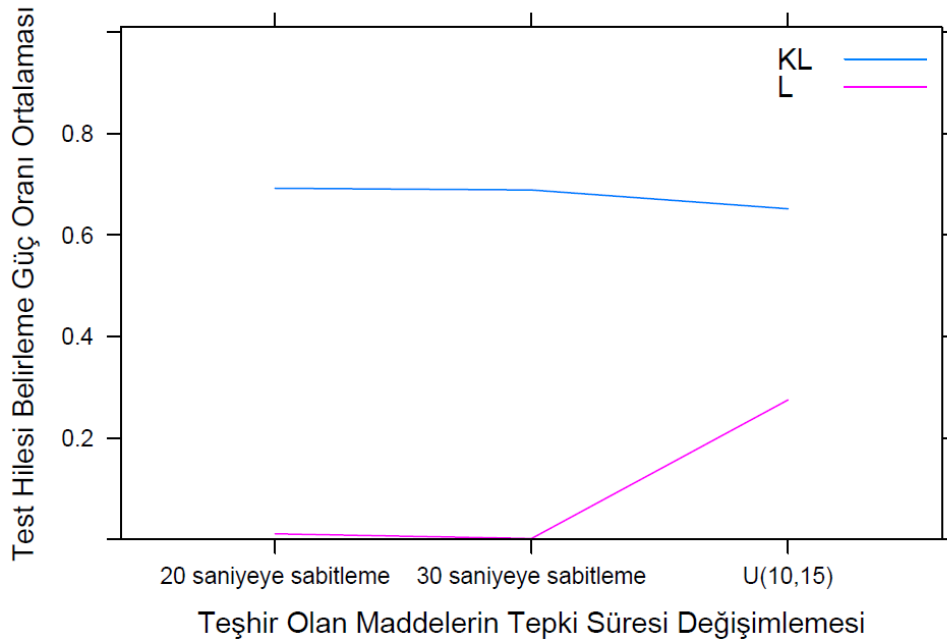
Teşhir olan maddelerin tepki süresi değişimlesi değişkeninin, madde tepkileri için *Kullback-Leibler* uzaklığının ve $L(l_s^y)$ birey uyum istatistiğinin test hilesi belirleme güç oranlarına temel etkilerine yönelik elde edilen bulgular Şekil XIV’de, madde tepki süreleri için *Kullback-Leibler* uzaklığının ve $L(l^t)$ birey uyum istatistiğinin test hilesi belirleme güç oranlarına temel etkilerine yönelik elde edilen bulgular Şekil XV’de gösterilmektedir.



Şekil XIV. Teşhir Olan Maddelerin Tepki Süresi Değişimlesine Göre Uygun Olmayan Madde Tepki Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin Güç Oranı Ortalaması Grafiği

Şekil XIV incelendiğinde, $L(l_s^y)$ birey uyum istatistiğinin test hilesi belirleme güç oranlarının ortalama değerlerinin düşük değerler aldığı ve teşhir olan maddelerin tepki sürelerinin 30 saniyeye sabitlendiği durumda en yüksek değerini, 20 saniyeye sabitlendiği ve 10 ile 15 saniye arasındaki uniform dağılımdan elde edildiği durumlarda yakın değerler gösterdiği görülmektedir. Böylelikle, HMTK-TS model kullanılarak

uygun olmayan madde tepki örüntülerinin belirlenmesinde, madde tepkilerinin olabilirliğine dayalı yöntemlerin performansının, teşhir olan maddelerin tepki sürelerinin daha yüksek zaman aralıklarına sabitlenmesi ile arttığı söylenebilir. Madde tepkileri için *Kullback-Leibler* uzaklığının, teşhir olan maddelerin tepki sürelerinin değişimlesinin bütün düzeylerinde yüksek ve birbirine yakın test hilesi belirleme güç oranı ortalama değerleri gösterdiği görülmekle birlikte teşhir olan maddelerin tepki sürelerinin 20 saniyeye sabitlendiği durumda en yüksek değeri aldığı görülmektedir. Bununla birlikte, HMTK-TS modelleme ile madde ön bilgisinden kaynaklı test hilesi belirlenmesinde uygun olmayan tepki örüntülerinin tespit edilmesinde, teşhir olan maddelerin tepki süresi değişimlesmesi koşulu altında, *Kullback-Leibler* uzaklığının, $L(l_s^y)$ birey uyum istatistiğine göre daha yüksek test hilesi belirleme güç oranları ortalama değerlerine sahip olduğu görülmektedir.



Şekil XV. Teşhir Olan Maddelerin Tepki Süresi Değişimlesmesine Göre Uygun Olmayan Madde Tepki Süresi Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin Güç Oranı Ortalaması Grafiği

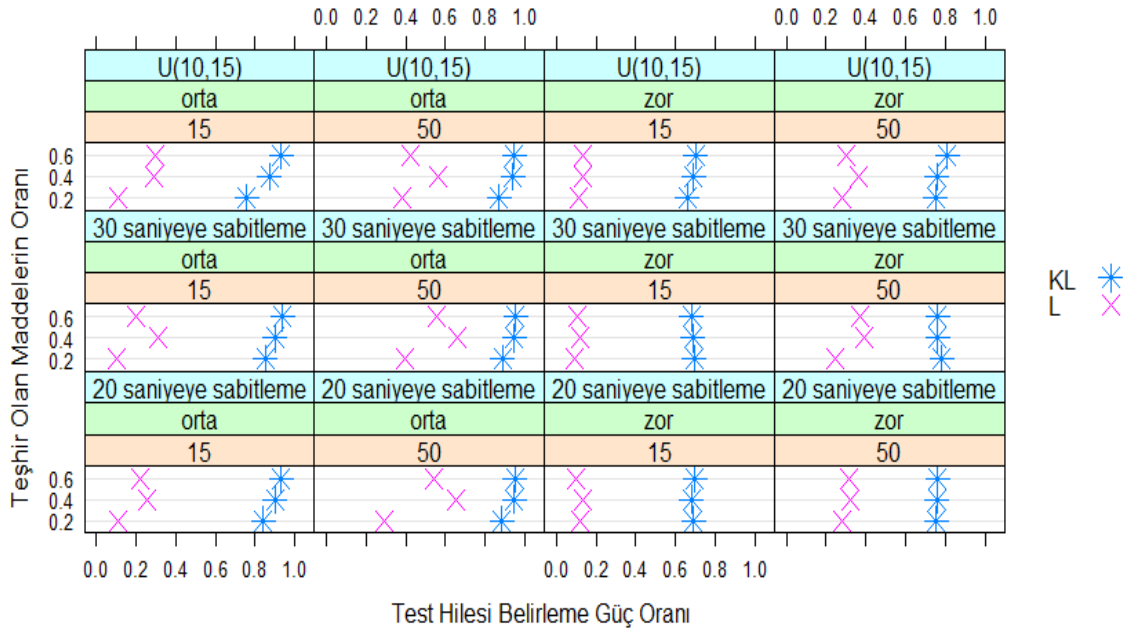
Şekil XV incelendiğinde, $L(l^t)$ birey uyum istatistiğinin test hilesi belirleme güç oranlarının ortalama değerlerinin düşük değerler aldığı ve teşhir olan maddelerin tepki sürelerinin 10 ile 15 saniye arasındaki uniform dağılımdan elde edildiği durumda en

yüksek değerini, 20 ve 30 saniyeye sabitlendiği durumlarda yakın değerler göstererek 0'a yakın değerler aldığı görülmektedir. Böylelikle, uygun olmayan madde tepki sürelerinin belirlenmesinde, madde tepkilerinin olabilirliğine dayalı yöntemlerin performansının, teşhir olan maddelerin cevaplanma sürelerinin daha kısa zaman aralıklarına sabitlendiği durumda arttığı söylenebilir. Madde tepki süreleri için *Kullback-Leibler* uzaklığının, teşhir olan maddelerin tepki süresi değişimlesinin bütün düzeylerinde yüksek ve birbirine yakın test hilesi belirleme güç oranı ortalama değerleri gösterdiği görülmekle birlikte teşhir olan maddelerin tepki sürelerinin 20 saniyeye sabitlendiği durumda en yüksek değeri aldığı görülmektedir. Bununla birlikte, HMTK-TS modelleme ile madde ön bilgisinden kaynaklı test hilesi belirlenmesinde uygun olmayan tepki örüntülerinin tespit edilmesinde, teşhir olan maddelerin tepki süresi değişimlesmesi koşulu altında, *Kullback-Leibler* uzaklığının, $L(l^t)$ birey uyum istatistiğine göre daha yüksek test hilesi belirleme güç oranı ortalama değerlerine sahip olduğu görülmektedir. Bu bulguların, Man ve diğerleri (2018)'nin çalışmalarında, madde ön bilgisine sahip olan bireylerin oranının %5 olduğu ve teşhir olan maddelerin tepki sürelerinin 10 ile 15 saniye arasındaki uniform dağılımdan elde edildiği koşulda, $L(l^t)$ birey uyum istatistiğinin test hilesi belirleme güç oranı ortalamalarının düşük değerler aldığı ve *Kullback-Leibler* uzaklığının, $L(l^t)$ birey uyum istatistiğine göre daha yüksek test hilesi belirleme güç oranı değerleri gösterdiği bulgularla benzerlik gösterdiği görülmektedir. Böylelikle, uygun olmayan madde tepki ve tepki süresi örüntülerinin belirlenmesinde, teşhir olan maddelerin tepki süresi değişimlesmesi koşulu altında, bireylerin yetenek parametrelerinin önsel ve sonsal dağılımları arasındaki farka dayalı yöntemlerin madde tepkilerinin ve tepki sürelerinin olabilirliğine dayalı yöntemlerden daha yüksek performans gösterdiği söylenebilir.

Koşulların Test Hilesi Belirleme Güç Oranına Ortak Etkisi

Test uzunluğu, teşhir olan maddelerin güçlük düzeyi, teşhir olan maddelerin oranı ve teşhir olan maddelerin tepki süresi değişimlesmesinin, yöntemlerin, madde ön bilgisinden kaynaklı test hilesini belirlemede güç oranlarına ortak etkileri şekiller kullanılarak gösterilmiştir. Değişimlenen koşulların, madde tepkileri için *Kullback-Leibler* uzaklığının ve $L(l_s^y)$ birey uyum istatistiğinin test hilesi belirleme güç oranlarına olan ortak etkilerine yönelik elde edilen bulgular Şekil XVI'da, madde tepki süreleri için *Kullback-Leibler* uzaklığının ve $L(l^t)$ birey uyum istatistiğinin test hilesini belirlemede

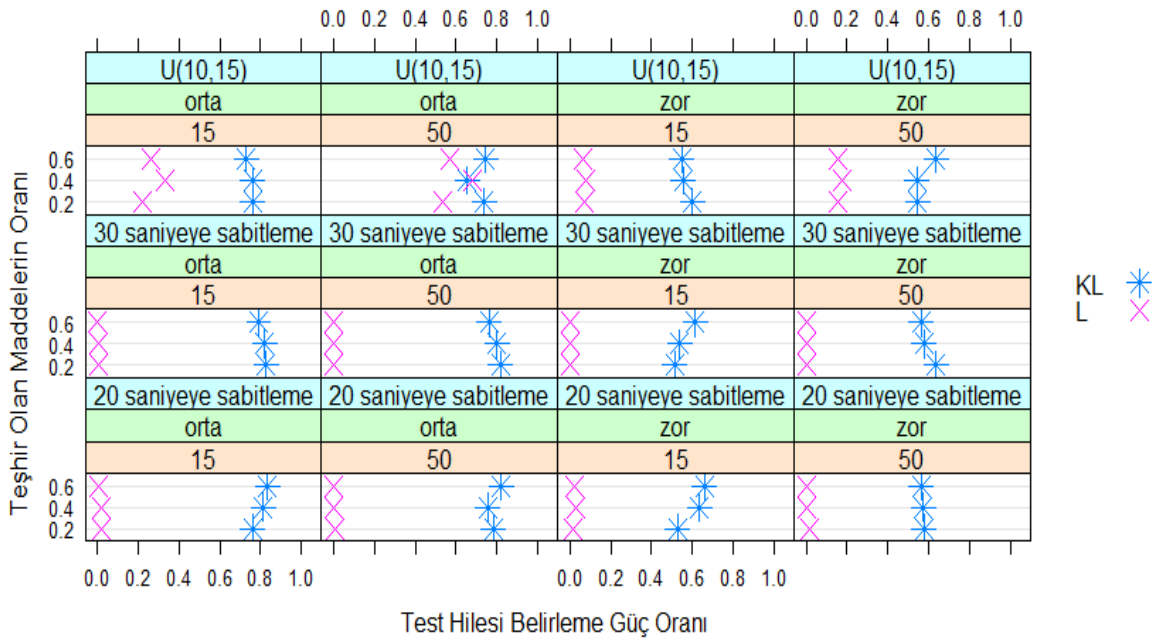
oranlarına olan ortak etkilerine yönelik elde edilen bulgular Şekil XVII'da gösterilmektedir.



Şekil XVI. Değişimlenen Koşullara Göre Uygun Olmayan Madde Tepki Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin Güç Oranları Grafiği

Değişimlenen koşulların, madde tepkileri için *Kullback-Leibler* uzaklığının ve $L(l_s^y)$ birey uyum istatistiğinin test hilesi belirleme güç oranına ortak etkisini gösteren Şekil XVI incelendiğinde, *Kullback-Leibler* uzaklığının, tüm koşullar altında, $L(l_s^y)$ birey uyum istatistiğine göre daha yüksek güç oranı değerlerine sahip olduğu görülmektedir. *Kullback-Leibler* uzaklığının, teşhir olan maddelerin tepki süresi değişimlemesinin bütün düzeylerinde, orta test uzunluğunda, orta güçlük düzeyindeki maddelerde ve teşhir olan maddelerin oranının yüksek olduğu durumda en yüksek güç oranı değerlerini gösterdiği görülmekle birlikte az fark ile teşhir olan maddelerin tepki sürelerinin 30 saniyeye sabitlendiği durumda diğer tepki süresi değişimleme düzeylerine göre daha yüksek güç oranı değerine sahip olduğu görülmektedir. Böylelikle, madde tepkileri için *Kullback-Leibler* uzaklığının güç oranı değerinin, test uzunluğu ve teşhir olan maddelerin oranı arttıkça ve teşhir olan maddelerin güçlük düzeyi azaldıkça arttığı görülmekle birlikte, teşhir olan maddelerin tepki sürelerini değişimlemede, madde tepki sürelerini daha uzun zaman değerlerine sabitlemenin, güç oranı değerlerini artırma eğiliminde olduğu sonucuna ulaşılmıştır. $L(l_s^y)$ birey uyum istatistiğinin ise

değişimlenen tüm koşullarda, güç oranı değerlerinin düşük olduğu görülmekle birlikte, yüksek güç değerlerinin, orta test uzunluğunda, orta güçlük düzeyindeki maddelerde ve teşhir olan maddelerin oranının orta düzeyde olduğu ve bu teşhir olan maddelerin tepki sürelerinin 20 ve 30 saniyeye sabitlendiği durumda gözlendiği görülmektedir. Ayrıca, teşhir olan maddelerin tepki sürelerinin 30 saniyeye sabitlendiği durumda 20 saniyeye sabitlendiği duruma göre az fark ile güç oranlarının yüksek olduğu görülmektedir. Böylelikle, $L(l_s^y)$ birey uyum istatistiğinin güç oranı değerinin, test uzunluğu arttıkça ve teşhir olan maddelerin güçlük düzeyi azaldıkça arttığı görülmekle birlikte, teşhir olan maddelerin oranının yüksek olduğu durumda azalma eğiliminde olduğu görülmektedir. Teşhir olan maddelerin tepki sürelerini değiştirmede ise madde tepki sürelerinin daha uzun zaman değerlerine sabitlenmenin, $L(l_s^y)$ birey uyum istatistiğinin güç oranı değerlerini artırma eğiliminde olduğu görülmektedir.



Şekil XVII. Değişimlenen Koşullara Göre Uygun Olmayan Madde Tepki Süresi Örüntülerinin Belirlenmesinde Kullanılan Yöntemlerin Güç Oranları Grafiği

Değişimlenen koşulların, madde tepki süreleri için *Kullback-Leibler* uzaklığının ve $L(l^t)$ birey uyum istatistiğinin test hilesi belirleme güç oranına ortak etkisini gösteren Şekil XVII incelendiğinde, *Kullback-Leibler* uzaklığının tüm koşullar altında $L(l^t)$ birey uyum istatistiğine göre daha yüksek güç oranı değerlerine sahip olduğu

görülmektedir. Bu bulguların, Man ve diğerleri (2018)'nin çalışmalarında elde edilen bulgularla benzerlik gösterdiği görülmektedir. Ayrıca, *Kullback-Leibler* uzaklığının, teşhir olan maddelerin tepki süresi değişimlemesinin bütün düzeylerinde, orta test uzunluğunda, orta güçlük düzeyindeki maddelerde ve teşhir olan maddelerin oranının yüksek olduğu durumda en yüksek güç oranı değerlerini gösterdiği görülmekle birlikte az fark ile teşhir olan maddelerin tepki sürelerinin 20 saniyeye sabitlendiği durumda en yüksek güç oranı değerine sahip olduğu görülmektedir. Böylelikle, *Kullback-Leibler* uzaklığının güç oranı değerinin, test uzunluğu ve teşhir olan maddelerin oranı arttıkça, teşhir olan maddelerin güçlük düzeyi azaldıkça arttığı görülmekle birlikte, teşhir olan maddelerin tepki sürelerini değişimlemede, madde tepki sürelerinin 20 saniyeye sabitlemenin, güç oranı değerlerini arttırma eğiliminde olduğu sonucuna ulaşılmıştır. $L(l^t)$ birey uyum istatistiğinin ise orta test uzunluğu, teşhir olan maddelerin oranının orta düzeyde, teşhir olan maddelerin ise orta güçlükte olduğu ve teşhir olan maddelerin tepki sürelerinin 10 ile 15 saniye arasındaki uniform dağılımından elde edildiği koşulu hariç diğer değişimlenen tüm koşullarda, güç oranı değerlerinin düşük olduğu görülmekle birlikte, daha yüksek güç oranı değerlerinin kısa ve orta test uzunluğunda, teşhir olan maddelerin oranının orta düzeyde olduğu ve teşhir olan maddelerin tepki sürelerinin 10 ile 15 saniye arasındaki uniform dağılımından elde edildiği durumda gözlemlendiği görülmektedir. Ayrıca, değişimlenen bütün koşullarda, teşhir olan maddelerin tepki sürelerinin 10 ile 15 saniye arasındaki uniform dağılımından elde edildiği durumda $L(l^t)$ birey uyum istatistiğinin daha yüksek güç oranları gösterdiği görülmektedir. Böylelikle, $L(l^t)$ birey uyum istatistiğinin güç oranı değerinin, test uzunluğu arttıkça ve teşhir olan maddelerin güçlük düzeyi azaldıkça arttığı görülmekle birlikte, teşhir olan maddelerin oranının yüksek olduğu durumda azalma eğiliminde olduğu görülmektedir. Teşhir olan maddelerin tepki sürelerini değişimlemede ise madde tepki sürelerinin daha kısa zaman değerlerine manipüle edilmesinin, güç oranı değerlerini arttırma eğiliminde olduğu görülmektedir.

BÖLÜM 5

SONUÇLAR VE ÖNERİLER

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

Sonuçlar

Bu çalışmada, bilgisayar temelli test uygulamalarında, madde tepkilerinin ve tepki sürelerinin birlikte hiyerarşik modellemesi (HMTK-TS) kullanılarak çoktan seçmeli maddelere ilişkin madde ön bilgisi kaynaklı test hilelerini tespit etmede, tepki süresi tabanlı *Kullback-Leibler* yaklaşımının ve madde tepkileri için $L(l_s^y)$ ve madde tepki süreleri için $L(l^t)$ birey uyum istatistiklerinin, test uzunluğu, teşhir olan maddelerin güçlük düzeyi, teşhir olan maddelerin oranı ve teşhir olan maddelerin tepki süresi değişimlemesi koşulları altında I. tip hata ve test hilesi belirleme güç oranlarının nasıl etkilendiği incelenmiştir. Çalışmadan elde edilen sonuçlar aşağıda özetlenmiştir.

1. HMTK-TS modelleme ile madde ön bilgisinden kaynaklı test hilesinin belirlenmesinde, uygun olmayan madde tepki ve tepki süresi örüntülerinin tespit edilmesinde, test uzunluğu koşulu altında, madde tepkileri ve tepki süreleri için *Kullback-Leibler* uzaklığının, madde tepkileri için $L(l_s^y)$ ve madde tepki süreleri için $L(l^t)$ birey uyum istatistiğine göre daha yüksek I. Tip hata ve güç oranı ortalama değerlerine sahip olduğu gözlenmiştir.
2. Madde tepkileri ve tepki süreleri için *Kullback-Leibler* uzaklığının, kısa ve orta test uzunluğunda yüksek I. Tip hata oranı ortalama değerleri gösterdiği görülmekle birlikte I. Tip hata oranları ortalamalarına göre madde ön bilgisi kaynaklı test hilesi yapmayan bireyleri test hilesi yapan birey olarak yanlış sınıflandırdığı gözlenmiştir. Madde tepkileri için *Kullback-Leibler* uzaklığının I. Tip hata oranları ortalamalarının test uzunluğu arttıkça küçük bir miktar arttığı, madde tepki süreleri için *Kullback-*

Leibler uzaklığının I. Tip hata oranları ortalamalarının ise küçük bir miktar azaldığı gözlenmiştir.

3. $L(l_s^y)$ ve $L(l^t)$ birey uyum istatistiklerinin I. Tip hata oranı ortalama değerlerinin, kısa ve orta test uzunluğunda düşük değerler aldığı görülmekle birlikte test uzunluğu arttıkça $L(l_s^y)$ birey uyum istatistiğinin I. Tip hata oranı ortalamalarında azalma, $L(l^t)$ birey uyum istatistiğinin I. Tip hata oranı ortalamalarında ise artış olduğu gözlenmiştir. Böylelikle, I. Tip hata oranları ortalamalarına göre $L(l_s^y)$ ve $L(l^t)$ birey uyum istatistiklerinin, madde ön bilgisi kaynaklı test hilesi yapmayan bireyleri doğru sınıflandırdığı gözlenmiştir. Bununla birlikte, test uzunluğu değişkeninin bütün düzeylerinde, L birey uyum istatistiğinin, uygun olmayan madde tepki örüntülerinin belirlenmesinde uygun olmayan madde tepki sürelerinin belirlenmesine göre daha düşük I. Tip hata oranı ortalama değerleri gösterdiği görülmüştür.
4. Madde tepkileri ve tepki süreleri için *Kullback-Leibler* uzaklığının, kısa ve orta test uzunluğunda yüksek test hilesi belirleme güç oranları ortalama değerleri gösterdiği görülmekle birlikte madde tepkileri için *Kullback-Leibler* uzaklığının test uzunluğu arttıkça güç oranları ortalama değerlerinde az farklar ile artış olduğu, madde tepki süreleri için *Kullback-Leibler* uzaklığının ise güç oranı değerlerinde az farklar ile azalma olduğu gözlenmiştir. Bununla birlikte, test uzunluğu değişkeninin bütün düzeylerinde, *Kullback-Leibler* uzaklığının, uygun olmayan madde tepki örüntülerinin belirlenmesinde uygun olmayan madde tepki süresi örüntülerinin belirlenmesine göre daha yüksek güç oranı ortalama değerleri gösterdiği görülmüştür.
5. $L(l_s^y)$ ve $L(l^t)$ birey uyum istatistiklerinin test hilesi belirleme güç oranlarının ortalama değerlerinin düşük değerler aldığı ve test uzunluğu arttıkça artış gösterdiği görülmüştür. Bununla birlikte, test uzunluğu değişkeninin bütün düzeylerinde, L birey uyum istatistiğinin, uygun olmayan madde tepki örüntülerinin belirlenmesinde uygun olmayan madde tepki sürelerinin belirlenmesine göre daha yüksek test hilesi belirleme güç oranı ortalama değerleri gösterdiği görülmüştür.
6. HMTK-TS modelleme ile madde ön bilgisinden kaynaklı test hilesinin belirlenmesinde, uygun olmayan tepki ve tepki süresi örüntülerinin tespit edilmesinde, teşhir olan maddelerin güçlük düzeyi koşulu altında, tepki süresi tabanlı

Kullback-Leibler uzaklığının, madde tepkileri için $L(l_s^y)$ ve madde tepki süreleri için $L(l^t)$ birey uyum istatistiğine göre daha yüksek I. Tip hata ve güç oranı ortalama değerlerine sahip olduğu gözlenmiştir.

7. Madde tepkileri ve tepki süreleri için *Kullback-Leibler* uzaklığının, teşhir olan maddelerin güçlük düzeyi değişkeninin bütün düzeylerinde, yüksek I. Tip hata oranı ortalama değerleri gösterdiği görülmekle birlikte I. Tip hata oranları ortalamalarına göre madde tepkileri ve madde tepki süreleri için *Kullback-Leibler* uzaklığının madde ön bilgisi kaynaklı test hilesi yapmayan bireyleri test hilesi yapan birey olarak yanlış sınıflandırdığı gözlenmiştir. Madde tepkileri için *Kullback-Leibler* uzaklığının I. Tip hata oranı ortalama değerlerinin teşhir olan maddelerin güçlük düzeyi arttıkça az farklar ile arttığı, madde tepki süreleri için *Kullback-Leibler* uzaklığının I. Tip hata oranı ortalama değerlerinin ise az farklar ile azaldığı gözlenmiştir. Bununla birlikte, teşhir olan maddelerin güçlük düzeyi değişkeninin bütün düzeylerinde, *Kullback-Leibler* uzaklığının, uygun olmayan madde tepki süresi örüntülerinin belirlenmesinde uygun olmayan madde tepkilerinin belirlenmesine göre daha düşük I. Tip hata oranı ortalama değerleri gösterdiği görülmüştür.
8. $L(l_s^y)$ ve $L(l^t)$ birey uyum istatistiklerinin I. Tip hata oranı ortalama değerlerinin, teşhir olan maddelerin güçlük düzeyi değişkeninin bütün düzeylerinde düşük değerler aldığı görülmekle birlikte teşhir olan maddelerin güçlük düzeyi arttıkça $L(l_s^y)$ ve $L(l^t)$ birey uyum istatistiklerinin I. Tip hata oranı ortalama değerlerinde artış olduğu gözlenmiştir. Böylelikle, I. Tip hata oranları ortalamalarına göre $L(l_s^y)$ ve $L(l^t)$ birey uyum istatistiklerinin, madde ön bilgisi kaynaklı test hilesi yapmayan bireyleri doğru sınıflandırdığı gözlenmiştir. Bununla birlikte, teşhir olan maddelerin güçlük düzeyi değişkeninin bütün düzeylerinde, L birey uyum istatistiğinin, uygun olmayan madde tepki süresi örüntülerinin belirlenmesinde uygun olmayan madde tepkilerinin belirlenmesine göre daha düşük I. Tip hata oranı ortalama değerleri gösterdiği görülmüştür.
9. Madde tepkileri ve tepki süreleri için *Kullback-Leibler* uzaklığının, teşhir olan maddelerin orta güçlük düzeyinde ve zor olduğu durumlarda, yüksek test hilesi belirleme güç oranları ortalama değerleri gösterdiği görülmekle birlikte teşhir olan

maddelerin güçlük düzeyi arttıkça güç oranları ortalama değerlerinde azalma olduğu gözlenmiştir. Bununla birlikte, teşhir olan maddelerin güçlük düzeyi değişkeninin bütün düzeylerinde, *Kullback-Leibler* uzaklığının, uygun olmayan madde tepki örüntülerinin belirlenmesinde uygun olmayan madde tepki süresi örüntülerinin belirlenmesine göre daha yüksek güç oranı ortalama değerleri gösterdiği görülmüştür.

10. $L(l_s^y)$ ve $L(l^t)$ birey uyum istatistiklerinin test hilesi belirleme güç oranlarının ortalama değerlerinin düşük değerler aldığı ve teşhir olan maddelerin güçlük düzeyi arttıkça azalma gösterdiği görülmüştür. Bununla birlikte, teşhir olan maddelerin güçlük düzeyi değişkeninin bütün düzeylerinde, L birey uyum istatistiğinin, uygun olmayan madde tepki örüntülerinin belirlenmesinde uygun olmayan madde tepki sürelerinin belirlenmesine göre daha yüksek test hilesi belirleme güç oranı ortalama değerleri gösterdiği görülmüştür.
11. HMTK-TS modelleme ile madde ön bilgisinden kaynaklı test hilesinin belirlenmesinde, uygun olmayan tepki ve tepki süresi örüntülerinin tespit edilmesinde, teşhir olan maddelerin oranı koşulu altında, tepki süresi tabanlı *Kullback-Leibler* uzaklığının, madde tepkileri için $L(l_s^y)$ ve madde tepki süreleri için $L(l^t)$ birey uyum istatistiğine göre daha yüksek güç oranı ortalama değerlerine sahip olduğu gözlenmiştir.
12. Madde tepkileri ve tepki süreleri için *Kullback-Leibler* uzaklığının, teşhir olan maddelerin oranı değişkeninin bütün düzeylerinde yüksek test hilesi belirleme güç oranları ortalama değerleri gösterdiği görülmekle birlikte teşhir olan maddelerin oranının arttıkça güç oranları ortalama değerlerinde de artışın olduğu gözlenmiştir. Bununla birlikte, teşhir olan maddelerin oranı değişkeninin bütün düzeylerinde, *Kullback-Leibler* uzaklığının, uygun olmayan madde tepki örüntülerinin belirlenmesinde uygun olmayan madde tepki süresi örüntülerinin belirlenmesine göre daha yüksek güç oranı ortalama değerleri gösterdiği görülmüştür.
13. $L(l_s^y)$ ve $L(l^t)$ birey uyum istatistiklerinin test hilesi belirleme güç oranlarının ortalama değerlerinin düşük değerler aldığı ve teşhir olan maddelerin oranının orta düzeyde olduğu durumda en yüksek güç oranı ortalama değerlerini gösterdiği görülmüştür. Bununla birlikte, teşhir olan maddelerin oranı değişkeninin bütün

düzeylerinde, L birey uyum istatistiğinin, uygun olmayan madde tepki örüntülerinin belirlenmesinde uygun olmayan madde tepki sürelerinin belirlenmesine göre daha yüksek test hilesi belirleme güç oranı ortalama değerleri gösterdiği görülmüştür.

14. HMTK-TS modelleme ile madde ön bilgisinden kaynaklı test hilesinin belirlenmesinde, uygun olmayan tepki ve tepki süresi örüntülerinin tespit edilmesinde, teşhir olan maddelerin tepki süresi değişimlemesi koşulu altında, tepki süresi tabanlı *Kullback-Leibler* uzaklığının, madde tepkileri için $L(l_s^y)$ ve madde tepki süreleri için $L(l^t)$ birey uyum istatistiğine göre daha yüksek güç oranı ortalama değerlerine sahip olduğu gözlenmiştir.
15. Madde tepkileri ve tepki süreleri için *Kullback-Leibler* uzaklığının, teşhir olan maddelerin tepki süresi değişimlemesi değişkeninin bütün düzeylerinde, yüksek test hilesi belirleme güç oranları ortalama değerleri gösterdiği görülmekle birlikte teşhir olan maddelerin tepki sürelerinin 20 saniyeye sabitlendiği durumda en yüksek güç oranları ortalama değerlerini gösterdiği gözlenmiştir. Bununla birlikte, teşhir olan maddelerin tepki süresi değişimlemesi değişkeninin bütün düzeylerinde, *Kullback-Leibler* uzaklığının, uygun olmayan madde tepki örüntülerinin belirlenmesinde uygun olmayan madde tepki süresi örüntülerinin belirlenmesine göre daha yüksek güç oranı ortalama değerleri gösterdiği görülmüştür.
16. $L(l_s^y)$ ve $L(l^t)$ birey uyum istatistiklerinin test hilesi belirleme güç oranlarının ortalama değerlerinin düşük değerler aldığı ve $L(l^t)$ birey uyum istatistiğinin, teşhir olan maddelerin tepki sürelerinin 10 ile 15 saniye arasındaki uniform dağılımdan elde edildiği durumda, $L(l^t)$ birey uyum istatistiğinin ise teşhir olan maddelerin tepki sürelerinin ise 30 saniyeye sabitlendiği durumda en yüksek güç oranı ortalama değerlerini gösterdiği görülmüştür. Böylelikle, L birey uyum istatistiğinin, uygun olmayan madde tepki örüntülerinin belirlenmesinde, teşhir olan maddelerin tepki sürelerinin daha kısa zaman değeri aralıklarında olmasının, uygun olmayan madde tepki süresi örüntülerinin belirlenmesinde ise tepki sürelerinin daha uzun zaman değerine sabitlenmesinin test hilesi belirleme güç oranı ortalama değerlerini arttırdığı görülmüştür.

17. Genel sonuç olarak, HMTK-TS modelleme ile madde ön bilgisinden kaynaklı test hilesinin belirlenmesinde, test uzunluğu, teşhir olan maddelerin güçlük düzeyi, oranı ve tepki süresi değişimlemesi koşulları altında, uygun olmayan madde tepki ve madde tepki süresi örüntülerinin tespit edilmesinde, *Kullback-Leibler* uzaklığının L birey uyum istatistiğine göre daha güçlü olduğu; ancak yüksek I. Tip hata oranı barındırdığı görülmüştür.

Öneriler

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

Uygulayıcılara Öneriler

1. Araştırmadan elde edilen bulgular ışığında, bireylerin, madde tepkileri, madde tepki süreleri ve ek kanıtlar kullanılarak madde ön bilgisi kaynaklı uygun olmayan madde tepki davranışlarının belirlenmesinde, *Kullback-Leibler* uzaklığının eşik değerinin farklı kesme puanı yöntemleri ile hesaplanıp I. Tip hata oranının kontrol altına alındığı koşullar gözetilerek kullanılması önerilmektedir. Özellikle geniş ölçekli ve yüksek riskli bilgisayar temelli test uygulamalarında, uygun olmayan madde tepki ve madde tepki süresi örüntülerine sahip bireylerin belirlenmesinde, yalancı pozitif oranları da içerebilen *Kullback-Leibler* uzaklık ölçüsü gibi yaklaşımların kullanımına ek olarak, geliştirilen yazılımlar ile elde edilen detaylı bilgilerin (bireylerin kameralarına erişilip göz hareketleri ve duruş pozisyonları vb.) de değerlendirilerek çok yönlü araştırmaların yapılması önerilmektedir.
2. Madde ön bilgisinden kaynaklı test hilesini belirlemede, uygun olmayan madde tepkileri ve madde tepki süresi örüntülerinin belirlenmesinde, *Kullback-Leibler* uzaklığının 1000 örneklem büyüklüğünde kullanımında, I. Tip hata oranı kontrolünün birçok koşulda sağlanamadığı göz önünde bulundurularak, bu yöntemin, 1000'den küçük örneklem büyüklükleri için kullanımı önerilmemektedir.

3. Araştırmanın sonucunda, madde ön bilgisi kaynaklı test hilesini belirlemede, uygun olmayan madde tepki ve tepki süresi örüntülerinin tespit edilmesinde, *Kullback-Leibler* uzaklığının ve L birey uyum istatistiğinin, zor maddelerde, daha yüksek I. Tip hata oranı değerlerine sahip olduğu görülmektedir. Böylelikle, teşhir olan maddelerin zor olduğu test uygulamalarında, uygun olmayan madde tepki ve tepki örüntülerine sahip olan bireylerin belirlenmesinde, yöntemlerin kullanımında, yüksek I. Tip hata oranı değerlerinin göz önünde bulundurulması önerilmektedir.
4. Araştırmanın sonucunda, madde tepkileri ve tepki süreleri için *Kullback-Leibler* uzaklığının, teşhir olan maddelerin tepki süresi değişimlemesi değişkeninin bütün düzeylerinde, test hilesi belirleme performansının yüksek olduğu görülmekle birlikte teşhir olan orta güçlük düzeyinde ve zor maddelerin tepki sürelerinin 20 saniyeye sabitlendiği durumda en yüksek performansa sahip olduğu görülmüştür. Böylelikle, teşhir olan maddelerin tepki sürelerinin 20 saniyeden daha az olduğu durumlarda, uygun olmayan madde tepki ve tepki süresi örüntülerinin belirlenmesinde *Kullback-Leibler* uzaklığının kullanılması önerilmemektedir.
5. Türkiye’de, günümüzde uygulaması devam eden, ÖSYM tarafından gerçekleştirilen geniş ölçekli Elektronik Yabancı Dil Sınavı (e-YDS) ve MEB tarafından gerçekleştirilen Elektronik Motorlu Taşıtları Sürücü Adayları Sınavı (e-sınav) uygulamalarında, bireylerin test maddelerine verdikleri tepki sürelerinin de kayıt altına alınarak madde ön bilgisinden kaynaklı test hilesini belirlemede, uygun olmayan madde tepkilerine ve tepki örüntülerine sahip olan bireylerin belirlenmesine yönelik analizlerin, araştırmanın bulguları doğrultusunda, *Kullback-Leibler* uzaklığının I. Tip hata oranının kontrol altına alındığı koşullar gözetilip kullanılarak yapılması ve sınav verilerinin analiz raporunda, bu analizlere yer verilmesi önerilmektedir. Böylelikle, kamuoyunda, ÖSYM ve MEB tarafından uygulanan geniş ölçekli ve yüksek riskli sınavların objektifliğine olan güvenin artacağı düşünülmektedir.

Araştırmacılara Öneriler

1. Araştırmanın sonucunda, bütün koşullarda, madde ön bilgisi kaynaklı test hilesinin belirlenmesinde, uygun olmayan madde tepki ve tepki süresi örüntülerinin tespit edilmesinde, madde tepkilerinin ve tepki sürelerinin olabilirliğine dayalı olan $L(l_s^y)$ ve $L(l^t)$ birey uyum istatistiklerinin performansının düşük olduğu ancak I. Tip hata oranını kontrol altına aldığı görülmüştür. Böylelikle, simülasyon çalışması olarak gerçekleştirilen bu çalışma, gerçek veriler kullanılarak L birey uyum istatistiğinin performansının artırılmasına yönelik olarak farklı alfa düzeylerinde tekrarlanabilir.
2. Çalışmada, sabit değişkenler olarak ele alınan, örneklem büyüklüğü ve test hilesi yapan birey oranı değişkenlerin düzeyleri, geniş ölçekli sınav uygulamalarındaki koşullar gözetilerek farklılaştırılıp, araştırmanın sonucunda, madde ön bilgisi kaynaklı test hilesini belirlemede düşük performans gösteren madde tepkileri ve madde tepki sürelerinin olabilirliğine dayalı L birey uyum istatistiğinin performansının artırılmasına yönelik simülasyon çalışmalarının yapılması önerilmektedir.
3. Araştırmanın sonucunda, madde tepkilerinin ve madde tepki sürelerinin birlikte hiyerarşik olarak modellenerek madde ön bilgisinden kaynaklı test hilesinin belirlenmesinde, test uzunluğu, teşhir olan maddelerin güçlük düzeyi, oranı ve tepki süresi değişimlemesi koşulları altında, uygun olmayan madde tepki ve madde tepki süresi örüntülerinin tespit edilmesinde, *Kullback-Leibler* uzaklığının L birey uyum istatistiğine göre daha güçlü olduğu; ancak yüksek I. Tip hata oranı barındırdığı görülmüştür. Böylelikle, madde tepkileri ve madde tepki süreleri için *Kullback-Leibler* uzaklık değerinin, bireylerin yetenek parametrelerindeki küçük farklılıklara karşı duyarlı olduğu ve uygun olmayan madde tepki ve tepki sürelerinin belirlenmesinde kullanılan kesme puanı yönteminin (ROC analiz yöntemlerinden *Youden İndeksi*) konservatif olduğu görülmüştür. Bu nedenle, benzer koşullarda, uygun olmayan madde tepkilerinin ve madde tepki sürelerinin belirlenmesinde kullanılan *Kullback-Leibler* uzaklık değerlerinin hesaplanmasında, birey yetenek parametrelerinin kestirimi işleminin yetenek kestirimlerindeki farklılıkları en aza

indirgeyecek şekilde farklı paket programlarında yapılarak ve kesme puanının belirlenmesinde farklı yöntemler kullanılarak I.Tip hata oranının kontrol altına alınmasını sağlamaya yönelik simülasyon çalışmalarının gerçekleştirilmesi önerilmektedir.

4. Araştırmanın sonuçları, madde tepkileri ve tepki süreleri için Kullback-Leibler uzaklığının kullanıldığı madde ön bilgisi kaynaklı test hilesi belirleme sürecinde, test hilesi yapılmadığı halde “test hilesi yapıldı” kararının verilme olasılığının yüksek olduğunu göstermektedir. Bu nedenle, benzer bir araştırmanın, madde ön bilgisi kaynaklı test hilesini belirlemede, uygun olmayan tepkileri ve tepki süresi örüntülerini tespit etmede kullanılan farklı uzaklık ölçüleri ve birey uyum istatistikleri ile gerçekleştirilmesi önerilmektedir.
5. Araştırmada, hile senaryosunda değişimlenen faktörler ve düzeylerinin yöntemlerin, test hilesini belirlemede. I. tip hata ve güç oranlarına olan etkisini gözlemlemek için örneklemden %5 oranında düşük yetenekli bireylerden madde ön bilgisine sahip bireyler seçilmiştir. Araştırmada kullanılan yöntemlerin performans ve I.Tip hata oranlarına dair ek bilgiler için , madde ön bilgisi kaynaklı test hilesi yapan bireylerin yetenek düzeyleri değişkeninin düzeyleri, çok düşük, orta ve yüksek yetenek düzeyi olarak farklılaştırılarak araştırmanın genişletilmesi önerilmektedir.
6. Araştırmadan elde edilen bulgular ışığında, madde ön bilgisi kaynaklı test hilesinin belirlenmesinde, uygun olmayan madde tepki ve tepki süresi örüntülerinin tespit edilmesinde, *Kullback-Leibler* uzaklığının I.Tip hata oranının kontrol altına alındığı koşullar gözetilip kopyacı ve kaynak çiftleri oluşturularak çeşitli koşullarda, cevap kopyalamadan kaynaklı test hilesini belirlemedeki performansını belirlemeye yönelik simülasyon çalışmalarının yapılması önerilmektedir.
7. Madde tepkilerinin ve madde tepki sürelerinin birlikte hiyerarşik olarak modellenmesiyle, cevap kopyalamadan ve madde ön bilgisinden kaynaklı grup düzeyindeki test hilelerini belirlemede, araştırma bulguları ışığında, uygun olmayan madde tepkileri ve tepki sürelerinin belirlenmesinde performansı yüksek olan *Kullback-Leibler* uzaklığının I.Tip hata oranının kontrol altına alındığı koşullar gözetilerek simülasyon çalışmalarının yapılması önerilmektedir.

8. Madde ön bilgisinden kaynaklı test hilesinin belirlenmesinde, uygun olmayan madde tepki ve tepki süresi örüntülerinin tespit edilmesinde, farklı madde tepki süresi modelleri kullanılarak *Kullback-Leibler* uzaklığının I.Tip hata oranının kontrol altına alındığı koşullar gözetilerek kullanımı ,önerilmektedir.
9. Araştırmanın sonucunda, madde tepkilerinin ve madde tepki sürelerinin birlikte hiyerarşik olarak modellenerek madde ön bilgisinden kaynaklı test hilesinin belirlenmesinde, teşhir olan maddelerin oranının %20, %40 ve %60 olduğu durumda, uygun olmayan madde tepki ve madde tepki süresi örüntülerinin tespit edilmesinde, *Kullback-Leibler* uzaklığının L birey uyum istatistiğine göre daha güçlü olduğu; ancak yüksek I. Tip hata oranı barındırdığı görülmüştür. Böylelikle, teşhir olan maddelerin oranının %80 ve üzeri olduğu durumlarda, yöntemlerin performansının ortaya konulmasına yönelik çalışmaların yapılması önerilmektedir.
10. Bu çalışmada, madde tepkileri 3PLM ile madde tepki süreleri ise log- normal model ile modellenmiştir. Madde tepkilerinin STM ile madde tepki sürelerinin de log-normal model ile modellendiği durumda, madde ön bilgisi kaynaklı test hilesinin belirlenmesinde, uygun olmayan madde tepki ve madde tepki süresi örüntülerinin tespit edilmesinde, *Kullback-Leibler* uzaklığının ve L birey uyum istatistiğinin performanslarını belirlemeye yönelik benzer çalışmaların yapılması önerilmektedir.

KAYNAKLAR

- Albert, J., & Chib, S.(1993). Bayesian analysis of binary and polychotomous response data. *Journal of the American Statistical Association*, 80(422), 669-679. Retrieved from <http://www.jstor.org/stable/2290350>
- Alpar, R. (2013). *Uygulamalı çok değişkenli istatistiksel yöntemler*. Ankara: Detay Yayıncılık.
- Altuner, F. (2019). *Madde İstatistikleri ve Madde Tepki Süreleri Arasındaki İlişkinin İncelenmesi*.(Yayımlanmamış Yüksek Lisans Tezi). Mersin Üniversitesi, Mersin, Türkiye.
- Angoff, W.(1974). The development of statistical indices for detecting cheaters. *Journal of American Statistical Association*, 69(34), 44-49. doi: 10.2307/2285498
- Armstrong, R., & Shi, M. (2009). A parametric cumulative sum statistic for person fit. *Applied Psychological Measurement*, 33(5), 391-410. doi:10.1177/0146621609331961
- Balta E., Uçar A., & Sahin A. (2019, June). *Detection of cheating behavior in online unproctored CATs via a validation test*. Paper presented at the 2019 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: ÖSYM Yayınları.
- Bellezza, F., & Bellezza, S.(1989). Detection of cheating on multiple choice tests by using error-similarity analysis. *Teaching of Psychology*, 16(3), 151-155. Retrieved from https://journals.sagepub.com/doi/10.1207/s15328023top1603_15
- Belov, D. I. (2011). Detection of answer copying based on the structure of a high-stakes test. *Applied Psychological Measurement*, 35(7), 495–517. doi:10.1177/0146621611420705
- Belov, D. I. (2013). Detection of test collusion via Kullback–Leibler divergence. *Journal of Educational Measurement*, 50 (2), 141–163. doi:10.1111/ 12008
- Belov, D. I. (2014). Detecting item preknowledge in computerized adaptive testing using information theory and combinatorial optimization. *Journal of Computerized Adaptive Testing* , 2(3), 37–58. doi:10.7333/jcat.v2i0.36

- Belov, D. (2015). Robust detection of examinees with aberrant answer changes. *Journal of Educational Measurement*, 52(4), 437–456. doi: 10.1111/jedm.12094
- Belov, D. (2016). Comparing the performance of eight item preknowledge detection statistics. *Applied Psychological Measurement*, 40(2), 83–90. doi: 10.1177/0146621615603327
- Belov, D., & Armstrong, R. (2010). Automatic detection of answer copying via Kullback–Leibler divergence and K-index. *Applied Psychological Measurement*, 34(6), 379–392. doi/10.1177/0146621610370453.
- Belov, D., & Armstrong, R. (2011). Distributions of the Kullback-Leibler divergence with applications. *British Journal of Mathematical and Statistical Psychology*, 64, 291–309. doi: 10.1348/000711010X522227
- Belov, D., Pashley, P., Lewis, C., & Armstrong, R. (2007). Detecting aberrant responses with Kullback–Leibler distance. In K. Shigemasu, A. Okada, T. Imaizumi & T. Hoshino (Eds.), *New trends in psychometrics*, (pp. 7–14). Tokyo, Japan: Universal Academy Press.
- Boughton, K., Smith, J., & Ren, H. (2017). Using response time data to detect compromised items and/or people. In G. J. Cizek & J. A. Wollack (Eds.), *Handbook of detecting cheating on tests*, (pp. 177–190). New York: Routledge.
- Bock, R. (1972). Estimating item parameters and latent ability when responses are scored in two or more nominal categories. *Psychometrika*, 37, 29–51. Retrieved from <https://link.springer.com/article/10.1007/BF02291411>
- Box, G. E. (1980). Sampling and Bayes' inference in scientific modelling and robustness. *Journal of the Royal Statistical Society. Series A (General)*, 143(4), 383–430. doi: 10.2307/2982063
- Box, G. E., Hunter, J. S., & Hunter, W. G. (1978). *Statistics for experimenters*. New York: Wiley.
- Birnbaum, A. (1968) Some latent trait models and their use in inferring an examinee's ability. In F.M Lord, M. R., Novick & A. Birnbaum (Eds.), *Statistical theories of mental test scores*, (pp.397–479). New York: Addison-Wesley.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. New York: Springer.
- Bishop, K. L., & Stephens, C.N. (2013). *Detecting unusual item response patterns based on likelihood of answer*. Paper presented at the 2013 Statistical Detection of Potential Test Fraud Conference. Madison, Wisconsin. Abstract retrieved from https://img1.wsimg.com/blobby/go/71ff6051-7987-4285-89a3-874c4bfcc71c/downloads/1d2ljaegt_777967.pdf

- Bhola, H.S. (1994). Adult education policy and performance: An international perspective. *Review of Policy Research*. 13(4), 315-318. doi: 10.1111/j.1541-1338.1994.tb00609.
- Bhola, D. S., Plake, B. S., & Roos, L. L. (1993, October). *Setting an optimum time limit for a computer-administered test*. Paper presented at the 1993 annual meeting of the Midwestern Educational Research Association, Chicago, IL. Abstract retrieved from <https://www.worldcat.org/title/abstracts-of-papers-of-the-annual-meeting/oclc/20053594>
- Bradlow, E. T., Weiss, R. E., & Cho, M. (1998). Bayesian identification of outliers in computerized adaptive testing. *Journal of the American Statistical Association*, 93(443), 910-919. Retrieved from <https://amstat.tandfonline.com/doi/abs/10.1080/01621459.1998.10473747>
- Burke, O. (2012). *MCMC Output Analysis*. Retrieved from: <http://www.stats.ox.ac.uk/~burke/Autocorrelation/MCMC%20Output.pdf>
- Casella, G., & Berger, R. L. (2002). *Statistical inference*. Pacific Grove, CA: Duxbury.
- Charalambides, C. A. (2005). *Combinatorial methods in discrete distributions*. New York: John Wiley & Sons, Inc
- Cizek, G., & Wollack, J. (2017). Identification of item preknowledge by the methods of information theory and combinatorial optimization. In G. Cizek & J. Wollack (Eds.), *Handbook of Quantitative Methods for Detecting Cheating on Tests*, (pp.217-233). New York:Routledge.
- Clark, J. Skorupski, W., & Murphy, S. (2017). Using nonlinear regression to identify unusual performance level classification rates. In Cizek, G. and Wollack, J. (Eds), *Handbook of Quantitative Methods for Detecting Cheating on Tests*, (pp 245-261). New York:Routledge.
- Cox, D. R., & Hinkley, D. V. (1974). *Theoretical statistics*. London, England: Chapman and Hall.
- Cowles, M. K., & Carlin, B. (1996). Markov chain monte carlo convergence diagnostics: a comparative review. *Journal of the American Statistical Association*, 91(434), 883-904. doi:10.1080/01621459.1996.10476956
- Crocker, L., & Algina, J. (1986). *Introduction to classical and modern test theory*. New York: Holt, Rinehart & Winston.
- Çokluk, Ö. S., Şekercioğlu, G. ve Büyüköztürk, S. (2014). *Sosyal Bilimler için Çok Değişkenli İstatistik: SPSS ve LISREL uygulamaları*. Ankara: Pegem Akademi Yayıncılık.
- de La Torre, J., & Deng, W. (2008). Improving person-fit assessment by correcting the ability estimate and its reference distribution. *Journal of Educational Measurement*, 45(2), 159–177. doi:10.1111/j.1745-3984.2008.00058.x

- DeMars, C. (2010) *Item Response Theory: Understanding Statistics Measurement*. Oxford University Press, Oxford.
- Deng, W. (2007). *An innovative use of the standardized log-likelihood statistic to evaluate person fit*. (Unpublished doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3277289).
- Donlan, T. F., & Fischer, F. E. (1968). An index of an individual's agreement with group determined item difficulties. *Educational and Psychological Measurement*, 28, 105–113. doi:10.1177/001316446802800110
- Dimitrov, D. M., & Smith, R. M. (2006). Adjusted rasch person-fit statistics. *Journal of Applied Measurement*, 7(2), 170-183. Retrieved from <https://www.researchgate.net/publication/7147714>
- Dinno, A. (2018). Package 'paran'. Retrieved from <https://cran.r-project.org/web/packages/paran/paran.pdf>
- 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. doi:10.1177/014662168701100105
- Drasgow, F., Levine, M. V., & McLaughlin, M. E. (1991). Appropriateness measurement for some multidimensional test batteries. *Applied Psychological Measurement*, 15 (2), 171–191. doi: 10.1177/014662169101500207
- Drasgow, F., Levine, M., & Williams, E. (1985). Appropriateness measurement with polychotomous item response models and standardized indices. *British Journal of Mathematical and Statistical Psychology*, 38(1), 67-86. doi 10.1111/j.2044-8317.1985.tb00817.x
- Drasgow, F., Levine, M. V., & Zickar, M. J. (1996). Optimal identification of mismeasured individuals. *Applied Measurement in Education*, 9(1), 47-64. doi: 10.1207/s15324818ame0901_5
- Eckerly, C. A., Babcock, B., & Wollack, J. A. (2015, April) *Preknowledge detection using ascalepurified deterministic gated IRT model*. Paper presented at the 2015 annual meeting of the National Conference on Measurement in Education, Chicago, IL. Abstract retrieved from https://higherlogicdownload.s3.amazonaws.com/NCME/c53581e4-9882-4137-987b4475f6cb502a/UploadedImages/Documents/NCME_2015_Program_WEB3.pdf
- Eckerly, C., Smith, R., & Lee, Y. (2018, October). *An introduction to item preknowledge detection with real data applications* . Paper presented at the 2018 Conference on Test Security, Park City, UT, United States. Abstract retrieved from <https://conferenceontestsecurity.org/cots-2018-1>

- Egberink, I., Meijer, R., Veldkamp, B., Schakel, L., & Smid, N. (2010). Detection of aberrant item score patterns in a computerized adaptive test: An empirical example using the CUSUM. *Personality and Individual Differences*, 48, 921-925. doi:10.1016/j.paid.2010.02.023
- Emons, W. H. M. (2008). Nonparametric person-fit analysis of polytomous item scores. *Applied Psychological Measurement*, 32(3), 224-247. doi: 10.1177/0146621607302479
- Flach, P. (2010). ROC Analysis. In Book, *Encyclopedia of Machine Learning* (pp. 869-875), New York: Springer.
- Fox, J.-P. (2010). *Bayesian item response modeling: Theory and applications*. New York: Springer.
- Fox, J.-P., Klein Entink, R., & Linden, W. (2007). Modeling of responses and response times with the package CIRT. *Journal of Statistical Software*, 20(7), 1-14. Retrieved from <https://www.jstatsoft.org/index>
- Fox, J.-P., Klotzke, K., & Klein Entink, R. H. (2019). Package 'LNIRT: Lognormal response time item response theory models'. Retrieved from <https://cran.r-project.org/web/packages/LNIRT/LNIRT.pdf>
- Fox, J.-P., & Marianti, S. (2017). Person-fit statistics for joint models for accuracy and speed. *Journal of Educational Measurement*, 54(2), 243-262. doi: 10.1111/jedm.12143
- Frary, R. B. (1993). Statistical detection of multiple-choice answer copying: Review and commentary. *Applied Measurement in Education*, 6(2), 153-165. doi: 10.1207/s15324818ame0602_4
- Frary, R. B., Tideman, T. N., & Watts, T. M. (1977). Indices of Cheating on Multiple-Choice Tests. *Journal of Educational Statistics*, 2(4), 235-256. doi:10.3102/10769986002004235
- Gaertner, M., & McBride, Y. (2017). Detecting unexpected change in pass rates – a comparison of two statistic approaches. In G. Cizek & Wollack, J. (Eds), *Handbook of Quantitative Methods for Detecting Cheating on Tests*, (pp 262-279). Routledge: New York.
- Geyer, C. J. (1992). Practical markov chain monte carlo. *Statistical Science*, 7(4), 473-511. Retrieved from <https://www.jstor.org/stable/2246094>
- Geweke, J. (1992). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, Dawid, A. P. & Smith, A. F. M. (Eds.), *Bayesian Statistics*, (pp.169-193). Oxford: Oxford University Press.
- Gill, J. (2015). *Bayesian Methods: A Social and the Behavioural Sciences*. Florida: Chapman& Hall /CRC Press.

- Gulliksen, H. (1950). *Theory of mental tests*. NewYork: John Wiley & Sons, Inc.
- Guo, J., & Drasgow, F. (2010). Identifying cheating on unproctored internet Tests: The Z-test and the likelihood ratio test. *International Journal of Selection and Assessment*, 18(4), 351-364. doi: 10.1111/j.1468-2389.2010.00518.x
- Guttman, L. (1944). A basis for scaling qualitative data. *American Sociological Review*, 9(2), 139-150. doi: 10.2307/2086306
- Guttman, L. (1950). The basis for scalogram analysis. In S. A. Stouffer, L. Guttman, E. A. Suchman, P. F. Lazarsfeld, S. A. Star & J. A. Clausen (Eds.), *Measurement and prediction*, (pp. 60–90). Princeton, NJ: Princeton University Press.
- Haberman, S., & Lee, Y. (2017). *A statistical procedure for testing unusually frequent exactly matching responses and nearly matching responses*. (Research Report No:RR-17-23). Princeton, NJ: Educational Testing Service. doi:10.1002/ets2.12150.
- Han, N., & Hambleton, R. (2004). *Detecting exposed test items in computer-based testing*. Paper presented at the 2004 annual meeting of the National Council on Measurement in Education, San Diego, CA. Abstract retrieved from <http://iacat.org/content/detecting-exposed-test-items-computer-based-testing>
- Harnisch, D. L., & Linn, R. L. (1981). Analysis of item response patterns: Questionable test data and dissimilar curriculum practices. *Journal of Educational Measurement*, 18(3), 133-146. doi: 10.1111/j.1745-3984.1981.tb00848.x
- Harpp, D.N., & Hogan, J.J. (1993). Crime in the classroom - detection and prevention of cheating on multiple-choice exams. *Journal of Chemical Education*, 70 (4), 306-311. doi:10.1021/ed070p306
- He, Q., & Tymms, P. (2014). The principal axis approach to value added calculation. *Educational Research and Evaluation*, 20(1), 25-43. doi: 10.1080/13803611.2013.860036
- He, Q., & Stockford, I. (2015). *Possible statistical techniques for identifying centres with unusual performance on non-exam assessment in GCSE Computer Science for moderation*. Ofqual Internal Report. <https://www.gov.uk/government/organisations/ofqual> adresinden erişilmiştir.
- Heck, R.H., & Thomas, S. L. (2015). *An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus*. NewYork: Routledge.
- Holland, P.W. (1996). Assessing unusual agreement between the incorrect answers of two examinees using the K-index: statistical theory and empirical support (*Research Report RR-94-4*) Princeton, NJ: Educational Testing Service. doi: 10.1002/j.2333- 8504.1996.tb01685.x

- Hosmer, D. W., & Lemeshow, S. (2013). *Applied logistic regression*. New York:Wiley.
- Huang, T. W. (2012). Aberrance detection powers of the BW and person-fit indices. *Educational Technology and Society*, 15(1), 28–37. Retrieved from https://www.researchgate.net/publication/286950181_Aberrance_Detection_Powers_of_the_BW_and_Person-Fit_Indices
- Hui, H.-F. (2008). *Stability and sensitivity of a model-based person-fit index in detecting item pre-knowledge in computerized adaptive test*. (Unpublished doctoral dissertation). Hong Kong: The Chinese University of Hong Kong.). Available from ProQuest Dissertations and Theses (UMI No.3377981.)
- Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. *The Quarterly Journal of Economics*, 118, 843–877. Retrieved from <https://www.nber.org/papers/w9413>
- Kaftandjieva, F. (2010). *Methods for setting cut scores in criterion-referenced achievement tests A comparative analysis of six recent methods with an application to tests of reading in EFL, EALTA, Cito, Arnhem*. https://www.ealta.eu.org/documents/resources/FK_second_doctorate.pdf adresinden erişilmiştir.
- Kane, M. T., & Brennan, R. L. (1980). Agreement coefficients as indices of dependability for domain-referenced tests. *Applied Psychological Measurement*, 4(1), 105–126.doi: 10.1177/014662168000400111
- Karabatsos, G. (2003). Comparing the aberrant response detection performance of thirty-six person-fit statistics. *Applied Measurement in Education*, 16(4), 277–298. doi:10.1207/s15324818ame1604_2
- Kasli, M., & Zopluoglu, C. (2018, October). *Do people with item pre-knowledge really respond faster to items they had prior access? An empirical investigation* Paper presented at the Conference on Test Security, Park City, UT, United States. Abstract retrieved from <https://conferenceontestsecurity.org/cots-2018-1>
- Kasli, M., Zopluoglu, C., & Toton, S. L. (2020,May). *A deterministic gated lognormal response time model to identify examinees with item preknowledge*. Retrieved from <https://doi.org/10.31234/osf.io/bqa3t>
- Keeves, J., & Alagumalai, S. (1999). Item banking. In G. Masters & J. Keeves (Eds.), *Advances in Measurement in Educational Research and Assessment*, (pp. 23-42). The Netherlands: Elsevier Science.
- Kim, D., Woo, A., & Dickison, P. (2017). Identifying and investigating aberrant responsesusing psychometrics-based and machine learning-based approaches. In G. Cizek & Wollack, J. (Eds.), *Handbook of Quantitative Methods for Detecting Cheating on Tests*, (pp.102-133). New York:Routledge.
- Kingston, N., & Clark, A. (2014). *Test fraud: Statistical detection and methodology*. New York, NY: Routledge.

- Klauer, K. C. (1991). An exact and optimal standardized person test for assessing consistency with the Rasch model. *Psychometrika*, 56(2), 535–547. doi:10.1007/bf02294459
- Klauer, K. C., & Rettig, K. (1990). An approximately standardized person test for assessing consistency with a latent trait model. *British Journal of Mathematical and Statistical Psychology*, 43(2), 193–206. doi: 10.1111/j.2044-8317.1990.tb00935.x
- Klein Entink, R. H., Fox, J.-P., & van der Linden, W. J. (2009). A multivariate multilevel approach to the modeling of accuracy and speed of test takers. *Psychometrika*, 74(1), 21–48. doi: 10.1007/S11336-008-9075-Y.
- Klein Entink, R. H., Kuhn, J. T., Hornke, L. F., & Fox, J.-P. (2009). Evaluating cognitive theory: A joint modeling approach using responses and response times. *Psychological Methods*, 14(1), 54–75. doi: 10.1037/a0014877.
- Krzanowski, W.J., & Hand, D. J. (2009). *ROC curves for continuous data*. Boca Raton: Chapman & Hall/CRC Press.
- Kullback, S., & Leibler, R. A. (1951). On information and sufficiency. *Annals of Mathematical Statistics*, 22(1), 79–86. doi:10.1214/aoms/1177729694
- Lee, S. Y. (2018). *A mixture model approach to detect examinees with item preknowledge*. (Unpublished doctoral dissertation). Available from ProQuest Dissertations and Theses (UMI No.10830593)
- Lee, Y., & Haberman, S. (2016). Investigating test-taking behaviors using timing and process data. *International Journal of Testing*, 16(3), 240–267. doi:10.1080/15305058.2015.108538
- Lehmann, E. L., & Romano, J. P. (2005). *Testing statistical hypotheses*. New York: Springer.
- Levine, M.V., & Drasgow, F. (1988). Optimal appropriateness measurement. *Psychometrika*, 53(2), 161–176. doi:10.1007/bf02294130
- Levine, M. V., & Rubin, D. B. (1979). Measuring the appropriateness of multiple-choice test scores. *Journal of Educational and Behavioral Statistics*, 4(4), 269–290. doi: 10.3102/10769986004004269
- Li, X, Huang, C., & Harris, D. (2014, October). *Examining individual and cluster test irregularities in mixed-format testing*. Paper presented at the 2014 Conference on Test Security. Iowa City, Iowa. Abstract retrieved from <https://conferenceontestsecurity.org/cots-2014>

- Liu, X., Liu, F., & Simon, M. (2014, April). *Are the score gains suspicious?-A Bayesian growth analysis approach*. Paper presented at the annual meeting of the National Council on Measurement in Education, Philadelphia, PA. Abstract retrieved from <https://www.ncme.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=970c04c8-1fb3-bdea-daf1-12b99d8db4b8&forceDialog=0>
- Linacre, J. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, 16(2), 878. Retrieved from <https://www.rasch.org/rmt/rmt162f.htm>
- 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
- Marianti, S., Fox, J.-P., Avetisyan, M., Veldkamp, B. P., & Tijmstra, J. (2014). Testing for aberrant behavior in response time modeling. *Journal of Educational and Behavioral Statistics*, 39 (6), 426–451. doi:10.3102/1076998614559412
- Maynes, D. D. (2009, April). *Combining statistical evidence for increased power in detecting cheating*. Presented at the 2009 annual conference of the National Council on Measurement in Education, San Diego, CA. Abstract retrieved from <https://www.ncme.org/viewdocument/2009-san-diego-meeting-program>
- Maynes, D. D. (2005). *M₄: A new answer-copying index*. Unpublished manuscript, Caveon Test Security, Midvale, UT. Retrieved from <https://www.caveon.com/>
- Maynes, D. D. (2013). Educator cheating and the statistical detection of group-based test security threats. In J. A. Wollack & J. J. Fremer (Eds.), *Handbook of Test Security* (pp. 173-199). New York: Routledge.
- Maynes, D. D (2014a). Detection of non-independent test taking by similarity analysis. In Kingston, N. M. & Clark, A. K. (Eds.): *Test Fraud: Statistical Detection and Methodology*, (pp 52-80). New York: Routledge.
- Maynes, D.D. (2014b). A method for measuring performance inconsistency by using score differences. In N. M., Kingston & A. K., Clark, (Eds.), *Test Fraud: Statistical Detection and Methodology*, (pp 186-199). New York: Routledge.
- Maynes, D. D. (2017). Detecting potential collusion among individual examinees using similarity analysis. In G.J. Cizek & J.A. Wollack (Eds.), *Handbook of Quantitative Methods for Detecting Cheating on Tests*, (pp. 47-69). New York: Routledge.
- McLeod, L. D., & Lewis, C. (1999). Detecting item memorization in the CAT environment. *Applied Psychological Measurement*, 23(2), 147–160. Retrieved from <https://core.ac.uk/download/pdf/206658489.pdf>

- Meng, X.-B., Tao, J., & Chang, H.-H. (2015). A conditional modeling approach for locally dependent item responses and response times. *Journal of Educational Measurement*, 52(1), 1-27. doi: 10.1111/jedm.12060
- 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. (2002). Outlier detection in high-stakes certification testing. *Journal of Educational Measurement*, 39(3), 219-233. doi: 10.2307/1435079
- Meijer, R. R., & Sotaridona, L. S. (2006). *Detection of advance item knowledge using response times in computer adaptive testing*. Law School Admission Council. (LSAC Research Report 03-03). Retrieved from https://ris.utwente.nl/ws/portalfiles/portal/5129730/LSAC_CT-03-03.pdf
- Meijer, R. R., & Sijtsma, K. (1995). Detection of aberrant item score patterns: A review of recent developments. *Applied Measurement in Education*, 8(3), 261–272. doi: 10.1207/s15324818ame0803_5
- Meijer, R. R., & Sijtsma, K. (2001). Methodology review: Evaluating person fit. *Applied Psychological Measurement*, 25(2), 107-135. doi:10.1177/01466210122031957
- 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*. Law School Admission Council. (LSAC Research Report 14-03). Retrieved from <https://www.lsac.org/data-research/research/use-person-fit-scores-high-stakes-educational-testing-how-use-them-and-what>
- Meyer, J. P. (2010). A mixture Rasch model with item response time components. *Applied Psychological Measurement*, 34(7), 521-538. doi:10.1177/0146621609355451
- McLeod, L. D., Lewis, C., & Thissen, D. (2003). A Bayesian method for the detection of item preknowledge in computerized adaptive testing. *Applied Psychological Measurement*, 27(2), 121–137. doi:10.1177/0146621602250534
- McLeod, L. D., & Schnipke, D. L. (2006). *Detecting items that have been memorized*. Law School Admission Council. (LSAC Research Report 99-05). Retrieved from <https://lawcat.berkeley.edu/record/437725>
- Molenaar, I. W., & Hoijtink, H. (1990). The many null distributions of person fit indices. *Psychometrika*, 55(1), 75–106. doi:10.1007/BF02294745
- Mousavi, A., Tendeiro, J. N., & Younesi, J. (2016). Person fit assessment using the Perfit package in R. *The Quantitative Methods for Psychology*, 12(3), 232-242. doi: 10.20982/tqmp.12.3.p232

- Murphy, K. P. (2012). *Machine Learning: A probabilistic perspective*. London: The MIT Press.
- Obregon, P. (2013). *A Bayesian approach to detecting compromised items*. Paper presented at the 2013 annual meeting of the National Council on Measurement in Education, San Francisco, CA. Abstract retrieved from <https://www.ncme.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=26c6a941-2f63-5bd5-9438-95911fa2a6c4&forceDialog=0>
- O'Leary, L. S., & Smith, R. W. (2013). Detecting candidate preknowledge and compromised content using differential person and item functioning. In G. J. Cizek & J. A. Wollack (Eds.), *Handbook of quantitative methods for detecting cheating on tests*, (pp 200-216). New York: Routledge.
- Qian, H., Staniewska, D., Reckase, M., & Woo, A. (2016). Using response time to detect item preknowledge in computer based licensure examinations. *Educational Measurement: Issues and Practice*, 35, 38–47. doi:10.1111/emip.12102
- Quintano, C., Castellano, R., & Longobardi, S. (2009). A Fuzzy clustering approach to improve the accuracy of Italian student data. Procedure to correct the impact of the outliers on assessment test Scores. *Statistica and Applicazioni*, 7 (2), 149–17 Retrieved from http://www.invalsi.it/invalsi/ri/sis/documenti/022013/longobardi_paper.pdf
- Plummer, M., Best, N., Cowles, K., & Vines, K. (2006). CODA: Convergence diagnosis and output analysis for MCMC, *R News*, 6(1), 7-11. Retrieved from https://cran.r-project.org/doc/Rnews/Rnews_2006-1.pdf#page=7
- Plummer, M., Best, N., Cowles, K., Vines, K., Sarkar, D., Bates, D., Almond, R., & Magnusson, A. Package 'coda'. Retrieved from <https://cran.r-project.org/web/packages/coda/coda.pdf>
- Rao, C. R. (1973). *Linear statistical inference and its applications*. New York: Wiley.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Danish Paedagogiske Institute: Copenhagen, Denmark.
- Raton, M. (2014). OptimalCutpoints: An R package for selecting optimal cutpoints in diagnostic tests. *Journal of Statistical Software*, 61(8), 1-36. doi: 10.18637/jss.v061.i08
- 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>

- Robitzsch, A. (2020). Package 'sirt'. Retrieved from <https://cran.r-project.org/web/packages/sirt/sirt.pdf>
- Roskam, E. E. (1997). Models for speed and time-limit tests. In W. J. van der Linden & R. K. Hambleton (Eds.), *Handbook of Modern Item Response Theory*, (pp. 187-208). New York: Springer.
- Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, 14(3), 271-282. doi:10.1177/014662169001400305
- Rouder, J. N., Sun, D., Speckman, P. L., Lu, J., & Zhou, D. (2003). A hierarchical Bayesian statistical framework for response time distributions. *Psychometrika*, 68(4), 589-606.. doi:10.1007/bf02295614
- Rubin, D. B. (1984). Comment: Assessing the fit of logistic regressions using the implied discriminant analysis. *Journal of the American Statistical Association*, 79(385), 79–80. doi:10.1080/01621459.1984.10477066
- Rupp, A. A. (2013). A systematic review of the methodology for person fit research in item response theory: Lessons about generalizability of inferences from the design of simulation studies. *Psychological Test and Assessment Modeling*, 55(1), 3–38. Retrieved from <https://psycnet.apa.org/record/2013-18917-001>
- Santos, M. A., Moala, F. A., & Tachibana, V. M. (2009). Approximate Bayesian methods for logistic regression model. *Revista Brasileira de Biometria*, 27(2), 288-300. Retrieved from http://jaguar.fcav.unesp.br/RME/fasciculos/v27/v27_n2/Maola.pdf
- Sato, T. (1975). *The construction and interpretation of S-P tables*. Tokyo: Meiji Tosho.
- Schnipke, D. L., & Scrams, D. J. (1997). Modeling item response times with a two-state mixture model: A new method of measuring speededness. *Journal of Educational Measurement*, 34(3), 213-232. doi: 10.2307/1435443
- 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).
- Shu, Z., Henson, R., & Luecht, R. (2013). Using deterministic, gated item response theory model to detect test cheating due to item compromise. *Psychometrika*, 78(3), 481-497. doi:10.1007/s11336-012-9311-3
- Skorupski, W., & Egan, K. (2012, May). *A hierarchical linear modeling approach for detecting cheating and aberrance*. Paper presented at the 2012 Conference on the Statistical Detection of Potential Test Fraud. Lawrence, KS. Abstract retrieved from https://img1.wsimg.com/blobby/go/71ff6051-7987-4285-89a3-874c4bfcc71c/downloads/1d2lkha9q_895129.pdf?ver=1602022349018

- Skorupski, W., & Egan, K. (2014). A Bayesian hierarchical linear modelling approach for detecting cheating and aberrance. . In N. M., Kingston & A. K., Clark, (Eds.) *Test Fraud: Statistical Detection and Methodology*, (pp 121-133). New York: Routledge.
- Skorupski, W., Fitzpatrick, J., & Egan, K. (2017). A Bayesian hierarchical model for detecting aberrant growth at group level. In G., Cizek & J. Wollack, (Eds.) *Handbook of Quantitative Methods for Detecting Cheating on Tests*, (pp 232-244). New York: Routledge.
- Simon, M. (2014). Local outlier detection in data forensics: data mining approach to flag unusual schools. In N. M., Kingston & A. K. , Clark, (Eds.) *Test Fraud: Statistical Detection and Methodology*, (pp 83-100). New York: Routledge.
- Smith, R. M. (1985). A comparison of rasch person analysis and robust estimators. *Educational and Psychological Measurement*, 45(3), 433–444. doi:10.1177/001316448504500301
- Smith, R. W., & Davis-Becker, S. (2011, April). *Detecting suspect candidates: An application of differential person functioning analysis*. Paper presented at the 2011 annual meeting of the National Council on Measurement in Education, New Orleans, LA. Abstract retrieved from <https://www.ncme.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=67a550a4-3f42-3ceb-cd65-b4275609c893&forceDialog=0>
- Snijders, T. A. B. (2001). Asymptotic null distribution of person fit statistics with estimated person parameter. *Psychometrika*, 66(3), 331-342. doi:10.1007/bf02294437
- Singmann, H. (2020). Complete environment for bayesian inference (LaplaceDemon) [Computer software manual]. Retrieved from <https://cran.r-project.org/web/packages/LaplacesDemon/LaplacesDemon.pdf>
- Sinharay, S. (2015a). Assessing person fit using l_z^* and the posterior predictive model checking method for dichotomous item response theory models. *International Journal of Quantitative Research in Education*, 2(3/4), 265-282. doi:10.1504/ijqre.2015.071730
- Sinharay, S. (2015b). Assessment of person fit for mixed-format tests. *Journal of Educational and Behavioral Statistics*, 40(4), 343-365. doi:10.3102/1076998615589128
- Sinharay, S. (2017a). Detection of item preknowledge using likelihood ratio test and score test. *Journal of Educational and Behavioral Statistics*, 42, 46–68. doi:10.3102/1076998616673872
- Sinharay, S. (2017b). Which statistic should be used to detect item preknowledge when the set of compromised items is known? *Applied Psychological Measurement*, 41(6), 403–421. doi: 10.1177/0146621617698453

- Sinharay, S. (2018). A new person-fit statistic for the lognormal model for response times. *Journal of Educational Measurement*, 55(4), 457–476. doi:10.1111/jedm.12188
- Sinharay, S. (2020). Detection of item preknowledge using response times. *Applied Psychological Measurement*, 44(5), 1-17. doi: 10.1177/0146621620909893
- Sijtsma, K. (1986). A coefficient of deviance of response patterns. *Kwantitatieve Methoden*, 22(7), 131–145. Retrieved from <https://research.tilburguniversity.edu/en/publications/a-coefficient-of-deviance-of-response-patterns>
- Sijtsma, K., & Meijer, R. R. (1992). A method for investigating the intersection of item response functions in Mokken's non-parametric IRT model. *Applied Psychological Measurement*, 16(2), 149–157. doi:10.1177/014662169201600204
- 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(2), 115–132. doi: 10.2307/1435251
- Sotaridona, L.S., & Meijer, R.R. (2003). Two new statistics to detect answer copying. *Journal of Educational Measurement*, 40(1), 53–69. doi: 10.2307/1435054
- Sotaridona, L. S., Wibowo, A., & Hendrawan, I. (2014). A parametric approach to detect a disproportionate number of identical item responses on a test. In N. M., Kingston, & A. K., Clark (Eds.) *Test Fraud: Statistical Detection and Methodology*, (pp 38-52). New York: Routledge
- Sotaridona, L. S., van der Linden, W.J., & Meijer, R.R. (2006). Detecting answer copying using the kappa statistic. *Applied Psychological Measurement*, 30(5), 412-431. doi: 10.1177/0146621606288891
- Sünbül, O., & Yormaz, S. (2018). 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
- Tatsuoka, K. (1984). Caution indices based on item response theory. *Psychometrika*, 49 (1), 95–110. doi:10.1007/bf02294208
- Tatsuoka, K. K., & Tatsuoka, M. M. (1980). A model for incorporating response-time data in the scoring of achievement tests. In D. J. Weiss (Eds.), *Proceedings of the 1979 Computerized Adaptive Testing Conference* (pp. 236-256). Minnesota, MN: University of Minnesota, Department of Psychology, Psychometric Methods Program.
- Tatsuoka, K. K., & Tatsuoka, M. M. (1983). Spotting erroneous rules of operation by the individual consistency index. *Journal of Educational Measurement*, 20(3), 221-230. doi:10.1111/j.1745-3984.1983.tb00201.x

- Thiessen, B. (2007). *Case study—policies to address educator cheating*. Cheating Policies. Retrieved from <http://www.bradthiessen.com/html5/docs/format.pdf>
- Thiessen, B. (2008). *Relationship between test security policies and test score manipulations*. (Unpublished doctoral thesis), University of Iowa, Iowa. Available from ProQuest Dissertations and Theses database. (UMI No. 3347249).
- Thissen, D. (1983). Timed testing: An approach using item response theory. In D. J. Weiss (Eds.), *New horizons in testing: Latent trait test theory and computerized adaptive testing*, (pp. 179-203). New York: Academic Press.
- Thomas, N. (2020). Package ‘R2OpenBUGS’. Retrieved from <https://cran.r-project.org/web/packages/R2OpenBUGS/R2OpenBUGS.pdf>
- Toton, S. L., & Maynes, D. D. (2019). Detecting examinees with pre-knowledge in experimental data using Conditional Scaling of Response Times. *Frontiers in Education*, 4(49), 1-18. doi: 10.3389/educ.2019.00049
- Trabin, T. E., & Weiss, D. J. (1983). The person response curve: Fit of individuals to item response theory models. In D. J. Weiss (Eds.), *New horizons in testing*, (pp. 83–108). New York, NY: Academic Press.
- 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>
- 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(2), 99-117. doi: 10.1111/j.1745-3984.2007.00055.x
- van der Flier, H. (1977). Environmental factors and deviant response patterns. In Y.P. Poortinga (Eds.), *Basic problems in cross-cultural psychology*, (pp. 30-35). Amsterdam: Swets and Zeitlinger.
- van der Flier, H. (1982). Deviant response patterns and comparability of test scores. *Journal of Cross-Cultural Psychology*, 13(1), 267–298. doi:10.1177/0022002182013003001
- van der Linden, W. J., Scrams, D. J., & Schnipke, D. L. (1999). Using response-time constraints to control for speededness in computerized adaptive testing. *Applied Psychological Measurement*, 23(3), 195-210. doi:10.1177/01466219922031329
- 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. (2009). Conceptual issues in response-time modeling. *Journal of Educational Measurement*, 46(3), 247–272. doi:10.1111/j.1745-3984.2009.00080.x
- van der Linden, W. J. (2010). Linking response-time parameters onto a common scale. *Journal of Educational Measurement*, 47(1), 92–114. doi:10.1111/j.1745-3984.2009.00101.x
- van der Linden, W. J. (2011). Modeling response times with latent variables: Principles and applications. *Psychological Test and Assessment Modeling*, 53(3), 334–358. Retrieved from https://ris.utwente.nl/ws/portalfiles/portal/15037134/05_vanderLinden.pdf
- van der Linden, W. J., & Guo, F. (2008). Bayesian procedures for identifying aberrant response-time patterns in adaptive testing. *Psychometrika*, 73, 365–384. doi:10.1007/s11336-007-9046-8.
- van der Linden, W. J., & Jeon, M. (2012). Modeling answer changes on test items. *Journal of Educational and Behavioral Statistics*, 37(1), 180–199. doi:10.3102/1076998610396899
- van der Linden, W. J., & Sotaridona, L.S. (2004). A statistical test for detecting answer copying on multiple-choice tests. *Journal of Educational Measurement*, 41(4), 361–378. doi: 10.2307/20461768
- van der Linden, W. J., & Lewis, C. (2015). Bayesian checks on cheating on tests. *Psychometrika*, 80, 689–706. doi:10.1007/s11336-014-9409-x
- van der Linden, W. J., & Sotaridona, L. (2006). Detecting answer copying when the regular response process follows a known response model. *Journal of Educational and Behavioral Statistics*, 31(3), 283–304. doi:10.3102/10769986031003283
- 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. doi:10.1007/bf02294800
- van der Linden, W. J., & Glas, C. A. W. (2010). Statistical tests of conditional independence between responses and/or response times on test items. *Psychometrika*, 75(1), 120–139. doi:10.1007/s11336-009-9129-9
- van der Linden, W. J., Klein Entink, R. H., & Fox, J.- P. (2010). IRT parameter estimation with response times as collateral information. *Applied Psychological Measurement*, 34(5), 327–347. doi: 10.1177/0146621609349800
- van der Linden, W. J., Scrams, D. J., & Schnipke, D. L. (1999). Using response-time constraints to control for speededness in computerized adaptive testing. *Applied Psychological Measurement*, 23(3), 195–210. doi:10.1177/01466219922031329

- van Krimpen-Stoop, E. M. L. A & Meijer, R. R. (2000). Detecting person misfit in, adaptive testing using statistical process control techniques. In W. J. van der Linden & C. A. W. Glas (Eds.), *Computerized adaptive testing: Theory and practice*, (pp. 201–219). Kluwer Academic Publishers: Dordrecht, the Netherlands.
- van 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. doi: 10.3102/10769986026002199
- van Krimpen-Stoop, E. M. L. A. & Meijer, R. R. (2002). Detection of person misfit in computerized adaptive tests with polytomous Items. *Applied Psychological Measurement*, 26(2), 164-180. doi:10.1177/01421602026002004
- Verhelst, N. D., Verstralen, H. H. F. M., & Jansen, M. G. H. (1997). A logistic model for timelimit tests. In W. J. van der Linden & R. K. Hambleton (Eds.) *Handbook of Modern Item Response Theory*, (pp. 169-185). New York: Springer.
- von Davier, M., & Rost, J. (2007). Mixture distribution item response models. In C. R. Rao & S. Sinharay (Eds.) *Handbook of Statistics*, (pp. 643-661). The Netherlands: Elsevier B. V.
- von Davier, M., & Yamamoto, K. (2006). Mixture-Distribution and HYBRID Rasch models. In M. von Davier & C. H. Carstensen (Eds.) *Multivariate and Mixture Distribution Rasch Models: Extensions and Applications*, (pp. 99-115). New York: Springer.
- 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
- Wang, C., Xu, G., & Shang, Z. (2018). A two-stage approach to differentiating normal and aberrant behavior in computer based testing. *Psychometrika*, 83(1), 223-254. doi:10.1007/s11336-016-9525-R
- Wang, C., Xu, G., Shang, Z., & Kuncel, N. (2018). Detecting aberrant behavior and item preknowledge: A comparison of mixture modeling method and residual method. *Journal of Educational and Behavioral Statistics*, doi: 1076998618767123
- Weiss, D. J. (1973). *The stratified adaptive computerized ability test*. (Research Report 73-3). Minneapolis MN: University of Minnesota, Department of Psychology.
- Wesolowsky, G. (2000a). Detecting excessive similarity in answers on multiple choice exams. *Journal of Applied Statistics*, 27(7), 909-921. doi:10.1080/02664760050120588
- Wesolowsky, G. (2000b). *Statistical detection of cheating (copying, collusion) on multiple choice tests and examinations*. Retrieved from <http://wesolows.wired4wisdom.com/>

- Wibowo, A., Sotaridona, L., & Hendrawan, I. (2013, April). *Statistical models for flagging unusual number of wrong-to-right erasures*. Paper presented at the 2013 Annual Meeting of the National Council on Measurement in Education. San Francisco, CA. Abstract retrieved from <https://www.ncme.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=26c6a941-2f63-5bd5-9438-95911fa2a6c4&forceDialog=0>
- 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
- Wise, S. L., Ma, L., Kingsbury, G. G., & Hauser, C. (2010, May). *An investigation of the relationship between time of testing and Test-Taking Effort*. Paper presented at the 2010 annual meeting of the National Council on Measurement in Education, Denver, CO. Abstract 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. doi:10.1207/s15324818ame1802_2
- Wollack, J. A. (1997). A nominal response model approach for detecting answer copying. *Applied Psychological Measurement*, 21(4), 307-320. doi:10.1177/01466216970214002
- Wollack, J. A., Cohen, A. S., & Eckerly, C. A. (2015). Detecting test tampering using item response theory. *Educational and Psychological Measurement*, 75(6), 931-953. doi: 10.1177/0013164414568716
- Wollack, J. A. & Maynes, D. (2011). *Detection of test collusion using item response data*. Paper presented at the 2011 Annual Meeting of the National Council on Measurement in Education. New Orleans, LA. Abstract retrieved from <https://www.ncme.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=67a550a4-3f42-3ceb-cd65-b4275609c893&forceDialog=0>
- Wollack, J. A., & Maynes, D. D. (2017). Detection of test collusion using cluster analysis. In G. J. Cizek & J. A. Wollack (Eds.), *Handbook of quantitative methods for detecting cheating on tests*, (pp. 124-150). New York, NY: Routledge.
- Wright, B., & Masters, G. (1982). *Rating scale analysis: Rasch measurement*. Chicago: MESA Press.
- Wright, B. D., & Panchapakesan, N. (1969). A procedure for sample-free item analysis. *Educational and Psychological Measurement*, 29(1), 23-48. doi:10.1177/001316446902900102
- Wright, B. & Stone, M. (1979). *Best test design: Rasch measurement*. Chicago: MESA Press

- Yamamoto, K. Y. (1989). *HYBRID model of IRT and latent class models* (ETS Research Report No. RR-89-41). Princeton, NJ: Educational Testing Service.
- Yormaz, S. (2014). *Kopya Belirleme İndekslerinin Çeşitli Koşullar Altında I. Tip Hata Oranlarının ve Güçlerinin Belirlenmesi*. (Yayımlanmamış Yüksek Lisans Tezi). Mersin Üniversitesi, Mersin, Türkiye.
- Yormaz, S. (2019). *Test Güvenliği Açısından Bireyler Arasındaki Olası İş Birliğinin İncelenmesi*. (Yayımlanmamış Doktora Tezi). Mersin Üniversitesi, Mersin, Türkiye.
- Yormaz, S. & Sünbül, O. (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
- Zhang, Y., Searcy, C. A., & Horn, L. (2011, April). *Mapping clusters of aberrant patterns in item responses*. Paper presented at the 2011 annual meeting of the National Council on Measurement in Education, New Orleans, LA. Abstract retrieved from <https://www.ncme.org/HigherLogic/System/DownloadDocumentFile.ashx?DocumentFileKey=67a550a4-3f42-3ceb-cd65-b4275609c893&forceDialog=0>
- Zhan, P., Jiao, H., Wang, W.-C., & Man, K. (2018). *A multidimensional hierarchical framework for modeling speed and ability in computer-based multidimensional tests*. arXiv preprint arXiv:1807.04003. Retrieved from <https://arxiv.org/abs/1807.04003>
- Zickar, M. J., & Drasgow, F. (1996). Detecting faking on a personality instrument using appropriateness measurement. *Applied Psychological Measurement*, 20(1), 71-87. doi:10.1177/014662169602000107
- Zopluoğlu, C. (2016). Classification performance of answer-copying indices under different types of IRT models. *Applied Psychological Measurement*, 40(8), 592–607. doi:10.1177/0146621616664724
- Zopluoğlu, C. (2017). Similarity, answer copying, and aberrance – understanding the status quo. In G. Cizek & Wollack, J. (Eds.) *Handbook of Quantitative Methods for Detecting Cheating on Tests*, (pp 25-46). New York: Routledge.
- Zopluoğlu, 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. doi: 10.1177/0013164412442941



EK 1. Etik Kurul Onayı

ANKARA ÜNİVERSİTESİ SOSYAL BİLİMLER ALT ETİK KURULU KARAR ÖRNEĞİ

Karar Tarihi : 15/01/2021

Toplantı Sayısı : 01

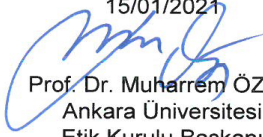
Karar Sayısı : 10

10-Eğitim Bilimleri Enstitüsü doktora öğrencisi Ebru Balta'nın "Madde Ön Bilgisine Dayalı Örüntü Benzerliğinin Farklı Modeller Kullanılarak Grup Düzeyinde İncelenmesi" başlıklı tezi ile ilgili "İnsan Üzerinde Yapılan Klinik Dışı Araştırmalar Başvuru Formu" Etik Kurulumuzca incelendi.

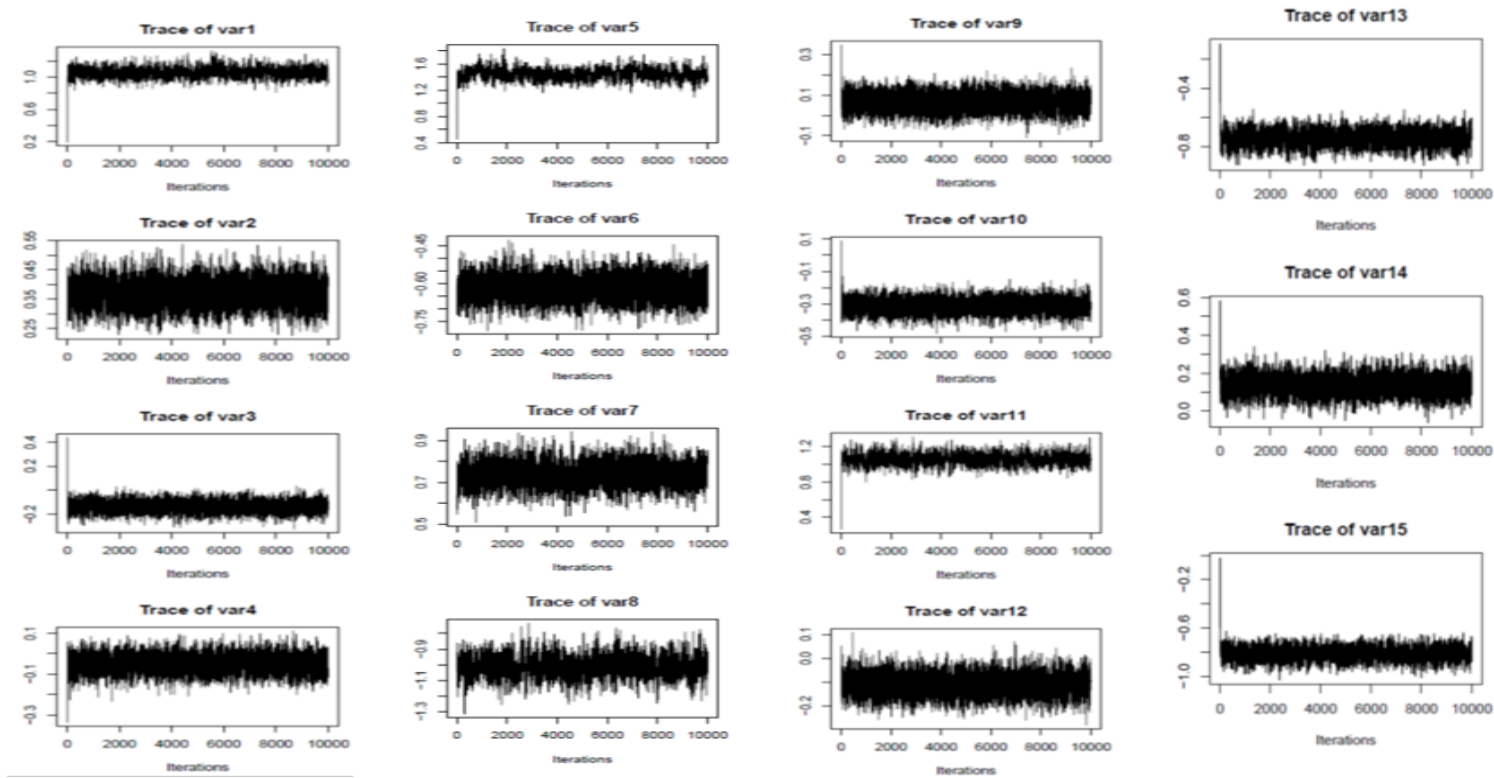
Eğitim Bilimleri Enstitüsü doktora öğrencisi **Ebru Balta'nın "Madde Ön Bilgisine Dayalı Örüntü Benzerliğinin Farklı Modeller Kullanılarak Grup Düzeyinde İncelenmesi" başlıklı tezinin araştırma protokolüne uyulması ve etik onay tarihinden itibaren geçerli olması koşuluyla uygulanmasının etik açıdan uygun olduğuna oybirliği ile karar verildi.**

ASLININ AYNIDIR

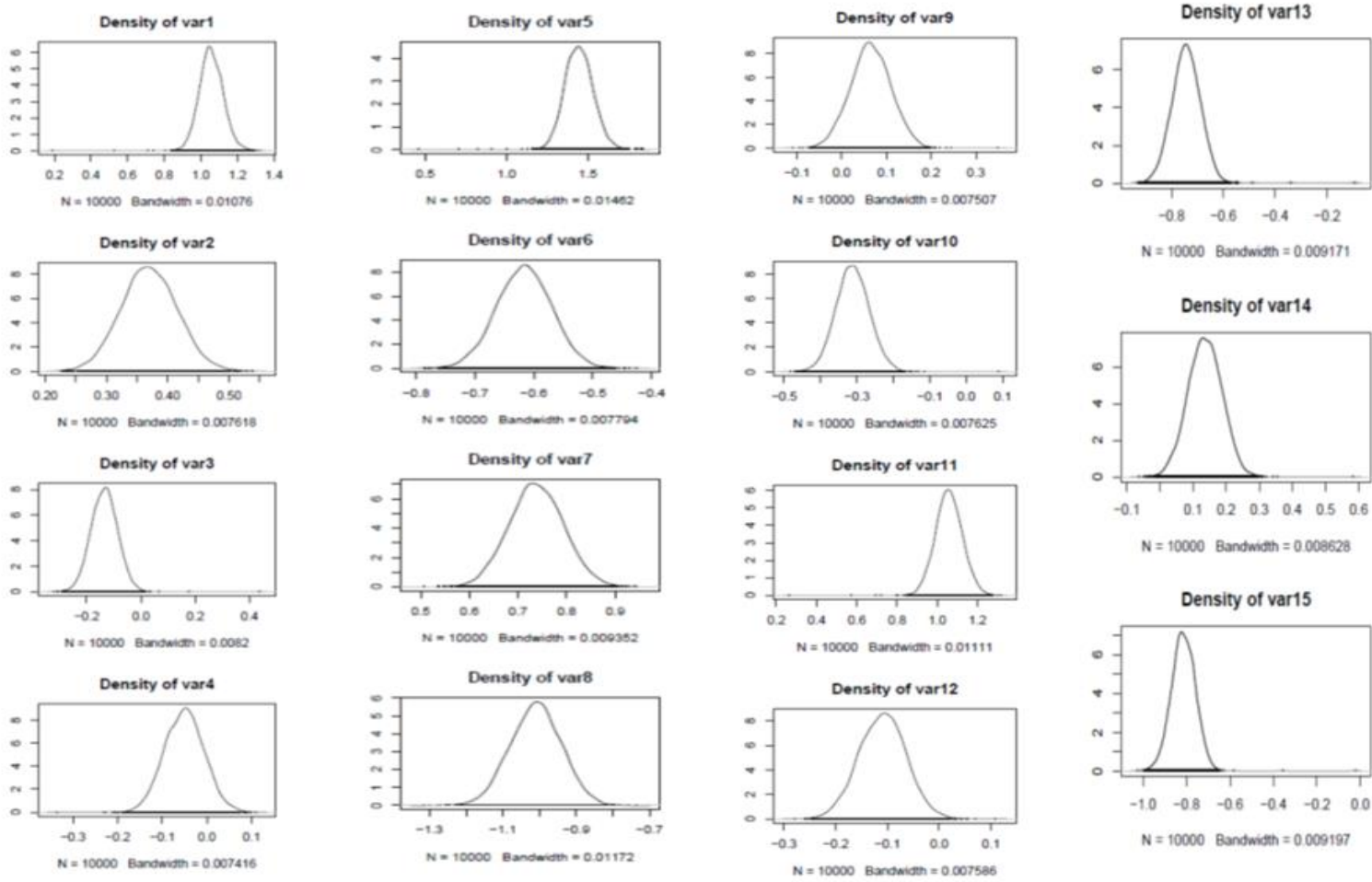
15/01/2021


Prof. Dr. Muharrem ÖZEN
Ankara Üniversitesi
Etik Kurulu Başkanı

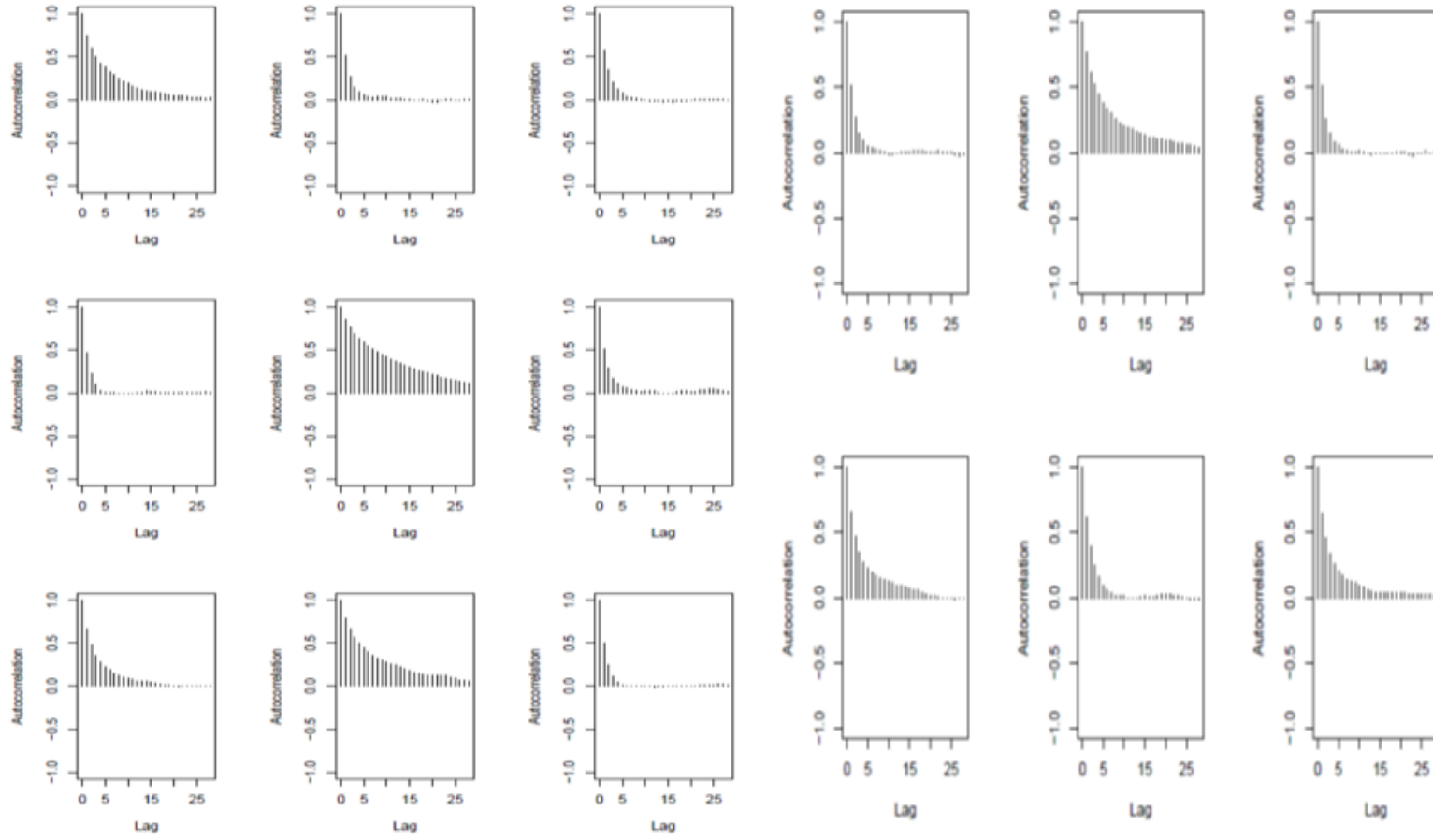
EK 2. Test Maddeleri İçin 10000 MCMC Yinelemesine İlişkin Madde Güçlük Parametresi İçin İz, Sonsal Olasılık Yoğunluk ve Otokorelasyon Grafikleri



Şekil XVIII. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Madde Güçlük Parametresi İçin İz Grafikleri

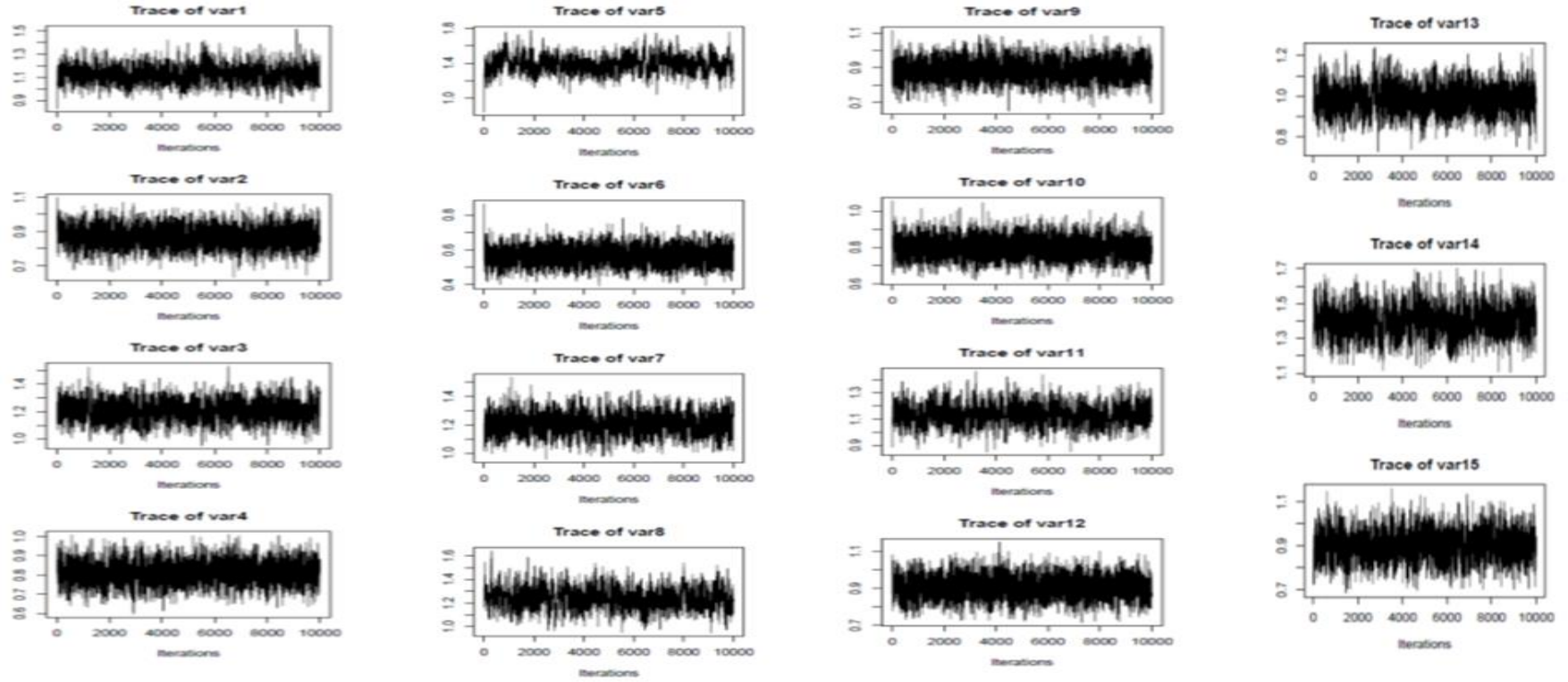


Şekil XIX. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Madde Güçlük Parametresi İçin Olasılık Yoğunluk Grafikleri

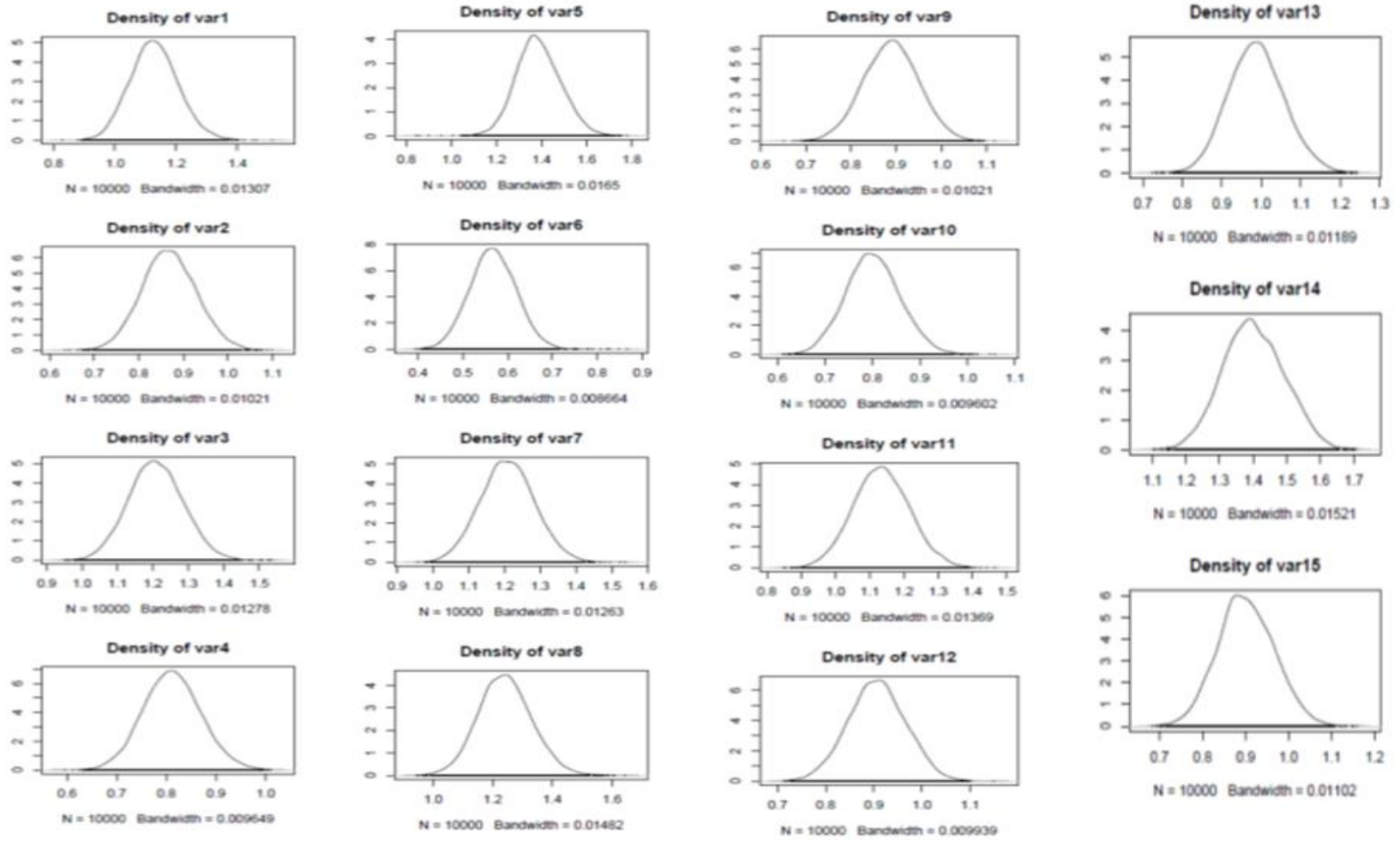


Şekil XX. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Madde Güçlük Parametresi İçin Otokorelasyon Grafikleri

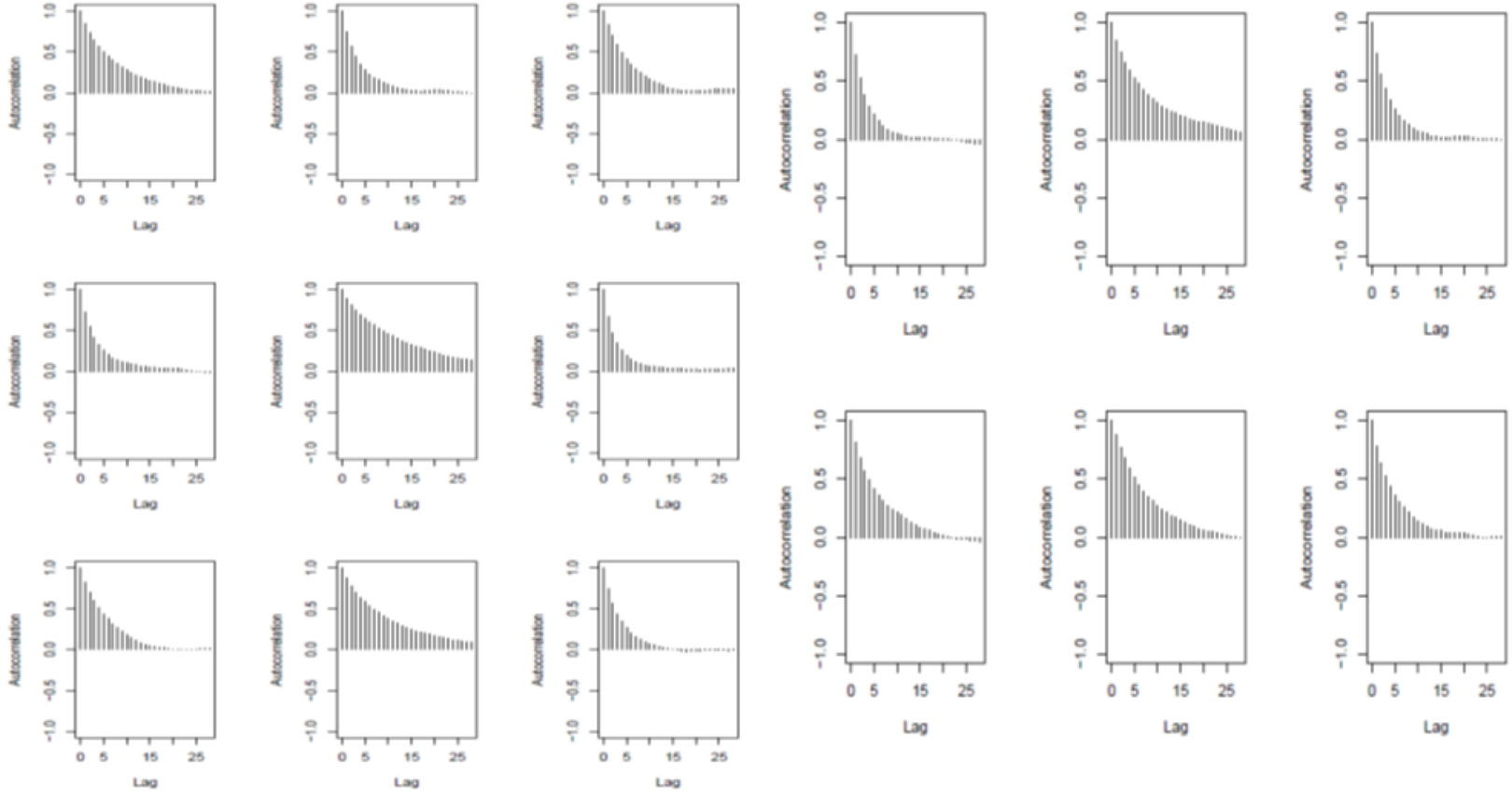
EK 3. Test Maddeleri İçin 10000 MCMC Yinelemesine İlişkin Madde Ayırtedicilik Parametresi İçin İz, Sonsal Olasılık Yoğunluk ve Otokorelasyon Grafikleri



Şekil XXI. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Madde Ayırtedicilik Parametresi İçin İz Grafikleri

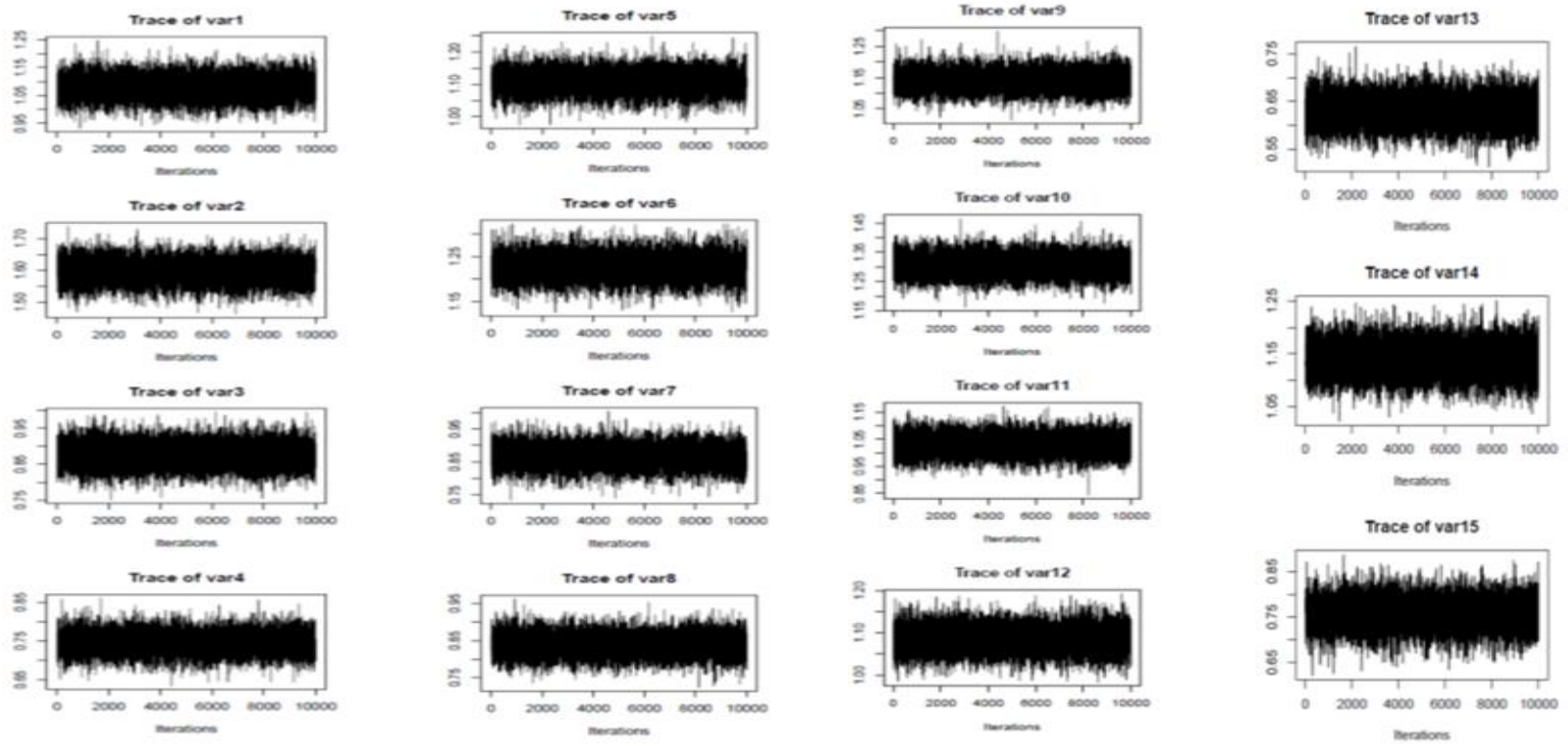


Şekil XXII. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Madde Ayırtedicilik Parametresi İçin Olasılık Yoğunluk Grafikleri

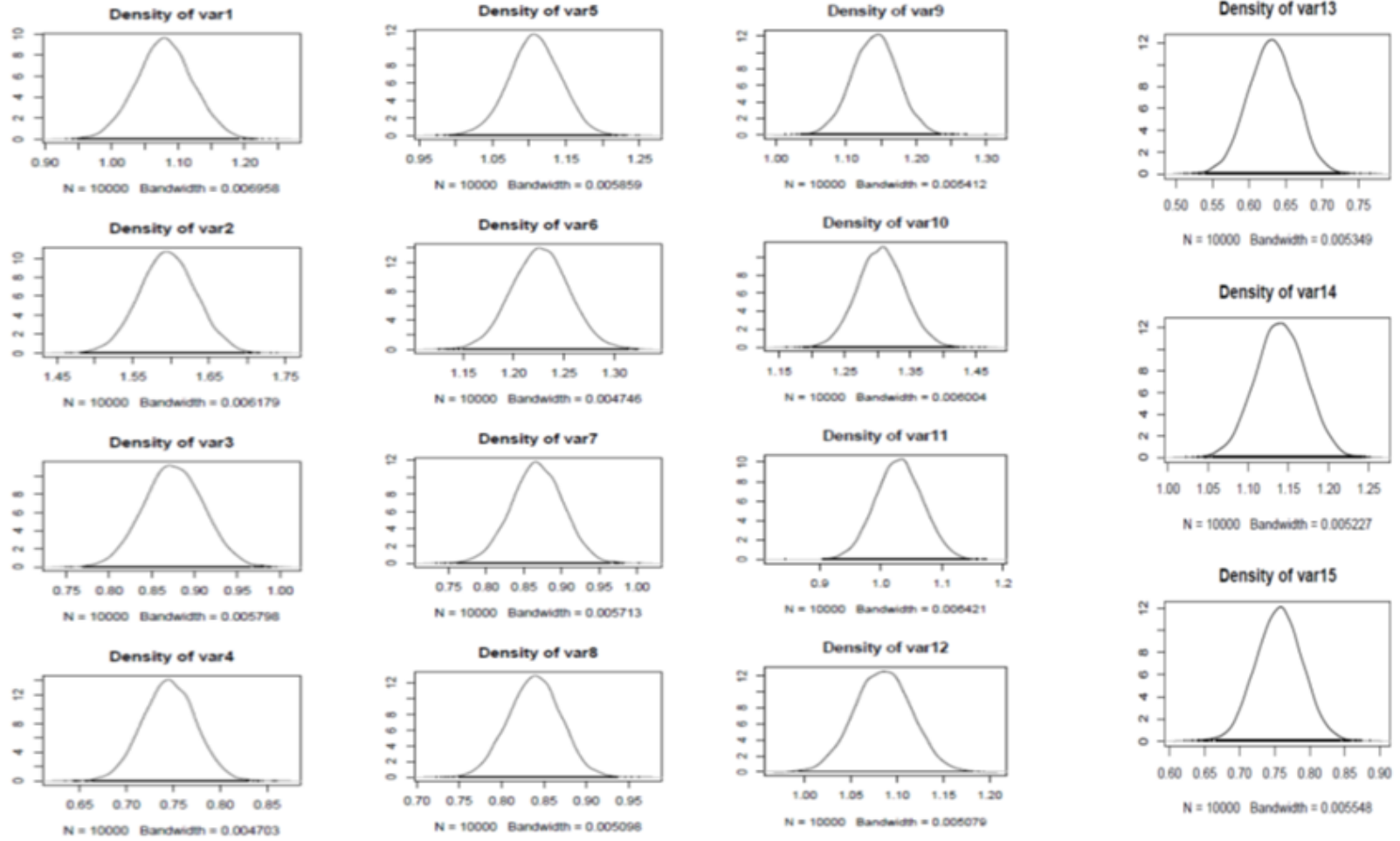


Şekil XXIII. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Madde Ayırtedicilik Parametresi İçin Otokorelasyon Grafikleri

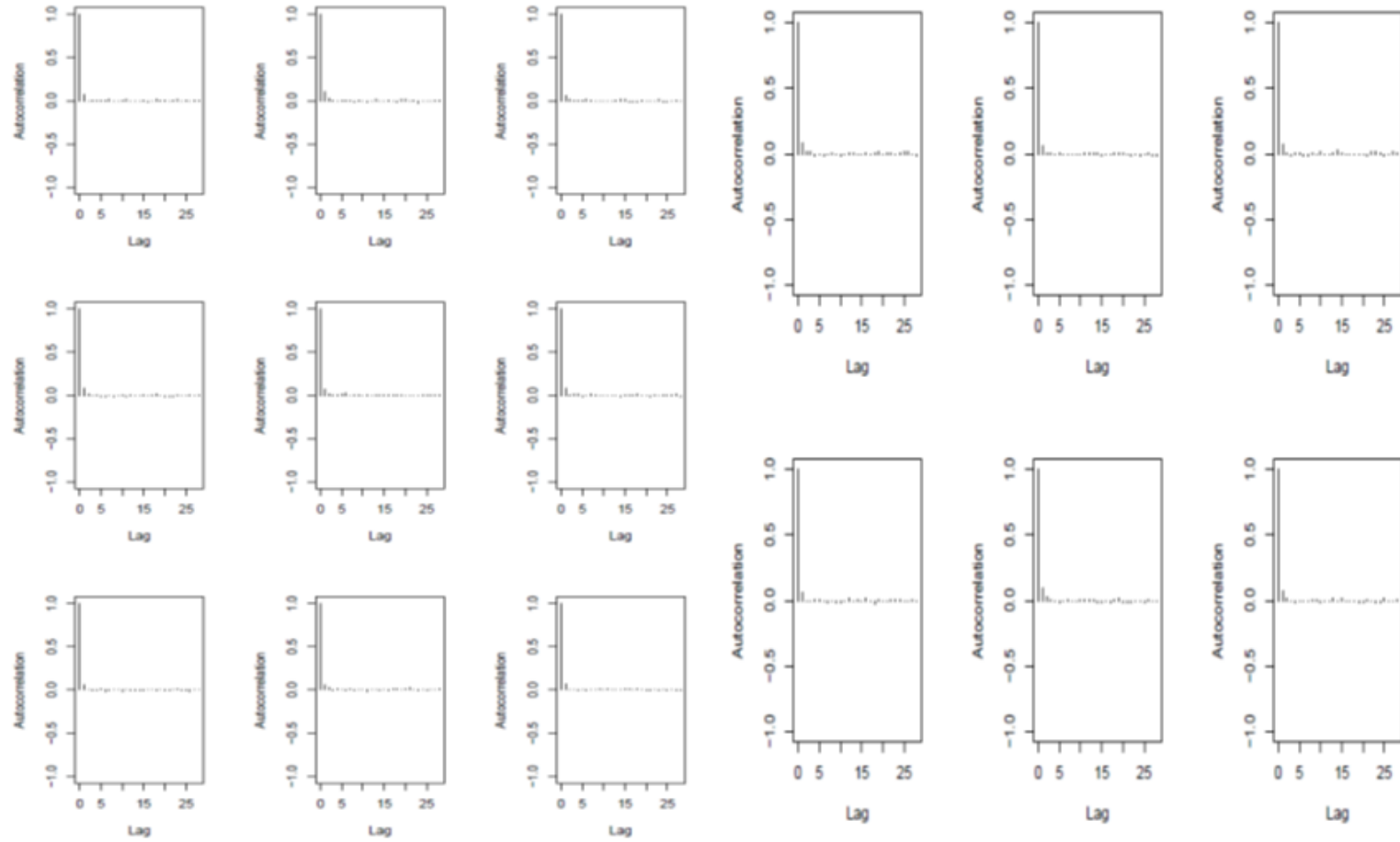
EK 4. Test Maddeleri İçin 10000 MCMC Yinelemesine İlişkin Zaman Ayırtedicilik Parametresi İçin İz, Sonsal Olasılık Yoğunluk ve Otokorelasyon Grafikleri



Şekil XXIV. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Zaman Ayırtedicilik Parametresi İçin İz Grafikleri

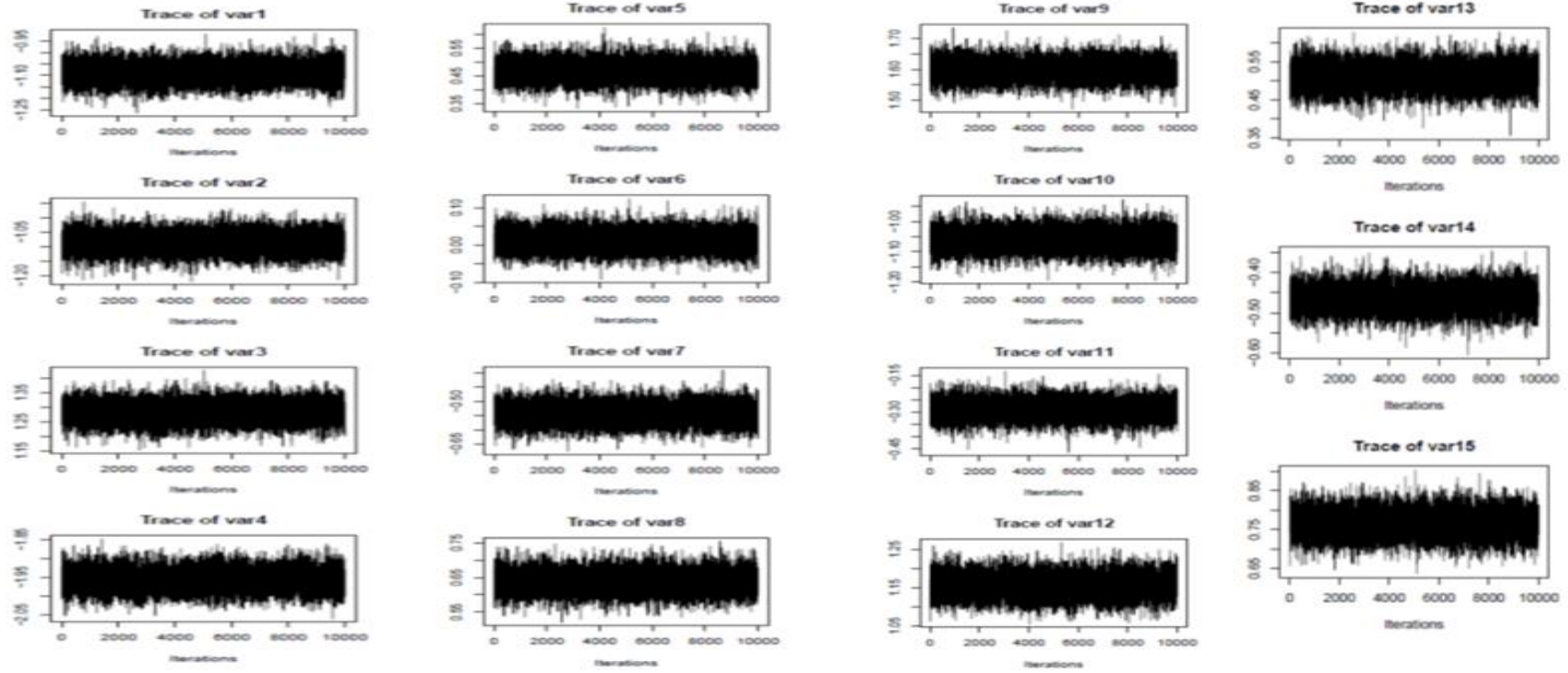


Şekil XXV. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Zaman Ayırtediclik Parametresi İçin Olasılık Yoğunluk Grafikleri

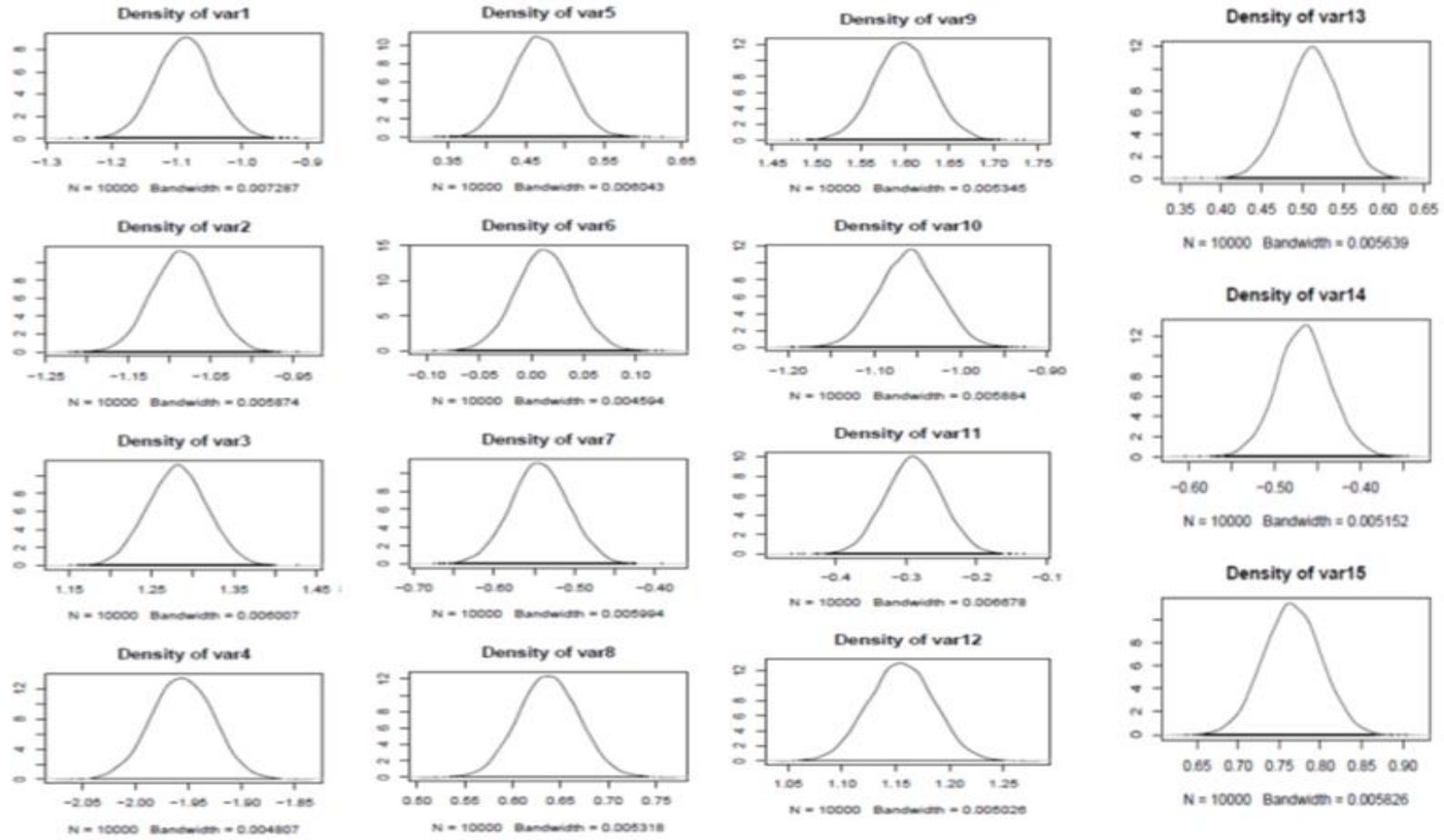


Şekil XXVI. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Zaman Ayırteçicilik Parametresi İçin Otokorelasyon Grafikleri

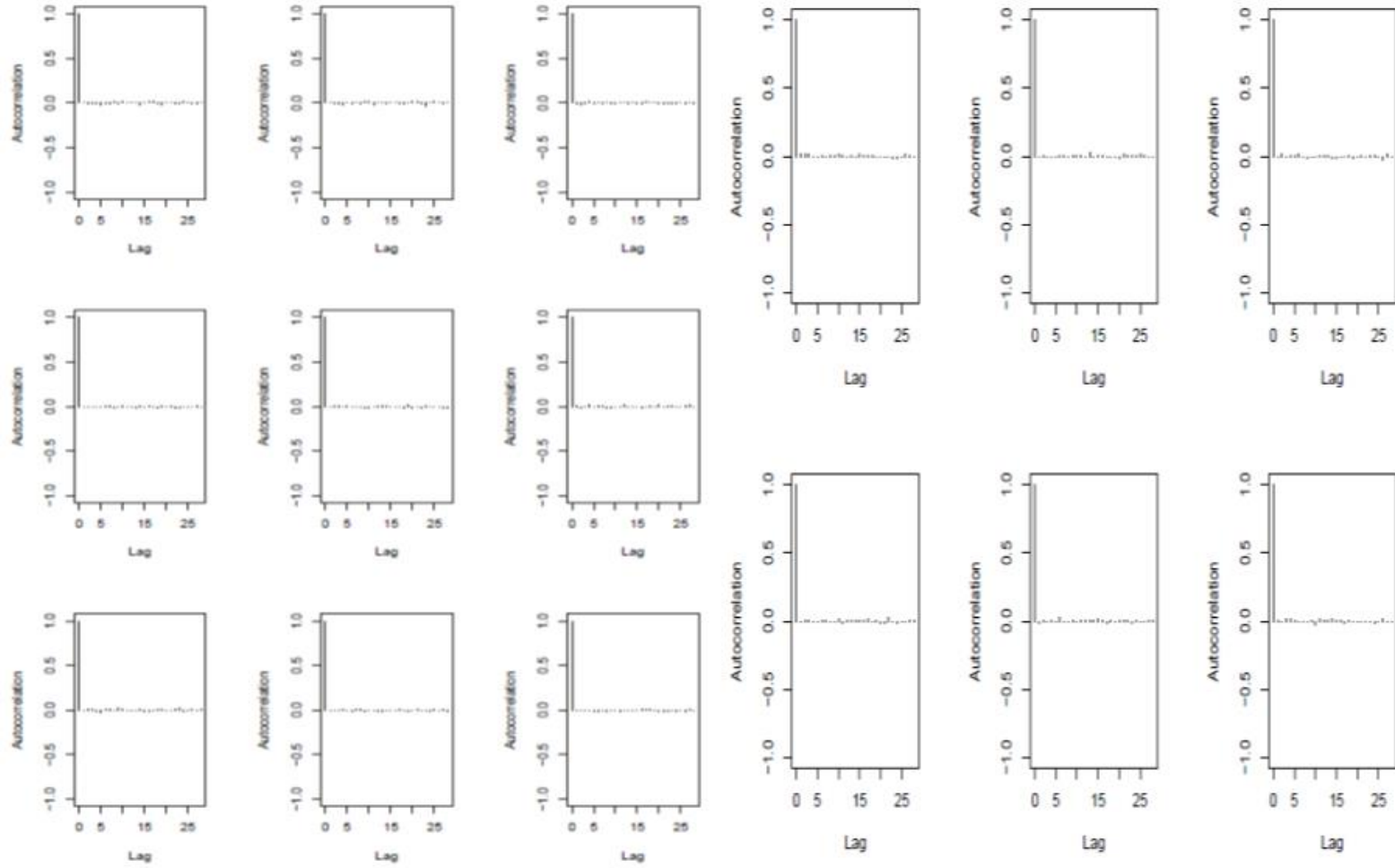
EK 5. Test Maddeleri İçin 10000 MCMC Yinelemesine İlişkin Zaman Yoğunluk Parametresi İçin İz, Sonsal Olasılık Yoğunluk ve Otokorelasyon Grafikleri



Şekil XXVII. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Zaman Yoğunluk Parametresi İçin İz Grafikleri



Şekil XXVIII. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Zaman Yoğunluk Parametresi İçin Olasılık Yoğunluk Grafikleri



Şekil XXIX. 15 Test Maddesi İçin 10000 MCMC Yinelemesine İlişkin Zaman Yoğunluk Parametresi İçin Otokorelasyon Grafikleri

EK 6. Geweke Yakınsaklık Test İstatistik Değerleri

Tablo 4

HMTK-TS Model Parametrelerine İlişkin Geweke Yakınsaklık Test İstatistik Değerleri

Madde Numarası	Madde Güçlük Parametresi	Madde Ayırtedicilik Parametresi	Zaman Ayırtedicilik Parametresi	Zaman Yoğunluk Parametresi
1	-0.302	-0.203	-0.843	0.002
2	1.037	1.852	1.018	-0.518
3	0.542	0.611	1.378	0.352
4	-0.479	0.475	1.431	1.021
5	0.213	0.002	-0.079	-0.082
6	0.604	-1.684	-0.097	0.762
7	-2.699	-1.158	1.906	-0.331
8	-1.360	1.817	1.217	-0.206
9	0.979	1.004	0.062	-0.159
10	0.979	0.676	0.578	0.819
11	-0.221	-0.347	-1.737	-0.317
12	0.755	-0.322	-1.772	0.264
13	1.017	0.432	-1.059	-0.198
14	-0.643	-0.027	-0.125	-0.342
15	1.281	-0.645	-1.162	-0.496

**EK 7. Koşullara Göre Uygun Olmayan Madde Tepkilerini Belirlemede
Kullanılan Yöntemlerin I. Tip Hata Oranları**

Tablo 5

Koşullara Göre Madde Tepkileri İçin Kullback-Leibler Uzaklığı ve L Birey Uyum İstatistiğinin I. Tip Hata Oranları Ortalaması

Yöntemler	Test Uzunluğu	Teşhir Olan Maddelerin Güçlük Düzeyi	I. Tip Hata Oranı
Madde tepkileri için KL uzaklığı	15 madde	Zor	0.378
		Orta Güçlük Düzeyi	0.364
	50 madde	Zor	0.387
		Orta Güçlük Düzeyi	0.366
$L(l_s^y)$ birey uyum istatistiği	15 madde	Zor	0.008
		Orta Güçlük Düzeyi	0.006
	50 madde	Zor	0.002
		Orta Güçlük Düzeyi	0.004

**EK 8. Koşullara Göre Uygun Olmayan Madde Tepki Sürelerini Belirlemede
Kullanılan Yöntemlerin I. Tip Hata Oranları**

Tablo 6

Koşullara Göre Madde Tepki Süreleri İçin Kullback-Leibler Uzaklığı ve L Birey Uyum İstatistiğinin I. Tip Hata Oranları Ortalaması

Yöntemler	Test Uzunluğu	Teşhir Olan Maddelerin Güçlük Düzeyi	I. Tip Hata Oranı
Madde tepki süreleri için KL uzaklığı	15 madde	Zor	0.160
		Orta Güçlük Düzeyi	0.167
	50 madde	Zor	0.157
		Orta Güçlük Düzeyi	0.161
$L(l^t)$ birey uyum istatistiği	15 madde	Zor	0.024
		Orta Güçlük Düzeyi	0.026
	50 madde	Zor	0.223
		Orta Güçlük Düzeyi	0.029

EK 9. Koşullara Göre Uygun Olmayan Madde Tepkilerini Belirlemede
Kullanılan Yöntemlerin Test Hilesi Belirleme Güç Oranları

Tablo 7

Koşullara Göre Madde Tepkileri İçin Kullback-Leibler Uzaklığı ve L Birey Uyum İstatistiğinin Test Hilesi Belirleme Güç Oranları Ortalaması

Yöntemler	Test Uzunluğu	Teşhir Olan Maddelerin Güçlük Düzeyi	Teşhir Olan Maddelerin Oranı	Teşhir Olan Maddelerin Tepki Süresi Değişimlemesi		
				Random olarak 10 ve 15 saniye arasındaki uniform dağılım	20 saniyeye sabitleme	30 saniyeye sabitleme
Madde tepkileri için KL uzaklığı	15 madde	Zor	%20	0.659	0.692	0.696
			%40	0.691	0.682	0.688
			%60	0.703	0.698	0.682
		Orta Güçlük Düzeyi	%20	0.758	0.841	0.852
			%40	0.872	0.899	0.902
			%60	0.931	0.932	0.938
	50 madde	Zor	%20	0.752	0.756	0.784
			%40	0.757	0.758	0.763
			%60	0.812	0.762	0.761
		Orta Güçlük Düzeyi	%20	0.869	0.885	0.887
			%40	0.940	0.947	0.948
			%60	0.948	0.952	0.954
$L(l_s^y)$ birey uyum istatistiği	15 madde	Zor	%20	0.114	0.121	0.091
			%40	0.132	0.134	0.118
			%60	0.134	0.101	0.107
		Orta Güçlük Düzeyi	%20	0.110	0.112	0.103
			%40	0.288	0.253	0.308
			%60	0.294	0.224	0.198
	50 madde	Zor	%20	0.279	0.279	0.246
			%40	0.364	0.322	0.392
			%60	0.304	0.312	0.374
		Orta Güçlük Düzeyi	%20	0.384	0.293	0.394
			%40	0.564	0.651	0.662
			%60	0.425	0.545	0.556

**EK 10. Koşullara Göre Uygun Olmayan Madde Tepki Sürelerini Belirlemede
Kullanılan Yöntemlerin Test Hilesi Belirleme Güç Oranları**

Tablo 8

Koşullara Göre Madde Tepki Süreleri İçin KL Uzaklığı ve $L(l^t)$ Birey Uyum İstatistiğinin Test Hilesi Belirleme Güç Oranları Ortalaması

Yöntemler	Test Uzunluğu	Teşhir Olan Maddelerin Güçlük Düzeyi	Teşhir Olan Maddelerin Oranı	Teşhir Olan Maddelerin Tepki Süresi Değişimlemesi		
				Random olarak 10 ve 15 saniye arasındaki uniform dağılım	20 saniyeye sabitleme	30 saniyeye sabitleme
Madde tepkileri için KL uzaklığı	15 madde	Zor	%20	0.601	0.529	0.512
			%40	0.556	0.636	0.539
			%60	0.547	0.658	0.615
		Orta Güçlük Düzeyi	%20	0.762	0.763	0.827
			%40	0.761	0.813	0.822
			%60	0.731	0.833	0.791
	50 madde	Zor	%20	0.547	0.577	0.638
			%40	0.543	0.573	0.578
			%60	0.636	0.564	0.565
		Orta Güçlük Düzeyi	%20	0.739	0.782	0.818
			%40	0.654	0.757	0.801
			%60	0.746	0.823	0.762
$L(l^t)$ birey uyum istatistiği	15 madde	Zor	%20	0.071	0.012	0.002
			%40	0.078	0.031	0.003
			%60	0.063	0.021	0.001
		Orta Güçlük Düzeyi	%20	0.219	0.019	0.007
			%40	0.331	0.017	0.005
			%60	0.263	0.002	0.001
	50 madde	Zor	%20	0.152	0.019	0.003
			%40	0.178	0.001	0.000
			%60	0.158	0.001	0.000
		Orta Güçlük Düzeyi	%20	0.536	0.009	0.002
			%40	0.684	0.002	0.000
			%60	0.569	0.001	0.000

BENZERLİK BİLDİRİMİ

“ Madde Ön-Bilgisi Kaynaklı Test Hilesinin Madde Tepki Süresi Modeli Kullanılarak Farklı Koşullar Altında 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	: 1483662279
Sayfa Sayısı	: 124
Sözcük Sayısı	: 30703
Karakter Sayısı	: 222150
Benzerlik Oranı	: %5
Savunma Tarihi	: 28.01.2021

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: Ebru BALTA

Tarih: 06.01.2021

İmza

ÖZGEÇMİŞ

Kişisel Bilgiler

Adı ve Soyadı : EBRU BALTA

Doğum Yeri/Tarihi :

E-Posta Adresi :

İş Deneyimi :

Unvan	Görev Yeri	Yıl
Araştırma Görevlisi	Ağrı İbrahim Çeçen Üniversitesi Eğitim Fakültesi/Eğitim Bilimleri Bölümü/ Eğitimde Ölçme ve Değerlendirme Anabilim Dalı (ÖYP)	2019- Devam Ediyor
Araştırma Görevlisi	Ankara Üniversitesi/ Eğitim Bilimleri Enstitüsü/ Eğitimde Ölçme ve Değerlendirme Anabilim Dalı	2016- 2019
Araştırma Görevlisi	Mersin Üniversitesi/ Eğitim Bilimleri Enstitüsü/ Eğitimde Ölçme ve Değerlendirme Anabilim Dalı	2013-2016
Araştırma Görevlisi	Selçuk Üniversitesi Eğitim Fakültesi/Ortaöğretim Fen ve Matematik Alanlar Eğitimi Bölümü/ Fizik Eğitimi Anabilim Dalı	2011- 2013

Akademik Bilgiler

Yabancı Diller: İngilizce,Almanca

Öğrenim Durumu:

Derece	Bölüm/Program	Üniversite	Yıl
Lisans	Ortaöğretim Fen ve Matematik Alanlar Eğitimi Bölümü	Hacettepe Üniversitesi(Tezsiz Yüksek Lisans)	2004-2010
Yüksek Lisans	Eğitimde Ölçme ve Değerlendirme	Mersin Üniversitesi	2013-2016
Doktora	Eğitimde Ölçme ve Değerlendirme	Ankara Üniversitesi	2016(Eylül)-2021(Şubat)

Uluslararası - ESCI kapsamındaki dergilerde yayımlanan tam makaleler:

1. Balta, E., & Sünbül Omür, S. (2017). An investigation of ordering test items differently depending on their difficulty level by differential item functioning. *Eurasian Journal of Educational Research*, 17(72), 23-42. doi: 10.14689/ejer.2017.72.2

Uluslararası hakemli dergilerde yayımlanan makaleler:

1. Gürçay, D., & Balta, E. (2013). The effect of Turkish students' motivational beliefs on their metacognitive self-regulation in Physics. *Asia-Pacific Forum on Science Learning and Teaching*, 14(2), 1-14. Retrieved from https://www.eduhk.hk/apfslt/download/v14_issue2_files/gurcay.pdf

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

1. Balta E., Uçar A.,& Dogan, C. D. (2020). *Detecting Performance of l^t Person Fit Index Based on The Response Time Model*. European Test Security Conference, Amersfood, The Netherlands.(COVID-19 pandemisi kaynaklı ertelendi.)
2. Uçar A.,& Balta E. (2019,June). *Comparison of Different Test Termination Rules in Terms of Measurement Precision in Computerized Adaptive Testing*. International Conference on Research in Teaching and Education ,Vienna,Austria. Abstract retrieved from <https://www.dpublication.com/wp-content/uploads/2019/06/25-3035-RTECONF.pdf>
3. Balta E., Ucar A., & Sahin A. (2019, June). *Detection of cheating behavior in online unproctored CATs via a validation test*. International Association for Computerized Adaptive Testing (IACAT) Conference, Minneapolis, MN, USA. Abstract retrieved from <http://iacat.org/sites/default/files/IACAT-Program-6June.pdf>
4. Balta E., & Omür Sünbül, S. (2018,May). *Güçlüklerine Göre Farklı Sıralanan Maddelerin Değişen Madde Fonksiyonlarının Cinsiyet Açısından İncelenmesi*. Vth International Eurasian Educational Research Congress, Antalya,Turkey. Abstract retrieved from <https://ejercongress.org/public/assets/images/B%C4%B0LD%C4%B0R%C4%B0%C3%96ZETLER%C4%B0.pdf>
5. Cokluk Bökeoglu O., Ucar A., & Balta E. (2018,May). *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. Abstract retrieved from <https://ejercongress.org/public/assets/images/B%C4%B0LD%C4%B0R%C4%B0%C3%96ZETLER%C4%B0.pdf>

6. Balta E.,Gürçay D.,Dogan O., (July,2012). *Investigation of the Effect of Students' Motivational Beliefs on their Metacognitive Self Regulation in Physics Courses*. The World Conference on Physics Education, İstanbul, Turkey. Abstract retrieved from http://www.wcpe2012.org/WCPE_Book_of_abstracts.pdf
7. Coskun, S., Uysal,M.,& Balta,E.(2012). Investigation of Pre-Service Teachers Attitudes Towards Argumentativeness. 3rd International Conference on New Trends in Education and Their Implications, Antalya,Turkey. Abstract retrieved from https://www.pegem.net/Akademi/kongrebildiri_detay.aspx?id=134838

Ulusal bilimsel toplantılarda sunulan ve bildiri kitaplarında basılan bildiriler:

1. Kozan S., Balta E., & Uzun B., (2015). *Diş Hekimliği Öğrencilerinin İletişim Becerilerinin Değerlendirilmesinde Genellenebilirlik Kuramının Uygulaması*. III. Sağlık Bilimlerinde Klinik ve İletişim Beceri Eğitimleri Kongresi, Ankara,Türkiye (Poster) .
<http://www.tipegitimi.org.tr/kibe2015/images/bildiriler.pdf>
2. Uçar A., Kozan, S., Balta, E., Uzun, B., & Odabaşı, O. (2015). *Diş Hekimliği Öğrencilerinin İletişim Becerilerinin Değerlendirilmesinde Kullanılacak Uygun Koşul Sayılarının Genellenebilirlik Teorisiyle Belirlenmesi*. III. Sağlık Bilimlerinde Klinik ve İletişim Beceri Eğitimleri Kongresi, Ankara,Türkiye.
<http://www.tipegitimi.org.tr/kibe2015/images/bildiriler.pdf>

Akademik İlgi Alanları

Test Güvenliği, Bireye Uyarlanmış Bilgisayarlı Test Uygulamaları, Madde Tepki Kuramı,Değişen Madde Fonksiyonu,Test Eşitleme

Bilgisayar Becerileri

Kullanılan Programlar: R, EduG, SPSS, ITEMAN, LISREL, SIMULCAT, WINGEN, BILOG, MULTILOG